Large-scale generalized linear longitudinal data models with grouped patterns of unobserved heterogeneity

Ando, Tomohiro and Bai, Jushan

23 December 2021

Large-scale generalized linear longitudinal data models with grouped patterns of unobserved heterogeneity*

Tomohiro Ando† and Jushan Bai‡

January 8, 2022

Abstract

This paper provides methods for flexibly capturing unobservable heterogeneity from longitudinal data in the context of an exponential family of distributions. The group memberships of individual units are left unspecified, and their heterogeneity is influenced by group-specific unobservable structures, as well as heterogeneous regression coefficients. We discuss a computationally efficient estimation method and derive the corresponding asymptotic theory. The established asymptotic theory includes verifying the uniform consistency of the estimated group membership. To test the heterogeneous regression coefficients within groups, we propose the Swamy-type test, which considers unobserved heterogeneity. We apply the proposed method to study the market structure of the taxi industry in New York City. Our method reveals interesting important insights from large-scale longitudinal data that consist of over 450 million data points.

Keywords: Clustering; Factor analysis; Generalized linear models; Longitudinal data; Unobserved heterogeneity.

*We would like to thank Joint Editor, Associate Editor and two anonymous referees for their insightful comments that greatly improved our paper. The authors would like to thank the participants at the 14th International Conference on Computational and Financial Econometrics 2020 and the China Meeting of the Econometric Society 2021, for their constructive comments.

†Melbourne Business School, Melbourne University, T.Ando@mbs.edu. 200 Leicester Street, Carlton, Victoria 3053, Australia.

‡Department of Economics, Columbia University, jb3064@columbia.edu. 1019 International Affairs Building 420 West 118 Street New York, NY 10027 USA

1 Introduction

"Clustering" is one of the most popular methods for grouping data; in a wide range of disciplines, various clustering approaches are employed. Examples include the work of Park and Park (2016), who used clustering for understanding and predicting online customers' store visits and purchase behaviors, a study by Kong et al. (2010) to improve the efficiency of a liver allocation system, and research by France and Ghose (2016) to identify market structures, and capture "varieties of capitalism" (Ahlquist and Breunig (2012)) and for the analysis of microarray gene expression data (Qin (2006), Heard et al. (2006), Chiou and Li (2007), Cai et al. (2019)). James and Suger (2003) developed a flexible model-based procedure for clustering functional data. Peng and Muller (2008) applied distance-based clustering to analyze online auction data. Hancock et al. (2007) analyzed social networks based on model-based clustering. Delaigle et al. (2012) developed approaches for clustering functional data. We refer to Hennig et al. (2015) for an effective overview of the growing number of interdisciplinary applications of cluster analysis.

This paper attempts to identify the latent group structure within longitudinal data in the context of an exponential family of distributions. There is a vast volume of literature on clustering individual units of "linear" longitudinal data (Lin and Ng (2012), Bonhomme and Manresa (2015), Su et al. (2016), Ando and Bai (2016, 2017), Vogt and Linton (2017), Wang et al. (2018), Lumsdaine et al. (2020) among others). Zhang (2019) proposed a quantile regression-based method for panel data to identify subgroups and estimate group-specific parameters, which does not always allow us to analyze data from an exponential family of distributions, for example, binary response.

In contrast to these previous studies, this paper employs generalized linear models for capturing grouped patterns of unobserved heterogeneity in longitudinal data. Studies on clustering longitudinal data with unobserved heterogeneity in the context of an exponential family of distributions are scant. Many previous studies on binary response longitudinal data with unobserved effects (for, e.g., Fernández-Val and Weidner (2016), Charbonneau (2017), Moon et al. (2018), Boneva and Linton (2017), Ando et al. (2021), Chen et al. (2021)) did not accommodate the unobserved group membership structure. Additionally, many studies have considered group-specific regression coefficients, and they all assumed that the regression coefficients are homogeneous across units within groups but can differ across groups. Wang and Su (2021) considered a procedure for identifying latent group structures in nonlinear longitudinal data models when some regression coefficients are heterogeneous across groups but homogeneous within a group. Although such setups often simplify the estimation and inference procedures, the resultant conclusions can be misleading when the group-specific homogeneity assumption does not hold. Our empirical analysis indicates that the group-specific homogeneous assumption is not supported by data. We, therefore, allow for heterogeneous regression coefficients even within groups.

We emphasize that the extension from a linear model to the generalized linear model is not a straightforward task. In terms of estimation, we need to develop an iterative algorithm that changes depending on whether it is used for group membership estimation, regression coefficient estimation or unobservable structure estimation. This extension also involves several theoretical challenges that must be overcome.

First, the group membership structure is assumed to be unknown and estimated from the data. A natural question is whether we can correctly identify the group membership structure. To address this important concern, we establish a "uniform" consistency for the estimated group membership. It should be noted that the number of individuals goes to infinity in our theoretical setting, and thus, the proof of "uniform" consistency is not straightforward. By addressing this challenge, we show that the proposed method can identify true latent group structures with probability approaching one. Second, our asymptotic theory also addresses the consistency of the estimated regression coefficients, as well as that of the estimated unobserved heterogeneity. We note that the regression coefficients vary across individual units, and the number of individual units is allowed to go to infinity. Thus, the study of asymptotic theory for regression coefficients is a difficult task. Third, with regard to checking whether the regression coefficients are heterogeneous within a group, no approach has been proposed for large-scale generalized linear longitudinal models with grouped patterns of unobserved heterogeneity. There are many studies that have tested for the homogeneity of regression coefficients in linear longitudinal data models, including those of Pesaran et al. (1996), Phillips and Sul (2003), Pesaran and Yamagata (2008), Blomquist and Westerlund (2013) and Ando and Bai (2015). However, studies testing for the homogeneity of regression coefficients in longitudinal data models under an exponential family of distributions are very limited. We propose the Swamy-type test and investigate the asymptotic distribution of our test statistic. This is the first result obtained by testing for the homogeneity of the regression coefficients in nonlinear longitudinal data models under an exponential family of distributions.

This paper applies the proposed method to large-scale longitudinal data from the taxi industry in New York City (NYC). In NYC, the taxi industry is highly regulated by the NYC Taxi & Limousine Commission (TLC). Some of the restrictions on the industry include the use of common pricing schemes and limits to the total number of taxis. Additionally, profitability may vary across firms, and this is mainly due to the efficiency with which taxi firms utilize their resources. For example, one of the ways in which firms can improve their financial performance is through the improvement of their capacity utilization rate. This rate, for example, can be measured by the fraction of time that a driver has a fare-paying passenger. From a managerial perspective, the performance evaluation of each individual taxi can prominently benefit a firm. It will allow the firm to compare its performance with that of its competitors, and it will yield information for making managerial decisions to improve performance. We empirically examine the efficiency of yellow cab medallion taxis in NYC by applying the proposed clustering approach.

In summary, our paper makes the following contributions. First, we introduce new nonlinear longitudinal data models with grouped patterns of unobserved heterogeneity. Second, a new model estimation and selection procedure is developed for the introduced model. Third, our established theory shows that the true group membership and the proposed estimator are asymptotically equivalent. This result considers the uniform consistency of the estimated group membership and thus identifies an attractive property of the proposed estimator. Fourth, we derive an asymptotic theory for the estimated model parameters. It is shown that the asymptotic distribution of the estimated regression coefficients is the multivariate normal distribution. Additionally, the asymptotic distribution of the estimated common factor is shown to be multivariate normal. Fifth, we propose the Swamy-type test to investigate whether the regression coefficients are heterogeneous within a group. Finally, we apply our methods to estimate the capacity utilization rates based on NYC yellow taxi data. Our method reveals interesting and important insights from a large-scale longitudinal dataset that consists of over 450 million data points.

The remainder of this paper is organized as follows. Section 2 introduces the nonlinear longitudinal data model with a factor structure, and Section 3 describes the modeling, estimation, and model selection procedures. Section 4 investigates the consistency of the proposed estimator. Its asymptotic behaviors are also investigated, and further theoretical studies are conducted. To save space, all technical proofs are provided in the online supplementary document, which also contains Monte Carlo simulation results. Section 5 applies the procedure to the analysis of the TLC dataset. Concluding remarks are provided in Section 6.

2 Model

In this section, we introduce a new model: a generalized linear longitudinal data model with grouped patterns of unobserved heterogeneity. Suppose that the response of an individual unit is measured over T time periods together with some observable explanatory variables. For the i-th individual unit ($i = 1, ..., N$), at time t, its response y_{it} is observed together with a $(p+1)$-

dimensional vector of explanatory variables $\boldsymbol{x}_{it} = (1, x_{it,1}, ..., x_{it,p})'$. We let S be the number of groups (which is unknown) and let $G = \{g_1, ..., g_N\}$ denote the group membership such that $g_i \in \{1, ..., S\}$. Additionally, we denote N_j as the number of individual units within group j $(j = 1, ..., S)$ such that $N = \sum_{j=1}^{S} N_j$.

To capture grouped patterns of unobserved heterogeneity in nonlinear longitudinal data, we consider an exponential family of distributions:

$$f(y_{it}|\theta_{it,g_i}) = \exp\left\{y_{it}\theta_{it,g_i} - b(\theta_{it,g_i}) + c(y_{it})\right\} \tag{1}$$

with the natural parameter θ_{it,g_i} being expressed as

$$\theta_{it,g_i} \equiv \boldsymbol{x}_{it}'\boldsymbol{b}_i + \eta_{it,g_i}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T, \tag{2}$$

where $\boldsymbol{b}_i = (b_{i,0}, b_{i,1}, ..., b_{i,p})'$ is a $(p+1)$-dimensional vector of regression coefficients and η_{it,g_i} denotes the unobservable effects that also explain the characteristics of the variability in y_{it}. In this paper, we assume that the unobserved structure η_{it,g_i}, which depends on the group membership g_i, is modeled with a factor structure:

$$\eta_{it,g_i} = \sum_{\ell=1}^{r} f_{\ell t,g_i}\lambda_{i\ell,g_i} = \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i}, \tag{3}$$

where $\boldsymbol{f}_{t,g_i}$ is an $r_{g_i} \times 1$ vector of group-specific unobservable factors and $\boldsymbol{\lambda}_{i,g_i}$ represents the factor loadings. We note that the explanatory variables may be correlated with the group-specific factors, factor loadings, or both. In such cases, ignoring the factor structure leads to inconsistent estimates of the regression coefficients (see, e.g., Bai (2009), Pesaran (2006)).

Combining (1), (2) and (3), our nonlinear longitudinal model is formulated as

$$f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}) = \exp\left\{y_{it}(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i}) - b(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i}) + c(y_{it})\right\}. \tag{4}$$

Below are some specific examples of our model (4).

Example 1 Consider the standard linear model $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{b}_i + \eta_{it,g_i} + \varepsilon_{it}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$. Assuming that ε_{it} follows a normal distribution with mean 0 and variance σ^2, the model (4) is given as

$$f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i})^2}{\sigma^2}\right\}.$$

Example 2 Let y_{ij} be a binary outcome such that y_{it} takes the value 0 or 1 and let Φ be a cumulative distribution function for a standard normal or logistic distribution. The model (4) is given as

$$f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}) = \Phi(\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i})^{y_{it}}(1 - \Phi(\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}))^{1-y_{it}}.$$

If the logistic distribution is used, then Φ is $\Phi(y_{it} = 1|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}) = \exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i})/\{1 + \exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i})\}$.

Example 3 Let y_{ij} be a count outcome. A Poisson model can be considered with the conditional expectation of y_{it} written as $\exp(y_{it}(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i}))$.

$$f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}) = \Psi(y_{it}, \exp(y_{it}(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i})),$$

where $\Psi(y, \alpha)$ is the probability mass function of a Poisson random variable with parameter α.

In addition to Example 1 – Example 3, other examples, such as the binomial model, inverse Gaussian model, and gamma model, can be considered. When y_{it} is a selection out of multiple possible choices, a multinomial choice model can be considered.

The unknown parameters are the regression coefficients $B = \{\boldsymbol{b}_1, ..., \boldsymbol{b}_N\}$, the group membership G, the group-specific factors $F_j = (\boldsymbol{f}_{1,j}, ..., \boldsymbol{f}_{T,j})'$ and the corresponding factor loadings $\Lambda_j = (\boldsymbol{\lambda}_{1,j}, ..., \boldsymbol{\lambda}_{N_j,j})'$ for $j = 1, ..., S$. The number of groups S and the dimensions of the group-specific factors are unspecified, and thus, we need to determine these values. These estimation and model selection problems are investigated in Section 3.

3 Estimation

3.1 Model building

In this section, we describe our model-building framework. This involves estimating model parameters as well as identifying the number of groups and the numbers of group-specific factors.

Given the number of groups S and the number of factors in group r_j ($j = 1, 2, ..., S$), the estimator $\{\hat{B}, \hat{G}, \hat{F}_1, ..., \hat{F}_S, \hat{\Lambda}_1, ..., \hat{\Lambda}_S\}$ is defined as the maximizer of the following log-likelihood function:

$$L(B, G, F_1, ..., F_S, \Lambda_1, ..., \Lambda_S) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\log f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}), \tag{5}$$

subject to normalization restrictions on F_j and Λ_j. Following Bai and Ng (2013), we impose the following restriction: $F_j'F_j/T = I_{r_j}$, and $\Lambda_j'\Lambda_j$ is the diagonal. Bai and Ng (2013) demonstrated that this restriction leads to the identification of F_j and Λ_j.

We first note that finding the exact maximizer of the likelihood function (5) subject to normalization restrictions is not an easy task. This is largely because the number of possible combinations of the group membership G can be enormous when the number of individuals N is large. In that case, it is computationally infeasible to explore all possible combinations of group membership. Because the parameters depend on one another, we need to update the set of parameters sequentially.

Given the group-specific factors $\bar{F}_j$ ($j = 1, 2, ..., S$), we can easily update the group membership g_i, the regression coefficients $\boldsymbol{b}_i$ and the factor loadings $\boldsymbol{\lambda}_{i,g_i}$ as

$$\{g_i, \boldsymbol{b}_i, \boldsymbol{\lambda}_{i,g_i}\} = \mathrm{argmax}\frac{1}{T}\sum_{t=1}^{T}\log f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \bar{\boldsymbol{f}}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}), \tag{6}$$

for $i = 1, ..., N$.

Given $\bar{B}, \bar{\Lambda}_1, ..., \bar{\Lambda}_S$ and $\bar{G}$, we update the group-specific factors $\boldsymbol{f}_{t,j}$ as

$$\boldsymbol{f}_{t,j} = \mathrm{argmax}\frac{1}{N_j}\sum_{i;\ \bar{g}_i=j}\log f(y_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{b}}_i, \boldsymbol{f}_{t,j}, \bar{\boldsymbol{\lambda}}_{i,j}), \tag{7}$$

for $j = 1, ..., S$. Then, we calculate the matrix $A_j = (\frac{1}{T}F_j'F_j)^{1/2}(\frac{1}{N_j}\Lambda_j'\Lambda_j)(\frac{1}{T}F_j'F_j)^{1/2}$ and its associated diagonalization $A_j = U_jV_jU_j'$, where U_j is an orthogonal matrix and V_j is a diagonal matrix. The group-specific common factors and the factor-loading matrices satisfying the normalization restriction are given as follows: $\Lambda_j^\dagger = \Lambda_j(\frac{1}{T}F_j'F_j)^{1/2}U_j$ and $F_j^\dagger = F_j(\frac{1}{T}F_j'F_j)^{-1/2}U_j$.

To capture the dependencies between the regressors and unobservable structures simultaneously due to the endogeneity problem, we use the following algorithm. Although our computation is approximate, as is mostly done in practice for clustering analyses, it allows us to obtain approximate solutions quickly.

Estimation algorithm

Step 1. Fix S and $\{r_1, ..., r_S\}$. Initialize the unknown parameters $B^{(0)} = \{\boldsymbol{b}_1^{(0)}, ..., \boldsymbol{b}_N^{(0)}\}$, $\{F_j^{(0)}, \Lambda_j^{(0)};\ j = 1, ..., S\}$ and $G^{(0)} = \{g_1^{(0)}, ..., g_N^{(0)}\}$.

Step 2. Given the values of $\{F_j,\ j = 1, ..., S\}$, update $\{g_i, \boldsymbol{b}_i, \boldsymbol{\lambda}_{i,g_i}\}$ for $i = 1, ..., N$.

Step 3. Given the values of B, $\{\Lambda_j,\ j = 1, ..., S\}$ and G, update $\boldsymbol{f}_{t,j}$ for $t = 1, ..., T$, $j = 1, ..., S$.

Step 4. Repeat Steps 2 $\sim$ 3 until convergence is achieved.

In Step 1, starting values are needed. To obtain the initial group membership $G^{(0)}$, we simply apply the well-known K-means algorithm to the longitudinal data of y_{it}. This algorithm divides the individual units into S groups. An initial estimate of $\boldsymbol{b}_i^{(0)}$ ($i = 1, ..., N$) is obtained via the standard maximum likelihood approach by ignoring the factor structures. Given $G^{(0)}$ and $\boldsymbol{b}_i^{(0)}$ ($i = 1, ..., N$), we obtain the starting values $\{F_j^{(0)}, \Lambda_j^{(0)}\}$ for group j by applying the standard principal component approach to y_{it}, which belongs to group j.

While the convergence of the algorithm to a local maximum is guaranteed, the proposed algorithm cannot guarantee convergence to a global optimum (see, e.g., Bai (2009)). To minimize the chance of a local minimum, multiple starting values from the K-means algorithm can be used for Step 1 in our estimation algorithm. Following Bai (2009), upon convergence, we choose the estimator that gives the largest likelihood function value.

In practice, the number of groups S and the number of group-specific factors $\{r_1, ..., r_S\}$ are unknown. We use an information criterion (IC) to select these quantities.

$$\begin{aligned} & \mathrm{IC}(S, k_1, ..., k_S) \\ = \; & -\frac{2}{NT}\sum_{j=1}^{S}\sum_{i;\hat{g}_i=j}\log f(y_{it}|\boldsymbol{x}_{it}, \hat{\boldsymbol{b}}_i, \hat{\boldsymbol{f}}_{\hat{g}_i,t}, \hat{\boldsymbol{\lambda}}_{\hat{g}_i,i}) + \sum_{j=1}^{S} k_j \frac{N_j}{N}\left(\frac{T+N_j}{TN_j}\right)\log\left(\frac{TN_j}{T+N_j}\right) \end{aligned} \quad (8)$$

By minimizing the IC, we can choose the number of groups S and the number of group-specific factors k_j $(j = 1, ..., S)$.

Remark 1 While this paper focuses on (4), a researcher may be interested in estimating a model having both global common factors across all groups and group-specific factors:

$$\begin{aligned} & f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,c}, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,c}, \boldsymbol{\lambda}_{i,g_i}) \\ & = \exp\left\{y_{it}(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_{t,c}'\boldsymbol{\lambda}_{i,c} + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i}) - b(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_{t,c}'\boldsymbol{\lambda}_{i,c} + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i}) + c(y_{it})\right\}, \end{aligned} \quad (9)$$

where $\boldsymbol{f}_{t,c}$ is an r_c-dimensional vector of global common factors across all groups and $\boldsymbol{\lambda}_{i,c}$ is the corresponding r_c-dimensional factor-loading vector. For the linear case, Ando and Bai (2017) studied its estimation procedure. By extending their estimation procedure to the exponential family of distributions, one can estimate the parameters in the model (9). To avoid repeating the similar argument provided in this section, an estimation procedure for (9) is delegated to Appendix A in the supplementary document.

3.2 Convergence to a local maximum

Given $B, \Lambda_1, ..., \Lambda_S$ and G, the log-likelihood function in (5) is concave in $F_1, ..., F_S$. Therefore, we have

$$L(B, G, F_1^{old}, ..., F_S^{old}, \Lambda_1, ..., \Lambda_S) \leq L(B, G, F_1^{new}, ..., F_S^{new}, \Lambda_1, ..., \Lambda_S), \quad (10)$$

where $F_1^{new}, ..., F_S^{new}$ can be obtained from (7). Under the fixed $\{F_1, ..., F_S\}$ and G, the log-likelihood function in (5) is concave in $B(G)$ and $\Lambda_1(G), ..., \Lambda_S(G)$, where we emphsize the dependency of B and $\Lambda_1, ..., \Lambda_S$ on G. This implies that, for any G, the following inequality holds:

$$L(B^{old}(G), G, F_1, ..., F_S, \Lambda_1^{old}(G), ..., \Lambda_S^{old}(G)) \leq L(B^{new}(G), G, F_1, ..., F_S, \Lambda_1^{new}(G), ..., \Lambda_S^{new}(G)),$$

where $B^{new}(G), \Lambda_1^{new}(G), ..., \Lambda_S^{new}(G)$ can be obtained from

$$\{\boldsymbol{b}_{g_i}, \boldsymbol{\lambda}_{g_i}\} = \text{argmax} \frac{1}{T} \sum_{t=1}^{T} \log f(y_{it} | \boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i})$$

with $\{F_1, ..., F_S\}$ and G being fixed. Thus, under the fixed value of the group-specific factors $\{F_1, ..., F_S\}$, (6) leads

$$\begin{aligned} &L(B^{old}(G^{old}), G^{old}, F_1, ..., F_S, \Lambda_1^{old}(G^{old}), ..., \Lambda_S^{old}(G^{old})) \\ &\leq L(B^{new}(G^{old}), G^{old}, F_1, ..., F_S, \Lambda_1^{new}(G^{old}), ..., \Lambda_S^{new}(G^{old})) \\ &\leq L(B^{new}(G^{new}), G^{new}, F_1, ..., F_S, \Lambda_1^{new}(G^{new}), ..., \Lambda_S^{new}(G^{new})). \end{aligned} \tag{11}$$

We therefore have

$$\begin{aligned} &L(B^{old}, G^{old}, F_1^{old}, ..., F_S^{old}, \Lambda_1^{old}, ..., \Lambda_S^{old}) \\ &\leq L(B^{old}, G^{old}, F_1^{new}, ..., F_S^{new}, \Lambda_1^{old}, ..., \Lambda_S^{old}) \\ &\leq L(B^{new}, G^{new}, F_1^{new}, ..., F_S^{new}, \Lambda_1^{new}, ..., \Lambda_S^{new}), \end{aligned}$$

where we used (10) for the first inequality and (11) for the last inequality. This implies that the convergence of the algorithm to a local maximum is guaranteed at least. However, it is known that convergence to a global optimum is not guaranteed for interactive effects panel data models (see, e.g., Bai (2009)).

4 Asymptotic properties

4.1 Assumptions

We first state the assumptions needed for the asymptotic analysis. We then define some notations. Let $\|A\| = [tr(A'A)]^{1/2}$ be the usual norm of the matrix A, where "tr" denotes the trace of a square matrix. Equation $a_n = O(b_n)$ states that the deterministic sequence a_n is, at most, of order b_n; $c_n = O_p(d_n)$ states that the random variable c_n is, at most, of order d_n in terms of probability, and $c_n = o_p(d_n)$ is of a smaller order in terms of probability.

The true regression coefficient is denoted by $B_0 = \{\boldsymbol{b}_{1,0}, .., \boldsymbol{b}_{N,0}\}$. The true group-specific factor and the factor loading of individual i with true group membership $g_i^0 = j$ are denoted as $F_{j,0} = (\boldsymbol{f}_{1,j,0}, ..., \boldsymbol{f}_{T,j,0})'$ and $\boldsymbol{\lambda}_{i,j,0}$, respectively. We denote the true factor-loading matrix for the j-th group as $\Lambda_{j,0} = [\boldsymbol{\lambda}_{1,j,0}, \ldots, \boldsymbol{\lambda}_{N_j,j,0}]'$. Because the dimensions of Λ_j and F_j are divergent, we cannot assume that the standard regularity conditions for likelihood functions hold. The set of regularity conditions is as follows:

Assumption A: Common factors

Let $\mathcal{F}_j$ be a compact subset of R^{r_j}. For each group $j = 1, .., S$, the group-specific factors $\boldsymbol{f}_{t,j,0} \in \mathcal{F}_j$ satisfy $T^{-1}\sum_{t=1}^{T} \boldsymbol{f}_{t,j,0}\boldsymbol{f}_{t,j,0}{}' = I_{r_j}$. Although correlations between $\boldsymbol{f}_{t,j,0}$ and $\boldsymbol{f}_{t,k,0}$ $(j \neq k)$ are allowed, they are not perfectly correlated.

Assumption B: Factor loadings and regression coefficients

(B1): Let $\mathcal{B}$ and $\mathcal{L}_j$ be compact subsets of R^p and R^{r_j}, respectively. The regression coefficient $\boldsymbol{b}_i$ and the factor-loading satisfy $\boldsymbol{b}_i \in \mathcal{B}$ and $\boldsymbol{\lambda}_{i,g_i} \in \mathcal{L}_j$. Additionally, the factor-loading matrix $\Lambda_{j,0}$ satisfies $N_j^{-1}\Lambda_{j,0}'\Lambda_{j,0} \to \Sigma_{\Lambda_j}$ as $N_j \to \infty$, where Σ_{Λ_j} is an $r_j \times r_j$ diagonal matrix with distinct diagonal elements and the smallest diagonal element bounded away from zero.

(B2): For each i, the factor-loading satisfies $\|\boldsymbol{\lambda}_{i,g_i^0}\| > 0$.

(B3): For each i and j, $\boldsymbol{f}_{t,j,0}'\boldsymbol{\lambda}_{i,j,0}$ are strongly mixing processes with mixing coefficients that satisfy $r(t) \leq \exp(-a_1 t^{b_1})$ with tail probability $P(|\boldsymbol{f}_{t,j,0}'\boldsymbol{\lambda}_{i,j,0}| > z) \leq \exp\{1 - (z/b_2)^{a_2}\}$, where a_1, a_2, b_1 and b_2 are positive constants.

Assumption C: Idiosyncratic error terms

(C1): Conditional on $\boldsymbol{x}_{it}$, $\boldsymbol{b}_{i,0}$, $\boldsymbol{f}_{t,g_i,0}$ and $\boldsymbol{\lambda}_{i,g_i,0}$, y_{it} is independently generated. Additionally, $\varepsilon_{it} = y_{it} - b'(\boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} + \boldsymbol{f}_{t,g_i,0}'\boldsymbol{\lambda}_{i,g_i,0})$ has zero mean $E[\varepsilon_{it}] = 0$ for all i and t.

(C2): For all i and t, ε_{it} are strongly mixing processes with mixing coefficients that satisfy $s(t) \leq \exp(-c_1 t^{d_1})$ with tail probability $P(|\varepsilon_{it}| > z) \leq \exp\{1 - (z/d_2)^{c_2}\}$, where c_1, c_2, d_1 and d_2 are positive constants.

Assumption D: Predictors and design matrix

(D1): For a positive constant, C, the predictors satisfy $\sup_{it} \|\boldsymbol{x}_{it}\| < C < \infty$.

(D2): For each i and all T, there exist positive constants C_1 and C_2 such that

$$0 < C_1 < \liminf_{T \to \infty} \lambda_{\min}(T^{-1}(X_i, F_{g_i^0,0})'(X_i, F_{g_i^0,0})),$$

where $\lambda_{\min}(A)$ is the smallest eigenvalue of a matrix A.

(D3): Define $A_i = \frac{1}{T} X_i' M_{F_{g_i}} X_i$, $B_i = (\boldsymbol{\lambda}_{i,g_i^0,0} \boldsymbol{\lambda}_{i,g_i^0,0}') \otimes I_T$, $C_i = \frac{1}{\sqrt{T}} \left[\boldsymbol{\lambda}_{i,g_i^0,0} \otimes (M_{F,g_i^0} X_i)\right]'$, and $M_{F_j} = I - F_j(F_j'F_j)^{-1}F_j'$. Let $\mathcal{A}$ be the collection of F_j such that $\mathcal{A} = \{F_j : F_j'F_j/T = I\}$. We assume

$$\inf_{F_j \in \mathcal{A}} \lambda_{\min} \left[\frac{1}{N_j} \sum_{i; g_i^0 = j} E_i(F_j)\right] > 0, \tag{12}$$

where $E_i(F_j) = B_i - C_i' A_i^{-1} C_i$.

(D.4): For N_j individuals belonging to the j-th group, let $\Upsilon(B_j)$ be an $N_j \times T$ matrix with its (i,t)th entry equal to $\boldsymbol{x}_{it}'\boldsymbol{b}_{i(j)}$, where $B_j = (\boldsymbol{b}_{1(j)}, \boldsymbol{b}_{2(j)}, \ldots, \boldsymbol{b}_{N_j(j)})'$ are the regression coefficients corresponding to individuals belonging to the j-th group. For any nonzero B_j, there exists a positive constant $c > 0$ such that, with probability approaching one,

$$\frac{1}{N_j T} \|M_{\Lambda_{j,0}} \Upsilon(B) M_{F_{j,0}}\|^2 \geq c \frac{1}{N_j} \sum_{i \in g_i = j} \|\boldsymbol{b}_i\|^2.$$

Assumption E: Number of units in each group

(E1): The true number of groups, S, is finite and independent of N and T.

(E2): All units are divided into a finite number of groups S, with each containing N_j units such that $0 < \underline{a} < N_j/N < \bar{a} < 1$; this implies that the number of units in the j-th group increases as the total number of units N grows.

Remark 2 Some comments regarding the assumptions are in order. Assumptions A and B1 imply the existence of r_j group-specific factors, $j = 1, ..., S$. To identify the true group membership, Assumption B2 is needed. Assumption B3 assumes that the unobservable factor structure is strongly mixing with a faster than polynomial decay rate and a restricted tail property. This condition is used to bound the misclassification probabilities. Assumption C1 assumes that the independence property of y_{it} is conditional on the factor structure. Assumption C2 restricts the tail property of the error. Assumption D1 requires some moment conditions for the explanatory variables. The explanatory variables can be correlated with the group-specific factors, factor loadings or both. Assumption D2 is necessary for the consistent estimation of regression coefficients even if the group-specific factors are known. Assumption D3 is used for proof of consistency when the factors and factor loadings are also estimated. D4 is used to ensure the consistency of $\boldsymbol{b}_i$. A similar condition was also used by Ando et al. (2021). Assumption E1 can be relaxed so that S also increases as $N, T \to \infty$. However, this is outside of the scope of this paper. Assumption E2 was also used by Ando and Bai (2017).

4.2 Asymptotic results

This section investigates some asymptotic properties of the proposed estimators. All proofs of the theorems described below are provided in the online supplementary document. The true parameter values $\{B_0, G_0, F_{1,0}, ..., F_{S,0}, \Lambda_{1,0}, ..., \Lambda_{S,0}\}$ are defined as the maximizer of the expected likelihood function $E[L(B, G, F_1, ..., F_S, \Lambda_1, ..., \Lambda_S)]$ subject to the identification restriction on F_j and Λ_j. Here, the expectation is taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N,\ t = 1, ..., T\}$, which is conditional on X, G_0, $F_{j,0}$ and $\Lambda_{j,0}$.

We note that all asymptotic results are obtained with $N, T \to \infty$. Our theoretical results are obtained by allowing the number of regression coefficient vectors N to diverge to infinity. The restrictions on the relative rates of convergence of N and T are specified below. We first claim that the estimated factors are consistent in the sense of an averaged norm.

Theorem 1 : Consistency. *Under Assumptions A–E, the estimator* $\{\hat{F}_j,\ j = 1, ..., S\}$ *is consistent in the sense of the following norm:*

$$T^{-1}\|\hat{F}_j - F_{j,0}\|^2 = o_p(1), \quad j = 1, ..., S, \tag{13}$$

This indicates that the space spanned by the common factor is estimated consistently. With this result, we argue that the estimated group membership is a consistent estimator of the true membership. The estimates of B, $\{F_j, \Lambda_j;\ j = 1, ..., S\}$, and G depend on each other. Following Ando and Bai (2016), we denote the estimator of group membership $\hat{g}_i$ as $\hat{g}_i(\hat{B}, \hat{F}, \hat{\Lambda})$, with $\hat{F} = \{\hat{F}_1, ..., \hat{F}_S\}$ and $\hat{\Lambda} = \{\hat{\Lambda}_1, ..., \hat{\Lambda}_S\}$. The next theorem claims that the estimated group memberships are consistent in the sense of an averaged norm.

Theorem 2 : Average consistency of the group membership estimator. *Suppose that the assumptions in Theorem 1 hold. Then,*

$$\frac{1}{N}\sum_{i=1}^{N} P\left(\hat{g}_i(\hat{B}, \hat{F}, \hat{\Lambda}) \neq g_i^0\right) = o(1).$$

Although Theorem 2 ensures that the group membership estimator is accurate in the average sense, it is ideal to show that the group membership estimator can accurately estimate the true membership across all individuals $i = 1, ..., N$. For all individuals $i = 1, ..., N$, due to endogeneity issues, it is also important to note that the regression coefficients can be estimated consistently under true group membership. The next theorem shows that the estimated group membership converges to the true group membership as T and N go to infinity.

Theorem 3 : Consistency of the estimator for group membership. *Suppose that the assumptions in Theorem 1 hold. Then, for all $\tau > 0$ and $T, N \to \infty$, we have*

$$P\left(\sup_{i \in \{1,\ldots,N\}} \left|\hat{g}_i(\hat{B}, \hat{F}, \hat{\Lambda}) - g_i^0\right| > 0\right) = o(1) + o(N/T^{\tau}).$$

We note that this uniform result holds across all individuals whose factor loadings are bounded away from zero. This is true from Assumption B. When $N/T^a \to 0$ for some $a > 0$, the true group membership g_i^0 and the estimator $\hat{g}_i(\hat{B}, \hat{F}, \hat{\Lambda})$ are asymptotically equivalent. Theorem 3 is similar to results obtained by Bonhomme and Manresa (2012) and Ando and Bai (2017) for 'linear" longitudinal models with unknown group memberships. The claim of our theorem 3 indicates that group membership consistency can be obtained even when the data are generated by an exponential family of distributions.

Next, we show that the asymptotic distribution of the estimated regression coefficients is a multivariate normal distribution. This finding is also true for the estimated factor loadings. We introduce $\boldsymbol{\gamma}_i = (\boldsymbol{b}_i', \boldsymbol{\lambda}_{i,g_i,0}')'$ and denote $\boldsymbol{\gamma}_{i,0} = (\boldsymbol{b}_{i,0}', \boldsymbol{\lambda}_{i,g_i^0,0}')'$ as its true value. The next theorem also states that the asymptotic distribution of the estimated common factors $\hat{\boldsymbol{f}}_{t,\tau}$ is a multivariate normal distribution.

Theorem 4 : Asymptotic distribution of $\boldsymbol{\gamma}_i$. *Under Assumptions A–E, $\sqrt{T}/N \to 0$ and $\sqrt{N}/T \to 0$, the asymptotic distribution of $T^{1/2}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0})$ is a multivariate normal with mean $\mathbf{0}$ and a covariance matrix of*

$$\Sigma_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} b''(\boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} + \boldsymbol{f}_{t,g_i^0}'\boldsymbol{\lambda}_{i,g_i^0,0})\boldsymbol{z}_{it,0}\boldsymbol{z}_{it,0}',$$

with $\boldsymbol{z}_{it,0} = (\boldsymbol{x}_{it}', \boldsymbol{f}_{t,g_i^0}')'$. Here $b''(\cdot)$ is the second derivate of the known function $b(\cdot)$ in (1). Additionally, the asymptotic distribution of $N^{1/2}(\hat{\boldsymbol{f}}_{t,j} - \boldsymbol{f}_{t,j,0})$ is normal with mean zero and a variance-covariance matrix of

$$\Psi_{t,j} = \lim_{N_j \to \infty} \frac{1}{N_j} \sum_{g_i^0 = j}^{N_j} b''(\boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} + \boldsymbol{f}_{t,g_i^0,0}'\boldsymbol{\lambda}_{i,g_i^0,0})\boldsymbol{\lambda}_{i,g_i^0,0}\boldsymbol{\lambda}_{i,g_i^0,0}'.$$

4.3 Testing the homogeneity of the slope

To test whether the regression coefficients within groups are homogeneous over individual units, we can consider a Swamy-type test. Let $\{\boldsymbol{b}_1, ..., \boldsymbol{b}_{N_j}\}$ be the regression coefficients of N_j individual units that belong to the j-th group. The null hypothesis of interest is

$$H_0 \ : \ \boldsymbol{b}_{1,0} = \boldsymbol{b}_{2,0} = \cdots = \boldsymbol{b}_{N_j,0} = \boldsymbol{b}_0^j \ \text{ for some } \boldsymbol{b}_0^j.$$

The alternative hypothesis is

$$H_1 \ : \ \boldsymbol{b}_{i,0} \neq \boldsymbol{b}_{0,k} \ \text{ for a nonzero fraction of pairwise slopes for } i \neq k.$$

Our Swamy's test statistic for the homogeneity of the regression coefficients within groups takes the form

$$\hat{\Xi}_j = \frac{T(\hat{\boldsymbol{s}}_j - \bar{\boldsymbol{s}}_j)' \left(\hat{\Gamma}_j - \frac{1}{N}\hat{L}_j'\right) \hat{\Omega}_j^{-1} \left(\hat{\Gamma}_j - \frac{1}{N}\hat{L}_j\right) (\hat{\boldsymbol{s}}_j - \bar{\boldsymbol{s}}_j) - N_j p}{\sqrt{2N_j p}}, \tag{14}$$

where $\hat{\boldsymbol{s}}' = (\hat{\boldsymbol{b}}_1', ..., \hat{\boldsymbol{b}}_{N_j}')$, $\bar{\boldsymbol{s}}' = (\bar{\boldsymbol{b}}_j', ..., \bar{\boldsymbol{b}}_j')$, $\bar{\boldsymbol{b}}_j = \sum_{i=1}^{N_j} \hat{\boldsymbol{b}}_i / N_j$, $\hat{\Gamma}_j$ is an $N_j p \times N_j p$ block diagonal matrix with the (i,i)th block being $T^{-1}\sum_{t=1}^{T} b''(\boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_{i,0} + \hat{\boldsymbol{f}}_{t,j}'\hat{\boldsymbol{\lambda}}_{i,j})\hat{\boldsymbol{z}}_{it}\hat{\boldsymbol{z}}_{it}'$ and $\hat{\boldsymbol{z}}_{it} = (\boldsymbol{x}_{it}', \hat{\boldsymbol{f}}_{t,j}')'$, $\hat{L}_j$ is an $N_j p \times N_j p$ block diagonal matrix with the (i,i)-th block being $T^{-1}\sum_{t=1}^{T} J_{it}\Psi_t^{-1}J_{it}'$ and $\Psi_t = N_j^{-1}\sum_{i}^{N_j} b''(\boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_{i,0} + \hat{\boldsymbol{f}}_{t,j}'\hat{\boldsymbol{\lambda}}_{i,j})\hat{\boldsymbol{\lambda}}_{i,j}\hat{\boldsymbol{\lambda}}_{i,j}'$, $J_{it} = b''(\boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_{i,0} + \hat{\boldsymbol{f}}_{t,j}'\hat{\boldsymbol{\lambda}}_{i,j})\boldsymbol{x}_{it}\hat{\boldsymbol{\lambda}}_{i,j}'$, and $\hat{\Omega}_j$ is the variance–covariance estimator of $\Omega_j \equiv E[\text{diag}\{\boldsymbol{\varepsilon}_1', ..., \boldsymbol{\varepsilon}_{N_j}'\}]$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})$.

Before we investigate a new hypothesis testing procedure, we state the assumptions needed for the asymptotic analysis.

Assumption F: Central limit theorem

As T goes to infinity,

$$\Omega_i^{-1/2}\frac{1}{\sqrt{T}}X_i'\boldsymbol{\varepsilon}_i \to_d N(\boldsymbol{0}, I_p),$$

where $\Omega_i = E[X_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'X_i]/T$.

Assumption G: Joint limit

Let $\boldsymbol{\zeta}_i = -\frac{1}{\sqrt{T}} X_i' \boldsymbol{\varepsilon}_i$, $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1', \boldsymbol{\zeta}_2', ..., \zeta_{N_j}')'$, and $\Omega = \text{block- diag}(\Omega_1, \Omega_2, ..., \Omega_{N_j})$, where Ω_i is defined in Assumption F. We assume, as $N, T \to \infty$,

$$\frac{\boldsymbol{\zeta}'\Omega^{-1}\boldsymbol{\zeta} - N_j p}{\sqrt{2N_j p}} \to N(0,1).$$

Remark 3 The variables in Assumption F do not involve any estimated quantities; they are random variables from the true model. It follows that the conditional mean of $\frac{1}{\sqrt{T}} X_i' M_{F^0} \boldsymbol{\varepsilon}_i$, conditional on (X_i, F^0), is equal to 0, and the conditional and unconditional variance of $\Omega_i^{-1/2} \frac{1}{\sqrt{T}} X_i' M_{F^0} \boldsymbol{\varepsilon}_i$ is an identity matrix. Assumption G imposes the central limit theorem for $\boldsymbol{\zeta}_i$. Moreover, the expected value of $\boldsymbol{\zeta}_i' \Omega_i^{-1} \boldsymbol{\zeta}_i$ is equal to p.

The following theorem provides the asymptotic distribution of $\hat{\Xi}_j$ under the null hypothesis.

Theorem 5 : Swamy-type test *Suppose that assumptions in Theorem 4, Assumption F and Assumption G hold. Then, under the null hypothesis H_0,*

$$\hat{\Xi}_j \to N(0,1) \quad \text{in distribution}$$

as $T, N \to \infty$.

Theorem 5 suggests that our proposed test statistic $\hat{\Xi}$ is asymptotically normal with mean 0 and standard deviation 1 when the null hypothesis of slope homogeneity is satisfied. Therefore, our proposed test is simple to implement because it has a limiting $N(0,1)$ distribution.

5 Application

5.1 NYC taxi data and an empirical model

The trip record data for yellow cab taxis are made publicly available by the NYC Taxi & Limousine Commission. We use a dataset made available by Brian and Dan (2016), spanning from 1st January 2010 through 31st December 2013. This dataset allows us to analyze individual taxi's trip records. For each trip, this dataset includes the yellow cab identifier medallion, precise location coordinates for where the trip started and ended and timestamps for when the trip started and ended. Figure 1 and Figure 2 provide a map of the Manhattan area in New York City and an intensity of taxi pick-up and drop-off locations in January 2010. Similar patterns were observed for the other months and years (See Appendix H in the online supplement document. It contains monthly plots for other months in 2010).

Because the pricing structure is strictly regulated by authorities, it is important for individual taxis to utilize their vehicles efficiently. To understand the efficiency, one of the measurements is the capacity utilization rate, which is measured by the fraction of time that a driver has a fare-paying passenger. For the i-th taxi, we let y_{it} be the capacity utilization rate at time t. More specifically, the capacity utilization rate is defined as the fraction of time that a driver has a fare-paying passenger every hour, i.e., $y_{it}/60$, where y_{it} is the total number of minutes per hour during which a cab drives a fare-paying passenger. Figure 3 shows the average capacity utilization rate at the time for a particular day of the week and time frame in 2010. Similar patterns were observed for the other years (See Appendix H in the online supplement document. It contains similar plots for 2011, 2012 and 2013). We can see that the capacity utilization rate varies over the time frame. On average, the capacity utilization rate at midnight is lower than that during the daytime on weekdays. In contrast, the capacity utilization rate at midnight during the weekend is relatively high.

To analyze the capacity utilization rate, we fit the following model:

$$\begin{aligned} & f(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}) \\ &= \binom{60}{y_{it}} \pi(\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i})^{y_{it}} \{1 - \pi(\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i})\}^{60-y_{it}}, \end{aligned} \tag{15}$$

where y_{it} is the fraction of the total number of minutes per hour during which a taxi is driving a fare-paying passenger for the time period t, and $\binom{60}{y_{it}}$ is the binomial coefficient. In our analysis, we employ the following information for $\boldsymbol{x}_{it}$: Time frame (every hour: MIDNIGHT–1 AM, 1 AM–2 AM, ..., 10 PM–11 PM, 11 PM–MIDNIGHT), Day of the week (Monday, Tuesday, ..., Friday, Saturday, Sunday). By combining "time frame" and "day of the week", we create a set of $168(= 24 \times 7)$ indicator variables. Because the regressors are common over individuals, $\pi(\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i})$ in (15) becomes

$$\pi(\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}) \equiv \pi(\boldsymbol{x}_t, \boldsymbol{b}_i, \boldsymbol{f}_{t,g_i}, \boldsymbol{\lambda}_{i,g_i}) = \frac{\exp\{\boldsymbol{x}_t'\boldsymbol{b}_i + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i}\}}{1 + \exp\{\boldsymbol{x}_t'\boldsymbol{b}_i + \boldsymbol{f}_{t,g_i}'\boldsymbol{\lambda}_{i,g_i}\}},$$

where $\boldsymbol{x}_t$ is the vector of indicator variables. To avoid multicollinearity, the intercept term is not included in $\boldsymbol{x}_t$.

5.2 Estimation results

To understand how the market structure evolved over time, we analyze the set of 48 longitudinal data. Here, 48 denotes 12 months $\times$ 4 years (2010, 2011, 2012 and 2013). For example, we analyze the longitudinal data with size $(T, N) = (744, 13341)$ for January 1st through January 31st in 2010. Here, $T = 744$ implies that the data period spans 31 days (24 hours/per day $\times$ 31 days), and there are $N = 13341$ individual medallion taxis. A similar operation is performed for each month in 2011, 2012 and 2013.

The formulated model (15) is estimated by maximizing the objective function (5). To determine the number of groups S and the number of group-specific common factors, we apply the information criterion in (8). The maximum number of groups is set to $S_{\max} = 5$ when we search for the best value based on the information criterion in (8).

Table 1 summarizes the estimated number of groups $\hat{S}$ for the set of 48 longitudinal data. Here, $\hat{r}_j$ is the selected number of common factors for the j-th group. We can see that the estimated number of groups $\hat{S}$ is two in almost all months, while there are some variations. Figure 4 shows the Sankey diagram, which represents how the set of individual taxis form groups from one month to the next in 2010 and 2013. If the market structure is stable, group membership should be stable over a 12-month period. However, the Sankey diagram indicates that the membership has been shuffled every month. Similar observations are obtained for 2011 and 2012. (See Appendix H in the online supplement document, which contains monthly plots for 2011 and 2012).

Table 2 summarizes the total sum of the number of common factors $\sum_{j=1}^{\hat{S}} \hat{r}_j$ for the set of 48 longitudinal data. From Table 2, we see that there are some unobservable variables that explain the performance. The model structure implies that such unobservable structures are a source of variation in group membership changes. The collection of all possible explanatory variables that may influence the response variables is sometimes costly due to restrictions on time, budget and so on. Our method, which automatically captures unobserved heterogeneity through the factor structure, is very useful because it can save costs for such practical issues.

Finally, we implement the proposed homogeneity test for the regression coefficients. According to our proposed test, for each group and each month from 2010 to 2013, the null hypothesis (the regression coefficients within groups are homogeneous across individual taxis) is rejected at the 5% level. This indicates that a model with some regression coefficients that are heterogeneous across groups but homogeneous within a group is not sufficient to capture the heterogeneity in the data. Thus, our proposed model and methods are very important because the group-specific homogeneous regression coefficients will lead to biased results.

6 Concluding remarks

The proposed method is a flexible approach for capturing grouped patterns of unobserved heterogeneity in nonlinear longitudinal data models. To handle the endogeneity issue, we developed a new estimation procedure with which the regression parameters, unobservable factor structure, and group membership were all estimated jointly. The consistency and asymptotic normality of the estimated regression coefficients were established even in the presence of unknown group memberships. We also examined the problem of testing the homogeneity of regression coefficients and developed a new testing procedure based on Swamy's (1970) test principle. If the test implies that the regression coefficients are homogenous within a group, the proposed model (4) with group-specific regression coefficients can be applied. Nevertheless, we showed that group membership can be estimated consistently.

The proposed procedure was applied to the NYC taxi dataset and revealed useful information for taxi firm management. Although we analyzed the NYC taxi dataset, our method can be applied to various datasets, including those in economics, finance, marketing, political science, medicine, the natural sciences and other areas.

Finally, advancements in information technology may create a situation where the number of explanatory variables is large. To accommodate this situation, our proposed method can be extended to address cases with large numbers of explanatory variables. One of the most commonly used ideas is to use shrinkage methods that simply add the regularization term into the log-likelihood function (5). For the penalty function, one can consider the lasso penalty (Tibshirani, 1996) and its variants (Zou, 2006; Yuan and Lin, 2006), the elastic net penalty (Zou and Hastie, 2005), the minimax concave penalty (Zhang 2010), the SCAD penalty of Fan and Li (2001) and Fan and Peng (2004), and so on. We would like to investigate this topic in a future study.

SUPPLEMENTARY MATERIAL

To save space, all technical proofs are provided in the online supplementary document, which also contains Monte Carlo simulation results. It also contains additional figures in the empirical analysis.

References

Ahlquist, J.S. and Breunig, C. (2012), "Model-based Clustering and Typologies in the Social Sciences," *Political Analysis*, **20**, 92-112.

Ando, T. and Bai, J. (2015), "A simple new test for slope homogeneity in panel data models with interactive effects," *Economics Letters*, **136**, 112–117.

Ando, T. and Bai, J. (2016), "Panel data models with grouped factor structures under unknown group membership," *Journal of Applied Econometrics*, **13**, 163–191.

Ando, T. and Bai, J. (2017), “Clustering huge number of time series: A panel data approach with high-dimensional predictors and factor structures,” *Journal of the American Statistical Association*, **112**, 1182–1198.

Ando, T., Bai, J. and Li, K. (2021), “Bayesian and maximum likelihood analysis of large scale panel choice models with unobserved heterogeneity,” *Journal of Econometrics*, forthcoming.

Bai, J. (2009), “Panel data models with interactive fixed effects,” *Econometrica*, **77**, 1229–1279.

Bai, J. and Ng, S. (2013), “Principal components estimation and identification of static factors,” *Journal of Econometrics*, **176**, 18–29.

Boneva, L. and O, Linton. (2017), “A discrete choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance,” *Journal of Applied Econometrics*, **32**, 1226–1243.

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

Brian, D. and Dan, W. (2016), New York City Taxi Trip Data (2010-2013). University of Illinois at Urbana-Champaign.

Blomquist, J. and Westerlund, J. (2013), “Testing slope homogeneity in large panels with serial correlation,” *Economics Letters*, **121**, 374–378

Cai, T., Ma, J. and Zhang, L. (2019), “CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality,” *Annals of Statistics*, **47**, 1234–1267.

Charbonneau, K. (2017), “Multiple Fixed Effects in Binary Response Panel Data Models,” *Econometrics Journal*, **20**, S1–S13.

Chen, M. Fernández-Val, I. and Weidner, M. (2021), “Nonlinear Panel Models with Interactive Effects,” *Journal of Econometrics*, **220**, 296–324.

Chiou, J.-M. and Li, P/-L. (2007), “Functional clustering and identifying substructures of longitudinal data,” *Journal of the Royal Statistical Society*, B**69**, 679–699

Delaigle, A., Hall, P. and Bathia, N. (2012), “Componentwise classification and clustering of functional data," *Biometrika*, **99**, 299313.

Fan, J. and Li, R. (2001), "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, **96**, 13481360.

Fan, J. and Peng, H. (2004), "Nonconcave penalized likelihood with a diverging number of parameters," *Annals of Statistics*, **32**, 928–961.

Fernández-Val, I. and Weidner, M. (2016), "Individual and time effects in nonlinear panel data models with large N, T," *Journal of Econometrics*, **192**, 291–312.

France, S.L. and Ghose, S. (2016), "An Analysis and Visualization Methodology for Identifying and Testing Market Structure," *Marketing Science*, **35**, 1-200.

Handcock, M., Raftery, A.E. and Jeremy, T. (2007), "Model-Based Clustering for Social Networks," *Journal of the Royal Statistical Society*, A**170**, 301–354.

Heard, N.A., Holmes, C.C. and Stephens, D.A. (2006), "A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: an application of Bayesian hierarchical clustering of curves," *Journal of the American Statistical Association*, **101**, 18-29.

Hennig, C. Meila, M. Murtagh, F. and Rocci, R. (2015), *Handbook of Cluster Analysis*. Chapman and Hall/CRC Press.

James, G.M. and Sugar, C.A. (2003), "Clustering for sparsely sampled functional data," *Journal of the American Statistical Association*, **98**, 397-408.

Kong, N., Schaefer, A.J., Hunsaker, B. and Roberts, M. S. (2010), "Maximizing the Efficiency of the U.S. Liver Allocation System Through Region Design," *Management Science*, **56**, 2111–2122.

Lin, C.-C. and Ng, S. (2012), "Estimation of panel data models with parameter heterogeneity when group membership is unknown," *Journal of Econometric Methods*, **1**, 42–55.

Lumsdaine, R.L., Okui, R. and Wang, W. (2020), "Estimation of panel group structure models with structural breaks in group memberships and coefficients," Working Paper.

Moon, H., Shum, M. and Weidner, M. (2018), "Estimation of random coefficients logit demand models with interactive fixed effects," *Journal of Econometrics*, **206**, 613–644.

Park, C.H. and Park, Y.-H. (2016), "Investigating Purchase Conversion by Uncovering Online Visit Patterns," *Marketing Science*, **35**, 894-914.

Peng, J. and Muller, H. G. (2008), "Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions," *Annals of Applied Statistics*, **2**, 1056-1077.

Pesaran, M. H. (2006), "Estimation and inference in large heterogeneous panels with a multi-factor error structure," *Econometrica*, **74**, 967–1012.

Pesaran, M. H. and Yamagata, T. (2008), "Testing slope homogeneity in large panels," *Journal of Econometrics*, **142**, 50–93.

Pesaran, H., Smith, R. and Im, K. S. (1996), Dynamic linear models for heterogenous panels. In: Matyas, L. & Sevestre, P. (Eds.), *Econometrics of Panel Data: A Handbook of the Theory with Applications*, second edition. Kluwer Academic Publishers, Dordrecht, 145–195.

Phillips, P. C. B. and Sul, D. (2003), "Dynamic panel estimation and homogeneity testing under cross section dependence," *Econometrics Journal* 6, 217–259.

Qin, Z.S. (2006), "Clustering microarray gene expression data using weighted Chinese restaurant process," *Bioinformatics*, **22**, 1988-1997.

Su, L., Shi, Z. and Phillips, P. (2016), "Identifying latent structures in panel data," *Econometrica*, **84**, 2215–2264.

Swamy, P. A. V. B. (1970), "Efficient inference in a random coefficient regression model," *Econometrica*, **38**, 311–323.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, B58, 267–288.

Vogt, M. and Linton, O. (2017), "Classification of nonparametric regression functions in heterogeneous panels," *Journal of the Royal Statistical Society*, **B79**, 5–27.

Wang, W., Phillips, P. C. B. and Su, L. (2018), "Homogeneity pursuit in panel data models:

theory and applications," *Journal of Applied Econometrics*, **33**, 797–815.

Wang, W. and Su, L. (2021), "Identifying latent group structures in nonlinear panels," *Journal of Econometrics*, **220**, 272–295.

Yuan, M. and Lin, Y. (2006), "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society*, B68, 49–67

Zhang, C. H. (2010), "Nearly unbiased variable selection under minimax concave penalty," *Annals of Statistics*, 38, 894–942.

Zhang, Y. Wang, H.J. and Zhu, Z. (2019), "Quantile-regression-based clustering for panel data," *Journal of Econometrics*, 213, 54–67

Zou, H. (2006), "The adaptive Lasso and its oracle properties," *Journal of the American Statistical Association*, 101, 1418–1429.

Zou, H. and Hastie, T. (2005), "Regularization and variable selection via the Elastic Net," *Journal of the Royal Statistical Society* B67, 301–320.

Table 1: The estimated number of groups $\hat{S}$ for the set of 48 longitudinal data.

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
2010	2	2	2	2	4	3	2	2	2	2	2	2
2011	2	4	2	3	2	4	2	2	2	2	2	2
2012	2	2	4	2	2	2	2	4	2	2	2	3
2013	2	2	2	3	3	3	2	2	2	3	2	4

Table 2: The estimated number of common factors $\sum_{j=1}^{\hat{S}} \hat{r}_j$ for the set of 48 longitudinal data.

	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
2010	6	10	10	8	20	14	8	9	6	7	8	6
2011	9	20	9	15	10	20	8	6	9	6	9	7
2012	9	9	20	9	6	9	6	20	8	9	9	15
2013	10	9	7	15	15	15	10	9	8	14	9	20

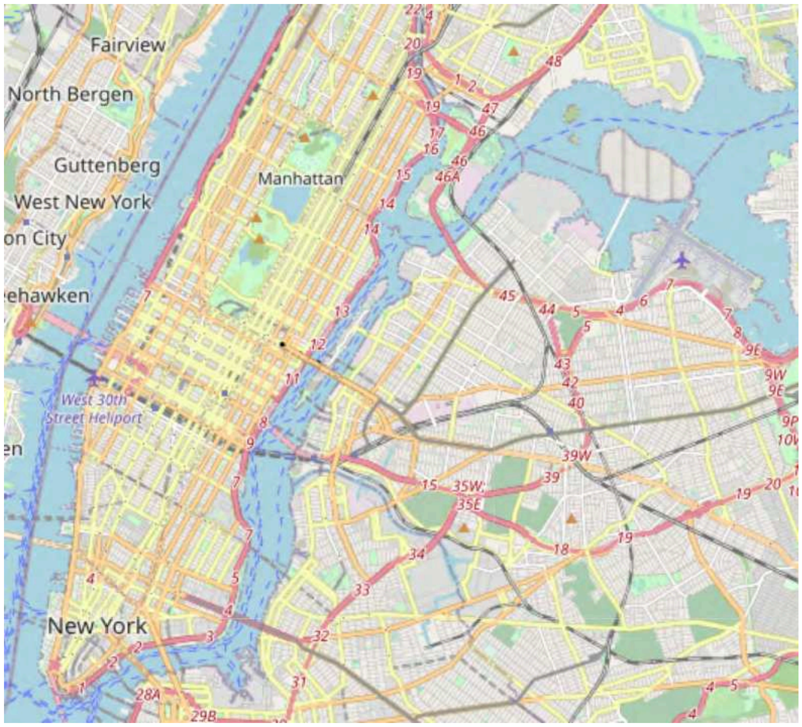
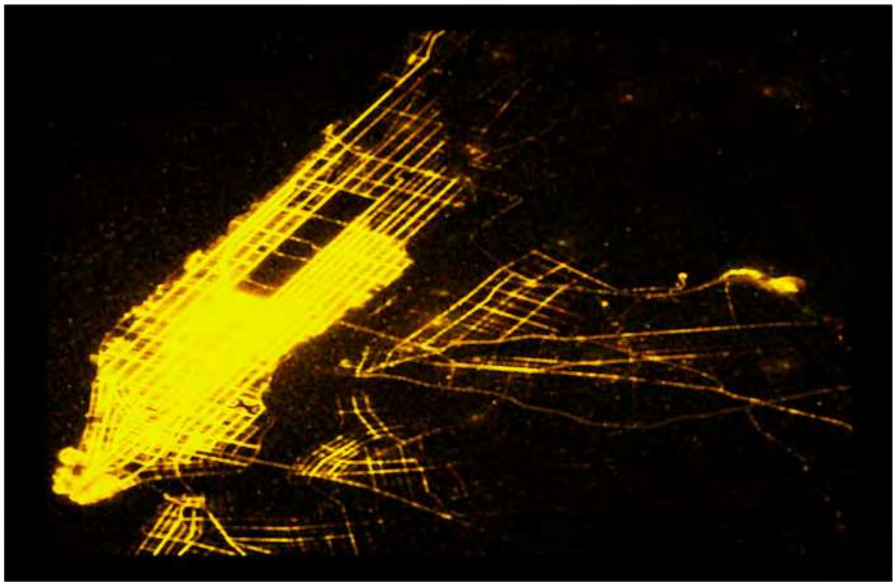
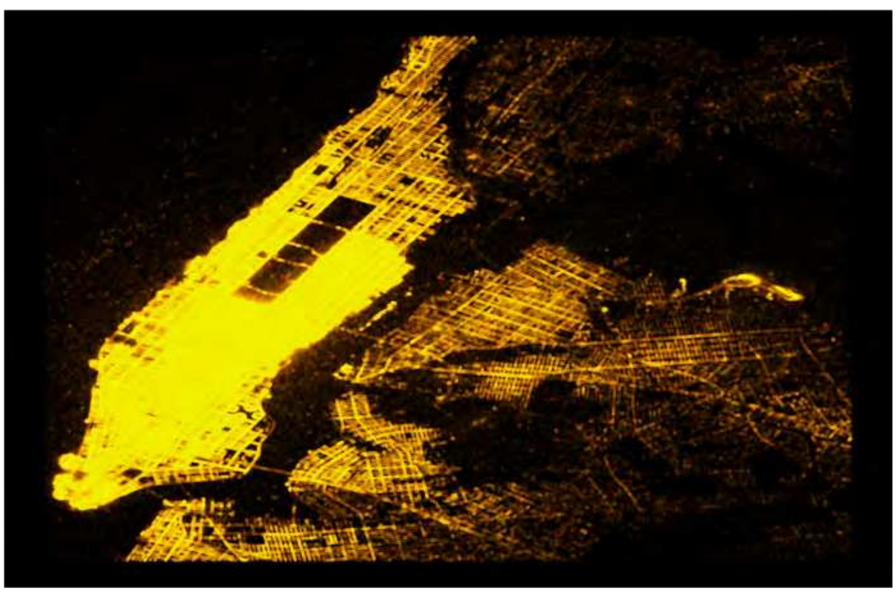


Figure 1: Map of the Manhattan area in New York City. This map was created by using the R package leaflet.



(a) Pick-ups



(b) Drop-offs

Figure 2: Every (a) pick-up and (b) drop-off location by a taxi in New York City from January 1 2010 to January 31 2010. Brighter regions (i.e., closer to yellow) indicate more taxi activity.

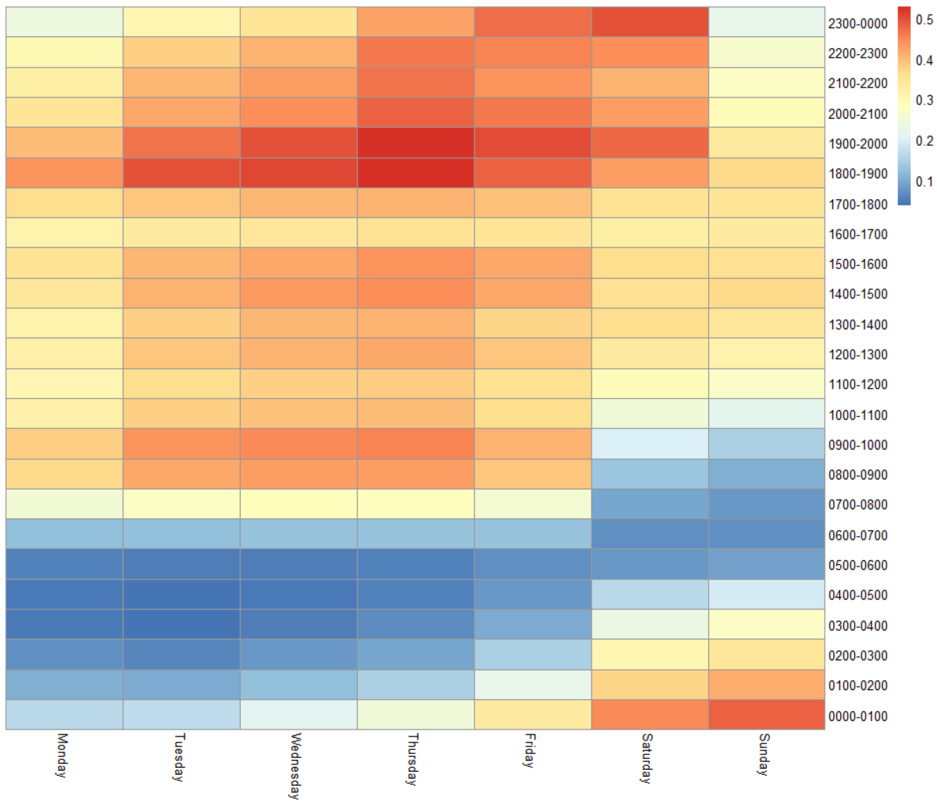
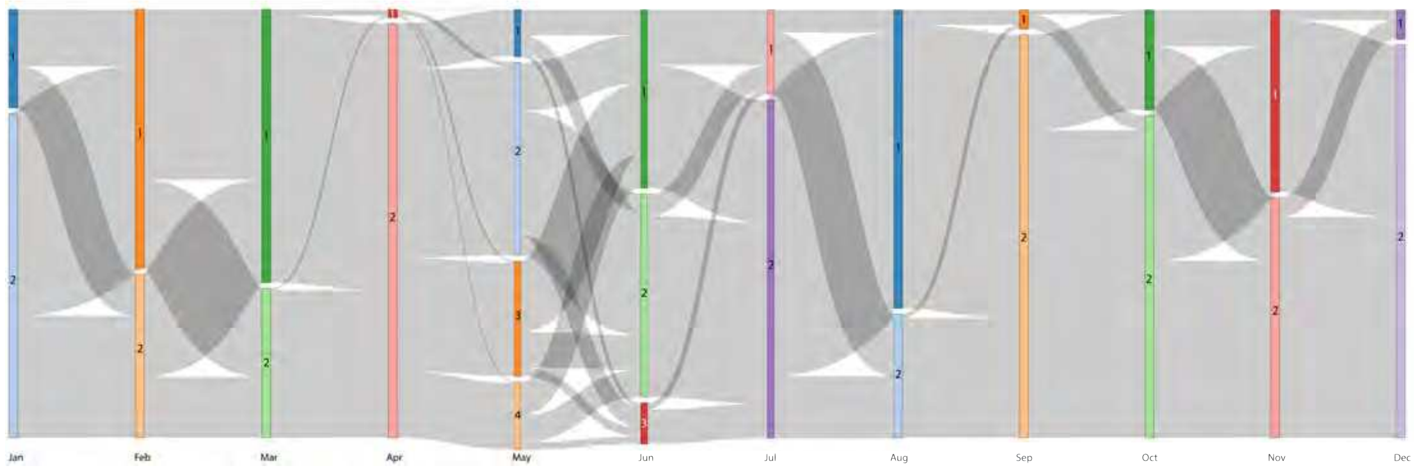
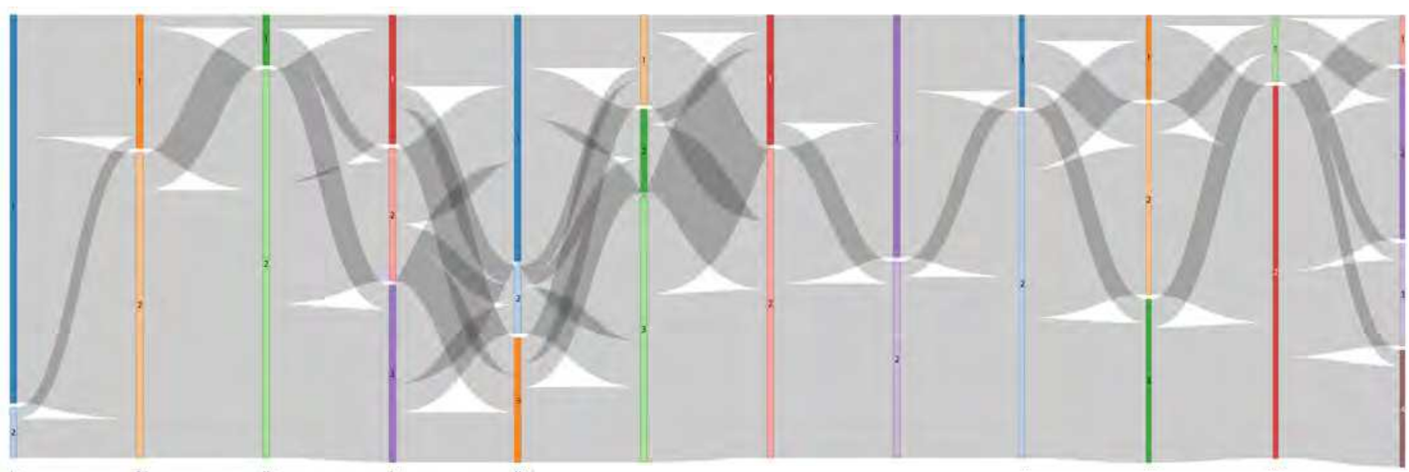


Figure 3: Heatmap of the average capacity utilization rates in 2010.



2010



2013

Figure 4: A Sankey diagram of the grouping results during 2010 and 2013. The most left and most right groups in the figures are formed groups in January and December, respectively.