Large scale panel choice model with unobserved heterogeneity

Tomohiro Ando [1] and Jushan Bai [2]

Summary

This paper considers the estimation and inference of a logistic panel regression model with interactive fixed effects, where multiple individual effects are allowed and the model is capable of capturing high-dimensional cross-section dependence. The proposed model also allows for heterogeneous regression coefficients. We propose maximum likelihood estimation via data augmentation. We investigate the asymptotic property of the estimated parameters. When both the cross section and time series dimensions of the panel go to infinity, we show the consistency and the asymptotic normality of the estimated regression coefficients and the estimated interactive fixed effects. An information criterion based on the likelihood function is proposed to estimate the dimension of the interactive fixed effects. It is shown that the criterion asymptotically selects the true dimension. Monte Carlo simulation documents the satisfactory performance of the proposed method. Finally, the method is applied to study the New York City medallion drivers' efficiency performance.

Keywords: Cross-sectional and series dependence; Endogeneity; Factor analysis; Heterogeneous panel; Nonlinear panel data.

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

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

1 Introduction

This paper studies panel logistic regression models with interactive fixed effects to analyze choices made by individuals facing a set of alternatives. We consider the maximum likelihood estimation with a new data augmentation algorithm. We also study the asymptotic properties of the maximum likelihood estimator, establishing consistency and asymptotic normality. The model allows for heterogeneous coefficients and multiple individual effects via a factor error structure. We also propose a new information criterion to determine the dimension of the interactive effects.

In the Bayesian statistics literature on cross-sectional logistic regression models (not panel data), a data-augmentation strategy is often employed (e.g., Polson and Scott (2013), Holmes and Held (2006)). This paper extends these studies to panel models with endogeneity, where the regressors are correlated with the unobserved interactive effects. Modeling this endogeneity is important to ensure asymptotically consistent estimation. To our knowledge, this is the first study to investigate a data-augmentation approach to the analysis of panel logistic regression models with interactive effects.

The analysis of consistency is nontrivial because the dimension of the parameter space grows with the dimensions of the panel data. We first establish an average consistency and then individual parameter consistency. Asymptotic normality is also obtained. Also to the best of our knowledge, this is the first study that rigorously develops these results for the panel logistic regression models with interactive fixed effects under increasing dimensions of the panel.

Several related studies include Chen et al (2014), Fernández-Val and Weidner (2016), Sun (2016), Moon et al. (2016), Charbonneau (2017) and Boneva and Linton (2017). Fernández-Val and Weidner (2016) and Sun (2016) studied nonlinear panel data models with individual and time effects. Charbonneau (2017) also studied nonlinear panel data models with additive effects and homogeneous regression coefficients. These studies are thus not about the nonlinear panel data models with the interactive fixed effects. Recently, Chen et al. (2014) also consider inference in panel logistic regression models with predetermined explanatory variables and interactive effects. They consider the interactive fixed effects with a single factor and impose "homogeneous" regression coefficients. Our paper allows multiple factors and heterogenous regression coefficients. Also, Chen et al. (2014) assumes the number of factors is known and being one. We propose the model selection criterion for determining the number of factors.

Boneva and Linton (2017) studied panel probit model with interactive effects. They proposed an estimator belonging to the class of common correlated effects estimators

(Pesaran (2006)). While their estimator is simple to compute, as Boneva and Linton (2017) pointed out, their approach is valid only if the unobserved factors are contained in the span of the observed factors and the cross-sectional averages of the regressors. In contrast to their approach, we permit the unobserved factors are outside of the span of the observed factors and the cross-sectional averages of the regressors. This is one desirable property of our proposed method. Also, we further consider inference in multinomial panel choice models, while Boneva and Linton (2017) studied a binary choice. Furthermore, we show how our idea can be extended to a direct inference approach for the panel probit model with interactive effects. This approach can be implemented without assuming that the unobserved factors are contained in the span of the observed factors and the cross-sectional averages of the regressors.

In summary, our contributions are as follows. First, we introduce heterogeneous panel logistic model with interactive effects. Second, a new parameter estimation procedure and model selection criterion are developed. Third, the consistency and the asymptotic distribution of our estimator are established. Fourth, the proposed model selection criterion is shown to consistently detect the true dimension of interactive effects.

Discrete choice models are widely used in social science studies. The proposed model will have wide range of potential applications, for example, the inference of market structures in marketing research (Elrod and Keane (1995)), the analysis of partisanship patterns of roll-call votes from the United States Senate (Hahn et al. (2012)), the association studies based on high-throughput single nucleotide polymorphism (SNP) data in genomic DNA study (Lee et al. (2010)), the study of firms' decisions to split their shares (Perez et al. (2015)), and so forth.

The paper is organized as follows. Section 2 introduces the panel logistic regression model with interactive fixed effects. In Section 3, we introduce the estimation and model selection method. Section 4 investigates some asymptotic properties of the proposed method. To save the space, all technical proofs are provided in the online supplementary document. Section 5 contains Monte Carlo simulation results. In Section 6, the proposed method is applied to taxi's capacity utilization data. Section 7 discuss the extension of our proposed procedures. Section 8 concludes.

2 Panel logistic regression model with interactive fixed effects

Suppose that there are $i = 1, ..., N$ individuals, facing binary choices. At time t, each individual chooses one of the alternatives, labeled alternative 1 and alternative 0. We

consider the random utility (the difference in utilities between alternative 1 and alternative 0) associated with the choice for individual i at time t:

$$u_{it} = \boldsymbol{x}_{it}'\boldsymbol{b}_i + \eta_{it} + \varepsilon_{it}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T, \tag{1}$$

where $\boldsymbol{x}_{it}$ is a p_i-dimensional vector of observed attributes of the alternatives or the observed individual characteristics; η_{it} denotes the unobserved structure of individual i's utility, which can vary across t; and ε_{it} denotes the non-modeled component of utility (or shocks to preference). Alternative 1 is chosen if and only if $u_{it} > 0$ (the corresponding utility is higher).

As one of the novelties of this paper, we focus on the case in which the unobserved structure η_{it} is modeled with a factor structure:

$$\eta_{it} = \sum_{\ell=1}^{r} f_{\ell t}\lambda_{i\ell} = \boldsymbol{f}_t'\boldsymbol{\lambda}_i, \tag{2}$$

where $\boldsymbol{f}_t$ is an $r \times 1$ vector of unobservable factors and $\boldsymbol{\lambda}_i$ represents the factor loadings. This is known as the interactive effect in the econometric literature (e.g., Bai, 2009). Note that the interactive effects are more general than conventional additive effects. To see this, suppose that there are two factors ($r = 2$), and consider the special factor $f_t = (1, \delta_t)'$ and the special loading $\lambda_i = (\alpha_i, 1)'$. Then $f_t'\lambda_i = \alpha_i + \delta_t$, reducing to the standard individual effect and time effect model (additive effects). In additive effects models, the influence of individual effects (α_i) is constant over time, and the influence of time effects (δ_t) is identical across individuals. In contrast, the interactive effects allow the unobserved individual characteristics (λ_i) to have time-varying effects (through f_t). Another interpretation of the interactive effects is that they allow a vector of common shocks or social trends (f_t) to impact individuals in a heterogeneous way (through λ_i).

An important feature of the model is that correlations between the unobserved factor structure η_{it} and the regressors $\boldsymbol{x}_{it}$ (endogeneity) are allowed, while the standard logistic regression model does not permit such a situation. This correlation arises because some of the explanatory variables are themselves decision variables, which are correlated with the unobserved individual effects. This endogeneity problem is common in economics and other social sciences. Ignoring endogeneity, if it exists, will cause inconsistent estimation of the traditional maximum likelihood estimator. Conventional panel data analysis often assumes cross-sectional independence. Interactive effects models provide a way of modeling cross-sectional dependence because individuals share the same common shocks f_t. These models are effective in modeling high-dimensional cross-sectional dependence.

Let $y_{it} \in \{0, 1\}$ denote the observed choice outcome, taking value 1 if alternative 1 is chosen, 0 otherwise. Alternative 1 will be chosen if and only of $u_{it} > 0$. For the logistic

specification of the idiosyncratic shock ε_{it}, the conditional probability of such a choice is given by

$$P(y_{it}=1|\boldsymbol{x}_{it},\boldsymbol{b}_i,\boldsymbol{f}_t,\boldsymbol{\lambda}_i)=\frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i+\boldsymbol{f}_t'\boldsymbol{\lambda}_i)}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i+\boldsymbol{f}_t'\boldsymbol{\lambda}_i)}. \tag{3}$$

Assuming that the errors ε_{it} are independently and identically distributed, the joint probability of observing the choices $Y\equiv\{y_{it}|i=1,...,N,t=1,...,T\}$, $L(Y|X,B,F,\Lambda)$, is

$$L(Y|X,B,F,\Lambda)=\prod_{i=1}^{N}\prod_{t=1}^{T}\left[\frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i+\boldsymbol{f}_t'\boldsymbol{\lambda}_i)}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i+\boldsymbol{f}_t'\boldsymbol{\lambda}_i)}\right]^{y_{it}}\left[\frac{1}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i+\boldsymbol{f}_t'\boldsymbol{\lambda}_i)}\right]^{1-y_{it}}, \tag{4}$$

where $X\equiv\{\boldsymbol{x}_{it}|i=1,...,N,t=1,...,T\}$, $\Lambda=(\boldsymbol{\lambda}_1,...,\boldsymbol{\lambda}_N)'$, $B=(\boldsymbol{b}_1,...,\boldsymbol{b}_N)'$ and $F=(\boldsymbol{f}_1,...,\boldsymbol{f}_T)'$.

Remark 1 Recently, there is an increasing literature on panel data models with unobserved factor structures where the dimensions of both cross section and time series of the panel are large (Ando and Bai (2015, 2016, 2017a), Bai (2009), Bai and Li (2014), Chudik and Pesaran (2015), Kapetanios et al. (2011), Moon and Weidner (2015), Pesaran (2006), Pesaran and Tosetti (2011), Song (2013) among others). Although a number of studies exist on the linear panel data regression models with interactive effects, studies on nonlinear panel models with unobserved factor structures are scant.

In Section 1, we discussed Chen et al (2014), Fernández-Val and Weidner (2016), Sun (2016), Moon et al. (2016), Charbonneau (2017) and Boneva and Linton (2017). Some other related studies are Li and Ansari (2014), Naik et al. (2010) and Ebbes et al. (2005). Li and Ansari (2014) also considered latent factor structures for choice modeling. In (1), we model the interactive effects η_{it}, which is unobservable, depends the individuals i and time t. In contract, Li and Ansari (2014) modeled the unobserved structure that depends on unobservable choice attributes and time t. Letting the interactive effects η_{it} in (1) being common over i, our model then reduces to their unobserved structure. Naik et al. (2010) considered the multi-index binary response (MBR) model. They pointed out that their model is related to approximate factor models (e.g., Bai and Ng (2002)) based on the fact that both involve linear combinations of the predictors. As well as their estimation procedure, our model estimation procedure explicitly incorporates the information in a choice data in the process of constructing common factors $\boldsymbol{f}_t$ and loadings $\boldsymbol{\lambda}_i$. In contrast to the MBR model, however, our interactive effects η_{it} is not restricted only on the space of linear combinations of the predictors. Thus, our model nests the model formulations of Li and Ansari (2014) and Naik et al. (2010). In the context of the linear instrumental variable regression where endogeneity is likely to be present, Ebbes et al. (2005) introduced the

latent instrumental variables method. Similar to Ebbes et al. (2005), our method does not rely on observable instruments. This can be achieved by taking account for dependencies between the predictors $\boldsymbol{x}_{it}$ and the interactive effects η_{it}.

In the next section, we introduce a new parameter estimation procedure. Then, we introduce a new information criterion for determining the dimension of interactive effects.

3 Estimation and model selection

3.1 Data-augmentation approach for parameter inference

Inference on the panel logistic regression with interactive effects is a challenging problem due to the analytically inconvenient form of the model structure. If the interactive effects can be ignored, the regression coefficients can be obtained by maximizing

$$L(Y|X,B) = \prod_{i=1}^{N}\prod_{t=1}^{T}\left[\frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i)}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i)}\right]^{y_{it}}\left[\frac{1}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i)}\right]^{1-y_{it}}, \tag{5}$$

However, this approach ignores the issue of "endogeneity", and the estimated regression coefficient is not consistent estimator.

Also, we need to impose the identification condition on the factors structure. Following Bai and Ng (2013), we consider a restriction $F'F/T = I_r$ and $\Lambda' = (\Lambda_1', \Lambda_2')'$, with Λ_1 being an invertible lower triangular matrix. We refer to Bai and Ng (2002, 2013) and Stock and Watson (2002) for the identification of the principal component estimator for the mean panel data model.

In this paper, we employ the Markov chain Monte Carlo approach and generate a set of posterior samples. Our computation approach is very attractive because the common factor structure can be easily investigated conditional on the individual effects and vice versa. For the data-augmentation approach, we need to specify the prior distribution of the parameters. For the ease of computation, we assume that the priors of the factors and factor loadings are mutually independent, i.e., $\pi(B,F,\Lambda) = \pi(B,\Lambda)\pi(F)$. Then, the posterior density will be

$$\pi(B,F,\Lambda|Y,X) \propto L(Y|X,F,\Lambda,B)\pi(B,\Lambda)\pi(F),$$

which does not provide analytical posterior density forms.

When we use the principal component framework (See, e.g., Bai (2009) and references therein), we usually analyze the unobservable common factor and its factor loadings jointly. Thus, the prior specification will take the form $\pi(B,F,\Lambda) = \pi(B)\pi(\Lambda,F)$. In this paper, in contrast, we analyze the regression coefficients and the factor loadings

jointly. This treatment will provide a convenient data augmentation for inference on these unknown parameters. Moreover, although one might conjecture that equation (1) allows us to easily derive the conditional posterior distributions of the interactive fixed-effect parameters (F, Λ), it does not lead to an easy method for sampling from their posterior distribution because the error term ε_{it} is not normal.

3.1.1 Prior specification and posterior analysis for B and Λ

Here, we specify the prior densities on B and Λ and derive their conditional posterior distributions, given F and Ω. First, the likelihood contribution of observation y_{it} can be expressed as

$$\begin{aligned}
& \left[\frac{\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i)}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i)}\right]^{y_{it}} \times \left[\frac{1}{1+\exp(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i)}\right]^{1-y_{it}} \\
= & \frac{\exp\{\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i\}^{y_{it}}}{1+\exp\{\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i\}} \\
\propto & \exp\{z_{it}(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i)\} \times \int_0^\infty \exp\{-\omega_{it}\{\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i\}^2/2\} p(\omega_{it}) d\omega_{it} \\
\equiv & \exp\{z_{it}\boldsymbol{v}_{it}'\boldsymbol{\gamma}_i\} \times \int_0^\infty \exp\{-\omega_{it}\{\boldsymbol{v}_{it}'\boldsymbol{\gamma}_i\}^2/2\} p(\omega_{it}) d\omega_{it},
\end{aligned}$$

where $z_{it} = y_{it} - 1/2$, and $p(\omega_{it})$ is the density of a Polya-Gamma random variable with parameters $(1, 0)$. In the cross-sectional context, this expression is obtained in Polson and Scott (2013). For simplicity of notation, we used $\boldsymbol{v}_{it} = (\boldsymbol{x}_{it}', \boldsymbol{f}_t')'$, and $\boldsymbol{\gamma}_i = (\boldsymbol{b}_i', \boldsymbol{\lambda}_i')'$.

When we wish to obtain the maximum likelihood estimator, we simply use the diffuse prior $\pi(\boldsymbol{\gamma}_i) \propto$ Const.. Then, the conditional posterior density of $\boldsymbol{\gamma}_i = (\boldsymbol{b}_i', \boldsymbol{\lambda}_i')'$ is

$$\begin{aligned}
\pi(\boldsymbol{\gamma}_i | Y, X, B_{-i}, \Lambda_{-i}, \Omega) & \propto \exp\{z_{it}\boldsymbol{v}_{it}'\boldsymbol{\gamma}_i - \omega_{it}\{\boldsymbol{v}_{it}'\boldsymbol{\gamma}_i\}^2/2\} \\
& \propto \exp\left\{-\frac{1}{2}(\boldsymbol{z}_i - W_i\boldsymbol{\gamma}_i)'\Omega_i(\boldsymbol{z}_i - W_i\boldsymbol{\gamma}_i)\right\},
\end{aligned}$$

which implies that the conditional posterior density of $\boldsymbol{\gamma}_i$ is the multivariate normal density with mean $(W_i'\Omega_i W_i)^{-1}W_i'\boldsymbol{z}_i$ and variance-covariance matrix $(W_i'\Omega_i W_i)^{-1}$. Here $\Omega \equiv \{\omega_{it} | i = 1, ..., N, t = 1, ..., T\}$, $W_i = (X_i, F)$ is the design matrix, $B_{-i} = (\boldsymbol{b}_1, ..., \boldsymbol{b}_{i-1}, \boldsymbol{b}_{i+1}, ..., \boldsymbol{b}_N)'$ and $\Lambda_{-i} = (\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_{i-1}, \boldsymbol{\lambda}_{i+1}, ..., \boldsymbol{\lambda}_N)'$. This implies that the conditional posterior density of $\boldsymbol{\gamma}_i$ is the multivariate normal density with mean $(W_i'\Omega_i W_i + A_{\gamma_i}^{-1})^{-1}W_i'\boldsymbol{z}_i$ and variance-covariance matrix $(W_i'\Omega_i W_i + A_{\gamma_i}^{-1})^{-1}$.

3.1.2 Conditional posterior density of ω_{it}

We can easily obtain the conditional posterior densities of ω_{it}, that is,

$$\pi(\omega_{it} | Y, X, B, \Lambda, \Omega_{-\omega_{it}}) \propto \exp\{-\omega_{it}\{\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i\}^2/2\} p(\omega_{it}),$$

which is a Polya-Gamma distribution with parameter $(1, \boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i)$. Again, we can easily draw a posterior sample of ω_{it} using the Gibbs sampler.

3.1.3 Prior specification and posterior analysis for F

Combining the terms from all observations yields the following expression for the conditional posterior of F:

$$\begin{aligned}
\pi(F|Y, X, B, \Lambda, \Omega) &\propto \pi(F) \prod_{i=1}^{N}\prod_{t=1}^{T}\left[\exp\{z_{it}\{\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i\}\} \times \exp\{-\omega_i\{\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i\}^2/2\}\right] \\
&\propto \pi(F) \prod_{i=1}^{N}\prod_{t=1}^{T}\exp\left\{-\frac{\omega_{it}}{2}\{z_{it}/\omega_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\}^2\right\} \\
&\propto \pi(F)\exp\left\{-\sum_{i=1}^{N}(\boldsymbol{z}_i^* - F\boldsymbol{\lambda}_i)'\Omega_i(\boldsymbol{z}_i^* - F\boldsymbol{\lambda}_i)\right\}, \qquad (6)
\end{aligned}$$

where $\Omega_i = \text{diag}\{\omega_{i1}, ..., \omega_{iT}\}$, $\boldsymbol{z}_i^* = (z_{i1}^*, ..., z_{iT}^*)$ with $z_{it}^* = z_{it}/\omega_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i = (y_{it} - 1/2)/\omega_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i$.

We further investigate the form of the conditional posterior of F: In this paper, the common factor F is subject to the normalization condition $F'F/T = I_r$ for identification purposes. From $F'F/T = I_r$, F belongs to a hyperball in T dimensions, and its support is restricted to be the Cartesian product of the T-dimensional hyperball. Furthermore, because of the orthogonality requirement, its support is then reduced to a Stiefel manifold $S_{T,r}$ of radius $\sqrt{T}$ (Khatri and Mardia (1977)). Therefore, the prior of F is a flat prior over the Stiefel manifold corresponding to orthogonal transformations and, hence, is invariant with respect to the orthogonal group. Specifically, the prior of F is

$$\pi(F) = \frac{1}{C(T, r)} \cdot 1(F \in S_{T,r}), \qquad (7)$$

where $1(\cdot)$ is the indicator function and

$$C(T, k) = \frac{2^k \pi^{kT/2} T^{k(2T-k-1)/4}}{\pi^{k(k-1)/4} \prod_{j=1}^{k} \Gamma\{(T - j + 1)/2\}}$$

is the normalizing constant with $\Gamma(\cdot)$ being the Gamma function.

However, under the prior $\pi(F)$ in (7), the analysis of the conditional posterior of F in (6) is still not straightforward. This is mainly because the diagonal matrix Ω_i prevents the derivation of an analytical conditional posterior of F. We therefore use the Metropolis-Hastings algorithm to generate the posterior sample of F. Given the posterior density $\pi(F|Y, X, B, \Lambda, \Omega)$, known up to a constant, and a proposal conditional density $p(F)$, we can generate the posterior sample of F in the following way.

To generate samples from $\pi(F|Y, X, B, \Lambda)$, the Metropolis-Hastings algorithm requires us to specify a proposal density $p(F)$. The algorithm then first draws a candidate parameter value F^{new} from the proposal density $p(F)$. The generated parameter value F^{new} will be accepted or rejected based on the acceptance probability

$$\alpha = \min\left\{1, \frac{L(Y|X, F^{new}, \Lambda, B)\pi(B, F^{new}, \Lambda)/p(F^{new})}{L(Y|X, F^{old}, \Lambda, B)\pi(B, F^{old}, \Lambda)/p(F^{old})}\right\},$$

where F^{old} is the current state of F.

In the practical implementation of the Metropolis-Hasting algorithm, we need to prepare a proposal density. Here, the random-walk Metropolis-Hastings algorithm is used. We draw a new candidate F^{new} from a proposal density

$$p(F) \propto \exp\{-\text{tr}\{(Z - F\Lambda)'(Z - F\Lambda)\}\} \cdot 1(F \in S_{T,r}).$$

where F is on the Stiefel manifold and $Z = (\boldsymbol{z}_1, ..., \boldsymbol{z}_N)$. Simulation of the matrix Bingham-von Mises-Fisher distribution can be found in Hoff (2009). In our simulation study, this proposal density works well.

3.1.4 Posterior sampling algorithm and the estimator

As discussed above, we can analytically obtain the conditional posterior distributions of B, Λ and Ω. Therefore, we easily draw the posterior samples by implementing the Gibbs sampling algorithm. To draw F, we can use the Metropolis-Hastings algorithm. Given value of the number of common factors r, our estimation algorithm is summarized as follows.

Posterior sampling algorithm:
Step 1. Initialize the parameters.
Step 2. Sample F from $\pi(F|Y, X, B, \Lambda, \Omega)$.
Step 3. Sample $\boldsymbol{\gamma}_i$ from $\pi(\boldsymbol{\gamma}_i|Y, X, B_{-i}, \Lambda_{-i}, \Omega)$.
Step 4. Sample ω_{it} from $\pi(\omega_{it}|Y, X, B, \Lambda, \Omega_{-\omega_{it}})$.
Step 5. Repeat Step 2 to Step 4 for a sufficiently large number of iterations.

In Step 1 we need to initialize the parameters. The details of our efficient initialization algorithm is given in online supplementary document. The outcomes of the above algorithm can be regarded as a random sample from the joint posterior density function after a burn-in period. We then obtain a set of H posterior samples $\{B^{(k)}, F^{(k)}, \Lambda^{(k)}; k = 1, ..., H\}$ for inference. Then, our estimator, $\{\hat{B}, \hat{F}, \hat{\Lambda}\}$, is given as

$$\{\hat{B}, \hat{F}, \hat{\Lambda}\} = \text{argmax}_{\{B^{(k)}, F^{(k)}, \Lambda^{(k)}\}; k=1,...,H} L(Y|X, B^{(k)}, F^{(k)}, \Lambda^{(k)}), \tag{8}$$

where the likelihood function is given in (4).

3.2 Model selection

In practice, we have to determine the dimension of the interactive effects, or equivalently, Because the presence of iterative effects, cross-validation can not be applied easily (Ando and Bai (2018)). Although Ando and Bai (2017a), Bai and Ng (2002), Hallin and Liška (2007) proposed some model selection criteria, these are applicable only for the linear panel data models, and thus can not be applied to panel choice models. In this paper, we propose a new information criterion. The dimension of the interactive effects is selected by minimizing the following information criterion

$$IC(r) = \log L(Y|X, \hat{B}^{(r)}, \hat{F}^{(r)}, \hat{\Lambda}^{(r)}) + r \times q(N,T) \tag{9}$$

where $\hat{B}^{(r)}$, $\hat{F}^{(r)}$ and $\hat{\Lambda}^{(r)}$ are the estimated model parameters under the dimension of the interactive effects being r. The function $q(N,T)$ is a penalty on the dimension of interactive effects. In numerical study, we specify the function as

$$q(N,T) = \log\left(\frac{NT}{N+T}\right)\left(\frac{N+T}{NT}\right). \tag{10}$$

The asymptotic performance of $IC(r)$ in (9) is investigated in the next section. As shown in Bai and Ng (2002), the penalty function (10) satisfies the conditions in Theorem 3, which is given in the next section. One can also consider alternative penalty function. However, this is out of scope of this paper.

4 Asymptotic results

There is a rich opportunity to apply the proposed method, but theoretical results are lacking in the literature. As a theoretical justification of the proposed method, this section provides the theoretical results.

4.1 Assumptions

We first state the assumptions needed for the asymptotic analysis. Because the dimensions of B, Λ and F are diverging, we cannot assume the standard regularity conditions for likelihood functions. The set of regularity conditions that are imposed on the proposed model are as follows:

We first 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. The 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 set of regularity conditions that are imposed on the proposed model are as follows:

Assumption A: Common factors

Let $\mathcal{F}$ be compact subset of R^r. The common factors $\boldsymbol{f}_t \in \mathcal{F}$ satisfy $T^{-1}\sum_{t=1}^{T} \boldsymbol{f}_t \boldsymbol{f}_t{}' \to \Sigma_F$ as $T \to \infty$, where Σ_F is an $r \times r$ positive definite matrix.

Assumption B: Factor loadings and regression coefficients

Let $\mathcal{B}$ and $\mathcal{L}$ be compact subsets R^p and R^r. The regression coefficient $\boldsymbol{b}_i$ and the factor-loading for the common factors satisfy $\boldsymbol{b}_i \in \mathcal{B}$ and $\boldsymbol{\lambda}_i \in \mathcal{L}$. Also, the factor-loading matrix $\Lambda = [\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N]'$ satisfies $N^{-1}\Lambda'\Lambda$ being $r \times r$ positive definite matrix.

Assumption C: Idiosyncratic error terms

Idiosyncratic error term u_{it} is i.i.d. over i and t, and is independent of $\boldsymbol{x}_{ks}$, $\boldsymbol{\lambda}_\ell$, $\boldsymbol{f}_m$ for all k, s, ℓ, m.

Assumption D: Predictors and design matrix

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

(D2): Under the normalization for F such that $F'F/T = I$, for each i and all T, there exist positive constants C_1 and C_2 such that

$$0 < C_1 < \lambda_{\min}(T^{-1}(X_i, F)'(X_i, F)) < \lambda_{\max}(T^{-1}(X_i, F)'(X_i, F)) < C_2 < \infty,$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest eigenvalue and largest eigenvalue of a matrix A, respectively.

(D3): Define $A_i = \frac{1}{T}X_i' M_F X_i$, $B_i = (\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i') \otimes I_T$, $C_i' = \frac{1}{\sqrt{T}}\boldsymbol{\lambda}_i' \otimes (X_i' M_F)$, $\boldsymbol{\eta} = \frac{1}{\sqrt{T}}\mathrm{vec}(M_F F)$, $M_F = I - F(F'F)^{-1}F'$. Let $\mathcal{F}$ be the collection of F such that $\mathcal{F} = \{F : F'F/T = I\}$. We assume

$$\inf_{F \in \mathcal{F}} \left[\frac{1}{N}\sum_{i=1}^{N} E_i(F)\right] \quad \text{is positive definite,} \tag{11}$$

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

Remark 2 The full rank assumption of Σ_F and Σ_Λ in Assumptions A and B is necessary for the number of common factors to be r. Also, the assumption on compact subset is not restrictive when F and Λ are both treated as fixed parameters. Assumption C may

be relaxed to allows cross-sectional and serial correlations and heteroskedasticities in the idiosyncratic errors ε_{it}. However, exploring the weaker condition is beyond the scope of this paper. Assumptions D is necessary to obtain the asymptotic distributions on the regression coefficients and the structure of interactive effects.

4.2 Theoretical results

First of all, all proofs are given in online supplementary document. Because the dimensions of the panel size N and T are diverging, a novel proof is developed. Now, we denote $B = (\boldsymbol{b}_{1,0}, ..., \boldsymbol{b}_{N,0})'$, $\Lambda = (\boldsymbol{\lambda}_{1,0}, ..., \boldsymbol{\lambda}_{N,0})'$, and $F = (\boldsymbol{f}_{1,0}, ..., \boldsymbol{f}_{T,0})'$ as the true parameter values. The following proposition provides the average consistency of $\hat{\boldsymbol{\gamma}}_i \equiv (\hat{\boldsymbol{b}}_i', \hat{\boldsymbol{\lambda}}_i')'$. Note that the estimated common factor $\hat{F}$ is consistent in a certain norm, which implies that the space spanned by F_0 and the space spanned by the estimated factors $\hat{F}$ are asymptotically the same.

Proposition 1 *Under Assumptions A–D and* $\log\log(N)/T \to 0$*, the following claims hold:*

$$N^{-1}\sum_{i=1}^{N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|^2 = o_p(1),$$
$$T^{-1}\|\hat{F} - F_0\|^2 = o_p(1).$$

Next, we prove that the estimated regression coefficients $\hat{\boldsymbol{b}}_i$ and the estimated factor loading $\hat{\boldsymbol{\lambda}}_i$ converge in probability to $\boldsymbol{b}_{i,0}$, $\boldsymbol{\lambda}_{i,0}$ and uniformly over $1 \le i \le N$. We also show that the estimated factor $\hat{\boldsymbol{f}}_t$ converges in probability to $\boldsymbol{f}_{t,0}$ uniformly over $1 \le t \le T$.

Theorem 1 *Under Assumption A – Assumption D,* $\log(N)/T \to 0$ *and* $\log(T)/N \to 0$*,* $\hat{\boldsymbol{b}}_i$ *and* $\hat{\boldsymbol{\lambda}}_i$ *are consistent*

$$\max_{1\le i\le N} \|\hat{\boldsymbol{b}}_i - \boldsymbol{b}_{i,0}\| = o_p(1),$$
$$\max_{1\le i\le N} \|\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_{i,0}\| = o_p(1).$$

Moreover, the estimated common factor is consistent

$$\max_{1\le t\le T} \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| = o_p(1).$$

Next, Theorem 2 shows that the asymptotic distribution of the estimated parameters, $\hat{\boldsymbol{\gamma}}_i$, is multivariate normal. Similarly, the asymptotic distribution of the estimated common factor $\hat{\boldsymbol{f}}_t$ is also multivariate normal distribution.

Theorem 2 *Under Assumption A – D,* $\log(N)N^{1/2}/T^{1-\alpha} \to 0$ *and* $\log(T)T^{1/2}/N^{1-\alpha} \to 0$ *with a small* α $(0 < \alpha < 1/2)$*, the asymptotic distribution of* $T^{1/2}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0})$ *is the multivariate normal with mean* $\mathbf{0}$ *and covariance matrix*

$$\Sigma_i = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \pi_{it,0}(1 - \pi_{it,0}) \boldsymbol{z}_{it,0} \boldsymbol{z}'_{it,0},$$

with $\boldsymbol{z}_{it,0} = (\boldsymbol{x}'_{it}, \boldsymbol{f}'_{t,0})'$ *and* $\pi_{it,0}$ *is the true choice probability. Moreover, the asymptotic distribution of* $N^{1/2}(\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0})$ *is the multivariate normal with mean* $\mathbf{0}$ *and covariance matrix*

$$\Theta_t \equiv \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \pi_{it,0}(1 - \pi_{it,0}) \boldsymbol{\lambda}_{i,0} \boldsymbol{\lambda}'_{i,0}.$$

Next, we provide a new solution to this issue and provide a theoretical justification for our proposed model selection criterion.

Theorem 3 *Suppose that the set of regularity conditions of Theorem 2 hold. Then, under the model selection criterion* $IC(r)$ *with penalty* $q(N,T)$ *that satisfies*

$$q(N,T) \to 0 \quad \text{and} \quad C_{NT}^{-1} \times q(N,T) \to \infty,$$

where $C_{NT} = \min\{N,T\}$*, the true number of common factors* r_0 *will be selected asymptotically.*

5 Simulation results

To demonstrate the usefulness of the proposed method, we use simulated data, for which the data generating process and the model parameters are known so that evaluation can be performed.

5.1 Performance of the estimation procedure

For the true data-generating process, we use

$$u_{it} = \boldsymbol{x}'_{it}\boldsymbol{b}_i + \boldsymbol{f}'_t\boldsymbol{\lambda}_i + \varepsilon_{it}, \quad i = 1,\ldots,N,\ t = 1,\ldots,T,$$

where the $r = 3$-dimensional factors $\boldsymbol{f}_t$ and the factor-loading vector $\boldsymbol{\lambda}_i$ are vector of $N(0,1)$ variables,

We investigate the performance of the proposed procedure when the regressors and the unobservable factor structures exhibit dependency. The method also worked well

in various settings. Here, we report the results for the challenging case in which the explanatory variables x_{it} are correlated with the unobservable factor structures. Setting $p_i = 10$, we generate the set of regressors as follows:

$$x_{it,1} = s_{it,1} + 0.05 f_{1t}, \quad x_{it,3} = s_{it,3} - 0.05 f_{2t}$$
$$x_{it,5} = s_{it,5} + 0.05 f_{3t}, \quad x_{it,k} = s_{it,k} \quad (k \neq 1, 3, 5),$$

where $u_{it,k}$ is generated from the uniform distribution over $[-1, 1]$. The method also worked well in various settings. Setting $p_i = 10$, the true parameter values of $\boldsymbol{\beta}_i$ are set to be $\boldsymbol{\beta}_{i,0} = (-1, 1, -1, 1, -1, 1, 0.5v_{i6}, 0.5v_{i7}, 0.5v_{i8}, 0.5v_{i9}, 0.5v_{i,10})'$, where v_{ij} is from uniform $[-1, 1]$. We simulate a large panel with N individuals and T time periods. We base our estimate with the true number of factors and assess the robustness of the proposed strategy to endogeneity.

For the error ε, we consider the followings:

DGP1: the idiosyncratic error term ε_{it} follows the standard logistic distribution.

DGP2: the idiosyncratic error term ε_{it} is generated from multivariate normal with mean 0 and covariance matrix $\sigma_{ij} = 0.5^{-|i-j|}$.

DGP3: the idiosyncratic error term ε_{it} exhibits a serial correlation such that $\varepsilon_{it} = 0.3\varepsilon_{i,t-1} + e_{it}$ with e_{it} is from the standard normal distribution $N(0, 1)$.

We compare the proposed inference procedure with the maximum likelihood estimator ignoring the factor structure (MLE without factor structure $\boldsymbol{f}_t'\boldsymbol{\lambda}_i$). The maximum likelihood estimator ignoring the factor structure is found as the maximizer of the standard likelihood function (5). Thus, the factor structure is ignored, which implies that the ignorance of endogeneity.

Estimation results are averaged over 100 simulated data sets and are reported in Table 1 $\sim$ Table 3. To save the space, Table 1 $\sim$ Table 3 are given in Appendix G (See the online supplementary document). The total number of Markov chain Monte Carlo iterations is 2,000, of which the first iterations are usually discarded as burn-in. Because our estimator $\{\hat{B}, \hat{F}, \hat{\Lambda}\}$ is the maximizer of (8), we picked up the k-th iteration that maximizes the $L(Y|X, B^{(k)}, F^{(k)}, \Lambda^{(k)})$ under $k = 1, ..., 2000$. Tables show the following mean squared errors (MSE)

$$\text{MSE}_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{X_i\boldsymbol{b}_{i,0} + F\boldsymbol{\lambda}_{i,0} - X_i\hat{\boldsymbol{b}}_i - \hat{F}\hat{\boldsymbol{\lambda}}_{it}\}^2,$$
$$\text{MSE}_2 = \frac{1}{NT}\sum_{i=1}^{N}\|X_i\boldsymbol{b}_{i,0} - X_i\hat{\boldsymbol{b}}_i\|^2,$$

$$\text{MSE}_3 \quad = \quad \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{b}_{i,0} - \hat{\boldsymbol{b}}_i\|^2.$$

These measures are also computed for the maximum likelihood estimators (without the factor structure). The MLE without factor structure ignores the endogeneity, and the true structure $\boldsymbol{\mu}_i^0 = X_i\boldsymbol{\beta}_i^0 + F_0\boldsymbol{\lambda}_i^0$ is estimated by $X_i\hat{\boldsymbol{\beta}}_i$, where $\hat{\boldsymbol{\beta}}_i$ is the parameter estimates. As shown in Table 1 $\sim$ Table 3, the proposed method is capable of capturing the true structures. In contrast, the standard argument of consistency of the MLE without factor structure no longer holds. Their MSE measures do not decrease even T and(or) N increase. Therefore, it is important to have the factor structure if it indeed exists.

5.2 Performance of model selection criterion

We investigate the performance of the proposed model-selection criterion, $IC(r)$, to select the dimension of the interactive effects. We set the possible dimension of the interactive effects (i.e., the numbers of common factors) r to range from 1 to 8. We use the same data-generating process described in the previous section. We generated the dataset under the various combinations of N and T. Table 4 $\sim$ Table 6 (See Appendix G in online supplementary document) report the percentages for the selected dimension of interactive effects r based on the proposed criterion. The results are obtained based on 100 repetitions. As shown in the tables, the proposed criterion is capable of selecting the true dimension of common factors. When the size of panel N and T increases, the procedure achieves the better performance in terms of identifying true of dimension of common factors.

6 Application

The concept of operational efficiency has been playing an important role in service management. One of the key managerial aspects that determines operational efficiency is to match demand and service capacity. Examples include hair salon, relaxation massage, restaurant, and taxi service. This matching process forces a manager to consider, among other factors, what predefined level of capacity/staffing will be needed in a given time period. It is obvious that for the decision, the service capacity/staffing levels should be reasonable such that the managers can meet the actual demand, which cannot be observed in advance. To track the performance from this aspect, it is beneficial to have an efficiency measure of capacity utilization. This key performance indicator will provide useful information for the managers. This section studies the efficient capacity utilization

of taxi industry. From regulatory perspective, the improvement of efficiency measures is important. This is because an inefficient taxi service management will lead to more the idle time and passengers will face longer waiting time.

6.1 Background information and data

We analyze the data collected by the New York City Taxi and Limousine Commission (TLC). The data contains Yellow medallion taxi ID, driver initial, shift number, pick-up time, drop-off time, and geographical location for trip origin and destination, travel distance and fare etc.

The New York City taxi market is highly regulated both from pricing and entry aspects. First, the medallion taxis use the same pricing scheme. Second, the number of medallions (legally permitted to operate taxi) is capped. Third, Yellow medallion taxis are not authorized to conduct pre-arranged pick-ups. Passengers pick up Yellow medallion taxis from the street. In other words, taxis and passengers need to find one another. From management and regulatory perspectives, it is thus important to understand the efficient capacity utilization of each Yellow medallion taxi.

Let y_{it} be the efficiency measure whether the taxi medallion i achieves the pre-specified level of capacity utilization rate (Achieve; Yes=1 or No=0) at time t. More specifically, we set the pre-specified level of capacity utilization every hour. If the medallion taxi i drives with passengers longer than 20 minutes during time period t (In this study every hour. For example, 10:00AM–11:00AM), then the pre-specified capacity utilization rate is successfully achieved. We note that an alternative pre-specified rate can be considered depending upon the management. By sampling 3000 medallion taxi in March 1 2013 – March 21 2013, we create a panel with size $(T, N) = (744, 3000)$. Here $T = 744$ implies that the data period spans 31 days (24 hours/per day $\times$ 31 days) and there are $N = 3000$ individual taxi medallions. Therefore, we analyze more than 2 million trip records.

Figure 1 provides examples of the measured performance from 3 medallion taxis. The horizontal axis is the dates and the vertical axis represents the hours. Colored cell indicates that the individual medallion taxi achieved the pre-specified capacity utilization rate. A fair amount heterogeneity can be observed in Figure 1.

Through our empirical analysis, we explore the following questions: Are there any performance variations over the medallion taxis? Is a particular medallion taxi doing better than the others? If not, how to improve the capacity utilization rate? Also, we address the question: how many common factors is under a specified explanatory variables? If the common factors exist, the standard MLE without factor structures

would lead to the inconsistent regression coefficients (See also Section 5.1). Thus, dealing with this issue is important to make sure that the performance measure (see below) is unbiased.

6.2 Model specification and estimation results

In our analysis, we use the following explanatory variables for $\boldsymbol{x}_{it}$: Time frame (Every hours; MIDNIGHT–1AM, 1AM–2AM,...,10PM–11PM, 11PM–MIDNIGHT), Weekday (Monday, Tuesday,..., Friday) and Weekend (Saturday, Sunday, Holiday). We use the indicator variables for these predictors.

To understand the data-generating structure, we obtained the estimator with the interactive effects. We generate 2,000 iterations using the proposed posterior sampling algorithms. In practice, we have to select the number of unobservable factors (the dimension of F) to adequately describe the information contained in the observed panel data. Here, we use the proposed model selection criterion $IC(r)$ in (9). For both panels, we select the best number of common factors as the minimizer of the $IC(r)$ score in (9). The best number of common factors is 2.

By setting the selected number of common factors, we compute the proposed estimators of $\boldsymbol{b}_i$ and factor loadings $\boldsymbol{\lambda}_i$ for each of the individual customers. Because our purpose is to compare the relative performance of 3000 medallion taxis, the estimated coefficients on standardized to have mean zero. Let $\hat{\boldsymbol{b}}_i$ be the coefficient vector on the meal times for the i-th customer. Then, it is standardized by subtracting the mean over i; $\bar{\boldsymbol{b}}_i \equiv \hat{\boldsymbol{b}}_i - \sum_{i=1}^{N} \hat{\boldsymbol{b}}_i/N$.

Figure 2 shows the estimated regression coefficients and the factor loadings. Each column in Figure 2 corresponds to the individual taxi's performance contributed by the predictors and to the common factors. On the top of the figure, trees based on the hierarchical clustering are presented. We can see that the individual taxi's performances are categorized into several segments.

6.3 Evaluation of efficient capacity utilization

By using the estimated model, we can evaluate the performance of 3000 medallion taxis. First, we evaluate the overall evaluation of 3000 medallion taxis by comparing their individual regression coefficients. Let $\bar{B} = (\bar{\boldsymbol{b}}_1, \bar{\boldsymbol{b}}_2, ..., \bar{\boldsymbol{b}}_{3000})'$ be the matrix of standardized regression coefficients, where $\bar{\boldsymbol{b}}_i$ is defined in the previous section. Then, an overall performance can be measured by

$$\boldsymbol{p}_{\text{overall}} \equiv \bar{B}\mathbf{1},$$

where $\mathbf{1}$ is a vector of ones. The i-th element of the vector $\boldsymbol{p}_{\text{overall}}$ corresponds to the overall performance of i-th medallion taxi. Large positive value of $p_{i,\text{overall}}$ implies that the i-th medallion taxi has been achieving a superior performance compared to the others, and vice-versa. After we sort the element of $\boldsymbol{p}_{\text{overall}}$ in decreasing order, we made a barplot given in Figure 3. We can see that there is a huge performance variation among the medallion taxis.

Can the inferior segment improve their performance? To identify a reason of inferior performance, we analyze the bottom 10% medallion taxis. We again apply the hierarchical clustering to $\bar{\boldsymbol{b}}_i$ with respect to the top 10% and the bottom 10% medallion taxis. Figure 4 (a) and (b) shows the clustering results for the top 10% and the bottom 10% medallion taxis, respectively. We can make the following observations. First, we can see the homogeneity of the top 10% medallion taxis from Figure 4 (a). This implies that there seems to be a good tactics to achieve the better performance. Second, there are roughly two segments among the bottom 10% medallion taxis in Figure 4 (b). With respect to the daytime performance, one segment (middle segment of Figure) is performing not well. This indicates that the taxi drivers need to improve their skills to be hailed by passengers as quickly as possible when they are vacant. The performance of the other segment (left side segment of Figure) for the daytime is comparable to the top 10% medallion taxis. However, for some reasons, their performance during the evening, before midnight, and after midnight is inferior to those of top 10% medallion taxis. In general, taxi drivers operate in one of two separate shifts (12 hours each including a time for meal and pass the taxi to the following driver). If a manager wants to improve the capacity utilization rate, it is recommended to make sure that both shifts are operated properly.

To further explore the performance after 5:00PM till 8:00AM next morning, the pickup locations are explored. Figure 5 and Figure 6 compare the density of pick up locations recorded by the top 10% and the bottom 10% medallion taxis. Darker points correspond to points of higher density. We can see that the bottom 10% medallion taxis tend to be hailed by passengers around Midtown Manhattan. Typically, they tended to be hailed around Times Square. In contrast, the top 10% medallion taxis were hailed in Lower Manhattan in addition to Midtown Manhattan. Also, the density around Times Square is lower than that of the bottom 10% medallion taxis. Similar patterns are observed for the day time (between 8:00AM and 5:00PM) in Figure 7 and Figure 8. To improve the performance, the bottom 10% medallion taxi drivers may explore a matching opportunity around Lower Manhattan.

7 Further extensions of the inference procedure

7.1 Probit model specification

Finally, we note that our idea can be extended to the direct estimation of the probit model. Assuming the normal idiosyncratic shock ε_{it} in (1), we obtain the probit specification of the conditional choice probability,

$$P(y_{it} = 1|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_t, \boldsymbol{\lambda}_i) = \Phi\left(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right),$$

where $\Phi(\cdot)$ is the distribution function of the standard normal. Putting this the conditional choice probability function into (4), we obtain the likelihood function. Assuming that the unobserved factors are contained in the span of the observed factors and the cross-sectional averages of the regressors, Boneva and Linton (2017) proposed the estimator belongs to the class of common correlated effects estimators. However, this is restrictive assumption in some cases. Here, we show that the direct inference can be implemented without imposing such assumptions. To save the space, the details are provided in online supplementary document (Appendix E). Because all conditional posterior densities are obtained analytically, one can simply use Gibbs sampling algorithm (See also Albert and Chib (1993) who developed the data augmentation procedure for the standard probit regression model in the cross sectional context).

We finally note that the two specifications (logistic versus probit) can be compared by using the $IC(r)$ score (9). More specifically, for each of the specifications, we can find the optimal number of common factors r based on the minimizer of $IC(r)$. Then, the achieved minimum scores under these two specifications are compared. However, the comparison of logistic versus probit formulation is out of the scope of this paper.

7.2 Multiple alternatives model

In this section, we discuss an extension of our proposed data-augmentation strategy. In general, the number of alternatives will exceed 2. Each individual makes a single choice among many alternatives, such as transportation modes and occupational fields, selecting one candidate out of many. We extend our data-augmentation strategy to this setting.

Suppose that there are $i = 1, ..., N$ individuals and $J+1$ alternatives labeled $\{0, 1, ..., J\}$. At time period t, each individual chooses one of the alternatives. Consider the random utility for the j-th alternative

$$u_{ijt} = \boldsymbol{x}_{it}'\boldsymbol{b}_{ij} + \eta_{ijt} + \varepsilon_{ijt}, \quad i = 1, \ldots, N;\ t = 1, \ldots, T; j = 0, 1, 2, ..., J \tag{12}$$

where η_{ijt} denotes the unobserved structure of individual i's choice j, which can vary across time t, and ε_{ijt} follows a type I Extreme Value distribution. Alternative j ($j = 0, 1, 2, ..., J$) will be chosen if and only of $u_{ijt} > u_{ikt}$ ($k \neq j$). Thus, at time t, an individual i chooses alternative j if it offers the highest utility among all alternatives. Similar to the arguments in Section 3, we assume that these unobserved structures vary across time and individuals according to a factor structure:

$$\eta_{ijt} = \sum_{\ell=1}^{r} f_{jt\ell}\lambda_{ij\ell} = \boldsymbol{f}'_{jt}\boldsymbol{\lambda}_{ij} \tag{13}$$

and $\boldsymbol{f}_{jt}$ is an $r_j \times 1$ vector of unobservable factors and $\boldsymbol{\lambda}_{ij}$ represents the factor loadings.

Let $y_{ijt} \in \{0, 1\}$ denote the observed choice outcome, taking value 1 if the corresponding alternative j is chosen and 0 otherwise. Let $\boldsymbol{b}_i = (\boldsymbol{b}'_{i1}, ..., \boldsymbol{b}'_{iJ})'$, $\boldsymbol{f}_t = (\boldsymbol{f}'_{1t}, ..., \boldsymbol{f}'_{Jt})'$, and $\boldsymbol{\lambda}_i = (\boldsymbol{\lambda}'_{i1}, ..., \boldsymbol{\lambda}'_{iJ})'$. After normalizing the coefficients $(\boldsymbol{b}'_{i0}, \boldsymbol{\lambda}'_{i0})'$ for alternative 0 to zero, we obtain the multinomial logit specification (See McFadden (1973)) with the following choice probabilities

$$\begin{aligned} P(y_{ijt} = 1|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_t, \boldsymbol{\lambda}_i) &= \frac{\exp(\boldsymbol{x}'_{it}\boldsymbol{b}_{ij} + \boldsymbol{f}'_{jt}\boldsymbol{\lambda}_{ij})}{1 + \sum_{k=1}^{J}\exp(\boldsymbol{x}'_{it}\boldsymbol{b}_{ik} + \boldsymbol{f}'_{kt}\boldsymbol{\lambda}_{ik})}, \quad j = 1, ..., J, \\ P(y_{i0t} = 1|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_t, \boldsymbol{\lambda}_i) &= \frac{1}{1 + \sum_{k=1}^{J}\exp(\boldsymbol{x}'_{it}\boldsymbol{b}_{ik} + \boldsymbol{f}'_{kt}\boldsymbol{\lambda}_{ik})}. \end{aligned} \tag{14}$$

Assuming that the errors ε_{ijt} are independently and identically distributed, the joint probability of observing the complete set of choices $Y \equiv \{y_{ijt}|i = 1, ..., N, t = 1, ..., T, j = 1, ..., J\}$ is

$$\begin{aligned} &L(Y|X, B, F, \Lambda) \\ &= \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J}\left[\frac{\exp(\boldsymbol{x}'_{it}\boldsymbol{b}_{ij} + \boldsymbol{f}'_{jt}\boldsymbol{\lambda}_{ij})}{1 + \sum_{k=1}^{J}\exp(\boldsymbol{x}'_{it}\boldsymbol{b}_{ik} + \boldsymbol{f}'_{kt}\boldsymbol{\lambda}_{ik})}\right]^{y_{ijt}}\left[\frac{1}{1 + \sum_{k=1}^{J}\exp(\boldsymbol{x}'_{it}\boldsymbol{b}_{ik} + \boldsymbol{f}'_{kt}\boldsymbol{\lambda}_{ik})}\right]^{1-\sum_{k'=1}^{J}y_{ik't}}, \end{aligned}$$

where $X \equiv \{\boldsymbol{x}_{it}|i = 1, ..., N, t = 1, ..., T\}$, $\Lambda = (\Lambda_1, ..., \Lambda_J)$ with $\Lambda_j = (\boldsymbol{\lambda}_{j1}, ..., \boldsymbol{\lambda}_{jN})'$, $B = (B_1, ..., B_J)$ with $B_j = (\boldsymbol{b}_{j1}, ..., \boldsymbol{b}_{jN})'$ and $F = (F_1, ..., F_J)$ with $F_j = (\boldsymbol{f}_{j1}, ..., \boldsymbol{f}_{jT})'$ is the common factor.

We first consider the posterior sampling procedure of B and Λ because these parts are nearly identical to those presented in the previous section. Similar to the idea of Holmes and Held (2006) and Polson and Scott (2013), we rewrite each probability (14) as

$$P^*(y_{ijt} = 1|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_t, \boldsymbol{\lambda}_i) = \frac{\exp(\boldsymbol{v}'_{ijt}\boldsymbol{\gamma}_{ij} - \log\{1 + \sum_{k\neq j}\exp(\boldsymbol{v}'_{ikt}\boldsymbol{\gamma}_{ik})\})}{1 + \exp(\boldsymbol{v}'_{ijt}\boldsymbol{\gamma}_{ij} - \log\{1 + \sum_{k\neq j}\exp(\boldsymbol{v}'_{ikt}\boldsymbol{\gamma}_{ik})\})},$$

for $j = 1, ..., J$. Here, $\boldsymbol{v}_{ijt} = (\boldsymbol{x}'_{it}, \boldsymbol{f}'_{jt})'$, $\boldsymbol{\gamma}_{ij} = (\boldsymbol{b}'_{ij}, \boldsymbol{\lambda}'_{ij})'$, and $P(y_{i0t} = 1|\boldsymbol{x}_{it}, \boldsymbol{b}_{ij}, \boldsymbol{f}_{jt}, \boldsymbol{\lambda}_{ij}) = 1 - \sum_{j=1}^{J}P^*(y_{ijt} = 1|\boldsymbol{x}_{it}, \boldsymbol{b}_i, \boldsymbol{f}_t, \boldsymbol{\lambda}_i)$. Note that this transformation makes use of the normalization on alternative 0 (usually referred to as the outside option), while the normalization

is not used in the multinomial choice probability in Holmes and Held (2006) or Polson and Scott (2013). Normalization is useful for parameter identification for choice-dependent coefficients; see Greene (2000, page 860). Appendix F in the supplementary document provides a posterior sampling procedure for this multiple choice model with interactive effects.

7.3 High-dimensional predictors

When the dimension of $\boldsymbol{x}_{it}$ is large, some shrinkage methods are useful. The use of a shrinkage prior on $\boldsymbol{\gamma}_i$ allows us to address high-dimensional predictors. In the context of Bayesian (cross-sectional) linear regression, Park and Casella (2008) study the Bayesian lasso to exploit model inference via posterior distributions. To extend the proposed algorithm by incorporating the shrinkage approach, we can employ the Bayesian adaptive lasso prior (Leng et al. (2014)). It is ideal to place a larger penalty on the coefficients of unimportant predictors. The Bayesian adaptive lasso prior allows for variable selection with more flexible penalties than the Bayesian lasso. The Bayesian adaptive lasso prior on $\boldsymbol{\gamma}_i$ is $\pi(\boldsymbol{\gamma}_i) = \prod_{k=1}^{p_i+r} \frac{\kappa_{ik}}{2} \exp[-\kappa_{ik}|\gamma_{ik}|]$, where κ_{ik} corresponds to the adaptive weights in the adaptive lasso framework (Zou (2006)). Intuitively, a small penalty will be applied to the set of regressors that are relevant to the choices and a large penalty will be applied to those that are irrelevant.

Appendix F in the supplementary document summarizes the posterior sampling procedure based on the adaptive lasso prior. This extended algorithm can be applied to the panel logit model with high-dimensional predictors in the presence of endogeneity.

8 Conclusion

In this paper we introduced a new panel logistic regression models with interactive fixed effects. The estimation of the interactive effects is challenging because the likelihood function has an inconvenient form in terms of model parameters. We proposed a simple parameter estimation procedure as well as a new information criterion for determining the dimension of interactive effects. Numerical results showed that the proposed procedure perform well.

This paper has made several theoretical contributions. First, we proved the consistency of the estimated model parameters under double asymptotics where the dimensions of both cross section and time series of the panel go to infinity. This part gives am important contribution to the literature because the dimension of the parameter space grows with the panel size. We also studied the asymptotic distribution of the estimated parameters.

Moreover, the model selection consistency is established for the proposed information criterion.

Acknowledgments The authors would like to thank seminar participants at Monash University, University of Melbourne and University of Sydney, and participants at the fourth annual conference of the International Association for Applied Econometrics 2017.

References

Albert, J. and Chib, S. (1993) Bayesian Analysis of Binary and Polychotomous Response Data *Journal of the American Statistical Association*, 88, 669–679.

Amengual, D. and Watson, M. W. (2007) Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business and Economic Statistics*, 25, 91–96.

Ando, T. and Bai, J. (2015) Asset pricing with a general multifactor structure. *Journal of Financial Econometrics*, 13, 556–604.

Ando, T. and Bai, J. (2016) "Panel data models with grouped factor structures under unknown group membership," *Journal of Applied Econometrics*, 136, 163–191.

Ando, T. and Bai, J. (2017a) 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. and Bai, J. (2017b) Quantile co-movement in financial markets; A panel quantile model with unobserved heterogeneity. SSRN Working Paper.

Ando, T. and Bai, J. (2018) Selecting the regularization parameters in high-dimensional panel data models: consistency and efficiency. *Econometric Reviews*, 37, 183–211.

Andrews, R.L. Currim, I.S. and Leeflang, P.S.H. (2011) A comparison of sales response predictions from demand models applied to store-level versus panel data. *Journal of Business and Economic Statistics*, 29, 319–326.

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

Bai, J. and Ng, S. (2002) Determining the number of factors in approximate factor models. *Econometrica*, 70, 191–221.

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

Bai, J. and Li, K. (2014) Theory and methods of panel data models with interactive effects. *Annals of Statistics*, 42, 142–170.

Boneva, L. and Linton, O. (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*, forthcoming.

Charbonneau, K. (2017) Multiple Fixed Effects in Binary Response Panel Data Models. *Econometrics Journal*, forthcoming.

Chen, M. Fernández-Val, I. and Weidner, M. (2014) Nonlinear Panel Models with Interactive Effects. Working Paper.

Chudik, A. and Pesaran, M.H. (2015) Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188, 393–420.

Connor, G. and Korajzcyk, R. (1986) Performance measurement with the arbitrage pricing theory: a new framework for analysis. *Journal of Financial Economics*, 15, 373–394.

Ebbes, P., Wedel, M., Böckenholt, U. and Steerneman, T. (2005) Solving and testing for regressor-error (in)dependence when no instrumental variables are available: with new evidence for the effect of education on income. *Quantitative Marketing and Economics*, 3, 365–392.

Elrod, T. and Keane, M. P. (1995) A factor-analytic Probit model for representing the market structure in panel data. *Journal of Marketing Research*, 32, 1–16.

Fan, J. Liao, Y. and Mincheva, M. (2011) High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, 39, 3320–3356.

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.

Greene, W. (2000) *Econometric Analysis*, 4th Edition, Prentice Hall.

Hahn, P.R. Scott, J. and Carvalho, C.M. (2012) A Sparse Factor-Analytic Probit Model for Congressional Voting Patterns. *Journal of Royal Statistical Society*, C61, 619–635.

Hallin, M., and R. Liška (2007) The generalized dynamic factor model: determining the number of factors. *Journal of the American Statistical Association*, 102, 603–617.

Hoff, P.D. (2009) Simulation of the Matrix Bingham-von Mises-Fisher Distribution, With Applications to Multivariate and Relational Data. *Journal of Computational and Graphical Statistics*, 18, 438–456,

Holmes, C. and Held, L. (2006) Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1, 145–168.

Kapetanios, G., Pesaran, M.H. and Yamagata, T. (2011) Panels with non-stationary multifactor error structures. *Journal of Econometrics*, 160, 326–348.

Khatri, C. G. and Mardia, K. V. (1977) The von Mises.Fisher distribution in orientation statistics. *Journal of the Royal Statistical Society*, B39, 95–106.

Lee, S., Huang, J.Z. and Hu, J. (2010) Sparse principal components analysis for binary data. *Annals of Applied Statistics*, 4, 1579–1601

Leng, C., Tran, M.N. and Nott, D. (2014) Bayesian Adaptive Lasso, *Annals of Institute of Statistical Mathematics*, 66, 221–244.

Li, Y. and Ansari, A. (2014) A Bayesian Semiparametric Approach for Endogeneity and Heterogeneity in Choice Models. *Management Science*, 60, 1161–1179.

McFadden, D. (1973) Conditional logit analysis qualitative choice behavior. In *Frontiers of Econometrics*, ed. by P. Zarembka, Academic Press, N.Y., pp. 105–42.

Moon, H. and Weidner, M. (2015) Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83, 1543–1579.

Moon, H., Shum, M. and Weidner, M. (2016) Estimation of random coefficients logit demand models with interactive fixed effects. *Journal of Econometrics*, forthcoming.

Naik, P.A., Wedel, M. and Kamakura, W. (2015) Multi-index binary response analysis of large data sets. *Journal of Business & Economic Statistics*, 28, 67–81.

Park, T. and Casella, G. (2008) The Bayesian Lasso, *Journal of the American Statistical Association*,103, 681–686.

Perez, M.F. Shkilko, A. and Sokolov, K. (2015) Factor models for binary financial data. *Journal of Banking and Finance*, forthcoming.

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

Pesaran, M.H. and Tosetti, E. (2011) Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161, 182–202.

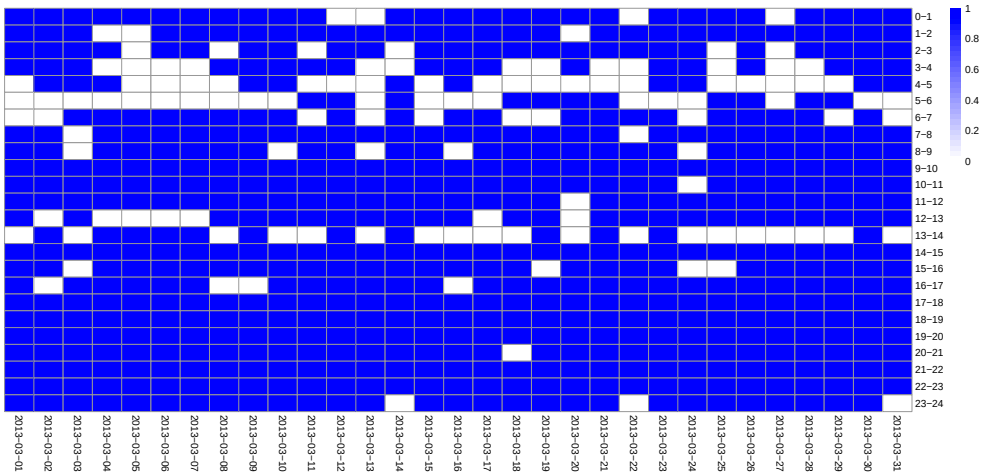
Polson, N.G. and Scott, J. (2013) Bayesian inference for logistic models using Polya-Gamma latent variables. *Journal of the American Statistical Association*, 108, 1339–1349.

Song, M. (2013) Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Working paper, Columbia University.

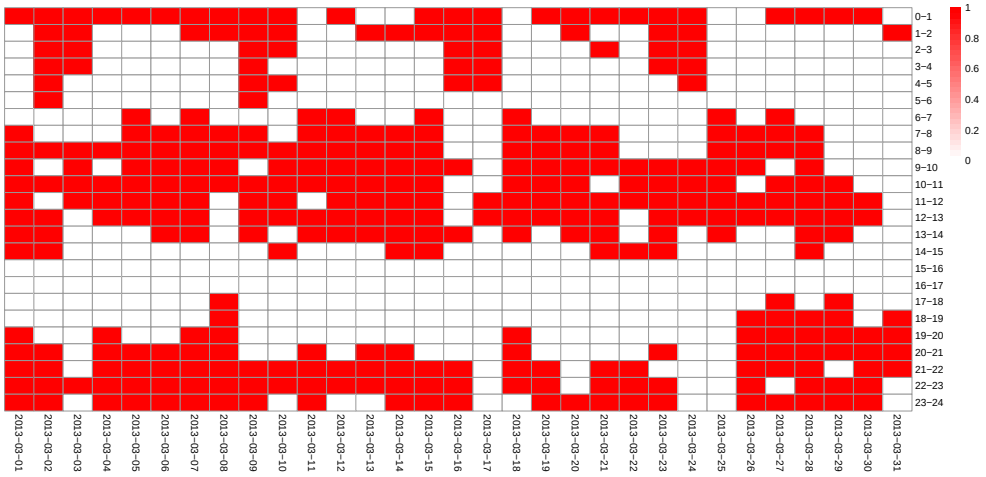
Stock, J. H., and Watson, M. W. (2002) Forecasting using principal components from a large number of observable factors. *Journal of the American Statistical Association*, 97, 1167–1179.

Sun, Y. (2016) Likelihood-based inference for nonlinear models with both individual and time effects. Discussion paper series, KU Leuven Department of Economics.

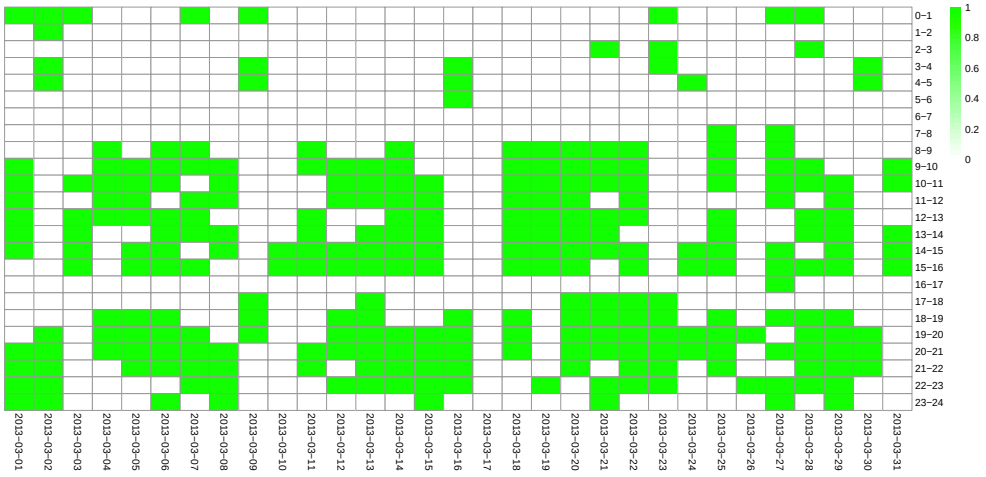
Zou, H. (2006) The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429.



Medallion taxi A.



Medallion taxi B.



Medallion taxi C.

Figure 1: Examples of performance achievements from 3 medallion taxis in March 2013. Horizontal axis; date, Vertical axis; hours. Colored cell indicates that the medallion taxi achieved the pre-specified level of performance.

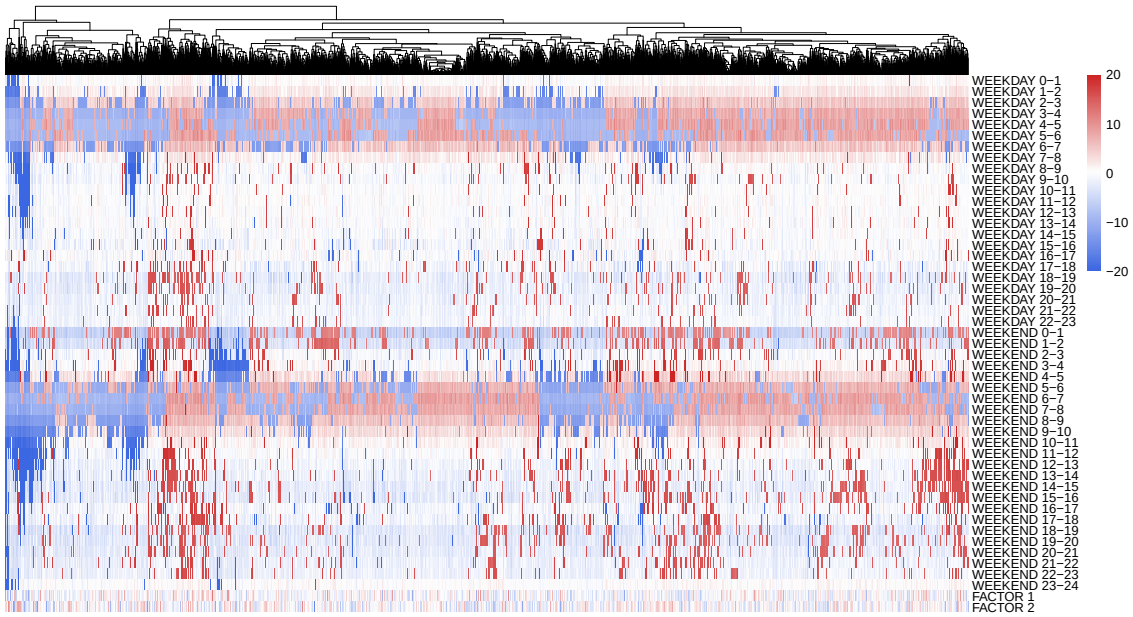


Figure 2: Hierarchical clustering results. Each column corresponds to the standardized individual's sensitivity to the predictors and to the common factors, $\bar{B}$. Trees on the top are from the hierarchical clustering.

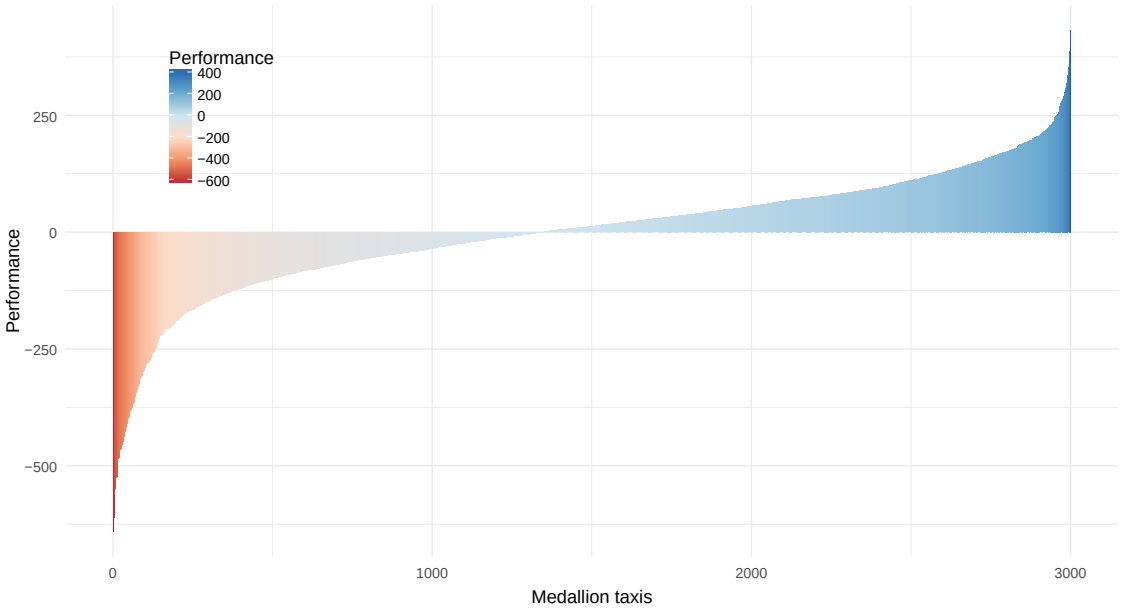
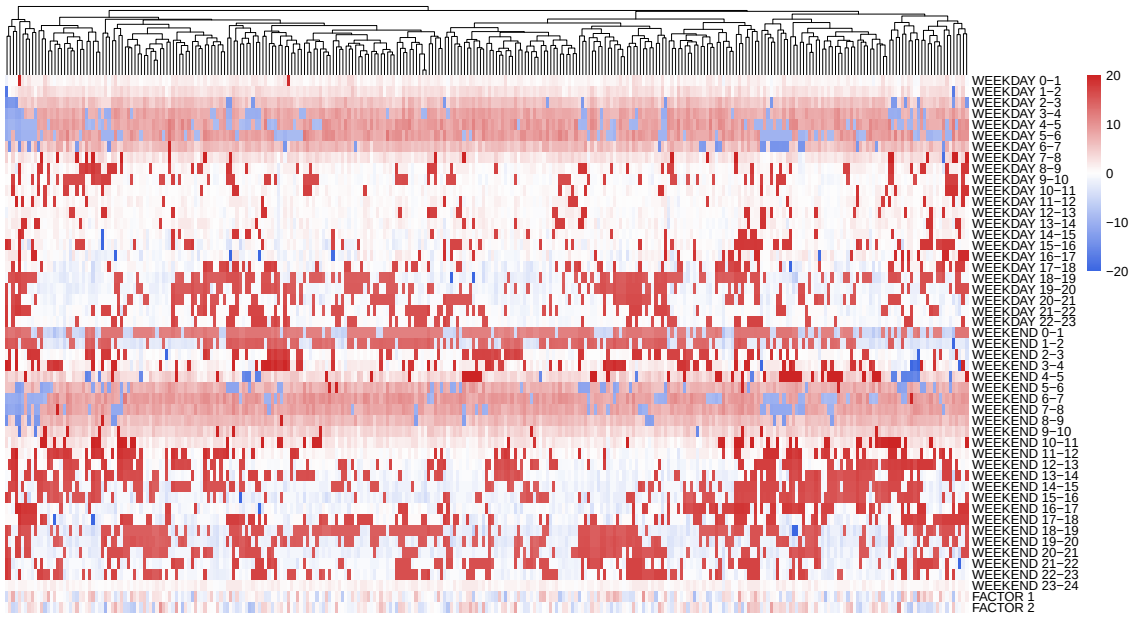
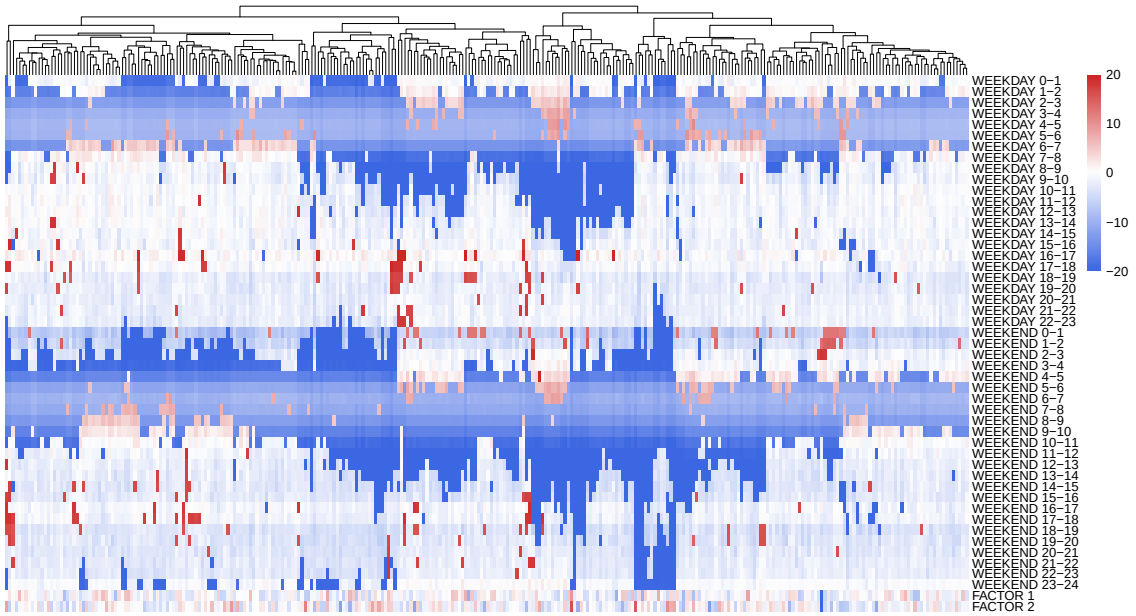


Figure 3: Overall performance of each taxi based on $\boldsymbol{p}_{\text{overall}}$.



(a) Top 10%



(b) Bottom 10%

Figure 4: Hierarchical clustering results of $\bar{\boldsymbol{b}}_i$ for the top 10% and the bottom 10% taxis. Each column corresponds to the standardized individual's sensitivity to the predictors and to the common factors. Trees on the top are from the hierarchical clustering.

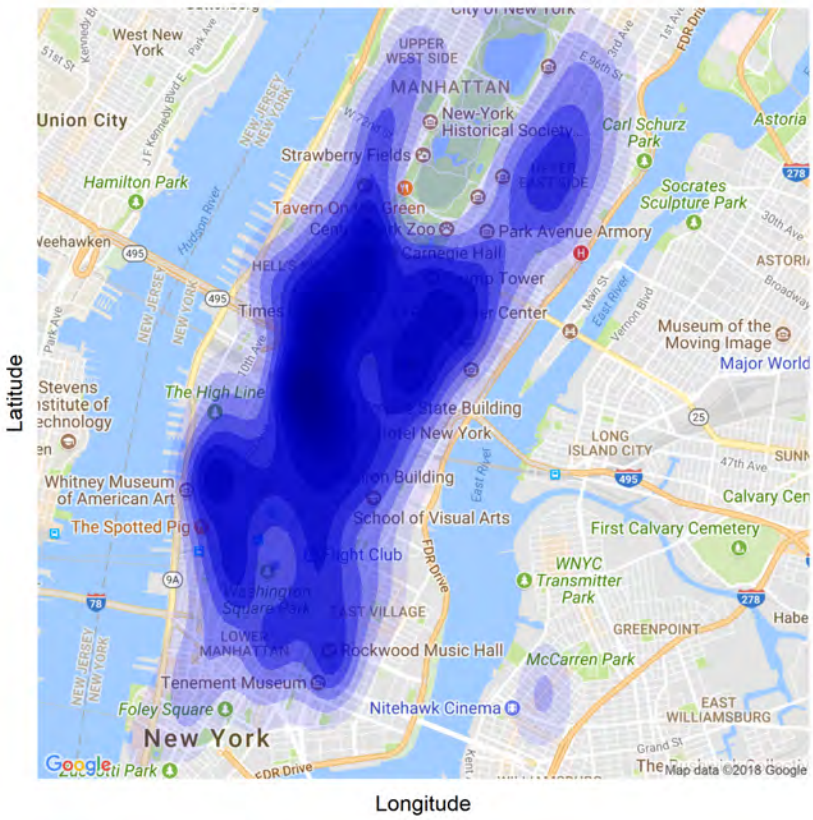


Figure 5: Density of pickup location of the top 10% taxis after 5:00PM till 8:00AM next morning. Darker points correspond to points of higher density.

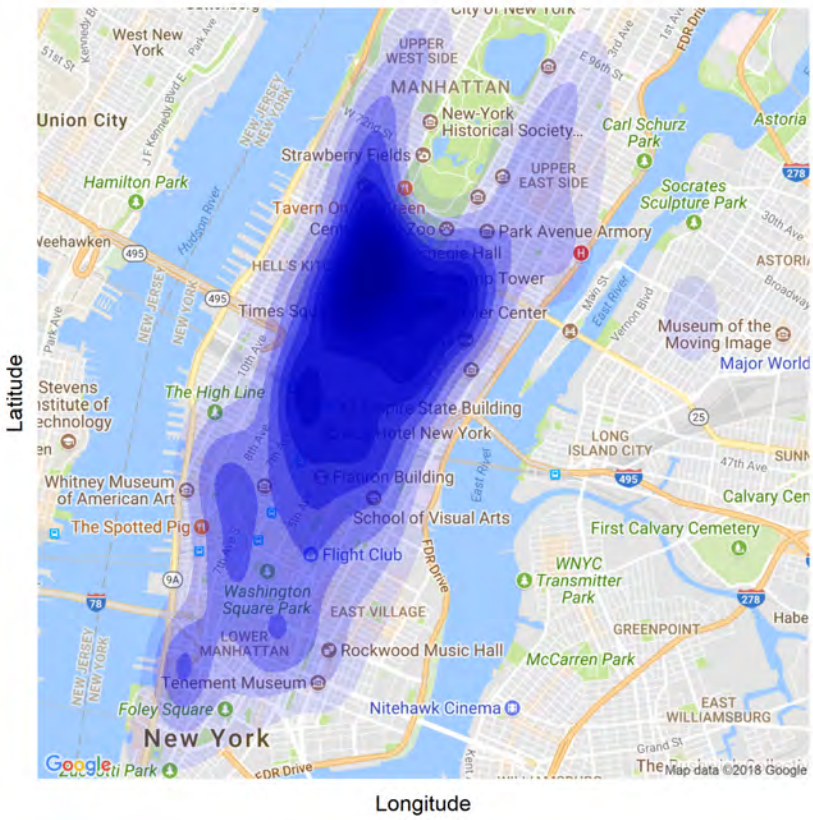


Figure 6: Density of pickup location of the bottom 10% taxis after 5:00PM till 8:00AM next morning. Darker points correspond to points of higher density.

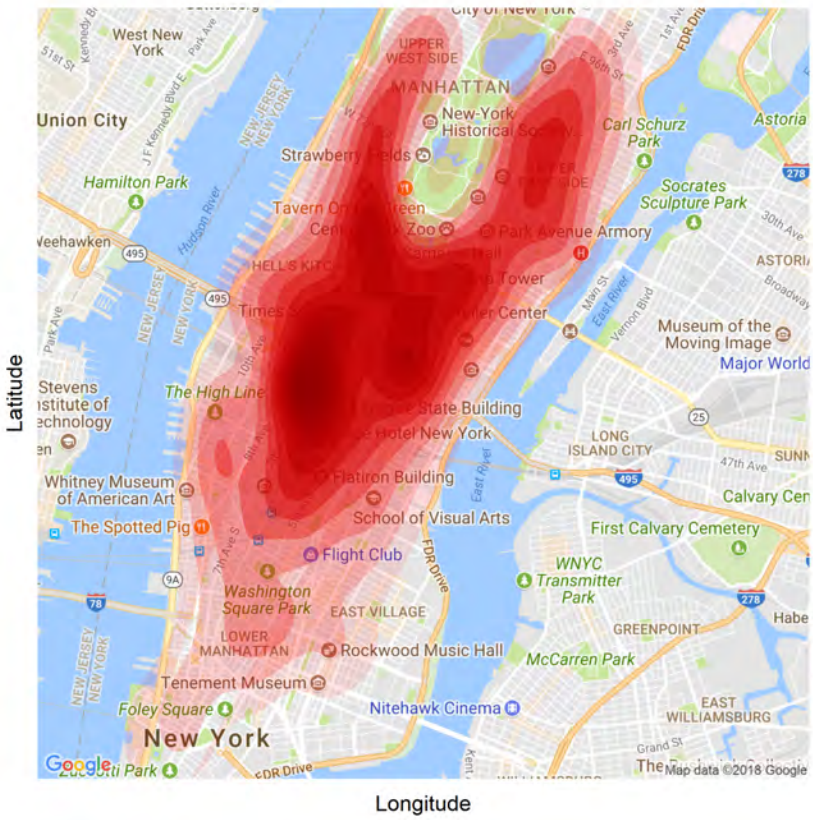


Figure 7: Density of pickup location of the top 10% taxis between 8:00AM and 5:00PM. Darker points correspond to points of higher density.

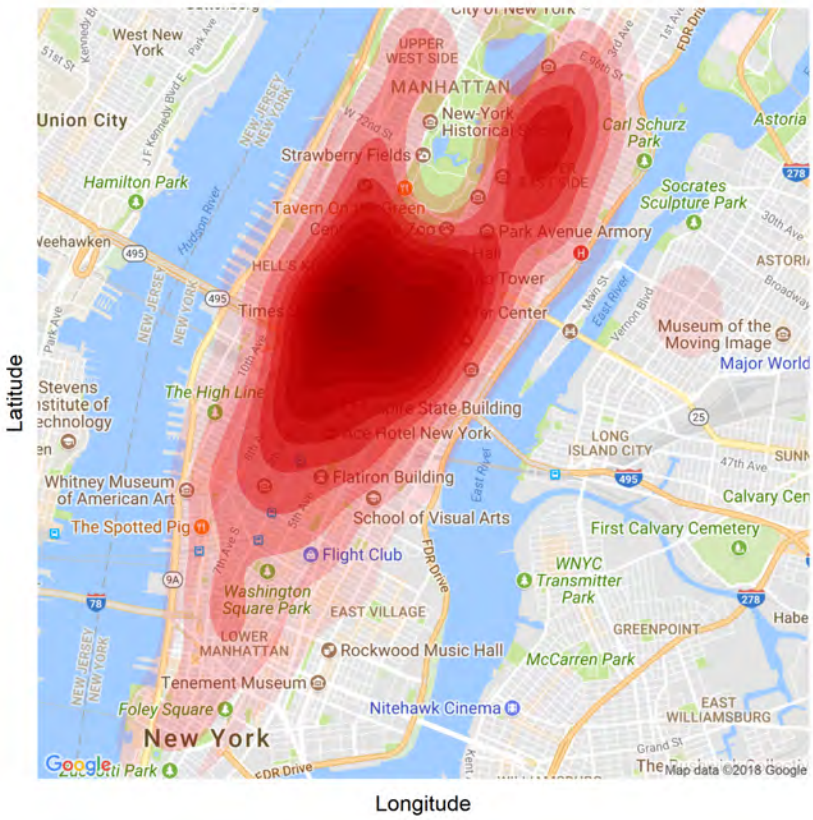


Figure 8: Density of pickup location of the bottom 10% taxis between 8:00AM and 5:00PM. Darker points correspond to points of higher density.