



Unknown latent structure and inefficiency in panel stochastic frontier models

Levent Kutlu¹ · Kien C. Tran ² · Mike G. Tsionas³

Published online: 25 July 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

This paper extends the fixed effect panel stochastic frontier models to allow group heterogeneity in the slope coefficients. We propose the first-difference penalized maximum likelihood (FDPML) and control function penalized maximum likelihood (CFPML) methods for classification and estimation of latent group structures in the frontier as well as inefficiency. Monte Carlo simulations show that the proposed approach performs well in finite samples. An empirical application is presented to show the advantages of data-determined identification of the heterogeneous group structures in practice.

Keywords Classification · Fixed effect · Group heterogeneity · Panel stochastic frontier · Penalized control function maximum likelihood · Penalized first-difference maximum likelihood.

JEL classification C13 · C23 · C36.

1 Introduction

Unobserved heterogeneity plays an important role in the estimation of panel stochastic frontier models, and since heterogeneity is a latent feature of the data, its extent is unknown a priori in empirical practices. Therefore, neglecting unobserved heterogeneity in the data can lead to inconsistent estimation of frontier parameters, and misleading inferences and predictions of the inefficiency indices. Greene (2005a; 2005b) pointed out that if individual-specific heterogeneity is not adequately controlled for, the predicted inefficiency may be picking up some, if not all, of the individual-specific heterogeneity. Thus, recent work on panel stochastic frontier models have focused on how to control for unobserved heterogeneity (see, for example, Guan et al. (2009), Wang and Ho (2010), Colombi et al. (2014), Chen et al. (2014), Kumbhakar et al. (2014),

Tsionas and Kumbhakar (2014), Kutlu et al. (2019), and Kutlu and Tran (2019) for reference therein).

However, all the papers mentioned above, except Kutlu et al. (2019), typically assumed complete slope homogeneity (i.e., the frontier parameters are the same across individuals), and unobserved heterogeneity is modeled through individual-specific effects. Kutlu et al. (2019) allow only a subset of variables to have different slopes for individuals. Tsionas (2002) considered a pooled panel stochastic frontier model that allowed for slope heterogeneity where the frontier parameters are random so that they are completely different for different individuals; however, he assumed that the intercept term is common for all individuals over time, and hence he did not control for individual-specific effects. Whilst allowing for cross-section slope heterogeneity may help to improve on the specification bias of the frontier, its main disadvantage is the loss of power due to cross-section averaging in the estimation of the response patterns that may be common across individuals (i.e., certain groups of individuals in the panel). Moreover, since the parameters are random, this model is subject to standard problems of random effects models, e.g., inconsistent parameter estimates when the slopes are correlated with the error term. Thus, for the panel stochastic frontier model, it is essential to control for unobserved heterogeneity in the data as well as for the potential heterogeneity

✉ Kien C. Tran
kien.tran@uleth.ca

¹ Department of Economics, University of Texas Rio Grande Valley, Edinburg, TX, USA

² Department of Economics, University of Lethbridge, 4401 University Drive W, Lethbridge, AB, Canada

³ Lancaster University Management School, Lancaster, UK

in the response mechanisms that characterizes within the model.

In this paper, we extend previous work on panel stochastic frontier models, and specifically the Wang and Ho (2010) model, to allow for both unobserved heterogeneity via individual-specific effects and for group heterogeneity in the slope parameters. In the standard panel regression models with individual-specific effects, Su et al. (2016) develop a new estimation and inference procedure when the regression parameters are heterogeneous across groups. They treat individual group membership as unknown and the group classification is determined empirically. We follow their lead in this paper and extend their approach to panel stochastic frontier models. Specifically, we use first-differencing transformation to remove the fixed effect, and then propose a penalized maximum likelihood estimation procedure to consistently estimate the frontier parameters, classification of groups and their memberships as well as technical inefficiency scores. Moreover, we also extend the model to allow for some or all regressors to be endogenous and propose a different estimation approach for which we term as penalized control function maximum likelihood.

Our proposed model is related to the class of the metafrontier model developed by Battese et al. (2004) and among others, in the sense that both models consider the group-wise heterogeneity in response coefficients. However, our proposed model differs from the metafrontier literature in the following ways. First, the number of groups is specified a priori in the metafrontier model whilst they are determined endogenously based on the data in our model. Second, unobserved individual-specific effects can be different even among the firms within each group, but the metafrontier model does not allow for such effects and our model certainly allows for it. Finally, the metafrontier model assumes there exists a deterministic frontier which envelopes the groups' frontiers. However, we do not make such an assumption in our model due to the presence of the general (unobserved) individual-specific effects. Nevertheless, we believe that it can be readily extended to allow for a such deterministic frontier. Thus, our proposed model is more general and flexible.

The remainder of the paper is organized as follows. Section 2 outlines the model and estimation procedure. Specifically, we consider first-differencing transformations in the estimation procedures to remove the individual-specific effects and show how to determine the number of groups, classification of group membership, and prediction of technical inefficiency scores. Section 3 extends the model to accommodate for endogenous regressors. A detailed computational algorithm of the proposed approach is given in Section 4. Section 5 provides some Monte Carlo simulations to examine the finite sample performance of the

proposed estimators. An empirical application is presented in Section 6, and finally, Section 7 concludes the paper.

2 The model with exogenous regressors

In order to fix the ideas, we will describe the estimation of a production function. However, with standard minor modifications in the model, a cost function can be estimated as well. Suppose we observed a panel data $\{(x_{it}, y_{it}): i = 1, \dots, N; t = 1, \dots, T\}$ where y_{it} is a scalar representing (log) output of firm i at time t and x_{it} is $k \times 1$ vector of (log) inputs of firm i at time t . The fixed effects stochastic frontier model with group-specific pattern heterogeneity can be written as:

$$y_{it} = \alpha_i + x'_{it}\beta_i + v_{it} - u_{it}, \quad (1a)$$

$$v_{it} \sim N(0, \sigma_v^2), \quad (1b)$$

$$u_{it} = h_{it}u_i^*, \quad (1c)$$

$$h_{it} = h(q'_{it}\delta), \quad (1d)$$

$$u_i^* \sim N^+(0, \sigma_u^2), \quad (1e)$$

where α_i are scalar individual effects, β_i is a $k_x \times 1$ vector of parameters of interest, v_{it} is a random symmetric error term representing factors that are beyond the firm's control, $u_{it} \geq 0$ is a one-sided stochastic variable representing a technical inefficiency component, h_{it} is a positive function of a $k_q \times 1$ vector of non-stochastic inefficiency determinants (q_{it}), and δ is a $k_q \times 1$ vector of unknown parameters. We assume that the random variable u_i^* is independent of all T observations on v_{it} , and both u_i^* and v_{it} are independent of all T observations on $\{x_{it}, q_{it}\}$. For identification purposes, we further assume that neither x_{it} nor q_{it} contains a constant term, and at least one variable in q_{it} is not time-invariant. Following Su et al. (2016) (hereafter SSP), we allow for β_i to follow a group-specific pattern of the general form:

$$\beta_i = \begin{cases} \gamma_1 & \text{if } i \in G_1 \\ \vdots & \\ \gamma_{J_0} & \text{if } i \in G_{J_0} \end{cases}, \quad (2)$$

where in (2), for any $j \neq l$, $\gamma_j \neq \gamma_l$, $G_j \cap G_l = \emptyset$, and $\cup_{j=1}^{J_0} G_j = \{1, 2, \dots, N\}$. Let $N_j = \#G_j$, $j = 1, \dots, J_0$, denote the cardinality of the set G_j . To simplify the discussion, for now we assume that the number of groups, J_0 is known and fixed but each individual's group membership is unknown. In addition, we implicitly assume that individual group membership does not vary over time. The above model can be thought of as an extension of the models of Wang and Ho (2010) and Chen et al. (2014), which allows for the

slopes to vary according to a specific group. Note that in (1d) we assume that δ is the same for all i . Allowing for δ to vary with i would complicate the analysis further since the group classification is now needed to be done simultaneously. It is beyond the scope of this paper and we will leave it for future research.

2.1 First-difference penalized likelihood (FDPL) estimation

Following Wang and Ho (2010), we first introduce the following notations. For any random variable r_{it} , let $\Delta r_{it} = r_{it} - r_{it-1}$, and $\Delta \tilde{r}_i = (\Delta r_{i2}, \dots, \Delta r_{iT})'$ for $i = 1, \dots, N$. In general, with a slight abuse of notation, $\Delta \tilde{r}_i$ represents a matrix with relevant columns obtained from each variable. For example, $\Delta \tilde{x}_i$ is a $(T - 1) \times k_x$ matrix. Then, taking the first difference of Eqs. (1a)–(1c), the model becomes:

$$\Delta \tilde{y}_i = \Delta \tilde{x}_i \beta_i + \Delta \tilde{v}_i - \Delta \tilde{u}_i, \tag{3a}$$

$$\Delta \tilde{v}_i \sim N(0, \sigma_v^2 \Sigma), \tag{3b}$$

$$\Delta \tilde{u}_i = \Delta \tilde{h}_i u_i^*, \tag{3c}$$

$$u_i^* \sim N^+(0, \sigma_u^2), \tag{3d}$$

where in Eq. (3b), the first-difference of v_{it} introduces correlations of Δv_{it} within the i^{th} panel and the $(T - 1) \times (T - 1)$ matrix Σ is given by:

$$\Sigma = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix}. \tag{4}$$

Note that, after the transformation, Eq. (3d) is the same as Eq. (1e) implying that the half-normality of u_i^* is unaffected by the transformation, and this is the key aspect of the model that leads to a tractable derivation of the likelihood function. Under the above assumptions, the marginal log-likelihood function of panel i in the model is given by:

$$\begin{aligned} \log L_{i,NT} = & -\frac{1}{2} \left[(T - 1)(\log(2\pi) + \log(\sigma_v^2)) + \log(T) + \frac{\Delta \tilde{e}' \Sigma^{-1} \Delta \tilde{e}_i}{\sigma_v^2} \right] \\ & + \frac{1}{2} \frac{\mu_i^2}{\sigma_u^2} + \log \left[\sigma_u \Phi \left(\frac{\mu_i}{\sigma_u} \right) \right] - \log \left(\frac{\sigma_u}{2} \right), \end{aligned} \tag{5}$$

where $\mu_{i*} = -\frac{\Delta \tilde{e}' \Sigma^{-1} \Delta \tilde{h}_i}{\Delta \tilde{h}' \Sigma^{-1} \Delta \tilde{h}_i + \sigma_v^2 / \sigma_u^2}$, $\sigma_{i*}^2 = \frac{\sigma_v^2}{\Delta \tilde{h}' \Sigma^{-1} \Delta \tilde{h}_i + \sigma_v^2 / \sigma_u^2}$, $\Delta \tilde{e}_i = \Delta \tilde{y}_i - \Delta \tilde{x}_i \beta_i$, and $\Phi(\cdot)$ is the standard normal CDF.

Let $\boldsymbol{\gamma} = (\gamma'_1, \dots, \gamma'_J)'$, $\boldsymbol{\beta} = (\beta'_1, \dots, \beta'_N)'$, and $\boldsymbol{\theta} = (\delta', \sigma_v^2, \sigma_u^2)'$. We estimate $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\theta}$ by maximizing the following FDPL criterion:

$$\log L_{NT, \eta_1}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^N \log L_{i,NT} + \frac{\eta_1}{N} \sum_{i=1}^N \prod_{j=1}^{J_0} \|\beta_i - \gamma_j\|, \tag{6}$$

where $\eta_1 = \eta_{1,NT}$ is a tuning parameter, $\|A\|$ denotes the Frobenius norm, and the second term on the right-hand side of Eq. (6) represents a penalty term. As in SSP, the penalty term takes a mixed additive-multiplicative form, which is different from the traditional penalized estimation (where the additive penalty term is normally used). The additive component is needed for the identification of $\{\beta_i\}$ and $\{\gamma_j\}$ jointly; and the main reason for the inclusion of the multiplicative term is that, for each i , β_i can take any one of the J_0 unknown values, $\gamma_1, \dots, \gamma_J$, and it is not known a priori to which point β_i should shrink. Maximizing Eq. (6) produces FDPL or Classifier-Lasso (C-Lasso) estimates $\hat{\boldsymbol{\beta}} = (\hat{\beta}'_1, \dots, \hat{\beta}'_N)'$, $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}'_1, \dots, \hat{\gamma}'_J)'$, and $\hat{\boldsymbol{\theta}} = (\hat{\delta}', \hat{\sigma}_v^2, \hat{\sigma}_u^2)'$ of $\boldsymbol{\gamma} = (\gamma'_1, \dots, \gamma'_J)'$, $\boldsymbol{\beta} = (\beta'_1, \dots, \beta'_N)'$, and $\boldsymbol{\theta} = (\delta', \sigma_v^2, \sigma_u^2)'$, respectively¹.

2.2 Determination of the number of groups

The discussion in the previous sub-section assumes that the number of groups J_0 is known a priori. However, in practice, the exact number of groups is rarely known and must be estimated. In this sub-section, we show how to determine the number of groups using an information criterion (IC) procedure. Our approach follows along the argument given in SSP. First, we assume that J_0 is bounded from above by a finite integer J_{max} . For a given $J \in \{1, \dots, J_{max}\}$, let $\{\hat{\beta}_i(J, \eta_1), \hat{\gamma}_j(J, \eta_1)\}$ and $\hat{\boldsymbol{\theta}}$ denote the FDPL (or C-Lasso) estimators of $\{\beta_i, \gamma_j\}$ and $\boldsymbol{\theta}$ discussed above; and individual i is classified into group $\hat{G}_j(J, \eta_1)$ according to $\hat{G}_j(J, \eta_1) = \{i \in \{1, 2, \dots, N\} : \hat{\beta}_i(J, \eta_1) = \hat{\gamma}_j(J, \eta_1)\}$ for $j = 1, \dots, J$. Finally, let $\hat{G}(J, \eta_1) = \{\hat{G}_1(J, \eta_1), \dots, \hat{G}_J(J, \eta_1)\}$ and $\hat{\gamma}_{\hat{G}_j(J, \eta_1)}$ denote the post-FDPL (or post-Lasso) estimator. Then, we select J so that it minimizes the following IC:

$$IC_1(J, \eta_1) = \frac{2}{NT} \sum_{j=1}^J \sum_{i \in \hat{G}_j(J, \eta_1)} \log L_{i,NT}(w_{it}; \hat{\gamma}_{\hat{G}_j(J, \eta_1)}, \hat{\boldsymbol{\theta}}(\hat{\gamma}_{\hat{G}_j(J, \eta_1)})) + \lambda_{1,NT} K_1 J, \tag{7}$$

¹ C-Lasso is termed by SSP.

where $\lambda_{1,NT}$ is a tuning parameter, and $K_1 = k_x + k_q + 2$. That is, the number of groups, J is chosen such that $\hat{J}(\eta_1) = \arg \min_{1 \leq J \leq J_{\max}} IC_1(J, \eta_1)$.

Remark 1: As noted by SSP, the choice of the tuning parameter $\lambda_{1,NT}$ and $\eta_{1,NT}$ respectively, can play an important role in determining the correct number of groups and post-FDPL estimates in practice. Following SSP, we impose the following conditions on the tuning parameters $\lambda_{1,NT}$ and $\eta_{1,NT}$.

A.1: As $(N, T) \rightarrow \infty$, $\lambda_{1,NT} \rightarrow 0$ and $\lambda_{1,NT}NT \rightarrow \infty$.

A.2: As $(N, T) \rightarrow \infty$, (i) $T\eta_1^2/(\ln T)^{6+2\nu} \rightarrow \infty$ and $\eta_1(\ln T)^\nu \rightarrow 0$ for some $\nu > 0$; (ii) $N^{1/2}T^{-1}(\ln T)^9 \rightarrow 0$ and $N^2T^{1-q/2} \rightarrow c \in [0, \infty)$ for some $q \geq 6$.

The condition A.1 reflects the conditions for consistency of model selection, i.e., $\lambda_{1,NT}$ cannot shrink to zero too quickly or too slowly. Condition A.2 holds if $\eta_{1,NT} \propto T^{-a}$ for any $a \in (0, 1/2)$.

In practice, under A.1, we can fine-tune $\lambda_{1,NT}$ over a finite set $\Lambda_1 = \{\lambda_1 = \kappa_l(NT)^{-1/2}, l = 1, \dots, L\}$ for some $\kappa_l > 0$. Similarly, under A.2, we also suggest to fine-tune $\eta_{1,NT}$ over a finite set $\mathfrak{N}_1 = \{\eta_1 = c_l T^{-1/3}, c_l = c_0 \zeta^l, l = 1, \dots, L\}$ for some $c_0 > 0$ and $\zeta > 1$. In essence, these tuning parameters are analogous to the bandwidth selections in the kernel smoothing.

Remark 2: Under certain regularity conditions, SSP derive the asymptotic properties of the post-Lasso estimators including the oracle property for the non-stochastic frontier models. It can be shown that our proposed estimator satisfies the regularity conditions set out in SSP, and hence it is consistent, asymptotically normal, and achieves the oracle property as well². For inference purposes, it is important to recognize that our post-FDPL estimator belongs to the class of M-estimators, and hence the asymptotic variance has the form: $avar(\hat{\psi}_j) = A_{0j}^{-1}B_{0j}A_{0j}^{-1}$ where $\hat{\psi}_j = (\hat{\theta}_j, \hat{\beta}_j, \hat{\gamma}_j)$, $A_{0j} = -E[\nabla_{\psi_j \psi_j} \log L_{NT, \eta_1}^{(J_0)}(\psi_{0j})]$ and $B_{0j} = E[\nabla_{\psi_j} \log L_{NT, \eta_1}^{(J_0)}(\psi_{0j}) \nabla_{\psi_j} \log L_{NT, \eta_1}^{(J_0)}(\psi_{0j})']$, with ∇_{ψ_j} and $\nabla_{\psi_j \psi_j}$ denoting the vector of first and second derivatives of the log-likelihood function, respectively, ψ_{0j} is the true parameter vector and $j = 1, \dots, J_0$. The estimated asymptotic variance can be obtained by replacing the true parameters with their estimates discussed above, and the expectation is replaced by the sample average over NT observations.

² Even if we do not formally establish the asymptotic properties of the FDPL estimator, it is worth pointing out that the results of our Monte Carlo simulations are consistent with the belief that these asymptotic properties hold. See Section 5 for more details.

2.3 Prediction of the inefficiency index

The primary interest in estimating model (1) is to obtain the prediction for technical inefficiency, u_{it} . The conditional expectation estimator $E(u_{it} | e_{it})$ proposed by Jondrow et al. (1982) is often used for this purpose. For our proposed model, a similar conditional expectation estimator can also be used but with one simple modification. As Wang and Ho (2010) pointed out, instead of conditioning on the level of e_{it} , it is more convenient to compute the expectation of u_{it} condition on $\Delta \tilde{e}_i$ since $\Delta \tilde{e}_i = \Delta \tilde{y}_i - \Delta \tilde{x}_i \beta_i$ does not depend on the estimates of individual-specific effect, $\hat{\alpha}_i$. In addition, the vector $\Delta \tilde{e}_i$ contains all the information of individuals i within each group in the sample. Thus, given the estimates of $\hat{\beta}_i$ and $\hat{\theta}$ discussed previously, the conditional expectation estimator $E(u_{it} | \Delta \tilde{e}_i)$ and efficiency estimate Eff_{it} can be written as:

$$\hat{u}_{it} = E(u_{it} | \Delta \tilde{e}_i) = h_{it} \left[\mu_{i^*} + \frac{\sigma_{i^*} \phi(\mu_{i^*} / \sigma_{i^*})}{\Phi(\mu_{i^*} / \sigma_{i^*})} \right] \tag{8}$$

$$Eff_{it} = \exp(-\hat{u}_{it}),$$

where μ_{i^*} and σ_{i^*} are defined previously, and the expression in Eq. (6) is evaluated at $\Delta \tilde{e}_i = \Delta \hat{e}_i$, $h_{it} = h(q'_{it} \hat{\delta})$, $\mu_{i^*} = \hat{\mu}_{i^*}$, and $\sigma_{i^*} = \hat{\sigma}_{i^*}$. The group-wise efficiency prediction can be computed as $Eff_{it}^{(J)} = \exp(-\hat{u}_{it}^{(J)})$ where $\hat{u}_{it}^{(J)} = E(u_{it}^{(J)} | \Delta \tilde{e}_i^{(J)})$.

3 Model with endogenous regressors

3.1 Control function penalized likelihood (CFPL) estimation

In this section, we relax the independence assumption between $\{x_{it}, q_{it}\}$ and v_{it} (See for example, Kutlu 2010; Tran and Tsionas 2013; Amsler et al. 2016, 2017; Karakaplan and Kutlu 2017; Kutlu et al. 2019). In particular, we assume that a $k_p \times 1$ sub-vector, p_{it} of $\{x_{it}, q_{it}\}$ is correlated with v_{it} . However, we assume that $\{x_{it}, q_{it}\}$ and u_i^* are independent. In addition, we assume that there is a $k_z \times 1$ vector of (strictly) exogenous instruments z_{it} , where $k_z \geq k_p$ in the sense that $E(v_{it} | z_{is}) = 0$ for all t and s . Under these assumptions, we use a single-stage control function approach to deal with the endogeneity issue. That is, we use the following system of equations for the stochastic frontier model:

$$y_{it} = \alpha_i + x'_{it} \beta_i + v_{it} - u_{it}, \tag{9a}$$

$$p_{it} = \Pi z_{it} + \varepsilon_{it}, \tag{9b}$$

$$u_{it} = h_{it} u_i^*, \tag{9c}$$

$$h_{it} = h(q'_{it}\delta), \tag{9d}$$

$$u_i^* \sim N^+(0, \sigma_u^2), \tag{9e}$$

where Π is a $k_p \times k_z$ matrix of unknown coefficients, ε_{it} is a $k_p \times 1$ vector of reduced form errors, and β_i follows a group-specific pattern as in Eq. (2).

Under the specification of the above model, the endogeneity problem is introduced by allowing the reduced form error term ε_{it} to be correlated with v_{it} . More explicitly, we assume that conditional on z_{it} ,

$$\begin{pmatrix} \varepsilon_{it}^* \\ v_{it} \end{pmatrix} = \begin{pmatrix} \Omega^{-1/2} \varepsilon_{it} \\ v_{it} \end{pmatrix} \sim MN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_{k_p} & \sigma_v \rho \\ \sigma_v \rho' & \sigma_v^2 \end{pmatrix} \right), \tag{10}$$

where Ω is a $k_p \times k_p$ variance-covariance matrix of ε_{it} , and ρ is a $k_p \times 1$ vector representing the correlation between ε_{it}^* and v_{it} . To simplify the discussion, we assume for now that the number of groups, J_0 is known and fixed but each individual's group membership is unknown. Taking the first difference of Eq. (9) to eliminate the fixed effects and stacking all $T - 1$ observations, we have:

$$\Delta \tilde{y}_i = \Delta \tilde{x}_i \beta_i + \Delta \tilde{v}_i - \Delta \tilde{u}_i, \tag{11a}$$

$$\Delta \tilde{p}_i = \Delta \tilde{z}_i \Pi + \Delta \tilde{\varepsilon}_i, \tag{11b}$$

$$\Delta \tilde{u}_i = \Delta \tilde{h}_i u_i^*, \tag{11c}$$

$$u_i^* \sim N^+(0, \sigma_u^2). \tag{11d}$$

Recall that $\Delta \tilde{x}_i$ and $\Delta \tilde{z}_i$ are $(T - 1) \times k_p$ matrices and $\Delta \tilde{z}_i$ is a $(T - 1) \times k_z$ matrix. By a Cholesky decomposition of the variance-covariance matrix of $(\varepsilon_{it}^*, v_{it})'$, we obtain:

$$\begin{pmatrix} \varepsilon_{it}^* \\ v_{it} \end{pmatrix} = \begin{pmatrix} I_{k_p} & 0 \\ \sigma_v \rho' & \sigma_v \sqrt{1 - \rho' \rho} \end{pmatrix} \begin{pmatrix} \varepsilon_{it}^* \\ w_{it}^* \end{pmatrix}, \tag{12}$$

where ε_{it}^* and $w_{it}^* \sim N(0, 1)$ are independent. Therefore, we have:

$$\begin{aligned} v_{it} &= \sigma_v \rho' \varepsilon_{it}^* + \sigma_v \sqrt{1 - \rho' \rho} w_{it}^* \\ &= \varepsilon'_{it} \varphi + w_{it}, \end{aligned} \tag{13}$$

where $\varphi = \frac{\sigma_w \Omega^{-1/2} \rho}{\sqrt{1 - \rho' \rho}}$, $\sigma_w = \sigma_v \sqrt{1 - \rho' \rho}$, and $w_{it} = \sigma_w w_{it}^*$. After first differencing, we get: $\Delta \tilde{v}_i = \Delta \tilde{\varepsilon}_i \varphi + \Delta \tilde{w}_i$. Then, the frontier Eq. (11a) can be written as:

$$\Delta \tilde{y}_i = \Delta \tilde{x}'_i \beta_i + (\Delta \tilde{x}_i - \Pi' \Delta \tilde{z}_i) \varphi + \Delta \tilde{\varepsilon}_i, \tag{14}$$

where $\Delta \tilde{\xi}_i = \Delta \tilde{w}_i - \Delta \tilde{u}_i$ and $\Delta \tilde{\varepsilon}_i \varphi$ is the endogeneity bias correction term. The density function of $\Delta \tilde{w}_i$ is given by³:

$$\begin{aligned} f_{\Delta \tilde{w}_i}(\Delta \tilde{w}_i) &= |2\pi \sigma_w^2 \Sigma|^{-1/2} \exp \left(-\frac{1}{2\sigma_w^2} \Delta \tilde{w}'_i \Sigma^{-1} \Delta \tilde{w}_i \right) \\ &= (2\pi)^{-(T-1)/2} T^{-1/2} (\sigma_w^2)^{-(T-1)/2} \exp \left(-\frac{1}{2\sigma_w^2} \Delta \tilde{w}'_i \Sigma^{-1} \Delta \tilde{w}_i \right). \end{aligned} \tag{15}$$

Similarly, the joint density function of $\Delta \tilde{\varepsilon}_i$ is given by:

$$\begin{aligned} f_{\text{vec}(\Delta \tilde{\varepsilon}'_i)}(\text{vec}(\Delta \tilde{\varepsilon}'_i)) &= |2\pi(\Sigma \otimes \Omega)|^{-1/2} \exp \left(-\frac{1}{2} \text{vec}(\Delta \tilde{\varepsilon}'_i) (\Sigma^{-1} \otimes \Omega^{-1}) \text{vec}(\Delta \tilde{\varepsilon}'_i) \right) \\ &= (2\pi)^{-(T-1)k_p/2} T^{-k_p/2} |\Omega|^{-(T-1)/2} \exp \left(-\frac{1}{2} \text{vec}(\Delta \tilde{\varepsilon}'_i) (\Sigma^{-1} \otimes \Omega^{-1}) \text{vec}(\Delta \tilde{\varepsilon}'_i) \right). \end{aligned} \tag{16}$$

Since $\Delta \tilde{\varepsilon}_i$ and $\Delta \tilde{\xi}_i$ are independent, after tedious but straightforward derivation, the marginal log-likelihood function of the panel i is given by:

$$\log L_{2i,NT} = \log L_{2i,NT}^{(1)} + \log L_{2i,NT}^{(2)}, \tag{17}$$

where

$$\begin{aligned} \log L_{2i,NT}^{(1)} &= -\frac{1}{2} \left[(T - 1) (\log(2\pi) + \log(\sigma_w^2)) + \ln(T) + \frac{\Delta \tilde{\xi}'_i \Sigma^{-1} \Delta \tilde{\xi}_i}{\sigma_w^2} \right] \\ &\quad + \frac{1}{2} \frac{\mu_{i**}^2}{\sigma_{i**}^2} + \log \left[\sigma_{i**} \Phi \left(\frac{\mu_{i**}}{\sigma_{i**}} \right) \right] - \log \left(\frac{\sigma_u}{2} \right), \end{aligned}$$

and

$$\begin{aligned} \log L_{2i,NT}^{(2)} &= -\frac{(T-1)k_p}{2} \log(2\pi) - \frac{k_p}{2} \log(T) - \frac{(T-1)}{2} \log(|\Omega|) \\ &\quad - \frac{1}{2} \text{tr} \left[(\Sigma^{-1} \otimes \Omega^{-1}) \text{vec}(\Delta \tilde{\varepsilon}'_i) \text{vec}(\Delta \tilde{\varepsilon}'_i)' \right], \end{aligned}$$

where

$$\begin{aligned} \mu_{i**} &= -\frac{\Delta \tilde{\xi}'_i \Sigma^{-1} \Delta \tilde{h}_i}{\Delta \tilde{h}'_i \Sigma^{-1} \Delta \tilde{h}_i + \sigma_w^2 / \sigma_u^2}, & \sigma_{i**}^2 &= \frac{\sigma_w^2}{\Delta \tilde{h}'_i \Sigma^{-1} \Delta \tilde{h}_i + \sigma_w^2 / \sigma_u^2}, \\ \Delta \tilde{\xi}_i &= \Delta \tilde{y}_i - \Delta \tilde{x}_i \beta_i - \Delta \tilde{\varepsilon}_i \varphi, & \text{and } \Delta \tilde{\varepsilon}_i &= \Delta \tilde{x}_i - \Pi \Delta \tilde{z}_i. \end{aligned}$$

As before, let $\beta = (\beta_1, \dots, \beta_N)'$, $\gamma = (\gamma_1, \dots, \gamma_J)'$, and $\theta^* = (\text{vec}(\Pi)', \delta', \rho', \sigma_v^2, \sigma_u^2)'$, then we propose to estimate β , γ , and θ^* by maximizing the following CFPL criterion:

$$\begin{aligned} \log L_{NT, \eta_2}^{(J)}(\theta^*, \beta, \gamma) &= \sum_{i=1}^N \log L_{2i,NT}^{(1)} + \sum_{i=1}^N \log L_{2i,NT}^{(2)} \\ &\quad + \frac{\eta_2}{N} \sum_{i=1}^N \prod_{j=1}^J \|\beta_i - \gamma_j\|, \end{aligned} \tag{18}$$

where $\eta_2 = \eta_{2,NT}$ is a tuning parameter. Maximizing Eq. (18) produces the CFPL estimates $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_N)'$, $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_J)'$, and $\hat{\theta}^* = (\text{vec}(\hat{\Pi})', \hat{\delta}', \hat{\rho}', \hat{\sigma}_v^2, \hat{\sigma}_u^2)'$.

3.2 Determination of the number of groups

Similar to the case of exogenous regressors, when the number of groups J is unknown, we replace J_0 by $J \in$

³ Note that $|\Sigma| = T$.

$(1, \dots, J_{\max})$ to obtain the CFPL (or C-Lasso) estimates $\{\widehat{\beta}_i(J, \eta_2), \widehat{\gamma}_j(J, \eta_2)\}$ and $\widehat{\theta}^*$ of $\{\beta_i, \gamma_j\}$ and θ^* discussed above. We then classify individual i into group $\widehat{G}_i(J, \eta_2)$ according to $\widehat{G}_i(J, \eta_2) = \{i \in \{1, 2, \dots, N\} : \widehat{\beta}_i(J, \eta_2) = \widehat{\gamma}_j(J, \eta_2)\}$. Let $\widehat{G}(J, \eta_2) = \{\widehat{G}_1(J, \eta_2), \dots, \widehat{G}_J(J, \eta_2)\}$ and $\widehat{\gamma}_{\widehat{G}_i(J, \eta_2)}$ denote the post-CFPL (or post-Lasso) estimator. Then, we suggest selecting J that minimizes the following IC:

$$IC_2(J, \eta_2) = \frac{2}{NT} \sum_{j=1}^J \sum_{i \in \widehat{G}_j(J, \eta_2)} \log L_{2i,NT}(w_{it}; \widehat{\gamma}_{\widehat{G}_j(J, \eta_2)}, \widehat{\theta}^*(\widehat{\gamma}_{\widehat{G}_j(J, \eta_2)})) + \lambda_{2,NT} K_2 J, \tag{19}$$

where $\lambda_{2,NT}$ is a tuning parameter, and $K_2 = k_x + k_q + k_z + 2$. That is, the number of groups J is chosen such that $\widehat{J}(\eta_2) = \arg \min_{1 \leq J \leq J_{\max}} IC_2(J, \eta_2)$. For the choice of the tuning parameters $\eta_{2,NT}$ and $\lambda_{2,NT}$ in practice as well as discussion on the asymptotic properties of the proposed estimator, see Remark 1 and Remark 2 above.

3.3 Prediction of inefficiency score

As earlier, given the CFPL estimates of $\widehat{\beta}_i$ and $\widehat{\theta}^*$, the conditional expectation estimator $E(u_{it} | \Delta \tilde{\xi}_i)$ can be written as:

$$E(u_{it} | \Delta \tilde{\xi}_i) = h_{it} \left[\mu_{i^{**}} + \frac{\sigma_{i^{**}} \phi(\mu_{i^{**}} / \sigma_{i^{**}})}{\Phi(\mu_{i^{**}} / \sigma_{i^{**}})} \right], \tag{20}$$

where $\mu_{i^{**}}$ and $\sigma_{i^{**}}$ are defined previously, and the expression in Eq. (20) is evaluated at $\Delta \tilde{\xi}_i = \Delta \tilde{\xi}_i$ where $\Delta \tilde{\xi}_i = \Delta \tilde{y}_i - \Delta \tilde{x}_i \widehat{\beta}_i - \Delta \tilde{e}_i \widehat{\theta}$, and $\Delta \tilde{e}_i = \Delta \tilde{x}_i - \overline{\Pi} \Delta \tilde{z}_i$; $h_{it} = h(q'_{it} \widehat{\delta})$, $\mu_{i^{**}} = \overline{\mu}_{i^{**}}$; and $\sigma_{i^{**}} = \overline{\sigma}_{i^{**}}$. As in the exogenous regressors case, the group-wise efficiency prediction can be computed as $Eff_{it}^{(J)} = \exp(-\widehat{u}_{it}^{(J)})$ where $\widehat{u}_{it}^{(J)} = E(u_{it}^{(J)} | \Delta \tilde{\xi}_i^{(J)})$.

4 Computational algorithm

In this section, we briefly outline an iterative numerical algorithm to obtain the FDPL estimates $\widehat{\gamma}$, $\widehat{\beta}$, and $\widehat{\theta}$ discussed in Section 2.1. For the CFPL estimation, a similar algorithm can be applied. Let s be the iteration index.

Step 1: Set $s = 1$ and start with the initial value $\widehat{\gamma}^{(0)} = (\widehat{\gamma}_1^{(0)}, \dots, \widehat{\gamma}_J^{(0)})$, $\widehat{\beta}^{(0)} = (\widehat{\beta}_1^{(0)}, \dots, \widehat{\beta}_N^{(0)})$, and $\widehat{\theta}^{(0)} = (\widehat{\delta}^{(0)}, \widehat{\sigma}_v^{2(0)}, \widehat{\sigma}_u^{2(0)})$ such that $\sum_{i=1}^N \|\widehat{\beta}_i^{(0)} - \widehat{\gamma}_j^{(0)}\| \neq 0$ for $j = 2, \dots, J$.

Step 2: Given $\widehat{\gamma}^{(r-1)}$, $\widehat{\beta}^{(r-1)}$, and $\widehat{\theta}^{(r-1)}$, choose $(\gamma_1, \beta, \theta)$ to maximize:

$$\log L_{NT, \eta_1}^{(j_0)}(\theta, \beta, \gamma) = \sum_{i=1}^N \log L_{i,NT}(\beta, \theta) + \frac{\eta_1}{N} \sum_{i=1}^N \|\beta_i - \gamma_1\| \prod_{j=1}^J \|\widehat{\beta}_i^{(r-1)} - \widehat{\gamma}_j^{(r-1)}\|,$$

and obtain the updated $(\widehat{\beta}^{(r)}, \widehat{\theta}^{(r)}, \widehat{\gamma}_1^{(r)})$. Repeat this procedure until we obtain the updated of $(\widehat{\beta}^{(r)}, \widehat{\theta}^{(r)}, \widehat{\gamma}_j^{(r)})$.

Step 3: Update s to $s + 1$ and repeat Step 2 until convergence.

Define the final estimate of γ as $\widehat{\gamma} = (\widehat{\gamma}_1^{(S)}, \dots, \widehat{\gamma}_J^{(S)})$ where S denotes the final iteration such that the convergence is achieved. Then, individual i is classified as a member of group \widehat{G}_j if $\widehat{\beta}_i^{(S,j)} = \widehat{\gamma}_j$; otherwise, $\widehat{\beta}_i$ is allocated to the $\gamma_j^{(S)}$ that is nearest to some $\widehat{\beta}_i^{(S,k)}$, for $k = 1, \dots, J$. The initial value in Step 1 is chosen based on random initial conditions to find the best starting value (10,000 searches). If a failure occurs, a new random search is used. We use this choice of initial value throughout our simulations as well as the empirical application below.

5 Monte Carlo simulations

5.1 Data generating process (DGP)

To examine the finite sample performance of the estimation and classification procedure, we consider two DGP that cover both exogenous and endogenous regressors. We consider sample sizes $N = \{250,500\}$ and time periods $T = \{10, 20, 40\}$. For each sample, the observations in each DGP are drawn from three groups with the proportion $\{N_1: N_2: N_3\} = \{0.4:0.4:0.3\}$. Throughout the experiments, the fixed effect α_i is standard normal, independent across i . The one-sided error u_i^* is generated as i.i.d. from a half-normal with $\sigma_u = \{1, 2\}$, and u_i^* is independent of α_i and all regressors.

DGP 1 (Exogenous Regressors): The observations (y_{it}, x_{it}, q_{it}) are generated from the model (1a)–(1e). The exogenous regressors $x_{it} = (x_{1it}, x_{2it})' = (0.5\alpha_i + \omega_{1it}, 0.5\alpha_i + \omega_{2it})'$ where $\omega_{1it}, \omega_{2it} \sim i.i.d.N(0, 1)$ are mutually independent, and independent of α_i . The two-sided error v_{it} is standard normal, independent across i and t , and independent of α_i, u_i^* and all the regressors. The environmental variable q_{it} is generated as i.i.d. from a Uniform distribution on $[-1, 1]$ and $h_{it} = \exp(0.2q_{it})$. Finally, the true coefficients for the three groups $(\beta_{11}, \beta_{12}), (\beta_{21}, \beta_{22}),$ and (β_{31}, β_{32}) are $(0.5, 1.5), (1, 1),$ and $(1.5, 0.5)$, respectively. FDPL will be used for this DGP.

DGP 2 (Endogenous Regressors): For this DGP, the observations are generated from the model (9a)–(9e). We

assume that $x_{2,it}$ is exogenous and is generated as in DGP 1, whilst $x_{1,it}$ is endogenous and is generated as $x_{1it} = 0.4\alpha_i + 0.8z_{it} + \varepsilon_{it}$ where $z_{it} \sim i.i.d.N(0, 1)$ and

$$\begin{pmatrix} \varepsilon_{it} \\ v_{it} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

In each experiment, we set $\rho = \{0.2, 0.4, 0.8\}$ which corresponds to a weak, moderate and strong correlation between v_{it} and ε_{it} , respectively. The environmental variable q_{it} and the true coefficients for the three groups are the same as in DGP 1. Finally, the Monte Carlo replications for each DGP is 500.

Our first simulation exercise is to assess how well the proposed IC selects the number of groups for each DGP. As discussed earlier, the choice of the fine-tune parameters η_{jNT} and λ_{jNT} for $j = 1, 2$, can be important in selecting the correct number of groups. For the tuning parameter η_{jNT} , we choose $\eta_j = c_j s_{\Delta y}^2 T^{-1/3}$ for $j = 1, 2$, where $s_{\Delta y}^2$ is the sample variance of Δy_{it} and $c_j \in \{0.125, 0.25, 0.5, 1, 2\}$. As for the tuning parameter λ_{jNT} we use the following $\lambda_j = \kappa (NT)^{-1/2}$ for $j = 1, 2$, where $\kappa \in \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}\}$. We experimented with many alternatives, and found that $c_j = 0.25$ and $\kappa_j = 2/3$ for $j = 1, 2$, work fairly well and they are used throughout the simulations and the empirical application. For DGP 1, we pick up from the set candidate values of η_1 that maximizes $IC_1(\hat{J}(\eta_1), \eta_1)$, and similarly for GDP 2, we pick up from the set candidate values of η_2 that maximizes $IC_2(\hat{J}(\eta_2), \eta_2)$. In all experiments, we use the initial starting value described in Section 4 and the BFGS numerical algorithm procedure from NETLIB in Fortran77, GNU compiler to maximize the log-likelihood function with the convergence criterion is set at 10^{-5} . In almost all cases the convergence was quick and stable.

5.2 Results

Tables 1 and 2 report the empirical probability that a particular group size from 1 to 5 is selected according to the proposed IC when the true number of groups is 3. In particular, Table 1

Table 1 Empirical probability of selecting $J = 1, \dots, 5$ when $J_0 = 3$ -DGP 1

N	J	T				
		1	2	3	4	5
250	10	0.000	0.158	0.842	0.000	0.000
250	20	0.000	0.088	0.912	0.000	0.000
250	40	0.000	0.063	0.937	0.000	0.000
500	10	0.000	0.055	0.955	0.000	0.000
500	20	0.000	0.023	0.977	0.000	0.000
500	40	0.000	0.000	1.000	0.000	0.000

shows the results for DGP 1 (exogenous case), whilst Table 2 display the results for DGP 2 (endogenous case). In line with our prior expectation, the correct classification percentage approaches 100% as T increases for both DGPs.

Next, we focus our attention on the classification of individual units and the point estimation of post-Lasso, given the true number of groups ($J_0 = 3$). Due to the space limitation, all tabulated results are produced using $c_j = 0.25, j = 1, 2$, albeit the outcomes are found to be robust over the specified range of constants. For comparison purposes, we also include the oracle estimator $\hat{\gamma}_{G_j}$ or $\bar{\gamma}_{G_j}$ which defines as the infeasible estimator that utilizes the true group identity G_j . For conservation of space, we only report the results for the first coefficient $\gamma_1 = (\gamma_{11}, \dots, \gamma_{1J_0})'$, and since γ_1 is a $J_0 \times 1$ vector, we use the average statistics over their weight $N_j/N, j = 1, \dots, J_0$. The results are depicted in Tables 3 and 4. We report the bias, root-mean-squared errors (RMSE), and the percentage of correct classification of the N units, computed as $N^{-1} \sum_{j=1}^3 \sum_{i \in \hat{G}_j} 1\{\beta_i = \gamma_k\}$ averaged over 500 replications. The results indicate that the estimated bias and RMSE of the oracle and post-Lasso estimators are decreasing as either N or

Table 2 Empirical probability of selecting $J = 1, \dots, 5$ when $J_0 = 3$ -DGP 2

N	J	T				
		1	2	3	4	5
(A)						
$(\rho = 0.2)$						
250	10	0.000	0.150	0.850	0.000	0.000
250	20	0.000	0.078	0.922	0.000	0.000
250	40	0.000	0.059	0.941	0.000	0.000
500	10	0.000	0.028	0.972	0.000	0.000
500	20	0.000	0.000	1.000	0.000	0.000
500	40	0.000	0.000	1.000	0.000	0.000
(B)						
$(\rho = 0.4)$						
250	10	0.000	0.108	0.892	0.000	0.000
250	20	0.000	0.056	0.944	0.000	0.000
250	40	0.000	0.029	0.971	0.000	0.000
500	10	0.000	0.011	0.989	0.000	0.000
500	20	0.000	0.000	1.000	0.000	0.000
500	40	0.000	0.000	1.000	0.000	0.000
(C)						
$(\rho = 0.8)$						
250	10	0.000	0.108	0.892	0.000	0.000
250	20	0.000	0.019	0.981	0.000	0.000
250	40	0.000	0.007	0.993	0.000	0.000
500	10	0.000	0.000	1.000	0.000	0.000
500	20	0.000	0.000	1.000	0.000	0.000
500	40	0.000	0.000	1.000	0.000	0.000

Table 3 Classification and point estimation of γ_1 -DGP 1

N	T	% of correct classification	Oracle		Post-FDPL	
			Bias	RMSE	Bias	RMSE
250	10	0.815	0.0124	0.0232	0.0172	0.0303
250	20	0.892	0.0071	0.0175	0.0092	0.0253
250	40	0.920	0.0055	0.0126	0.0076	0.0211
500	10	0.947	0.0032	0.0081	0.0055	0.0131
500	20	0.969	0.0028	0.0065	0.0034	0.0077
500	40	0.995	0.0011	0.0052	0.0013	0.0063

Table 4 Classification and point estimation of γ_1 -DGP 2

N	T	% of correct classification	Oracle		Post-CFPL	
			Bias	RMSE	Bias	RMSE
(A)						
$(\rho = 0.2)$						
250	10	0.813	0.0174	0.0254	0.0229	0.0442
250	20	0.891	0.0085	0.0196	0.0099	0.0320
250	40	0.925	0.0071	0.0177	0.0082	0.0291
500	10	0.948	0.0063	0.0090	0.0077	0.0172
500	20	0.968	0.0044	0.0072	0.0052	0.0091
500	40	0.997	0.0020	0.0046	0.0026	0.0055
(B)						
$(\rho = 0.4)$						
250	10	0.820	0.017	0.0271	0.025	0.0440
250	20	0.897	0.0082	0.0213	0.0098	0.0381
250	40	0.928	0.0070	0.0195	0.0081	0.0272
500	10	0.955	0.0045	0.0094	0.0053	0.0130
500	20	0.973	0.0022	0.0075	0.0031	0.0092
500	40	0.998	0.0008	0.0049	0.0013	0.0055
(C)						
$(\rho = 0.8)$						
250	10	0.898	0.010	0.0294	0.0194	0.0431
250	20	0.916	0.0051	0.0236	0.0092	0.0343
250	40	0.945	0.0032	0.0207	0.0049	0.0291
500	10	0.981	0.0025	0.0097	0.0032	0.0131
500	20	0.992	0.0019	0.0078	0.0027	0.0085
500	40	0.999	0.0007	0.0051	0.0011	0.0060

T decreases. Moreover, the bias and RMSE of the oracle’s estimator are slightly smaller than those of post-Lasso. The main reason for these (mild) discrepancies of the RMSE is that the estimated bias and standard deviation of the post-Lasso estimator are inflated by some misclassification units, which mask as outliers against most of the group members. Nevertheless, our results seem to confirm the oracle properties of the proposed estimators.

6 An empirical application

In this section we present an empirical application of US banks to illustrate the usefulness of our proposed models and estimation methods discussed earlier.

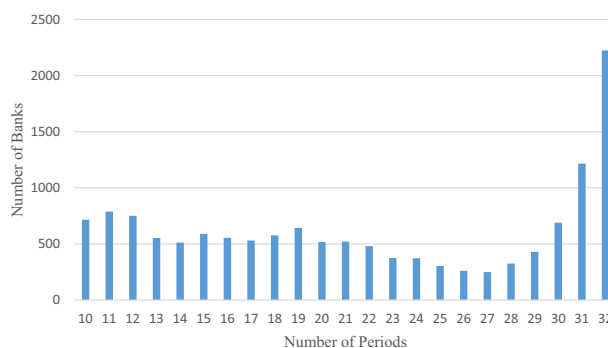


Fig. 1 Frequency distribution of the data

6.1 The data

The data we use in this paper is the annual year-end bank-level which include all FDIC-insured commercial banks from 1976–2007. The data was taken from Koetter et al. (2012) and the detailed description of the data is given in their paper.

For the input prices, we truncated these variables at the 1st and 99th percentiles of their respective empirical distributions to mitigate the influence of outliers. In addition, we use the 2005 Consumer Price Index for all urban consumption (published by the U.S. Bureau of Labor Statistics) to deflate all nominal quantities. Since our proposed approach requires large N and T, we use an unbalanced panel data and first include all banks with time dimension $T_i \geq 10$. There are $N = 14,168$ banks included in our data set for the analysis. The average length of time periods for all banks is about $N^{-1} \sum_{i=1}^N T_i \approx 21.9$. Figure 1 plots the frequency distribution of T_i .

Following Sealey and Lindley (1997), the bank’s production technology is modeled using the “intermediation approach”. Labour, physical capital as well as liabilities are used as inputs to the bank’s production process, whilst assets (other than physical) are considered as outputs. The following variables are used as outputs (y) and inputs (x) in the production technology: Securities (y_1) and loans (y_2); fixed assets (x_1), labor (x_2), borrowed funds (x_3) as well as equity capital (x_4). The inclusion of equity capital as an additional input can be argued that banks may use it to guard against losses, and hence it can be considered as a source of loanable funds.

In order to contextualize the economic environment in which banks operate, we include the following variables, both internally and externally, to capture bank’s characteristics in our instrumental variables (z_{it}) and environmental variables (q_{it}). The instrumental variables z_{it} include: (1) the bank’s total assets as a proxy for its size and scale of operation; (2) the bank’s asset market share in a given state to capture its dominance in the market (see, for example, Stiroh and Strahan 2003; Boyd and DeNicolò 2005); (3) the bank’s ratio of equity to total assets; (4) the bank’s ratio of securities to total assets;

(5) the share of non-interest income (Koetter et al. 2012); (6) the share of loan-loss provisions and loan-loss reserves in the bank’s total loans to proxy for credit risk; and (7) macro-economic variables such as the disposable personal income and the state’s unemployment rate. The environmental variables (q_{it}) include: (1) the number of bank mergers in the state in a given year; (2) the bank’s z-score to proxy for the overall risk of bank failure (see Laeven and Levin 2009); (3) the Hirschman-Herfindahl index across the banks’ different types of loans; (4) an indicator for the top-hundred banks in a given year; and (5) three indicators to capture the institutional changes in states that correspond to deregulation in the intrastate branching, the interstate expansion and the post-IBBEA interstate banking.

The above chosen variables have important implications for bank efficiency and market power since they are more than likely to influence bank’s business strategies in their quest to maximize its franchise value (Demsetz and Strahan 1997; DeYoung and Rice 2004). For more details on the construction and rationale behind these variables, see Koetter et al. (2012).

Production technology is described by an output distance function (ODF). Suppose the inputs are X_1, \dots, X_K , and the outputs are Y_1, \dots, Y_M . Define $x_k = \ln X_k, k = 1, \dots, K, y_1 = \ln Y_1$ and $y_m = \ln(Y_m/Y_1), m = 2, \dots, M$. This transformation is used to ensure that the IDF is homogeneous of degree one in outputs. Then, the output distance function (ODF) takes the following form:

$$y_{1,it} = f(x_{1,it}, \dots, x_{K,it}, y_{2,it}, \dots, y_{M,it}) + \alpha_i + v_{it} - u_{it},$$

$$i = 1, \dots, n; t = 1, \dots, T_i.$$

Let $\tilde{y}_{it} = (y_{2,it}, \dots, y_{M,it})'$ and $x_{it} = (x_{1,it}, \dots, x_{K,it})'$. If we adopt a translog specification, we have:

$$y_{1,it} = \alpha_i + \beta'_{xi}x_{it} + \frac{1}{2}x'_{it}B_{xi}x_{it} + \beta'_{yi}\tilde{y}_{it} + \frac{1}{2}\tilde{y}'_{it}B_{yi}\tilde{y}_{it}$$

$$+ x'_{it}B_{xyi}\tilde{y}_{it} + v_{it} - u_{it},$$

where B_{xi}, B_{yi} , and B_{xyi} are parameter matrices, β_{xi} and β_{yi} are parameter vectors. We can write the ODF as follows:

$$y_{1,it} = \alpha_i + x'_{*,it}\beta_i + v_{it} - u_{it},$$

where $x_{*,it} = [1, x'_{it}, \text{vech}(x_{it} \otimes x_{it})', \tilde{y}'_{it}, \text{vech}(\tilde{y}_{it} \otimes \tilde{y}_{it})']'$. We treat both log inputs and log output ratios as endogenous. Under certain economic assumptions, log output ratios can be treated as predetermined. However, econometrically, there is no compelling reason to adopt this as conclusive evidence that they can be treated as exogenous. In fact, profit maximization would imply that both inputs and outputs are economically (and, in all likelihood, econometrically, as well) endogenous. In addition to the instruments listed above, we also include interactions among the

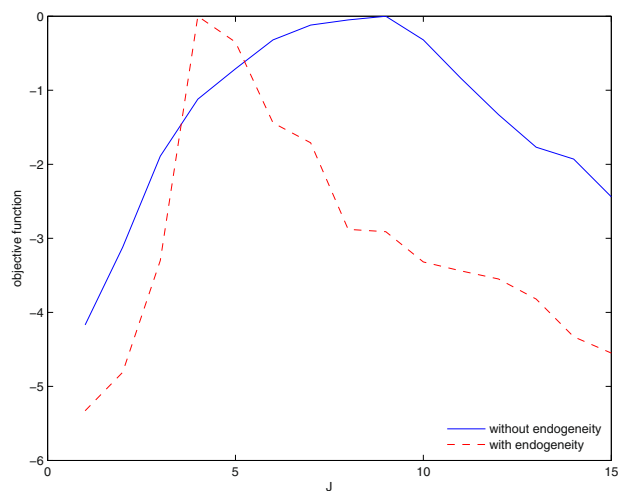


Fig. 2 Identification of number of groups

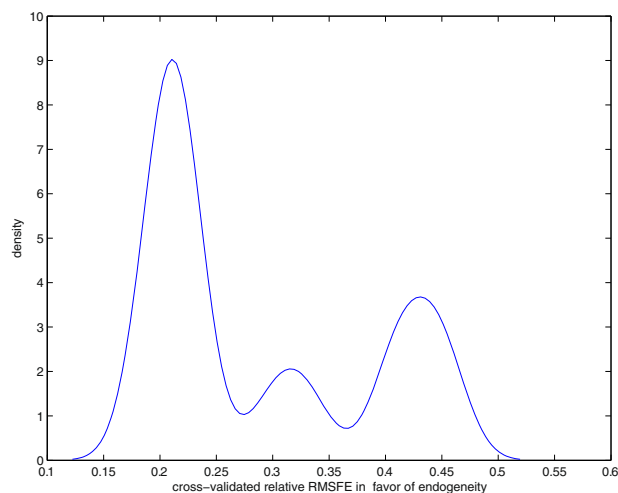


Fig. 3 Density plot of the ratio of RMSFE

variables in z_{it} , interactions of time dummies with all other time-invariant variables in z_{it} as well as the lagged values of time-varying instruments along with their interactions with all other variables as additional instruments.

6.2 Results

We first determine the appropriate number of groups using Eqs. (7) and (19) for the case of exogenous and endogenous regressors, respectively, and the results are depicted in Fig. 2. For the exogenous regressor case, the optimal number of groups is seven, whilst for the endogenous regressors’ case, the optimal group is four implying that endogeneity may be an issue for our model. To check for the endogeneity problem, we plot the density of the ratio of the root mean squares forecast errors (RMSFE) and our results show that the regressors are indeed endogenous (see Fig. 3). As a further evidence for the endogeneity of these

Table 5 RTS, efficiency change, technical change and productivity growth results

Group	RTS	Eff. change	Tech. change	Prod. growth
Homogenous	0.9694*** (0.0220)	0.0029 (0.0145)	0.0007 (0.0115)	0.0036 (0.0240)
1	0.6257*** (0.0676)	0.0091** (0.0045)	0.0088 (0.0071)	0.0179*** (0.0068)
2	0.8878*** (0.1390)	0.0025 (0.0018)	0.0157* (0.0087)	0.0182** (0.0088)
3	0.8349*** (0.0622)	0.0089*** (0.0016)	0.0101** (0.0046)	0.0189*** (0.0045)
4	0.6558*** (0.0427)	0.0044* (0.0023)	0.0038* (0.0021)	0.0082*** (0.0031)

Standard errors are given the parentheses.

***1% significant, **5% significant, *10% significant.

variables, we conduct a simple test for $\varphi = 0$ in (14) using F-statistics (see, for example, Amsler et al. (2016) and Karakaplan and Kutlu (2017)). The value of the F-statistics is 64.78 with a p -value of 0.000 indicating that the null hypothesis of $\varphi = 0$ is rejected at a one percent significant level. Thus, in what follows we only report the results for the endogenous case. For comparison purposes, we also estimate the homogenous coefficient stochastic model with endogenous regressors using the approach developed in Kutlu et al. (2019).

For the sake of convenience and simplicity, Table 5 and Fig. 5 report only the summary of estimated groups' productivity measures, which include returns to scale (RTS), efficiency change, technical change, and productivity growth⁴. For comparison purposes, the estimated productivity measures for the homogenous-coefficients model are given in the second row of Table 5 and Fig. 4a, b, respectively. Our results indicate that RTS for the homogenous-coefficient model is close to one (i.e., constant RTS) and statistically significant at the 1% level, whilst other measures such as efficiency change (EC), technical change (TC), and productivity growth are not statistically significant. In contrast, the results based on our proposed approach, the CFPL classifies the banks into four groups based on the variables that are related to the banks' size and scale operation, market shares, as well as the ratio of equity and securities to total assets. The results in Table 5 and Fig. 5 show that (RTS) vary over the four groups, and they average 0.627, 0.888, 0.835, and 0.656 indicating that different groups have different properties in terms of RTS, albeit in all groups, we have decreasing returns to scale. Efficiency change is mostly positive for all groups. The densities of efficiency change exhibit bimodal for groups one and four with a dominant mode at near-zero value for group two, and a positive value of approximately 0.0065 for group four. The average efficiency scores for the four groups are 0.821, 0.924, 0.935, and 0.845, respectively. For the most part, technical change is positive and consequently, productivity growth is positive except for group 4, which has significantly less technical change and

⁴ Detailed results for the estimated frontier parameters are available from the authors up request.

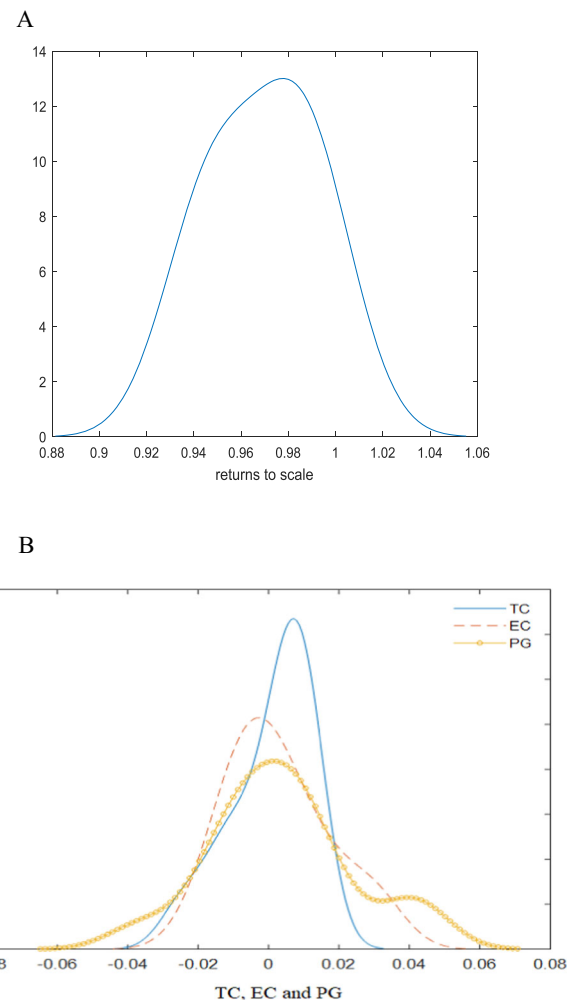
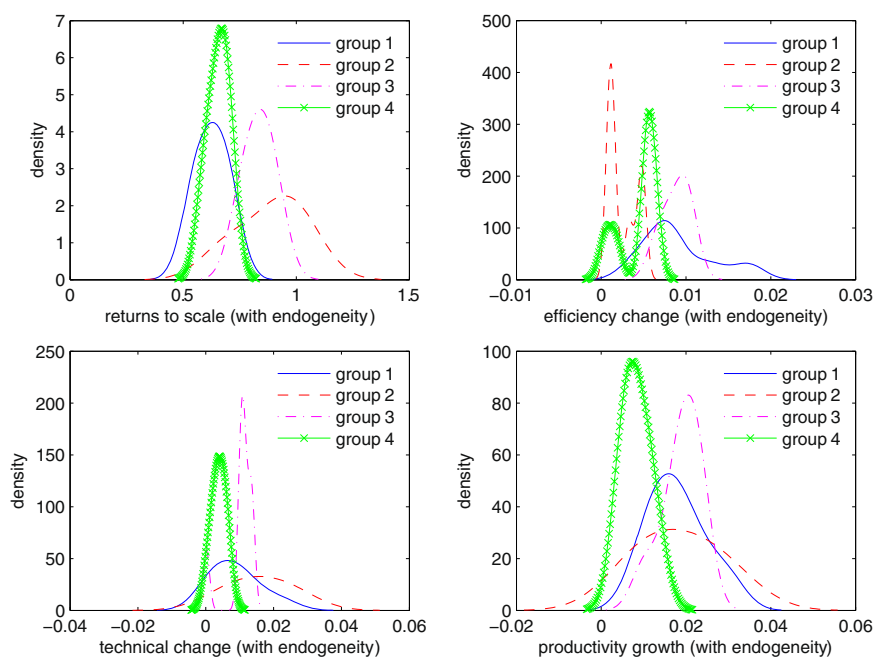


Fig. 4 a Density plot of returns to scale: homogenous case. b Density plot of TC, EC and PG: homogenous case

productivity growth compared to the other groups. As the groups are different in terms of RTS, technical change, efficiency change, and productivity growth, any policy measures will have heterogeneous effects on specific banks according to the group to which they belong. Consequently, ignoring the group-wise heterogeneity when it is present, can provide misleading estimates of productivity and efficiency measures which may have negative consequences on policy and banking supervisions.

Fig. 5 Density plots of productivity and efficiency change measurements



Finally, as a robustness check, we also consider the case where $T_i \geq 9$ and $T_i \geq 8$. In these cases, the number of banks (N) increased to 14,974 and 15,729, respectively. Using our approach for both cases, the optimal number of groups obtained is still 4 and the productivity measures are similar to those in Table 5. For the conservation of space, we do not report these results here but available from the authors upon request.

7 Concluding remarks

This paper extends the fixed effects panel stochastic frontier model of Wang and Ho (2010) to allow group heterogeneity in the slope coefficients. We propose the first-difference penalized maximum likelihood (FDPML) and control function penalized maximum likelihood (CFPML) methods for classification and estimation of latent group structures in the frontier as well as inefficiency. Monte Carlo simulations show that the proposed approach performs well in finite samples. An empirical application indicates the advantages of data-determined identification of the heterogeneous group structures in practice.

The approach in this paper can also be adapted to the Chen et al. (2014) model where they leave the inefficiency term u_{it} unspecified. In this case, the density of the transformation errors (using either first-difference or within transformation) can be obtained with a similar approach as in Chen et al. (2014) using the closed skew normal results. Also, it would be interesting to extend our approach to the four-component stochastic frontier models of Colombi et al. (2014), Kumbhakar et al. (2014), and Tsionas and Kumbhakar (2014).

Finally, under the endogenous regressors case, it is possible to extend the current method to allow for some or all inputs to be correlated with both u_{it} and v_{it} . Using Amsler et al. (2017) approach, one can construct the joint density of all the errors in the model via copula function. Alternatively, the correlation between some or all inputs and inefficiency can be modeled using the correlated effects as it has been done in Griffiths and Hajargasht (2016). We will leave these topics for future research.

Acknowledgements We would like to thank the Editor, the Associate Editor and two anonymous referees for constructive comments and suggestions that helped improve this paper. An earlier draft of this paper was presented at The EcoSta Conference in Hong Kong, June 2017. We would like to thank Artem Prokhorov, Dan Henderson, and the participants in our invited session for their comments and suggestions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Amsler C, Prokhorov A, Schmidt P (2016) Endogeneity in stochastic frontier models. *J Econ* 190:280–288
- Amsler C, Prokhorov A, Schmidt P (2017) Endogenous environmental variables in stochastic frontier models. *J Econ* 199:131–140
- Battese GE, Rao DSP, O'Donnell CJ (2004) A metafrontier production function for estimation of technical efficiencies and technology

- gap for firms operating under different technologies. *J Prod Anal* 21:91–103
- Boyd JH, De Nicrolo G (2005) The theory of bank risk taking and competition revisited. *J Financ* 60:1329–1343
- Chen Y-Y, Schmidt P, Wang H-J (2014) Consistent estimation of the fixed effects stochastic frontier model. *J Econ* 181:65–76
- Colombi R, Kumbhakar S, Martini G, Vittandini G (2014) Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *J Prod Anal* 42:123–136
- Demsetz RS, Strahan PE (1997) Diversification, size, and risk at bank holding companies. *J Money Credit Bank* 29:300–313
- DeYoung R, Rice T (2004) Non-interest income and financial performance at U.S. commercial banks. *Financ Rev* 39:101–127
- Greene WH (2005a) Fixed and random effects in stochastic frontier models. *J Prod Anal* 23:7–32
- Greene WH (2005b) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *J Econ* 126:269–303
- Griffiths WE, Hajargasht G (2016) Some models for stochastic frontiers with endogeneity. *J Econ* 190:341–348
- Guan Z, Kumbhakar SC, Myers RJ, Lansink AO (2009) Measuring excess capital capacity in agricultural production. *Am J Agric Econ* 91:765–776
- Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. *J Econ* 19:233–238
- Karakaplan MU, Kutlu L (2017) Handling endogeneity in stochastic frontier analysis. *Econ Bull* 37:889–901
- Koetter M, Kolari JW, Spierdijk L (2012) Enjoying the quiet life under deregulation? Evidence from adjusted lerner indices for U.S. banks. *Rev Econ Stat* 94:462–480
- Kumbhakar SC, Lien G, Hardaker JB (2014) Technical efficiency in competing panel data models: a study of Norwegian grain farming. *J Prod Anal* 41:321–337
- Kutlu L (2010) Battese-Coelli estimator with endogenous regressors. *Econ Lett* 109:79–81
- Kutlu L, Tran KC, Tsionas MG (2019) A time-varying true individual effects model with endogenous regressors. *J Econ* 211:539–559
- Kutlu L, Tran KC (2019) Heterogeneity and endogeneity in panel stochastic frontier models. In: Tsionas EG (ed) *Panel data econometrics, Volume I: Theory*, Chapter 5. Elsevier Publisher, Amsterdam, The Netherland
- Laeven L, Levin R (2009) Bank governance, regulation and risk taking. *J Financial Econ* 93:259–275
- Sealey Jr. CW, Lindley JT (1997) Inputs, outputs, and theory of production cost at depository financial institutions. *J Financ* 32:1251–1266
- Stiroh KJ, Strahan PE (2003) Competitive dynamics of competition: evidence from U.S. banking. *J Money Credit Bank* 35:801–828
- Su L, Shi Z, Phillips PCB (2016) Identifying latent structures in panel data. *Econometrica* 84:2215–2264
- Tran KC, Tsionas EG (2013) GMM estimation of stochastic frontier model with endogenous regressors. *Econ Lett* 118:233–236
- Tsionas EG (2002) Stochastic frontier models with random coefficients. *J Appl Econ* 17:121–47
- Tsionas EG, Kumbhakar SC (2014) Firm heterogeneity, persistent and transient technical inefficiency: a generalized true random effects model. *J Appl Econ* 29:110–132
- Wang HJ, Ho CW (2010) Estimating fixed-effect panel stochastic frontier models by model transformation. *J Econ* 157:286–296