Supplementary file to
"Heterogenous structural breaks in panel data models"

Ryo Okui* and Wendun Wang†

April 2019

This supplementary file contains additional theoretical, simulation, and empirical results that are omitted from the main text. We first extend the simulation experiments to better understand the performance of GAGFL in Section S.1. Then we present the empirical analysis of the democracy–income relationship with different methods and under alternative specifications in Section S.2, and an additional application on the determinants of cross-country difference in savings behavior in Section S.3. In Section S.4, we present the proofs for the main result that are not included in the main text. In Section S.5, we provide the theoretical analysis for models with individual specific fixed effects. Section S.6 discusses models in which a part of coefficients are fully time varying. Finally, Section S.7 considers the non-iterative estimator obtained by applying AGFL given the initial estimates of group membership structure.

S.1 Additional simulation studies

In this section, we present additional simulation results. There are five simulation exercises. First we consider the Bayesian information criterion (BIC) based on initial estimates for choosing the number of groups. Second, we examine cases with group-dependent regressors. Third, we consider designs with small groups and those with close break dates. Fourth, we

*Corresponding author.d Department of Economics and the Institute of Economic Research, Seoul National University Building 16, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, South Korea; and Department of Economics, University of Gothenburg, P.O. Box 640, SE-405 30 Gothenburg, Sweden. Email: okui.ryo.3@gmail.com

†Econometric Institute, Erasmus University Rotterdam, Burg. Oudlaan 50, 3062PA, Rotterdam; and Tinbergen Institute. Email: wang@ese.eur.nl

investigate the performance of GAGFL when group heterogeneity is small and when break sizes are small. Lastly, performances of initial estimates are evaluated.

S.1.1 Determining the number of groups based on initial estimates

We examine the performance of BIC based on the initial estimates. Although using the initial and final (iterative) estimates to compute the BIC are asymptotically equivalent, we argue that the potential disadvantage of using initial estimates is that less efficient coefficient estimates may result in less accurate selection of G in finite samples.

Table S.1 investigates the performance of BIC based on the initial estimates. All DGPs remain the same as in the paper. Compared with the BIC based on the final (iterative) estimates, using initial estimates indeed lowers the frequency of selecting the correct number of groups in almost all cases, although to a small extent, except in DGP.3 when $N = 50$ and $T = 10$. In DGP.3 with small samples, break estimation is less precise since data are first–differenced, and this further contaminates the determination of G. On the contrary, initial estimates that do not detect breaks lead to slightly more accurate selection of G.

S.1.2 Group-dependent regressors

We examine the cases where the regressor is group dependent, and generate x_{it} such that $x_{it} = f_{h_i} + \nu_{it}$, where f_h is drawn from a standard normal distribution for $h = 1, \ldots, H$ but common for units within a group, and ν_{it} is idiosyncratic and also follows a standard normal distribution for all i and t. We consider two subcases. First, f_h and $\beta_{g,t}$ share the same the group structure (i.e., $h_i = g_i, \forall i$). Second, the group structure of f_h differs from that of $\beta_{g,t}$. In the latter case, we generate f_h from four groups with $N_1 : N_2 : N_3 : N_4 = 0.2 : 0.2 : 0.2 : 0.4$, while the group structure of $\beta_{g,t}$ remains the same as specified in the paper (three groups with $N_1 : N_2 : N_3 = 0.3 : 0.3 : 0.4$). h_i and g_i are independent. To save space, we report the performance of clustering and break detection in two leading cases, DGP.1 and DGP.3.

Table S.2 presents the average misclustering rate when the regressor is characterized by group dependence. It shows that allowing for group dependent regressors does not affect clustering accuracy in DGP.1, and the misclustering rates are similar to those in the case of independent regressors. In DGP.3 with small T, the misclustering rates in the presence

Table S.1: Group number selection frequency using BIC based in initial estimates ($G_0 = 3$)

σ_ϵ	N	T	1	2	3	4	5	1	2	3	4	5
					DGP.1					DGP.2		
0.5	50	10	0.000	0.000	**0.976**	0.024	0.000	0.000	0.000	**0.996**	0.004	0.000
	50	20	0.000	0.000	**0.984**	0.016	0.000	0.000	0.000	**0.988**	0.012	0.000
	50	40	0.000	0.000	**0.984**	0.016	0.000	0.000	0.000	**0.992**	0.008	0.000
	100	10	0.000	0.000	**0.988**	0.012	0.000	0.000	0.000	**0.984**	0.012	0.004
	100	20	0.000	0.000	**0.996**	0.004	0.000	0.000	0.000	**0.976**	0.024	0.000
	100	40	0.000	0.000	**0.984**	0.016	0.000	0.000	0.000	**0.984**	0.016	0.000
0.75	50	10	0.000	0.008	**0.976**	0.016	0.000	0.000	0.004	**0.972**	0.024	0.000
	50	20	0.000	0.072	**0.924**	0.004	0.000	0.000	0.012	**0.952**	0.036	0.000
	50	40	0.000	0.060	**0.940**	0.000	0.000	0.000	0.012	**0.956**	0.032	0.000
	100	10	0.000	0.000	**0.996**	0.004	0.000	0.000	0.000	**0.992**	0.008	0.000
	100	20	0.000	0.000	**0.964**	0.036	0.000	0.000	0.000	**0.960**	0.040	0.000
	100	40	0.000	0.000	**0.968**	0.032	0.000	0.000	0.000	**0.984**	0.016	0.000
					DGP.3					DGP.4		
0.5	50	10	0.000	0.000	**0.998**	0.002	0.000	0.000	0.000	**0.922**	0.024	0.004
	50	20	0.000	0.002	**0.986**	0.012	0.000	0.000	0.000	**0.968**	0.032	0.000
	50	40	0.000	0.000	**0.990**	0.010	0.000	0.000	0.000	**0.984**	0.016	0.000
	100	10	0.000	0.000	**0.998**	0.002	0.000	0.000	0.000	**0.796**	0.180	0.024
	100	20	0.000	0.000	**0.996**	0.004	0.000	0.000	0.000	**0.952**	0.048	0.000
	100	40	0.000	0.000	**0.992**	0.008	0.000	0.000	0.000	**0.964**	0.032	0.004
0.75	50	10	0.000	0.056	**0.936**	0.008	0.000	0.000	0.000	**0.916**	0.084	0.000
	50	20	0.000	0.222	**0.778**	0.000	0.000	0.000	0.016	**0.904**	0.076	0.004
	50	40	0.000	0.158	**0.842**	0.000	0.000	0.000	0.040	**0.924**	0.032	0.004
	100	10	0.000	0.000	**0.996**	0.004	0.000	0.000	0.000	**0.720**	0.252	0.028
	100	20	0.000	0.000	**0.996**	0.004	0.000	0.000	0.000	**0.888**	0.104	0.008
	100	40	0.000	0.000	**0.998**	0.002	0.000	0.000	0.000	**0.944**	0.044	0.012

		$N=50$			$N=100$		
		$T=10$	$T=20$	$T=40$	$T=10$	$T=20$	$T=40$
		Same group structure as β					
DGP.1	$\sigma_\epsilon=0.5$	0.0064	0.0040	0.0010	0.0050	0.0008	0.0021
	$\sigma_\epsilon=0.75$	0.0228	0.0098	0.0014	0.0219	0.0117	0.0039
DGP.3	$\sigma_\epsilon=0.5$	0.1083	0.0375	0.0072	0.0986	0.0240	0.0078
	$\sigma_\epsilon=0.75$	0.1483	0.0696	0.0298	0.1422	0.0575	0.0175
		Different group structure from β					
DGP.1	$\sigma_\epsilon=0.5$	0.0053	0.0023	0.0010	0.0063	0.0017	0.0023
	$\sigma_\epsilon=0.75$	0.0208	0.0081	0.0036	0.0213	0.0109	0.0053
DGP.3	$\sigma_\epsilon=0.5$	0.1208	0.0389	0.0108	0.1166	0.0289	0.0112
	$\sigma_\epsilon=0.75$	0.1535	0.0768	0.0273	0.1585	0.0644	0.0151

Table S.2: Average misclustering rate: Regressors with group dependence

of group dependent regressors tend to be higher than those in the case of independent regressors. This is not surprising since first-differencing removes variation of regressors, and the lack of variation is more severe when regressors are cross-sectionally dependent and the sample size is small. Even in this difficult case with limited variation of regressors, our approach can still correctly classify roughly 90% of units in the smallest sample when $\sigma_\epsilon = 0.5$ and 85% when $\sigma_\epsilon = 0.75$. As T increases, the misclustering rate drops quickly, and reaches a similarly low level to that in the case of independent regressors when $T = 40$. These results hold no matter whether the group structure of regressors coincides that of coefficients.

Next, we examine the accuracy of break detection when regressors are group dependent. Table S.3 provides the average frequency of correct estimation of the number of breaks in the presence of group dependent regressors. The estimated numbers of breaks are similar to those in the case of independent regressors no matter whether the group structure of regressors coincides the structure of slope coefficients. Particularly, the estimated number of breaks is fairly accurate except in DGP.3 with large errors and small samples. When the number of breaks is not estimated correctly, it is typically overestimated, and the accuracy improves rapidly as either N or T increases. Table S.4 presents the Hausdorff error of break-date estimates when regressors are group dependent. Similarly, the break dates can be estimated accurately in almost all cases and the error is as small as in the

Table S.3: Average frequency of correct estimation of the number of breaks: Regressors with group dependence

	σ_ϵ	Group	$N=50$			$N=100$		
			$T=10$	$T=20$	$T=40$	$T=10$	$T=20$	$T=40$
			Same group structure as β					
DGP.1	0.5	G1 ($m_{1,0}^0=2$)	0.993	0.997	0.990	0.997	0.997	0.995
		G2 ($m_{2,0}^0=2$)	0.993	0.998	0.993	0.997	0.998	0.996
		G3 ($m_{3,0}^0=0$)	0.988	0.997	0.997	0.998	0.999	0.999
	0.75	G1 ($m_{1,0}^0=2$)	0.862	0.936	0.976	0.976	0.992	0.992
		G2 ($m_{2,0}^0=2$)	0.858	0.978	0.992	0.970	0.992	0.992
		G3 ($m_{3,0}^0=0$)	0.854	0.948	0.964	0.974	0.996	0.998
DGP.3	0.5	G1 ($m_{1,0}^0=2$)	0.630	0.847	0.959	0.774	0.924	0.982
		G2 ($m_{2,0}^0=2$)	0.644	0.889	0.974	0.814	0.966	0.994
		G3 ($m_{3,0}^0=0$)	0.716	0.868	0.933	0.880	0.954	0.974
	0.75	G1 ($m_{1,0}^0=2$)	0.260	0.378	0.618	0.382	0.658	0.876
		G2 ($m_{2,0}^0=2$)	0.162	0.404	0.756	0.342	0.712	0.936
		G3 ($m_{3,0}^0=0$)	0.292	0.445	0.608	0.454	0.710	0.890
			Different group structure from β					
DGP.1	0.5	G1 ($m_{1,0}^0=2$)	0.992	0.996	0.998	0.998	0.998	0.999
		G2 ($m_{2,0}^0=2$)	0.998	0.996	0.998	0.999	0.999	0.995
		G3 ($m_{3,0}^0=0$)	0.988	1.000	1.000	1.000	0.999	1.000
	0.75	G1 ($m_{1,0}^0=2$)	0.868	0.958	0.982	0.978	0.988	0.996
		G2 ($m_{2,0}^0=2$)	0.858	0.982	0.990	0.974	0.990	0.995
		G3 ($m_{3,0}^0=0$)	0.854	0.948	0.964	0.972	0.996	0.999
DGP.3	0.5	G1 ($m_{1,0}^0=2$)	0.596	0.854	0.951	0.738	0.946	0.978
		G2 ($m_{2,0}^0=2$)	0.664	0.902	0.989	0.774	0.957	0.996
		G3 ($m_{3,0}^0=0$)	0.738	0.878	0.929	0.902	0.957	0.962
	0.75	G1 ($m_{1,0}^0=2$)	0.220	0.362	0.636	0.582	0.664	0.860
		G2 ($m_{2,0}^0=2$)	0.176	0.384	0.750	0.598	0.710	0.940
		G3 ($m_{3,0}^0=0$)	0.314	0.442	0.602	0.670	0.748	0.882

Table S.4: Hausdorff error of break date estimates: Regressors with group dependence

			$N = 50$			$N = 100$		
	σ_ϵ	Group	$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$
			Same group structure as β					
DGP.1	0.5	G1 ($m^0_{1,0} = 2$)	0.0016	0.0016	0.0017	0.0004	0.0000	0.0005
		G2 ($m^0_{2,0} = 2$)	0.0026	0.0004	0.0005	0.0000	0.0001	0.0003
	0.75	G1 ($m^0_{1,0} = 2$)	0.0255	0.0126	0.0027	0.0046	0.0012	0.0021
		G2 ($m^0_{2,0} = 2$)	0.0222	0.0080	0.0062	0.0046	0.0014	0.0004
DGP.3	0.5	G1 ($m^0_{1,0} = 2$)	0.0511	0.0246	0.0075	0.0243	0.0112	0.0030
		G2 ($m^0_{2,0} = 2$)	0.0374	0.0187	0.0097	0.0158	0.0071	0.0044
	0.75	G1 ($m^0_{1,0} = 2$)	0.1121	0.1126	0.0655	0.0821	0.0478	0.0191
		G2 ($m^0_{2,0} = 2$)	0.0790	0.0731	0.0553	0.0526	0.0390	0.0142
			Different group structure from β					
DGP.1	0.5	G1 ($m^0_{1,0} = 2$)	0.0010	0.0010	0.0003	0.0006	0.0002	0.0008
		G2 ($m^0_{2,0} = 2$)	0.0016	0.0003	0.0003	0.0008	0.0000	0.0004
	0.75	G1 ($m^0_{1,0} = 2$)	0.0222	0.0085	0.0026	0.0026	0.0025	0.0007
		G2 ($m^0_{2,0} = 2$)	0.0228	0.0080	0.0066	0.0046	0.0011	0.0008
DGP.3	0.5	G1 ($m^0_{1,0} = 2$)	0.0511	0.0195	0.0097	0.0236	0.0051	0.0041
		G2 ($m^0_{2,0} = 2$)	0.0374	0.0158	0.0124	0.0118	0.0081	0.0072
	0.75	G1 ($m^0_{1,0} = 2$)	0.1121	0.1131	0.0651	0.0611	0.0463	0.0240
		G2 ($m^0_{2,0} = 2$)	0.0790	0.0705	0.0564	0.0448	0.0349	0.0158

Notes: HD ratios of GAGFL estimates for G3 (with no breaks) are not reported because they are all zeros.

case of independent regressors except in DGP.3 with large noise. The similar performance of clustering and break detection in the two cases of dependent and independent regressors jointly leads to a similar performance of coefficient estimates.

S.1.3 Small groups and close break dates

To better understand how cross-sectional variation plays a role in the performance of GAGFL, we consider two experiments: one with small groups and the other with closer break dates.

Table S.5: GAGFL clustering and break detection: $N_1 : N_2 : N_3 = 0.1 : 0.8 : 0.1$

	$N = 50$			$N = 100$		
	$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$
			DGP.1, $\sigma_\epsilon = 0.5$			
Misclustering rate	0.1615	0.1489	0.0778	0.1046	0.0970	0.0760
Freq. of correct estimation of number of breaks						
G_1 ($m^0_{1,0} = 2$)	0.5980	0.6340	0.5080	0.8960	0.9000	0.9160
G_2 ($m^0_{2,0} = 2$)	0.9880	0.9960	1.0000	0.9960	1.0000	1.0000
G_3 ($m^0_{3,0} = 0$)	0.4800	0.5960	0.7860	0.7340	0.7640	0.8280
Hausdorff error						
G_1 ($m^0_{1,0} = 2$)	0.0788	0.0691	0.0808	0.0190	0.0162	0.0148
G_2 ($m^0_{2,0} = 2$)	0.0002	0.0005	0.0000	0.0008	0.0000	0.0000
			DGP.3, $\sigma_\epsilon = 0.5$			
Misclustering rate	0.1872	0.1816	0.1133	0.1308	0.1283	0.1145
Freq. of correct estimation of number of breaks						
G_1 ($m^0_{1,0} = 2$)	0.2840	0.3120	0.3480	0.5320	0.6140	0.7020
G_2 ($m^0_{2,0} = 2$)	0.8820	0.9500	0.9760	0.9220	0.9920	0.9920
G_3 ($m^0_{3,0} = 0$)	0.0880	0.2320	0.5640	0.3660	0.5080	0.6820
Hausdorff error						
G_1 ($m^0_{1,0} = 2$)	0.1439	0.1407	0.1267	0.0826	0.0581	0.0465
G_2 ($m^0_{2,0} = 2$)	0.0088	0.0029	0.0000	0.0060	0.0001	0.0005

First, we consider the case where the number of individual units in each group follows $N_1 : N_2 : N_3 = 0.1 : 0.8 : 0.1$, and small groups that contain only a few units emerge.

Since coefficient estimation makes use of within-group cross-sectional variation, a small group (with a few units) is expected to result in less accurate coefficient estimates, and further inaccurate clustering and break detection. Tables S.5 summarizes the average misclustering rate, frequency of correct estimation of the number of breaks, and Hausdorff error of break-date estimates in DGP.1 and DGP.3 in the presence of small groups. As expected, the misclustering rates in both DGPs are higher than those in the cases of similarly sized groups ($N_1 : N_2 : N_3 = 0.3 : 0.3 : 0.4$). With the smallest sample $N = 50$ and $T = 10$, GAGFL can correctly classify roughly 84% of units in DGP.1 and 81% in DGP.3. Further examination reveals that the relatively poorer performance of clustering is due to misclassification of units from Groups 1 and 3, the two smallest groups. Increasing T helps improve classification accuracy. But unlike in the case of similarly sized groups, the accuracy also increases significantly by enlarging N because cross-sectional variation in the small groups is increased.

The number of breaks is also estimated less accurately for the small groups, but not for the large group. Particularly, GAGFL tends to overestimate the number of groups in Group 1 and 3, while it can still correctly estimate two breaks in Group 2. Compared with break-date estimates in the cases of similarly sized groups, the estimated break dates in the small group, namely Group 1, are less precise, but the estimates in the large group, namely Group 2, are more accurate due to a large cross-sectional sample. The conclusions from DGP.2 and DGP.4 are highly similar and thus not reported.

Next, we study the performance of GAGFL when the break dates are close. Again, we demonstrate this situation using DGP.1 and DGP.3, but change the break dates. The breaks in the first group occur at $\lfloor T/2 \rfloor$ and $\lfloor 2T/3 \rfloor$, and in the second group at $\lfloor T/3 \rfloor$ and $\lfloor T/2 \rfloor$, where $\lfloor \cdot \rfloor$ takes the integer part. Now the difference between the two breaks in both groups is just $\lfloor T/6 \rfloor$, i.e. 1 when $T = 10$, 3 when $T = 20$, and 6 when $T = 40$. For the third group, the slope coefficient is stable without a break.

Table S.6 evaluates the performance of GAGFL in DGP.1 and DGP.3 with closer break dates, again based on its misclustering rate, frequency of correct estimation of the number of breaks, and Hausdorff error of break date estimates. It shows that the misclustering rate and the accuracy of break estimation are hardly affected by shrinking the interval between two breaks. This is because we make use of cross-sectional variation for coefficient estimation. As long as there are sufficiently many individual units in each group, we can consistently estimate slope coefficients (and further the groups and breaks) even when the two breaks are consecutive.

Table S.6: GAGFL clustering and break detection: Close break dates

	$N = 50$			$N = 100$		
	$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$
			DGP.1, $\sigma_\epsilon = 0.5$			
Misclustering rate	0.0072	0.0052	0.0047	0.0081	0.0021	0.0010
Freq. of correct estimation of number of breaks						
G_1 $(m^0_{1,0} = 2)$	0.9760	0.9840	0.9820	0.9960	0.9960	0.9980
G_2 $(m^0_{2,0} = 2)$	0.9880	0.9940	0.9920	0.9960	0.9980	0.9940
G_3 $(m^0_{3,0} = 0)$	0.9900	0.9820	0.9900	1.0000	1.0000	0.9980
Hausdorff error						
G_1 $(m^0_{1,0} = 2)$	0.0052	0.0026	0.0041	0.0010	0.0007	0.0008
G_2 $(m^0_{2,0} = 2)$	0.0020	0.0031	0.0032	0.0004	0.0000	0.0006
			DGP.3, $\sigma_\epsilon = 0.5$			
Misclustering rate	0.0119	0.0004	0.0000	0.0095	0.0005	0.0000
Freq. of correct estimation of number of breaks						
G_1 $(m^0_{1,0} = 2)$	0.8080	0.9120	0.9620	0.9520	0.9820	1.0000
G_2 $(m^0_{2,0} = 2)$	0.8180	0.9040	0.9760	0.9620	0.9940	0.9980
G_3 $(m^0_{3,0} = 0)$	0.7560	0.9280	0.9960	0.9240	1.0000	1.0000
Hausdorff error						
G_1 $(m^0_{1,0} = 2)$	0.0335	0.0136	0.0036	0.0082	0.0037	0.0000
G_2 $(m^0_{2,0} = 2)$	0.0405	0.0158	0.0031	0.0084	0.0015	0.0001

S.1.4 Small degree of group heterogeneity and break size

Now we consider the case where the breaks are small and groups are more alike. We generate the data with the same specifications as in the paper but with different values of parameters as follows:

$$\beta_{1,t} = \begin{cases} 1 & \text{if } 1 \le t < \lfloor T/2 \rfloor \\ 1.5 & \text{if } \lfloor T/2 \rfloor \le t < \lfloor 5T/6 \rfloor \\ 2 & \text{if } \lfloor 5T/6 \rfloor \le t \le T \end{cases}, \quad \beta_{2,t} = \begin{cases} 2 & \text{if } 1 \le t < \lfloor T/3 \rfloor \\ 2.5 & \text{if } \lfloor T/3 \rfloor \le t < \lfloor 5T/6 \rfloor \\ 3 & \text{if } \lfloor 5T/6 \rfloor \le t \le T \end{cases},$$

and

$$\beta_{3,t} = 1.5 \quad \text{for all } 1 \le t \le T.$$

Table S.7 evaluates the performance of GAGFL in DGP.1 and DGP.3 but with small breaks and group heterogeneity. The misclustering rate is roughly 12% in DGP.1 and 15% in DGP.3 when $N = 50$ and $T = 10$, both higher than the cases with well-separated coefficients reported in the paper. But this rate quickly decreases to less than 1.5% when T increases to 40. Small breaks also affect break detection to some extent. When $N = 50$ and $T = 10$, the frequency of correct estimation of the number of breaks is roughly 30% less in DGP.1 and 38% less in DGP.3 than in the cases with large breaks as reported in the paper. But again, the accuracy improves quickly as N or T increases. When $N = 100$, the frequency of correct estimation of the number of breaks is always more than 0.94. Similar observations are found for the break date estimates. Small breaks and group heterogeneity lead to less precise estimates of break dates, but the precision improves quickly as the sample size increases.

S.1.5 Performance of initial estimates

Finally, we discuss the iterative feature of the algorithm. In particular, we examine the number of iterations for the algorithm to converge, and compare the final (iterative) estimates with the non-iterative estimates, say $\dot{\gamma}$ and $\beta^{(0)}$, defined in Alogrithm 1.

Table S.8 reports the average number of iterations for Algorithm 1 to converge. On average, the algorithm takes 2-3 steps to converge in most cases. We find that more iterations are needed in DGP.3, since accurate estimation is more difficult for the first differenced data. The number of iterations also increases when the variance of the error is

Table S.7: GAGFL clustering and break detection: Smaller group heterogeneity and break sizes

	$N = 50$			$N = 100$		
	$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$
			DGP.1, $\sigma_\epsilon = 0.5$			
Misclustering rate	0.1226	0.0433	0.0150	0.0903	0.0396	0.0103
Freq. of correct estimation of numer of breaks						
G_1 ($m^0_{1,0} = 2$)	0.7040	0.8360	0.7980	0.9440	0.9660	0.9640
G_2 ($m^0_{2,0} = 2$)	0.8060	0.8500	0.7720	0.9820	0.9780	0.9660
G_3 ($m^0_{3,0} = 0$)	0.8800	0.9780	0.9900	0.9800	0.9900	0.9880
Hausdorff error						
G_1 ($m^0_{1,0} = 2$)	0.0452	0.0376	0.0377	0.0144	0.0076	0.0075
G_2 ($m^0_{2,0} = 2$)	0.0229	0.0282	0.0434	0.0046	0.0039	0.0058
			DGP.3, $\sigma_\epsilon = 0.5$			
Misclustering rate	0.1569	0.0570	0.0115	0.1346	0.0440	0.0078
Freq. of correct estimation of numer of breaks						
G_1 ($m^0_{1,0} = 2$)	0.4803	0.5640	0.4780	0.7640	0.8300	0.8160
G_2 ($m^0_{2,0} = 2$)	0.6352	0.6300	0.4440	0.8780	0.8760	0.7720
G_3 ($m^0_{3,0} = 0$)	0.4803	0.9020	0.9900	0.8020	0.9840	1.0000
Hausdorff error						
G_1 ($m^0_{1,0} = 2$)	0.0960	0.0801	0.0860	0.0506	0.0320	0.0258
G_2 ($m^0_{2,0} = 2$)	0.0729	0.0749	0.1011	0.0245	0.0205	0.0364

Table S.8: Average number of iterations

		$N = 50$			$N = 100$		
		$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$
DGP.1	$\sigma_\epsilon = 0.5$	2.204	2.118	2.042	2.288	2.110	2.058
	$\sigma_\epsilon = 0.75$	2.750	2.740	2.256	2.796	2.594	2.170
DGP.2	$\sigma_\epsilon = 0.5$	2.132	2.062	2.036	2.120	2.034	2.090
	$\sigma_\epsilon = 0.75$	2.464	2.338	2.092	2.542	2.298	2.066
DGP.3	$\sigma_\epsilon = 0.5$	2.425	2.170	2.024	2.460	2.151	2.017
	$\sigma_\epsilon = 0.75$	2.744	2.755	2.305	2.941	2.744	2.192
DGP.4	$\sigma_\epsilon = 0.5$	2.264	2.068	2.024	2.256	2.024	2.020
	$\sigma_\epsilon = 0.75$	2.868	2.624	2.200	2.704	2.312	2.020

large, but decreases as T grows. Moreover, we find that it takes more steps to converge when we generate the data with closer break dates, small degree of group heterogeneity and breaks, and small groups containing only very few units. To summarize, the iterative algorithm is especially useful in finite samples when the clustering and break detection is difficult due to, for example, high serial correlation, large noise, and more alike coefficients.

To compare the performance of non-iterative estimates with the iterative ones, we first report the misclustering rate of the non-iterative estimates, namely $\dot{\gamma}$ in Table S.9. For comparison convenience, we also copy the misclustering rate of the iterative estimators $\hat{\gamma}$ from the paper. We find the misclustering rate of iterative estimates consistently lower than the rate of non-iterative estimates, suggesting that the clustering performance is improved by iteration, sometimes to a large extent.

Next, we report the RMSE and the coverage probability of the non-iterative coefficient estimates $\beta^{(0)}$ in Table S.10. Again, we also list the statistics of the iterative version $\hat{\beta}$ for the convenience of comparison. In general, we find that iterative estimates produce lower RMSEs but higher coverage probability than non-iterative ones. In DGP.1, the difference between the RMSEs of the two versions of estimates enlarges as T increases, and can reach 70% when $\sigma_\varepsilon = 0.75$. In DGP.2, the difference is smaller but can still reach more than 30% in some cases. In DGP.3, the difference in the RMSE is smaller than in DGP.1, but the coverage probability of iterative estimates is much higher than that of non-iterative estimates, especially when the T and σ_ϵ are large. Similar superiority of iterative estimates is found in most cases of DGP.4.

Table S.9: Average misclustering rate: Comparison of non-iterative and iterative estimation

		$N = 50$			$N = 100$		
		$T = 10$	$T = 20$	$T = 40$	$T = 10$	$T = 20$	$T = 40$
		Non-iterative estimates					
DGP.1	$\sigma_\epsilon = 0.5$	0.0145	0.0184	0.0072	0.0215	0.0123	0.0157
	$\sigma_\epsilon = 0.75$	0.0567	0.0316	0.0190	0.0411	0.0285	0.0174
DGP.2	$\sigma_\epsilon = 0.5$	0.0099	0.0068	0.0075	0.0036	0.0084	0.0278
	$\sigma_\epsilon = 0.75$	0.0370	0.0207	0.0221	0.0320	0.0213	0.0117
DGP.3	$\sigma_\epsilon = 0.5$	0.0208	0.0043	0.0013	0.0167	0.0033	0.0030
	$\sigma_\epsilon = 0.75$	0.0781	0.0342	0.0150	0.0487	0.0273	0.0033
DGP.4	$\sigma_\epsilon = 0.5$	0.0093	0.0028	0.0034	0.0060	0.0019	0.0040
	$\sigma_\epsilon = 0.75$	0.0449	0.0198	0.0094	0.0348	0.0069	0.0012
		Iterative estimates					
DGP.1	$\sigma_\epsilon = 0.5$	0.0104	0.0026	0.0010	0.0097	0.0015	0.0000
	$\sigma_\epsilon = 0.75$	0.0448	0.0177	0.0027	0.0377	0.0140	0.0042
DGP.2	$\sigma_\epsilon = 0.5$	0.0048	0.0025	0.0010	0.0040	0.0022	0.0001
	$\sigma_\epsilon = 0.75$	0.0296	0.0076	0.0028	0.0206	0.0081	0.0042
DGP.3	$\sigma_\epsilon = 0.5$	0.0179	0.0024	0.0001	0.0171	0.0028	0.0020
	$\sigma_\epsilon = 0.75$	0.0663	0.0240	0.0041	0.0484	0.0161	0.0013
DGP.4	$\sigma_\epsilon = 0.5$	0.0074	0.0005	0.0024	0.0059	0.0004	0.0040
	$\sigma_\epsilon = 0.75$	0.0357	0.0114	0.0016	0.0327	0.0046	0.0002

Table S.10: RMSE and coverage probability of coefficient estimates: Comparison between Non-iterative and iterative estimation

				RMSE		Coverage	
	σ	N	T	Non-iterative	Iterative	Non-iterative	Iterative
DGP.1	0.5	50	10	0.1232	0.1161	0.9224	0.9237
		50	20	0.0961	0.0611	0.9196	0.9349
		50	40	0.0574	0.0388	0.9438	0.9477
		100	10	0.1179	0.1022	0.9146	0.9265
		100	20	0.0650	0.0493	0.9341	0.9408
		100	40	0.0536	0.0429	0.9295	0.9394
	0.75	50	10	0.2468	0.2347	0.8275	0.8396
		50	20	0.1720	0.1612	0.8816	0.8940
		50	40	0.2078	0.0771	0.9164	0.9401
		100	10	0.1983	0.1916	0.8794	0.8860
		100	20	0.1129	0.1051	0.9175	0.9267
		100	40	0.1174	0.0467	0.9268	0.9406
DGP.2	0.5	50	10	0.1071	0.0787	0.9205	0.9276
		50	20	0.0578	0.0435	0.9327	0.9441
		50	40	0.0439	0.0270	0.9373	0.9435
		100	10	0.0651	0.0708	0.9355	0.9367
		100	20	0.0547	0.0347	0.9365	0.9444
		100	40	0.0456	0.0428	0.9343	0.9438
	0.75	50	10	0.1860	0.1725	0.8703	0.8826
		50	20	0.1104	0.0892	0.9202	0.9333
		50	40	0.0811	0.0718	0.9160	0.9330
		100	10	0.1517	0.1441	0.9049	0.9152
		100	20	0.1254	0.0732	0.9310	0.9412
		100	40	0.0521	0.0321	0.9313	0.9417

Table S.10 (cont.): RMSE and coverage probability of coefficient estimates: Comparison between Non-iterative and iterative estimation

				RMSE		Coverage	
	σ	N	T	Non-iterative	Iterative	Non-iterative	Iterative
DGP.3	0.5	50	10	0.1681	0.1588	0.8357	0.8444
		50	20	0.0796	0.0747	0.8912	0.8919
		50	40	0.0463	0.0438	0.9024	0.9175
		100	10	0.1336	0.1339	0.8879	0.8870
		100	20	0.0578	0.0573	0.9200	0.9298
		100	40	0.0341	0.0301	0.9225	0.9319
	0.75	50	10	0.3357	0.3212	0.7156	0.7375
		50	20	0.2150	0.2010	0.7838	0.7744
		50	40	0.1243	0.1027	0.8308	0.8755
		100	10	0.2492	0.2428	0.7981	0.7987
		100	20	0.1373	0.1294	0.8526	0.8806
		100	40	0.0632	0.0564	0.9016	0.9150
DGP.4	0.5	50	10	0.4933	0.0478	0.9037	0.9140
		50	20	0.0611	0.0426	0.9354	0.9416
		50	40	0.0338	0.0329	0.9464	0.9478
		100	10	0.0406	0.0317	0.9199	0.9312
		100	20	0.0176	0.0156	0.9368	0.9386
		100	40	0.0097	0.0096	0.9443	0.9482
	0.75	50	10	0.0831	0.0766	0.8653	0.8771
		50	20	0.0468	0.0391	0.9019	0.9060
		50	40	0.0403	0.0368	0.9257	0.9327
		100	10	0.0587	0.0552	0.9034	0.9107
		100	20	0.0316	0.0299	0.9340	0.9444
		100	40	0.0157	0.0178	0.9440	0.9435

S.2 Additional empirical analysis in democracy–income application

In this section, we present additional results for the democracy–income application whose main results are given in Section 8 of the main text. In particular, we provide the results of the initial estimates, estimates with a larger number of groups, and the results from an alternative specification in which the intercepts (group fixed effects) are fully time varying. We also present confidence sets for group membership.

S.2.1 Analysis with initial estimates

We examine the democracy–income relation based on the initial estimators defined in (3). The BIC computed from the initial estimates again selects four groups, in line with the choice based on the iterative estimators. Thus we analyze the democracy–income relation using the initial estimator for $G = 4$.

Table S.11 lists the group memberships of all countries based on initial and final (iterative) estimators. The group structure produced by the initial estimators resembles but not precisely coincides the iterative estimators. Particularly, 15 countries switch group memberships during iteration, most of which are Latin American and African countries, such as Bolivia, Colombia, Costa Rica, South Africa, Tunisia, Uganda, Venezuela, and Zambia.

We then examine the initial coefficient estimates that vary at each time period as presented in Table S.12. The initial estimates are generally in line with the iterative ones, although sometimes less efficient. Particularly, the estimated coefficients in Group 1 are indeed highly stable in each period for all variables. Although the lagged income effect turns negative in the last period, different from the estimates in the previous periods, the negative coefficient is highly insignificant. Pooling stable coefficients along the time (as reported in Table 8 of the paper) improves the efficiency of the estimates. The fully time varying coefficient estimates in Group 2 exhibit a clear structural break in the second period, where the lag and income effects both turn from negative to strongly positive. Again, the iterative estimates that pool periods 2–7 seem more efficient than the initial estimates. In Group 3, the initial estimate of the intercept exhibits an obvious change in the 5th period, explaining the break detected by GAGFL that occurs in the same period. The average democracy and income effect in Group 4 seem rather volatile, and this explains the two breaks reported by GAGFL.

Table S.11: Initial and iterative estimates of group memberships

	Init.	Iter.		Init.	Iter.		Init.	Iter.
Algeria	1	1	Ghana	3	3	Nigeria	3	3
Argentina	3	3	Greece	2	2	Norway	1	1
Australia	1	1	Guatemala	4	3	Panama	2	2
Austria	1	1	Guinea	1	1	Paraguay	1	1
Belgium	1	1	Honduras	2	2	Peru	2	2
Benin	4	4	Iceland	1	1	Philippines	4	4
Bolivia	4	4	India	1	1	Portugal	2	2
Brazil	3	3	Indonesia	2	2	Romania	1	1
Burkina Faso	2	3	Iran	1	1	Rwanda	1	2
Burundi	1	2	Ireland	1	1	Sierra Leone	2	3
Cameroon	1	1	Israel	1	1	Singapore	1	1
Canada	1	1	Italy	1	1	South Africa	3	3
Central African Rep.	4	4	Jamaica	1	1	Spain	2	2
Chad	1	1	Japan	1	1	Sri Lanka	1	1
Chile	3	3	Jordan	1	1	Sweden	1	1
China	1	1	Kenya	1	1	Switzerland	1	1
Colombia	1	1	Korea, Rep.	3	3	Syrian Arab Rep.	1	1
Congo, Dem. Rep.	1	2	Luxembourg	1	1	Taiwan	2	2
Congo, Rep.	4	2	Madagascar	4	4	Tanzania	2	3
Costa Rica	1	1	Malawi	4	4	Thailand	3	2
Cote d'Ivoire	1	1	Malaysia	1	1	Togo	1	2
Cyprus	3	1	Mali	4	4	Trinidad & Tobago	1	1
Denmark	1	1	Mauritania	1	1	Tunisia	1	1
Dominican Rep.	2	1	Mexico	2	3	Turkey	4	4
Ecuador	2	2	Morocco	1	1	Uganda	2	2
Egypt, Arab Rep.	1	1	Nepal	2	2	United Kingdom	1	1
El Salvador	4	4	Netherlands	1	1	United States	1	1
Finland	1	1	New Zealand	1	1	Uruguay	3	4
France	1	1	Nicaragua	3	3	Venezuela, RB	1	1
Gabon	1	1	Niger	4	4	Zambia	4	3

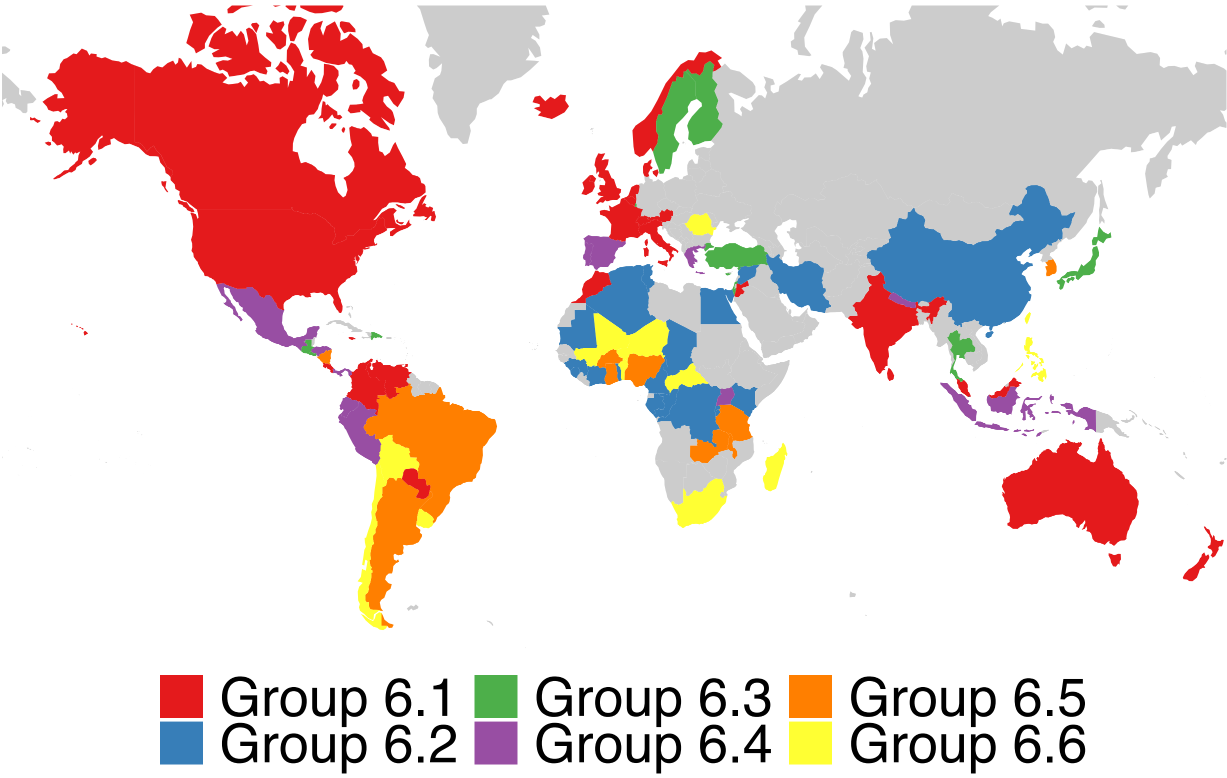
Table S.12: Income and democracy: Initial coefficient estimates of $G = 4$

	Regime	1	2	3	4	5	6	7
Group 1	Intercept	−0.3813	−0.0586	0.0799	−0.0346	−0.0471	−0.1451	0.0690
		(0.0490)	(0.0373)	(0.0387)	(0.0398)	(0.0409)	(0.0428)	(0.0445)
	Democracy$_{t-1}$	1.0802	0.9764	0.8792	0.8672	0.7715	0.7680	1.0161
		(0.0721)	(0.0548)	(0.0531)	(0.0549)	(0.0578)	(0.0670)	(0.0634)
	Income$_{t-1}$	0.2676	0.0056	0.0554	0.1431	0.2032	0.2167	−0.0837
		(0.0614)	(0.0656)	(0.0626)	(0.0610)	(0.0628)	(0.0680)	(0.0619)
Group 2	Intercept	−0.9062	−0.2692	0.2832	−0.0312	0.1068	−0.0239	0.5262
		(0.1344)	(0.1914)	(0.1792)	(0.1135)	(0.1084)	(0.1104)	(0.1043)
	Democracy$_{t-1}$	−0.3515	0.4194	0.0685	1.2802	0.5904	0.7961	0.1464
		(0.1435)	(0.1700)	(0.2032)	(0.2125)	(0.1361)	(0.1863)	(0.1559)
	Income$_{t-1}$	−0.0106	0.3704	0.5406	−0.1158	0.4076	0.0258	0.4566
		(0.1330)	(0.1209)	(0.1314)	(0.1485)	(0.1316)	(0.1637)	(0.1300)
Group 3	Intercept	−0.3595	−0.5152	−0.1504	−0.1789	0.5102	0.3272	0.5464
		(0.1440)	(0.1210)	(0.1369)	(0.1230)	(0.1297)	(0.1370)	(0.1415)
	Democracy$_{t-1}$	0.9587	−0.7031	0.2391	0.8500	0.5936	−0.2558	0.4772
		(0.2191)	(0.1647)	(0.1787)	(0.2679)	(0.2024)	(0.2216)	(0.3388)
	Income$_{t-1}$	0.1070	1.0474	−0.9453	1.5832	0.2060	1.1095	−0.0219
		(0.2270)	(0.2334)	(0.2060)	(0.2560)	(0.2655)	(0.2341)	(0.2968)
Group 3	Intercept	0.2513	−0.2064	−0.8996	0.5646	0.3242	−0.1536	0.2243
		(0.1812)	(0.1793)	(0.1625)	(0.2992)	(0.1668)	(0.1855)	(0.1582)
	Democracy$_{t-1}$	0.4868	0.7959	0.3501	0.4393	0.5674	0.4604	1.4734
		(0.1605)	(0.2012)	(0.2016)	(0.2838)	(0.1674)	(0.2281)	(0.2800)
	Income$_{t-1}$	1.0548	−0.0027	−0.0633	0.9766	0.5050	−0.9039	0.6153
		(0.1828)	(0.2902)	(0.2700)	(0.1906)	(0.2420)	(0.2839)	(0.1778)

Table S.13: Income and democracy: Coefficient and regime estimates of $G = 4$ after first few iteration

<table>
<tr><th></th><th>Regime</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th></tr>
<tr><td colspan="9">Estimates after first iteration</td></tr>
<tr><td>Group 4.1</td><td>Intercept</td><td colspan="7">−0.0728 (0.0205)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td colspan="7">0.8504 (0.0274)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td colspan="7">0.1572 (0.0320)</td></tr>
<tr><td>Group 4.2</td><td>Intercept</td><td>−0.9743 (0.1253)</td><td colspan="6">0.1974 (0.0518)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td>−0.2224 (0.1622)</td><td colspan="6">0.4809 (0.0786)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td>0.1509 (0.0884)</td><td colspan="6">0.4028 (0.0745)</td></tr>
<tr><td>Group 4.3</td><td>Intercept</td><td colspan="4">−0.3501 (0.0943)</td><td colspan="3">0.3360 (0.0721)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td colspan="4">−0.1256 (0.1544)</td><td colspan="3">0.3000 (0.1428)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td colspan="4">0.3589 (0.0916)</td><td colspan="3">0.3866 (0.1403)</td></tr>
<tr><td>Group 4.4</td><td>Intercept</td><td colspan="3">−0.3813 (0.2223)</td><td colspan="2">0.5879 (0.1650)</td><td colspan="2">0.2623 (0.1980)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td colspan="3">0.4854 (0.1173)</td><td colspan="2">0.3651 (0.1954)</td><td colspan="2">0.0440 (0.0890)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td colspan="3">0.3407 (0.1736)</td><td colspan="2">0.9277 (0.2378)</td><td colspan="2">−0.2385 (0.1582)</td></tr>
<tr><td colspan="9">Estimates after second iteration</td></tr>
<tr><td>Group 4.1</td><td>Intercept</td><td colspan="7">−0.0651 (0.0219)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td colspan="7">0.8617 (0.0268)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td colspan="7">0.1442 (0.0315)</td></tr>
<tr><td>Group 4.2</td><td>Intercept</td><td>−1.1311 (0.0755)</td><td colspan="6">0.1794 (0.0484)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td>−0.3269 (0.0947)</td><td colspan="6">0.5123 (0.0779)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td>0.2448 (0.0663)</td><td colspan="6">0.4056 (0.0748)</td></tr>
<tr><td>Group 4.3</td><td>Intercept</td><td colspan="4">−0.3551 (0.0979)</td><td colspan="3">0.2936 (0.0735)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td colspan="4">−0.1309 (0.1579)</td><td colspan="3">0.2699 (0.1479)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td colspan="4">0.3545 (0.0901)</td><td colspan="3">0.4284 (0.1444)</td></tr>
<tr><td>Group 4.4</td><td>Intercept</td><td colspan="3">−0.3722 (0.1943)</td><td colspan="2">0.5531 (0.1375)</td><td colspan="2">0.3421 (0.1818)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td colspan="3">0.4718 (0.0957)</td><td colspan="2">0.3758 (0.1951)</td><td colspan="2">0.0640 (0.0871)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td colspan="3">0.3574 (0.1650)</td><td colspan="2">0.8917 (0.2323)</td><td colspan="2">−0.1912 (0.1568)</td></tr>
</table>

Figure S.1: Estimates of group membership ($G = 6$)



S.2.2 Democracy and income: Analysis under $G = 6$

In this section, we analyze the democracy–income relationship using GAGFL and discuss the results when we set $G = 6$. In Section 7.1 of the main text, we examines the same application but mainly discuss the results under four groups. Setting $G = 4$ corresponds to the minimum value of BIC, and it leads to a large group containing both democratic and autocratic countries. Given the short time dimension of the data set, the number of groups might be incorrectly estimated and we examine how our results are sensitive to a different choice of the number of groups here. Specifically, we set $G = 6$ which corresponds the second least value of the BIC. We are particularly interested in how the clustering results would be affected if we allow for more heterogeneity by specifying a larger G. The data set and the model are explained in Section 8.1.

Figure S.1 illustrates the clustering results and Table S.14 summarizes the estimated

Table S.14: Income and democracy: Coefficient and regime estimates of $G = 6$

<table>
<tr><th></th><th>Regime</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th></tr>
<tr><td>Group 6.1</td><td>Intercept</td><td colspan="7">0.0155 (0.0394)</td></tr>
<tr><td></td><td>Democracy_{t-1}</td><td colspan="7">0.8852 (0.0417)</td></tr>
<tr><td></td><td>Income_{t-1}</td><td colspan="7">0.0651 (0.0283)</td></tr>
<tr><td>Group 6.2</td><td>Intercept</td><td colspan="7">−0.7344 (0.0606)</td></tr>
<tr><td></td><td>Democracy_{t-1}</td><td colspan="7">0.2231 (0.0631)</td></tr>
<tr><td></td><td>Income_{t-1}</td><td colspan="7">0.1794 (0.0406)</td></tr>
<tr><td>Group 6.3</td><td>Intercept</td><td colspan="7">0.3371 (0.1015)</td></tr>
<tr><td></td><td>Democracy_{t-1}</td><td colspan="7">0.0446 (0.1453)</td></tr>
<tr><td></td><td>Income_{t-1}</td><td colspan="7">0.4971 (0.0944)</td></tr>
<tr><td>Group 6.4</td><td>Intercept</td><td>−0.9990 (0.1011)</td><td colspan="6">0.2142 (0.0594)</td></tr>
<tr><td></td><td>Democracy_{t-1}</td><td>−0.3562 (0.0938)</td><td colspan="6">0.4964 (0.0927)</td></tr>
<tr><td></td><td>Income_{t-1}</td><td>0.2428 (0.0735)</td><td colspan="6">0.3169 (0.0776)</td></tr>
<tr><td>Group 6.5</td><td>Intercept</td><td colspan="4">−0.5105 (0.1029)</td><td>0.3600 (0.0512)</td><td colspan="2">0.4103 (0.1126)</td></tr>
<tr><td></td><td>Democracy_{t-1}</td><td colspan="4">−0.4437 (0.1324)</td><td>−0.0379 (0.1881)</td><td colspan="2">0.0265 (0.1889)</td></tr>
<tr><td></td><td>Income_{t-1}</td><td colspan="4">0.4635 (0.0816)</td><td>1.0259 (0.1598)</td><td colspan="2">0.3670 (0.1304)</td></tr>
<tr><td>Group 6.6</td><td>Intercept</td><td colspan="3">−0.5148 (0.1527)</td><td colspan="4">0.3940 (0.0761)</td></tr>
<tr><td></td><td>Democracy_{t-1}</td><td colspan="3">0.4324 (0.1320)</td><td colspan="4">0.6841 (0.0875)</td></tr>
<tr><td></td><td>Income_{t-1}</td><td colspan="3">0.3296 (0.0954)</td><td colspan="4">0.1085 (0.1046)</td></tr>
</table>

values of coefficients and their standard errors. We denote the six groups in this case as Groups 6.1–6.6. As discussed in the main text, the group with stable coefficients under $G = 4$ (i.e., Group 4.1) is further divided when we set $G = 6$. Recall that Group 4.1 contains countries with different levels of democracy. When we set $G = 6$, poorly and highly democratic countries are sharply separated into three groups, which we label as Groups 6.1–6.3. For example, US and China belong to different groups. Nonetheless, their slope coefficients are all stable over time.

Countries in Group 6.1 are all highly democratic with stable political system, except for Jordan, Morocco, and Paraguay with moderate but still persistent democracy level. The persistency in democracy is demonstrated by a high value of the coefficient on the lagged dependent variable. The income level in this group varies. Most of countries are characterized by high income, such as a number of European countries, US, and Canada, while there are a few countries with high democracy but moderately low income, for example, Colombia, Costa Rica, India, Jamaica, and Malaysia. Hence, the association between income and democracy of this group is not strong.

Group 6.2 is also characterized by stable slope coefficients, but with much lower average level of democracy (estimated intercept being –0.7344). This group contains countries with low democracy level, such as a large share of African countries, Middle Eastern countries, and several autocratic Asian countries. These countries are generally poor. They have a higher income effect than Group 6.1, but the effect is still small.

Group 6.3 contains eleven countries. All of them are democratic but their Freedom House score fluctuates over time to different extents. We also observe that the income level of countries in this group varies jointly with their democracy level, such that the relation between income and democracy remains stable. These features are well embodied in the weak persistence of democracy and strong income effect.

It may be argued that Groups 6.4—6.6 here correspond to Groups 4.2—4.4, respectively. Particularly, 91% (10 out of 11) of the countries in Group 6.4 correspond to Group 4.2, and they are characterized by one structural break in the early part of the period. Their democracy level increases and becomes more stable after the break. Group 6.5 contains almost identical members as Group 4.3, whose democracy transition happens in the later part of the period. The detection of two breaks (one in the middle and one in the late), in contrast to one break found under $G = 4$, is possibly due to fluctuation of democracy and income during transition. Group 6.6 consists of 12 countries, and is largely similar to Group 4.4, with 8 countries in common. This group experiences one structural break in the middle of the period, and it differs from Group 6.5 in that the dynamic persistence of democracy is positive before the break and becomes even stronger after the break, while Group 6.5 has a negative dynamic effect in the first part of the period. Besides, the income effect is much weaker after the break in Group 6.6 than in Group 6.5.

We conclude that our main results hold under a different specification of G qualitatively. There exists a substantial degree of heterogeneity in the income effect on democracy. Moreover, there are heterogeneous structural breaks. There are countries who did not experience structural breaks, while there are also countries that experienced breaks in the effect. Even among those countries that exhibit breaks, the timings and the magnitudes are markedly different.

S.2.3 Specification with fully time varying group fixed effects

We consider the setup where $\alpha_{g_i,t}$ is not penalized, and thus varies at each time period. In this case, the break detection is purely based on slope coefficients. The estimates can be obtained by minimizing the following objective function:

$$\underset{(\alpha,\beta,\gamma)}{\arg\min} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \alpha_{g_i,t} - x_{it}'\beta_{g_i,t})^2 + \lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{g,t} \left\| \beta_{g,t} - \beta_{g,t-1} \right\|, \tag{S.1}$$

where $y_{it} = \text{democracy}_{it}$ and $x_{it} = (\text{democracy}_{i,t-1}, \text{income}_{i,t-1})$. Note that this specification is a special case of those considered in Section S.6 where the theoretical properties of the estimator are investigated.

Table S.15: Income and democracy: Fully time varying intercept and $G = 2$

<table>
<tr><th></th><th>Regime</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th></tr>
<tr><td>Group 1</td><td>Intercept</td><td>−0.2958
(0.0619)</td><td>−0.0707
(0.0304)</td><td>0.1080
(0.0442)</td><td>0.0081
(0.0307)</td><td>0.0144
(0.0389)</td><td>0.0203
(0.0744)</td><td>0.0154
(0.0398)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td colspan="7">0.8633
(0.0251)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td colspan="7">0.1046
(0.0315)</td></tr>
<tr><td>Group 2</td><td>Intercept</td><td>−0.3121
(0.1490)</td><td>−0.4141
(0.1307)</td><td>−0.2189
(0.1336)</td><td>−0.0661
(0.1218)</td><td>0.2133
(0.0860)</td><td>0.2163
(0.1271)</td><td>0.4719
(0.0705)</td></tr>
<tr><td></td><td>Democracy$_{t-1}$</td><td colspan="4">−0.0025
(0.1062)</td><td colspan="3">0.3875
(0.0930)</td></tr>
<tr><td></td><td>Income$_{t-1}$</td><td colspan="4">0.4192
(0.0726)</td><td colspan="3">0.3083
(0.0825)</td></tr>
</table>

When we allow the intercept to be fully time varying, the BIC selects two groups. Table S.15 presents the GAGFL estimates of fully time varying intercept and slope coefficients for the two groups. Group 1 is characterized by stable slope coefficients, and the lag and income effects are both positive and strong. The average level of democracy generally improves over the years, but the estimates after the 3rd period are all insignificant. This suggests that, on one hand, fully time varying estimates are rather inefficient due to limited observations, and on the other hand, there may still exist heterogeneity in this group, at least on the average level of democracy. Group 2 exhibits a structural break in slope coefficients in the later period of the sample. The persistence of democracy becomes stronger while the income effect becomes weaker after the break. The average level of democracy also improves over the years as in Group 1, but to a larger extent and the estimates are more significant. Table S.17 displays the group membership estimates in this case. Group 1 contains stable countries and a few transitory countries that becomes democratic at different stages, e.g. Benin, Bolivia, Costa Rica, but their income effects are all highly persistent. Group 2 contains late transition countries whose democracy level and income effect both change remarkably after the transition.

Table S.16: Income and democracy: Fully time varying intercepts with $G = 4$

<table>
<tr><th></th><th>Regime</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th></tr>
<tr><td rowspan="6">Group 1</td><td>Intercept</td><td>−0.2841</td><td>−0.0793</td><td>0.0697</td><td>−0.0137</td><td>−0.0354</td><td>−0.1225</td><td>−0.0202</td></tr>
<tr><td></td><td>(0.0628)</td><td>(0.0292)</td><td>(0.0320)</td><td>(0.0326)</td><td>(0.0398)</td><td>(0.0489)</td><td>(0.0396)</td></tr>
<tr><td>Democracy$_{t-1}$</td><td colspan="7">0.8792</td></tr>
<tr><td></td><td colspan="7">(0.0274)</td></tr>
<tr><td>Income$_{t-1}$</td><td colspan="7">0.1259</td></tr>
<tr><td></td><td colspan="7">(0.0310)</td></tr>
<tr><td rowspan="6">Group 2</td><td>Intercept</td><td>−0.6387</td><td>−0.6561</td><td>0.1413</td><td>0.0994</td><td>0.0679</td><td>0.0303</td><td>0.4804</td></tr>
<tr><td></td><td>(0.1646)</td><td>(0.1708)</td><td>(0.1635)</td><td>(0.1308)</td><td>(0.0859)</td><td>(0.1240)</td><td>(0.1041)</td></tr>
<tr><td>Democracy$_{t-1}$</td><td colspan="3">0.0042</td><td colspan="4">0.6025</td></tr>
<tr><td></td><td colspan="3">(0.1345)</td><td colspan="4">(0.0920)</td></tr>
<tr><td>Income$_{t-1}$</td><td colspan="3">0.2996</td><td colspan="4">0.2413</td></tr>
<tr><td></td><td colspan="3">(0.0859)</td><td colspan="4">(0.0699)</td></tr>
<tr><td rowspan="6">Group 3</td><td>Intercept</td><td>−0.3768</td><td>−0.6127</td><td>−0.2270</td><td>−0.3395</td><td>0.3574</td><td>0.2572</td><td>0.5417</td></tr>
<tr><td></td><td>(0.2505)</td><td>(0.1084)</td><td>(0.0786)</td><td>(0.1812)</td><td>(0.1619)</td><td>(0.1409)</td><td>(0.1456)</td></tr>
<tr><td>Democracy$_{t-1}$</td><td>0.9002</td><td>−0.8079</td><td>0.2096</td><td colspan="4">0.0730</td></tr>
<tr><td></td><td>(0.3239)</td><td>(0.1797)</td><td>(0.1271)</td><td colspan="4">(0.1226)</td></tr>
<tr><td>Income$_{t-1}$</td><td>0.1254</td><td>1.0632</td><td>−0.9570</td><td colspan="4">0.6535</td></tr>
<tr><td></td><td>(0.3788)</td><td>(0.2164)</td><td>(0.1159)</td><td colspan="4">(0.1448)</td></tr>
<tr><td rowspan="6">Group 4</td><td>Intercept</td><td>0.2513</td><td>−0.1083</td><td>−0.4886</td><td>0.1054</td><td>0.3242</td><td>−0.1536</td><td>0.2243</td></tr>
<tr><td></td><td>(0.2568)</td><td>(0.1879)</td><td>(0.1701)</td><td>(0.1907)</td><td>(0.1081)</td><td>(0.1568)</td><td>(0.1291)</td></tr>
<tr><td>Democracy$_{t-1}$</td><td>0.4868</td><td colspan="3">0.3364</td><td>0.5674</td><td>0.4604</td><td>1.4734</td></tr>
<tr><td></td><td>(0.1602)</td><td colspan="3">(0.1798)</td><td>(0.1312)</td><td>(0.1122)</td><td>(0.2865)</td></tr>
<tr><td>Income$_{t-1}$</td><td>1.0548</td><td colspan="3">0.4953</td><td>0.5050</td><td>−0.9039</td><td>0.6153</td></tr>
<tr><td></td><td>(0.1930)</td><td colspan="3">(0.2147)</td><td>(0.2140)</td><td>(0.1733)</td><td>(0.1529)</td></tr>
</table>

Since the model with two groups may not fully capture the heterogeneity, we also estimate the model with four groups, but allowing the intercepts to be fully time varying. The coefficient estimates are presented in Table S.16, and the estimated group memberships are provided in Table S.17. The estimated group structure is largely similar to that in the case in which the intercepts are also penalized. The estimated group membership in this case is even closer to that of initial estimates under $G = 4$ with only one country switching. Again, Group 1 is featured by its stability of slope coefficients and the intercept. Compared to the case of fully time varying intercepts with $G = 2$, the estimated intercepts under $G = 4$ are more significant, confirming that heterogeneity is better controlled by allowing more groups. Group 2 contains countries of early transition, whose average level of democracy improves remarkably in period 2, and the lag and income effects exhibit a break in period 3. The slope coefficients of Group 3 are quite volatile in the first three periods, but become stable afterwards. The time path of the average level of democracy

Table S.17: Estimated group memberships when intercepts are fully time varying

	$G=2$	$G=4$		$G=2$	$G=4$		$G=2$	$G=4$
Algeria	1	1	Ghana	2	3	Nigeria	2	3
Argentina	2	3	Greece	2	2	Norway	1	1
Australia	1	1	Guatemala	2	4	Panama	2	2
Austria	1	1	Guinea	1	1	Paraguay	1	1
Belgium	1	1	Honduras	2	2	Peru	2	2
Benin	1	4	Iceland	1	1	Philippines	2	4
Bolivia	2	4	India	1	1	Portugal	1	2
Brazil	2	3	Indonesia	2	2	Romania	1	1
Burkina Faso	2	2	Iran	1	1	Rwanda	1	1
Burundi	1	1	Ireland	1	1	Sierra Leone	2	2
Cameroon	1	1	Israel	1	1	Singapore	1	1
Canada	1	1	Italy	1	1	South Africa	2	3
Central African Rep.	1	4	Jamaica	1	1	Spain	2	2
Chad	1	1	Japan	1	1	Sri Lanka	1	1
Chile	2	3	Jordan	1	1	Sweden	1	1
China	1	1	Kenya	1	1	Switzerland	1	1
Colombia	1	1	Korea, Rep.	2	3	Syrian Arab Rep.	1	1
Congo, Dem. Rep.	1	1	Luxembourg	1	1	Taiwan	1	2
Congo, Rep.	1	4	Madagascar	2	4	Tanzania	2	2
Costa Rica	1	1	Malawi	2	4	Thailand	2	3
Cote d'Ivoire	1	1	Malaysia	1	1	Togo	1	1
Cyprus	1	1	Mali	1	4	Trinidad & Tobago	1	1
Denmark	1	1	Mauritania	1	1	Tunisia	1	1
Dominican Rep.	1	2	Mexico	2	2	Turkey	2	4
Ecuador	2	2	Morocco	1	1	Uganda	2	2
Egypt, Arab Rep.	1	1	Nepal	1	2	United Kingdom	1	1
El Salvador	2	4	Netherlands	1	1	United States	1	1
Finland	1	1	New Zealand	1	1	Uruguay	2	3
France	1	1	Nicaragua	2	3	Venezuela, RB	1	1
Gabon	1	1	Niger	1	4	Zambia	2	4

labels Group 3 as late transition countries, in line with the results when the intercept is penalized. Finally, the income effect and democracy level of Group 4 fluctuate in the later periods of the sample, and the estimated intercept is insignificant in most periods.

S.3 Understanding cross-country savings differences

This section presents a second application that studies the determinants of the cross-country differences in savings behavior. The literature on international differences in savings can be dated back to Feldstein (1980). As acknowledged by several recent studies, countries at different economic and social development stages are characterized by disparate institutions, customs, and social norms, which further causes the savings rate to respond differently to its various determinants, especially in the short-run (see, for example, Pesaran et al., 2000; Loayza et al., 2000). In addition, institutional and social norms are likely to change over time, leading to time varying effects of the determinants of savings behavior, and this time varying feature is country-specific. Hence, it is important to incorporate the heterogeneous time varying behavior when studying the determinants of cross-country savings differences.

Thus motivated, we examine the effects of typical savings rate determinants by allowing them to be heterogeneous across countries and time varying. The dependent variable is the ratio of savings to GDP (S). Following Su et al. (2016) we consider the following determinants: the CPI-based inflation rate (I), real interest rate (R), and per capita GDP growth rate (G). We consider the specification with additive time-invariant country-specific effects:

$$S_{it} = \mu_i + \theta_{1,g_i,t} I_{it} + \theta_{2,g_i,t} R_{it} + \theta_{3,g_i,t} G_{it} + \varepsilon_{it}.$$

We estimate this model by applying GAGFL to the first-differenced data. We use the same data set as Su et al. (2016) that contains a balanced yearly panel of 56 countries over the time span of 1995–2010. Here, we range λ in the interval of $[0.01, 20]$, and use the same information criterion as in the simulation exercises and the first application to determine the tuning parameter. The number of groups is selected by the BIC as mentioned above. The minimum BIC corresponds to three groups, which is different from two groups selected by Su et al. (2016) based on C-Lasso estimation. The group composition produced by the two methods also deviates from each other to some extent. The difference mainly stems from allowing structural breaks in the coefficients and forming the grouping based on both the magnitude of the coefficient estimates and the breakpoints. In fact, two of our estimated three groups are characterized by stable slope coefficients, and they resemble the two groups reported by Su et al. (2016), while the extra third group is featured by one structural break. We find that our estimated clustering is to some extent related to the economic status of the countries. For example, Group 1 covers most of the developing countries, while Group 2 is mainly composed of countries with good economic performance.

Table S.18: GAGFL estimates of determinants of savings rate

	Regime	1–2	3–15
Group 1	Inflation		−0.0437 (0.0112)
	Interest rate		−0.1130 (0.0286)
	GDP growth		−0.0118 (0.0178)
Group 2	Inflation		0.1580 (0.0468)
	Interest rate		0.0969 (0.0201)
	GDP growth		0.1396 (0.0142)
Group 3	Inflation	−2.1285 (0.3821)	−0.2664 (0.0816)
	Interest rate	0.7922 (0.1120)	−0.3816 (0.0555)
	GDP growth	−0.1720 (0.2265)	0.0890 (0.0218)

Table S.18 displays the coefficient and structural break estimates of the three groups. Group 1 contains 26 countries, which are mainly emerging economies, i.e. most Southeast Asian countries and a large number of African and Latin American countries. Countries in this group are characterized by stable estimated coefficients, although their significance varies. In particular, inflation and interest rate are significantly and negatively associated with savings rate for this group, while the relationship between GDP growth rate and savings rate is ambiguous.

Group 2 is also characterized by stable slope coefficient estimates, but the effects of the determinants are in sharp contrast to those of Group 1. In particular, the effects of inflation and interest rate are strong and positive, and higher income growth is significantly associated with a higher savings rate. The opposite effect of inflation and interest rate in two groups is consistent with Su et al. (2016), and they also found that the significance of income growth varies across the two groups. This group contains 24 countries, and is mainly composed of developed countries, such as Canada, US, and West European countries. It also includes a few rapidly growing countries, e.g. China and Armenia.

Group 3 contains six countries, namely Bangladesh, Indonesia, Israel, Japan, Mongolia, and Sri Lanka. This group is featured by one structural break at the beginning of the sample period, namely 1997. The association between inflation and savings rate is significantly negative in both regimes, but the size is much weaker after the break, from −2.1285 to

−0.2664. The interest rate has a positive correlation with the savings rate before the break, but switches to negative after the break. Income growth is insignificantly related to the savings rate in the first regime, but the association is much stronger and becomes positive in the second regime. Further examination of these countries reveals that they all experienced a big drop in their income growth and savings rate at the breakpoint of 1997. Before the break, their savings rate moves closely with inflation and interest rate, while their income growth is relatively volatile. Nevertheless, the co-movement between income growth and savings rate is strengthened greatly after the break, while the association with the other two determinants is weakened. Interestingly, we note that this group contains mostly Asian countries except Israel, which were all severely affected by the Asia financial crisis in 1997. The crisis year corresponds to our estimated breakpoint precisely, and it thus well explains the sharp change in their economic performance and savings rate.

This application again confirms the importance of incorporating heterogeneous structural breaks. On one hand, the impact of a financial crisis and the time varying pattern of slope coefficients cannot be captured by standard classification approaches. Classification that ignores structural breaks tends to merge Groups 1 and 3. On the other hand, it is clear that not all countries are affected by the crisis, and thus assuming that the shock influences all individual units is not appropriate.

S.4 Proofs omitted in the main text

In this section, we present the proofs omitted in the main text. For ease of reference, we restate the assumptions. We also provide the statements of the lemmas before we present their proofs. We present these results in the supplement because they are similar to those in Bonhomme and Manresa (2015) or in Qian and Su (2016). Note that their results do not cover our case, and that the results presented here are new.

S.4.1 Assumptions

We first present the assumptions used in the lemmas. The explanations of these assumptions are given in the main text.

Assumption 1.

1. *$\mathcal{B}$ is compact.*

2. $E(\epsilon_{it}x_{it}) = 0$ *for all* i *and* t.

3. *There exists* $M > 0$ *such that for any* N *and* T,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}E\left(\epsilon_{it}\epsilon_{jt}x'_{it}x_{jt}\right)\right| < M$$

4. *There exists* $M > 0$ *such that for any* N *and* T,

$$\left|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}Cov\left(\epsilon_{it}\epsilon_{jt}x'_{it}x_{jt},\epsilon_{is}\epsilon_{js}x'_{is}x_{js}\right)\right| < M$$

5. *There exists* $M > 0$ *such that for any* N *and* T, $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(\|x_{it}\|^4\right) < M$.

Assumption 2.

1. *Let*

$$M(\gamma, g, \tilde{g}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0 = g\right\}\mathbf{1}\left\{g_i = \tilde{g}\right\}\begin{pmatrix} x_{i1}x'_{i1} & 0 & \dots & 0 \\ 0 & x_{i2}x'_{i2} & \dots & \dots \\ \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & x_{iT}x'_{iT} \end{pmatrix}.$$

Let $\hat{\rho}(\gamma, g, \tilde{g})$ *be the minimum eigenvalue of* $M(\gamma, g, \tilde{g})$. *There exist* $\hat{\rho}$ *and* $\rho > 0$ *such that* $\hat{\rho} \to_p \rho$ *and* $\forall g$,

$$\min_{\gamma\in\mathbb{G}^N}\max_{\tilde{g}\in\mathbb{G}}\hat{\rho}(\gamma, g, \tilde{g}) > \hat{\rho}.$$

2. *Let*

$$D_{g\tilde{g}i} = \frac{1}{T}\sum_{t=1}^{T}\left(x'_{it}(\beta^0_{g,t} - \beta^0_{\tilde{g},t})\right)^2.$$

For all $g \neq \tilde{g}$, *there exists a* $c_{g,\tilde{g}} > 0$ *such that*

$$\operatorname*{plim}_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}D_{g\tilde{g}i} > c_{g,\tilde{g}}$$

and for all i,

$$\operatorname*{plim}_{T\to\infty} D_{g\tilde{g}i} > c_{g,\tilde{g}}.$$

Assumption 3.

1. *There exists a constant M^*_{ex} such that as $N, T \to \infty$, for all $\delta > 0$,*

$$\sup_{1\le i\le N} \Pr\left(\frac{1}{T}\sum_{t=1}^{T} \|\epsilon_{it}x_{it}\|^2 \ge M^*_{ex}\right) = O(T^{-\delta}).$$

2. *There exists a constant M^*_x such that as $N, T \to \infty$, for all $\delta > 0$,*

$$\sup_{1\le i\le N} \Pr\left(\frac{1}{T}\sum_{t=1}^{T} \|x_{it}\|^4 \ge M^*_x\right) = O(T^{-\delta}).$$

3. *There exist constants $a > 0$ and $d_1 > 0$ and a sequence $\alpha[t] < \exp(-at^{d_1})$ such that, for all $i = 1, \ldots, N$ and $(g, \tilde{g}) \in \mathbb{G}^2$ such that $g \ne \tilde{g}$, $\{x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})\}_t$, $\{x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})\epsilon_{it}\}_t$ are strongly mixing process with mixing coefficients $\alpha[t]$. Moreover, $E(x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})\epsilon_{it}) = 0$.*

4. *There exist constants $b_x > 0$, $b_e > 0$, $d_{2x} > 0$ and d_{2e} such that $\Pr(|x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})| > m) \le \exp(1 - (m/b_x))^{d_{2x}}$ and $\Pr(|x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})\epsilon_{it}| > m) \le \exp(1 - (m/b_e))^{d_{2e}}$, for any i, t and $m > 0$.*

Assumption 4.

1. $\sqrt{N}T\lambda\left(\sum_{g\in\mathbb{G}} m^0_g\right) J^{-\kappa}_{\min} = O_p(1)$.

2. $\sqrt{N}T\lambda N^{-\kappa/2} \to_p \infty$.

3. $\sqrt{N}J_{\min} \to \infty$.

Assumption 5. *Suppose that Σ_x and Ω are well-defined, their minimum eigenvalues are bounded away from zero and their maximum eigenvalues are bounded uniformly over T.*

$N_g/N \to \pi_g > 0$ *for any* $g \in \mathbb{G}$. *Let*

$$d_{g,NT} = \frac{1}{\sqrt{N_g}} \sum_{g_i^0 = g} \left(\sum_{t=1}^{T_{g,1}^2 - 1} x_{it}\epsilon_{it}/\sqrt{I_{g,1}}, \dots, \sum_{t=T^0_{g,m_g^0}}^{T} x_{it}\epsilon_{it}/\sqrt{I_{g,m_g^0+1}} \right)'.$$

For a $l \times \sum_{g=1}^{G}(m_g^0+1)k$ *matrix* D, *where* l *does not depend on* T *and* $\lim_{T\to\infty} D\Omega D'$ *exists and is positive definite,* $D(d'_{1,NT}, \dots, d'_{G,NT})' \to_d N(0, \lim_{T\to\infty} D\Omega D')$.

Assumption 6. $N \sum_{g=1}^{G}(m_g^0)\lambda^2 I_{\min}^{-1} J_{\min}^{-2\kappa} = o_p(1)$.

S.4.2 Proofs for Section A.1

We present the proofs of the lemmas in Section A.1 of the main text. We also provide the statement of each lemma before the proof for ease of reference.

Recall that

$$\dot{Q}_{NT}(\beta, \gamma) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x'_{it}\beta_{g_i,t})^2$$

and

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

The Cauchy–Schwarz inequality is abbreviated as the CS inequality. M denotes a generic universal constant.

Lemma 3. *Suppose that Assumptions 1.1–4 hold.*

$$\sup_{(\beta,\gamma)\in\mathcal{B}^{GT}\times\Gamma_G} \left| \dot{Q}_{NT}(\beta, \gamma) - \tilde{Q}_{NT}(\beta, \gamma) \right| = o_p(1).$$

Proof. We observe that

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

We have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}x_{it}'\beta_{g_i,t} = \sum_{g=1}^{G}\frac{1}{NT}\mathbf{1}\{g_i=g\}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}x_{it}'\beta_{g,t}$$
$$= \sum_{g=1}^{G}\frac{1}{T}\sum_{t=1}^{T}\beta_{g,t}'\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\epsilon_{it}x_{it}.$$

For any $g \in \mathbb{G}$, the CS inequality implies that

$$\left(\frac{1}{T}\sum_{t=1}^{T}\beta_{g,t}'\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\epsilon_{it}x_{it}\right)^2$$
$$\leq \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\beta_{g,t}'\right\|^2\right) \times \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\epsilon_{it}x_{it}\right\|^2\right).$$

Assumption 1.1 implies that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\beta_{g,t}'\right\|^2 < M.$$

We also have

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\epsilon_{it}x_{it}\right\|^2$$
$$=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}\{g_i=g\}\mathbf{1}\{g_j=g\}\epsilon_{it}\epsilon_{jt}x_{it}'x_{jt}$$
$$=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}\{g_i=g\}\mathbf{1}\{g_j=g\}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}\epsilon_{jt}x_{it}'x_{jt}$$
$$\leq\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}\epsilon_{jt}x_{it}'x_{jt}\right|$$
$$\leq\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}E\left(\epsilon_{it}\epsilon_{jt}x_{it}'x_{jt}\right)\right|$$
$$+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\epsilon_{it}\epsilon_{jt}x_{it}'x_{jt}-E\left(\epsilon_{it}\epsilon_{jt}x_{it}'x_{jt}\right)\right)\right|$$

$$=o_p(1),$$

where the last equality follows from Assumptions 1.3 and 1.4. Thus, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}x_{it}'\beta_{g_i,t}=o_p(1)$$

uniformly over $\mathcal{B}^{GT}\times\Gamma_G$. Similarly we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}x_{it}'\beta^0_{g_i^0,t}=o_p(1).$$

Therefore, we have the desired result. □

We consider the following HD in $\mathcal{B}^{GT}$ such that

$$d_H(\beta^a,\beta^b)=\max\left\{\max_{g\in\mathbb{G}}\left(\min_{\tilde{g}\in\mathbb{G}}\frac{1}{T}\sum_{t=1}^{T}\left\|\beta^a_{\tilde{g},t}-\beta^b_{g,t}\right\|^2\right),\max_{\tilde{g}\in\mathbb{G}}\left(\min_{g\in\mathbb{G}}\frac{1}{T}\sum_{t=1}^{T}\left\|\beta^a_{\tilde{g},t}-\beta^b_{g,t}\right\|^2\right)\right\}$$

Lemma 4. *Suppose that Assumptions 1.1–4 and 2 hold.*

$$d_H(\dot{\beta},\beta^0)=o_p(1).$$

Proof. From Lemma 3, we have

$$\tilde{Q}(\dot{\beta},\dot{\gamma})=\dot{Q}(\dot{\beta},\dot{\gamma})+o_p(1)\le\dot{Q}(\beta^0,\gamma^0)+o_p(1)=\tilde{Q}(\beta^0,\gamma^0)+o_p(1).$$

Because $\tilde{Q}(\beta,\gamma)$ is minimized at $\beta=\beta^0$ and $\gamma=\gamma^0$, we have

$$\tilde{Q}(\dot{\beta},\dot{\gamma})-\tilde{Q}(\beta^0,\gamma^0)=o_p(1).$$

On the other hand, we have

$$\begin{aligned}\tilde{Q}(\beta,\gamma)-\tilde{Q}(\beta^0,\gamma^0)=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}'(\beta^0_{g_i^0,t}-\beta_{g_i,t})\right)^2\\=&\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\frac{1}{T}\left(\beta^0_g-\beta_{\tilde{g}}\right)'M(\gamma,g,\tilde{g})\left(\beta^0_g-\beta_{\tilde{g}}\right)\end{aligned}$$

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

Note that $\hat{\rho}$ is asymptotically bounded away from zero by Assumption 2.1.

Therefore we have

$$
\max_{g\in\mathbb{G}} \left(\min_{\tilde{g}\in\mathbb{G}} \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \beta_{g,t}^0 - \dot{\beta}_{\tilde{g},t} \right\|^2 \right) \right) = o_p(1). \tag{S.2}
$$

Let

$$
\sigma(g) = \arg\min_{\tilde{g}\in\mathbb{G}} \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \beta_{g,t}^0 - \dot{\beta}_{\tilde{g},t} \right\|^2 \right).
$$

Then we have for $\tilde{g} \neq g$,

$$
\begin{aligned}
\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it}'(\dot{\beta}_{\sigma(g),t} - \dot{\beta}_{\sigma(\tilde{g}),t}) \right)^2 \right)^{1/2} \geq & \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it}'(\beta_{g,t}^0 - \beta_{\tilde{g},t}^0) \right)^2 \right)^{1/2} \\
& - \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it}'(\dot{\beta}_{\sigma(g),t}^0 - \beta_{g,t}^0) \right)^2 \right)^{1/2} \\
& - \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it}'(\dot{\beta}_{\sigma(\tilde{g}),t} - \beta_{\tilde{g},t}^0) \right)^2 \right)^{1/2}.
\end{aligned}
$$

Assumption 2.2 states that the first term on the right hand side of the inequality is bounded away from zero. Equation (S.2) implies that the second and third terms are $o_p(1)$. Therefore, we have $\sigma(g) \neq \sigma(\tilde{g})$ with probability approaching one, which implies that with probability approaching one σ is bijective and has the inverse which is denoted as σ^{-1}.

Thus, we have

$$\min_{g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\beta^0_{g,t}-\dot{\beta}_{\tilde{g},t}\right\|^2\right)\geq\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\beta^0_{\sigma^{-1}(\tilde{g}),t}-\dot{\beta}_{\tilde{g},t}\right\|^2\right)$$
$$=\min_{h\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\beta^0_{\sigma^{-1}(\tilde{g}),t}-\dot{\beta}_{h,t}\right\|^2\right)=o_p(1),$$

where the last equality follows from (S.2). Therefore we have

$$\max_{\tilde{g}\in\mathbb{G}}\left(\min_{g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\beta^0_{g,t}-\dot{\beta}_{\tilde{g},t}\right\|^2\right)\right)=o_p(1).$$

We thus have the desired result. □

The proof of Lemma 4 shows that there exists a permutation σ such that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\beta^0_{\sigma(g),t}-\dot{\beta}_{g,t}\right\|^2=o_p(1).$$

We obtain $\sigma(g)=g$ by relabeling.

Define

$$\mathcal{N}_\eta=\left\{\beta\in\mathcal{B}^{GT}:\frac{1}{T}\sum_{t=1}^{T}\left\|\beta^0_{g,t}-\beta_{g,t}\right\|^2<\eta,\forall g\in\mathbb{G}\right\}.$$

Let

$$\hat{g}_i(\beta)=\arg\min_{g\in\mathbb{G}}\sum_{t=1}^{T}(y_{it}-x_{it}'\beta_{g,t})^2. \tag{S.3}$$

Lemma 5. *Suppose that Assumptions 2.2 and 3 are satisfied. For $\eta>0$ small enough, we have, $\forall\delta>0$,*

$$\sup_{\beta\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\beta)\neq g_i^0\right\}=o_p(T^{-\delta}).$$

Proof. For any $g \in \mathbb{G}$, we have

$$\mathbf{1}\left\{\hat{g}_i(\beta) = g\right\} \leq \mathbf{1}\left\{\sum_{t=1}^{T}(y_{it} - x_{it}'\beta_{g,t})^2 \leq \sum_{t=1}^{T}(y_{it} - x_{it}'\beta_{g_i^0,t})^2\right\}.$$

Thus, we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\beta) \neq g_i^0\right\} &= \sum_{g=1}^{G}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0 \neq g\right\}\mathbf{1}\left\{\hat{g}_i(\beta) = g\right\} \\ &\leq \sum_{g=1}^{G}\frac{1}{N}\sum_{i=1}^{N}Z_{ig}(\beta),\end{aligned}$$

where

$$Z_{ig}(\beta) = \mathbf{1}\left\{g_i^0 \neq g\right\}\mathbf{1}\left\{\sum_{t=1}^{T}(y_{it} - x_{it}'\beta_{g,t})^2 \leq \sum_{t=1}^{T}(y_{it} - x_{it}'\beta_{g_i^0,t})^2\right\}.$$

We now bound $Z_{ig}(\beta)$. We have

$$\begin{aligned}&Z_{ig}(\beta) \\ \leq&\mathbf{1}\left\{g_i^0 \neq g\right\} \\ &\times \mathbf{1}\left\{\sum_{t=1}^{T}x_{it}'(\beta_{g_i^0,t} - \beta_{g,t}))\left(x_{it}'\beta_{g_i^0,t}^0 + \epsilon_{it} - \frac{x_{it}'(\beta_{g_i^0,t} + \beta_{g,t})}{2}\right) \leq 0\right\} \\ \leq& \max_{\tilde{g}\in\mathbb{G}\setminus\{g\}}\mathbf{1}\left\{\sum_{t=1}^{T}x_{it}'(\beta_{\tilde{g},t} - \beta_{g,t})\left(x_{it}'\beta_{\tilde{g},t}^0 + \epsilon_{it} - \frac{x_{it}'(\beta_{\tilde{g},t} + \beta_{g,t})}{2}\right) \leq 0\right\}.\end{aligned}$$

Let

$$\begin{aligned}A_T =&\Bigg|\sum_{t=1}^{T}x_{it}'(\beta_{\tilde{g},t} - \beta_{g,t})\left(x_{it}'\beta_{\tilde{g},t}^0 + \epsilon_{it} - \frac{x_{it}'(\beta_{\tilde{g},t} + \beta_{g,t})}{2}\right) \\ &- \sum_{t=1}^{T}x_{it}'(\beta_{\tilde{g},t}^0 - \beta_{g,t}^0)\left(x_{it}'\beta_{\tilde{g},t}^0 + \epsilon_{it} - \frac{x_{it}'(\beta_{\tilde{g},t}^0 + \beta_{g,t}^0)}{2}\right)\Bigg|.\end{aligned}$$

Then we have

$$A_T \leq \left|\sum_{t=1}^{T}x_{it}'(\beta_{\tilde{g},t} - \beta_{g,t})\epsilon_{it} - \sum_{t=1}^{T}x_{it}'(\beta_{\tilde{g},t}^0 - \beta_{g,t}^0)\epsilon_{it}\right|$$

$$+\left|\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}-\beta_{g,t})x_{it}'\beta_{\tilde{g},t}^{0}-\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}^{0}-\beta_{g,t}^{0})x_{it}'\beta_{\tilde{g},t}^{0}\right|$$
$$+\left|\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}-\beta_{g,t})\frac{x_{it}'(\beta_{\tilde{g},t}+\beta_{g,t})}{2}-\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}-\beta_{g,t})\frac{x_{it}'(\beta_{\tilde{g},t}^{0}+\beta_{g,t}^{0})}{2}\right|$$
$$+\left|\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}-\beta_{g,t})\frac{x_{it}'(\beta_{\tilde{g},t}^{0}+\beta_{g,t}^{0})}{2}-\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}^{0}-\beta_{g,t}^{0})\frac{x_{it}'(\beta_{\tilde{g},t}^{0}+\beta_{g,t}^{0})}{2}\right|.$$

Thus, when $\beta \in \mathcal{N}_\eta$, the CS inequality implies that

$$A_T \leq 2T\left(\frac{1}{T}\sum_{t=1}^{T}\|\epsilon_{it}x_{it}\|^2\right)^{1/2}\sqrt{\eta}+2T\left(\frac{1}{T}\sum_{t=1}^{T}(x_{it}'\beta_{\tilde{g},t}^{0})^2\|x_{it}\|^2\right)^{1/2}\sqrt{\eta}$$
$$+T\left(\frac{1}{T}\sum_{t=1}^{T}(x_{it}'(\beta_{\tilde{g},t}-\beta_{g,t}))^2\|x_{it}\|^2\right)^{1/2}\sqrt{\eta}+T\left(\frac{1}{T}\sum_{t=1}^{T}(x_{it}'(\beta_{\tilde{g},t}^{0}+\beta_{g,t}^{0}))^2\|x_{it}\|^2\right)^{1/2}\sqrt{\eta}.$$

As $\mathcal{B}$ is bounded, we have, for $\beta \in \mathcal{N}_\eta$,

$$A_T \leq C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|\epsilon_{it}x_{it}\|^2\right)^{1/2}+C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^4\right)^{1/2},$$

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

Thus, we have

$$Z_{ig}(\beta) \leq \max_{\tilde{g}\in\mathbb{G}\setminus\{g\}} \mathbf{1}\Big\{\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}^{0}-\beta_{g,t}^{0})\left(x_{it}'\beta_{\tilde{g},t}^{0}+\epsilon_{it}-\frac{x_{it}'(\beta_{\tilde{g},t}^{0}+\beta_{g,t}^{0})}{2}\right)$$
$$\leq C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|\epsilon_{it}x_{it}\|^2\right)^{1/2}+C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^4\right)^{1/2}\Big\}.$$

Let

$$\tilde{Z}_{ig} = \max_{\tilde{g}\in\mathbb{G}\setminus\{g\}} \mathbf{1}\Big\{\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}^{0}-\beta_{g,t}^{0})\left(x_{it}'\beta_{\tilde{g},t}^{0}+\epsilon_{it}-\frac{x_{it}'(\beta_{\tilde{g},t}^{0}+\beta_{g,t}^{0})}{2}\right)$$
$$\leq C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|\epsilon_{it}x_{it}\|^2\right)^{1/2}+C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^4\right)^{1/2}\Big\}.$$

Thus, we have

$$\sup_{\beta\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\beta)\neq g_i^0\right\}\leq\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}Z_{ig}(\beta)\leq\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}\tilde{Z}_{ig}.$$

Note that $\tilde{Z}_{ig}$ does not depend on β.

We now bound $\Pr(\tilde{Z}_{ig}=1)$. We observe

$$\begin{aligned}
&\Pr(\tilde{Z}_{ig}=1)\\
&\leq\sum_{\tilde{g}\in\mathbb{G}\setminus\{g\}}\Pr\Big(\sum_{t=1}^{T}x_{it}'(\beta_{\tilde{g},t}^0-\beta_{g,t}^0)\epsilon_{it}\leq-\frac{1}{2}\sum_{t=1}^{T}\left(x_{it}'(\beta_{\tilde{g},t}^0-\beta_{g,t}^0)\right)^2\\
&\quad+C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|\epsilon_{it}x_{it}\|^2\right)^{1/2}+C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^4\right)^{1/2}\Big).
\end{aligned}$$

Let $M^*>\max(M_{ex}^*,M_x^*)$, where M_{ex}^* and M_x^* are defined in Assumptions 3.1 and 3.2, respectively. Thus, we have

$$\begin{aligned}
&\Pr(\tilde{Z}_{ig}=1)\\
&\leq\sum_{\tilde{g}\in\mathbb{G}\setminus\{g\}}\Bigg[\Pr\left(\frac{1}{T}\sum_{t=1}^{T}\|\epsilon_{it}x_{it}\|^2\geq M^*\right)+\Pr\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^4\geq M^*\right)\\
&\quad+\Pr\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}'(\beta_{\tilde{g},t}^0-\beta_{g,t}^0)\right)^2\leq\frac{c_{g,\tilde{g}}}{2}\right)\\
&\quad+\Pr\left(\sum_{t=1}^{T}x_{it}'(\beta_{\tilde{g},t}^0-\beta_{g,t}^0)\epsilon_{it}\leq-T\frac{c_{g,\tilde{g}}}{4}+TC_3\sqrt{\eta}\sqrt{M^*}\right)\Bigg],
\end{aligned}$$

where C_3 is a constant that is independent of η and T. Assumptions 3.1 and 3.2 give that

$$\Pr\left(\frac{1}{T}\sum_{t=1}^{T}\|\epsilon_{it}x_{it}\|^2\geq M^*\right)+\Pr\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^4\geq M^*\right)=O(T^{-\delta}).$$

Under Assumptions 2.2, 3 and 4, a very similar argument to Bonhomme and Manresa (2015, page 1176) implies that

$$\Pr\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}'(\beta_{\tilde{g},t}^0-\beta_{g,t}^0)\right)^2\leq\frac{c_{g,\tilde{g}}}{2}\right)=O(T^{-\delta}).$$

Take η such that

$$\eta \leq \left(\frac{\min_{\tilde{g} \in \mathbb{G} \setminus \{g\}} c_{g,\tilde{g}}}{8C_3\sqrt{M^*}} \right).$$

We then have

$$\begin{aligned}
&\Pr\left(\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}^0 - \beta_{g,t}^0)\epsilon_{it} \leq -T\frac{c_{g,\tilde{g}}}{4} + TC_3\sqrt{\eta}\sqrt{M^*} \right) \\
\leq &\Pr\left(\frac{1}{T}\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}^0 - \beta_{g,t}^0)\epsilon_{it} \leq -\frac{c_{g,\tilde{g}}}{8} \right) \\
\leq &\Pr\left(\left| \frac{1}{T}\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}^0 - \beta_{g,t}^0)\epsilon_{it} \right| > \frac{c_{g,\tilde{g}}}{8} \right).
\end{aligned}$$

Under Assumptions 3 and 4, a very similar argument to Bonhomme and Manresa (2015, page 1177), in particular the use of exponential inequalities in Bonhomme and Manresa (2015, Lemma B.5), implies that

$$\Pr\left(\left| \frac{1}{T}\sum_{t=1}^{T} x_{it}'(\beta_{\tilde{g},t}^0 - \beta_{g,t}^0)\epsilon_{it} \right| > \frac{c_{g,\tilde{g}}}{8} \right) = O(T^{-\delta}).$$

We thus have

$$\Pr(\tilde{Z}_{ig} = 1) \leq (G-1)O(T^{-\delta}).$$

This implies that

$$\begin{aligned}
E\left(\sup_{\beta \in \mathcal{N}_\eta} \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left\{ \hat{g}_i(\beta) \neq g_i^0 \right\} \right) \leq &\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N} E\left(\tilde{Z}_{ig}\right) \\
= &\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N} \Pr\left(\tilde{Z}_{ig} = 1\right) \\
= &G(G-1)O(T^{-\delta}) = O(T^{-\delta}).
\end{aligned}$$

The Markov inequality implies the desired result.

□

Let

$$\check{\beta} = \arg\min_{\beta \in \mathcal{B}^{GT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}' \beta_{g_i^0, t})^2.$$

Note that $\check{\beta}$ is the estimator of β when the group memberships (i.e., γ^0) are known. Let

$$\check{Q}(\beta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}' \beta_{g_i^0, t})^2.$$

Note that $\check{Q}(\beta) = \dot{Q}(\beta, \gamma^0)$ and that $\check{\beta} = \arg\min_{\beta \in \mathcal{B}^{GT}} \check{Q}(\beta)$.

Lemma 6. *Suppose that Assumptions 1 and 2.1 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. Then it follows that for all g and t,*

$$\check{\beta}_{g,t} - \beta_{g,t}^0 = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Proof. Observe that $\check{\beta}_{g,t}$ is a least squares estimator whose objective function is

$$\sum_{g_i^0 = g} (y_{it} - x_{it}' \beta)^2.$$

The result holds by the standard argument for OLS. $\square$

Let

$$\dot{Q}(\beta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}' \beta_{\hat{g}_i(\beta), t})^2.$$

Note also that $\dot{Q}(\beta) = \dot{Q}(\beta, \hat{\gamma}(\beta))$ and that $\dot{\beta} = \arg\min_{\beta \in \mathcal{B}^{GT}} \dot{Q}(\beta)$

Lemma 7. *Suppose that Assumptions 1, 2, and 3 are satisfied. As $N, T \to \infty$, for any $\delta > 0$, it holds that*

$$\dot{\beta}_{g,t} = \check{\beta}_{g,t} + o_p(T^{-\delta}),$$

for all g and t.

Proof. We first evaluate the difference between $\check{Q}(\beta)$ and $\dot{Q}(\beta)$. We note that

$$\begin{aligned}\check{Q}(\beta)-\dot{Q}(\beta)=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{\hat{g}_i(\beta)\neq g_i^0\}\left((y_{it}-x_{it}'\beta_{g_i^0,t})^2-(y_{it}-x_{it}'\beta_{\hat{g}_i(\beta),t})^2\right)\\=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{\hat{g}_i(\beta)\neq g_i^0\}\left((y_{it}-x_{it}'\beta_{g_i^0,t})^2-(y_{it}-x_{it}'\beta_{\hat{g}_i(\beta),t})^2\right)\end{aligned}$$

By Assumptions 1.1, 1.3 and 1.5 and Lemma 5, we have, for sufficiently small η,

$$\sup_{\beta\in\mathcal{N}_\eta}\left|\check{Q}(\beta)-\dot{Q}(\beta)\right|=o_p(T^{-\delta}). \tag{S.4}$$

We then see that

$$\Pr\left(\left|\check{Q}(\dot{\beta})-\dot{Q}(\dot{\beta})\right|>\epsilon T^{-\delta}\right)\leq \Pr(\dot{\beta}\notin\mathcal{N}_\eta)+\Pr\left(\sup_{\beta\in\mathcal{N}_\eta}\left|\check{Q}(\beta)-\dot{Q}(\beta)\right|>\epsilon T^{-\delta}\right)=o(1),$$

by that $\dot{\beta}$ is consistent from Lemma 4, and (S.4). Therefore we have

$$\check{Q}(\dot{\beta})-\dot{Q}(\dot{\beta})=o_p(T^{-\delta}).$$

Similarly, we have

$$\check{Q}(\check{\beta})-\dot{Q}(\check{\beta})=o_p(T^{-\delta}).$$

Next, we evaluate the difference between $\check{\beta}$ and $\dot{\beta}$. By the definition of $\check{\beta}$ and $\dot{\beta}$, we have

$$0\leq\check{Q}(\dot{\beta})-\check{Q}(\check{\beta})=\dot{Q}(\dot{\beta})-\dot{Q}(\check{\beta})+o_p(T^{-\delta})\leq o_p(T^{-\delta}).$$

Thus we have

$$\check{Q}(\dot{\beta})-\check{Q}(\check{\beta})=o_p(T^{-\delta}) \tag{S.5}$$

We observe that

$$\check{Q}(\dot{\beta})-\check{Q}(\check{\beta})=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}'\dot{\beta}_{g_i^0,t})^2-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}'\check{\beta}_{g_i^0,t})^2$$

$$
\begin{aligned}
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((y_{it}-x_{it}'\check{\beta}_{g_i^0,t}+x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t}))^2-(y_{it}-x_{it}'\check{\beta}_{g_i^0,t})^2\right)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t}))^2+2(y_{it}-x_{it}'\check{\beta}_{g_i^0,t})(x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t}))\right)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t}))^2,
\end{aligned}
$$

where the last equality holds because $\check{\beta}$ is an OLS estimator and satisfies $\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}'\check{\beta}_{g_i^0,t})x_{it}=0$. Moreover, we have

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t}))^2=&\frac{1}{T}\sum_{g\in\mathbb{G}}(\check{\beta}_g-\dot{\beta}_g)'M(\gamma^0,g,g)(\check{\beta}_g-\dot{\beta}_g)\\
\geq&\hat{\rho}\frac{1}{T}\sum_{g\in\mathbb{G}}\left\|\dot{\beta}_g-\check{\beta}_g\right\|^2.
\end{aligned}
$$

We thus have, by (S.5) and Assumption 2.1,

$$
\frac{1}{T}\sum_{g\in\mathbb{G}}\left\|\dot{\beta}_g-\check{\beta}_g\right\|^2=o_p(T^{-\delta}).
$$

This implies that

$$
\left\|\dot{\beta}_{g,t}-\check{\beta}_{g,t}\right\|^2=o_p(T^{1-\delta})
$$

for any δ. Thus we have the desired result.

□

We can now consider the rate of convergence of the elements of $\dot{\beta}$.

Theorem 1. *Suppose that Assumptions 1, 2 and 3 hold. Suppose that $N_g/N\to\pi_g>0$ for any $g\in\mathbb{G}$. Then it follows that for all g and t,*

$$
\dot{\beta}_{g,t}-\beta_{g,t}^0=O_p\left(\frac{1}{\sqrt{N}}\right).
$$

Proof. The theorem follows from Lemmas 6 and 7.

□

S.4.3 Proofs for Section A.2

We present the proofs of the lemmas in Section A.2 of the main text. We also provide the statement of each lemma before the proof for ease of reference.

Recall that

$$\mathring{Q}(\beta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x'_{it}\beta_{g_i^0,t})^2 + \lambda \sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t} - \beta_{g,t-1}\right\|,$$

and that $\mathring{\beta} = \arg\min_{\beta\in\mathcal{B}^{GT}} \mathring{Q}(\beta)$.

Lemma 8. *Suppose that Assumptions 1.3, 2.1 and 4.1 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. We have, as $N, T \to \infty$,*

$$\frac{1}{T}\left\|\mathring{\beta}_g - \beta_g^0\right\|^2 = O_p\left(\frac{1}{N}\right)$$

for any $g \in \mathbb{G}$. We also have, as $N, T \to \infty$,

$$\mathring{\beta}_{g,t} - \beta_{g,t}^0 = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Proof. This lemma contains two parts. We first consider the first part regarding the norm of coefficient difference. Let $b_t = N^{-1/2}(\beta_{g,t} - \beta_{g,t}^0)$ and $b = (b'_1, \cdots b'_T)' = N^{-1/2}(\beta_g - \beta_g^0)$. Let

$$Q_g(\beta_g) = \frac{1}{NT}\sum_{g_i^0=g}\sum_{t=1}^{T}(y_{it} - x'_{it}\beta_{g,t})^2 + \lambda\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t} - \beta_{g,t-1}\right\|.$$

Note that $\mathring{\beta}_g = \arg\min Q_g(\beta_g)$.

We have

$$\begin{aligned}
N(Q_g(\beta_g) - Q_g(\beta_g^0)) =& \frac{1}{T}\sum_{g_i^0=g}\sum_{t=1}^{T}(y_{it} - x'_{it}\beta_{g,t})^2 + N\lambda\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t} - \beta_{g,t-1}\right\| \\
&- \frac{1}{T}\sum_{g_i^0=g}\sum_{t=1}^{T}(y_{it} - x'_{it}\beta_{g,t}^0)^2 - N\lambda\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t}^0 - \beta_{g,t-1}^0\right\| \\
=& \frac{1}{NT}\sum_{g_i^0=g}\sum_{t=1}^{T}b'_t x_{it}x'_{it}b_t - \frac{2}{\sqrt{NT}}\sum_{g_i^0=g}\sum_{t=1}^{T}\epsilon_{it}x'_{it}b_t
\end{aligned}$$

$$+ N\lambda \sum_{t \in \mathcal{T}^0_{m^0_g,g}} \dot{w}_{g,t} \left(\left\| \beta^0_{g,t} - \beta^0_{g,t-1} + N^{-1/2}(b_t - b_{t-1}) \right\| - \left\| \beta^0_{g,t} - \beta^0_{g,t-1} \right\| \right)$$
$$+ N\lambda \sum_{t \in \mathcal{T}^{0c}_{m^0_g,g}} \dot{w}_{g,t} \left\| N^{-1/2}(b_t - b_{t-1}) \right\|. \tag{S.6}$$

We consider the four terms on the right hand side of the last equality of (S.6) in turn. For the first term, we note, by Assumption 2.1, that

$$\frac{1}{NT} \sum_{g_i^0 = g} \sum_{t=1}^{T} b_t' x_{it} x_{it}' b_t = \frac{1}{T} b' M(\gamma^0, g, g) b \geq \hat{\rho} \frac{1}{T} \|b\|^2.$$

For the second term, we have

$$\frac{2}{\sqrt{N}T} \sum_{g_i^0 = g} \sum_{t=1}^{T} \epsilon_{it} x_{it}' b_t = O_p\left(\frac{1}{\sqrt{T}} \|b\| \right),$$

by Assumption 1.3. Next, we consider the third term. We have by the Jensen, triangular and CS inequalities,

$$\begin{aligned}
&N\lambda \sum_{t \in \mathcal{T}^0_{m^0_g,g}} \dot{w}_{g,t} \left(\left\| \beta^0_{g,t} - \beta^0_{g,t-1} + N^{-1/2}(b_t - b_{t-1}) \right\| - \left\| \beta^0_{g,t} - \beta^0_{g,t-1} \right\| \right) \\
\leq &N\lambda \sum_{t \in \mathcal{T}^0_{m^0_g,g}} \dot{w}_{g,t} \left\| N^{-1/2}(b_t - b_{t-1}) \right\| \\
\leq &m^0 \sqrt{N} \lambda \max_{s \in \mathcal{T}^0_{m^0_g,g}} (\dot{w}_{g,s}) \frac{1}{m^0} \sum_{t \in \mathcal{T}^0_{m^0_g,g}} \|b_t - b_{t-1}\| \\
\leq &m^0 \sqrt{N} \lambda \max_{s \in \mathcal{T}^0_{m^0_g,g}} (\dot{w}_{g,s}) \left(\frac{1}{m^0} \sum_{t \in \mathcal{T}^0_{m^0_g,g}} \|b_t - b_{t-1}\|^2 \right)^{1/2} \\
\leq &2\sqrt{m^0 N} \lambda \max_{s \in \mathcal{T}^0_{m^0_g,g}} (\dot{w}_{g,s}) \|b\|.
\end{aligned}$$

By Assumption 4.1, this third term is $O_p(T^{-1/2}\|b\|)$. Lastly, we consider the fourth term, and we have

$$N\lambda \sum_{t\in\mathcal{T}^{0c}_{m^0_g,g}} \dot{w}_{g,t} \left\|N^{-1/2}(b_t - b_{t-1})\right\| \geq 0.$$

Summing the four parts up, we have

$$0 \geq N(Q_g(\beta_g) - Q_g(\beta^0_g)) \geq \hat{\rho}\frac{1}{T}\|b\|^2 - O_p(T^{-1/2}\|b\|).$$

If $\|b\|^2/T$ is not stochastically bounded, then the first term, which is positive, dominates and the first inequality does not hold asymptotically. We thus have the desired result.

Next, we consider the second part regarding the difference between two coefficient vectors. The above argument implies that

$$\begin{aligned}0 \geq N(Q_g(\beta_g) - Q_g(\beta^0_g)) \geq & \frac{1}{NT}\sum_{g^0_i=g}\sum_{t=1}^{T} b'_t x_{it} x'_{it} b_t - \frac{2}{\sqrt{NT}}\sum_{g^0_i=g}\sum_{t=1}^{T}\epsilon_{it} x'_{it} b_t \\ & - \sqrt{N}\lambda \max_{s\in\mathcal{T}^0_{m^0_g,g}} (\dot{w}_{g,s}) \sum_{t\in\mathcal{T}^0_{m^0_g,g}} \|b_t - b_{t-1}\|.\end{aligned}$$

We thus have

$$\begin{aligned}0 \geq N(Q_g(\beta_g) - Q_g(\beta^0_g)) \geq & \frac{1}{NT}\sum_{g^0_i=g}\sum_{t=1}^{T} b'_t x_{it} x'_{it} b_t - \frac{2}{\sqrt{NT}}\sum_{g^0_i=g}\sum_{t=1}^{T}\epsilon_{it} x'_{it} b_t \\ & - \sqrt{N}\lambda \max_{s\in\mathcal{T}^0_{m^0_g,g}} (\dot{w}_{g,s}) \sum_{t=1}^{T} 2\|b_t\|.\end{aligned}$$

The right hand side of the inequality can be analyzed for each t. If $|b_t|$ is not $O_p(1)$, then it can be seen that it cannot be optimal. The desired result thus follows.

□

Recall $\mathring{\theta}_{g,1} = \mathring{\beta}_{g,1}$ and $\mathring{\theta}_{g,t} = \mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1}$.

Lemma 9. *Suppose that Assumptions 1.3, 2.1, 1.5, and 4 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. It follows that*

$$\Pr\left(\left\|\mathring{\theta}_{g,t}\right\| = 0, \forall t \in \mathcal{T}^{0c}_{m^0_g,g}, g \in \mathbb{G}\right) \to 1$$

as $N \to \infty$.

Proof. The proof is by contradiction. Suppose that $\exists(t,g)$ such that $2 \le t \le T-1$ and $\mathring{\theta}_{g,t} \neq 0$ for sufficiently large N. Note that $\|\theta\|$ is differentiable at $\mathring{\theta}_{g,t}$ if $\mathring{\theta}_{g,t} \neq 0$. We thus have the following first order condition (FOC) for $\mathring{\beta}_{g,t}$.

$$-2\frac{1}{NT}\sum_{g_i^0=g}(y_{it}-x_{it}'\mathring{\beta}_{g,t})x_{it}+\lambda\dot{w}_{g,t}\frac{\mathring{\theta}_{g,t}}{\|\mathring{\theta}_{g,t}\|}-\lambda\dot{w}_{g,t+1}e_{t+1}=0,$$

where $e_{t+1}=\mathring{\theta}_{g,t+1}/\|\mathring{\theta}_{g,t+1}\|$ if $\mathring{\theta}_{g,t+1}\neq 0$ and $\|e_{t+1}\|\le 1$ otherwise. Multiplying both sides of the equation by $\sqrt{N}T$ and use $y_{it}=x_{it}'\beta^0_{g_i^0,t}+\epsilon_{it}$, we have

$$2\frac{1}{\sqrt{N}}\sum_{g_i^0=g}x_{it}x_{it}'(\mathring{\beta}_{g,t}-\beta^0_{g,t})-2\frac{1}{\sqrt{N}}\sum_{g_i^0=g}\epsilon_{it}x_{it}$$
$$+\sqrt{N}T\lambda\dot{w}_{g,t}\frac{\mathring{\theta}_{g,t}}{\|\mathring{\theta}_{g,t}\|}-\sqrt{N}T\lambda\dot{w}_{g,t+1}e_{t+1}=0.$$

The first term is $O_p(1)$ by Lemma 8 and Assumption 1.5. The second term is $O_p(1)$ by Assumption 1.3. For the third term, we observe that the absolute value of at least one element of $\mathring{\theta}_{g,t}/\|\mathring{\theta}_{g,t}\|$ exceeds $1/\sqrt{k}$, and that $\dot{w}_{g,t}^{-1}=O_p(N^{-\kappa/2})$ because $t\in\mathcal{T}^{0c}_{m_g^0,g}$ and because of Theorem 1. Therefore, the third term is $O_p(\sqrt{N}T\lambda N^{-\kappa/2}/\sqrt{k})$ and this tends to infinity by Assumption 4.2.

We now consider the term $\sqrt{N}T\lambda\dot{w}_{g,t+1}e_{t+1}$. Suppose that $t+1\in\mathcal{T}^0_{m_g^0,g}$. In this case, $\dot{w}_{g,t+1}=O_p(J_{\min}^{-\kappa})$ because of Theorem 1. This and Assumption 4.1 implies that $\sqrt{N}T\lambda\dot{w}_{g,t}e_{t+1}=O_p(1)$. In this case, the third term explodes but other terms are stochastically bounded, and the first order condition cannot hold.

Next, we consider the case where $t+1\in\mathcal{T}^{0c}_{m_g^0,g}$. We note that from the argument in the previous paragraph, if $t=T^0_{g,j}-1\in\mathcal{T}^{0c}_{m_g^0,g}$ for some j, then $\Pr(\|\mathring{\theta}_{g,t}\|=0)\to 1$ and $\sqrt{N}T\lambda\dot{w}_{g,t+1}e_{t+1}=O_p(1)$. This implies that for $t=T^0_{g,j}-2\in\mathcal{T}^{0c}_{m_g^0,g}$, $\Pr(\|\mathring{\theta}_{g,t}\|=0)\to 1$ and $\sqrt{N}T\lambda\dot{w}_{g,t+1}e_{t+1}=O_p(1)$ too. Applying this argument recursively until $t=T^0_{g,j-1}+1\in\mathcal{T}^{0c}_{m_g^0,g}$, we have for all $t\in\mathcal{T}^{0c}_{m_g^0,g}$, $\Pr(\|\mathring{\theta}_{g,t}\|=0)\to 1$.

Lastly, we consider the case in which $t=T$. In this case, the first order condition is

$$2\frac{1}{\sqrt{N}}\sum_{g_i^0=g}x_{iT}x_{iT}'(\mathring{\beta}_{g,T}-\beta^0_{g,T})-2\frac{1}{\sqrt{N}}\sum_{g_i^0=g}\epsilon_{iT}x_{iT}+\sqrt{N}T\lambda\dot{w}_{g,T}e_T=0,$$

and there is no fourth term. We can apply the argument above and obtain $\Pr(\|\mathring{\theta}_{g,t}\| = 0) \to 1$.

$\square$

Lemma 10. *Suppose that Assumptions 1.3, 2.1, 1.5, and 4 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. It holds that, as $N \to \infty$,*

$$\Pr(\mathring{m}_g = m_g^0, \forall g \in \mathbb{G}) \to 1,$$

and

$$\Pr(\mathring{T}_{g,j} = T_{g,j}^0, \forall j \in \{1, \dots, m_g^0\}, g \in \mathbb{G} \mid \mathring{m}_g = m_g^0, \forall g \in \mathbb{G}) \to 1$$

Proof. The proof is based on an argument essentially identical to the proof of Corollary 3.4 in Qian and Su (2016) and is thus omitted. $\square$

Recall that $\mathring{\alpha}_{g,j} = \mathring{\beta}_{g,t}$ for $T_{g,j}^0 \le t < T_{g,j+1}^0$.

Lemma 11. *Suppose that Assumptions 1.3, 2.1, 1.5, 4, 5 and 6 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. Let A be a diagonal matrix whose diagonal elements are $(I_{1,1}, \dots, I_{1,m_1^0+1}, I_{2,1}, \dots, I_{2,m_2^0+1}, I_{3,1} \dots, I_{G-1,m_{G-1}^0+1}, I_{G,1}, \dots, I_{G,m_G^0+1})$. Let Π be a $\sum_{g=1}^G (m_g^0 + 1)k \times \sum_{g=1}^G (m_g^0 + 1)k$ block diagonal matrix whose g-th diagonal block is a $(m_g^0 + 1)k \times (m_g^0 + 1)k$ diagonal matrix whose diagonal elements are π_g.*

Conditional on $\mathring{m}_g = m_g^0$ for all $g \in \mathbb{G}$, we have, if $(\max_{g\in\mathbb{G}} m_g^0)^2/(I_{\min} \min_{g\in\mathbb{G}} N_g) \to 0$,

$$D\sqrt{N}A^{1/2}(\mathring{\alpha} - \mathring{\alpha}^0) \to_d N(0, D\Sigma_x^{-1}\Pi^{-1/2}\Omega\Pi^{-1/2}\Sigma_x^{-1}D').$$

Proof. We note that $\mathring{\alpha}_{g,j}$ satisfies the following FOC:

$$\frac{1}{NT} \sum_{g_i^0 = g} \sum_{t=T_{g,j}^0}^{T_{g,j+1}^0 - 1} (y_{it} - x_{it}'\mathring{\alpha}_{g,j})x_{it} + R_{g,j},$$

where $R_{g,1} = -\lambda \dot{w}_{g,T_{g,1}^0} e_{T_{g,1}^0}$, $R_{g,j} = \lambda(\dot{w}_{g,T_{g,j-1}^0} e_{T_{g,j-1}^0} - \dot{w}_{g,T_{g,j}^0} e_{T_{g,j}^0})$ for $2 \le j \le m_g^0$ and

$R_{g,m_g^0+1} = \lambda \dot{w}_{g,T^0_{m_g^0}} e_{T^0_{m_g^0}}$. We thus have

$$\mathring{\alpha}_{g,j} = \left(\sum_{g_i^0=g} \sum_{t=T^0_{g,j}}^{T^0_{g,j+1}-1} x_{it}x_{it}' \right)^{-1} \left(\sum_{g_i^0=g} \sum_{t=T^0_{g,j}}^{T^0_{g,j+1}-1} x_{it}y_{it} \right) + \left(\sum_{g_i^0=g} \sum_{t=T^0_{g,j}}^{T^0_{g,j+1}-1} x_{it}x_{it}' \right)^{-1} R_{g,j}.$$

Let

$$\hat{\Sigma}_{x,g,j} = \frac{1}{N_g}\frac{1}{I_{g,j}} \sum_{g_i^0=g} \sum_{t=T^0_{g,j}}^{T^0_{g,j+1}-1} x_{it}x_{it}',$$

$\hat{\Sigma}_{x,g}$ be a $(m_g^0+1)k \times (m_g^0+1)k$ block diagonal matrix whose t-th diagonal block is $\hat{\Sigma}_{x,g,j}$ and $\hat{\Sigma}_x$ be a $\sum_{g=1}^G (m_g^0+1)k \times \sum_{g=1}^G (m_g^0+1)k$ block diagonal matrix whose g-th diagonal block is $\hat{\Sigma}_{x,g}$. We can write

$$\sqrt{N} A^{1/2}(\mathring{\alpha} - \alpha^0) = (\hat{\Sigma}_x)^{-1}\Pi^{-1/2}(d_{1,NT}', \ldots, d_{G,NT}')' + (\hat{\Sigma}_x)^{-1}\Pi^{-1}N^{-1/2}R,$$

where R is a vector of $R_{g,j}$s. The first term converges to a normal distribution with the variance-covariance matrix specified in the statement of the theorem. We show that the second term is $o_p(1)$. We have

$$\begin{aligned}
&\|R\|^2 \\
\leq& \lambda^2 \sum_{g=1}^G \left(I_{g,1}^{-1} \left\| \dot{w}_{g,T^0_{g,1}} e_{g,T^0_{g,1}} \right\|^2 + \sum_{j=2}^{m_g^0} I_{g,j}^{-1} \left\| \dot{w}_{g,T^0_{g,j-1}} e_{g,T^0_{g,j-1}} \right\|^2 + I_{g,m_g^0+1}^{-1} \left\| \dot{w}_{g,T^0_{m_g^0,1}} e_{g,T^0_{m_g^0,1}} \right\|^2 \right) \\
\leq& 4 \sum_{g=1}^G (m_g^0+1)\lambda^2 I_{\min}^{-1} \max_{g\in\mathbb{G}, t\in\mathcal{T}^0_{m_g^0,g}} \|\dot{w}_{g,t}\|^2 \\
=& O_p\left(\sum_{g=1}^G (m_g^0)\lambda^2 I_{\min}^{-1} J_{\min}^{-2\kappa} \right).
\end{aligned}$$

By Assumptions 5 and 6, the second term is $o_p(1)$.

□

S.5 Models with individual-specific fixed effects

In this section, we derive the asymptotic properties of GAGFL in models with individual-specific fixed effects. This extension is discussed in Section 5 of the main text but we delegate the theoretical analysis to this supplement.

Recall that the model considered is

$$y_{it} = \mu_i + x'_{it}\beta_{g_i^0,t} + \epsilon_{it},$$

and we estimate the first-differenced model:

$$\Delta y_{it} = x'_{it}\beta_{g_i^0,t} - x'_{i,t-1}\beta_{g_i^0,t-1} + \Delta\epsilon_{it}.$$

The estimator is

$$(\hat{\beta}, \hat{\gamma}) = \underset{(\beta,\gamma)\in\mathcal{B}^{GT}\times\mathbb{G}^N}{\arg\min} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it} - x'_{it}\beta_{g_i,t} + x'_{i,t-1}\beta_{g_i,t-1})^2 + \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t} - \beta_{g,t-1}\right\|.$$

S.5.1 Assumptions

In this section, we present the assumptions. Note that they are similar to Assumptions 1–5, but we modify them for models with individual fixed effects. Additional assumptions are also made to use the results of Qian and Su (2016) directly.

Assumption 7.

1. *$\mathcal{B}$ is compact.*

2. *$E(\Delta\epsilon_{it}x_{it}) = E(\Delta\epsilon_{it}x_{i,t-1}) = 0$ for all i and t.*

3.

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=2}^{T}E\left(\Delta\epsilon_{it}\Delta\epsilon_{jt}x'_{it}x_{jt}\right)\right| < M$$

4.

$$\left|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{T}\sum_{t=2}^{T}\sum_{s=2}^{T}Cov\left(\Delta\epsilon_{it}\Delta\epsilon_{jt}x'_{it}x_{jt}, \Delta\epsilon_{is}\Delta\epsilon_{js}x'_{is}x_{js}\right)\right| < M$$

5. There exists $M > 0$ such that for any N and T,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E(\|x_{it}\|^4) < M.$$

This assumption except for the second item is satisfied under Assumption 1 in the main text. However, we still make this assumption to facilitate our theory. We note that Assumption 7.2 is different from Assumption 1.2.

Assumption 8.

1. Let

$$\begin{aligned} M^F(\gamma, g, \tilde{g}) =& \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}\mathbf{1}\left\{g_i = \tilde{g}\right\} \\ &\times \begin{pmatrix} x_{i1}x_{i1}' & -x_{i1}x_{i2}' & 0 & \dots & 0 \\ -x_{i2}x_{i1} & 2x_{i2}x_{i2}' & -x_{i2}x_{i3}' & \dots & \dots \\ 0 & -x_{i3}x_{i2} & \dots & \dots & 0 \\ \dots & \dots & \dots & 2x_{i,T-1}x_{i,T-1}' & -x_{i,T-1}x_{iT} \\ 0 & \dots & 0 & -x_{iT}x_{i,T-1} & x_{iT}x_{iT}' \end{pmatrix}. \end{aligned}$$

Let $\hat{\rho}^F(\gamma, g, \tilde{g})$ be the minimum eigenvalue of $M^F(\gamma, g, \tilde{g})$. There exits a $\hat{\rho}^F$ such that $\hat{\rho}^F \to_p \rho^F > 0$ and $\forall g$,

$$\min_{\gamma\in\mathbb{G}^N}\max_{\tilde{g}\in\mathbb{G}} \hat{\rho}^F(\gamma, g, \tilde{g}) > \hat{\rho}^F.$$

2. Let

$$D_{g\tilde{g}i}^F = \frac{1}{T}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(x_{it}'(\beta_{g,t}^0 - \beta_{\tilde{g},t}^0) - x_{i,t-1}'(\beta_{g,t-1}^0 - \beta_{\tilde{g},t-1}^0).\right)^2$$

For all $g \neq \tilde{g}$, there exists a $c_{g,\tilde{g}}^F > 0$ such that

$$\operatorname*{plim}_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N} D_{g\tilde{g}i}^F > c_{g,\tilde{g}}^F,$$

and for all i,

$$\operatorname*{plim}_{T\to\infty} D^F_{g\tilde{g}i} > c^F_{g,\tilde{g}}.$$

This assumption plays a role similar to Assumption 2 in the main text and it guarantees the identification of the parameters.

Assumption 9.

1. *There exists a constant M^*_{ex} such that as $N, T \to \infty$, for all $\delta > 0$,*

$$\sup_{1\le i\le N} \Pr\left(\frac{1}{T}\sum_{t=2}^{T} \|\Delta\epsilon_{it} x_{it}\|^2 \ge M^*_{ex}\right) = O(T^{-\delta}),$$

and

$$\sup_{1\le i\le N} \Pr\left(\frac{1}{T}\sum_{t=2}^{T} \|\Delta\epsilon_{it} x_{i,t-1}\|^2 \ge M^*_{ex}\right) = O(T^{-\delta}).$$

2. *There exists a constant M^*_{x} such that as $N, T \to \infty$, for all $\delta > 0$,*

$$\sup_{1\le i\le N} \Pr\left(\frac{1}{T}\sum_{t=1}^{T} \|x_{it}\|^4 \ge M^*_{x}\right) = O(T^{-\delta}).$$

3. *There exist constants $a > 0$ and $d_1 > 0$ and a sequence $\alpha[t] < \exp(-at^{d_1})$ such that, for all $i = 1, \dots, N$ and $(g, \tilde{g}) \in \mathbb{G}^2$ such that $g \neq \tilde{g}$, $\{x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t}) - x_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1})\}_t$, $\{(x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t}) - x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1}))\Delta\epsilon_{it}\}_t$ are strongly mixing process with mixing coefficients $\alpha[t]$. Moreover, $E(x_{it}\Delta\epsilon_{it}) = 0$ and $E(x_{i,t-1}\Delta\epsilon_{it}) = 0$.*

4. *There exist constants $b_x > 0$, $b_e > 0$, $d_{2x} > 0$ and d_{2e} such that $\Pr(|x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t}) - x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1})| > m) \le \exp(1 - (m/b_x))^{d_{2x}}$ and $\Pr(|(x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t}) - x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1}))\Delta\epsilon_{it}| > m) \le \exp(1 - (m/b_e))^{d_{2e}}$, for any i, t and $m > 0$.*

This assumption is similar to Assumption 3 in the main text. It imposes restrictions on the tail properties of the variables and guarantees that clustering errors are small. Assumptions 7–9 are needed for accurate group estimation in the presence of individual fixed effects.

For break detection, it is not necessary to introduce new assumptions on breaks since Assumptions 4 and 6 in the main text are also sufficient here. Nevertheless, to directly apply the results of Qian and Su (2016) for AGFL in the presence of individual fixed effects, we introduce the following new assumption.

Assumption 10.

1. *$\{(x_{i1}, \epsilon_{i1}), \ldots, (x_{iT}, \epsilon_{iT})\}$'s are independent over i.*

2. *$\max_{1\leq i\leq N} \max_{1\leq t\leq T} E(\|x_{it}\|^{2\tau_0}) < C < \infty$ and $\max_{1\leq i\leq N} \max_{1\leq t\leq T} E(\|\epsilon_{it}\|^{2\tau_0}) < C < \infty$ for some C and $\tau_0 \geq 2$.*

3. *For τ_0 that satisfies condition 2, there exists $\varepsilon_0 > 0$ such that $N^{1-\tau_0}T(\ln T)^{\epsilon_0\tau_0} \to 0$ as $N, T \to \infty$.*

Lastly, we make an assumption to derive the asymptotic distribution of GAGFL. Let $\mathrm{TriD}(A, D)_M$ denote the symmetric block tridiagonal matrix defined by A and D:

$$\mathrm{TriD}(A, D)_M = \begin{pmatrix} D_1 & -A_2' & & & & & \\ -A_2 & D_2 & -A_3' & & & & \\ & -A_3 & D_3 & -A_4' & & & \\ & \ddots & \ddots & \ddots & & & \\ & & & & -A_{M-1} & D_{M-1} & -A_M' \\ & & & & & -A_M & D_M \end{pmatrix},$$

where D_m's are symmetric, A_m's are square matrices and the blank blocks are zero matrices. Let

$$\Phi^{\dagger}_{g,l+1} = \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_l} x'_{i,T_l-1},$$

for $l = 1, \ldots, m_g^0$. Define

$$\begin{aligned}
\Phi_{g,1} =& \frac{1}{N_g} \sum_{g_i^0=g} \sum_{t=1}^{T_1-1} \Delta x_{it} \Delta x'_{it} + \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_1-1} x'_{i,T_1-1}, \\
\Phi_{g,l} =& \frac{1}{N_g} \sum_{g_i^0=g} \sum_{t=T_{l-1}}^{T_l-1} \Delta x_{it} \Delta x'_{it} + \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_l-1} x'_{i,T_l-1} + \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_{l-1}} x'_{i,T_{l-1}},
\end{aligned}$$

$$
\begin{aligned}
&\text{for } l = 2, \ldots, m_g^0, \\
\Phi_{g,m_g^0+1} =& \frac{1}{N_g} \sum_{g_i^0=g} \sum_{t=T_{m_g^0}}^{T} \Delta x_{it} \Delta x_{it}' + \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_{m_g^0}} x_{i,T_{m_g^0}}'.
\end{aligned}
$$

Then we denote

$$
\hat{\Sigma}_{x,g} = \text{TriD}(\Phi_g^\dagger, \Phi_g)_{m_g^0+1},
$$

and let $\hat{\Sigma}_x$ be the block diagonal matrix whose g-th block is $\hat{\Sigma}_{x,g}$. Furthermore, define

$$
\begin{aligned}
\Psi_{g,1} =& \frac{1}{N_g} \sum_{g_i^0=g} \sum_{t=1}^{T_1-1} \Delta x_{it} \Delta \epsilon_{it}' - \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_1-1} \Delta \epsilon_{i,T_1}, \\
\Psi_{g,l} =& \frac{1}{N_g} \sum_{g_i^0=g} \sum_{t=T_{l-1}}^{T_l-1} \Delta x_{it} \Delta \epsilon_{it} - \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_l-1} \Delta \epsilon_{i,T_l} + \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_{l-1}} \Delta \epsilon_{i,T_{l-1}}, \\
&\text{for } l = 2, \ldots, m_g^0, \\
\Psi_{g,m_g^0+1} =& \frac{1}{N_g} \sum_{g_i^0=g} \sum_{t=T_{m_g^0}}^{T} \Delta x_{it} \Delta \epsilon_{it} + \frac{1}{N_g} \sum_{g_i^0=g} x_{i,T_{m_g^0}} \Delta \epsilon_{i,T_{m_g^0}}'.
\end{aligned}
$$

Then we denote

$$
\Psi_g = \left(\Psi_{g,1}', \ldots, \Psi_{g,m_g^0}' \right)',
$$

and let Ψ be the vector of Ψ_g's.

Assumption S.1. *Let A be a diagonal matrix whose diagonal elements are $(I_{1,1}, \ldots, I_{1,m_1^0+1}, I_{2,1}, \ldots, I_{2,m_2^0+1}, I_{3,1} \ldots, I_{G-1,m_{G-1}^0+1}, I_{G,1}, \ldots, I_{G,m_G^0+1})$. There exists Σ_x such that the spectral norm of $A^{-1/2}\hat{\Sigma}_x A^{-1/2} - \Sigma_x$ converges to 0 in probability. $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. Let Π be a $\sum_{g=1}^G (m_g^0+1)k \times \sum_{g=1}^G (m_g^0+1)k$ block diagonal matrix whose g-th diagonal block is a $(m_g^0+1)k \times (m_g^0+1)k$ diagonal matrix with elements being π_g. For a $l \times \sum_{g=1}^G (m_g^0+1)k$ matrix D, where l does not depend on T, we have*

$$
\sqrt{N} D \Sigma_x^{-1} \Pi^{-1/2} \Psi \to_d N(0, \Omega_D),
$$

for some positive definite matrix Ω_D.

While this condition may look complicated, it simply states that the standard assumptions for least squares are satisfied for each group and each span of periods between two breaks.

S.5.2 Asymptotic results

Under these new sets of assumptions, we show that the GAGFL estimator applied to first differenced models has asymptotic properties similar to those in the case of level models without individual specific intercepts. The lemmas and theorems are compactly presented in this subsection, while their proofs are delegated in the next subsection S.5.3.

We first show that the estimation error in group structure has a limited impact on the estimation of the coefficients. Let

$$\mathring{\beta} = \arg\min_{\beta \in \mathcal{B}^{GT}} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{it} - x'_{it}\beta_{g_i^0,t} + x'_{i,t-1}\beta_{g_i^0,t-1})^2 + \lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{g,t} \left\| \beta_{g,t} - \beta_{g,t-1} \right\| \right).$$

Note that $\mathring{\beta}$ is the estimator of β when the group memberships (i.e., γ^0) are known. With the assumptions stated above, Lemma 2 states that the difference between $\hat{\beta}$ and $\mathring{\beta}$ is small:

Lemma 2. *Suppose that Assumptions 7, 8 and 9 are satisfied. As $N, T \to \infty$, for any $\delta > 0$, it holds that*

$$\hat{\beta}_{g,t} = \mathring{\beta}_{g,t} + o_p(T^{-\delta}),$$

for all g and t.

Note that δ in the theorem can be arbitrary large.

The following theorem states that our method consistently spots dates in which there is no break. Let $\hat{\theta}_{g,t} = \hat{\beta}_{g,t} - \hat{\beta}_{g,t-1}$.

Theorem S.1. *Suppose that Assumptions 4, 7, 8, 9 and 10 hold. It follows that*

$$\Pr\left(\left\| \hat{\theta}_{g,t} \right\| = 0, \forall t \in \mathcal{T}^{0c}_{m_g^0,g}, g \in \mathbb{G} \right) \to 1,$$

as $N, T \to \infty$ with $N/T^{\delta} \to 0$ for some δ.

This theorem requires that $N/T^{\delta} \to 0$ for some δ. Since δ is arbitrary, this condition is satisfied as long as N is of geometric order of T but it does not hold if N is of exponential order of T. Note that the probability in this theorem is unconditional in the sense that it does not depend on knowing the true group membership.

Next, we present the follow theorem that states that the number of breaks and break dates are consistently estimated.

Theorem S.2. *Suppose that Assumptions 4, 7, 8, 9 and 10 hold. It holds that, as* $N, T \to \infty$ *with* $N/T^\delta \to 0$ *for some* $\delta > 0$,

$$\Pr(\hat{m}_g = m_g^0, \forall g \in \mathbb{G}) \to 1,$$

and

$$\Pr(\hat{T}_{g,j} = T_{g,j}^0, \forall j \in \{1, \ldots, m_g^0\}, g \in \mathbb{G} \mid \hat{m}_g = m_g^0, \forall g \in \mathbb{G}) \to 1.$$

Lastly, we present the asymptotic distribution of $\hat{\beta}$. The asymptotic distribution is the same as that of the least squares estimator with known group memberships and known breaks. To state the theorem, we introduce the following notation. Denote N_g as the number of individual units in group g and $I_{g,j}$ as the number of time observations between $T_{g,j}^0$ and $T_{g,j+1}^0$ for group g. The asymptotic distribution is given in the following theorem.

Theorem S.3. *Suppose that Assumptions 4, 6, 7, 8, 9, 10 and S.1 hold. Conditional on* $\hat{m}_g = m_g^0$ *for all* $g \in \mathbb{G}$, *we have, if* $(\max_{g\in\mathbb{G}} m_g^0)^2/(I_{\min} \min_{g\in\mathbb{G}} N_g) \to 0$,

$$D\sqrt{N}A^{1/2}(\hat{\alpha} - \alpha^0) \to_d N(0, \Omega_D).$$

S.5.3 Proofs for the results in Section S.5.2

We present the proofs of the results in Section S.5.2. The structure of the proofs is similar to the technical appendix attached to the main text. We first consider the GFE-type estimator in S.5.3.1. Then we examine the GAGFL estimator with known group membership in S.5.3.2. Note that this is just an application of the AGFL estimator in Qian and Su (2016) to each group and we rely on their analysis. Lastly, asymptotic properties of GAGFL are discussed in S.5.3.3.

S.5.3.1 Asymptotic properties of the GFE-type estimator

The argument given in this section are similar to those in Bonhomme and Manresa (2015).

Let

$$\dot{Q}_{NT}(\beta,\gamma) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}(\Delta y_{it} - x'_{it}\beta_{g_i,t} + x'_{i,t-1}\beta_{g_i,t-1})^2$$

and

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

As in the case of the previous section, the Cauchy-Schwarz inequality is abbreviated as the CS inequality and M denotes a generic universal constant.

Lemma S.1. *Suppose that Assumptions 7.1–4 hold.*

$$\sup_{(\beta,\gamma)\in\mathcal{B}^{GT}\times\Gamma_G}\left|\dot{Q}_{NT}(\beta,\gamma) - \tilde{Q}_{NT}(\beta,\gamma)\right| = o_p(1).$$

Proof. We observe that

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

We have

$$\begin{aligned}&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\epsilon_{it}x'_{it}\beta_{g_i,t}\\ =&\sum_{g=1}^{G}\frac{1}{NT}\mathbf{1}\{g_i = g\}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\epsilon_{it}x'_{it}\beta_{g,t}\\ =&\sum_{g=1}^{G}\frac{1}{T}\sum_{t=2}^{T}\beta'_{g,t}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i = g\}\Delta\epsilon_{it}x_{it}.\end{aligned}$$

For any $g \in \mathbb{G}$, the CS inequality implies that

$$\begin{aligned}&\left(\frac{1}{T}\sum_{t=2}^{T}\beta_{g,t}'\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\Delta\epsilon_{it}x_{it}\right)^2\\ \leq&\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta_{g,t}'\right\|^2\right)\times\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\Delta\epsilon_{it}x_{it}\right\|^2\right).\end{aligned}$$

Assumption 7.1 implies that

$$\frac{1}{T}\sum_{t=2}^{T}\left\|\beta_{g,t}'\right\|^2 < M.$$

We also have

$$\begin{aligned}&\frac{1}{T}\sum_{t=2}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\Delta\epsilon_{it}x_{it}\right\|^2\\ =&\frac{1}{T}\sum_{t=2}^{T}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}\{g_i=g\}\mathbf{1}\{g_j=g\}\Delta\epsilon_{it}\Delta\epsilon_{jt}x_{it}'x_{jt}\\ =&\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}\{g_i=g\}\mathbf{1}\{g_j=g\}\frac{1}{T}\sum_{t=2}^{T}\Delta\epsilon_{it}\Delta\epsilon_{jt}x_{it}'x_{jt}\\ \leq&\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=2}^{T}\Delta\epsilon_{it}\Delta\epsilon_{jt}x_{it}'x_{jt}\right|\\ \leq&\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=2}^{T}E\left(\Delta\epsilon_{it}\Delta\epsilon_{jt}x_{it}'x_{jt}\right)\right|\\ &+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=2}^{T}\left(\Delta\epsilon_{it}\Delta\epsilon_{jt}x_{it}'x_{jt}-E\left(\Delta\epsilon_{it}\Delta\epsilon_{jt}x_{it}'x_{jt}\right)\right)\right|\\ =&o_p(1),\end{aligned}$$

where the last equality follows by Assumptions 7.3 and 7.4. Thus, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\epsilon_{it}x_{it}'\beta_{g_i,t}=o_p(1)$$

uniformly over over $\mathcal{B}^{GT} \times \Gamma_G$. Similarly we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\epsilon_{it}x'_{it}\beta^0_{g_i^0,t} = o_p(1),$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\epsilon_{it}x'_{i,t-1}\beta^0_{g_i^0,t-1} = o_p(1),$$

and

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\epsilon_{it}x'_{it}\beta_{g_i,t-1} = o_p(1).$$

Therefore, we have the desired result. □

We consider the following Hausdorff distance in $\mathcal{B}^{GT}$ such that

$$d_H(\beta^a, \beta^b) = \max\left\{\max_{g\in\mathbb{G}}\left(\min_{\tilde{g}\in\mathbb{G}}\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^a_{\tilde{g},t} - \beta^b_{g,t}\right\|^2\right), \max_{\tilde{g}\in\mathbb{G}}\left(\min_{g\in\mathbb{G}}\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^a_{\tilde{g},t} - \beta^b_{g,t}\right\|^2\right)\right\}.$$

Lemma S.2. *Suppose that Assumptions 7.1–4 and 8 hold.*

$$d_H(\dot{\beta}, \beta^0) = o_p(1).$$

Proof. By Lemma S.1, we have

$$\tilde{Q}(\dot{\beta}, \dot{\gamma}) = \dot{Q}(\dot{\beta}, \dot{\gamma}) + o_p(1) \le \dot{Q}(\beta^0, \gamma^0) + o_p(1) = \tilde{Q}(\beta^0, \gamma^0) + o_p(1).$$

Because $\tilde{Q}(\beta, \gamma)$ is minimized at $\beta = \beta^0$ and $\gamma = \gamma^0$, we have

$$\tilde{Q}(\dot{\beta}, \dot{\gamma}) - \tilde{Q}(\beta^0, \gamma^0) = o_p(1).$$

On the other hand, we have

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

$$
\begin{aligned}
&=\sum_{g=1}^{G}\sum_{\tilde g=1}^{G}\frac{1}{T}\left(\beta_g^0-\beta_{\tilde g}\right)' M^F(\gamma,g,\tilde g)\left(\beta_g^0-\beta_{\tilde g}\right)\\
&\geq\sum_{g=1}^{G}\sum_{\tilde g=1}^{G}\hat\rho^F(\gamma,g,\tilde g)\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta_{g,t}^0-\beta_{\tilde g,t}\right\|^2\right)\\
&\geq\sum_{g=1}^{G}\max_{\tilde g\in\mathbb{G}}\hat\rho^F(\gamma,g,\tilde g)\min_{\tilde g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta_{g,t}^0-\beta_{\tilde g,t}\right\|^2\right)\\
&\geq\sum_{g=1}^{G}\hat\rho^F\min_{\tilde g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta_{g,t}^0-\beta_{\tilde g,t}\right\|^2\right)\\
&\geq\hat\rho^F\max_{g\in\mathbb{G}}\left(\min_{\tilde g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta_{g,t}^0-\beta_{\tilde g,t}\right\|^2\right)\right).
\end{aligned}
$$

Note that $\hat\rho^F$ is asymptotically bounded away from zero by Assumption 8.1.

Therefore we have

$$
\max_{g\in\mathbb{G}}\left(\min_{\tilde g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta_{g,t}^0-\dot\beta_{\tilde g,t}\right\|^2\right)\right)=o_p(1). \tag{S.7}
$$

Let

$$
\sigma(g)=\arg\min_{\tilde g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta_{g,t}^0-\dot\beta_{\tilde g,t}\right\|^2\right).
$$

Then we have for $\tilde g\neq g$,

$$
\begin{aligned}
&\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(x_{it}'(\dot\beta_{\sigma(g),t}-\dot\beta_{\sigma(\tilde g),t})-x_{i,t-1}'(\dot\beta_{\sigma(g),t-1}-\dot\beta_{\sigma(\tilde g),t-1})\right)^2\right)^{1/2}\\
\geq&\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(x_{it}'(\beta_{g,t}^0-\beta_{\tilde g,t}^0)-x_{i,t-1}'(\beta_{g,t-1}^0-\beta_{\tilde g,t-1}^0)\right)^2\right)^{1/2}\\
&-\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(x_{it}'(\dot\beta_{\sigma(g),t}-\beta_{g,t}^0)-x_{i,t-1}'(\dot\beta_{\sigma(g),t-1}-\beta_{g,t-1}^0)\right)^2\right)^{1/2}\\
&-\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(x_{it}'(\dot\beta_{\sigma(\tilde g),t}-\beta_{\tilde g,t}^0)-x_{i,t-1}'(\dot\beta_{\sigma(\tilde g),t-1}-\beta_{\tilde g,t-1}^0)\right)^2\right)^{1/2}.
\end{aligned}
$$

Assumption 8.2 states that the first term in the right hand side of the inequality is bounded away from zero. Equation (S.7) implies that the second and third terms are $o_p(1)$. Therefore, we have $\sigma(g) \neq \sigma(\tilde{g})$ with probability approaching one, which implies that with probability approaching one σ is bijective and has the inverse σ^{-1}. Thus, we have

$$\begin{aligned}
&\min_{g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^0_{g,t}-\dot{\beta}_{\tilde{g},t}\right\|^2\right)\\
\geq&\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^0_{\sigma^{-1}(\tilde{g}),t}-\dot{\beta}_{\tilde{g},t}\right\|^2\right)\\
=&\min_{h\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^0_{\sigma^{-1}(\tilde{g}),t}-\dot{\beta}_{h,t}\right\|^2\right)=o_p(1),
\end{aligned}$$

where the last equality follows by (S.7). Therefore we have

$$\max_{\tilde{g}\in\mathbb{G}}\left(\min_{g\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^0_{g,t}-\dot{\beta}_{\tilde{g},t}\right\|^2\right)\right)=o_p(1).$$

We thus have the desired result. $\square$

The proof of Lemma S.2 shows that there exists a permutation σ such that

$$\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^0_{\sigma(g),t}-\dot{\beta}_{g,t}\right\|^2=o_p(1).$$

We take $\sigma(g)=g$ by relabeling.

Define

$$\mathcal{N}_\eta=\left\{\beta\in\mathcal{B}^{GT}:\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^0_{g,t}-\beta_{g,t}\right\|^2<\eta,\forall g\in\mathbb{G}\right\}.$$

Let

$$\hat{g}_i(\beta)=\arg\min_{g\in\mathbb{G}}\sum_{t=2}^{T}(\Delta y_{it}-x'_{it}\beta_{g,t}+x'_{i,t-1}\beta_{g,t-1})^2. \tag{S.8}$$

Lemma S.3. *Suppose that Assumptions 8.2 and 9 are satisfied. For $\eta>0$ small enough,* *we have,* $\forall \delta > 0$,

$$\sup_{\beta \in \mathcal{N}_\eta} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left\{\hat{g}_i(\beta) \neq g_i^0\right\} = o_p(T^{-\delta}).$$

Proof. For any $g \in \mathbb{G}$, we have

$$\mathbf{1}\left\{\hat{g}_i(\beta) = g\right\} \leq \mathbf{1}\left\{\sum_{t=2}^{T}(\Delta y_{it} - x_{it}'\beta_{g,t} + x_{i,t-1}'\beta_{g,t-1})^2 \leq \sum_{t=2}^{T}(\Delta y_{it} - x_{it}'\beta_{g_i^0,t} + x_{i,t-1}'\beta_{g_i^0.t-1})^2\right\}.$$

Thus, we have

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\beta) \neq g_i^0\right\} =& \sum_{g=1}^{G}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0 \neq g\right\}\mathbf{1}\left\{\hat{g}_i(\beta) = g\right\} \\
\leq& \sum_{g=1}^{G}\frac{1}{N}\sum_{i=1}^{N} Z_{ig}(\beta),
\end{aligned}$$

where

$$Z_{ig}(\beta) = \mathbf{1}\left\{g_i^0 \neq g\right\}\mathbf{1}\left\{\sum_{t=2}^{T}(\Delta y_{it} - x_{it}'\beta_{g,t} + x_{i,t-1}'\beta_{g,t-1})^2 \leq \sum_{t=2}^{T}(\Delta y_{it} - x_{it}'\beta_{g_i^0,t} + x_{i,t-1}'\beta_{g_i^0,t-1})^2\right\}.$$

We now bound $Z_{ig}(\beta)$. Let

$$\begin{aligned}
z_{it}(g,\tilde{g}) =& (x_{it}'(\beta_{\tilde{g},t} - \beta_{g,t}) - x_{i,t-1}'(\beta_{\tilde{g},t-1} - \beta_{g,t-1})) \\
&\times\left(x_{it}'\beta_{\tilde{g},t}^0 - x_{i,t-1}'\beta_{\tilde{g},t-1}^0 + \Delta\epsilon_{it} - \frac{x_{it}'(\beta_{\tilde{g},t} + \beta_{g,t}) - x_{i,t-1}'(\beta_{\tilde{g},t-1} + \beta_{g,t-1})}{2}\right).
\end{aligned}$$

We have

$$\begin{aligned}
Z_{ig}(\beta) \leq& \mathbf{1}\left\{g_i^0 \neq g\right\} \times \mathbf{1}\left\{\sum_{t=2}^{T} z_{it}(g, g_i^0) \leq 0\right\} \\
\leq& \max_{\tilde{g} \in \mathbb{G}\setminus\{g\}} \mathbf{1}\left\{\sum_{t=2}^{T} z_{it}(g,\tilde{g}) \leq 0\right\}.
\end{aligned}$$

Let

$$z_{it}^0(g,\tilde{g}) = (x_{it}'(\beta_{\tilde{g},t}^0 - \beta_{g,t}^0) - x_{i,t-1}'(\beta_{\tilde{g},t-1}^0 - \beta_{g,t-1}^0))$$

$$\times \left(x'_{it}\beta^0_{\tilde{g},t} - x'_{i,t-1}\beta^0_{\tilde{g},t-1} + \Delta\epsilon_{it} - \frac{x'_{it}(\beta^0_{\tilde{g},t} + \beta^0_{g,t}) - x'_{i,t-1}(\beta^0_{\tilde{g},t-1} + \beta^0_{g,t-1})}{2} \right).$$

Let

$$A_T = \left| \sum_{t=2}^{T} (z_{it}(g,\tilde{g}) - z^0_{it}(g,\tilde{g})) \right|.$$

Then we have

$$\begin{aligned}
A_T \leq & \left| \sum_{t=2}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})\Delta\epsilon_{it} - \sum_{i=1}^{T} x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})\Delta\epsilon_{it} \right| \\
& + \left| \sum_{t=2}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1} - \beta_{g,t-1})\Delta\epsilon_{it} - \sum_{i=1}^{T} x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1})\Delta\epsilon_{it} \right| \\
& + \left| \sum_{t=2}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})x'_{it}\beta^0_{\tilde{g},t} - \sum_{i=1}^{T} x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})x'_{it}\beta^0_{\tilde{g},t} \right| \\
& + \left| \sum_{t=2}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1} - \beta_{g,t-1})x'_{it}\beta^0_{\tilde{g},t} - \sum_{i=1}^{T} x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1})x'_{it}\beta^0_{\tilde{g},t} \right| \\
& + \left| \sum_{t=2}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})x'_{i,t-1}\beta^0_{\tilde{g},t-1} - \sum_{i=1}^{T} x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})x'_{i,t-1}\beta^0_{\tilde{g},t-1} \right| \\
& + \left| \sum_{t=2}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1} - \beta_{g,t-1})x'_{i,t-1}\beta^0_{\tilde{g},t-1} - \sum_{i=1}^{T} x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1})x'_{i,t-1}\beta^0_{\tilde{g},t-1} \right| \\
& + \left| \sum_{t=2}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})\frac{x'_{it}(\beta_{\tilde{g},t} + \beta_{g,t})}{2} - \sum_{i=1}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})\frac{x'_{it}(\beta^0_{\tilde{g},t} + \beta^0_{g,t})}{2} \right| \\
& + \left| \sum_{t=2}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})\frac{x'_{it}(\beta^0_{\tilde{g},t} + \beta^0_{g,t})}{2} - \sum_{i=1}^{T} x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})\frac{x'_{it}(\beta^0_{\tilde{g},t} + \beta^0_{g,t})}{2} \right| \\
& + \left| \sum_{t=2}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1} - \beta_{g,t-1})\frac{x'_{it}(\beta_{\tilde{g},t} + \beta_{g,t})}{2} - \sum_{i=1}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1} - \beta_{g,t-1})\frac{x'_{it}(\beta^0_{\tilde{g},t} + \beta^0_{g,t})}{2} \right| \\
& + \left| \sum_{t=2}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1} - \beta_{g,t-1})\frac{x'_{it}(\beta^0_{\tilde{g},t} + \beta^0_{g,t})}{2} - \sum_{i=1}^{T} x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1})\frac{x'_{it}(\beta^0_{\tilde{g},t} + \beta^0_{g,t})}{2} \right| \\
& + \left| \sum_{t=2}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})\frac{x'_{i,t-1}(\beta_{\tilde{g},t-1} + \beta_{g,t-1})}{2} - \sum_{i=1}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})\frac{x'_{i,t-1}(\beta^0_{\tilde{g},t-1} + \beta^0_{g,t-1})}{2} \right| \\
& + \left| \sum_{t=2}^{T} x'_{it}(\beta_{\tilde{g},t} - \beta_{g,t})\frac{x'_{i,t-1}(\beta^0_{\tilde{g},t-1} + \beta^0_{g,t-1})}{2} - \sum_{i=1}^{T} x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t})\frac{x'_{i,t-1}(\beta^0_{\tilde{g},t-1} + \beta^0_{g,t-1})}{2} \right|
\end{aligned}$$

$$+\left|\sum_{t=2}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1}-\beta_{g,t-1})\frac{x'_{i,t-1}(\beta_{\tilde{g},t-1}+\beta_{g,t-1})}{2}-\sum_{i=1}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1}-\beta_{g,t-1})\frac{x'_{i,t-1}(\beta^0_{\tilde{g},t-1}+\beta^0_{g,t-1})}{2}\right|$$
$$+\left|\sum_{t=2}^{T} x'_{i,t-1}(\beta_{\tilde{g},t-1}-\beta_{g,t-1})\frac{x'_{i,t-1}(\beta^0_{\tilde{g},t-1}+\beta^0_{g,t-1})}{2}-\sum_{i=1}^{T} x'_{i,t-1}(\beta^0_{\tilde{g},t-1}-\beta^0_{g,t-1})\frac{x'_{i,t-1}(\beta^0_{\tilde{g},t-1}+\beta^0_{g,t-1})}{2}\right|.$$

Thus, when $\beta \in \mathcal{N}_\eta$, the CS inequality implies that

$$\begin{aligned}
A_T \leq & 2T\left(\frac{1}{T}\sum_{t=2}^{T}\|\Delta\epsilon_{it}x_{it}\|^2\right)^{1/2}\sqrt{\eta} \\
& + 2T\left(\frac{1}{T}\sum_{t=2}^{T}\|\Delta\epsilon_{it}x_{i,t-1}\|^2\right)^{1/2}\sqrt{\eta} \\
& + 2T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{it}\beta^0_{\tilde{g},t})^2\|x_{it}\|^2\right)^{1/2}\sqrt{\eta} \\
& + 2T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{i,t-1}\beta^0_{\tilde{g},t-1})^2\|x_{it}\|^2\right)^{1/2}\sqrt{\eta} \\
& + 2T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{it}\beta^0_{\tilde{g},t})^2\|x_{i,t-1}\|^2\right)^{1/2}\sqrt{\eta} \\
& + 2T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{i,t-1}\beta^0_{\tilde{g},t-1})^2\|x_{i,t-1}\|^2\right)^{1/2}\sqrt{\eta} \\
& + T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{it}(\beta_{\tilde{g},t}-\beta_{g,t}))^2\|x_{it}\|^2\right)^{1/2}\sqrt{\eta} \\
& + T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{i,t-1}(\beta_{\tilde{g},t-1}-\beta_{g,t-1}))^2\|x_{it}\|^2\right)^{1/2}\sqrt{\eta} \\
& + T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{it}(\beta_{\tilde{g},t}-\beta_{g,t}))^2\|x_{i,t-1}\|^2\right)^{1/2}\sqrt{\eta} \\
& + T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{i,t-1}(\beta_{\tilde{g},t-1}-\beta_{g,t-1}))^2\|x_{i,t-1}\|^2\right)^{1/2}\sqrt{\eta} \\
& + T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{it}(\beta^0_{\tilde{g},t}+\beta^0_{g,t}))^2\|x_{it}\|^2\right)^{1/2}\sqrt{\eta}
\end{aligned}$$

$$
\begin{aligned}
&+ T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{i,t-1}(\beta^0_{\tilde{g},t-1}+\beta^0_{g,t-1}))^2\left\|x_{it}\right\|^2\right)^{1/2}\sqrt{\eta}\\
&+ T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{it}(\beta^0_{\tilde{g},t}+\beta^0_{g,t}))^2\left\|x_{i,t-1}\right\|^2\right)^{1/2}\sqrt{\eta}\\
&+ T\left(\frac{1}{T}\sum_{t=2}^{T}(x'_{i,t-1}(\beta^0_{\tilde{g},t-1}+\beta^0_{g,t-1}))^2\left\|x_{i,t-1}\right\|^2\right)^{1/2}\sqrt{\eta}.
\end{aligned}
$$

Since $\mathcal{B}$ is bounded, we have, for $\beta \in \mathcal{N}_\eta$,

$$
\begin{aligned}
A_T \leq & C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\Delta\epsilon_{it}x_{it}\right\|^2\right)^{1/2}\\
&+ C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\Delta\epsilon_{it}x_{i,t-1}\right\|^2\right)^{1/2} + C_3\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\left\|x_{it}\right\|^4\right)^{1/2},
\end{aligned}
$$

where C_1, C_2 and C_3 are constants independent of η and T.

Thus, we have

$$
\begin{aligned}
Z_{ig}(\beta) \leq \max_{\tilde{g}\in\mathbb{G}\setminus\{g\}} \mathbf{1}\Big\{ &\sum_{t=2}^{T} z^0_{it}(g,\tilde{g}) \leq C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\Delta\epsilon_{it}x_{it}\right\|^2\right)^{1/2}\\
&+ C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\Delta\epsilon_{it}x_{i,t-1}\right\|^2\right)^{1/2} + C_3\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\left\|x_{it}\right\|^4\right)^{1/2}\Big\}.
\end{aligned}
$$

Let

$$
\begin{aligned}
\tilde{Z}_{ig} = \max_{\tilde{g}\in\mathbb{G}\setminus\{g\}} \mathbf{1}\Big\{ &\sum_{t=2}^{T} z^0_{it}(g,\tilde{g}) \leq C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\Delta\epsilon_{it}x_{it}\right\|^2\right)^{1/2}\\
&+ C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\Delta\epsilon_{it}x_{i,t-1}\right\|^2\right)^{1/2} + C_3\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\left\|x_{it}\right\|^4\right)^{1/2}\Big\}.
\end{aligned}
$$

Thus, we have

$$
\sup_{\beta\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\beta)\neq g^0_i\right\} \leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N} Z_{ig}(\beta) \leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}\tilde{Z}_{ig}.
$$

Note that $\tilde{Z}_{ig}$ does not depend on β.

We now bound $\Pr(\tilde{Z}_{ig} = 1)$. We observe

$$
\begin{aligned}
&\Pr(\tilde{Z}_{ig} = 1) \\
\leq &\sum_{\tilde{g} \in \mathbb{G} \setminus \{g\}} \Pr\Bigg(\sum_{t=2}^{T} z_{it}^{0}(g, \tilde{g}) \leq C_1 \sqrt{\eta} T \left(\frac{1}{T} \sum_{t=2}^{T} \|\Delta \epsilon_{it} x_{it}\|^2 \right)^{1/2} \\
&+ C_2 \sqrt{\eta} T \left(\frac{1}{T} \sum_{t=2}^{T} \|\Delta \epsilon_{it} x_{i,t-1}\|^2 \right)^{1/2} + C_3 \sqrt{\eta} T \left(\frac{1}{T} \sum_{t=1}^{T} \|x_{it}\|^4 \right)^{1/2} \Bigg).
\end{aligned}
$$

Let $M^* > \max(M_{ex}^*, M_x^*)$, where M_{ex}^* and M_x^* are defined in Assumptions 9.1 and 9.2, respectively. Thus, we have

$$
\begin{aligned}
&\Pr(\tilde{Z}_{ig} = 1) \\
\leq &\sum_{\tilde{g} \in \mathbb{G} \setminus \{g\}} \Bigg[\Pr\left(\frac{1}{T} \sum_{t=2}^{T} \|\Delta \epsilon_{it} x_{it}\|^2 \geq M^* \right) + \Pr\left(\frac{1}{T} \sum_{t=2}^{T} \|\Delta \epsilon_{it} x_{i,t-1}\|^2 \geq M^* \right) \\
&+ \Pr\left(\frac{1}{T} \sum_{t=1}^{T} \|x_{it}\|^4 \geq M^* \right) \\
&+ \Pr\left(\frac{1}{T} \sum_{t=2}^{T} \left(x_{it}'(\beta_{\tilde{g},t}^{0} - \beta_{g,t}^{0}) - x_{i,t-1}'(\beta_{\tilde{g},t-1}^{0} - \beta_{g,t-1}^{0}) \right)^2 \leq \frac{c_{g,\tilde{g}}^{F}}{2} \right) \\
&+ \Pr\left(\sum_{t=2}^{T} (x_{it}'(\beta_{\tilde{g},t}^{0} - \beta_{g,t}^{0}) - x_{i,t-1}'(\beta_{\tilde{g},t-1}^{0} - \beta_{g,t-1}^{0})) \Delta \epsilon_{it} \leq -T \frac{c_{g,\tilde{g}}^{F}}{4} + T C_3 \sqrt{\eta} \sqrt{M^*} \right) \Bigg],
\end{aligned}
$$

where C_3 is a constant independent of η and T. Assumptions 9.1 and 9.2 give that

$$
\begin{aligned}
&\Pr\left(\frac{1}{T} \sum_{t=2}^{T} \|\Delta \epsilon_{it} x_{it}\|^2 \geq M^* \right) + \Pr\left(\frac{1}{T} \sum_{t=2}^{T} \|\Delta \epsilon_{it} x_{i,t-1}\|^2 \geq M^* \right) + \Pr\left(\frac{1}{T} \sum_{t=1}^{T} \|x_{it}\|^4 \geq M^* \right) \\
=&O(T^{-\delta}).
\end{aligned}
$$

Under Assumptions 8.2, 9.3 and 9.4, a very similar argument to Bonhomme and Manresa (2015, page 1176) implies that

$$
\Pr\left(\frac{1}{T} \sum_{i=1}^{T} \left(x_{it}'(\beta_{\tilde{g},t}^{0} - \beta_{g,t}^{0}) - x_{i,t-1}'(\beta_{\tilde{g},t-1}^{0} - \beta_{g,t-1}^{0}) \right)^2 \leq \frac{c_{g,\tilde{g}}^{F}}{2} \right) = O(T^{-\delta}).
$$

Take η such that

$$\eta \leq \left(\frac{\min_{\tilde{g} \in \mathbb{G} \setminus \{g\}} c^F_{g,\tilde{g}}}{8C_3\sqrt{M^*}} \right).$$

We then have

$$\begin{aligned}
&\Pr\left(\sum_{t=2}^{T} (x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t}) - x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1}))\Delta\epsilon_{it} \leq -T\frac{c^F_{g,\tilde{g}}}{4} + TC_3\sqrt{\eta}\sqrt{M^*} \right) \\
\leq &\Pr\left(\frac{1}{T}\sum_{t=2}^{T} (x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t}) - x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1}))\Delta\epsilon_{it} \leq -\frac{c^F_{g,\tilde{g}}}{8} \right) \\
\leq &\Pr\left(\left| \frac{1}{T}\sum_{t=2}^{T} (x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t}) - x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1}))\Delta\epsilon_{it} \right| > \frac{c^F_{g,\tilde{g}}}{8} \right).
\end{aligned}$$

Under Assumptions 9.3 and 9.4, a very similar argument to Bonhomme and Manresa (2015, page 1177) implies that

$$\Pr\left(\left| \frac{1}{T}\sum_{t=2}^{T} (x'_{it}(\beta^0_{\tilde{g},t} - \beta^0_{g,t}) - x'_{i,t-1}(\beta^0_{\tilde{g},t-1} - \beta^0_{g,t-1}))\Delta\epsilon_{it} \right| > \frac{c^F_{g,\tilde{g}}}{8} \right) = O(T^{-\delta}).$$

We thus have

$$\Pr(\tilde{Z}_{ig} = 1) \leq (G-1)O(T^{-\delta}).$$

This implies that

$$\begin{aligned}
&E\left(\sup_{\beta \in \mathcal{N}_\eta} \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left\{ \hat{g}_i(\beta) \neq g^0_i \right\} \right) \\
\leq &\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N} E\left(\tilde{Z}_{ig}\right) \\
= &\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N} \Pr\left(\tilde{Z}_{ig} = 1\right) \\
= &G(G-1)O(T^{-\delta}) \\
= &O(T^{-\delta}).
\end{aligned}$$

The Markov inequality implies the desired result.

$\square$

Let

$$\check{\beta} = \arg \min_{\beta \in \mathcal{B}^{GT}} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{it} - x'_{it}\beta_{g_i^0,t} + x'_{i,t-1}\beta_{g_i^0,t-1})^2.$$

Note that $\check{\beta}$ is the estimator of β when the group memberships (i.e., γ^0) are known. Let

$$\check{Q}(\beta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{it} - x'_{it}\beta_{g_i^0,t} + x'_{i,t-1}\beta_{g_i^0,t-1})^2.$$

Note that $\check{Q}(\beta) = \dot{Q}(\beta, \gamma^0)$ and that $\check{\beta} = \arg\min_{\beta \in \mathcal{B}^{GT}} \check{Q}(\beta)$. We have the following result.

Lemma S.4. *(Qian and Su, 2016, Lemma B.2). Suppose that Assumptions 7, 8 and 10 are satisfied. As $N, T \to \infty$,*

$$\check{\beta}_{g,t} - \beta^0_{g,t} = O_p\left(\frac{1}{\sqrt{N}}\right).$$

for each $t = 1, 2, \ldots, T$ and for all $g \in \mathbb{G}$.

Let

$$\dot{Q}(\beta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{it} - x'_{it}\beta_{\hat{g}_i(\beta),t} + x'_{i,t-1}\beta_{\hat{g}_i(\beta),t-1})^2.$$

Note also that $\dot{Q}(\beta) = \dot{Q}(\beta, \hat{\gamma}(\beta))$ and that $\dot{\beta} = \arg\min_{\beta \in \mathcal{B}^{GT}} \dot{Q}(\beta)$

Lemma S.5. *Suppose that Assumptions 7, 8 and 9 are satisfied. As $N, T \to \infty$, for any $\delta > 0$, it holds that*

$$\dot{\beta}_{g,t} = \check{\beta}_{g,t} + o_p(T^{-\delta}),$$

for all g and t.

Proof. We first evaluate the difference between $\check{Q}(\beta)$ and $\dot{Q}(\beta)$. We note that

$$\begin{aligned}
&\check{Q}(\beta) - \dot{Q}(\beta) \\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbf{1}\{\hat{g}_i(\beta) \neq g_i^0\} \\
&\times \left((\Delta y_{it} - x'_{it}\beta_{g_i^0,t} + x'_{i,t-1}\beta_{g_i^0,t-1})^2 - (\Delta y_{it} - x'_{it}\beta_{\hat{g}_i(\beta),t} + x'_{i,t-1}\beta_{\hat{g}_i(\beta),t-1})^2 \right).
\end{aligned}$$

By Assumptions 7.1, 7.3 and 7.5 and Lemma S.3, we have, for sufficiently small η,

$$\sup_{\beta \in \mathcal{N}_\eta} \left| \check{Q}(\beta) - \dot{Q}(\beta) \right| = o_p(T^{-\delta}). \tag{S.9}$$

We then see that

$$\Pr\left(\left| \check{Q}(\dot{\beta}) - \dot{Q}(\dot{\beta}) \right| > \epsilon T^{-\delta} \right) \leq \Pr(\dot{\beta} \notin \mathcal{N}_\eta) + \Pr\left(\sup_{\beta \in \mathcal{N}_\eta} \left| \check{Q}(\beta) - \dot{Q}(\beta) \right| > \epsilon T^{-\delta} \right) = o(1),$$

by that $\dot{\beta}$ is consistent by Lemma S.2, and (S.9). Therefore we have

$$\check{Q}(\dot{\beta}) - \dot{Q}(\dot{\beta}) = o_p(T^{-\delta}).$$

Similarly, we have

$$\check{Q}(\check{\beta}) - \dot{Q}(\check{\beta}) = o_p(T^{-\delta}).$$

Next, we evaluate the difference between $\check{\beta}$ and $\dot{\beta}$. By the definition of $\check{\beta}$ and $\dot{\beta}$, we have

$$0 \leq \check{Q}(\dot{\beta}) - \check{Q}(\check{\beta}) = \dot{Q}(\dot{\beta}) - \dot{Q}(\check{\beta}) + o_p(T^{-\delta}) \leq o_p(T^{-\delta}).$$

Thus we have

$$\check{Q}(\dot{\beta}) - \check{Q}(\check{\beta}) = o_p(T^{-\delta}). \tag{S.10}$$

We observe that

$$\check{Q}(\dot{\beta}) - \check{Q}(\check{\beta})$$

$$
\begin{aligned}
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it}-x_{it}'\dot{\beta}_{g_i^0,t}+x_{i,t-1}'\dot{\beta}_{g_i^0,t-1})^2-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it}-x_{it}'\check{\beta}_{g_i^0,t}+x_{i,t-1}'\check{\beta}_{g_i^0,t-1})^2\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\big((\Delta y_{it}-x_{it}'\check{\beta}_{g_i^0,t}+x_{i,t-1}'\check{\beta}_{g_i^0,t-1}+x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t})-x_{i,t-1}(\check{\beta}_{g_i^0,t-1}-\dot{\beta}_{g_i^0,t-1}))^2\\
&-(\Delta y_{it}-x_{it}'\check{\beta}_{g_i^0,t}+x_{i,t-1}'\check{\beta}_{g_i^0,t-1})^2\big)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\big((x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t})-x_{i,t-1}(\check{\beta}_{g_i^0,t-1}-\dot{\beta}_{g_i^0,t-1}))^2\\
&+2(\Delta y_{it}-x_{it}'\check{\beta}_{g_i^0,t}+x_{i,t-1}'\check{\beta}_{g_i^0,t-1})(x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t})-x_{i,t-1}'(\check{\beta}_{g_i^0,t-1}-\dot{\beta}_{g_i^0,t-1}))\big)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t})-x_{i,t-1}(\check{\beta}_{g_i^0,t-1}-\dot{\beta}_{g_i^0,t-1}))^2,
\end{aligned}
$$

where the last equality holds because $\check{\beta}$ is an OLS estimator and satisfies

$$
\begin{aligned}
&\sum_{i=1}^{N}(\Delta y_{it}-x_{it}'\check{\beta}_{g_i^0,t}+x_{i,t-1}'\check{\beta}_{g_i^0,t-1})x_{it}\mathbf{1}\{t>1\}\\
&-\sum_{i=1}^{N}(\Delta y_{i,t+1}-x_{i,t+1}'\check{\beta}_{g_i^0,t}+x_{it}'\check{\beta}_{g_i^0,t})x_{it}\mathbf{1}\{t<T\}=0.
\end{aligned}
$$

Moreover, we have

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(x_{it}'(\check{\beta}_{g_i^0,t}-\dot{\beta}_{g_i^0,t})-x_{i,t-1}(\check{\beta}_{g_i^0,t-1}-\dot{\beta}_{g_i^0,t-1}))^2\\
=&\frac{1}{T}\sum_{g\in\mathbb{G}}(\check{\beta}_g-\dot{\beta}_g)'M^F(\gamma^0,g,g)(\check{\beta}_g-\dot{\beta}_g)\\
\geq&\hat{\rho}^F\frac{1}{T}\sum_{g\in\mathbb{G}}\left\|\dot{\beta}_g-\check{\beta}_g\right\|^2.
\end{aligned}
$$

We thus have, by (S.10) and Assumption 8.1,

$$
\frac{1}{T}\sum_{g\in\mathbb{G}}\left\|\dot{\beta}_g-\check{\beta}_g\right\|^2=o_p(T^{-\delta}).
$$

This implies that

$$
\left\|\dot{\beta}_{g,t}-\check{\beta}_{g,t}\right\|^2=o_p(T^{1-\delta})
$$

for any δ. Thus we have the desired result.

$\square$

We can now have the asymptotic distribution of $\dot{\beta}$.

Theorem S.4. *Suppose that Assumptions 7, 8, 9 and 10 are satisfied. As $N, T \to \infty$ with $N/T^a \to 0$ for some $a > 0$,*

$$\dot{\beta}_{g,t} - \beta^0_{g,t} = O_p\left(\frac{1}{\sqrt{N}}\right),$$

for each $t = 1, 2, \ldots, T$ and for all $g \in \mathbb{G}$.

Proof. The theorem follows by Lemmas S.4 and Lemma S.5.

$\square$

S.5.3.2 Asymptotic properties of the GAGFL estimator with known group membership

The results presented here are directly implied by the results of Qian and Su (2016).

Let

$$\mathring{Q}(\beta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it} - x'_{it}\beta_{g_i^0,t} + x'_{i,t-1}\beta_{g_i^0,t-1})^2 + \lambda \sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t} - \beta_{g,t-1}\right\|.$$

Note that $\mathring{Q}(\beta) = \hat{Q}(\beta, \gamma^0)$ and that $\mathring{\beta} = \arg\min_{\beta\in\mathcal{B}^{GT}} \mathring{Q}(\beta)$.

We derive the asymptotic distribution of $\mathring{\beta}$.

Lemma S.6. *Suppose that Assumptions 4.1, 7, 8 and 10 hold. We have, as $N, T \to \infty$,*

$$\frac{1}{T}\left\|\mathring{\beta}_g - \beta^0_g\right\|^2 = O_p\left(\frac{1}{N}\right)$$

for any $g \in \mathbb{G}$. We also have, as $N, T \to \infty$,

$$\mathring{\beta}_{g,t} - \beta^0_{g,t} = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Proof. The proof is almost the same as that of Theorem 3.2(i) in Qian and Su (2016).

$\square$

Let $\mathring{\theta}_{g,1} = \mathring{\beta}_{g,1}$ and $\mathring{\theta}_{g,t} = \mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1}$.

Lemma S.7. *Suppose that Assumptions 4, 7, 8 and 10 hold. It follows that*

$$\Pr\left(\left\|\mathring{\theta}_{g,t}\right\| = 0, \forall t \in \mathcal{T}^{0c}_{m^0_g,g}, g \in \mathbb{G}\right) \to 1$$

as $N \to \infty$.

Proof. The proof is almost the same as that of Theorem 3.3 in Qian and Su (2016). □

Lemma S.8. *Suppose that Assumptions 4, 7, 8 and 10 hold. It holds that, as $N \to \infty$,*

$$\Pr(\mathring{m}_g = m^0_g, \forall g \in \mathbb{G}) \to 1,$$

and

$$\Pr(\mathring{T}_{g,j} = T^0_{g,j}, \forall j \in \{1, \ldots, m^0\}, g \in \mathbb{G} \mid \mathring{m} = m^0_g, \forall g \in \mathbb{G}) \to 1$$

Proof. The proof is almost the same as that of Corollary 3.4 in Qian and Su (2016). □

We now obtain the asymptotic distribution of $\mathring{\beta}$. Let $\mathring{\alpha}_{g,j} = \mathring{\beta}_{g,t}$ for $T^0_{g,j} \leq t < T^0_{g,j+1}$.

Lemma S.9. *Suppose that Assumptions 4, 6, 7, 8 10 and S.1 hold. Let A be a diagonal matrix whose diagonal elements are $(I_{1,1}, \ldots, I_{1,m^0_1+1}, I_{2,1}, \ldots, I_{2,m^0_2+1}, I_{3,1} \ldots, I_{G-1,m^0_{G-1}+1}, I_{G,1}, \ldots, I_{G,m^0_G+1})$. Let Π be a $\sum_{g=1}^G (m^0_g + 1)k \times \sum_{g=1}^G (m^0_g + 1)k$ block diagonal matrix whose g-th diagonal block is a $(m^0_g + 1)k \times (m^0_g + 1)k$ diagonal matrix whose diagonal elements are $\sqrt{\pi_g}$.*

Conditional on $\mathring{m}_g = m^0_g$ for all $g \in \mathbb{G}$, we have, if $(\max_{g \in \mathbb{G}} m^0_g)^2 / (I_{\min} \min_{g \in \mathbb{G}} N_g) \to 0$,

$$D\sqrt{N} A^{1/2} (\mathring{\alpha} - \mathring{\alpha}^0) \to_d N(0, D\Sigma_x^{-1} \Pi^{-1/2} \Omega \Pi^{-1/2} \Sigma_x^{-1} D').$$

Proof. The proof of this lemma is similar to that of Theorem 3.5(i) in Qian and Su (2016) and is thus omitted. □

S.5.3.3 Asymptotic properties of GAGFL under unknown group membership

Let

$$\hat{Q}_{NT}(\beta,\gamma) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it} - x_{it}'\beta_{g_i,t} + x_{i,t-1}'\beta_{g_i,t-1})^2 + \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t}-\beta_{g,t-1}\right\|.$$

Let

$$\begin{aligned}\tilde{\hat{Q}}_{NT}(\beta,\gamma) =& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(x_{it}'(\beta^0_{g_i^0,t}-\beta_{g_i,t}) - x_{i,t-1}'(\beta^0_{g_i^0,t-1}-\beta_{g_i,t-1}))^2\\ &+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t}-\beta_{g,t-1}\right\|.\end{aligned}$$

Lemma S.10. *Suppose that Assumptions 7.1–4 hold.*

$$\sup_{(\beta,\gamma)\in\mathcal{B}^{GT}\times\Gamma_G}\left|\hat{Q}_{NT}(\beta,\gamma)-\tilde{\hat{Q}}_{NT}(\beta,\gamma)\right| = o_p(1).$$

Proof. Note that

$$\hat{Q}_{NT}(\beta,\gamma)-\tilde{\hat{Q}}_{NT}(\beta,\gamma) = \dot{Q}_{NT}(\beta,\gamma)-\tilde{Q}_{NT}(\beta,\gamma).$$

Lemma S.1 implies the desired result.

□

Lemma S.11. *Suppose that Assumptions 4.1, 7, 8 and 10 hold.*

$$d_H(\hat{\beta},\beta^0) = o_p(1).$$

Proof. By Lemma S.10, we have

$$\tilde{\hat{Q}}(\hat{\beta},\hat{\gamma}) = \hat{Q}(\hat{\beta},\hat{\gamma}) + o_p(1) \le \hat{Q}(\beta^0,\gamma^0)+o_p(1) = \tilde{\hat{Q}}(\beta^0,\gamma^0)+o_p(1).$$

Because $\tilde{Q}(\beta,\gamma)$ is minimized at $\beta=\beta^0$ and $\gamma=\gamma^0$, we have

$$\tilde{\hat{Q}}(\hat{\beta},\hat{\gamma}) - \tilde{\hat{Q}}(\beta^0,\gamma^0) = o_p(1).$$

On the other hand, we have

$$
\begin{aligned}
&\tilde{\hat{Q}}(\beta,\gamma)-\tilde{\hat{Q}}(\beta^0,\gamma^0)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(x_{it}'(\beta^0_{g_i^0,t}-\beta_{g_i,t})-x_{i,t-1}'(\beta^0_{g_i^0,t-1}-\beta_{g_i,t-1})\right)^2\\
&+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t}-\beta_{g,t-1}\right\|-\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta^0_{g,t}-\beta^0_{g,t-1}\right\|\\
=&\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\frac{1}{T}\left(\beta^0_g-\beta_{\tilde{g}}\right)'M^F(\gamma,g,\tilde{g})\left(\beta^0_g-\beta_{\tilde{g}}\right)\\
&+\lambda\sum_{g\in\mathbb{G}}\sum_{t\in\mathcal{T}^0_{m_g^0,g}}\dot{w}_{g,t}\left(\left\|\beta_{g,t}-\beta_{g,t-1}\right\|-\left\|\beta^0_{g,t}-\beta^0_{g,t-1}\right\|\right)\\
&+\lambda\sum_{t\in\mathcal{T}^{0c}_{m_g^0,g}}\dot{w}_{g,t}\left\|\beta_{g,t}-\beta_{g,t-1}\right\|.
\end{aligned}
$$

In the proof of Lemma S.2, we have shown that

$$
\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\frac{1}{T}\left(\beta^0_g-\beta_{\tilde{g}}\right)'M^F(\gamma,g,\tilde{g})\left(\beta^0_g-\beta_{\tilde{g}}\right)\geq\hat{\rho}^F\max_{g\in\mathbb{G}}\left(\min_{\tilde{g}\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=2}^{T}\left\|\beta^0_{g,t}-\beta_{\tilde{g},t}\right\|^2\right)\right).
$$

Note that $\hat{\rho}$ is asymptotically bounded away from zero by Assumption 8.1.

We have, by the Jensen, triangular and CS inequalities,

$$
\begin{aligned}
&\left|\lambda\sum_{g\in\mathbb{G}}\sum_{t\in\mathcal{T}^0_{m_g^0,g}}\dot{w}_{g,t}\left(\left\|\beta_{g,t}-\beta_{g,t-1}\right\|-\left\|\beta^0_{g,t}-\beta^0_{g,t-1}\right\|\right)\right|\\
\leq&\lambda\sum_{g\in\mathbb{G}}\sum_{t\in\mathcal{T}^0_{m_g^0,g}}\dot{w}_{g,t}\left\|\beta_{g,t}-\beta_{g,t-1}-(\beta^0_{g,t}-\beta^0_{g,t-1})\right\|\\
\leq&\lambda\max_{s\in\mathcal{T}^0_{m_g^0,g},g\in\mathbb{G}}(\dot{w}_{g,s})\sum_{t\in\mathcal{T}^0_{m_g^0,g}}\left\|\beta_{g,t}-\beta_{g,t-1}-(\beta^0_{g,t}-\beta^0_{g,t-1})\right\|\\
=&O_p\left(\lambda\left(\sum_{g\in\mathbb{G}}m_g^0\right)J_{\min}^{-\kappa}\right)=o_p(1).
\end{aligned}
$$

where the last equality follows by Assumptions 7.1 and 4.1.

Lastly note that

$$\lambda \sum_{t \in \mathcal{T}^{0c}_{m^0_g, g}} \dot{w}_{g,t} \left\| \beta_{g,t} - \beta_{g,t-1} \right\| \geq 0.$$

Therefore unless we have

$$\max_{g \in \mathbb{G}} \left(\min_{\tilde{g} \in \mathbb{G}} \left(\frac{1}{T} \sum_{t=2}^{T} \left\| \beta^0_{g,t} - \hat{\beta}_{\tilde{g},t} \right\|^2 \right) \right) = o_p(1),$$

$\tilde{\hat{Q}}(\beta, \gamma) - \tilde{\hat{Q}}(\beta^0, \gamma^0) < o_p(1)$ does not hold.

We then follow the argument made in in the proof of Lemma S.2 to obtain

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

We thus have the desired result.

□

As in the case for $\dot{\beta}$, the above result implies that there exists a permutation σ such that

$$\frac{1}{T} \sum_{t=2}^{T} \left\| \beta^0_{\sigma(g),t} - \hat{\beta}_{g,t} \right\|^2 = o_p(1)$$

and we take $\sigma(g) = g$ by relabeling.

We observe that given β, the second term of $\hat{Q}_{NT}(\beta, \gamma)$ does not affect the estimation of γ. Therefore, $\hat{g}_i(\beta)$ defined in (S.8) is also the estimate of g_i given β even if $\hat{Q}_{NT}(\beta, \gamma)$ is our objective function. It follows that Lemma S.3 can apply for the GAGFL estimator.

Let

$$\hat{Q}(\beta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{it} - x'_{it} \beta_{\hat{g}_i(\beta),t} + x'_{i,t-1} \beta_{\hat{g}_i(\beta),t-1})^2 + \lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{g,t} \left\| \beta_{g,t} - \beta_{g,t-1} \right\|.$$

Note also that $\hat{Q}(\beta) = \hat{Q}(\beta, \hat{\gamma}(\beta))$ and that $\hat{\beta} = \arg\min_{\beta \in \mathcal{B}^{GT}} \hat{Q}(\beta)$.

Proof of Lemma 2

Proof. We first evaluate the difference between $\mathring{Q}(\beta)$ and $\hat{Q}(\beta)$. Note that

$$\mathring{Q}(\beta) - \hat{Q}(\beta) = \check{Q}(\beta) - \dot{Q}(\beta).$$

Thus by the proof of Lemma S.5 implies that

$$\mathring{Q}(\hat{\beta}) - \hat{Q}(\hat{\beta}) = o_p(T^{-\delta}).$$

Similarly, we have

$$\mathring{Q}(\mathring{\beta}) - \hat{Q}(\mathring{\beta}) = o_p(T^{-\delta}).$$

Next, we evaluate the difference between $\mathring{\beta}$ and $\hat{\beta}$. By the definition of $\mathring{\beta}$ and $\hat{\beta}$, we have

$$0 \leq \mathring{Q}(\hat{\beta}) - \mathring{Q}(\mathring{\beta}) = \hat{Q}(\hat{\beta}) - \hat{Q}(\mathring{\beta}) + o_p(T^{-\delta}) \leq o_p(T^{-\delta}).$$

Thus we have

$$\mathring{Q}(\hat{\beta}) - \mathring{Q}(\mathring{\beta}) = o_p(T^{-\delta}).$$

We observe that

$$\begin{aligned}
&\mathring{Q}(\hat{\beta}) - \mathring{Q}(\mathring{\beta}) \\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it} - x_{it}'\hat{\beta}_{g_i^0,t} + x_{i,t-1}'\hat{\beta}_{g_i^0,t-1})^2 + \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\hat{\beta}_{g,t} - \hat{\beta}_{g,t-1}\right\| \\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it} - x_{it}'\mathring{\beta}_{g_i^0,t} + x_{i,t-1}'\mathring{\beta}_{g_i^0,t-1})^2 - \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1}\right\| \\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(x_{it}'(\mathring{\beta}_{g_i^0,t} - \hat{\beta}_{g_i^0,t}) - x_{i,t-1}(\mathring{\beta}_{g_i^0,t-1} - \hat{\beta}_{g_i^0,t-1}))^2 \\
&+2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it} - x_{it}'\hat{\beta}_{g_i^0,t} + x_{i,t-1}'\hat{\beta}_{g_i^0,t-1})(x_{it}'(\mathring{\beta}_{g_i^0,t} - \hat{\beta}_{g_i^0,t}) - x_{i,t-1}'(\mathring{\beta}_{g_i^0,t-1} - \hat{\beta}_{g_i^0,t-1}))
\end{aligned}$$

$$+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\hat{\beta}_{g,t}-\hat{\beta}_{g,t-1}\right\|-\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\right\|.$$

By the first order condition for $\mathring{\beta}_{g,t}$, we have

$$\begin{aligned}
&-2\frac{1}{NT}\sum_{g_i^0=g}(\Delta y_{it}-x_{it}'\mathring{\beta}_{g_i^0,t}+x_{i,t-1}'\mathring{\beta}_{g_i^0,t-1})x_{it}\mathbf{1}\{t>1\}\\
&+2\frac{1}{NT}\sum_{g_i^0=g}(\Delta y_{i,t+1}-x_{i,t+1}'\mathring{\beta}_{g_i^0,t}+x_{it}'\mathring{\beta}_{g_i^0,t})x_{it}\mathbf{1}\{t<T\}\\
&+\lambda\dot{w}_{g,t}e_{g,t}-\lambda\dot{w}_{g,t+1}e_{g,t+1}=0,
\end{aligned}$$

where $e_{g,1}=e_{g,T+1}=0$, for $2\leq t\leq T$, $e_{g,t}=(\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1})/\left\|\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\right\|$ if $\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\neq 0$ and $\|e_{g,t}\|\leq 1$ otherwise. We thus have

$$\begin{aligned}
&2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\Delta y_{it}-x_{it}'\mathring{\beta}_{g_i^0,t}+x_{i,t-1}'\mathring{\beta}_{g_i^0,t-1})(x_{it}'(\mathring{\beta}_{g_i^0,t}-\hat{\beta}_{g_i^0,t})-x_{i,t-1}'(\mathring{\beta}_{g_i^0,t-1}-\hat{\beta}_{g_i^0,t-1}))\\
=&\lambda\sum_{g\in\mathbb{G}}\sum_{t=1}^{T}(\dot{w}_{g,t}e_{g,t}-\dot{w}_{g,t+1}e_{g,t+1})'(\mathring{\beta}_{g_i^0,t}-\hat{\beta}_{g_i^0,t})\\
=&\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}e_{g,t}'((\mathring{\beta}_{g_i^0,t}-\mathring{\beta}_{g_i^0,t-1})-(\hat{\beta}_{g_i^0,t}-\hat{\beta}_{g_i^0,t-1})).
\end{aligned}$$

Let $\mathcal{T}_{m_g,g}$ be the set of t such that $\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\neq 0$ and $\mathcal{T}_{m_g,g}^c=\{2,\ldots,T\}\backslash\mathcal{T}_{m_g,g}$.

We have

$$\begin{aligned}
&\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}e_{g,t}'((\mathring{\beta}_{g_i^0,t}-\mathring{\beta}_{g_i^0,t-1})-(\hat{\beta}_{g_i^0,t}-\hat{\beta}_{g_i^0,t-1}))\\
&+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\hat{\beta}_{g,t}-\hat{\beta}_{g,t-1}\right\|-\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\right\|\\
=&\lambda\sum_{g\in\mathbb{G}}\sum_{t\in\mathcal{T}_{m_g,g}^c}\dot{w}_{g,t}\left(\left\|\hat{\beta}_{g,t}-\hat{\beta}_{g,t-1}\right\|-e_{g,t}'(\hat{\beta}_{g,t}-\hat{\beta}_{g,t-1})\right)\\
&+\lambda\sum_{g\in\mathbb{G}}\sum_{t\in\mathcal{T}_{m_g,g}}\dot{w}_{g,t}\left(\left\|\hat{\beta}_{g,t}-\hat{\beta}_{g,t-1}\right\|-\frac{(\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1})'(\hat{\beta}_{g,t}-\hat{\beta}_{g,t-1})}{\left\|\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\right\|}\right)\geq 0,
\end{aligned}$$

where the last inequality follows by the CS inequality. We thus have

$$\begin{aligned}
\mathring{Q}(\hat{\beta})-\mathring{Q}(\mathring{\beta}) \geq & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(x_{it}'(\mathring{\beta}_{g_i^0,t}-\hat{\beta}_{g_i^0,t})-x_{i,t-1}(\mathring{\beta}_{g_i^0,t-1}-\hat{\beta}_{g_i^0,t-1}))^2 \\
= & \frac{1}{T}\sum_{g\in\mathbb{G}}(\mathring{\beta}_g-\hat{\beta}_g)' M^F(\gamma^0,g,g)(\mathring{\beta}_g-\hat{\beta}_g) \\
\geq & \hat{\rho}\frac{1}{T}\sum_{g\in\mathbb{G}}\left\|\hat{\beta}_g-\mathring{\beta}_g\right\|^2 .
\end{aligned}$$

We thus have, by (S.10) and Assumption 8.1,

$$\frac{1}{T}\sum_{g\in\mathbb{G}}\left\|\hat{\beta}_g-\mathring{\beta}_g\right\|^2 = o_p(T^{-\delta}).$$

This implies that

$$\left\|\hat{\beta}_{g,t}-\mathring{\beta}_{g,t}\right\|^2 = o_p(T^{1-\delta})$$

for any δ. Thus we have the desired result.

□

Proof of Theorem S.1

Proof. Since $\mathring{\beta}$ minimizes $\hat{Q}(\beta,\gamma^0)$, $\mathring{\beta}=\hat{\beta}$ if $\hat{\gamma}=\gamma^0$. We note that

$$\Pr(\hat{\gamma}\neq\gamma^0)=\Pr\left(\max_{1\leq i\leq N}\mathbf{1}\{\hat{g}_i(\hat{\beta})\neq g_i^0\}=1\right)\leq\sum_{i=1}^{N}E\left(\mathbf{1}\{\hat{g}_i(\hat{\beta})\neq g_i^0\}\right).$$

By Lemmas 2 and S.6, we have $\Pr(\hat{\beta}\in\mathcal{N}_\eta)\to 1$ for any η. Together with this, the argument made in the proof of Lemma S.3 shows that $\max_{1\leq i\leq N}E\left(\mathbf{1}\{\hat{g}_i(\hat{\beta})\neq g_i^0\}\right)=O(T^{-\delta})$ for any $\delta>0$. This means that

$$\Pr(\hat{\gamma}\neq\gamma^0)\leq N\max_{1\leq i\leq N}E\left(\mathbf{1}\{\hat{g}_i(\hat{\beta})\neq g_i^0\}\right)=o(NT^{-\delta})$$

for any δ. Thus under the condition of the theorem, by Lemma S.7, we have

$$\Pr\left(\left\|\hat{\theta}_{g,t}\right\|\neq 0,\exists t\in\mathcal{T}^{0c}_{m_g^0,g}, g\in\mathbb{G}\right)$$

$$
\begin{aligned}
&\leq \Pr\left(\left\{\left\|\hat{\theta}_{g,t}\right\| \neq 0, \exists t \in \mathcal{T}^{0c}_{m^0_g,g}, g \in \mathbb{G}\right\}, \{\hat{\gamma} = \gamma^0\}\right) + \Pr\left(\hat{\gamma} \neq \gamma^0\right) \\
&= \Pr\left(\left\{\left\|\mathring{\theta}_{g,t}\right\| \neq 0, \exists t \in \mathcal{T}^{0c}_{m^0_g,g}, g \in \mathbb{G}\right\}, \{\hat{\gamma} = \gamma^0\}\right) + \Pr\left(\hat{\gamma} \neq \gamma^0\right) \\
&\leq \Pr\left(\left\|\mathring{\theta}_{g,t}\right\| \neq 0, \exists t \in \mathcal{T}^{0c}_{m^0_g,g}, g \in \mathbb{G}\right) + \Pr\left(\hat{\gamma} \neq \gamma^0\right) \to 0.
\end{aligned}
$$

We therefore has the desired result.

□

Proof of Theorem S.2

Proof. Given Lemma 2 and Theorem S.1, the proof is based on an argument essentially identical to the proof of Corollary 3.4 in Qian and Su (2016) and is thus omitted.

□

Proof of Theorem S.3

Proof. The theorem holds by Lemmas 2 and S.9.

□

S.6 Models with a subset of coefficients fully time varying

In this section, we provide theoretical analysis of the GAGFL estimator applied to models in which a subset of coefficients are fully time varying.

We divide the set of regressors and the coefficients such that $x_{it} = (x'_{1,it}, x'_{2,it})'$ and $\beta_{g,t} = (\beta'_{1,g,t}, \beta'_{2,g,t})'$. The model is written as

$$
y_{it} = x'_{1,it}\beta_{1,g,t} + x'_{1,it}\beta_{2,g,t} + \epsilon_{it}.
$$

We assume that $\beta_{1,g,t}$ is fully time varying and $\beta_{2,g,t}$ may exhibit structural breaks. The GAGFL objective function is modified such that the penalty part includes only $\beta_{2,g,t}$:

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x'_{it}\beta_{g_i,t})^2 + \lambda \sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\beta_{2,g,t} - \beta_{2,g,t-1}\right\|,
$$

where

$$\dot{w}_{2,g,t} = \|\beta_{2,g,t} - \beta_{2,g,t-1}\|^{-\kappa}.$$

S.6.1 Assumptions

We first present assumptions. We use the same first set of assumptions, namely Assumptions 1-3, as those used for our main algorithm. However, we slightly modify the second set of assumptions that are related to structural breaks. The modified assumptions concern $\beta_{2,g,t}$ only.

Let $J_{2,\min} = \min_{g\in\mathbb{G},1\leq j\leq m_g^0} \|\alpha_{2,g,j+1} - \alpha_{2,g,j}\|$.

Assumption S.2.

1. $\sqrt{N}T\lambda\left(\sum_{g\in\mathbb{G}} m_g^0\right) J_{2,\min}^{-\kappa} = O_p(1)$.

2. $\sqrt{N}T\lambda N^{-\kappa/2} \to_p \infty$.

3. $\sqrt{N}J_{2,\min} \to \infty$.

Let k_1 and k_2 be the dimensions of $x_{1,it}$ and $x_{2,it}$ respectively. Let $\Sigma_{x,g,j}$ be a $(I_{g,j}k_1 + k_2) \times (I_{g,j}k_1 + k_2)$ matrix whose upper-left block is a block diagonal matrix of $\text{plim}_{N\to\infty} \sum_{g_i=g} x_{1,it}x_{1,it}'/N_g$ for $t = T_{g,j}^0, \ldots, T_{g,j+1}^0 - 1$, whose upper-right block is a matrix consisting of $I_{g,j}$ blocks with each block being $\text{plim}_{N\to\infty} \sum_{g_i=g} x_{1,it}x_{2,it}'/(N_g\sqrt{I_{g,j}})$, whose lower-left block is the transpose of the upper-right block, and whose lower-right block is $\text{plim}_{N,T\to\infty} \sum_{g_i=g} \sum_{t=T_{g,j}^0}^{T_{g,j+1}^0-1} x_{2,it}x_{2,it}'/(N_g I_{g,j})$. Let $\Sigma_{x,g}$ be a block diagonal matrix whose j-th block is $\Sigma_{x,g,j}$. Lastly, Σ_x is a block diagonal matrix whose g-th block is $\Sigma_{x,g}$.

Let $d_{g,j,NT} = \sum_{g_i=g}(x_{1,iT_{g,j}^0}'\epsilon_{iT_{g,j}^0}, \ldots, x_{1,i,T_{g,j+1}^0-1}'\epsilon_{i,T_{g,j+1}^0-1}, \sum_{t=T_{g,j}^0}^{T_{g,j+1}^0-1} x_{2,it}'\epsilon_{it}/\sqrt{I_{g,j}})/\sqrt{N_g}$. Let $d_{g,NT} = (d_{g,0,NT}, \ldots, d_{g,m_g^0,NT})$. Lastly, let $d_{NT} = (d_{1,NT}, \ldots, d_{G,NT})$. Let $\Omega = \lim_{N,T\infty} E(d_{NT}'d_{NT})$.

Assumption S.3. *Suppose that Σ_x and Ω are well-defined, their minimum eigenvalues are bounded away from zero and their maximum eigenvalues are bounded uniformly over T. $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. For a $l \times (GTk_1 + \sum_{g=1}^G (m_g^0 + 1)k_2)$ matrix D, where l does not depend on T and $\lim_{T\to\infty} D\Omega D'$ exists and is positive definite, $Dd_{NT}' \to_d N(0, \lim_{T\to\infty} D\Omega D')$.*

Assumption S.4. $N\sum_{g=1}^G (m_g^0)\lambda^2 I_{\min}^{-1} J_{2,\min}^{-2\kappa} = o_p(1)$.

S.6.2 Theoretical results

This section presents the theoretical results about the asymptotic behaviors of the GAGFL estimator when only a subset of the coefficients are penalized. A set of very similar results to those in the standard case are obtained.

Lemma S.12. *Suppose that Assumptions 1, 2, and 3 are satisfied. Suppose also that $N_g/N \to \pi_g$ for some $0 < \pi_g < 1$ for all $g \in \mathbb{G}$. As $N, T \to \infty$, for any $\delta > 0$, it holds that*

$$\hat{\beta}_{g,t} = \mathring{\beta}_{g,t} + o_p(T^{-\delta}),$$

for all g and t.

Theorem S.5. *Suppose that Assumptions 1, 2, 3, and S.2 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. It follows that*

$$\Pr\left(\left\|\hat{\theta}_{2,g,t}\right\| = 0, \forall t \in \mathcal{T}^{0c}_{m_g^0,g}, g \in \mathbb{G}\right) \to 1$$

as $N, T \to \infty$ with $N/T^{\delta} \to 0$ for some δ.

Theorem S.6. *Suppose that Assumptions 1, 2, 3, and S.2 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. It holds that, as $N, T \to \infty$ with $N/T^{\delta} \to 0$ for some $\delta > 0$,*

$$\Pr(\hat{m}_g = m_g^0, \forall g \in \mathbb{G}) \to 1,$$

and

$$\Pr\left(\hat{T}_{g,j} = T^0_{g,j}, \forall j \in \{1, \ldots, m_g^0\}, g \in \mathbb{G} \mid \hat{m}_g = m_g^0, \forall g \in \mathbb{G}\right) \to 1.$$

Theorem S.7. *Suppose that Assumptions 1, 2, 3, S.2, S.3 and S.4 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. Let A be a diagonal matrix whose diagonal elements are A_g for $g = 1, \ldots, G$ and A_g is a diagonal matrix whose elements are $((\iota_{k_1}, I_{g,1}\iota_{k_2}), \ldots, (\iota_{k_1}, I_{g,m_1^0+1}\iota_{k_2})$ where ι_l is the l-dimensional row vector of ones. Let Π be a $(TGk_1 + \sum_{g=1}^G (m_g^0 + 1)k_2) \times (TGk_1 + \sum_{g=1}^G (m_g^0 + 1)k_2)$ block diagonal matrix whose g-th diagonal block is an $(m_g^0 + 1)k \times (m_g^0 + 1)k$ diagonal matrix with the elements being π_g.*

Conditional on $\hat{m}_g = m_g^0$ for all $g \in \mathbb{G}$, we have, if $(\max_{g\in\mathbb{G}} m_g^0)^2/(I_{\min} \min_{g\in\mathbb{G}} N_g) \to 0$,

$$D\sqrt{N}A^{1/2}(\hat{\alpha} - \alpha^0) \to_d N(0, D\Sigma_x^{-1}\Pi^{-1/2}\Omega\Pi^{-1/2}\Sigma_x^{-1}D').$$

S.6.3 Proofs

The proofs are almost identical to those for the main model. Note that the initial estimator, $\dot{\beta}$, is identical to the standard case and we start the proofs from the analysis of the AGFL estimator.

We firs consider the estimator under the true group membership strucutre. Let

$$\mathring{Q}(\beta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\beta_{g_i^0,t})^2 + \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\beta_{2,g,t} - \beta_{2,g,t-1}\right\|,$$

and that $\mathring{\beta} = \arg\min_{\beta\in\mathcal{B}^{GT}}\mathring{Q}(\beta)$.

Lemma S.13. *Suppose that Assumptions 1.3, 2.1 and S.2.1 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. We have, as $N, T \to \infty$,*

$$\frac{1}{T}\left\|\mathring{\beta}_g - \beta_g^0\right\|^2 = O_p\left(\frac{1}{N}\right)$$

for any $g \in \mathbb{G}$. We also have, as $N, T \to \infty$,

$$\mathring{\beta}_{g,t} - \beta_{g,t}^0 = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Proof. This lemma contains two parts. We first consider the first part regarding the norm of coefficient difference. Let $b_t = N^{-1/2}(\beta_{g,t} - \beta_{g,t}^0)$ and $b = (b_1', \cdots b_T')' = N^{-1/2}(\beta_g - \beta_g^0)$. Let also $b_{2,t} = N^{-1/2}(\beta_{2,g,t} - \beta_{2,g,t}^0)$ and $b_2^* = (b_{2,1}', \ldots, b_{2,T}')'$. Let

$$Q_g(\beta_g) = \frac{1}{NT}\sum_{g_i^0=g}\sum_{t=1}^{T}(y_{it} - x_{it}'\beta_{g,t})^2 + \lambda\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\beta_{2,g,t} - \beta_{2,g,t-1}\right\|.$$

Note that $\mathring{\beta}_g = \arg\min Q_g(\beta_g)$.

We have

$$\begin{aligned}
&N(Q_g(\beta_g) - Q_g(\beta_g^0)) \\
=&\frac{1}{T}\sum_{g_i^0=g}\sum_{t=1}^{T}(y_{it} - x_{it}'\beta_{g,t})^2 + N\lambda\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\beta_{2,g,t} - \beta_{2,g,t-1}\right\|
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{T}\sum_{g_i^0=g}\sum_{t=1}^{T}(y_{it}-x_{it}'\beta_{g,t}^0)^2 - N\lambda\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\beta_{2,g,t}^0-\beta_{2,g,t-1}^0\right\| \\
=&\frac{1}{NT}\sum_{g_i^0=g}\sum_{t=1}^{T}b_t'x_{it}x_{it}'b_t-\frac{2}{\sqrt{NT}}\sum_{g_i^0=g}\sum_{t=1}^{T}\epsilon_{it}x_{it}'b_t \\
&+N\lambda\sum_{t\in\mathcal{T}_{m_g^0,g}^0}\dot{w}_{2,g,t}\left(\left\|\beta_{2,g,t}^0-\beta_{2,g,t-1}^0+N^{-1/2}(b_{2,t}-b_{2,t-1})\right\|-\left\|\beta_{2,g,t}^0-\beta_{2,g,t-1}^0\right\|\right) \\
&+N\lambda\sum_{t\in\mathcal{T}_{m_g^0,g}^{0c}}\dot{w}_{2,g,t}\left\|N^{-1/2}(b_{2,t}-b_{2,t-1})\right\|. \qquad \text{(S.11)}
\end{aligned}
$$

We consider the four terms on the right hand side of the last equality of (S.11) in turn. For the first term, we note, by Assumption 2.1, that

$$
\frac{1}{NT}\sum_{g_i^0=g}\sum_{t=1}^{T}b_t'x_{it}x_{it}'b_t=\frac{1}{T}b'M(\gamma^0,g,g)b\geq\hat{\rho}\frac{1}{T}\|b\|^2.
$$

For the second term, we have

$$
\frac{2}{\sqrt{NT}}\sum_{g_i^0=g}\sum_{t=1}^{T}\epsilon_{it}x_{it}'b_t=O_p\left(\frac{1}{\sqrt{T}}\|b\|\right),
$$

by Assumption 1.3. Next, we consider the third term. We have by the Jensen, triangular and CS inequalities,

$$
\begin{aligned}
&N\lambda\sum_{t\in\mathcal{T}_{m_g^0,g}^0}\dot{w}_{2,g,t}\left(\left\|\beta_{2,g,t}^0-\beta_{2,g,t-1}^0+N^{-1/2}(b_{2,t}-b_{2,t-1})\right\|-\left\|\beta_{2,g,t}^0-\beta_{2,g,t-1}^0\right\|\right) \\
\leq&N\lambda\sum_{t\in\mathcal{T}_{m_g^0,g}^0}\dot{w}_{2,g,t}\left\|N^{-1/2}(b_{2,t}-b_{2,t-1})\right\| \\
\leq&m^0\sqrt{N}\lambda\max_{s\in\mathcal{T}_{m_g^0,g}^0}(\dot{w}_{2,g,s})\frac{1}{m^0}\sum_{t\in\mathcal{T}_{m_g^0,g}^0}\|b_{2,t}-b_{2,t-1}\| \\
\leq&m^0\sqrt{N}\lambda\max_{s\in\mathcal{T}_{m_g^0,g}^0}(\dot{w}_{2,g,s})\left(\frac{1}{m^0}\sum_{t\in\mathcal{T}_{m_g^0,g}^0}\|b_{2,t}-b_{2,t-1}\|^2\right)^{1/2}
\end{aligned}
$$

$$\leq 2\sqrt{m^0 N}\lambda \max_{s\in\mathcal{T}^0_{m^0_g,g}}(\dot{w}_{2,g,s})\|b^*_2\|.$$

By Assumption S.2.1, this third term is $O_p(T^{-1/2}\|b\|)$. Lastly, we consider the fourth term, and we have

$$N\lambda \sum_{t\in\mathcal{T}^{0c}_{m^0_g,g}} \dot{w}_{2,g,t}\left\|N^{-1/2}(b_{2,t}-b_{2,t-1})\right\| \geq 0.$$

Summing the four parts up, we have

$$0 \geq N(Q_g(\beta_g)-Q_g(\beta^0_g)) \geq \hat{\rho}\frac{1}{T}\|b\|^2 - O_p(T^{-1/2}\|b\|).$$

If $\|b\|^2/T$ is not stochastically bounded, then the first term, which is positive, dominates and the first inequality does not hold asymptotically. We thus have the desired result.

Next, we consider the second part regarding the difference between two coefficient vectors. The above argument implies that

$$\begin{aligned}0 \geq N(Q_g(\beta_g)-Q_g(\beta^0_g)) \geq& \frac{1}{NT}\sum_{g^0_i=g}\sum_{t=1}^{T} b_t' x_{it}x_{it}' b_t - \frac{2}{\sqrt{NT}}\sum_{g^0_i=g}\sum_{t=1}^{T}\epsilon_{it}x_{it}'b_t \\ &- \sqrt{N}\lambda \max_{s\in\mathcal{T}^0_{m^0_g,g}}(\dot{w}_{2,g,s})\sum_{t\in\mathcal{T}^0_{m^0_g,g}}\|b_{2,t}-b_{2,t-1}\|.\end{aligned}$$

We thus have

$$\begin{aligned}0 \geq N(Q_g(\beta_g)-Q_g(\beta^0_g)) \geq& \frac{1}{NT}\sum_{g^0_i=g}\sum_{t=1}^{T} b_t' x_{it}x_{it}' b_t - \frac{2}{\sqrt{NT}}\sum_{g^0_i=g}\sum_{t=1}^{T}\epsilon_{it}x_{it}'b_t \\ &- \sqrt{N}\lambda \max_{s\in\mathcal{T}^0_{m^0_g,g}}(\dot{w}_{2,g,s})\sum_{t=1}^{T}2\|b_{2,t}\|.\end{aligned}$$

The right hand side of the inequality can be analyzed for each t. If $|b_t|$ is not $O_p(1)$, then it can be seen that it cannot be optimal. The desired result thus follows.

□

Let $\mathring{\theta}_{2,g,1}=\mathring{\beta}_{2,g,1}$ and $\mathring{\theta}_{2,g,t}=\mathring{\beta}_{2,g,t}-\mathring{\beta}_{2,g,t-1}$.

Lemma S.14. *Suppose that Assumptions 1.3, 2.1, 1.5, and S.2 hold. Suppose that $N_g/N \to$*

$\pi_g > 0$ *for any* $g \in \mathbb{G}$. *It follows that*

$$\Pr\left(\left\|\mathring{\theta}_{2,g,t}\right\| = 0, \forall t \in \mathcal{T}^{0c}_{m^0_g,g}, g \in \mathbb{G}\right) \to 1$$

as $N \to \infty$.

Proof. The proof is by contradiction. Suppose that $\exists(t,g)$ such that $2 \le t \le T-1$ and $\mathring{\theta}_{2,g,t} \neq 0$ for sufficiently large N. Note that $\|\theta\|$ is differentiable at $\mathring{\theta}_{2,g,t}$ if $\mathring{\theta}_{2,g,t} \neq 0$. We thus have the following first order condition (FOC) for $\mathring{\beta}_{2,g,t}$:

$$-2\frac{1}{NT}\sum_{g^0_i=g}(y_{it} - x'_{it}\mathring{\beta}_{g,t})x_{it} + \lambda\dot{w}_{2,g,t}\frac{\mathring{\theta}_{2,g,t}}{\|\mathring{\theta}_{2,g,t}\|} - \lambda\dot{w}_{2,g,t+1}e_{t+1} = 0,$$

where $e_{t+1} = \mathring{\theta}_{2,g,t+1}/\|\mathring{\theta}_{2,g,t+1}\|$ if $\mathring{\theta}_{2,g,t+1} \neq 0$ and $\|e_{t+1}\| \le 1$ otherwise. Multiplying both sides of the equation by $\sqrt{N}T$ and use $y_{it} = x'_{it}\beta^0_{g^0_i,t} + \epsilon_{it}$, we have

$$\begin{aligned}&2\frac{1}{\sqrt{N}}\sum_{g^0_i=g}x_{it}x'_{it}(\mathring{\beta}_{g,t} - \beta^0_{g,t}) - 2\frac{1}{\sqrt{N}}\sum_{g^0_i=g}\epsilon_{it}x_{2,it}\\&+\sqrt{N}T\lambda\dot{w}_{2,g,t}\frac{\mathring{\theta}_{2,g,t}}{\|\mathring{\theta}_{2,g,t}\|} - \sqrt{N}T\lambda\dot{w}_{2,g,t+1}e_{t+1} = 0.\end{aligned}$$

The first term is $O_p(1)$ by Lemma S.13 and Assumption 1.5. The second term is $O_p(1)$ by Assumption 1.3. For the third term, we observe that the absolute value of at least one element of $\mathring{\theta}_{2,g,t}/\|\mathring{\theta}_{2,g,t}\|$ exceeds $1/\sqrt{k}$, and that $\dot{w}^{-1}_{2,g,t} = O_p(N^{-\kappa/2})$ because $t \in \mathcal{T}^{0c}_{m^0_g,g}$. Therefore, the third term is $O_p(\sqrt{N}T\lambda N^{-\kappa/2}/\sqrt{k})$ and this tends to infinity by Assumption S.2.2.

We now consider the term $\sqrt{N}T\lambda\dot{w}_{2,g,t+1}e_{t+1}$. Suppose that $t+1 \in \mathcal{T}^0_{m^0_g,g}$. In this case, $\dot{w}_{2,g,t+1} = O_p(J^{-\kappa}_{2,\min})$. This and Assumption S.2.1 implies that $\sqrt{N}T\lambda\dot{w}_{2,g,t}e_{t+1} = O_p(1)$. In this case, the third term explodes but other terms are stochastically bounded, and the first order condition cannot hold.

Next, we consider the case where $t+1 \in \mathcal{T}^{0c}_{m^0_g,g}$. We note that from the argument in the previous paragraph, if $t = T^0_{g,j} - 1 \in \mathcal{T}^{0c}_{m^0_g,g}$ for some j, then $\Pr(\|\mathring{\theta}_{2,g,t}\| = 0) \to 1$ and $\sqrt{N}T\lambda\dot{w}_{2,g,t+1}e_{t+1} = O_p(1)$. This implies that for $t = T^0_{g,j} - 2 \in \mathcal{T}^{0c}_{m^0_g,g}$, $\Pr(\|\mathring{\theta}_{2,g,t}\| = 0) \to 1$ and $\sqrt{N}T\lambda\dot{w}_{2,g,t+1}e_{t+1} = O_p(1)$ too. Applying this argument recursively until $t = T^0_{g,j-1} + 1 \in \mathcal{T}^{0c}_{m^0_g,g}$, we have for all $t \in \mathcal{T}^{0c}_{m^0_g,g}$, $\Pr(\|\mathring{\theta}_{2,g,t}\| = 0) \to 1$.

Lastly, we consider the case in which $t = T$. In this case, the first order condition for

$\beta_{2,g,t}$ is

$$2\frac{1}{\sqrt{N}}\sum_{g_i^0=g} x_{iT}x_{iT}'(\mathring{\beta}_{g,T}-\beta_{g,T}^0)-2\frac{1}{\sqrt{N}}\sum_{g_i^0=g}\epsilon_{iT}x_{2,iT}+\sqrt{N}T\lambda\dot{w}_{2,g,T}e_T=0,$$

and there is no fourth term. We can apply the argument above and obtain $\Pr(\|\mathring{\theta}_{2,g,t}\|=0)\to 1$.

$\square$

Lemma S.15. *Suppose that Assumptions 1.3, 2.1, 1.5, and S.2 hold. Suppose that $N_g/N\to\pi_g>0$ for any $g\in\mathbb{G}$. It holds that, as $N\to\infty$,*

$$\Pr(\mathring{m}_g=m_g^0,\forall g\in\mathbb{G})\to 1,$$

and

$$\Pr(\mathring{T}_{g,j}=T_{g,j}^0,\forall j\in\{1,\dots,m^0\},g\in\mathbb{G}\mid \mathring{m}_g=m_g^0,\forall g\in\mathbb{G})\to 1$$

Proof. The proof is based on an argument essentially identical to the proof of Corollary 3.4 in Qian and Su (2016) and is thus omitted. $\square$

Recall that $\mathring{\alpha}_{2,g,j}=\mathring{\beta}_{2,g,t}$ for $T_{g,j}^0\le t<T_{g,j+1}^0$.

Lemma S.16. *Suppose that Assumptions 1.3, 2.1, 1.5, S.2, S.3 and S.4 hold. Suppose that $N_g/N\to\pi_g>0$ for any $g\in\mathbb{G}$. Let A and Π be as defined in Theorem S.7.*

Conditional on $\mathring{m}_g=m_g^0$ for all $g\in\mathbb{G}$, we have, if $(\max_{g\in\mathbb{G}}m_g^0)^2/(I_{\min}\min_{g\in\mathbb{G}}N_g)\to 0$,

$$D\sqrt{N}A^{1/2}(\mathring{\alpha}-\mathring{\alpha}^0)\to_d N(0,D\Sigma_x^{-1}\Pi^{-1/2}\Omega\Pi^{-1/2}\Sigma_x^{-1}D').$$

Proof. We note that $\mathring{\alpha}_{g,j}$ satisfies the following FOC:

$$\begin{aligned}&\frac{1}{NT}\sum_{g_i^0=g}(y_{it}-x_{it}'\mathring{\alpha}_{g,j})x_{1,it}\quad\text{for}\quad t=T_{g,j}^0,\dots,T_{g,j+1}^0-1,\\&\frac{1}{NT}\sum_{g_i^0=g}\sum_{t=T_{g,j}^0}^{T_{g,j+1}^0-1}(y_{it}-x_{it}'\mathring{\alpha}_{g,j})x_{2,it}+R_{g,j},\end{aligned}$$

where $R_{g,1}=-\lambda\dot{w}_{2,g,T_{g,1}^0}e_{T_{g,1}^0}$, $R_{g,j}=\lambda(\dot{w}_{2,g,T_{g,j-1}^0}e_{T_{g,j-1}^0}-\dot{w}_{2,g,T_{g,j}^0}e_{T_{g,j}^0})$ for $2\le j\le m_g^0$ and $R_{g,m_g^0+1}=\lambda\dot{w}_{2,g,T_{m_g^0}^0}e_{T_{m_g^0}^0}$.

It is thus sufficient to show that $||R||$ is $o_p(1)$ where R is a vector of $R_{g,j}$s. We have

$$
\begin{aligned}
&\|R\|^2\\
\leq&\lambda^2\sum_{g=1}^{G}\left(I_{g,1}^{-1}\left\|\dot{w}_{2,g,T_{g,1}^0}e_{g,T_{g,1}^0}\right\|^2+\sum_{j=2}^{m_g^0}I_{g,j}^{-1}\left\|\dot{w}_{2,g,T_{g,j-1}^0}e_{g,T_{g,j-1}^0}\right\|^2+I_{g,m_g^0+1}^{-1}\left\|\dot{w}_{2,g,T_{m_g^0,1}^0}e_{g,T_{m_g^0,1}^0}\right\|^2\right)\\
\leq&4\sum_{g=1}^{G}(m_g^0+1)\lambda^2 I_{\min}^{-1}\max_{g\in\mathbb{G},t\in\mathcal{T}_{m_g^0,g}^0}\|\dot{w}_{2,g,t}\|^2\\
=&O_p\left(\sum_{g=1}^{G}(m_g^0)\lambda^2 I_{\min}^{-1}J_{2,\min}^{-2\kappa}\right).
\end{aligned}
$$

By Assumptions S.3 and S.4, the second term is $o_p(1)$.

$\square$

Let

$$\hat{Q}_{NT}(\beta,\gamma)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}'\beta_{g_i,t})^2+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\beta_{2,g,t}-\beta_{2,g,t-1}\right\|.$$

and

$$\tilde{\hat{Q}}_{NT}(\beta,\gamma)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}'(\beta_{g_i^0,t}^0-\beta_{g_i,t}))^2+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\beta_{2,g,t}-\beta_{2,g,t-1}\right\|.$$

Lemma S.17. *Suppose that Assumption 1 hold.*

$$\sup_{(\beta,\gamma)\in\mathcal{B}^{GT}\times\Gamma_G}\left|\hat{Q}_{NT}(\beta,\gamma)-\tilde{\hat{Q}}_{NT}(\beta,\gamma)\right|=o_p(1).$$

Proof. Note that

$$\hat{Q}_{NT}(\beta,\gamma)-\tilde{\hat{Q}}_{NT}(\beta,\gamma)=\dot{Q}_{NT}(\beta,\gamma)-\tilde{Q}_{NT}(\beta,\gamma).$$

Lemma 3 implies the desired result. $\square$

Lemma S.18. *Suppose that Assumptions 1, 2 and S.2.1 hold. Suppose that $N_g/N\to\pi_g>$*

0 *for any* $g \in \mathbb{G}$. *It holds that as* $N, T \to \infty$,

$$d_H(\hat{\beta}, \beta^0) = o_p(1).$$

Proof. From Lemma S.17, we have

$$\tilde{\hat{Q}}(\hat{\beta}, \hat{\gamma}) = \hat{Q}(\hat{\beta}, \hat{\gamma}) + o_p(1) \leq \hat{Q}(\beta^0, \gamma^0) + o_p(1) = \tilde{\hat{Q}}(\beta^0, \gamma^0) + o_p(1).$$

Because $\tilde{Q}(\beta, \gamma)$ is minimized at $\beta = \beta^0$ and $\gamma = \gamma^0$, we have

$$\tilde{\hat{Q}}(\hat{\beta}, \hat{\gamma}) - \tilde{\hat{Q}}(\beta^0, \gamma^0) = o_p(1).$$

On the other hand, we have

$$\begin{aligned}
&\tilde{\hat{Q}}(\beta, \gamma) - \tilde{\hat{Q}}(\beta^0, \gamma^0) \\
=&\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it}'(\beta^0_{g_i^0,t} - \beta_{g_i,t}) \right)^2 \\
&+ \lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{2,g,t} \left(\|\beta_{2,g,t} - \beta_{2,g,t-1}\| - \left\|\beta^0_{2,g,t} - \beta^0_{2,g,t-1}\right\| \right) \\
=&\sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} \frac{1}{T} \left(\beta^0_g - \beta_{\tilde{g}}\right)' M(\gamma, g, \tilde{g}) \left(\beta^0_g - \beta_{\tilde{g}}\right) \\
&+ \lambda \sum_{g \in \mathbb{G}} \sum_{t \in \mathcal{T}^0_{m^0_g,g}} \dot{w}_{2,g,t} \left(\|\beta_{2,g,t} - \beta_{2,g,t-1}\| - \left\|\beta^0_{2,g,t} - \beta^0_{2,g,t-1}\right\| \right) \\
&+ \lambda \sum_{g \in \mathbb{G}} \sum_{t \in \mathcal{T}^{0c}_{m^0_g,g}} \dot{w}_{2,g,t} \left\|\beta_{2,g,t} - \beta_{2,g,t-1}\right\|.
\end{aligned} \tag{S.12}$$

We now examine the three terms on the left hand side of (S.12) in turn. For the first term, in the proof of Lemma 4, we have shown that

$$\sum_{g=1}^{G} \sum_{\tilde{g}=1}^{G} \frac{1}{T} \left(\beta^0_g - \beta_{\tilde{g}}\right)' M(\gamma, g, \tilde{g}) \left(\beta^0_g - \beta_{\tilde{g}}\right) \geq \hat{\rho} \max_{g \in \mathbb{G}} \left(\min_{\tilde{g} \in \mathbb{G}} \left(\frac{1}{T} \sum_{t=1}^{T} \left\|\beta^0_{g,t} - \beta_{\tilde{g},t}\right\|^2 \right) \right).$$

Note that $\hat{\rho}$ is asymptotically bounded away from zero by Assumption 2.1.

For the second term, we have, by the Jensen, triangular and CS inequalities,

$$\begin{aligned}
&\left|\lambda \sum_{g\in\mathbb{G}} \sum_{t\in\mathcal{T}^0_{m^0_g,g}} \dot{w}_{2,g,t}\left(\|\beta_{2,g,t}-\beta_{2,g,t-1}\| - \left\|\beta^0_{2,g,t}-\beta^0_{2,g,t-1}\right\|\right)\right| \\
\leq &\lambda \sum_{g\in\mathbb{G}} \sum_{t\in\mathcal{T}^0_{m^0_g,g}} \dot{w}_{2,g,t}\left\|\beta_{2,g,t}-\beta_{2,g,t-1}-(\beta^0_{2,g,t}-\beta^0_{2,g,t-1})\right\| \\
\leq &\lambda \max_{s\in\mathcal{T}^0_{m^0_g,g}, g\in\mathbb{G}} (\dot{w}_{2,g,s}) \sum_{t\in\mathcal{T}^0_{m^0_g,g}} \left\|\beta_{2,g,t}-\beta_{2,g,t-1}-(\beta^0_{2,g,t}-\beta^0_{2,g,t-1})\right\| \\
= &O_p\left(\lambda\left(\sum_{g\in\mathbb{G}} m^0_g\right) J_{2,\min}^{-\kappa}\right) = o_p(1).
\end{aligned}$$

where the last equality follows from Assumptions 1.1 and S.2.1.

Finally, for the third term, note that

$$\lambda \sum_{t\in\mathcal{T}^{0c}_{m^0_g,g}} \dot{w}_{2,g,t}\|\beta_{2,g,t}-\beta_{2,g,t-1}\| \geq 0$$

Therefore unless we have

$$\max_{g\in\mathbb{G}}\left(\min_{\tilde{g}\in\mathbb{G}}\left(\frac{1}{T}\sum_{t=1}^T \left\|\beta^0_{g,t}-\hat{\beta}_{\tilde{g},t}\right\|^2\right)\right) = o_p(1),$$

$\tilde{\hat{Q}}(\beta,\gamma) - \tilde{\hat{Q}}(\beta^0,\gamma^0) < o_p(1)$ does not hold. We then follow the argument made in the proof of Lemma 4 to obtain

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

We thus have the desired result. □

As in the case for $\dot{\beta}$, the above result implies that there exists a permutation σ such that

$$\frac{1}{T}\sum_{t=1}^T \left\|\beta^0_{\sigma(g),t}-\hat{\beta}_{g,t}\right\|^2 = o_p(1)$$

and we take $\sigma(g) = g$ by relabeling. Moreover, we observe that given β, the second term of $\hat{Q}_{NT}(\beta, \gamma)$ does not affect the estimation of γ. Therefore, $\hat{g}_i(\beta)$ defined in (S.3) is also the estimate of g_i given β even if $\hat{Q}_{NT}(\beta, \gamma)$ is our objective function. It follows that Lemma 5 applies for the GAGFL estimator.

S.6.3.1 Proof of Lemma S.12

Let

$$\hat{Q}(\beta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x'_{it}\beta_{\hat{g}_i(\beta),t})^2 + \lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{2,g,t} \left\| \beta_{2,g,t} - \beta_{2,g,t-1} \right\| .$$

Note also that $\hat{Q}(\beta) = \hat{Q}(\beta, \hat{\gamma}(\beta))$ and that $\hat{\beta} = \arg\min_{\beta \in \mathcal{B}^{GT}} \hat{Q}(\beta)$.

Proof. We first evaluate the difference between $\mathring{Q}(\beta)$ and $\hat{Q}(\beta)$. Note that

$$\mathring{Q}(\beta) - \hat{Q}(\beta) = \check{Q}(\beta) - \dot{Q}(\beta).$$

Thus the proof of Lemma 7 implies that

$$\mathring{Q}(\hat{\beta}) - \hat{Q}(\hat{\beta}) = o_p(T^{-\delta}).$$

Similarly, we have

$$\mathring{Q}(\mathring{\beta}) - \hat{Q}(\mathring{\beta}) = o_p(T^{-\delta}).$$

Next, we evaluate the difference between $\mathring{\beta}$ and $\hat{\beta}$. By the definition of $\mathring{\beta}$ and $\hat{\beta}$, we have

$$0 \leq \mathring{Q}(\hat{\beta}) - \mathring{Q}(\mathring{\beta}) = \hat{Q}(\hat{\beta}) - \hat{Q}(\mathring{\beta}) + o_p(T^{-\delta}) \leq o_p(T^{-\delta}).$$

Thus we have

$$\mathring{Q}(\hat{\beta}) - \mathring{Q}(\mathring{\beta}) = o_p(T^{-\delta}). \tag{S.13}$$

We observe that

$$
\begin{aligned}
\mathring{Q}(\hat{\beta})-\mathring{Q}(\mathring{\beta})=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}'\hat{\beta}_{g_i^0,t})^2+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\hat{\beta}_{2,g,t}-\hat{\beta}_{2,g,t-1}\right\|\\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}'\mathring{\beta}_{g_i^0,t})^2-\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\mathring{\beta}_{2,g,t}-\mathring{\beta}_{2,g,t-1}\right\|\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}'(\mathring{\beta}_{g_i^0,t}-\hat{\beta}_{g_i^0,t}))^2\\
&+2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}'\mathring{\beta}_{g_i^0,t})(x_{it}'(\mathring{\beta}_{g_i^0,t}-\hat{\beta}_{g_i^0,t}))\\
&+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\hat{\beta}_{2,g,t}-\hat{\beta}_{2,g,t-1}\right\|-\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}\left\|\mathring{\beta}_{2,g,t}-\mathring{\beta}_{2,g,t-1}\right\|.
\end{aligned}
$$

By the first order condition for $\mathring{\beta}_{g,t}$, we have

$$
\begin{aligned}
&-2\frac{1}{NT}\sum_{g_i^0=g}(y_{it}-x_{it}'\mathring{\beta}_{g_i^0,t})x_{1,it}=0,\\
&-2\frac{1}{NT}\sum_{g_i^0=g}(y_{it}-x_{it}'\mathring{\beta}_{g_i^0,t})x_{2,it}+\lambda\dot{w}_{2,g,t}e_{g,t}-\lambda\dot{w}_{2,g,t+1}e_{g,t+1}=0,
\end{aligned}
$$

where $e_{g,1}=e_{g,T+1}=0$, for $2\le t\le T$, $e_{g,t}=(\mathring{\beta}_{2,g,t}-\mathring{\beta}_{2,g,t-1})/\left\|\mathring{\beta}_{2,g,t}-\mathring{\beta}_{2,g,t-1}\right\|$ if $\mathring{\beta}_{2,g,t}-\mathring{\beta}_{2,g,t-1}\neq 0$ and $\|e_{g,t}\|\le 1$ otherwise. We thus have

$$
\begin{aligned}
&2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}'\mathring{\beta}_{g_i^0,t})(x_{it}'(\mathring{\beta}_{g_i^0,t}-\hat{\beta}_{g_i^0,t}))\\
=&\lambda\sum_{g\in\mathbb{G}}\sum_{t=1}^{T}(\dot{w}_{2,g,t}e_{g,t}-\dot{w}_{2,g,t+1}e_{g,t+1})'(\mathring{\beta}_{2,g_i^0,t}-\hat{\beta}_{2,g_i^0,t})\\
=&\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{2,g,t}e_{g,t}'((\mathring{\beta}_{2,g_i^0,t}-\mathring{\beta}_{2,g_i^0,t-1})-(\hat{\beta}_{2,g_i^0,t}-\hat{\beta}_{2,g_i^0,t-1})).
\end{aligned}
$$

Let $\mathcal{T}_{m_g,g}$ be the set of t such that $\mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1} \neq 0$ and $\mathcal{T}^c_{m_g,g} = \{2, \ldots, T\} \backslash \mathcal{T}_{m_g,g}$. We have

$$\begin{aligned}
&\lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{2,g,t} e'_{g,t} ((\mathring{\beta}_{2,g_i^0,t} - \mathring{\beta}_{2,g_i^0,t-1}) - (\hat{\beta}_{g_i^0,t} - \hat{\beta}_{g_i^0,t-1})) \\
&+ \lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{2,g,t} \left\| \hat{\beta}_{2,g,t} - \hat{\beta}_{2,g,t-1} \right\| - \lambda \sum_{g \in \mathbb{G}} \sum_{t=2}^{T} \dot{w}_{2,g,t} \left\| \mathring{\beta}_{2,g,t} - \mathring{\beta}_{2,g,t-1} \right\| \\
=& \lambda \sum_{g \in \mathbb{G}} \sum_{t \in \mathcal{T}^c_{m_g,g}} \dot{w}_{2,g,t} \left(\left\| \hat{\beta}_{2,g,t} - \hat{\beta}_{2,g,t-1} \right\| - e'_{g,t} (\hat{\beta}_{2,g,t} - \hat{\beta}_{2,g,t-1}) \right) \\
&+ \lambda \sum_{g \in \mathbb{G}} \sum_{t \in \mathcal{T}_{m_g,g}} \dot{w}_{2,g,t} \left(\left\| \hat{\beta}_{2,g,t} - \hat{\beta}_{2,g,t-1} \right\| - \frac{(\mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1})'(\hat{\beta}_{g,t} - \hat{\beta}_{g,t-1})}{\left\| \mathring{\beta}_{2,g,t} - \mathring{\beta}_{2,g,t-1} \right\|} \right) \geq 0,
\end{aligned}$$

where the last inequality follows by the CS inequality. This implies that

$$\begin{aligned}
\mathring{Q}(\hat{\beta}) - \mathring{Q}(\mathring{\beta}) \geq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (x'_{it} (\mathring{\beta}_{g_i^0,t} - \hat{\beta}_{g_i^0,t}))^2 =& \frac{1}{T} \sum_{g \in \mathbb{G}} (\mathring{\beta}_g - \hat{\beta}_g)' M(\gamma^0, g, g) (\mathring{\beta}_g - \hat{\beta}_g) \\
\geq& \hat{\rho} \frac{1}{T} \sum_{g \in \mathbb{G}} \left\| \hat{\beta}_g - \mathring{\beta}_g \right\|^2.
\end{aligned}$$

Hence, by (S.13) and Assumption 2.1, we have that,

$$\frac{1}{T} \sum_{g \in \mathbb{G}} \left\| \hat{\beta}_g - \mathring{\beta}_g \right\|^2 = o_p(T^{-\delta}),$$

which further implies that

$$\left\| \hat{\beta}_{g,t} - \mathring{\beta}_{g,t} \right\|^2 = o_p(T^{1-\delta})$$

for any δ. This gives the desired result.

□

S.6.3.2 Proof of Theorem S.5

Proof. As $\mathring{\beta}$ minimizes $\hat{Q}(\beta, \gamma^0)$, $\mathring{\beta} = \hat{\beta}$ if $\hat{\gamma} = \gamma^0$. We note that

$$\Pr(\hat{\gamma} \neq \gamma^0) = \Pr\left(\max_{1 \leq i \leq N} \mathbf{1}\{\hat{g}_i(\hat{\beta}) \neq g_i^0\} = 1 \right) \leq \sum_{i=1}^{N} E\left(\mathbf{1}\{\hat{g}_i(\hat{\beta}) \neq g_i^0\} \right).$$

From Lemmas S.12 and S.13, we have $\Pr(\hat{\beta} \in \mathcal{N}_\eta) \to 1$ for any η. Together with this, the argument made in the proof of Lemma 5 shows that $\max_{1\le i\le N} E\left(\mathbf{1}\{\hat{g}_i(\hat{\beta}) \neq g_i^0\}\right) = O(T^{-\delta})$ for any $\delta > 0$. This means that

$$\Pr(\hat{\gamma} \neq \gamma^0) \le N \max_{1\le i\le N} E\left(\mathbf{1}\{\hat{g}_i(\hat{\beta}) \neq g_i^0\}\right) = o(NT^{-\delta})$$

for any δ. Thus under the condition of the theorem, from Lemma S.14, we have

$$\begin{aligned}
&\Pr\left(\left\|\hat{\theta}_{2,g,t}\right\| \neq 0, \exists t \in \mathcal{T}^{0c}_{m_g^0,g}, g \in \mathbb{G}\right)\\
\le &\Pr\left(\left\{\left\|\hat{\theta}_{2,g,t}\right\| \neq 0, \exists t \in \mathcal{T}^{0c}_{m_g^0,g}, g \in \mathbb{G}\right\}, \{\hat{\gamma} = \gamma^0\}\right) + \Pr\left(\hat{\gamma} \neq \gamma^0\right)\\
= &\Pr\left(\left\{\left\|\mathring{\theta}_{2,g,t}\right\| \neq 0, \exists t \in \mathcal{T}^{0c}_{m_g^0,g}, g \in \mathbb{G}\right\}, \{\hat{\gamma} = \gamma^0\}\right) + \Pr\left(\hat{\gamma} \neq \gamma^0\right)\\
\le &\Pr\left(\left\|\mathring{\theta}_{2,g,t}\right\| \neq 0, \exists t \in \mathcal{T}^{0c}_{m_g^0,g}, g \in \mathbb{G}\right) + \Pr\left(\hat{\gamma} \neq \gamma^0\right) \to 0.
\end{aligned}$$

We therefore have the desired result.

□

S.6.3.3 Proof of Theorem S.6

Proof. Given Lemma S.12 and Theorem S.5, the proof is based on an argument essentially identical to the proof of Corollary 3.4 in Qian and Su (2016) and is thus omitted.

□

S.6.3.4 Proof of Theorem S.7

Proof. The theorem holds using Lemmas S.12 and S.16.

□

S.7 Properties of the one-step estimator

Here we present the asymptotic properties of the one-step estimator $\beta^{(0)}$. We show that it has the same asymptotic properties as those of the iterative estimator, $\hat{\beta}$. In the following, $\theta_{g,t}^{(0)}$, $\hat{m}_g^{(0)}$, $T_{g,j}^{(0)}$ and $\hat{\alpha}^{(0)}$ denote the estimators of $\beta_{g,t} - \beta_{g,t-1}$, the number of breaks for group g, the j-th break date for group g and α, based on $\beta^{(0)}$, respectively.

S.7.1 Asymptotic results

The following results show that $\beta^{(0)}$ possesses the same asymptotic properties as those of $\hat{\beta}$.

Lemma S.19. *Suppose that Assumptions 1, 2, and 3 are satisfied. Suppose also that $N_g/N \to \pi_g$ for some $0 < \pi_g < 1$ for all $g \in \mathbb{G}$. As $N, T \to \infty$, for any $\delta > 0$, it holds that*

$$\beta_{g,t}^{(0)} = \mathring{\beta}_{g,t} + o_p(T^{-\delta}),$$

for all g and t.

Theorem S.8. *Suppose that Assumptions 1, 2, 3, and 4 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. It follows that*

$$\Pr\left(\left\|\hat{\theta}_{g,t}^{(0)}\right\| = 0, \forall t \in \mathcal{T}_{m_g^0,g}^{0c}, g \in \mathbb{G}\right) \to 1$$

as $N, T \to \infty$ with $N/T^{\delta} \to 0$ for some δ.

Theorem S.9. *Suppose that Assumptions 1, 2, 3, and 4 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. It holds that, as $N, T \to \infty$ with $N/T^{\delta} \to 0$ for some $\delta > 0$,*

$$\Pr(\hat{m}_g^{(0)} = m_g^0, \forall g \in \mathbb{G}) \to 1,$$

and

$$\Pr\left(\hat{T}_{g,j}^{(0)} = T_{g,j}^0, \forall j \in \{1, \ldots, m_g^0\}, g \in \mathbb{G} \mid \hat{m}_g^{(0)} = m_g^0, \forall g \in \mathbb{G}\right) \to 1.$$

Theorem S.10. *Suppose that Assumptions 1, 2, 3, 4, 5 and 6 hold. Suppose that $N_g/N \to \pi_g > 0$ for any $g \in \mathbb{G}$. Let A be a diagonal matrix whose diagonal elements are $(I_{1,1}, \ldots, I_{1,m_1^0+1}, I_{2,1}, \ldots, I_{2,m_2^0+1}, I_{3,1} \ldots, I_{G-1,m_{G-1}^0+1}, I_{G,1}, \ldots, I_{G,m_G^0+1})$. Let Π be a $\sum_{g=1}^{G}(m_g^0 + 1)k \times \sum_{g=1}^{G}(m_g^0 + 1)k$ block diagonal matrix whose g-th diagonal block is an $(m_g^0 + 1)k \times (m_g^0 + 1)k$ diagonal matrix with the elements being π_g.*

Conditional on $\hat{m}_g = m_g^0$ for all $g \in \mathbb{G}$, we have, if $(\max_{g\in\mathbb{G}} m_g^0)^2/(I_{\min} \min_{g\in\mathbb{G}} N_g) \to 0$,

$$D\sqrt{N}A^{1/2}(\hat{\alpha}^{(0)} - \alpha^0) \to_d N(0, D\Sigma_x^{-1}\Pi^{-1/2}\Omega\Pi^{-1/2}\Sigma_x^{-1}D').$$

S.7.2 Proofs

Note that $\beta^{(0)}$ can be written as

$$\beta^{(0)} = \arg\min_{\beta \in \mathcal{B}^{NT}} \hat{Q}^{(0)}_{NT}(\beta) \tag{S.14}$$

where

$$\hat{Q}^{(0)}_{NT}(\beta) = \hat{Q}_{NT}(\beta, \dot{\gamma}). \tag{S.15}$$

Let

$$\dot{Q}^{(0)}(\beta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x'_{it}\beta_{\hat{g}_i(\dot{\beta}),t})^2. \tag{S.16}$$

S.7.2.1 Proof of Lemma S.19

Proof. The proof is very similar to the proof of Lemma 1 of the main text. The only difference is that we consider $\hat{g}_i(\dot{\beta})$ instead of $\hat{g}_i(\beta)$

We first evaluate the difference between $\mathring{Q}(\beta)$ and $\hat{Q}^{(0)}(\beta)$. Note that

$$\mathring{Q}(\beta) - \hat{Q}^{(0)}(\beta) = \check{Q}(\beta) - \dot{Q}^{(0)}(\beta).$$

We have

$$\check{Q}(\beta) - \dot{Q}^{(0)}(\beta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{\hat{g}_i(\dot{\beta}) \neq g_i^0\}\left((y_{it} - x'_{it}\beta_{g_i^0,t})^2 - (y_{it} - x'_{it}\beta_{\hat{g}_i(\dot{\beta}),t})^2\right).$$

Because $\dot{\beta}$ is consistent from Lemmas 6 and 7, Lemma 5 implies that $\sum_{i=1}^{N}\mathbf{1}\{\hat{g}_i(\dot{\beta}) \neq g_i^0\}/N = o_p(T^{-\delta})$. Thus, together with Assumptions 1.1, 1.3 and 1.5, we have

$$\mathring{Q}(\beta^{(0)}) - \hat{Q}^{(0)}(\beta^{(0)}) = o_p(T^{-\delta}).$$

Similarly, we have

$$\mathring{Q}(\mathring{\beta}) - \hat{Q}^{(0)}(\mathring{\beta}) = o_p(T^{-\delta}).$$

Next, we evaluate the difference between $\mathring{\beta}$ and $\beta^{(0)}$. By the definition of $\mathring{\beta}$ and $\beta^{(0)}$, we have

$$0 \leq \mathring{Q}(\beta^{(0)}) - \mathring{Q}(\mathring{\beta}) = \hat{Q}^{(0)}(\beta^{(0)}) - \hat{Q}^{(0)}(\mathring{\beta}) + o_p(T^{-\delta}) \leq o_p(T^{-\delta}).$$

Thus we have

$$\mathring{Q}(\beta^{(0)}) - \mathring{Q}(\mathring{\beta}) = o_p(T^{-\delta}). \tag{S.17}$$

We observe that

$$\begin{aligned}
\mathring{Q}(\beta^{(0)}) - \mathring{Q}(\mathring{\beta}) =& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\beta^{(0)}_{g_i^0,t})^2 + \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta^{(0)}_{g,t} - \beta^{(0)}_{g,t-1}\right\| \\
&- \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\mathring{\beta}_{g_i^0,t})^2 - \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1}\right\| \\
=& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}'(\mathring{\beta}_{g_i^0,t} - \beta^{(0)}_{g_i^0,t}))^2 \\
&+ 2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\mathring{\beta}_{g_i^0,t})(x_{it}'(\mathring{\beta}_{g_i^0,t} - \beta^{(0)}_{g_i^0,t})) \\
&+ \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta^{(0)}_{g,t} - \beta^{(0)}_{g,t-1}\right\| - \lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1}\right\|.
\end{aligned}$$

By the first order condition for $\mathring{\beta}_{g,t}$, we have

$$-2\frac{1}{NT}\sum_{g_i^0=g}(y_{it} - x_{it}'\mathring{\beta}_{g_i^0,t})x_{it} + \lambda\dot{w}_{g,t}e_{g,t} - \lambda\dot{w}_{g,t+1}e_{g,t+1} = 0,$$

where $e_{g,1} = e_{g,T+1} = 0$, for $2 \leq t \leq T$, $e_{g,t} = (\mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1})/\left\|\mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1}\right\|$ if $\mathring{\beta}_{g,t} - \mathring{\beta}_{g,t-1} \neq 0$ and $\|e_{g,t}\| \leq 1$ otherwise. We thus have

$$\begin{aligned}
&2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\mathring{\beta}_{g_i^0,t})(x_{it}'(\mathring{\beta}_{g_i^0,t} - \beta^{(0)}_{g_i^0,t})) \\
=&\lambda\sum_{g\in\mathbb{G}}\sum_{t=1}^{T}(\dot{w}_{g,t}e_{g,t} - \dot{w}_{g,t+1}e_{g,t+1})'(\mathring{\beta}_{g_i^0,t} - \beta^{(0)}_{g_i^0,t})
\end{aligned}$$

$$=\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}e_{g,t}'((\mathring{\beta}_{g_i^0,t}-\mathring{\beta}_{g_i^0,t-1})-(\beta_{g_i^0,t}^{(0)}-\beta_{g_i^0,t-1}^{(0)})).$$

Let $\mathcal{T}_{m_g,g}$ be the set of t such that $\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\neq 0$ and $\mathcal{T}_{m_g,g}^c=\{2,\ldots,T\}\backslash\mathcal{T}_{m_g,g}$. We have

$$\begin{aligned}
&\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}e_{g,t}'((\mathring{\beta}_{g_i^0,t}-\mathring{\beta}_{g_i^0,t-1})-(\beta_{g_i^0,t}^{(0)}-\beta_{g_i^0,t-1}^{(0)}))\\
&+\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\beta_{g,t}^{(0)}-\beta_{g,t-1}^{(0)}\right\|-\lambda\sum_{g\in\mathbb{G}}\sum_{t=2}^{T}\dot{w}_{g,t}\left\|\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\right\|\\
=&\lambda\sum_{g\in\mathbb{G}}\sum_{t\in\mathcal{T}_{m_g,g}^c}\dot{w}_{g,t}\left(\left\|\beta_{g,t}^{(0)}-\beta_{g,t-1}^{(0)}\right\|-e_{g,t}'(\beta_{g,t}^{(0)}-\beta_{g,t-1}^{(0)})\right)\\
&+\lambda\sum_{g\in\mathbb{G}}\sum_{t\in\mathcal{T}_{m_g,g}}\dot{w}_{g,t}\left(\left\|\beta_{g,t}^{(0)}-\beta_{g,t-1}^{(0)}\right\|-\frac{(\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1})'(\beta_{g,t}^{(0)}-\beta_{g,t-1}^{(0)})}{\left\|\mathring{\beta}_{g,t}-\mathring{\beta}_{g,t-1}\right\|}\right)\geq 0,
\end{aligned}$$

where the last inequality follows by the CS inequality. This implies that

$$\begin{aligned}
\mathring{Q}(\beta^{(0)})-\mathring{Q}(\mathring{\beta})\geq\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}'(\mathring{\beta}_{g_i^0,t}-\beta_{g_i^0,t}^{(0)}))^2=&\frac{1}{T}\sum_{g\in\mathbb{G}}(\mathring{\beta}_g-\beta_g^{(0)})'M(\gamma^0,g,g)(\mathring{\beta}_g-\beta_g^{(0)})\\
\geq&\hat{\rho}\frac{1}{T}\sum_{g\in\mathbb{G}}\left\|\beta_g^{(0)}-\mathring{\beta}_g\right\|^2.
\end{aligned}$$

Hence, by (S.17) and Assumption 2.1, we have that,

$$\frac{1}{T}\sum_{g\in\mathbb{G}}\left\|\beta_g^{(0)}-\mathring{\beta}_g\right\|^2=o_p(T^{-\delta}),$$

which further implies that

$$\left\|\beta_{g,t}^{(0)}-\mathring{\beta}_{g,t}\right\|^2=o_p(T^{1-\delta})$$

for any δ. This gives the desired result.

□

S.7.2.2 Proof of Theorem S.8

Proof. This proof is very similar to that of Theorem 1 in the main text. The only difference is that we consider $\dot{\gamma}$ instead of $\hat{\gamma}$.

As $\mathring{\beta}$ minimizes $\hat{Q}(\beta,\gamma^0)$, $\mathring{\beta}=\beta^{(0)}$ if $\dot{\gamma}=\gamma^0$. We note that

$$\Pr(\dot{\gamma}\neq\gamma^0)=\Pr\left(\max_{1\le i\le N}\mathbf{1}\{\hat{g}_i(\dot{\beta})\neq g_i^0\}=1\right)\le\sum_{i=1}^{N}E\left(\mathbf{1}\{\hat{g}_i(\dot{\beta})\neq g_i^0\}\right).$$

From Lemmas 6 and 7, we have $\Pr(\dot{\beta}\in\mathcal{N}_\eta)\to 1$ for any η. Together with this, the argument made in the proof of Lemma 5 shows that $\max_{1\le i\le N}E\left(\mathbf{1}\{\hat{g}_i(\dot{\beta})\neq g_i^0\}\right)=O(T^{-\delta})$ for any $\delta>0$. This means that

$$\Pr(\dot{\gamma}\neq\gamma^0)\le N\max_{1\le i\le N}E\left(\mathbf{1}\{\hat{g}_i(\dot{\beta})\neq g_i^0\}\right)=o(NT^{-\delta})$$

for any δ. Thus under the condition of the theorem, from Lemma S.14, we have

$$\begin{aligned}
&\Pr\left(\left\|\hat{\theta}^{(0)}_{g,t}\right\|\neq 0,\exists t\in\mathcal{T}^{0c}_{m_g^0,g},g\in\mathbb{G}\right)\\
\le&\Pr\left(\left\{\left\|\hat{\theta}^{(0)}_{g,t}\right\|\neq 0,\exists t\in\mathcal{T}^{0c}_{m_g^0,g},g\in\mathbb{G}\right\},\left\{\dot{\gamma}=\gamma^0\right\}\right)+\Pr\left(\dot{\gamma}\neq\gamma^0\right)\\
=&\Pr\left(\left\{\left\|\mathring{\theta}_{g,t}\right\|\neq 0,\exists t\in\mathcal{T}^{0c}_{m_g^0,g},g\in\mathbb{G}\right\},\left\{\dot{\gamma}=\gamma^0\right\}\right)+\Pr\left(\dot{\gamma}\neq\gamma^0\right)\\
\le&\Pr\left(\left\|\mathring{\theta}_{g,t}\right\|\neq 0,\exists t\in\mathcal{T}^{0c}_{m_g^0,g},g\in\mathbb{G}\right)+\Pr\left(\dot{\gamma}\neq\gamma^0\right)\to 0.
\end{aligned}$$

We therefore have the desired result.

□

S.7.2.3 Proof of Theorem S.9

Proof. Given Lemma S.19 and Theorem S.8, the proof is based on an argument essentially identical to the proof of Corollary 3.4 in Qian and Su (2016) and is thus omitted.

□

S.7.2.4 Proof of Theorem S.10

Proof. The theorem holds using Lemmas S.19 and 11.

□

References

S. Bonhomme and E. Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83:1147–1184, 2015.

M. Feldstein. International differences in social security and saving. *Journal of Public Economics*, 14:225–244, 1980.

N. Loayza, K. Schmidt-Hebbel, and L. Servén. Saving in developing countries: An overview. *The World Bank Economic Review*, 14:393–414, 2000.

H. M. Pesaran, N. Haqile, and S. Sharma. Neglected heterogeneity and dynamics in cross-country savings regressions. In J. Krishnakumar and E. Ronchetti, editors, *Panel Data Econometrics—Future Direction: Papers in Honour of Professor Pietro Balestra*. Amsterdam: North Holland, 2000.

J. Qian and L. Su. Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, 191:86–109, 2016.

L. Su, Z. Shi, and P. C. B. Phillips. Identifying latent structures in panel data. *Econometrica*, 84:2215–2264, 2016.