Section I

연구논문

Grouping and clustering methods in econometrics[*]

Ryo Okui[**]

This paper reviews recent developments in the application of clustering methods in econometrics. In particular, we discuss how the *k-means* algorithm can be extended and applied to econometric problems. Models with group structures are useful to describe heterogeneity across units. Such models can be estimated by extensions of the *k-means* algorithm. We also discuss inference methods for group memberships and methods to incorporate possible structural breaks.

Keywords: Grouping method, clustering method, panel data, heterogeneity, structural break.

JEL classification: C23, C38, C51

1. Introduction

This paper reviews recent developments in econometric applications of clustering methods. Clustering methods are data-driven methods to divide a sample into several groups. They do not assume prior knowledge of group memberships of the observations. Although clustering methods have a long history in statistics and machine learning, econometricians have only recently shown particular interest in these methods to analyze heterogeneity across observational units.

In economic applications, heterogeneity across observational units needs to be

---

Received: Dec. 1, 2020. Accepted: Dec. 5, 2020.

* The author gratefully acknowledges the financial support of the School of Social Sciences and a New Faculty Startup Grant at Seoul National University and from the Housing and Commercial Bank Economic Research Fund in the Institute of Economic Research at Seoul National University. The usual disclaimer applies.

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

adequately treated to have valid statistical inferences, and heterogeneity itself may also be of research interest. For example, consider an empirical analysis of the effect of democracy on economic growth using country-level data (Acemoglu et al. 2008). Unobserved heterogeneity across countries, such as their cultural differences and historical background, would affect both their political and economic situations. It would thus cause omitted variable bias when we measure the effect of democracy on economic growth. A consistent estimation of the effect of democracy needs to control for such heterogeneity. Moreover, there might also be heterogeneity in the effect of democracy; for some countries, a democratic government helps economic development but for other countries, an authoritarian regime may be better for economic growth.

Panel data are useful in handling heterogeneity across units. It enables us to control time-invariant unobserved heterogeneity through the introduction of unit fixed effects. It can also control for unobserved time effects that are common across units but evolve over time. Nonetheless, traditional modeling strategies for panel data still face limitations in terms of handling unobserved heterogeneity. First, introducing unit and time fixed effects does not solve the omitted variable bias problem completely because there might be time-varying unobserved heterogeneity that is correlated with the regressor and affects the outcome. Second, while unobserved heterogeneity in the effect of a variable of interest can be modeled using unit-specific coefficients, this approach causes the number of parameters to be very large; estimates may not behave well and are difficult to interpret. There are several ways to summarize unit-specific coefficients (Fernández-Val and Lee 2013; Okui and Yanagi 2019, 2020), but they are not particularly useful in investigating the effect for each unit.

The latent group structure is a useful modeling device to address these issues. It assumes that units are divided into groups. Units in the same group share the values of coefficients and two units in different groups have different values of coefficients. It is a restricted form of heterogeneity but it provides several advantages. First, group-specific coefficients can be time-varying. Second, it can control for time-varying unobserved heterogeneity, which is not feasible in approaches with individual and time effects. Third, group-specific coefficients can be estimated precisely because the estimation utilizes observations in the same group. Because group memberships can be estimated consistently, one can infer the coefficients for specific units.

This review starts with an overview of the k-means algorithm in Section 2. It is the most representative algorithm for clustering in statistics and machine learning. It is designed to divide a sample of vectors into several groups and members of each group have a common mean vector. We then extend it to panel data models in Section 3. We consider panel data models with group-specific coefficients. An extension of the k-means method to panel data models by Bonhomme and Manresa (2015) is introduced in Section 3.1. It provides an estimator of both group memberships and group-specific coefficients. Moreover, Section 3.2 reviews methods for statistical inferences for group membership structure developed by Dzemski and Okui (2020a).

We then discuss methods to detect structural breaks in Section 4. While time-varying group coefficients are allowed even in the method discussed in Section 3, it is useful to assume that group coefficients change only at breakpoints. We can have a more efficient estimator by combining information from time periods within each regime. We discuss the method developed by Okui and Wang (2020) in Section 4.1. We also consider a structural break that affects the group membership structure in Section 4.2. Lumsdaine, Okui, and Wang (2020) show that a least-squares method can be used to estimate the breakpoint, the group membership structure, and the group-specific coefficients. We then conclude in Section 5.

Models with latent group structure are clearly related to finite mixture models which have been studied intensively in econometrics, statistics, and machine learning. We do not discuss them in this review but focus on applications and extensions of the *k-means* algorithm in recent econometric literature. Compared with finite mixture models and their estimation through the EM algorithm (Dempster, Laird, and Rubin 1977), the grouping approach discussed in this review has several advantages. First, it does not require parametric assumptions on the error terms. Second, our approach provides a relatively straightforward inference method for group-specific coefficients. A disadvantage is that we typically need to assume that $T \to \infty$ to justify the procedures. Thus, the grouping methods are proven to be useful in long panels, but their usefulness in short panels is still to be seen.

While the word "cluster" is also used in the context of clustered standard errors (Angrist and Pischke 2008Chapter 8), this paper does not discuss it. It refers to the problem of constructing valid standard errors when observations are correlated within a cluster. In that context, we typically know which observations belong to which cluster and the problem we consider here is different.

Given the space limitation, this review article does not discuss the details of the asymptotic theory behind the procedures. We do not discuss how to choose the number of groups. It is typically chosen by information criteria and is discussed in papers on which the discussion for each section is based.

2. The *k-means* algorithm

We first discuss the *k-means* algorithm. It divides observations into several groups. This clustering algorithm is discussed in many statistical and computer science textbooks (Imai 2018; Murphy 2012). It is one of the most representative algorithms for unsupervised learning (that is, a problem without an outcome variable) in machine learning.

Suppose that we have a sample $\{Y_i\}_{i=1}^N$, where $Y_i = (y_{1i},\ldots,y_{mi})'$ is an m-dimensional vector. Our purpose is to cluster $\{Y_i\}_{i=1}^N$ into G groups. We assume that

$$Y_i = \mu_{g_i} + \epsilon_i,$$

where $\mu_g = (\mu_{1g},\ldots,\mu_{mg})'$ is an m-dimensional vector of parameters for group g and is called the centroid for group g, g_i is the group membership for unit i, and ϵ_i is an m-dimensional vector of error terms.

We estimate μ_g, for $g = 1,\ldots,G$ and g_i for $i = 1,\ldots,N$ simultaneously. The estimation minimizes the least-squares objective function:

$$(\{\hat{\mu}_g\}_{g=1}^G, \{\hat{g}_i\}_{i=1}^N) = \underset{(\{\mu_g\}_{g=1}^G, \{g_i\}_{i=1}^N)}{\arg\min} \sum_{i=1}^N \sum_{j=1}^m (y_{ji} - \mu_{jg_i})^2.$$

The minimization problem is usually solved by the following *k-means* algorithm.

1. Set an initial value of the centroids: $\mu_g^{(0)}$, for $g = 1,\ldots,G$. Set $s = 1$.
2. Let $g_i^{(s)} = \min_{g=1,\ldots,G} \sum_{j=1}^m (x_{ji} - \mu_{jg}^{(s-1)})^2$.
3. Let $\mu_g^{(s)} = \min_{\mu_g} \sum_{g_i^{(s)}=g} \sum_{j=1}^m (x_{ji} - \mu_{jg})^2$. Note that $\mu_{jg}^{(s)} = \sum_{g_i^{(s)}=g} x_{ji} / \#\{g_i^{(s)} = \mathrm{g}\}$.
4. Set $s = s + 1$.
5. Repeat 2 to 4 until convergence.

The algorithm converges for each initial value because the objective function decreases in each iteration. However, there is no guarantee that it achieves the global minimum and often it converges to a local minimum. We thus need to try many initial values and it might involve a large computational cost in particular when the number of groups (G) is large. Nonetheless, the algorithm seems to work well when G is not large (say, G is less than 10).

The asymptotic properties of the *k-means* algorithm have been discussed in the statistics literature. Pollard (1981) proves the probability limit of the estimator. The asymptotic framework considered is $N \to \infty$. Note that the distribution of the data does not need to be a finite mixture of some underlying distributions. Therefore, the probability limit may be considered as a pseudo true value that minimizes the probability limit of the *k-means* objective function. The asymptotic normality around the pseudo true value is discussed in Pollard (1982).

3. Group heterogeneity in panel data

We now discuss the application of the *k-means* algorithm to panel data models. In particular, we consider linear panel data models with group-specific coefficients. It is estimated by the grouped fixed effects estimator by Bonhomme and Manresa (2015). We also discuss how to make inferences about group membership structure following

Dzemski and Okui (2020a).

3.1. Grouped fixed effects estimator

Suppose that we have panel data (y_{it}, x_{it}) for $i = 1,\ldots,N$ and $t = 1,\ldots,T$. i is the unit indicator and t indicates time. The panel data contain information about N units for T time periods. y_{it} is the scalar dependent variable for unit i at time t, and x_{it} is a vector of regressors. We consider the effect of x_{it} on y_{it}.

Units in the panel data are divided into G groups. Units in the same group share the values of coefficients and their values of the coefficients are different from those in different groups. We consider the following model.

$$y_{it} = x_{it}'\,\beta_{g_i,t} + \epsilon_{it} \tag{1}$$

where $g_i \in \{1,\ldots,G\}$ is the group membership of unit i, $\beta_{g_i,t}$ is the group-specific and possibly time-varying coefficient vector, and ϵ_{it} is the error term. We assume that $E(\epsilon_{it} \mid x_{it}) = 0$ so that the coefficient vector $\beta_{g_i,t}$ captures the effect of a change in x_{it} on y_{it}.

The main part of Bonhomme and Manresa (2015) considers a version of the model in which the intercept exhibits a group structure but other coefficients are homogeneous:

$$y_{it} = w_{it}'\theta + \alpha_{g_i,t} + \epsilon_{it}, \tag{2}$$

where the regressors $(w_{it}',1) = x_{it}$ include the constant. The coefficients of nonconstant regressors (θ) are common across units and time-invariant while the intercept exhibits grouped heterogeneity and is time-varying. In this model, $\alpha_{g,t}$ captures time-varying unobserved heterogeneity whose omission would cause an omitted variable bias in the estimation of θ. Note that the *k-means* model considered in Section 2 is a further special case of (2). It is obtained by setting w_{it} null and t is interpreted as m in the previous section.

We estimate model (1) by minimizing the following objective function:

$$(\{\{\hat{\beta}_{g,t}\}_{t=1}^{T}\}_{g=1}^{G},\{\hat{g}_i\}_{i=1}^{N}) = \underset{(\{\{\beta_{g,t}\}_{t=1}^{T}\}_{g=1}^{G},\{g_i\}_{i=1}^{N})}{\arg\min} \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}{}'\beta_{g,t})^2. \tag{3}$$

This estimator is called the group fixed effects (GFE) estimator.

To compute the estimator, we adopt the following *k-means*-type algorithm.

1. Set an initial value of the coefficient: $\beta_{g,t}^{(0)}$, for $t = 1,\ldots,T$ and $g = 1,\ldots,G$. Set $s = 1$.
2. Let $g_i^{(s)} = \min_{g=1,\ldots,G} \sum_{t=1}^{T}(y_{it} - x_{it}{}'\hat{\beta}_{g,t}^{(s-1)})^2$.
3. Let $\{\hat{\beta}_{g,t}^{(s)}\}_{t=1}^{T} = \min_{\{\beta_{g,t}\}_{t=1}^{T}} \sum_{g_i^{(s)}=g} \sum_{t=1}^{T}(y_{it} - x_{it}{}'\beta_{g,t})^2$.
4. Set $s = s + 1$.
5. Repeat 2 to 4 until convergence.

In Step 2, each unit is assigned to the group that gives the smallest sum of squared residuals for a given value of the coefficient parameters. Step 3 is just an ordinary least-squares (OLS) estimation of the coefficient for each group using observations belonging to that group given the group membership structure at the iteration.

Similar remarks to the *k-means* algorithm apply here. There is no guarantee that the algorithm converges to the global minimum of the objective function and many initial values need to be tried. The GFE estimator may thus suffer from a computational burden, particularly when the number of groups is large. The supplemental appendix of Bonhomme and Manresa (2015) discusses an alternative algorithm to mitigate the computational concern. Alternatively, several estimators, whose computational costs are substantially lower than that of the GFE estimator, have been proposed. Su, Shi, and Phillips (2016) proposed the CLasso estimator which applies the idea of Lasso (Tibshirani 1996) to panel data models with group heterogeneity. Chen (2019) proposed the hierarchical agglomerative clustering method which is also computationally attractive. Wang, Phillips, and Su (2018) discusses an alternative method.

The group membership estimator $\{\hat{g}_i\}_{i=1}^{N}$ is consistent in the sense that P

$(\hat{g}_i = g_i, \text{for all } i) \to 1$ under a relatively weak condition when both $N \to \infty$ and $T \to \infty$. Accordingly, the coefficient estimator $\{\{\beta_{g,t}\}_{t=1}^{T}\}_{g=1}^{G}$ has an asymptotic distribution which is identical to that under known group memberships. Asymptotic properties are developed in the appendix of Okui and Wang (2020). Note that Bonhomme and Manresa (2015) considered model (2) in the main text and other variants in the supplemental appendix but this most general version is not considered there.

It is easy to modify the algorithm for cases in which some or all of the coefficients are time-invariant and some coefficients are common across groups. Various extensions are discussed in Bonhomme and Manresa (2015).

3.2. Inference for group memberships

This section considers how to make statistical inferences for group memberships. While the estimator for group membership structure is equal to the true group membership structure with probability approaching one, they may not be equal in finite samples. Indeed, simulation results given in, for example, Bonhomme and Manresa (2015) show that it is quite common that there are some units whose groups are misclassified. Note that the consistency of the group membership estimator for all units is not necessary for the coefficient estimator to have good asymptotic properties (Dzemski and Okui 2020b).

We consider the following panel data model with group-specific coefficients:

$$y_{it} = x_{it}'\beta_{g_i,t} + \epsilon_{it}.$$

The key idea is that if the group membership of unit i is g, then the values of the coefficients for group g give the smallest mean squared error. That is, the null hypothesis $H_0 : g_i = g$ is equivalent to

$$\frac{1}{T}\sum_{t=1}^{T} E\left[(y_{it} - x_{it}'\beta_{g,t})^2\right] \le \frac{1}{T}\sum_{t=1}^{T} E\left[(y_{it} - x_{it}'\beta_{h,t})^2\right]$$

for all $h \in \{1,\ldots,G\}\backslash\{g\}$. Let

$$d_{it}(g,h) = \frac{1}{2}\Big((y_{it} - x_{it}'\beta_{g,t})^2 - (y_{it} - x_{it}'\beta_{h,t})^2 + (x_{it}'(\beta_{g,t} - \beta_{h,t}))^2\Big),$$

where the third term in the parentheses $(x_{it}'(\beta_{g,t} - \beta_{h,t}))^2$ is the negative of the mean value of the sum of the first two terms under the null hypothesis. With $d_{it}(g,h)$, $H_0 : g_i^0 = g$ is equivalently written as $E(d_{it}(g,h)) = 0$ for all $h \neq g$. The alternative $H_1 : g_i^0 \neq g$ is equivalent to $\frac{1}{T}\sum_{t=1}^{T} E(d_{it}(g,h)) > 0$ for some $h \neq g$. This observation implies that a group membership structure can be tested by a one-sided test of moment conditions. A confidence set for group membership structure can be obtained by inverting such a test.

We now construct a test statistic. Let $\hat{\beta}_{g,t}$ be a consistent estimator for $\beta_{g,t}$ for $t = 1,\ldots,T$ and $g = 1,\ldots,G$. The sample version of $d_{it}(g,h)$ is

$$\hat{d}_{it}(g,h) = \frac{1}{2}\Big((y_{it} - x_{it}'\hat{\beta}_{g,t})^2 - (y_{it} - x_{it}'\hat{\beta}_{h,t})^2 + (x_{it}'(\hat{\beta}_{g,t} - \hat{\beta}_{h,t}))^2\Big).$$

The t statistic based on $\hat{d}_{it}(g,h)$ is

$$\hat{D}_i(g,h) = \frac{\sum_{t=1}^{T} \hat{d}_{it}(g,h)}{\sqrt{\sum_{t=1}^{T} (\hat{d}_{it}(g,h) - (\bar{\hat{d}}_i(g,h))^2}}$$

where $\bar{\hat{d}}_i(g,h) = T^{-1}\sum_{t=1}^{T} \hat{d}_{it}(g,h)$. Because we reject the null hypothesis only for a positive value of $\hat{D}_i(g,h)$, our test statistic is the maximum of the t statistics:

$$\hat{T}_i(g) = \max_{h \in \mathbb{G}\backslash\{g\}} \hat{D}_i(g,h).$$

The critical value may be obtained using a Bonferroni method. Let α be the significance level. A critical value suggested by Dzemski and Okui (2020a) is

$$c_{\alpha,1,i}(g) = \sqrt{\frac{T-1}{T}} t_{T-1}^{-1}\left(1 - \frac{\alpha}{(G-1)}\right),$$

where $t_{T\text{-}1}(\cdot)$ is the cumulative distribution function of Student's t distribution with $T-1$ degrees of freedom. This critical value is robust to within-unit correlation and can be computed easily in most of statistical and programming software. Alternatively, one may use a multivariate t distribution which takes within-unit correlation into account explicitly and is thus expected to achieve a higher power. This strategy is discussed in Dzemski and Okui (2020a).

A $(1-\alpha)$ confidence set for i's group membership is thus

$$\hat{C}_{\alpha,1,i}\{g \in \{1,\ldots,G\} : \hat{T}_i(g) \le c_{\alpha,1,i}(g)\} \cup \{\hat{g}_i\}. \tag{4}$$

We add the estimated group membership $\hat{g}_i$ to the confidence set because otherwise the confidence set may be empty.

We can also construct a $(1-\alpha)$ confidence set for the entire group membership structure:

$$\hat{C}_{\alpha} = \mathop{\times}_{1\le i\le N} \{g \in \{1,\ldots,G\} : \hat{T}_i(g) \le c_{\alpha,N,i}(g)\} \cup \{\hat{g}_i\}, \tag{5}$$

where

$$c_{\alpha,N,i}(g) = \sqrt{\frac{T-1}{T}} t_{T-1}^{-1}\left(1 - \frac{\alpha}{(G-1)N}\right). \tag{6}$$

$\hat{C}_{\alpha}$ is the Cartesian product of unit-wise confidence sets, but the confidence set for each unit-wise confidence set is adjusted using the Bonferroni correction. It is computationally easy to compute a confidence set even for all units.

Dzemski and Okui (2020a) show that under mild conditions the confidence set asymptotically covers the true group membership structure with probability more than

$1 - \alpha$:

$$\liminf_{N,T\to\infty} P(\{g_i\}_{i=1}^N \in \hat{C}_\alpha) \geq 1-\alpha.$$

The proof of this claim is mathematically quite involved because of the high-dimensional nature of the problem. Nonetheless, an intuition is simple, there are $(G - 1)N$ one-sided t tests and we combine them using the Bonferroni method.

4. Structural breaks

In this section, we consider how to handle structural breaks. In Section 4.1, structural breaks affect the values of the coefficients for each group. In Section 4.2, a structural break also changes the group membership structure.

4.1. Structural breaks in coefficients

In this section, we consider structural breaks in the group-specific coefficients. Considering possible structural changes is important in panel data analysis, in particular when the length of time series is large. While we have allowed time-varying parameters in Section 3, it is often more convenient to model time-varying pattern via structural breaks rather than to consider cases in which coefficients change at every period. Exploiting structural breaks allows us to combine information from multiple time periods between two breaks. We obtain a more efficient estimator. Detecting breaks also facilitates interpretation of the results. Here, we use Lasso to detect breaks following Okui and Wang (2020).

As before, we have data (y_{it}, x_{it}) for $i = 1,\ldots,N$ and $t = 1,\ldots,T$ and consider the following model:

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

where $g_i \in \{1,\ldots,G\}$ is unit i's group membership and ϵ_{it} is an error term. We assume that for each group g, the time-varying pattern of coefficients $\{\beta_{g,1},\ldots,\beta_{g,T}\}$ is characterized by structural breaks. Suppose that group g exhibits J_g breakpoints and $\{T_{g,j}\}_{j=0}^{J_g} \subset \{1,\ldots,T\}$, where $T_{g,0} = 1$ and $T_{g,J_g} = T$ denote the breakpoints for group g. The value of $\beta_{g,t}$ changes only at break dates and remains constant in a regime between two breaks:

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

We allow consecutive structural breaks, that is, breaks to occur in two adjacent periods. A break can occur at the end of the time series. These features are difficult to handle in individual time series analysis or panel data models with complete heterogeneity. Allowing these features is one of the advantages of using panel data models with latent group structure.

We allow break dates to be group specific. Note that existing methods for structural breaks in panel data often assume common break dates across units. For example, Qian and Su (2016a) consider common structural breaks in models with homogeneous coefficients. Baltagi, Feng, and Kao (2016) allow coefficients to be individual specific but keep the assumption of a common breakpoint. We consider a setting that is more general in terms of structural breaks, but more restrictive in terms of heterogeneity, than that of Baltagi, Feng, and Kao (2016).

Okui and Wang (2020) propose a method to jointly estimate the group membership, break dates, and coefficients, called the grouped adaptive group fused lasso (GAGFL). The idea of the GAGFL method is to combine the GFE method by Bonhomme and Manresa (2015) for group membership estimation and adaptive group fused lasso (AGFL) proposed by Qian and Su (2016a) (see also Qian and Su 2016b) for break detection.

The GAGFL estimator minimizes the following objective function.

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

where λ is a tuning parameter and $\dot{w}_{g,t}$ is a data-driven weight

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

κ is a user-specific constant and $\dot{\beta}_{g,t}$ is a preliminary estimate of $\beta_{g,t}$. The tuning parameter λ may be chosen by an information criterion. κ may be set to be 2. For $\dot{\beta}_{g,t}$, we may use the GFE estimates discussed in the previous section.

The first part of the objective function is the sum of squared residuals and is used for the GFE estimator. The second part of the objective function is the penalty term. It is an example of an L_1 penalty and can induce a sparsity property of the coefficients. Because of this penalty, the coefficient estimates may be $\hat{\beta}_{g,t} = \hat{\beta}_{g,t-1}$. Note that these two cannot be the same in the case of the GFE estimation. Breakpoints are estimated to be time periods t with $\hat{\beta}_{g,t} \neq \hat{\beta}_{g,t-1}$.

We adopt the following iterative algorithm for minimization.

1. Let $\{g_i^{(0)}\}_{i=1}^{N}$ be the initial GFE estimate of group membership structure. Set $s = 1$.
2. For the given $\{g_i^{(s-1)}\}_{i=1}^{N}$, compute for each g

$$\{\beta_{g,t}^{(s)}\}_{t=1}^{T} = \underset{\{\beta_{g,t}\}_{t=1}^{T}}{\arg\min} \frac{1}{NT} \sum_{g_i^{(s-1)}=g} \sum_{t=1}^{T} (y_{it} - x_{it}' \beta_{g,t})^2 + \lambda \sum_{t=2}^{T} \dot{w}_{g,t} \, \| \beta_{g,t} - \beta_{g,t-1} \|.$$

3. Compute for all $i \in \{1,\ldots,N\}$,

$$g_i^{(s)} = \underset{g \in \{1,\ldots,G\}}{\arg\min} \sum_{t=1}^{T} (y_{it} - x_{it}' \beta_{g,t}^{(s)})^2.$$

4. Set $s = s + 1$. Go to Step 2 until numerical convergence.

Step 2 is the AGFL estimation using observations belonging to a group according to the group membership assignment at the iteration. Step 3 assigns a group to each unit by minimizing the sum of squared residuals given the value of coefficients. Note that if we use the GFE estimator as a preliminary estimate, then iterations are not needed to obtain estimators with desirable asymptotic properties. Just one step of an application of the AGFL estimator is sufficient. Nonetheless, Okui and Wang (2020) show through their simulations that the estimator obtained after the iteration converges performs better. Note also that the algorithm converges quickly, which is expected if we use the GFE estimator as an initial value. The computational burden of this algorithm mainly comes from the preliminary estimation using the GFE method.

Okui and Wang (2020) prove the consistency of the estimator for the group membership structure, breakpoints, and the coefficients. Because the group membership structure and breakpoints estimators are equal to the true values with probability approaching one, the asymptotic distribution of the coefficient estimator is equivalent to that under known group membership structure and breakpoints.

Note that it is easy to modify the algorithm for situations in which some of the coefficients are known to be fully time-varying. In that case, those coefficients are not included in the penalty term. The algorithm can also be modified to treat cases in which some coefficients are known to be time-invariant and/or common across groups. Those extensions are discussed in Okui and Wang (2020).

4.2. Structural breaks in group membership structure

This section considers situations in which a structural break changes the group membership structure in addition to the coefficients. We have already seen that time-varying coefficients can be handled, and a method is available for detecting structural breaks that affect the value of the coefficients. However, they assume that the group membership structure is fixed over time. Lumsdaine, Okui, and Wang (2020) consider a break that may affect the group membership structure and proposes a statistical procedure to detect it.

Again, suppose that we have data (y_{it}, x_{it}) for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. We consider the following model:

$$y_{it} = x_{it}'\beta_{g_{it},t} + \epsilon_{it},$$

where $\beta_{g,t}$ is the coefficient for group g at time t and g_{it} is the group membership for unit i at time t. The coefficient is time-varying and so is the group membership.

We assume that the time-varying pattern of both the group membership and the coefficient value are determined by a single breakpoint. We assume that there is a break at k^0 common to all groups and the values of the coefficients are

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

where $g_i(B) \in \{1,\ldots,G^B\}$ is unit i's group membership before the break and $g_i(A) \in \{1,\ldots,G^A\}$ stands for unit i's group membership after the break. Note that the number of groups can change at the breakpoint. A caveat is that the labeling of groups is arbitrary and there may not be any relationship between the group labels before the break and those after the break. For example, group 1 before the break may not have any relationship with group 1 after the break.

We estimate the vectors of the coefficients before and after the break, $(\{\beta_{g,B}\}_{g=1}^{G^B}, \{\beta_{g,A}\}_{g=1}^{G^A})$, the group membership structure, $(\{g_i(B), g_i(A)\}_{i=1}^{N})$, and the breakpoint $k \in \{2,\ldots,T-1\}$. To estimate the parameters, we minimize the sum of squared residuals:

$$\begin{aligned} &(\{\hat{\beta}_{g,B}\}_{g=1}^{G^B}, \{\hat{\beta}_{g,A}\}_{g=1}^{G^A}, (\{\hat{g}_i(B), \hat{g}_i(A)\}_{i=1}^{N}), \hat{k}) \\ &= \underset{\{\beta_{g,B}\}_{g=1}^{G^B}, \{\beta_{g,A}\}_{g=1}^{G^A}, (\{g_i(B), g_i(A)\}_{i=1}^{N}), k}{\arg\min} \left(\sum_{t=1}^{k-1}\sum_{i=1}^{N} (y_{it} - x_{it}'\beta_{g_i(B),B})^2 + \sum_{t=k}^{T}\sum_{i=1}^{N} (y_{it} - x_{it}'\beta_{g_i(A),A})^2 \right). \end{aligned}$$

The minimization problem is solved using the following iterative algorithm.

1. Set $s = 1$. For a given $k \in \{2,\ldots,T-1\}$, initialize group structures in both regimes as $\{g_i(B)^{(0)}, g_i(A)^{(0)}\}_{i=1}^{N}$.
2. For given $\{g_i(B)^{(s-1)}, g_i(A)^{(s-1)}\}_{i=1}^{N}$, and k, estimate the slope coefficient

$\{\beta_{g,B}\}_{g=1}^{G^B}, \{\beta_{g,A}\}_{g=1}^{G^A}$ in the two regimes by

$$\beta_{g,B}^{(s)} = \arg\min_{\beta_{g,B}} \sum_{t=1}^{k-1} \sum_{g_i(B)^{(s-1)}=g} (y_{it} - x_{it}'\beta_{g,B})^2,$$

$$\text{and } \beta_{g,A}^{(s)} = \arg\min_{\beta_{g,A}} \sum_{t=k}^{T} \sum_{g_i(A)^{(s-1)}=g} (y_{it} - x_{it}'\beta_{g,A})^2.$$

3. Given $\{\beta_{g,B}^{(s)}\}_{g=1}^{G^B}$, $\{\beta_{g,A}^{(s)}\}_{g=1}^{G^A}$, find the optimal group for individual i in each regime by

$$g_i(B)^{(s)} = \arg\min_{g} \sum_{t=1}^{k-1} (y_{it} - x_{it}'\beta_{g,B}^{(s)})^2,$$

$$\text{and } g_i(A)^{(s)} = \arg\min_{g} \sum_{t=k}^{T} (y_{it} - x_{it}'\beta_{g,A}^{(s)})^2.$$

4. Iterate Steps 2 and 3 until numerical convergence. Let the values of parameters obtained after convergence be $(\{\hat{\beta}_{g,B}(k)\}_{g=1}^{G^B}, \{\hat{\beta}_{g,A}(k)\}_{g=1}^{G^A}, \{\hat{g}_i(B,k), \hat{g}_i(A,k)\}_{i=1}^{N})$.

5. Let k vary from 2 to T, and estimate the breakpoint by

$$\hat{k} = \arg\min_{k\in\{2,\cdots,T-1\}} \sum_{t=1}^{k-1}\sum_{i=1}^{N} (y_{it} - x_{it}'\hat{\beta}_{\hat{g}_i(B,k),B}(k))^2 + \sum_{t=k}^{T}\sum_{i=1}^{N} (y_{it} - x_{it}'\hat{\beta}_{\hat{g}_i(A,k),A}(k))^2.$$

Step 1 initializes the parameters. In Step 2, we compute an estimate of β given a breakpoint and a group membership structure. Note that this is just computing an OLS estimate. Step 3 updates the group membership structure. It is a least-squares estimation of group memberships. Step 4 states that we iterate Steps 2 and 3 until convergence, and this is essentially the *k-means* algorithm. In other words, we apply the GFE estimation of Bonhomme and Manresa (2015) to data before and after the break separately. We repeat Steps 1–4 for each k. Lastly, in Step 5, we find the value of k that minimizes the objective function. This break estimation approach is a least-squares method, and the idea is similar to those discussed in Bai (1997).

Lumsdaine, Okui, and Wang (2020) show that the breakpoint and the group membership structure can be estimated consistently. In particular, it is a "super" consistency result in the sense that with probability approaching one, the probability of the breakpoint and the group membership estimates are exactly equal to the true values. This super-consistency result implies that the coefficient estimator asymptotically behaves as if the breakpoint and the group membership structure were known. That is, the coefficient estimator is the OLS estimator applied to each regime and each group, and statistical inferences may be conducted as in the case of the usual OLS estimator.

5. Conclusion

This review summarizes the recent development of grouping methods in econometrics. Grouping methods are useful to analyze heterogeneity in panel data. They allow us to obtain precise estimates of group-specific coefficients and to interpret the results easily.

We expect that grouping methods will be applied to many different economic problems and there will be more econometric works in this field. The GFE estimator and other related methods have been extended to other types of models such as nonlinear models (see, e.g., Gu and Volgushev 2019; Liu et al. 2020) and nonstationary models (see, e.g., Huang, Jin, and Su 2019). We expect other different models and settings to be considered in the near future. It is also important to examine further the inference issues for the estimators obtained by grouping methods. Because of the super-consistency of the group membership estimator, the usual methods for statistical inferences for the OLS estimator may be used, but it is not clear if such statistical inferences are uniformly valid for a class of data generating processes. In addition, there has not been much discussion on how to make statistical inferences on the number of groups. An important exception is Lu and Su (2017), and more research is desirable.

Reference

Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. 2008. "Income and Democracy." American Economic Review 98 (3): 808–42.

Angrist, Jushua D., and Jörn-Steffen Pischke. 2008. Mostly Harmless Econometrics. Princeton University Press.

Bai, Jushan. 1997. "Estimation of a Change Point in Multiple Regression Models." The Review of Economics and Statistics 79: 551–63.

Baltagi, Badi H., Qu Feng, and Chihwa Kao. 2016. "Estimation of Heterogeneous Panels with Structural Breaks." Journal of Econometrics 191: 176–95.

Bonhomme, Stéphane, and Elena Manresa. 2015. "Grouped Patterns of Heterogeneity in Panel Data." Econometrica 83 (3): 1147–84.

Chen, Jia. 2019. "Estimating Latent Group Structure in Time-Varying Coefficient Panel Data Models." The Econometrics Journal 22 (3): 223–40.

Dempster, Arthur, Nan Laird, and Donald Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society. Series B (Methodological), 1–38.

Dzemski, Andreas, and Ryo Okui. 2020a. "Confidence Set for Group Membership."

______. 2020b. "Convergence Rate of Estimators of Clustered Panel Models with Misclassification."

Fernández-Val, Iván, and Joonhwah Lee. 2013. "Panel Data Models with Nonadditive Unobserved Heterogeneity: Estimation and Inference." Quantitative Economics 4: 453–81.

Gu, Jiaying, and Stanislav Volgushev. 2019. "Panel Data Quantile Regression with Grouped Fixed Effects." Journal of Econometrics 213 (1): 68–91.

Huang, Wenxin, Sainan Jin, and Liangjun Su. 2019. "Panel Cointegration with Latent Group Structures and an Application to the PPP Theory." Econometric Theory, Forthcoming.

Imai, Kosuke. 2018. Quantitative Social Science: An Introduction. Princeton University Press.

Liu, Ruiqi, Zuofeng Shang, Yonghui Zhang, and Qiankun Zhou. 2020. "Identification and Estimation in Panel Models with Overspecified Number of Groups." Journal of Econometrics 215 (2): 574–90.

Lu, Xun, and Liangjun Su. 2017. "Determining the Number of Groups in Latent Panel Structures with an Application to Income and Democracy." Quantitative Economics 8 (3): 729–60.

Lumsdaine, Robin L., Ryo Okui, and Wendun Wang. 2020. "Estimation of Panel Group Structure Models with Structural Breaks in Group Memberships and Coefficients."

Murphy, Kevin P. 2012. Machine Learning: A Probabilistic Perspective. MIT press.

Okui, Ryo, and Wendun Wang. 2020. "Heterogeneous Structural Breaks in Panel Data Models." Journal of Econometrics.

Okui, Ryo, and Takahide Yanagi. 2019. "Panel Data Analysis with Heterogeneous Dynamics." Journal of Econometrics 212 (2): 451–75.

______. 2020. "Kernel Estimation for Panel Data with Heterogeneous Dynamics." The Econometrics Journal 23 (1): 156–75.

Pollard, David. 1981. "Strong Consistency of k-Means Clustering." The Annals of Statistics, 135–40.

______. 1982. "A Central Limit Theorem for *k*-Means Clustering." The Annals of Probability 10 (4): 919–26.

Qian, Junhui, and Liangjun Su. 2016a. "Shrinkage Estimation of Common Breaks in Panel Data Models via Adaptive Group Fused Lasso." Journal of Econometrics 191: 86–109.

______. 2016b. "Shrinkage Estimation of Regression Models with Multiple Structural Changes." Econometric Theory 32: 1376–1433.

Su, Liangjun, Zhentao Shi, and Peter C. B. Phillips. 2016. "Identifying Latent Structures in Panel Data." Econometrica 84 (6): 2215–64.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological), 267–88.

Wang, Wuyi, Peter C. B. Phillips, and Liangjun Su. 2018. "Homogeneity Pursuit in

Panel Data Models: Theory and Application.” Journal of Applied Econometrics 33: 797–815.