High dimensional panel mixed data sampling logit model

Lixiong Yang[‡]
School of Management, Lanzhou University, Lanzhou, China.
E-mail address: ylx@lzu.edu.cn

January 2, 2025

[‡]Corresponding author: Lixiong Yang. School of Management, Lanzhou University, 222 South Tianshui Road, Lanzhou 730000, China. E-mail address: ylx@lzu.edu.cn. Phone: +86 13669327501

High dimensional panel threshold mixed data sampling logit model

Abstract

This paper introduces a high dimensional panel threshold mixed data sampling probit (PT-MIDAS-probit or logit??) model, which allows for a covariate-dependent threshold and unobserved individual-specific threshold effects. We propose a sparse group LASSO (sg-LASSO) estimator which can not only select variables but also select a model between linear and threshold models based on CRE device. The extension to a setting with a high-frequency threshold variable and the extension to dynamic panels are also investigated.

Keywords : Logit; Panel threshold effect; Mixed data sampling (MIDAS); Estimation; Testing.

JEL classification : C12, C13, C15, C51, C53.

1 Introduction

High dimensional mixed frequency data involve certain data structures, an these structures can be typically represented by groups covering lagged dependent variables and groups of lags for a high frequency covariate (e.g., Babii et al., 2022). Taking these structures into account can improve the quality of the estimates, and the predictive performance in small samples; hence, Baii et al. (2022) propose a high dimensional mixed-frequency time series regression, and Babii et al. (2023) propose a high dimensional mixed-frequency panel data regression.

This paper adds to this strand of literature by proposing a high dimensional panel threshold mixed data sampling (HDPT-MIDAS) model, which allows for unobserved individual-specific threshold effects. To take advantage of the mixed frequency time series panel data structures, we suggest a sparse group LASSO (sg-LASSO) estimator simultaneously selects variables and selects a model between linear and threshold models, which is important as a panel threshold MIDAS model with a set of covariates can be overturned by a linear panel MIDAS model with a different set of covariates, and thus, the presence of threshold effect can be falsely rejected if independent variables are chosen incorrectly.

The remainder of this paper is organized as follows: Section 2 introduces the panel threshold mixed data sampling probit (PT-MIDAS-probit) model, develops the estimation procedure for model parameters, and constructs test statistics for relevance of high-frequency predictors, threshold effect, equal weighting scheme and equal forecasting accuracy. Section 3 extends the model to the framework with a covariate-dependent threshold (PCT-MIDAS-logit), and proposes model estimation procedure and test statistic for threshold constancy. In addition, we discuss the case with higher-frequency threshold variable, and the extension to dynamic setting. Section 4 presents Monte Carlo simulations to evaluate the finite sample properties of the estimation and testing procedures. Section 5 presents an empirical application, Section 6 concludes.

2 The model

Suppose that we have the mixed frequency data $\{y_{it}, \boldsymbol{x}_{it}^{(m)}, q_{it}, \boldsymbol{z}_{it}\}$, where y_{it}, q_{it} and $\boldsymbol{z}_{it}$ are low frequency variables observed at $t = 1, 2, \ldots, T$ for $i = 1, 2, ..., N$, $\boldsymbol{x}_{it}^{(m)} = (x_{1,it}^{(m)}, \ldots, x_{p_x,it}^{(m)})'$ is a p_x-dimensional vector of high frequency data observed m times between $t-1$ and t. Then, following the literature on binary choice models (e.g., Honore and Kyriazidou, 2000; Bartolucci and Nigro, 2012; Wooldridge and Zhu, 2020), we begin with the following binary data model with a latent index function allowing for unobserved heterogeneity, that is

$$y_{it}^* = \begin{cases} \boldsymbol{\alpha}_1^{*\prime}\boldsymbol{z}_{it} + \sum_{k=1}^{p_x} B(L^{1/m}; \boldsymbol{b}_{1k})x_{k,it} + \mu_{1i} + u_{it}, & \text{if } q_{it} \leq \gamma \\ \boldsymbol{\alpha}_2^{*\prime}\boldsymbol{z}_{it} + \sum_{k=1}^{p_x} B(L^{1/m}; \boldsymbol{b}_{2k})x_{k,it} + \mu_{2i} + u_{it}, & \text{if } q_{it} > \gamma \end{cases}, \tag{1}$$

with

$$y_{it} = I(y_{it}^* > 0), i = 1, ..., N; t = 1, ..., T, \tag{2}$$

where $I(.)$ is the indicator function, y_{it}, $\boldsymbol{z}_{it}$, $\boldsymbol{x}_{it}(\boldsymbol{\theta})$ and q_{it} are assumed to be weakly dependent variables, $\boldsymbol{z}_{it}$ is a p_z-dimensional vector of regressors sampled at the low frequency, y_{it} is a dummy variable, q_{it} is the threshold variable, and γ is the threshold parameter. μ_{1i} and μ_{2i} are fixed effects, and we allow that arbitrary correlations between the fixed effects and the regressors. $L^{j/m}$ is the high-frequency lag operator such that $L^{j/m}x_{k,i,t}^{(m)} = x_{k,i,t-(j/m)}^{(m)}$ for $j = 1, 2, ..., m$. $B(L^{1/m}; \boldsymbol{b}_{sk})$ is a high-frequency lag polynomial

$$B(L^{1/m}; \boldsymbol{b}_{sk})x_{k,t} = \frac{1}{m}\sum_{j=1}^{m} \omega(\frac{j-1}{m}; \boldsymbol{b}_{sk})x_{k,t-(j-1)/m}, \quad s = 1, 2 \tag{3}$$

where ω is a nonlinear function mapping the higher-frequency data into a low frequency, $\boldsymbol{b}_{sk} = (b_{sk,1}, b_{sk,2}, \ldots, b_{sk,L})'$ is a L-dimensional vector with $L \leq m$. Following Babii et al. (2022), we can approximate the weight function using orthogonal Legendre polynomials on $[0, 1]$ as $\omega(u; \beta_k) \approx \sum_{l=1}^{L} b_{sk,l}w_{l-1}(u), \quad u \in [0, 1]$, in which $w_l(u) = \frac{1}{l!}\frac{\mathrm{d}^l}{\mathrm{d}u^l}\left(u^2 - u\right)^l$. Then, we can rewrite (1) as

$$\begin{aligned} y_{it}^* \;&\approx\; \begin{cases} \boldsymbol{a}_1^{*\prime}\mathbf{z}_{it} + \frac{1}{m}\sum_{k=1}^{p_x}\sum_{j=1}^{m}\sum_{l=1}^{L} b_{1k,l}w_{l-1}(\frac{j-1}{m})x_{k,i,t-(j-1)/m} + \mu_{1i} + u_{it}, q_{it} \leq \gamma \\ \boldsymbol{a}_2^{*\prime}\mathbf{z}_{it} + \frac{1}{m}\sum_{k=1}^{p_x}\sum_{j=1}^{m}\sum_{l=1}^{L} b_{2k,l}w_{l-1}(\frac{j-1}{m})x_{k,i,t-(j-1)/m} + \mu_{2i} + u_{it}, q_{it} > \gamma \end{cases} \\ &=\; \begin{cases} \boldsymbol{a}_1^{*\prime}\mathbf{z}_{it} + \boldsymbol{\beta}_1^{*\prime}\boldsymbol{x}_{it} + \mu_{1i} + u_{it}, q_{it} \leq \gamma \\ \boldsymbol{a}_2^{*\prime}\mathbf{z}_{it} + \boldsymbol{\beta}_2^{*\prime}\boldsymbol{x}_{it} + \mu_{2i} + u_{it}, q_{it} > \gamma \end{cases}, \end{aligned} \tag{4}$$

in which $\boldsymbol{\beta}_s^* = (\boldsymbol{b}_{s1}', \boldsymbol{b}_{s2}', \ldots, \boldsymbol{b}_{sp_x}')'$ for $s = 1, 2$. $\boldsymbol{x}_{it} = (\widetilde{x}_{1it}, \widetilde{x}_{2it}, \ldots, \widetilde{x}_{p_xit})$, $\widetilde{x}_{kit} = \widetilde{X}_k W_k$ for $k = 1, 2, ..., p_x$, $\widetilde{X}_k = \left(x_{k,it}, x_{k,i,t-1/m}, x_{k,i,t-2/m}, \ldots, x_{k,t-(m-1)/m}\right)'$, $W = (w_{l-1}((j-1)/m)/m)_{m\times L}$ is an $m \times L$ matrix of weights.

Assume that u_{it} is the error term with the standard normal distribution. Then, our panel threshold mixed data sampling model can be defined by

$$E(y_{it} = 1|\boldsymbol{z}_{it}, \boldsymbol{x}_{it}, q_{it}, \mu_{1i}, \mu_{2i}) = \begin{cases} \Phi(\boldsymbol{\alpha}_1^{*\prime}\boldsymbol{z}_{it} + \boldsymbol{\beta}_1^{*\prime}\boldsymbol{x}_{it} + \mu_{1i}), & \text{if } q_{it} \leq \gamma \\ \Phi(\boldsymbol{\alpha}_2^{*\prime}\boldsymbol{z}_{it} + \boldsymbol{\beta}_2^{*\prime}\boldsymbol{x}_{it} + \mu_{2i}), & \text{if } q_{it} > \gamma \end{cases}, \tag{5}$$

in which $\Phi(\bullet)$ may denote the cumulative distribution function of the standard logistic distribution (PT-MIDAS-logit), or the cumulative distribution function of a standard normal density (PT-MIDAS-probit). When $\boldsymbol{\alpha}_1^* = \boldsymbol{\alpha}_2^*$, and $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_2^* = 0$, the PT-MIDAS-probit model degenerates to the model studied by Wooldridge and Zhu (2020), and the PT-MIDAS-logit model degenerates to the model studied by Audrino et al. (2019).

2.1 CRE device

As illustrated by Yu et al. (2022), the inclusion of unobserved individual-specific threshold effects can cuase the failure of the traditional estimation methods based on the within-group transformation. To overcome this difficulty, Yu et al. (2022) suggest to take the correlated random effects (CRE) model and use Chamberlain-Mundlak CRE device to control the endogeneity caused by the unobserved individual-specific effects. We develop an estimation method by adopting this strategy.

Following Mundlak (1978) and Yu et al. (2022), we assume that μ_{1i} and μ_{2i} in model (1) are given as follows

$$\mu_{li} = \boldsymbol{\psi}_l^{\prime}\mathbf{h}_i + a_{li}\ (l = 1, 2) \text{ with } E[a_{li}|W_i] = 0, \text{and } E[u_{it}|W_i] = 0, \tag{6}$$

where $\mathbf{h}_i = (\bar{\mathbf{w}}_i^{\prime}, \underline{\mathbf{h}}_i^{\prime})^{\prime}$ with $\bar{\mathbf{w}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{w}_{it}$, $\mathbf{w}_{it}$= $(\boldsymbol{z}_{it}^{\prime}, q_{it}, \boldsymbol{x}_{it}^{\prime}, \boldsymbol{s}_{it}^{\prime})^{\prime}$. $W_i = (\mathbf{w}_{i1}^{\prime}, ..., \mathbf{w}_{iT}^{\prime}, \underline{\mathbf{h}}_i^{\prime})^{\prime}$ in which $\underline{\mathbf{h}}_i$ contains the time-invariant variables such as the constant 1. Here, $\underline{\mathbf{h}}_i$ is used to control the time-invariant effect, and we allow that $Cov(a_{1i}, a_{2i}) \neq 0$, and $a_{1i} \neq a_{2i}$; thus, the correlation between α_{1i} and α_{2i} is through $\mathbf{w}_i$ or (a_{1i}, a_{2i}). In addition, a_{1i} and a_{2i} can be correlated with u_{it}. $a_{li}|W_i \sim N(0, \sigma_a^2)$.

Following the threshold literature (e.g., Hansen, 1999), we develop a two-step estimation procedure. For convenience, a more compacted form of the model defined in (18) is considered. Define $\boldsymbol{\alpha}^* = [\boldsymbol{\alpha}_2^{*\prime}, \boldsymbol{\alpha}_1^{*\prime} - \boldsymbol{\alpha}_2^{*\prime}]^{*\prime}$, $\boldsymbol{\beta}^* = [\boldsymbol{\beta}_2^{*\prime}, \boldsymbol{\beta}_1^{*\prime} - \boldsymbol{\beta}_2^{*\prime}]^{*\prime}$, $\boldsymbol{\psi} = [\boldsymbol{\psi}_2^{\prime}, \boldsymbol{\psi}_1^{\prime} - \boldsymbol{\psi}_2^{\prime}]^{\prime}$, $\boldsymbol{z}_{it}(\gamma) = [\boldsymbol{z}_{it}^{\prime}, \boldsymbol{z}_{it}^{\prime}I(q_{it} \leq \gamma)]^{\prime}$, , $\boldsymbol{x}_{it}(\boldsymbol{\theta}, \gamma) = [\boldsymbol{x}_{it}^{\prime}(\boldsymbol{\theta}), \boldsymbol{x}_{it}^{\prime}(\boldsymbol{\theta})I(q_{it} \leq \gamma)]^{\prime}$, and $\mathbf{h}_i(\gamma) = [\mathbf{h}_i^{\prime}, \mathbf{h}_i^{\prime}I(q_{it} \leq \gamma)]^{\prime}$. Then, the model defined in (1)-(3) can be rewritten as

$$\begin{aligned} P(y_{it} = 1|W_i) &= P[a_i + u_{it} > -(\boldsymbol{\alpha}^{*\prime}\boldsymbol{z}_{it}(\gamma) + \boldsymbol{\beta}^{*\prime}\boldsymbol{x}_{it}(\gamma) + \boldsymbol{\psi}^{\prime}\mathbf{h}_i(\gamma))|W_i] \\ &= \Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_{it}(\gamma) + \boldsymbol{\beta}^{**\prime}\boldsymbol{x}_{it}(\gamma) + \boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\gamma)), \end{aligned} \tag{7}$$

where $\boldsymbol{\alpha}^{**} = \frac{\boldsymbol{\alpha}^*}{\sqrt{1+\sigma_a^2}}$, $\boldsymbol{\beta}^{**} = \frac{\boldsymbol{\beta}^*}{\sqrt{1+\sigma_a^2}}$, $\boldsymbol{\psi}^{**} = \frac{\boldsymbol{\psi}}{\sqrt{1+\sigma_a^2}}$. As illustrated by Papke and Wooldriage (2008), Wooldridge and Zhu (2020), we can focus on these scaled parameters.

It is worth noting that the model in (20) would degenerates into a high dimensional panel probit model for any given γ, which will be used to develop an estimation strategy in the later.

Following the literature on logit models, the density of y_{it} given $\boldsymbol{z}_{it}(\gamma)$, $\boldsymbol{x}_{it}(\gamma)$ and $\mathbf{h}_i(\gamma)$ can be written as

$$
\begin{aligned}
& f(y_{it}|\boldsymbol{x}_t, W_i; \boldsymbol{\alpha}^{**}, \boldsymbol{\beta}^{**}, \boldsymbol{\psi}^{**}, \gamma) \\
= \; & p_{it}^{y_{it}}(1-p_{it})^{1-y_{it}} \\
= \; & [\Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_t(\gamma) + \boldsymbol{\beta}^{**\prime}\boldsymbol{x}_t(\boldsymbol{\gamma})) + \boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\gamma)]^{y_t}[1 - \Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_t(\gamma) + \boldsymbol{\beta}^{**\prime}\boldsymbol{x}_t(\gamma)) + \boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\gamma)]^{1-y_t}
\end{aligned}
\tag{8}
$$

and the log-likelihood function is

$$
\begin{aligned}
\ln L(\boldsymbol{\alpha}, \gamma) \;\equiv\; & \ln L(\boldsymbol{\alpha}^{**}, \boldsymbol{\beta}^{**}, \boldsymbol{\psi}^{**}, \gamma) \\
= \; & \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it} \ln[\Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_{it}(\gamma) + \boldsymbol{\beta}^{**\prime}\boldsymbol{x}_{it}(\gamma)) + \boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\gamma)] \\
+ \; & \sum_{i=1}^{N}\sum_{t=1}^{T}(1-y_{it}) \ln[1 - \Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_{it}(\gamma) + \boldsymbol{\beta}^{**\prime}\boldsymbol{x}_{it}(\gamma)) + \boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\gamma)].
\end{aligned}
\tag{9}
$$

2.1.1 Sparse Group LASSO (sg-LASSO) Estimator

Then, we suggest to use sparse-group LASSO (sg-LASSO) estimator of Babii et al. (2022) to achieve variable selection and estimation by assuming that most of the true slope coefficients are zero (sparsity). Unlike the standard group LASSO, the sg-LASSO contains the LASSO and the group LASSO as special cases, encourages sparsity between and within groups, which is particularly important in the MIDAS context as explained by Babii et al. (2022). Typically, all high-frequency lags of a single covariate constitute a group. Thus, the penalized likelihood function is given by

$$
\mathrm{Q}_{NT}\left(\boldsymbol{\alpha}, \gamma\right) \equiv \mathrm{Q}_{NT}\left(\boldsymbol{\alpha},^{**} \beta^{**}, \boldsymbol{\psi}^{**}, \gamma\right) = -\ln L(\boldsymbol{\alpha}, \gamma) + \lambda\Omega(\boldsymbol{\alpha}). \tag{10}
$$

And the sg-LASSO estimator is defined by

$$(\hat{\boldsymbol{\alpha}}, \hat{\gamma}) = \underset{\boldsymbol{\alpha}\in\mathcal{A},\gamma\in\boldsymbol{\Gamma}}{\arg\min}\, \mathrm{Q}_T(\boldsymbol{\alpha}, \gamma), \tag{11}$$

in which $\Omega(\boldsymbol{\alpha}) = a|\boldsymbol{\alpha}|_1 + (1-a)|\boldsymbol{\alpha}|_{2,1}$, $|\boldsymbol{\alpha}|_{2,1} = \sum_{G\in\mathcal{G}} |\boldsymbol{\alpha}_G|_2$, and $\mathcal{G}$ is a group structure which can be specified by the econometrician and applied researchers by a particular problem. a is a weight parameter determining the relative importance of the sparsity and the group structure. When $a = 1$, we obtain the LASSO estimator, while $a = 0$ leads to the group LASSO estimator.

For each fixed $\gamma \in \Gamma$, define the sg-LASSO solution as

$$\widehat{\boldsymbol{\alpha}}(\gamma) = \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\arg\min}\, \mathrm{Q}_T(\boldsymbol{\alpha}, \gamma). \tag{12}$$

Substituting (11) into (9) yields $\mathrm{Q}_T(\boldsymbol{\gamma}) := \mathrm{Q}_T(\widehat{\boldsymbol{\alpha}}(\gamma), \gamma)$, hence the threshold estimates can be defined as the minimum value of $\mathrm{Q}_T(\gamma)$, and we can implement this minimization using the grid search following the threshold literature (e.g. Hasen, 1999).

3 Extension to a more flexible setting

Suppose that we have the mixed frequency data $\{y_{it}, \boldsymbol{z}_{it}, \boldsymbol{x}_{it}^{(m)}, q_{it}, \boldsymbol{s}_{it}\}$, where $y_{it}, \boldsymbol{z}_{it}$, q_{it} and $\boldsymbol{s}_{it}$ are low frequency variables observed at $t = 1, 2, \ldots, T$ for $i = 1, 2, ..., N$, $\boldsymbol{x}_{it}^{(m)} = (x_{1,it}^{(m)}, \ldots, x_{p_x,it}^{(m)})'$ is a p_x-dimensional vector of high frequency data observed m times between $t-1$ and t. Then, following the literature on binary choice models (e.g., Honore and Kyriazidou, 2000; Bartolucci and Nigro, 2012; Wooldridge and Zhu, 2020), we begin with the following binary data model with a latent index function allowing for unobserved heterogeneity, that is

$$y_{it}^* = \begin{cases} \boldsymbol{\alpha}_1^{*\prime}\boldsymbol{z}_{it} + \sum_{k=1}^{p_x} B(L^{1/m}; \boldsymbol{b}_{1k})x_{k,it} + \mu_{1i} + u_{it}, & \text{if } q_{it} \leq \boldsymbol{\gamma}'\boldsymbol{s}_{it} \\ \boldsymbol{\alpha}_2^{*\prime}\boldsymbol{z}_{it} + \sum_{k=1}^{p_x} B(L^{1/m}; \boldsymbol{b}_{2k})x_{k,it} + \mu_{2i} + u_{it}, & \text{if } q_{it} > \boldsymbol{\gamma}'\boldsymbol{s}_{it} \end{cases}, \tag{13}$$

with

$$y_{it} = I(y_{it}^* > 0), i = 1, ..., N; t = 1, ..., T, \tag{14}$$

where $I(.)$ is the indicator function, y_{it}, $\boldsymbol{z}_{it}$, $\boldsymbol{x}_{it}(\boldsymbol{\theta})$ and q_{it} are assumed to be weakly dependent variables, $\boldsymbol{z}_{it}$ is a p_z-dimensional vector of regressors sampled at the low frequency, y_{it} is a dummy variable, q_{it} is the threshold variable, and γ is the threshold parameter. μ_{1i} and μ_{2i} are fixed effects, and we allow that arbitrary correlations between the fixed effects and the regressors. $\boldsymbol{s}_{it} = (1, \boldsymbol{s}'_{1,it})' \in \mathbb{R}^{k+1}, \boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}'_1)'$, and $\boldsymbol{s}_{it}$ is a $(k+1)$-dimensional vector of covariates shaping the regime separation. It is worth noting that the model with a high frequency threshold variable $q_{it}^{(m)}$ can be modeled as

$$y_{it}^* = \begin{cases} \boldsymbol{\alpha}_1^{*\prime} \boldsymbol{z}_{it} + \sum_{k=1}^{p_x} B(L^{1/m}; \boldsymbol{b}_{1k}) x_{k,it} + \mu_{1i} + u_{it}, & \text{if } q_{it}(\tilde{\theta}) \leq \gamma \\ \boldsymbol{\alpha}_2^{*\prime} \boldsymbol{z}_{it} + \sum_{k=1}^{p_x} B(L^{1/m}; \boldsymbol{b}_{2k}) x_{k,it} + \mu_{2i} + u_{it}, & \text{if } q_{it}(\tilde{\theta}) > \gamma \end{cases}, \tag{15}$$

where $q_{it}(\tilde{\theta}) = \sum_{j=1}^{J} \gamma_{qj} L^{j/m} q_{it}^{(m)}$. It is easily seen that, when we normalize $\gamma_{q1} = 1$, the model in (15) is essentially equal to model (13). $L^{j/m}$ is the high-frequency lag operator such that $L^{j/m} x_{k,i,t}^{(m)} = x_{k,i,t-(j/m)}^{(m)}$ for $j = 1, 2, ..., m$. $B(L^{1/m}; \boldsymbol{b}_{sk})$ is a high-frequency lag polynomial

$$B(L^{1/m}; \boldsymbol{b}_{sk}) x_{k,t} = \frac{1}{m} \sum_{j=1}^{m} \omega(\frac{j-1}{m}; \boldsymbol{b}_{sk}) x_{k,t-(j-1)/m}, \quad s = 1, 2, \tag{16}$$

where ω is a nonlinear function mapping the higher-frequency data into a low frequency, $\boldsymbol{b}_{sk} = (b_{sk,1}, b_{sk,2}, \ldots, b_{sk,L})'$ is a L-dimensional vector with $L \leq m$. Following Babii et al. (2022), we can approximate the weight function using orthogonal Legendre polynomials on $[0, 1]$ as $\omega(u; \beta_k) \approx \sum_{l=1}^{L} b_{sk,l} w_{l-1}(u), \quad u \in [0, 1]$, in which $w_l(u) = \frac{1}{l!} \frac{\mathrm{d}^l}{\mathrm{d}u^l} (u^2 - u)^l$. Then, we can rewrite (1) as

$$\begin{aligned} y_{it}^* \quad &\approx \quad \begin{cases} \boldsymbol{a}_1^{*\prime} \mathbf{z}_{it} + \frac{1}{m} \sum_{k=1}^{p_x} \sum_{j=1}^{m} \sum_{l=1}^{L} b_{1k,l} w_{l-1}(\frac{j-1}{m}) x_{k,i,t-(j-1)/m} + \mu_{1i} + u_{it}, q_{it} \leq \boldsymbol{\gamma}' \boldsymbol{s}_{it} \\ \boldsymbol{a}_2^{*\prime} \mathbf{z}_{it} + \frac{1}{m} \sum_{k=1}^{p_x} \sum_{j=1}^{m} \sum_{l=1}^{L} b_{2k,l} w_{l-1}(\frac{j-1}{m}) x_{k,i,t-(j-1)/m} + \mu_{2i} + u_{it}, q_{it} > \boldsymbol{\gamma}' \boldsymbol{s}_{it} \end{cases} \\ &= \quad \begin{cases} \boldsymbol{a}_1^{*\prime} \mathbf{z}_{it} + \boldsymbol{\beta}_1^{*\prime} \boldsymbol{x}_{it} + \mu_{1i} + u_{it}, q_{it} \leq \boldsymbol{\gamma}' \boldsymbol{s}_{it} \\ \boldsymbol{a}_2^{*\prime} \mathbf{z}_{it} + \boldsymbol{\beta}_2^{*\prime} \boldsymbol{x}_{it} + \mu_{2i} + u_{it}, q_{it} > \boldsymbol{\gamma}' \boldsymbol{s}_{it} \end{cases}, \end{aligned} \tag{17}$$

in which $\boldsymbol{\beta}_s^* = (\boldsymbol{b}'_{s1}, \boldsymbol{b}'_{s2}, \ldots, \boldsymbol{b}'_{sp_x})'$ for $s = 1, 2$. $\boldsymbol{x}_{it} = (\widetilde{x}_{1it}, \widetilde{x}_{2it}, \ldots, \widetilde{x}_{p_x it})$, $\widetilde{x}_{kit} = \widetilde{X}_k W_k$ for $k = 1, 2, ..., p_x$, $\widetilde{X}_k = \left(x_{k,it}, x_{k,i,t-1/m}, x_{k,i,t-2/m}, \ldots, x_{k,t-(m-1)/m}\right)'$, $W = (w_{l-1}((j - 1)/m)/m)_{m\times L}$ is an $m \times L$ matrix of weights.

Assume that u_{it} is the error term with the standard normal distribution. Then, our panel threshold mixed data sampling logit (PT-MIDAS-probit) model can be defined by

$$E(y_{it} = 1|\boldsymbol{z}_{it}, \boldsymbol{x}_{it}, q_{it}, \mu_{1i}, \mu_{2i}) = \begin{cases} \Phi(\boldsymbol{\alpha}_1^{*\prime}\boldsymbol{z}_{it} + \boldsymbol{\beta}_1^{*\prime}\boldsymbol{x}_{it} + \mu_{1i}), & \text{if } q_{it} \leq \boldsymbol{\gamma}'\boldsymbol{s}_{it} \\ \Phi(\boldsymbol{\alpha}_2^{*\prime}\boldsymbol{z}_{it} + \boldsymbol{\beta}_2^{*\prime}\boldsymbol{x}_{it} + \mu_{2i}), & \text{if } q_{it} > \boldsymbol{\gamma}'\boldsymbol{s}_{it} \end{cases}, \tag{18}$$

in which $\Phi(\bullet)$ denotes the cumulative distribution function of a standard normal density. When $\boldsymbol{\alpha}_1^* = \boldsymbol{\alpha}_2^*$, and $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_2^* = 0$, our model degenerates to the model studied by Wooldridge and Zhu (2020).

3.1 CRE device

As illustrated by Yu et al. (2022), the inclusion of unobserved individual-specific threshold effects can cuase the failure of the traditional estimation methods based on the within-group transformation. To overcome this difficulty, Yu et al. (2022) suggest to take the correlated random effects (CRE) model and use Chamberlain-Mundlak CRE device to control the endogeneity caused by the unobserved individual-specific effects. We develop an estimation method by adopting this strategy.

Following Mundlak (1978) and Yu et al. (2022), we assume that μ_{1i} and μ_{2i} in model (1) are given as follows

$$\mu_{li} = \boldsymbol{\psi}_l'\mathbf{h}_i + a_{li}\ (l = 1, 2) \text{ with } E[a_{li}|W_i] = 0, \text{and } E[u_{it}|W_i] = 0, \tag{19}$$

where $\mathbf{h}_i = (\bar{\mathbf{w}}_i', \underline{\mathbf{h}}_i')'$ with $\bar{\mathbf{w}}_i = \frac{1}{T}\sum_{t=1}^T \mathbf{w}_{it}$, $\mathbf{w}_{it}$= $(\boldsymbol{z}_{it}', q_{it}, \boldsymbol{x}_{it}', \boldsymbol{s}_{it}')'$. $W_i = (\mathbf{w}_{i1}', ..., \mathbf{w}_{iT}', \underline{\mathbf{h}}_i')'$ in which $\underline{\mathbf{h}}_i$ contains the time-invariant variables such as the constant 1. Here, $\underline{\mathbf{h}}_i$ is used to control the time-invariant effect, and we allow that $Cov(a_{1i}, a_{2i}) \neq 0$, and $a_{1i} \neq a_{2i}$; thus, the correlation between α_{1i} and α_{2i} is through $\mathbf{w}_i$ or $(a_{1i},\ a_{2i})$. In addition, a_{1i} and a_{2i} can be correlated with u_{it}. $a_{li}|W_i \sim N(0, \sigma_a^2)$.

Following the threshold literature (e.g., Hansen, 1999), we develop a two-step estimation procedure. For convenience, a more compacted form of the model defined in (18) is considered. Define $\boldsymbol{\alpha}^* = [\boldsymbol{\alpha}_2^{*\prime}, \boldsymbol{\alpha}_1^{*\prime} - \boldsymbol{\alpha}_2^{*\prime}]^{*\prime}$, $\boldsymbol{\beta}^* = [\boldsymbol{\beta}_2^{*\prime}, \boldsymbol{\beta}_1^{*\prime} - \boldsymbol{\beta}_2^{*\prime}]^{*\prime}$, $\boldsymbol{\psi} = [\boldsymbol{\psi}_2', \boldsymbol{\psi}_1' - \boldsymbol{\psi}_2']'$, $\boldsymbol{z}_{it}(\boldsymbol{\gamma}) = [\boldsymbol{z}_{it}', \boldsymbol{z}_{it}' I(q_{it} \leq \boldsymbol{\gamma}'\boldsymbol{s}_{it})]'$, , $\boldsymbol{x}_{it}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = [\boldsymbol{x}_{it}'(\boldsymbol{\theta}), \boldsymbol{x}_{it}'(\boldsymbol{\theta}) I(q_{it} \leq \boldsymbol{\gamma}'\boldsymbol{s}_{it})]'$, and $\mathbf{h}_i(\boldsymbol{\gamma}) = [\mathbf{h}_i', \mathbf{h}_i' I(q_{it} \leq \boldsymbol{\gamma}'\boldsymbol{s}_{it})]'$. Then, the model defined in (1)-(3) can be rewritten as

$$\begin{aligned} P(y_{it} = 1|W_i) &= P[a_i + u_{it} > -(\boldsymbol{\alpha}^{*\prime}\boldsymbol{z}_{it}(\boldsymbol{\gamma}) + \boldsymbol{\beta}^{*\prime}\boldsymbol{x}_{it}(\boldsymbol{\gamma}) + \boldsymbol{\psi}'\mathbf{h}_i(\boldsymbol{\gamma}))|W_i] \\ &= \Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_{it}(\boldsymbol{\gamma}) + \boldsymbol{\beta}^{**\prime}\boldsymbol{x}_{it}(\boldsymbol{\gamma}) + \boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\boldsymbol{\gamma})), \end{aligned} \tag{20}$$

where $\boldsymbol{\alpha}^{**} = \frac{\boldsymbol{\alpha}^*}{\sqrt{1+\sigma_a^2}}$, $\boldsymbol{\beta}^{**} = \frac{\boldsymbol{\beta}^*}{\sqrt{1+\sigma_a^2}}$, $\boldsymbol{\psi}^{**} = \frac{\boldsymbol{\psi}}{\sqrt{1+\sigma_a^2}}$ for the probit case; for the logit case, we can not derive the the precise scaling because the true response probability does not even have a logit form. Thus, in this latter case, the likelihood function is misspecified. In a low dimensional setting, Kwak et al. (2023) show that the CRE logit model is competitive in estimating average partial effect (APE) even if the likelihood function is misspecified; especially, the CRE logit model is robust to the typically used conditional independence assumption (i.e., given the observed covariates and unobserved heterogeneity, the binary responses are independent over time).

As illustrated by Papke and Wooldriage (2008), Wooldridge and Zhu (2020), we can focus on these scaled parameters.

It is worth noting that the model in (20) would degenerates into a high dimensional panel probit model for any given $\boldsymbol{\gamma}$, which will be used to develop an estimation strategy in the later.

Following the literature on logit models, the density of y_{it} given $\boldsymbol{z}_{it}(\boldsymbol{\gamma})$, $\boldsymbol{x}_{it}(\boldsymbol{\gamma})$ and $\mathbf{h}_i(\boldsymbol{\gamma})$ can be written as

$$\begin{aligned} & f(y_{it}|\boldsymbol{x}_t, W_i; \boldsymbol{\alpha}^{**}, \boldsymbol{\beta}^{**}, \boldsymbol{\psi}^{**}, \boldsymbol{\gamma}) \\ = \;& p_{it}^{y_{it}}(1 - p_{it})^{1-y_{it}} \\ = \;& [\Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_t(\boldsymbol{\gamma}) + \boldsymbol{\beta}^{**\prime}\boldsymbol{x}_t(\boldsymbol{\gamma})) + \boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\boldsymbol{\gamma})]^{y_t}[1 - \Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_t(\boldsymbol{\gamma}) + \boldsymbol{\beta}^{**\prime}\boldsymbol{x}_t(\boldsymbol{\gamma})) + \boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\boldsymbol{\gamma})]^{1-y_t} \end{aligned} \tag{21}$$

and the log-likelihood function is

$$\begin{aligned}
\ln L(\boldsymbol{\alpha},\boldsymbol{\gamma}) &\equiv \ln L(\boldsymbol{\alpha}^{**},\boldsymbol{\beta}^{**},\boldsymbol{\psi}^{**},\boldsymbol{\gamma}) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}\ln[\Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_{it}(\boldsymbol{\gamma})+\boldsymbol{\beta}^{**\prime}\boldsymbol{x}_{it}(\boldsymbol{\gamma}))+\boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\boldsymbol{\gamma})] \\
&+ \sum_{i=1}^{N}\sum_{t=1}^{T}(1-y_{it})\ln[1-\Phi(\boldsymbol{\alpha}^{**\prime}\boldsymbol{z}_{it}(\boldsymbol{\gamma})+\boldsymbol{\beta}^{**\prime}\boldsymbol{x}_{it}(\boldsymbol{\gamma}))+\boldsymbol{\psi}^{**\prime}\mathbf{h}_i(\boldsymbol{\gamma})]. \quad (22)
\end{aligned}$$

3.1.1 Sparse Group LASSO (sg-LASSO) Estimator

Then, we suggest to use sparse-group LASSO (sg-LASSO) estimator of Babii et al. (2022) to achieve variable selection and estimation by assuming that most of the true slope coefficients are zero (sparsity). Unlike the standard group LASSO, the sg-LASSO contains the LASSO and the group LASSO as special cases, encourages sparsity between and within groups, which is particularly important in the MIDAS context as explained by Babii et al. (2022). Typically, all high-frequency lags of a single covariate constitute a group. Thus, the penalized likelihood function is given by

$$\mathrm{Q}_{NT}\left(\boldsymbol{\alpha},\boldsymbol{\gamma}\right)\equiv \mathrm{Q}_{NT}\left(\boldsymbol{\alpha},^{**}\boldsymbol{\beta}^{**},\boldsymbol{\psi}^{**},\boldsymbol{\gamma}\right)=-\ln L(\boldsymbol{\alpha},\boldsymbol{\gamma})+\lambda\Omega(\boldsymbol{\alpha}). \quad (23)$$

And the sg-LASSO estimator is defined by

$$(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\gamma}})=\underset{\boldsymbol{\alpha}\in\mathcal{A},\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}{\arg\min}\,\mathrm{Q}_{T}\left(\boldsymbol{\alpha},\boldsymbol{\gamma}\right), \quad (24)$$

in which $\Omega(\boldsymbol{\alpha})=a|\boldsymbol{\alpha}|_1+(1-a)|\boldsymbol{\alpha}|_{2,1}$, $|\boldsymbol{\alpha}|_{2,1}=\sum_{G\in\mathcal{G}}|\boldsymbol{\alpha}_G|_2$, and $\mathcal{G}$ is a group structure which can be specified by the econometrician and applied researchers by a particular problem. a is a weight parameter determining the relative importance of the sparsity and the group structure. When $a=1$, we obtain the LASSO estimator, while $a=0$ leads to the group LASSO estimator.

For each fixed $\boldsymbol{\gamma}\in\boldsymbol{\Gamma}$, define the sg-LASSO solution as

$$\widehat{\boldsymbol{\alpha}}(\boldsymbol{\gamma})=\underset{\boldsymbol{\alpha}\in\mathcal{A}}{\arg\min}\,\mathrm{Q}_{T}\left(\boldsymbol{\alpha},\boldsymbol{\gamma}\right). \quad (25)$$

Substituting (11) into (9) yields $\mathrm{Q}_T(\boldsymbol{\gamma}) := \mathrm{Q}_T(\widehat{\boldsymbol{\alpha}}(\boldsymbol{\gamma}), \boldsymbol{\gamma})$, hence the threshold estimates can be defined as the minimum value of $\mathrm{Q}_T(\boldsymbol{\gamma})$. A popular method in the threshold literature for this minimization is the grid search (e.g. Hasen, 1999, Lee et al., 2016); however, the grid search approach performs well only when the number of variables shaping the threshold is very small, and it may be computationally troublesome when the dimension of $\boldsymbol{\gamma}$ is large. Following Yu and Fan (2021), we suggest an algorithm based on the MCMC algorithm to improve computational efficiency in estimating the threshold parameters.

Algorithm 1. Estimation based on the MCMC technique.

Step 1: For any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, define

$$P(\boldsymbol{\gamma}) = \frac{\exp\{-\mathrm{Q}_T(\boldsymbol{\gamma})\} I(\boldsymbol{\gamma} \in \boldsymbol{\Gamma})}{\int_{\boldsymbol{\Gamma}} \exp\{-\mathrm{Q}_T(\boldsymbol{\gamma})\} d\boldsymbol{\gamma}} \tag{26}$$

which is a quasi-posterior of $\boldsymbol{\gamma}$ with a uniform prior on $\boldsymbol{\Gamma}$.

Step 2: Use the MCMC technique to draw a Markov chain

$$S = \left(\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, \cdots \boldsymbol{\gamma}^{(B)}\right). \tag{27}$$

whose marginal density is approximately given by $P(\boldsymbol{\gamma})$.

Step 3: For each $\boldsymbol{\gamma}^{(b)}, b = 1, 2, \ldots, B$, calculate $\mathrm{Q}_T(\boldsymbol{\gamma}^{(b)})$. Define the initial estimates as $\widehat{\boldsymbol{\gamma}}_I = \underset{\boldsymbol{\gamma} \in S}{\operatorname{argmin}} \mathrm{Q}_T(\boldsymbol{\gamma})$, which may be a set of $\boldsymbol{\gamma}$ values. Then, we can define $\widehat{\boldsymbol{\gamma}}$ as the mode of the points in $\widehat{\boldsymbol{\gamma}}_I$.

Step 4: If desired, update the parameter space in Step 1 from $\boldsymbol{\Gamma}$ to a neighbourhood of the initial estimates of $\widehat{\boldsymbol{\gamma}}_I$. Repeat *Steps 2 and 3* to obtain an updated set of $\boldsymbol{\gamma}$ estimation, say, $\widehat{\boldsymbol{\gamma}}_U$. Then $\widehat{\boldsymbol{\gamma}}$ is defined as the average of the points in $\widehat{\boldsymbol{\gamma}}_U$.

4 Monte Carlo simulations

In this section, we conduct two Monte Carlo experiments to examine the finite sample performance of the estimation procedure proposed in section 2. The following data generating process (DGP) is given as

$$
\begin{aligned}
y_{it}^{*} &\approx \begin{cases} \boldsymbol{a}_1^{*\prime}\mathbf{z}_{it} + \frac{1}{m}\sum_{k=1}^{p_x}\sum_{j=1}^{m}\sum_{l=1}^{L} b_{1k,l}w_{l-1}(\frac{j-1}{m})x_{k,i,t-(j-1)/m} + \mu_{1i} + u_{it}, q_{it} \leq \gamma \\ \boldsymbol{a}_2^{*\prime}\mathbf{z}_{it} + \frac{1}{m}\sum_{k=1}^{p_x}\sum_{j=1}^{m}\sum_{l=1}^{L} b_{2k,l}w_{l-1}(\frac{j-1}{m})x_{k,i,t-(j-1)/m} + \mu_{2i} + u_{it}, q_{it} > \gamma \end{cases} \\
&\equiv \begin{cases} \mathbf{a}_1'\mathbf{z}_{it} + \mathbf{b}_1'\boldsymbol{x}_{it} + \mu_{1i} + u_{it}, q_{it} \leq \gamma \\ \mathbf{a}_2'\mathbf{z}_{it} + \mathbf{b}_2'\boldsymbol{x}_{it} + \mu_{2i} + u_{it}, q_{it} > \gamma \end{cases}, \qquad (28)
\end{aligned}
$$

with

$$
y_{it} = I(y_{it}^{*} > 0), i = 1, ..., N; t = 1, ..., T, \qquad (29)
$$

where $\mathbf{x}_{it} = (x_{1it}, x_{2it}, \ldots, x_{kit})'$, $x_{kit} = x_{ki}W$, $x_{ki} = \left(x_{k,it}, x_{k,it-1/m}, x_{k,it-2/m}, \ldots, x_{k,it-(m-1)/m}\right)'$, W is an $m \times L$ matrix of weights whose (j,l)-th element is $w_l((j-1)/m)/m$. $\mathbf{z}_{it} \sim i.i.d.N(0,1)$. Following Babii et al. (2021), the DGP for high-frequency covariates $\left\{x_{k,it-(j-1)/m} : j \in [m], k = 1, ..., p_x\right\}$ are generated as follows: p_x $i.i.d$ realizations of the univariate autoregressive process $x_{ih} = \tau x_{ih-1} + \epsilon_{ih}$, where $\tau = 0.2$ and $\epsilon_h \overset{i.i.d}{\sim} N(0, \sigma_\epsilon^2)$, $\sigma_\epsilon^2 = 1$. We initiate the process as $x_0 \sim N(0, \sigma^2/(1-\tau^2))$ and treat the first 200 observations as burn-in. We consider quarterly and monthly data in our simulation, and use four quarters of data for the high-frequency regressors so that $m = 12$. We use a dictionary consisting of Legendre polynomials up to degree $L = 3$ shifted to the $[0,1]$ interval and the groups are defined as in section 2. The number of relevant high-frequency regressors is set at 3. We select the tuning parameter λ, weight parameter α and thresholded constant C_5 using BIC.[1] In the simulations, we set $\boldsymbol{\beta}_1' = (a_1, \boldsymbol{b}_1') = (1, 0, 1, 1, 0, \ldots, 0)$, $\boldsymbol{\beta}_2' = (a_2, \boldsymbol{b}_2') = (2, 2, 0, 2, 0, 0, \ldots, 0)$. So $\boldsymbol{\delta}' = \boldsymbol{\beta}_1' - \boldsymbol{\beta}_2' = (-1, -2, 1, -1, 0, \ldots, 0)$, and $\boldsymbol{\beta}^{*'} = (\boldsymbol{\beta}_1', \boldsymbol{\delta}')$, $(\gamma_0, \boldsymbol{\gamma}_1') = (0.1, 0.5, 0.6, 0, 0, 0)$, and run experiments on a range of sample sizes ($T = 100, 200, 500$). The number of

[1] The most commonly used methods for selecting regularization parameters include cross-validation and BIC. We choose to use BIC to select the tuning parameter λ and weight parameter ρ, as the computation cost is heavy in our covariate-dependent threshold model, and cross-validation would suffer from heavy computational cost problem in our model. We calculate BIC (which is implemented by the R package *midasml*) following Hirose et al. (2011). We search the weight parameter ρ over a grid from 0 to 1 by steps of 0.1. We select λ by grid search over 100 equispaced points on $[0, \lambda_{\max}]$ for any give $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, in which $\lambda_{\max}$ is calculated following Striaukas et al. (2022): (1) If $\rho = 1, \lambda_{\max} = \lambda_{\text{Lasso}} := \frac{1}{T}|\mathbf{X}(\boldsymbol{\gamma})'\mathbf{y}|_\infty$; (2) If $\rho = 0$, $\lambda_{\max} = \lambda_{\text{g-Lasso}} := \max_{G \in \mathcal{G}} \frac{1}{T}|\mathbf{X}_G(\boldsymbol{\gamma})'\mathbf{y}|_2/PF_G$, where $\mathbf{X}_G(\boldsymbol{\gamma})$ is the sub-matrix of $\mathbf{X}(\boldsymbol{\gamma})$ whose columns are corresponding to the G-th group, and PF_G is the penalty factor for the G-th group, which is defined by the square root of the length of each group; (3) If $0 < \rho < 1$, $\lambda_{\max} = \max_{G \in \mathcal{G}} \left(\frac{\lambda_{\text{Lasso}}}{\rho} \vee \frac{\frac{1}{T}|\mathbf{X}_G(\boldsymbol{\gamma})'\mathbf{y} - \rho PF_G \mathbf{1}_G|_2}{(1-\rho)PF_G} \right)$, where $\mathbf{1}_G$ is the all-ones vector matching the length of group G.

replications is 100 to save simulation time.

The simulation results are reported in Table 1. We can see that the mean of the threshold parameter is very close to the true value 0.1 for all considered combinations of N and T, and the parameter estimation errors of sg-LASSO and thresholded sg-LASSO (thsg-LASSO-TMIDAS) are fairly close (according to the statistics $E|\hat{\beta}^* - \beta_0^*|_1$ and $E|\hat{\beta}^* - \beta_0^*|_\infty$). The perfect selection (PS) statistic shows that perfect variable selection is more than 90% of the iteration when $N \times T \geq 1000$ for thsg-LASSO, but not for sg-LASSO, indicating the thresholded sg-LASSO works well in selecting variables, consistent with our theory; moreover, the simulation indicates that sg-LASSO and thsg-LASSO share a similar prediction performance according to the prediction risk. Additionally, we also report the number of zero parameters falsely included (FZ) and the number of nonzero parameters falsely excluded (FNZ), and find that thsg-LASSO can reduce the possibility of irrelevant variables falsely included according the FNZ statistic. In sum, these simulation results support that our proposed thsg-LASSO approach outperforms sg-LASSO as thsg-LASSO can achieve more accurate variable selection without the loss of the prediction and estimation accuracy; this is important as we can identify the main factors determining the forecast in applications, and hence the thsg-LASSO can be viewed as an interpretable machine learning approach.

T	N		γ	PS	FZ	FNZ	$E\|\hat{\beta}^* - \beta_0^*\|_1$	$E\|\hat{\beta}^* - \beta_0^*\|_\infty$	pred1
2	100	sg-LASSO	0.099	0.050	0.280	3.760	29.426	5.476	0.168
		thsg-LASSO		0.350	0.590	1.240	27.942	5.476	0.162
	200	sg-LASSO	0.100	0.060	0.000	2.940	26.436	4.934	0.143
		thsg-LASSO		0.690	0.080	0.280	25.496	4.934	0.139
	500	sg-LASSO	0.100	0.030	0.000	3.470	28.937	5.162	0.131
		thsg-LASSO		0.910	0.010	0.080	28.182	5.162	0.129
5	100	sg-LASSO	0.100	0.030	0.000	4.392	29.879	5.165	0.134
		thsg-LASSO		0.773	0.000	0.148	29.141	5.165	0.134
	200	sg-LASSO	0.100	0.030	0.000	3.480	28.781	5.080	0.137
		thsg-LASSO		0.920	0.000	0.130	28.111	5.080	0.137
	500	sg-LASSO	0.100	0.030	0.000	3.680	29.698	5.135	0.133
		thsg-LASSO		0.930	0.000	0.070	29.195	5.135	0.132