ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Econometrics

Heterogeneous structural breaks estimation in panel data models with endogenous regressors using instrumental variables GMM

Thijs Bon - 538764

Supervisor: dr. Wendun Wang

Second Assessor: dr. Hanno Reuvers

March 2022

Abstract

This paper puts forward a new model and estimation method that estimates heterogeneous structural breaks in static linear panel data models with endogenous regressors. It extends the grouped adaptive group fused Lasso framework to account for endogenous regressors. This is achieved by replacing least-squares-based coefficient estimation with an instrumental variables GMM estimator that is efficient under heteroskedasticity. Monte Carlo results indicate generally good performance in finite samples and suggest consistent estimation of latent group structure and structural break pattern. The coverage probability of the two-sided nominal 95% confidence interval of the coefficient estimate declines when the time series grows longer. However, the root mean squared error of the coefficient estimate moves substantially towards zero with both a larger number of cross-sectional observations and a longer time series, indicating accurate estimation. Moreover, comparing the Monte Carlo results of two GMM estimation methods suggest that structural break induced heteroskedasticity is important to take into account for consistent estimation of structural breaks when regressors are not fixed.

Contents

Acronyms

IV	instrumental variable
GMM	generalized method of moments
EGMM	efficient generalized method of moments
2SLS	two-stage least-squares
OLS	ordinary least-squares
AGFL	adaptive group fused Lasso
GAGFL	grouped adaptive group fused Lasso
GFE	grouped fixed effects
CCE	common correlated effects
BIC	bayesian information criterion
IC	information criterion
DGP	data generating process
HD	hausdorff distance
RMSE	root mean squared error

1 Introduction

Panel data models allow to effectively make use of the information contained in panel data sets by imposing an underlying data generating process for the observations over both time and individuals in a single comprehensive manner. This explains the increased popularity of their usage in economics and finance.

There are two main considerations that I want to address here regarding the model specification. First, it is important to account for structural changes in slope coefficients which can result from financial crises, disruptive technologies, or economic reform, for example. Additionally, these structural changes may have heterogeneous effects across individuals as both the direction and magnitude of the structural changes will typically differ significantly across individuals (De Wachter & Tzavalis, 2012). Second, the presence of endogenous independent variables leads to inconsistent coefficient estimates for least-squares-based estimators. An example of this is given by Calderón et al. (2015), who analyze the output contribution of infrastructure. They state that unobserved shocks represent a major source of simultaneity as they may affect both firm output and input choices of labor, capital, and infrastructure.

This paper contributes a heterogeneous structural break detection method for linear panel data models with endogenous regressors. To this end, I propose a new model and estimation method. It can be considered an extension of the model and estimation method by Okui & Wang (2021), allowing for time-varying and heterogeneous coefficients. Individual heterogeneity is modeled through a latent group structure, where coefficients are the same within groups and may vary between groups. The time-varying nature of the coefficients is restricted to structural breaks, differing per group. Group membership for each individual is estimated from the data. This allows for the desired model ability of heterogeneous structural breaks in both magnitude and timing.

To deal with the endogeneity of the regressors, instrumental variables (IVs) are used. Provided that valid instruments are available, the corresponding moment conditions are applied in a generalized method of moments (GMM) approach. In contrast with the conventional IV (or two-stage least-squares (2SLS)) estimator, the GMM estimator is efficient under heteroskedasticity (Baum et al., 2003). This is why it is also called the efficient GMM (EGMM)

estimator. Due to structural breaks in the slope coefficients, the variance of the error term is indirectly affected through the variance of the explanatory variable. This happens when the explanatory variable is not fixed and it could induce substantial heteroskedasticity. As a result, there may be heteroskedasticity between the period before and the period after a structural break. In the same manner, heteroskedasticity might exist between groups because the slope coefficients might differ.

In practice, the independent variable might display a variance and hence be random rather than fixed. For example, imprecise measurement or instances in which the independent variable has to be estimated because it cannot be directly observed will give rise to such a variance. Then, the structural break induced heteroskedasticity will be present.

Although 2SLS slope coefficient estimates are still consistently estimated under heteroskedasticity, estimates in finite samples might be poor because the estimation method is inefficient. Especially structural break estimates might suffer because failing to properly take heteroskedasticity into account might obscure any structural break or increase the chance of false detection. Moreover, standard error estimates are inconsistent for 2SLS in the case of heteroskedasticity. It is possible to account for this by adjusting only the standard error estimation method, for example to Newey-West estimates (Newey & West, 1987). By contrast, EGMM coefficient estimates and their standard error estimates are consistent already. They are a generalized version of the Newey-West standard errors.

Model estimation that incorporates EGMM is carried out by several adaptations of the grouped adaptive group fused Lasso (GAGFL) method from Okui & Wang (2021). GAGFL is a hybrid procedure of the grouped fixed effects (GFE) method by Bonhomme & Manresa (2015) and the adaptive group fused Lasso (AGFL) method proposed by Qian & Su (2016). Before estimating group structure and coefficient values, the number of groups is estimated and the tuning parameter of a penalization term involved in structural break detection is determined. Both follow from the minimization of their respective information criterion.

Given the number of groups and the value of the tuning parameter, three algorithms are applied sequentially to arrive at the final parameter estimates. For the first algorithm, I adapt the GFE method to estimate fully time-varying coefficients through EGMM iteratively. Here, many different group pattern initializations (around 100) are tried. Each iteration consists

of three steps. First, I minimize the EGMM quadratic criterion function with respect to the coefficients, conditioning on the previous group memberships estimate. Second, each individual is assigned to the group for which its contribution to the sum of squared errors is minimized, conditioning on the previously computed coefficient estimates. Given correct group memberships, the coefficient estimates are consistent. In turn, provided with the correct coefficient values, group allocation performed through least-squares is consistent as well. In this way, I propose to iteratively solve both objective functions simultaneously. This unconventional approach is observed to provide substantially better parameter estimates than using only the EGMM quadratic criterion function to update parameter estimates. Third, the weighting matrices are updated using the previously computed parameter estimates. These steps are repeated until numerical convergence to obtain preliminary coefficient and group membership estimates.

To obtain intermediary parameter estimates (given the preliminary parameter estimates), a second algorithm is employed. Here, the time-varying nature of the coefficients is given through a pattern of structural breaks. Again, model estimation is conducted iteratively. First, a penalized EGMM objective function as in Qian & Su (2016) is minimized, conditioning on the previous group memberships estimates. Second, group membership is determined by assigning each individual to the group that gives the smallest contribution to the sum of squared errors, conditioning on the slope coefficient estimates from the previous step. Third, the weighting matrices are updated using the parameter estimates from the first two steps. Repeating these steps until numerical convergence gives the intermediary coefficient and group membership estimates. The final estimates are obtained using a non-iterative post-Lasso EGMM estimator, where group structure and breakpoints are fixed at the estimates from the second algorithm and only the coefficients are updated. The complete procedure is titled EGMM GAGFL.

Important characteristics of GAGFL are (i) consistent estimation of the latent group membership structure, (ii) automatic determination of the number of breaks and consistent estimation of breakpoints for each group in one joint step, and (iii) consistent estimation of the regression coefficients with group-specific structural breaks (Okui & Wang, 2021). By changing least-squares-based estimators to EGMM-based estimators, it is not certain whether

the above characteristics still hold. Therefore, I conduct a Monte Carlo simulation study to indicate how applicable these characteristics are to EGMM GAGFL. Moreover, the Monte Carlo simulation is used to study finite sample properties.

The Monte Carlo simulation results indicate good finite sample performance and suggest that the first two of the above three characteristics still hold. The third characteristic (consistent coefficient estimation) might not apply because EGMM GAGFL possibly gives rise to a small bias, as indicated by the failure of the two-sided nominal 95% confidence interval to grow towards one with an increasing length of the time series. However, the root mean squared error of the coefficient estimates is found to be quite close to zero and becomes substantially closer with both a larger number of cross-sectional observations and a longer time series. Furthermore, comparing the break estimation accuracy between EGMM GAGFL and 2SLS GAGFL also indicate that it is important to take heteroskedasticity into account in the case of random regressors.

The remainder of this paper is organized as follows. First, I present a literature review. Second, the methodology is explained. This consists of describing the model setup and the estimation method. Thereafter, the Monte Carlo simulation study is described and the results that follow are presented. Finally, there is a section with concluding remarks.

2 Literature review

The literature underlying the EGMM GAGFL method is threefold. One stream of research relates to the (IV-)GMM literature. Another part is about testing and dating structural breaks. The final part concerns modeling individual heterogeneity. The EGMM GAGFL method builds upon these three streams of literature.

First, I treat the (IV-)GMM literature. 2SLS IV was independently developed by Theil (1953) and Basmann (1957). Sargan (1958) formulated 2SLS as an optimal IV estimator under conditional homoskedasticity. GMM, introduced by Hansen (1982), uses orthogonality conditions to obtain parameter estimates. 2SLS is a special case of GMM. They are the same in the exactly identified case, where you have the same number of instruments as regressors. In the overidentified case and when the errors satisfy all classical assumptions (i.e., the opti-

mal weighting matrix is proportional to the identity matrix) they coincide too (Baum et al., 2003). According to Baum et al. (2003), the usual approach for IV under heteroskedasticity is GMM. Concerning panel data, GMM is widely applied in the context of a dynamic panel. An IV GMM estimator to solve the problem of endogeneity in a dynamic model setup was first proposed by Anderson & Hsiao (1981). Later, Blundell & Bond (1998) advocate to use extra moment conditions as suggested by Arellano & Bover (1995). Endogenous regressors in panel data have been considered by Neal (2015), using 2SLS as an IV method within the framework of the common correlated effects (CCE) estimator (Pesaran, 2006).

Second, modelling and estimation of individual heterogeneity have been performed often. Examples are the random coefficients model (Swamy, 1970), mean group estimation (Pesaran & Smith, 1995), pooled mean group estimation (Pesaran et al., 1999), and the CCE estimator Pesaran (2006). These examples allow for different slope coefficients between any pair of individuals. By contrast, I assume a latent group structure is present in the panel data. Other work that has been done regarding latent group structure modelling and estimation include Sun (2005), Hahn & Moon (2010), Lin & Ng (2012), Bonhomme & Manresa (2015), Su et al. (2016), Ando & Bai (2016), **?**, Wang et al. (2018), Miao et al. (2020), and Wang & Su (2021), among others. All these studies assume a stable panel without structural change.

Third, much research relates to testing and dating common structural breaks in panel data. These include Bai (2010), Kim (2011), Qian & Su (2016), Li et al. (2016), and Baltagi et al. (2017), among others. By contrast, I consider structural breaks that may vary between individuals in both magnitude and timing.

Finally, I discuss research that encompasses a combination of the three streams of literature. An example of research that incorporates a latent group structure to model heterogeneous structural change is Su et al. (2019), who model structural change through continuous time-varying slope coefficients. De Wachter & Tzavalis (2012) study structural breaks in dynamic panel data models with endogenous regressors. However, they do not consider heterogeneous coefficients. Okui & Wang (2021) do consider heterogeneous structural breaks using a latent group structure, but they assume exogenous regressors. Qian & Su (2016) allow for endogenous regressors. However, they consider structural breaks to be common across individuals. I contribute to the literature by proposing a model and estimation method that models indi-

vidual heterogeneity of both slope coefficients and structural breaks through a latent group structure in the presence of endogenous covariates.

3 Methodology

In this section, I treat the methodology. The methodology consists of the model setup and the estimation method. First, the model setup describes the assumed panel data model. Afterward, the estimation method details the employed method to estimate the model parameters.

3.1 Model setup

The model setup is as follows. I consider a static linear panel data model with time-varying and heterogeneous coefficients. For model parsimony, heterogeneity of coefficients is restricted to a grouped pattern. Moreover, structural breaks are considered for each group to restrict how coefficients vary over time. The panel data model is

$$y_{it} = \beta'_{g_i,t} x_{it} + \varepsilon_{it}, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{1}$$

where y_{it} is the scalar dependent variable for observational unit i at time t. The total number of cross-sectional observations is N, and T denotes the length of the time series. The independent variables are captured in the $k \times 1$ vector $x_{it} = (1, x'_{it,exo}, x'_{it,pred}, x'_{it,endo})'$. The 1 is typically included as a first element to incorporate a constant in the model. In this manner, a group-specific fixed effect (which might even have structural breaks) may be included as an alternative to the individual-specific fixed effect. Each independent variable may be exogenous, predetermined or endogenous. They are incorporated into the vector $x_{it,exo}$, $x_{it,pred}$, or $x_{it,endo}$, respectively. The length of these column vectors corresponds to the number of exogenous, predetermined, and endogenous independent variables, respectively. The idiosyncratic error term (with a mean of zero) is ε_{it}, it has a variance of σ^2. In the case that only the intercept is allowed to vary over time, the model coincides with the GFE model (Bonhomme & Manresa, 2015).

The coefficients to be estimated for every group and time step combination are captured by the regressors' $k \times 1$ coefficient vector $\beta_{g_i,t}$. They are homogeneous within groups and are allowed to be heterogeneous between groups. I assume that there are G groups, let $\mathbb{G} = \{1, 2, \ldots, G\}$ be the set of groups and let $g_i \in \mathbb{G}$ denote the group membership of unit i. The number of groups G is estimated using a Bayesian information criterion (BIC), this is described in Section 3.3. The number of individuals in group g is given by N_g. All individuals within a group share $\beta_{g,t}$ as their time-varying coefficients, where $g \in \mathbb{G}$. Group membership structure $\{g_i\}_{i=1}^{N}$ is unknown and has to be estimated.

The time-varying nature of the coefficients is as follows. I assume that, for each group, structural breaks characterize the pattern $\{\beta_{\mathrm{g},1}, \beta_{\mathrm{g},2}, \ldots, \beta_{\mathrm{g},T}\}$. The coefficients are restricted to change only at a break date, as such they remain the same in between two consecutive break dates. Each group has m_g break dates, let $\mathcal{T}_{m_g,g} = \{T_{g,1}, T_{g,2}, \ldots, T_{g,m_g}\}$ denote the set of break dates for group g. Both m_g and $\mathcal{T}_{m_g,g}$ are unknown and have to be estimated from the data. Let $\mathcal{T}_m = \{T_1, T_2, \ldots, T_m\}$ denote the set of all break dates, which consists of every date on which a break occurs within any of the groups. Then, there are m break dates in total. Coefficient values in between break dates are expressed as follows. Let $\alpha_{g,j},\ j = 1, 2, \ldots, m_{g+1}$ denote the $k \times 1$ coefficient vector for regime j, which ranges from break date $j-1$ up until but not including break date j for group g,

$$\alpha_{g,j} = \beta_{\mathrm{g},t}, \quad \text{if} \quad T_{\mathrm{g},j-1} \leq t < T_{\mathrm{g},j}, \tag{2}$$

where $T_{g,0} = 1$ and $T_{g,m_g+1} = T + 1$ allow for coefficient values in the first and last period, respectively. Similarly, $T_0 = 1$ and $T_{m+1} = T + 1$.

Suppose one or more of the independent variables is not fixed but rather random. Then, heteroskedasticity might exist, following the heterogeneous structure of groups and structural breaks. This is illustrated by

$$Var(\beta'_{g_i,t} x_{it}) = (\beta'_{g_i,t})^2 Var(x_{it}) \neq 0, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T. \tag{3}$$

Imposing a group structure and restricting coefficients to vary in time only at break dates allows for a flexible and parsimonious way to model individual heterogeneity and structural change. This way, I can make use of cross-sectional variation in break date and coefficient

estimation (Bai, 2010).

3.2 Estimation method

To estimate the parameters discussed in Section 3.1, an estimation method is required. The estimation procedure is named EGMM GAGFL. It will be treated in this section. EGMM GAGFL is an extension of GAGFL (Okui & Wang, 2021) to endogenous variable settings. GAGFL combines GFE by Bonhomme & Manresa (2015) and AGFL by Qian & Su (2016). EGMM refers to the usage of GMM within the GAGFL framework in ensuring consistent coefficient estimates in the case of endogeneity issues, while being efficient ('E'GMM) under heteroskedasticity.

I first introduce some notation. Let $\beta = (\beta'_{1,1}, \beta'_{1,2}, \ldots, \beta'_{1,T}, \beta'_{2,1}, \beta'_{2,2}, \ldots, \beta'_{\mathrm{G},T})'$ denote the vector stacking all $\beta_{g,t}$ and let $\mathcal{B} \subset \mathbb{R}^k$ be the parameter space for each $\beta_{g,t}$. It follows that the parameter space for β is $\mathcal{B}^{GT}$. Let $\beta_g = (\beta'_{\mathrm{g},1}, \beta'_{\mathrm{g},2}, \ldots, \beta'_{\mathrm{g},T})'$ denote the vector stacking all $\beta_{g,t}$ for group g. Let θ denote the vector that contains the group membership information for each individual, i.e., $\theta = \{g_1, g_2, \ldots, g_N\}$. Hence, the parameter space for θ is $\mathbb{G}^N$.

To estimate (β, θ), I propose to jointly optimize two objective functions iteratively. For now, I assume that the number of groups G is known. To start, a random group pattern initialization is used. Conditioning on the group membership structure, fully time-varying coefficient estimates can be obtained through EGMM—provided that enough valid moment conditions are available. I assume the existence of an $m \times 1$ vector of instrumental variables z_{it}. Then, the moment conditions are

$$E\left[z_{it}\left(\varepsilon_{it}\right)\right] = E\left[z_{it}\left(y_{it} - \beta'_{g_i,t} x_{it}\right)\right] = 0, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{4}$$

To perform EGMM estimation of β, it is required that $m \geq k$, so that the coefficients are identified. In the exactly identified case ($m = k$), it suffices to solve (4) directly. By contrast, when $m > k$, coefficient estimates are obtained by minimizing

$$Q_{NT}(\beta, \theta) = \sum_{g=1}^{G} \sum_{t=1}^{T} \left(\frac{1}{N_g} \sum_{i|g_i=g} f_{it}(\beta_{g_i,t}) \right)' W_{g,t} \left(\frac{1}{N_g} \sum_{i|g_i=g} f_{it}(\beta_{g_i,t}) \right), \tag{5}$$

with $f_{it}(\beta_{g_i,t}) = z_{it}\varepsilon_{it} = z_{it}\left(y_{it} - \beta'_{g_i,t} x_{it}\right)$ an $m \times 1$ vector containing the moment condition

functions for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, and $W_{g,t}$ an $m \times m$ symmetric and positive definite weighting matrix for $g = 1, 2, \ldots, G$ and $t = 1, 2, \ldots, T$. Assuming that correct group membership estimates are known, Q_{NT} is quadratic in the parameter vector. Subsequently, minimization of (5) is a convex optimization problem, and solving it thus gives consistent coefficient estimates. Furthermore, the choice of weighing matrix does not influence consistency but rather efficiency. This is indicated by the fact that 2SLS, for which $W_{g,t} = I_m$ for $g = 1, 2, \ldots, G$ and $t = 1, 2, \ldots, T$, also gives consistent coefficient estimates.

It can be analytically proven that using the inverse of the covariance matrix of the moment conditions as weighing matrices yields asymptotically efficient parameter estimates. Thus, $W_{g,t} = (E[f_{g_i,t}(\beta_{g_i,t}) f_{g_i,t}(\beta_{g_i,t})'])^{-1}$. For example, suppose some estimates $(\tilde{\beta}, \tilde{\theta})$ exist. Then, the corresponding sample analogue for $W_{g,t}$ is $\tilde{W}_{g,t} = \left(\frac{1}{N_g}(\sum_{i|g_i=g} f_{it}(\tilde{\beta}_{g_i,t}) f_{it}(\tilde{\beta}_{g_i,t})')\right)^{-1}$, for $g = 1, 2, \ldots, G$ and $t = 1, 2, \ldots, T$, where $\sum_{i|g_i=g}$ sums over all individuals within group g. Let $W = (W_{1,1}, W_{1,2}, \ldots, W_{1,T}, W_{2,1}, W_{2,2}, \ldots, W_{G,T})'$ denote the vector stacking all $W_{g,t}$. I use time-varying weight matrices because structural breaks might lead to substantial heteroskedasticity, this would in turn make it difficult to accurately detect structural breaks for weighting matrices that are constant over time.

In contrast with coefficient estimation through minimization of the EGMM objective function (5), estimation of θ is performed using the least-squares objective function

$$LS_{NT}(\beta, \theta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \beta'_{g_i,t} x_{it}\right)^2. \tag{6}$$

Therefore, iterative optimization of two objective functions with interdependent parameter estimates is conducted simultaneously. The reasoning behind the convergence of both is that both estimators provide consistent estimates given correct parameter estimates from the other estimator. As such, they are likely to enable the consistency of each other, at least for certain initial grouping estimates. Consequently, I propose to perform group membership estimation by minimizing (6).

Moreover, I attempted to update group membership using the EGMM objective function (5) instead of performing group allocation through least-squares. This way I would iteratively optimize a single objective function instead of two. This would make the method coherent by providing a more justifiable theoretical motivation. The implementation specifics are

discussed in the last paragraph of Section 3.2.1. Surprisingly, Monte Carlo simulations with this estimation method setup gave very poor parameter estimates. Hence, I decided to abandon this approach in favor of the least-squares approach for group allocation.

3.2.1 EGMM estimation

Presuming the existence of enough valid instruments, consistent fully time-varying parameter estimates can be obtained by a GFE-type estimator. Let $\dot{\beta}$ be the preliminary estimate of β, resulting from a minimization of the EGMM quadratic criterion function (5). Bonhomme & Manresa (2015) use a least-squares objective function in an estimation method that is similar but assumes exogenous regressors. Let $\dot{\theta}$ be the preliminary estimate of θ and let $\dot{W}$ be the preliminary estimate of W. Preliminary parameter estimates result jointly from

$$\dot{\beta} = \underset{\beta \in \mathcal{B}^{GT}}{\arg\min}\ Q_{NT}(\beta, \theta), \tag{7}$$

and,

$$\dot{\theta} = \underset{\theta \in G^N}{\arg\min}\ LS_{NT}(\beta, \theta). \tag{8}$$

$\dot{W}$ is computed directly from $(\dot{\beta}, \dot{\theta})$. Computationally, I propose to use Algorithm 1 to obtain the preliminary parameter estimates iteratively. Algorithm 1 starts by generating a random group pattern by assigning each individual to one of the groups randomly. This pattern forms the initial group membership estimates. Afterward, Algorithm 1 enters a loop in which the parameters that minimize Q_{NT} are computed iteratively.

The steps in this loop are as follows. First, coefficient estimates are updated by minimizing Q_{NT} with respect to the coefficients, conditioning on the most recent estimates of group memberships and weighting matrices. To have fully time-varying coefficients, data and weighting matrices are transformed. I will show how this is done for X, for Z and W a similar operation is performed. For regressor k,

$$X^{(k)}_{g,tot} = (diag(x^{(k)}_{g_{i_1},1}, \ldots, x^{(k)}_{g_{i_1},T}), diag(x^{(k)}_{g_{i_2},1}, \ldots, x^{(k)}_{g_{i_2},T}), \ldots, diag(x^{(k)}_{g_{i_{N_g}},1}, \ldots, x^{(k)}_{g_{i_{N_g}},T}))',$$

where $x^{(k)}_{g_{i_n},t} = x^{(k)}_{it}$ for individual n in group g and period t. Then, $X_{g,tot} = (X^{(1)}_{g,tot}, \ldots, X^{(K)}_{g,tot})$.

Algorithm 1: EGMM estimator

Parameters: Regressors' coefficient vector (β), grouped pattern vector (θ), and weight matrices vector (W).

Hyper parameters: Number of groups (G), maximum number of iterations (s_{max}), and number of simulations (N_{sim}).

Input: Dependent variable $NT \times 1$ vector (y), independent variable $NT \times k$ matrix (X), and $NT \times m$ instrument matrix (Z).

Output: Preliminary parameter estimates $(\dot{\beta}, \dot{\theta}, \dot{W})$.

Initialization: Set $n = 1$ and $R_{min} = 10^7$.

1. **Generate a random group pattern:** For $i \in \{1, 2, \ldots, N\}$, set $g_i = random(\mathbb{G})$. Then, let $\theta^{(0)}$ be the resulting initial group membership estimates. Set $s = 0$, $\beta^{(0)} = 0$, and $W_{g,t}^{(0)} = I_m$ for $g = 1, 2, \ldots, G$ and $t = 1, 2, \ldots, T$.

2. **Obtain parameter estimates iteratively:**

   (a) **Update coefficients:** Given $\theta^{(s)}$ and $W^{(s)}$, set

   $$\beta_g^{(s+1)} = \begin{cases} (Z'_{g,tot} X_{g,tot})^{-1} Z'_{g,tot} y_g & \text{if m=k} \\ (X'_{g,tot} Z_{g,tot} W_{g,tot}^{(s)} Z'_{g,tot} X_{g,tot})^{-1} X'_{g,tot} Z_{g,tot} W_{g,tot}^{(s)} Z'_{g,tot} y_g & \text{if m} > \text{k,} \end{cases} \tag{9}$$

   for all $g \in \mathbb{G}$.

   (b) **Update group memberships:** For all $i \in \{1, 2, \ldots, N\}$, set

   $$g_i^{(s+1)} = \arg\min_{g \in \mathbb{G}} \sum_{t=1}^{T} \left(y_{it} - \beta_{g_i,t}^{(s+1)\prime} x_{it} \right)^2. \tag{10}$$

   Then, $\theta^{(s+1)}$ follows.

   (c) **If m > k, update weighting matrices:**

   $$W_{g,t}^{(s+1)} = \left(\frac{1}{N_g} \left(\sum_{i|g_i=g} f_{it}(\beta_{g_i,t}^{(s+1)}) f_{it}(\beta_{g_i,t}^{(s+1)})' \right) \right)^{-1}, \tag{11}$$

   for $g = 1, 2, \ldots, G$ and $t = 1, 2, \ldots, T$.

   (d) **Check stopping criteria:** If $\left\| \beta^{(s+1)} - \beta^{(s)} \right\|_2 > 0$ and $s + 1 < s_{max}$, set $s = s + 1$ and go to Step 2a. Otherwise, set $(\beta, \theta, W)_n = (\beta, \theta, W)^{(s+1)}$ and continue to Step 3.

3. **Obtain estimates and check stopping criteria:** If $\sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \beta'_{g_i,t} x_{it} \right)^2 < R_{min}$, update R_{min} and set $(\dot{\beta}, \dot{\theta}, \dot{W}) = (\beta, \theta, W)_n$. If $n < N_{sim}$, set $n = n + 1$ and go to Step 1. Otherwise (when $n = N_{sim}$), stop.

Second, each individual is assigned to the group for which its contribution to the sum of squared errors is minimized, conditioning on the previously computed coefficient estimates. As previously explained, group allocation is performed based on least-squares instead of EGMM.

Third, the weighting matrices are updated using the new parameter estimates from the previous two steps. These three steps are repeated until numerical convergence (applying the L_2 norm) or until the maximum number of iterations has been reached. After iterating, I check if the obtained parameter estimates result in the lowest value of the sum of squared errors thus far. If that is the case, the minimum value of the sum of squared errors and the preliminary parameter estimates are updated accordingly.

This whole simulation process is repeated by generating a random group pattern again and following all subsequent steps—that is until the maximum number of simulations has been reached. By running multiple simulations (that have different initial group memberships estimates), Algorithm 1 likely avoids getting preliminary parameter estimates corresponding to a local minimum that is markedly greater than the global minimum. Running a large number of simulations is the main computational effort, computation time increases linearly with this number (Okui & Wang, 2021).

The alternative approach to estimate group membership by minimizing the EGMM objective function (5) is as follows. I compute individual contributions in Step 2b of Algorithm 1 (per group) by

$$\begin{aligned} Q_{NT_{i^*,g}}(\beta_g, \theta) = &\sum_{t=1}^{T} \left(\frac{1}{N_g} \sum_{i|g_i=g} f_{it_{+i^*}}(\beta_{g,t}) \right)' W_{g,t} \left(\frac{1}{N_g} \sum_{i|g_i=g} f_{it_{+i^*}}(\beta_{g,t}) \right) \\ &- \sum_{t=1}^{T} \left(\frac{1}{N_g} \sum_{i|g_i=g} f_{it_{-i^*}}(\beta_{g,t}) \right)' W_{g,t} \left(\frac{1}{N_g} \sum_{i|g_i=g} f_{it_{-i^*}}(\beta_{g,t}) \right), \end{aligned} \tag{12}$$

where $f_{it_{+i^*}}(\beta_{g,t}) = 1$ for $i = i^*$ and $f_{it_{-i^*}}(\beta_{g,t}) = 0$ for $i = i^*$. Equation (12) subtracts the contribution of group g to the objective function in case i^* is included from the case in which i^* is not included. Since individual contributions are interdependent, this indirect approach allows to obtain the individual contributions per group. But, as aforementioned, Monte Carlo simulations indicated poor performance.

3.2.2 Penalized EGMM estimation

To restrict the time-varying nature of the coefficient estimates to a pattern of structural breaks, I follow Qian & Su (2016) by introducing a penalty term to (5). Subsequently, I jointly estimate (β, θ) by

$$\hat{\beta} = \underset{\beta \in \mathcal{B}^{GT}}{\arg\min} \left(Q_{NT}(\beta, \theta) + \sum_{g=1}^{G} \lambda_g \sum_{t=1}^{T} \dot{w}_{g,t} \left\| \beta_{g,t} - \beta_{g,t-1} \right\|_F \right), \tag{13}$$

and,

$$\hat{\theta} = \underset{\theta \in \mathbb{G}^N}{\arg\min}\ LS_{NT}(\beta, \theta). \tag{14}$$

The left-hand side of (13) term is the EGMM quadratic objective function, whereas the right-hand side term gives the penalty term. The group-dependent tuning parameter of the penalty term is λ_g, which is discussed in more detail in Section 3.3. The data-driven weight $\dot{w}_{g,t}$ is

$$\dot{w}_{g,t} = \left\| \dot{\beta}_{g,t} - \dot{\beta}_{g,t-1} \right\|_{k+1}^{-\kappa} = \left(\left\{ \sum_{p=1}^{k} \left(\left\| \dot{\beta}_{g,t} - \dot{\beta}_{g,t-1}) \right\|_1^k \right)_p \right\}^{k^{-1}} \right)^{-\kappa}, \tag{15}$$

where κ is a user specified constant (typically $\kappa = 2$) and $\left(\left\| \dot{\beta}_{g,t} - \dot{\beta}_{g,t-1}) \right\|_1^k \right)_p$ denotes entry number p within the first differenced (in absolute terms) coefficient vector, after exponentiation by k. Furthermore, $\|\cdot\|_F$ denotes the Frobenius norm.

The use of the Frobenius norm (instead of the usual L_1 norm) in the penalization term in (13) for the first difference vector $\beta_{g,t} - \beta_{g,t-1}$ generalizes the fused Lasso (Tibshirani et al., 2005) to the group fused Lasso. Usage of the adaptive weights $\dot{w}_{g,t}$ is what gives rise to the procedure name of adaptive group fused Lasso (AGFL) (Qian & Su, 2016). Building on this, Okui & Wang (2021) consider heterogeneous structural breaks through heterogeneous coefficients in a grouped pattern. To estimate the grouped pattern, they employ a GFE estimator (Bonhomme & Manresa, 2015). Okui & Wang (2021) title the resulting estimation procedure GAGFL. By allowing for endogenous regressors, and by adapting the estimation procedure accordingly, I extend the existing literature. I do this by replacing least-squares-based coefficient estimation with efficient IV-GMM, which is where the name EGMM GAGFL

originates from. Moreover, to achieve this, iterative estimation of the weighting matrices is incorporated within the iterative parameter estimation.

The motivation for the use of the Frobenius norm within EGMM GAGFL, instead of the L_1 norm, is to have accurate structural break detection. By using the Frobenius norm, the differences between adjacent (in a temporal sense) coefficient vectors are penalized instead of differences between scalar coefficients within the vectors. To have accurate detection of structural breaks, the used norm should give a good indication of the overall similarity between coefficient vectors. Thus, the Frobenius norm is most applicable.

Having obtained preliminary parameter estimates $(\dot{\beta}, \dot{\theta}, \dot{W})$ using Algorithm 1, the weights $\dot{w}_{g,t}$ can be computed as in (15). Subsequently, intermediary parameter estimates $(\hat{\beta}, \hat{\theta}, \hat{W})$ are obtained employing the penalized EGMM objective function (13) and the least-squares objective function (14). I propose to use Algorithm 2, very similar to Algorithm 1, to compute the intermediary parameter estimates.

Instead of randomly sampling multiple initial group patterns, Algorithm 2 utilizes the preliminary group memberships estimate $\dot{\theta}$ from Algorithm 1 as the single initial group memberships estimate. Provided that $\dot{\beta}$ and $\dot{\theta}$ are consistent, the initial parameter estimates are close to the (global) minimum and hence Algorithm 2 converges fast and there is no need for multiple initial group patterns.

Iterative minimization is performed in the same manner as the iteration steps in Algorithm 1, except that the penalized EGMM objective function is used in the coefficients update step instead of just the EGMM objective function (7). Instead of analytically solving the objective function, a block coordinate descent algorithm is used to numerically solve (16). Group memberships in Step 2 are updated by minimizing the sum of squared residuals. Furthermore, the penalization term does not depend directly on the group pattern and therefore need not be included in the group assignment step.

The estimated break dates are periods for which $\hat{\beta}_{g,t} - \hat{\beta}_{g,t-1} \neq 0$. Let the set of estimated break dates for group g be denoted by $\hat{\mathcal{T}}_{\hat{m}_g,g} = \left\{ t \in \{1, 2, \ldots, T\} \mid \hat{\beta}_{g,t} - \hat{\beta}_{g,t-1} \neq 0 \right\}$. The estimated number of breaks for group g ($\hat{m}_g$) is then given by the cardinality of $\hat{\mathcal{T}}_{\hat{m}_g,g}$. Consequently, the estimated number of unique coefficient vectors for group g is $\hat{m}_g + 1$. Furthermore, let $\hat{\mathcal{T}}_{\hat{m}} = \left\{ t \in \{1, 2, \ldots, T\} \mid \hat{\beta}_{g,t} - \hat{\beta}_{g,t-1} \neq 0 \; \forall \; g \in \mathbb{G} \right\}$ denote the set of all

Algorithm 2: penalized EGMM estimator

Parameters: Regressors' coefficient vector (β), grouped pattern vector (θ), and weight matrices vector (W).

Hyper parameters: Number of groups (G), maximum number of iterations (s_{max}), and tuning parameter (λ_g).

Input: Dependent variable $NT \times 1$ vector (y), independent variable $NT \times k$ matrix (X), and $NT \times m$ instrument matrix (Z). Preliminary parameter estimates $(\dot{\beta}, \dot{\theta}, \dot{W})$.

Output: Intermediary parameter estimates $(\hat{\beta}, \hat{\theta}, \hat{W})$.

Initialization: Set $\beta^{(0)} = \dot{\beta}$, $\theta^{(0)} = \dot{\theta}$, $W^{(0)} = \dot{W}$, and $s = 0$.

1. **Update coefficients:** Given $\theta^{(s)}$ and $W^{(s)}$, set

$$\beta^{(s+1)} = \underset{\beta \in \mathcal{B}^{GT}}{\arg\min}\ Q_{NT}(\beta, \theta^{(s)}) + \sum_{g=1}^{G} \lambda_g \sum_{t=1}^{T} \dot{w}_{g,t} \left\| \beta_{g,t} - \beta_{g,t-1} \right\|_F , \tag{16}$$

2. **Update group memberships:** For $i \in \{1, 2, \ldots, N\}$, set

$$g_i^{(s+1)} = \underset{g \in \mathbb{G}}{\arg\min}\ \sum_{t=1}^{T} \left(y_{it} - \beta_{g_i,t}^{(s+1)\prime} x_{it} \right)^2 . \tag{17}$$

   Then, $\theta^{(s+1)}$ follows.

3. **If m > k, update weighting matrices:**

$$W_{g,t}^{(s+1)} = \left(\frac{1}{N_g} \Big(\sum_{i | g_i = g} f_{it}(\beta_{g_i,t}^{(s+1)}) f_{it}(\beta_{g_i,t}^{(s+1)})' \Big) \right)^{-1} , \tag{18}$$

   for $g = 1, 2, \ldots, G$ and $t = 1, 2, \ldots, T$.

4. **Check stopping criteria:** If $\left\| \beta^{(s+1)} - \beta^{(s)} \right\|_2 > 0$ and $s + 1 < s_{max}$, set $s = s + 1$ and go to Step 1. Otherwise, set $(\hat{\beta}, \hat{\theta}, \hat{W}) = (\beta, \theta, W)^{(s+1)}$ and stop.

unique estimated break dates over all groups, where $\forall$ should be interpreted as 'for any'. It follows that the estimated total number of breaks $\hat{m}$ is given by the cardinality of $\hat{\mathcal{T}}_{\hat{m}}$.

Consistent estimation of break dates is achieved through appropriate weights ($\dot{w}_{g,t}$) that result from consistent preliminary coefficient estimates ($\dot{\beta}_{g,t}$). To illustrate this, suppose $\beta_{g,t} - \beta_{g,t-1} = 0$, then it is likely that $\dot{\beta}_{g,t} - \dot{\beta}_{g,t-1} \approx 0$. If that is the case, it can be seen from (15) that $\dot{w}_{g,t}$ will be large. That results in a heavy penalty, and presumably no

estimated breakpoint for the corresponding period. This facilitates break dates to be sparse and consistently estimated.

3.2.3 Post-Lasso EGMM estimation

After obtaining intermediary group pattern estimate $\hat{\theta}$, intermediary weight matrices estimate $\hat{W}$, and estimated set of break dates $\hat{\mathcal{T}}_{\hat{m}_g,g}$ for each group g, post-Lasso coefficient estimates are computed. Let $\alpha_{g,m} = \left(\alpha'_{g,1}, \alpha'_{g,2}, \ldots, \alpha'_{g,m_g+1}\right)'$ denote the vector stacking all $\alpha_{g,j}$ for group g. Post-Lasso coefficient estimates are obtained by performing EGMM for each period in each group. The objective function for group g is

$$V_{NT,g}(\alpha_{g,m}, \mathcal{T}_{m_g,g}, W) = \sum_{j=1}^{m_g+1} \sum_{t=T_{g,j-1}}^{T_{g,j}-1} \left(\frac{1}{N_g} \sum_{i|g_i=g} f_{it}(\alpha_{g,j}) \right)' W_{g,t} \left(\frac{1}{N_g} \sum_{i|g_i=g} f_{it}(\alpha_{g,j}) \right), \quad (19)$$

where $f_{it}(\alpha_{g,j}) = z_{it}\left(y_{it} - \alpha'_{g,j} x_{it}\right)$. Minimization of (19), which is performed for all $g \in \mathbb{G}$, gives $\hat{\alpha}_{g,m}(\hat{\mathcal{T}}_{\hat{m}_g,g}, \hat{W}) = \left(\hat{\alpha}'_{g,1}, \hat{\alpha}'_{g,2}, \ldots, \hat{\alpha}'_{g,m_g+1}\right)'$. Similar to Step 2a in Algorithm 1, coefficients are computed by

$$\hat{\alpha}_{g,j} = \begin{cases} (Z'_{g,j} X_{g,j})^{-1} Z'_{g,j} y_{g,j} & \text{if m=k} \\ \sum_{t=T_{g,j-1}}^{T_{g,j}-1} (X'_{g,t} Z_{g,t} W_{g,t} Z'_{g,t} X_{g,t})^{-1} X'_{g,t} Z_{g,t} W_{g,t} Z'_{g,t} y_{g,t} & \text{if m} > \text{k}, \end{cases} \quad (20)$$

for all $j \in \{1, 2, \ldots, m_{g+1}\}$ in each $g \in \mathbb{G}$. Here, $X_{g,j}$ consists of all panel data x_{it} for which both $i \in \{1, 2, \ldots, N\} \mid g_i = g$, and $t \in \{1, 2, \ldots, T\} \mid T_{\mathrm{g},j-1} \leq t < T_{\mathrm{g},j}$. Furthermore, $X_{g,t}$ consists of all panel data x_{it} for which $i \in \{1, 2, \ldots, N\} \mid g_i = g$ and $t = t$. The other terms in (20) are formed in a similar manner. Standard error estimates of the coefficient estimates are obtained by computing GMM standard errors as in Section 4.4.3 of Heij et al. (2004).

3.3 Selecting the number of groups and the penalty term tuning parameter

Prerequisites of using EGMM GAGFL are the selection of the number of groups G and selecting the tuning parameter of the penalty term λ_g in (13). Those issues are discussed in this section.

First, to select the number of groups G, a Bayesian information criterion is used,

$$BIC(G) = \frac{1}{NT} \sum_{g=1}^{G} \sum_{j=1}^{m_g+1} \sum_{i|g_i=g} \sum_{t=T_{g,j-1}}^{T_{g,j}-1} \left(y_{it} - \hat{\alpha}'_{g,j} x_{it}\right)^2 + \hat{\sigma}^2 \frac{n_p(G) + N}{NT} \ln NT, \tag{21}$$

which is adapted from Bonhomme & Manresa (2015). Here, $\hat{\sigma}^2$ is a scaling parameter that can be obtained from an estimate of the variance of the idiosyncratic error term. I use $\hat{\sigma}^2 = \frac{1}{NT} \sum_{j=1}^{m_1+1} \sum_{i=1}^{N} \sum_{t=T_{g,j-1}}^{T_{g,j}-1} \left(y_{it} - \hat{\alpha}'_{1,j} x_{it}\right)^2$, where the coefficient estimates $\hat{\alpha}_{1,j}$ result from applying EGMM GAGFL to a homogeneous panel ($G = 1$). Furthermore, $n_p(G) = \sum_{g=1}^{G}((\hat{m}_g + 1)k)$ gives the total number of estimated coefficients. To implement the selection of G, compute $BIC(G)$ for $G \subset \{1, 2, \ldots, G_{max}\}$, where G_{max} is an upper bound that needs to be known. Selection of G_{max} is not considered here. Subsequently, the estimated number of groups $\hat{G}$ is given by: $\underset{G \subset \{1,2,\ldots,G_{max}\}}{\arg\min} \quad BIC(G)$.

Since the squared residuals might differ among both groups and regimes (heteroskedasticity), and because the number of groups might be larger than one, $\hat{\sigma}^2$ probably yields an inconsistent estimate of σ^2. However, the reason for including $\hat{\sigma}^2$ is to scale the penalty term such that it is unaffected by the variation of the data (Okui & Wang, 2021). Okui & Wang (2021) find that using the upper bound for the number of groups (G_{max}) to estimate the variance of the idiosyncratic error term, as suggested for GFE by Bonhomme & Manresa (2015), may lead to unstable results. They advocate the use of a single group to estimate σ^2.

The BIC represents a tradeoff between model fitness and model parsimony. Generally, a larger value of G leads to a larger number of parameters and a better fit. However, this is not necessarily the case since more groups might result in fewer breaks over all groups and thus a lower number of coefficients and a worse model fit. Nevertheless, according to Okui & Wang (2021), the model that coincides with the data generating process gives the lowest BIC value, provided an appropriate choice of the tuning parameters λ_g.

Next, selection of tuning parameter λ_g is discussed. Similar to Okui & Wang (2021), I minimize the following information criterion (IC):

$$IC_g(\lambda_g) = \frac{1}{NT} \sum_{j=1}^{m_g+1} \sum_{i|g_i=g} \sum_{t=T_{g,j-1}}^{T_{g,j}-1} \left(y_{it} - \hat{\alpha}'_{g,j} x_{it}\right)^2 + \rho_{NT} \left(\hat{m}_{\lambda_g} + 1\right) k, \tag{22}$$

where $\hat{m}_{\lambda_g}$ is the estimated number of breaks when tuning parameter λ_g is used, and ρ_{NT} is a

tuning parameter. Here, I choose $\rho_{NT} = c\ln(NT)/\sqrt{NT}$ with $c = 0.05$, following Qian & Su (2016). The tuning parameter differs per group because the variance of the idiosyncratic error term and number of breaks might differ per group as well. Therefore, the tuning parameter value which minimizes (22) might differ too. Both Okui & Wang (2021) and Qian & Su (2016) verify via simulation that the performance of their method is robust to the choice of c as long as it lies in a reasonable range. To implement the selection of λ_g, I compute $IC(\lambda_g)$ for $\lambda_g \subset \{\lambda_{g,min}, \ldots, \lambda_{g,max}\}$, where $\lambda_{g,min}$ and $\lambda_{g,max}$ are bounds that need to be known. In between those bounds, there are a number of evenly spaced values—that is, on a logarithmic scale. The exact determination of the bounds and the number of values in between is not considered here. Subsequently, the value of the tuning parameter λ_g is given by $\underset{\lambda_g \subset \{\lambda_{g,min},\ldots,\lambda_{g,max}\}}{\arg\min} IC(\lambda_g)$.

4 Monte Carlo simulation

In this section, I evaluate the finite sample performance of the proposed EGMM GAGFL method for linear and static panel data with endogenous regressors by conducting a set of Monte Carlo experiments. Moreover, I perform a comparison between EGMM GAGFL and two other estimation methods. One of those methods is GAGFL, which differs from EGMM GAGFL in the sense that it uses ordinary least-squares (OLS) for coefficient estimation. The other method is 2SLS GAGFL and it discerns itself from the other two by employing 2SLS for coefficient estimation. The 2SLS and EGMM coefficient estimation techniques are equivalent when the number of instruments is equal to the number of parameters. In this case, they both boil down to a simple IV estimator which directly solves the moment condition.

For each method, following Qian & Su (2016), the bounds of the tuning parameter λ_g (i.e., $\lambda_{g,max}$ and $\lambda_{g,min}$) are selected so that $\lambda_{g,max}$ would yield no breaks and $\lambda_{g,min}$ would yield breaks in all time points. Furthermore, 50 logarithmically evenly distributed values in the interval of $[\lambda_{g,min}, \lambda_{g,max}]$ are used. Subsequently, the optimal tuning parameter is computed as described in Section 3.3. Following adaptive Lasso literature, I set $\kappa = 2$ to construct the weights $\dot{w}_{g,t}$.

4.1 Data generating process

Each method is estimated on three data generating processes (DGPs), which differ in terms of the presence of serial correlation in the independent variable and instruments, and the number of available valid instruments. The basis of the specifications resembles a combination of the Monte Carlo simulation studies of Qian & Su (2016) and Okui & Wang (2021). The basic form of each DGP is

$$y_{it} = \beta'_{g_i,t} x_{it} + \sigma_\varepsilon \varepsilon_{it}, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T. \tag{23}$$

Further specifications are

DGP 1 $x_{it} = \xi_{it} + 0.3\varepsilon_{it}$, $\varepsilon_{it} \sim$ i.i.d. $N(0,1)$, $\xi_{it} \sim$ i.i.d. $N(0,1)$, $z_{it}^{(1)} = \xi_{it} + 0.3u_{it}$, $u_{it} \sim$ i.i.d. $N(0,1)$, $z_{it}^{(2)} = \xi_{it}^3 + 0.3v_{it}$, $v_{it} \sim$ i.i.d. $N(0,1)$, ξ_{it} and ε_{it} are mutually independent, $z_{it}^{(1)}$ and $z_{it}^{(2)}$ are independent of ε_{it}.

DGP 2 Same as DGP 1, except that $\xi_{it} \sim AR(1)$ for each i : $\xi_{it} = 0.5\xi_{i,t-1} + u_{it}$, $u_{it} \sim$ i.i.d. $N(0, 0.75)$.

DGP 3 Same as DGP 1, except that the only available valid instrument is $z_{it}^{(1)}$.

DGP 1 is the benchmark case, both $z_{it}^{(1)}$ and $z_{it}^{(2)}$ are valid instruments for the endogenous regressor x_{it}. Since both of them are used in estimation, the endogenous regressor is overidentified. Furthermore, the exogenous part of the independent variable and all error terms are i.i.d. over i and t, and the idiosyncratic error process is strong white noise. DGP 2 differs from the benchmark in the sense that it displays serial correlation in the regressor and instruments. DGP 3 is the same as the benchmark case except that it has just a single instrument. Therefore, it corresponds to the exactly identified case. Regarding the instrument $z_{it}^{(2)}$, it is generated in such a way that it provides a valid instrument while preventing multicollinearity with $z_{it}^{(1)}$ so that the instruments provide valuable information that is different from each other. This entails a high correlation with the endogenous regressor while ensuring that the correlation between the instruments is not too high. Due to the exponentiation of ξ_{it}, the instruments will differ in scale. However, standardization is not necessary because 2SLS and EGMM both account for scale differences of the instruments and OLS does not

use the instruments. Furthermore, note that the independent variables and instruments are random instead of fixed and as such I expect there to be heteroskedasticity that follows the heterogeneous structure of groups and structural breaks.

In addition to three different DGPs, I evaluate performance at different noise levels. I consider $\sigma_\varepsilon \in (0.5, 0.75)$. These values correspond to signal-to-noise ratios of 4 and $\frac{16}{9}$, respectively. Moreover, to investigate the relationship between performance and data size characteristics, two cross-sectional sample sizes and three time series lengths are considered. These are $N = (50, 100)$ and $T = (10, 20, 40)$, respectively. This means that a total of 36 different model setups are considered since estimation method performance is assessed for every possible combination of three DGPs, two noise levels, two cross-sectional sample sizes, and three time series lengths. For every model setup, 1000 replications are performed for each estimation method.

In accordance with Okui & Wang (2021), I let there be three groups, and I let that be known. Let N_j ($j \in \{1, 2, 3\}$) denote the number of units in group j, as such $N = N_1 + N_2 + N_3$. Both N_1 and N_2 are a third of N, rounded to the nearest integer. N_3 accounts for the remaining part of the units. The coefficients of the three groups are

$$\beta_{1,t} = \begin{cases} 1 & \text{if } 1 \leq t < \lfloor T/2 \rfloor \\ 2 & \text{if } \lfloor T/2 \rfloor \leq t < \lfloor 5T/6 \rfloor, \\ 3 & \text{if } \lfloor 5T/6 \rfloor \leq t \leq T \end{cases} \qquad \beta_{2,t} = \begin{cases} 3 & \text{if } 1 \leq t < \lfloor T/3 \rfloor \\ 4 & \text{if } \lfloor T/3 \rfloor \leq t < \lfloor 5T/6 \rfloor, \\ 5 & \text{if } \lfloor 5T/6 \rfloor \leq t \leq T \end{cases}$$

$$\beta_{3,t} = \ 1.5 \quad \text{for } 1 \leq t \leq T,$$

where $\lfloor \cdot \rfloor$ is an operator that takes the integer part. The first two groups both display two structural breaks. However, the first break date differs among the two groups. The slope coefficient of the third group is constant over the entire observational period.

4.2 Evaluation criteria

The evaluation criteria to assess performance are all taken from Okui & Wang (2021). However, in contrast with Okui & Wang (2021), I do not evaluate the selection of the number of groups because it would lead to many more simulations and consequently prohibitive computation time. For example, suppose that I also consider five different number of groups. That

would mean four new situations for each of the 36 model setups, leading to 144 extra model setups for which I need the final parameter estimates. Since the adaption of the GAGFL method to coefficient estimation through EGMM or 2SLS leads to substantial increases in computation time already, the total computation time of 144 extra model setups becomes prohibitive.

It seems reasonable to assume that the performance of EGMM GAGFL in selecting the right number of groups does not differ much from the observed performance of GAGFL by Okui & Wang (2021) if the other parameters are accurately estimated. This is because the procedure to select G based on the BIC is the same for both methods and only parameter estimates are performed differently. Also, as mentioned before, the influence of $\hat{\sigma}^2$ is presumably very limited. For the DGP similar to my DGPs, Okui & Wang (2021) find that in the worst case ($\sigma_\varepsilon = 0.75$, $N = 50$, and $T = 10$), the correct number of groups is selected in over 97% of the cases.

The performance on the four remaining criteria is quantified through five performance indicators. The criteria are classification accuracy, break detection accuracy, breakpoint estimation accuracy, and coefficient estimation accuracy. Each criterion has one performance indicator, except for the latter which is evaluated by two distinct indicators. The respective performance indicators are the ratio of misclassified units to the total number of units, frequency of correctly estimating the number of breaks for each group, relative Hausdorff distance (HD) between break date estimates and true set of brake dates for each group, and root mean squared error (RMSE) and coverage probability of the two-sided nominal 95% confidence interval.

For the last three performance indicators, I provide some further explanation. The Hausdorff distance between any two sets A and B is defined as $\mathrm{HD}(A, B) \equiv \max\{\mathcal{D}(A, B), \mathcal{D}(B, A)\}$, where $\mathcal{D}(A, B) \equiv \sup_{b \in B} \inf_{a \in A} |a - b|$. The Hausdorff error of of an estimated set of break dates is defined by its HD to the true set of break dates (i.e., $\mathrm{HD}(\hat{\mathcal{T}}_{\hat{m}_g,g}, \mathcal{T}_{m_g,g})$, for group g). Then, the relative (with respect to T) Hausdorff error is equal to $\mathrm{HD}(\hat{\mathcal{T}}_{\hat{m}_g,g}, \mathcal{T}_{m_g,g})/T$. The RMSE of $\widehat{\beta}_{g_i,t}$ is computed by

$$\mathrm{RMSE}(\widehat{\beta}_{g_i,t}) = \sqrt{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\widehat{\beta}_{g_i,t} - \beta_{g_i,t}\right)^2},$$

and its coverage probability is

$$\text{Coverage}(\widehat{\beta}_{g_i,t}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} I\left(\widehat{\beta}_{g_i,t} - 1.96\widehat{\sigma}_{\beta_{g_i,t}} \leq \beta_{g_i,t} \leq \widehat{\beta}_{g_i,t} + 1.96\widehat{\sigma}_{\beta_{g_i,t}}\right),$$

where $\widehat{\sigma}_{\beta_{g_i,t}}$ is the estimated standard deviation of $\widehat{\beta}_{g_i,t}$. The estimates of standard deviation for OLS and 2SLS are Newey-West estimates (Newey & West, 1987) to ensure consistency under heteroskedasticity. As mentioned in Section 3.2.3, GMM standard errors are computed as in Heij et al. (2004) for EGMM. All performance indicators are averaged over 1000 simulation replications.

4.3 Results

Starting with the classification accuracy, EGMM performs well. The misclassification frequency is around 5% for the worst case and drops to around only 2% with increasing T. As can be seen from Table 1, the other two estimators perform quite well too, there is not much difference in performance. However, OLS outperforms EGMM and 2SLS, while 2SLS performs slightly worse than EGMM. OLS performs very well because the upward bias in coefficient estimation affects all coefficient estimates in the same direction. Therefore, the (absolute) differences in coefficient estimates between groups are the same as when strictly the exogenous part of x_{it} would be used in coefficient estimation. For the other two methods, the error process included in the instrument results in extra noise and thus a larger misclassification accuracy. In practice, the nature of endogeneity and heterogeneity between individuals may lead to results that contrast the better performance observed here.

Overall, misclassification frequency seems unaffected by changes in N, whereas it reduces markedly with increasing T. Furthermore, the results suggest that serial correlation in the instrument and independent variable negatively affects the classification accuracy. This may be explained as follows. In the case of serial correlation, outliers heavily influence close remaining periods in the same direction as the outlier. Consequently, the slope coefficient will be overestimated or underestimated for those periods, depending on the direction of the outlier. The over- or underestimation can coincide with the coefficient value of a different group, increasing the chance of misclassification. Without serial correlation, outliers affect

Table 1: Misclassification frequency group allocation estimates (in percentages)

DGP	σ_ε	Method	N = 50			N = 100		
			T = 10	T = 20	T = 40	T = 10	T = 20	T = 40
1	0.5	EGMM	1.16	0.22	0.00	1.04	0.22	0.01
		2SLS	1.18	0.24	0.00	1.05	0.24	0.00
		OLS	0.91	0.15	0.00	0.87	0.15	0.00
	0.75	EGMM	4.64	1.86	0.25	4.12	1.72	0.24
		2SLS	5.75	1.96	0.26	4.27	1.82	0.27
		OLS	3.49	1.29	0.12	3.40	1.36	0.14
2	0.5	EGMM	1.38	0.33	0.00	1.40	0.30	0.01
		2SLS	1.48	0.42	0.01	1.46	0.33	0.01
		OLS	1.07	0.27	0.00	1.20	0.23	0.01
	0.75	EGMM	5.42	2.48	0.39	4.79	2.22	0.35
		2SLS	6.13	2.48	0.41	5.04	2.29	0.40
		OLS	4.16	1.71	0.23	3.97	1.77	0.23
3	0.5	IV	1.09	0.22	0.01	1.12	0.22	0.00
		OLS	0.87	0.19	0.01	0.90	0.15	0.00
	0.75	IV	4.87	1.80	0.25	4.28	1.82	0.26
		OLS	3.44	1.31	0.13	3.54	1.37	0.12

Note: *For DGP 3, the EGMM and 2SLS estimators are equivalent and equal to a simple IV estimator because it represent the exactly identified case.*

just a single period.

Moving on to break detection accuracy, the results suggest that EGMM performs satisfactorily because it accounts for heteroskedasticity. In the worst case, as can be seen from Table 2, the frequency of correctly estimating the number of breaks is 70%. This occurs for the combination of large error size ($\sigma_\varepsilon = 0.75$), smaller cross-sectional sample size ($N = 50$), and overidentification of the endogenous regressor (DGP 1 and DGP 2). However, only changing the error size to be moderate ($\sigma_\varepsilon = 0.50$) increases performance to around 94% or more, while increasing N from 50 to 100 gives an accuracy of around 85% or higher. Again, OLS outperforms the other two methods with accuracy measures of 85% or higher for all cases. 2SLS performs very poorly. Its accuracy measures for groups with breaks are 90% at best and only 20% at worst. As previously mentioned, this can be attributed to the failure of 2SLS to account for heteroskedasticity, for which I found the following evidence.

As can be observed from Table 2, the number of breaks in group 3 (no breaks) is quite well

estimated, so 2SLS generally does not seem to falsely detect breaks when there are none. For the other groups (which have two breaks each), 2SLS tends to falsely detect an extra break. This was concluded after a closer inspection of the results on the number of breaks estimated, which revealed an overall overestimation of the number of breaks for 2SLS. The main difference between 2SLS and EGMM is that the former does not account for heteroskedasticity. The results thus suggest that there is some form of heteroskedasticity induced by the structural breaks which increases the chance of false break detection. The mechanism behind this can be explained as an increase in the variance of the error term resulting from an increase in the slope coefficient at a breakpoint. This works through the idiosyncratic variance of the independent variable. An increase in the error variance will lead to bigger deviations from the expected value for the dependent variable and hence a false break detection will become more likely. Further evidence for this is the lower accuracy of 2SLS for group 2 than for group 1. The main difference between these groups is that the coefficients of group 2 are higher than those of group 1. EGMM accounts for heteroskedasticity by having the influence of observations inversely proportional to their variance.

In line with the better performance of OLS observed for classification accuracy, OLS performs comparatively well again for break detection accuracy. Similar reasoning applies, focusing on the difference between regimes now. Supporting evidence is that the OLS results shown in Table 2 display no notable difference with the corresponding results found by Okui & Wang (2021).

All methods' break detection accuracy improves substantially with increasing N. For OLS this is also the case for increasing T, whereas the performance of 2SLS decreases with increasing T. Regarding EGMM, its accuracy generally improves with increasing T. Cases in which a decrease in performance is observed, associated with an increase in T, may be explained by the slight chance for a false break detection for every additional period since the number of breaks is allowed to grow unrestricted (Qian & Su, 2016).

Finally, in the exactly identified case (DGP 3), the IV estimators' accuracies become high (83% at worst and over 88% elsewhere) and very similar to those of OLS, albeit slightly worse. This difference may be attributed to the error term in the instrument, as the expected value of the (absolute) difference between regimes is the same for both methods.

Table 2: Frequency of correctly estimating the number of breaks (in percentages)

σ_ε	Method	Group	N = 50			N = 100		
			T = 10	T = 20	T = 40	T = 10	T = 20	T = 40
			DGP 1					
0.5	EGMM	G1$\left(m_1^0=2\right)$	94.5	96.8	97.8	98.3	99.5	100
		G2$\left(m_2^0=2\right)$	93.8	96.5	96.9	98.6	99.4	99.9
		G3$\left(m_3^0=0\right)$	94.4	97.0	98.9	99.0	99.6	99.9
	2SLS	G1$\left(m_1^0=2\right)$	74.4	67.1	53.6	88.8	85.6	79.1
		G2$\left(m_2^0=2\right)$	42.0	29.5	21.9	59.7	48.2	35.9
		G3$\left(m_3^0=0\right)$	99.2	100	100	100	100	100
	OLS	G1$\left(m_1^0=2\right)$	99.7	100	99.9	100	100	100
		G2$\left(m_2^0=2\right)$	99.6	100	100	100	100	100
		G3$\left(m_3^0=0\right)$	99.8	100	100	100	100	100
0.75	EGMM	G1$\left(m_1^0=2\right)$	70.6	77.9	79.9	86.0	92.1	97.2
		G2$\left(m_2^0=2\right)$	69.5	96.2	75.1	89.0	92.7	96.2
		G3$\left(m_3^0=0\right)$	76.3	98.2	93.6	88.0	95.1	98.8
	2SLS	G1$\left(m_1^0=2\right)$	59.7	54.4	42.8	78.0	77.8	65.1
		G2$\left(m_2^0=2\right)$	40.6	29.8	19.7	54.0	40.8	30.6
		G3$\left(m_3^0=0\right)$	80.9	95.7	99.9	94.5	99.7	100
	OLS	G1$\left(m_1^0=2\right)$	85.8	94.6	97.3	96.8	99.9	99.0
		G2$\left(m_2^0=2\right)$	91.9	96.2	95.6	97.7	99.6	100
		G3$\left(m_3^0=0\right)$	87.0	98.2	99.9	98.1	99.9	100
			DGP 2					
0.5	EGMM	G1$\left(m_1^0=2\right)$	93.1	95.9	98.0	98.3	99.5	99.7
		G2$\left(m_2^0=2\right)$	92.7	95.9	95.8	98.4	99.8	99.8
		G3$\left(m_3^0=0\right)$	95.0	97.5	98.6	98.5	99.7	99.9
	2SLS	G1$\left(m_1^0=2\right)$	75.9	68.0	55.2	89.6	85.0	80.0
		G2$\left(m_2^0=2\right)$	44.1	35.0	23.4	57.4	49.3	36.5
		G3$\left(m_3^0=0\right)$	99.5	99.9	100	99.9	100	100
	OLS	G1$\left(m_1^0=2\right)$	99.8	99.8	99.0	100	100	100
		G2$\left(m_2^0=2\right)$	99.4	100	100	100	100	100
		G3$\left(m_3^0=0\right)$	99.8	100	100	100	100	100
0.75	EGMM	G1$\left(m_1^0=2\right)$	69.1	78.6	78.1	87.0	93.2	95.8
		G2$\left(m_2^0=2\right)$	69.2	76.7	74.9	84.3	92.5	96.9
		G3$\left(m_3^0=0\right)$	76.2	88.6	95.4	89.7	95.5	98.7
	2SLS	G1$\left(m_1^0=2\right)$	58.8	55.7	41.5	77.9	76.4	70.2
		G2$\left(m_2^0=2\right)$	40.5	28.7	20.4	53.3	44.2	33.3
		G3$\left(m_3^0=0\right)$	77.4	94.4	99.7	93.6	99.4	100
	OLS	G1$\left(m_1^0=2\right)$	86.6	95.8	94.8	97.0	99.7	100
		G2$\left(m_2^0=2\right)$	90.8	97.0	96.7	97.9	100	99.9
		G3$\left(m_3^0=0\right)$	85.6	97.3	99.9	96.9	99.9	100
			DGP 3					
0.5	IV	G1$\left(m_1^0=2\right)$	99.1	99.0	98.8	100	99.9	100
		G2$\left(m_2^0=2\right)$	98.8	99.3	98.6	99.9	100	100
		G3$\left(m_3^0=0\right)$	99.4	100	100	100	100	100
	OLS	G1$\left(m_1^0=2\right)$	99.7	99.9	100	100	100	100
		G2$\left(m_2^0=2\right)$	99.6	100	99.9	100	100	100
		G3$\left(m_3^0=0\right)$	99.5	100	100	100	100	100
0.75	IV	G1$\left(m_1^0=2\right)$	82.9	88.7	88.0	95.6	98.9	99.6
		G2$\left(m_2^0=2\right)$	84.0	90.0	86.9	96.9	98.1	98.9
		G3$\left(m_3^0=0\right)$	84.6	96.9	99.6	96.6	99.8	100
	OLS	G1$\left(m_1^0=2\right)$	85.6	94.4	95.9	97.5	99.9	99.9
		G2$\left(m_2^0=2\right)$	89.2	96.0	94.9	99.2	99.8	99.9
		G3$\left(m_3^0=0\right)$	86.0	98.3	99.8	97.8	99.8	100

Note: *For DGP 3, the EGMM and 2SLS estimators are equivalent and equal to a simple IV estimator because it represent the exactly identified case. The Group column shows for each group the true number of breaks m_g^0 in parentheses.*

The third evaluation criterion is breakpoint estimation accuracy, for which EGMM scores well with a relative HD of around 5% in the worst case ($\sigma_\varepsilon = 0.75$, $N = 50$, and $T = 10$). The Hausdorff errors in Table 3 are reported only for cases in which the correct number of breaks has been estimated. Substantial improvements are observed with increasing N for all methods. Much of the results are very similar to those found for break detection accuracy. OLS performs best, whereas 2SLS performs markedly worse than the other methods (error for group 2 is always $> 6\%$). The performance of IV is really good again. I observe once more no notable impact of serial correlation. Increasing T generally has a positive effect on performance, but incidentally not for EGMM. For 2SLS it is mostly a negative effect. This is somewhat surprising because the Hausdorff errors are given relative to T, in contrast with the frequency of correctly estimating the number of breaks.

Finally, I consider coefficient estimation accuracy. Here, the benefit of using EGMM instead of OLS, in the presence of endogenous regressors, becomes apparent. The EGMM estimates are relatively close to the true coefficient values, as observed from its low RMSE in Table 4. Conversely, the low coverage probabilities suggest poor performance of EGMM. However, further inspection revealed that can likely be attributed to a combination of a very low estimated variance and a slight bias, which is not that problematic generally.

As is well known, $RMSE = \sqrt{Variance + bias^2}$. The RMSE of EGMM and 2SLS are very similar, but the coverage probability of 2SLS is substantially higher. This suggests that the portion of bias in the RMSE is higher for EGMM than for 2SLS. Moreover, the coverage probability of EGMM decreases with increasing T, accompanied by a simultaneous substantial decrease in RMSE. Furthermore, I looked into the unreported standard error estimates of EGMM and 2SLS, which showed that those of 2SLS are markedly greater. A reason for this is that 2SLS often falsely detects breaks, presumably as a consequence of not accounting for heteroskedasticity. This leads to overfitting on the data through more flexibility in coefficient estimation, which leads to a bigger variance. Concluding, the above suggests that EGMM coefficient estimates have a very low estimated variance and a slight bias, giving rise to poor coverage probabilities.

As aforementioned, the coefficient estimation accuracy motivates the use of EGMM instead of OLS when dealing with endogenous regressors. OLS performs very poorly on both RMSE

Table 3: Ratio of Hausdorff distance to time for break date estimates (in percentages)

σ_ε	**Method**	**Group**	**N = 50**			**N = 100**		
			T = 10	**T = 20**	**T = 40**	**T = 10**	**T = 20**	**T = 40**
			DGP 1					
0.5	EGMM	G1$\left(m_1^0=2\right)$	0.87	0.46	0.26	0.21	0.13	0.00
		G2$\left(m_2^0=2\right)$	0.94	0.50	0.48	0.22	0.11	0.01
	2SLS	G1$\left(m_1^0=2\right)$	3.68	3.81	5.00	1.39	1.33	1.78
		G2$\left(m_2^0=2\right)$	10.2	11.1	12.1	6.34	7.79	8.94
	OLS	G1$\left(m_1^0=2\right)$	0.08	0.01	0.01	0.00	0.00	0.00
		G2$\left(m_2^0=2\right)$	0.06	0.01	0.01	0.00	0.00	0.00
0.75	EGMM	G1$\left(m_1^0=2\right)$	4.97	3.56	2.92	2.12	1.30	0.46
		G2$\left(m_2^0=2\right)$	4.84	3.61	3.31	1.56	0.94	0.59
	2SLS	G1$\left(m_1^0=2\right)$	6.56	6.50	7.38	2.98	2.51	3.36
		G2$\left(m_2^0=2\right)$	10.9	12.2	12.4	7.77	8.98	10.1
	OLS	G1$\left(m_1^0=2\right)$	2.36	0.81	0.36	0.46	0.02	0.01
		G2$\left(m_2^0=2\right)$	1.39	0.62	0.58	0.32	0.06	0.01
			DGP 2					
0.5	EGMM	G1$\left(m_1^0=2\right)$	1.09	0.57	0.24	0.28	0.04	0.04
		G2$\left(m_2^0=2\right)$	1.03	0.53	0.51	0.19	0.02	0.02
	2SLS	G1$\left(m_1^0=2\right)$	3.72	3.97	4.70	1.26	1.34	1.57
		G2$\left(m_2^0=2\right)$	9.86	10.5	11.7	7.03	7.50	9.16
	OLS	G1$\left(m_1^0=2\right)$	0.04	0.02	0.01	0.00	0.00	0.00
		G2$\left(m_2^0=2\right)$	0.10	0.01	0.01	0.00	0.00	0.00
0.75	EGMM	G1$\left(m_1^0=2\right)$	5.07	3.52	2.96	2.09	1.16	0.47
		G2$\left(m_2^0=2\right)$	4.53	3.50	3.36	2.19	0.96	0.39
	2SLS	G1$\left(m_1^0=2\right)$	7.00	6.19	7.25	3.27	2.82	2.97
		G2$\left(m_2^0=2\right)$	11.1	12.0	12.6	8.15	8.39	9.71
	OLS	G1$\left(m_1^0=2\right)$	2.38	0.67	0.64	0.54	0.09	0.01
		G2$\left(m_2^0=2\right)$	1.54	0.54	0.46	0.30	0.00	0.00
			DGP 3					
0.5	IV	G1$\left(m_1^0=2\right)$	0.21	0.12	0.13	0.00	0.01	0.00
		G2$\left(m_2^0=2\right)$	0.23	0.11	0.24	0.01	0.00	0.00
	OLS	G1$\left(m_1^0=2\right)$	0.09	0.01	0.00	0.00	0.00	0.00
		G2$\left(m_2^0=2\right)$	0.04	0.01	0.02	0.00	0.00	0.00
0.75	IV	G1$\left(m_1^0=2\right)$	3.17	1.63	1.31	0.74	0.10	0.04
		G2$\left(m_2^0=2\right)$	2.66	1.55	1.81	0.41	0.15	0.11
	OLS	G1$\left(m_1^0=2\right)$	2.60	0.90	0.51	0.43	0.03	0.01
		G2$\left(m_2^0=2\right)$	1.63	0.62	0.70	0.13	0.01	0.02

Note: *For DGP 3, the EGMM and 2SLS estimators are equivalent and equal to a simple IV estimator because it represent the exactly identified case. The Group column shows for each group the true number of breaks m_g^0 in parentheses. No results are reported for group 3 because all Hausdorff errors are naturally zero.*

and coverage probability, due to the high bias resulting from the endogeneity of the regressor. Looking at the results for the combinations of moderate error size ($\sigma_\varepsilon = 0.50$), and high T ($T = 20$ and $T = 40$), OLS seems to converge asymptotically to a squared bias value which is relatively large. By contrast, the RMSEs of 2SLS and EGMM appear to be converging to zero, or a value close to zero. Moreover, the dominance of a large bias squared term shows from the coverage probability of OLS decreasing substantially with both N and T, becoming practically zero for $N = 100$ and $T = 40$.

Except for the coverage probabilities of OLS and EGMM, substantial performance improvements are observed with increasing T, which can mainly be attributed to a decrease in variance. Performance improvements with increasing N are less substantial but occur more uniformly across the model setups and estimation methods. Only for the coverage probability of OLS does the performance decrease with increasing N, due to the large bias involved. The increase in coverage probability of EGMM with increasing N is opposite to the effect of increasing T.

In the exactly identified case (DGP 3), the IV estimator performs especially well. The coverage probability is 85% in the worst case ($\sigma_\varepsilon = 0.75$, $N = 50$, and $T = 10$), and rises greatly to $> 90\%$ when increasing T. Furthermore, the RMSEs are substantially lower than those of EGMM and 2SLS (DGP 1 and DGP 2).

5 Conclusion

I propose a new estimation method that allows for heterogeneous structural changes in the slope coefficients of panel data models in the presence of endogenous independent variables. Individual heterogeneity is modeled through a latent group structure. The time pattern of the coefficients is characterized by structural breaks in the slope coefficients that may vary in size and timing between groups. My EGMM GAGFL method is an extension of the GAGFL method. It uses EGMM instead of OLS to estimate coefficients, to correct for the endogeneity bias.

Monte Carlo simulation results show that EGMM GAGFL generally performs well in finite samples, despite that a slight bias in coefficient estimation is possibly present. This is

Table 4: Root mean squared error and coverage probability of coefficient estimates

DGP	σ	N	T	RMSE			Coverage		
				EGMM	2SLS	OLS	EGMM	2SLS	OLS
1	0.5	50	10	0.1455	0.1518	0.1733	0.6622	0.8810	0.2444
		50	20	0.0952	0.1114	0.1465	0.6284	0.8991	0.0746
		50	40	0.0737	0.0901	0.1410	0.5633	0.9149	0.0108
		100	10	0.1142	0.1164	0.1667	0.7506	0.9108	0.0842
		100	20	0.0644	0.0740	0.1439	0.7267	0.9250	0.0127
		100	40	0.0411	0.0548	0.1391	0.6795	0.9288	0.0004
	0.75	50	10	0.2615	0.2627	0.2920	0.6276	0.8168	0.2585
		50	20	0.1750	0.1759	0.2364	0.6227	0.8619	0.0791
		50	40	0.1259	0.1198	0.2142	0.5563	0.8968	0.0125
		100	10	0.2186	0.2206	0.2759	0.7188	0.8556	0.0881
		100	20	0.1327	0.1370	0.2307	0.7084	0.8986	0.0118
		100	40	0.0742	0.0859	0.2108	0.6629	0.9214	0.0004
2	0.5	50	10	0.1523	0.1599	0.1825	0.6835	0.8794	0.2321
		50	20	0.0995	0.1166	0.1526	0.6340	0.8945	0.0725
		50	40	0.0728	0.0893	0.1421	0.5680	0.9155	0.0104
		100	10	0.1299	0.1336	0.1805	0.7310	0.9030	0.0748
		100	20	0.0687	0.0752	0.1487	0.7140	0.9188	0.0129
		100	40	0.0421	0.0555	0.1408	0.6818	0.9328	0.0003
	0.75	50	10	0.2810	0.2843	0.3082	0.6257	0.7984	0.2441
		50	20	0.1869	0.1848	0.2457	0.6113	0.8559	0.0810
		50	40	0.1277	0.1245	0.2180	0.5546	0.8940	0.0146
		100	10	0.2411	0.2450	0.2948	0.7087	0.8449	0.0775
		100	20	0.1451	0.1496	0.2412	0.6963	0.8847	0.0123
		100	40	0.0792	0.0883	0.2143	0.6678	0.9188	0.0004
3	0.5	50	10	0.1172		0.1718	0.9239		0.2461
		50	20	0.0609		0.1470	0.9392		0.0791
		50	40	0.0365		0.1410	0.9431		0.0111
		100	10	0.1076		0.1679	0.9296		0.0833
		100	20	0.0490		0.1435	0.9361		0.0125
		100	40	0.0220		0.1391	0.9521		0.0004
	0.75	50	10	0.2311		0.2895	0.8568		0.2546
		50	20	0.1356		0.2366	0.9128		0.0847
		50	40	0.0736		0.2146	0.9316		0.0134
		100	10	0.2087		0.2789	0.8738		0.0848
		100	20	0.1181		0.2305	0.9205		0.0126
		100	40	0.0533		0.2104	0.9438		0.0005

Note: *For DGP 3, the EGMM and 2SLS estimators are equivalent and equal to a simple IV estimator because it represent the exactly identified case.*

indicated by the fact that the coverage probability of the two-sided nominal 95% confidence interval of the coefficient estimate declines when the time series grows longer. Conversely, the root mean squared error of the coefficient estimate moves substantially towards zero with both a larger number of cross-sectional observations and a longer time series, which suggests accurate estimation. EGMM GAGFL seems to outperform GAGFL and 2SLS GAGFL. The better performance of GAGFL (which employs OLS coefficient estimation) for classification accuracy and break estimation accuracy can be attributed to the specification of endogeneity and heterogeneity in my Monte Carlo simulation. In practice, this advantage might not be there. By contrast, the poorer performance on coefficient estimation accuracy will exist in practice too. Moreover, comparing break estimation accuracy between EGMM GAGFL and 2SLS GAGFL indicates that it is important to take heteroskedasticity into account in the case of random regressors. Finally, the Monte Carlo results suggest that latent group membership, the number of breaks, and the breakpoints are consistently estimated by EGMM GAGFL.

For future work, the method can likely be improved by using regime-dependent weighting matrices in post-Lasso estimation instead of fully time-variant matrices, because the errors may be assumed homoskedastic within a regime. Furthermore, the method may be extended to allow for a dynamic model component and individual-specific fixed effect. This requires first differenced data and appropriate instruments for the first-order lag of the dependent variable.

References

Anderson, T. W., & Hsiao, C. (1981). Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association*, *76*(375), 598–606. doi: 10.1080/01621459.1981.10477691

Ando, T., & Bai, J. (2016). Panel Data Models with Grouped Factor Structure Under Unknown Group Membership. *Journal of Applied Econometrics*, *31*(1), 163–191. doi: 10.1002/jae.2467

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*(1), 29–51. doi: 10.1016/0304-4076(94)01642-D

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, *157*(1), 78–92. doi: 10.1016/j.jeconom.2009.10.020

Baltagi, B. H., Kao, C., & Liu, L. (2017). Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term. *Econometric Reviews*, *36*(1-3), 85–102. doi: 10.1080/07474938.2015.1114262

Basmann, R. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica*, *25*(1), 77–83. doi: 10.2307/1907743

Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *The Stata Journal*, *3*(1), 1–31. doi: 10.1177/1536867X0300300101

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115–143. doi: 10.1016/S0304-4076(98)00009-8

Bonhomme, S., & Manresa, E. (2015). Grouped Patterns of Heterogeneity in Panel Data. *Econometrica*, *83*(3), 1147–1184. doi: 10.3982/ECTA11319

Calderón, C., Moral-Benito, E., & Servén, L. (2015). Is infrastructure capital productive? A dynamic heterogeneous approach. *Journal of Applied Econometrics*, *30*(2), 177–198. doi: 10.1002/jae.2373

De Wachter, S., & Tzavalis, E. (2012). Detection of structural breaks in linear dynamic panel data models. *Computational Statistics & Data Analysis*, *56*(11), 3020–3034. doi: 10.1016/j.csda.2012.02.025

Hahn, J., & Moon, H. R. (2010). Panel data models with finite number of multiple equilibria. *Econometric Theory*, *26*(3), 863–881. doi: 10.1017/S0266466609990132

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*(4), 1029–1054. doi: 10.2307/1912775

Heij, C., de Boer, P., Franses, P. H., Kloek, T., & van Dijk, H. K. (2004). *Econometric methods with applications in business and economics*. Oxford University Press.

Kim, D. (2011). Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics*, *164*(2), 310–330. doi: 10.1016/j.jeconom.2011.06.018

Li, D., Qian, J., & Su, L. (2016). Panel Data Models With Interactive Fixed Effects and Multiple Structural Breaks. *Journal of the American Statistical Association*, *111*(516), 1804–1819. doi: 10.1080/01621459.2015.1119696

Lin, C.-C., & Ng, S. (2012). Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown. *Journal of Econometric Methods*, *1*(1), 42–55. doi: 10.1515/2156-6674.1000

Miao, K., Su, L., & Wang, W. (2020). Panel threshold regressions with latent group structures. *Journal of Econometrics*, *214*(2), 451–481. doi: 10.1016/j.jeconom.2019.07.006

Neal, B. T. (2015). *Estimating Heterogeneous Coefficients in Panel Data Models with Endogenous Regressors and Common Factors.*

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708. doi: 10.2307/1913610

Okui, R., & Wang, W. (2021). Heterogeneous structural breaks in panel data models. *Journal of Econometrics*, *220*(2), 447–473. doi: 10.1016/j.jeconom.2020.04.009

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, *74*(4), 967–1012. doi: 10.1111/j.1468-0262.2006.00692.x

Pesaran, M. H., Shin, Y., & Smith, R. P. (1999). Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association*, *94*(446), 621–634. doi: 10.1080/01621459.1999.10474156

Pesaran, M. H., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, *68*(1), 79–113. doi: 10.1016/0304-4076(94)01644-F

Qian, J., & Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused Lasso. *Journal of Econometrics*, *191*(1), 86–109. doi: 10.1016/j.jeconom.2015.09.004

Sargan, J. D. (1958). The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, *26*(3), 393–415. doi: 10.2307/1907619

Su, L., Shi, Z., & Phillips, P. C. B. (2016). Identifying Latent Structures in Panel Data. *Econometrica*, *84*(6), 2215–2264. doi: 10.3982/ecta12560

Su, L., Wang, X., & Jin, S. (2019). Sieve Estimation of Time-Varying Panel Data Models With Latent Structures. *Journal of Business & Economic Statistics*, *37*(2), 334–349. doi: 10.1080/07350015.2017.1340299

Sun, Y. (2005). Estimation and Inference in Panel Structure Models. *SSRN Electronic Journal*. doi: 10.2139/SSRN.794884

Swamy, P. A. V. B. (1970). Efficient Inference in a Random Coefficient Regression Model. *Econometrica*, *38*(2), 311–323. doi: 10.2307/1913012

Theil, H. (1953). Repeated least squares applied to complete equation systems. *The Hague: central planning bureau*.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(1), 91–108. doi: 10.1111/j.1467-9868.2005.00490.x

Wang, W., Phillips, P. C., & Su, L. (2018). Homogeneity pursuit in panel data models: Theory and application. *Journal of Applied Econometrics*, *33*(6), 797–815. doi: 10.1002/jae.2632

Wang, W., & Su, L. (2021). Identifying latent group structures in nonlinear panels. *Journal of Econometrics*, *220*(2), 272–295. doi: 10.1016/j.jeconom.2020.04.003