ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Accounting for Unobserved Time-varying Group-fixed Effects for Binary Dependent Variables

Victor Gosselink (512548)

Supervisor: Dr. Wendun Wang

Bachelor Thesis

Econometrics and Operations Research

July 3, 2021

Abstract

Time-varying heterogeneity is a common problem for financial research. I take an existing model for time-varying group effects, the GFE model, and show how this model can be modified to account for binary dependent variables in two different ways. The estimation methods use an iterative procedure which alternates between estimating parameters and grouping firms with the estimated parameters. I investigate the accuracy of the models though simulations and find that for dichotomous dependent variables, firms can be regrouped with an accuracy of up to 82 percent. Linear probability models with group fixed effects are an effective way of dealing with time-varying group heterogeneity, when little is known about the source of heterogeneity.

1 Introduction

Heterogeneity is a common problem when modeling financial data and can change the outcomes of research if not treated properly. An omitted variable bias occurs when the explanatory variables are correlated with the source of heterogeneity, this leads to inconsistent parameter estimates if the heterogeneity is not correctly incorporated in the model.

Current literature explores several ways to treat heterogeneity, but it often requires knowledge on the source of heterogeneity. Recently the group fixed effects (GFE) class of models has been developed, which group firms based on their reaction to shocks over time. GFE models assume that firms in a group react similarly to heterogeneous shocks, but that firms across groups can react differently. Firms are grouped together if their residuals move together over time. The model can be estimated by an iterative procedure that alternates between estimating groupings and parameters.

Sojli, Tham, and Wang (2018) implement a GFE model and show that it performs very well. Their model places firms in the correct groups over 98 percent of the time, with as few as 10 observations available per firm. The model gives consistent and unbiased estimates. It performs almost on par with infeasible models that assume the groupings known. In practice it is almost impossible to manually assign these groupings perfectly.

The existing literature on models for binary dependent variables assumes groupings known and firms are often grouped based on observable characteristics, such their industry. When groups are assumed to be heterogeneously affected through a (unknown) group structure, it is infeasible to manually assign these groupings. This leads to the following research question:

Can the GFE class of models be extended to incorporate binary dependent variables?

It is not as easy to disentangle heterogeneous effects from other effects, as the effect of the shocks are not always visible in the dependent variable. I combine the GFE estimation procedure with a logit model for fixed effects and a linear probability model. Through simulations I determine how well the firms can recover these groupings. The GFE estimation procedure for a linear probability model can recover over 80 percent of the original groupings, the logit GFE model can recover about 75 percent of the groupings.

The main contribution of this paper is that it extends the ideas of the GFE model to the binary class of dependent variables. It uses logit and linear probability models with fixed effects that have been used in the past to deal with heterogeneity and combines them with the iterative procedure used to estimate GFE models.

Section 2 summarizes literature on fixed effects for continuous and binary variables in panel data. Section 3 describes the estimation methods used for the groupings and parameters. Section 4 describes the simulations used to determine the relative performance of the models. Section 5 describes the simulation results. Section 6 presents a conclusion and some discussion points that could be a starting point for further research.

2 Literature review

In this section I discuss the various ways existing literature incorporates fixed effects in models. Fixed effects have been modeled as time-varying, varying intercepts for individuals or groups. Recently the GFE model has been developed to estimate fixed effects as time-varying shocks that affect groups heterogeneously. I start with continuous models, as previous literature assumes groupings known for binary variables.

2.1 Models for unobserved heterogeneity

When dealing with heterogeneity in finance, the dependent variable y is assumed to be affected in the same way within groups and differently across groups. Existing models are the two-way fixed effects (TFE) model, the the interacted fixed effects (IFE) model and the group fixed effects (GFE) model, discussed in Sojli et al. (2018).

Two-way fixed effects

The first model is the TFE, in which y_{it} depends on explanatory variables X in the following manner:

$$y_{it} = \alpha_i + \lambda_t + X_{it}'\beta + \varepsilon_{it} \qquad \text{for} \quad i = 1, ..., N; \quad t = 1, ..., T; \tag{1}$$

where λ_t are time-fixed effects and α_i are the unit-fixed effects. The regressors can be correlated with α_i and λ_t. The parameters are estimated with OLS after demeaning X and y in the cross-section and in the time series. This leads to elements of the form: $\tilde{X}_{it} = (X_{it} - \bar{X}_{i.} - \bar{X}_{.t} + \bar{X}_{..})$, where $\bar{X}_{i.}$ is the average of value of X for firm i over the time series, $\bar{X}_{.t}$ is the average value of X at time t and $\bar{X}_{..}$ is the average value of X. Y is transformed in a homologous manner. The model assumes that all individuals respond to the time-varying shocks in the same manner, which need not be the case. TFE estimation leads to biased results when the individuals respond differently to the time-varying shocks (Sojli et al., 2018).

Interacted fixed effects

Another model to estimate fixed effects is the IFE model. This model lets the effect of time-varying shocks vary over groups, which leads to unbiased estimates. A major limitation is that the model has no way of estimating these groups and requires a priori knowledge on the groupings. Firms are

often grouped by industry or another observably characteristic.
In practice this means that the model is infeasible as it is unlikely that a researcher can classify all firms in their original groups. Misspecification of the groups leads to inconsistent estimates if the group effects are correlated with X (Sojli et al., 2018).

Group-fixed effects

Sojli et al. (2018) describe a linear model with components $\theta_{gi,t}$ to account for time-varying group effects. The model assumes that responses to time-varying shocks differ over G distinct groups, but are similar within groups. There are i firms in N groups with $t = 1...T$ as the time of the observation. The model model assumes that y is generated as follows:

$$y_{it} = \alpha_i + \theta_{g_i,t} + X_{it}'\beta + \epsilon_{it}, \qquad g_i \in \{1, ..., G\}, \tag{2}$$

where α_i captures individual levels for firms, $\theta_{gi,t}$ is the unobserved time-varying group effect, g_i is the group firm i belongs to and X_{it} contains the explanatory variables. Sojli et al. jointly estimate the coefficients β and group parameters $\theta_{gi,t}$ with Algorithm 1, introduced by Bonhomme and Manresa (2015). Sojli et al. find that the algorithm can retrieve the original groupings almost perfectly through simulations. This estimation methods leads to no significant bias in $\hat{\beta}$. Let $\dot{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\dot{X}_{it} = X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}$. The estimation algorithm can be described as follows:

Algorithm 1

1. Let $g^{(0)}$ be an initial value of grouping. Set s=0.
2. For the given $g^{(s)}$, compute:
$$(\theta^{(s+1)}, \beta^{(s+1)}) = \underset{\beta,\theta}{\arg\min} \sum_{i=1}^{N}\sum_{t=1}^{T}(\dot{y}_{it} - \dot{X}_{it}'\beta - \theta_{g_i^{(s)},t})^2$$
3. compute for all $i \in \{1, ..., N\}$
$$g_i^{(s+1)} = \underset{g\in 1,...,G}{\arg\min} \sum_{i=1}^{N}\sum_{t=1}^{T}(\dot{y}_{it} - \dot{X}_{it}'\beta^{(s+1)} - \theta_{g_i,t}^{(s+1)})^2$$
4. Set $s = s + 1$ and go to Step 2 (until numerical convergence)

The algorithm is used because of the large number of possible combinations of groupings. As the the objective function (3) does not linearly depend on the groupings, it is not possible to find an analytical solution that minimizes Q_{NT} (Sojli et al., 2018).

The least squares objective function is:

$$Q_{NT} = \min_{\theta,\beta,g} \sum_{i=1}^{N} \sum_{t=s}^{T} (\dot{y}_{it} - \dot{X}_{it}'\beta - \theta_{g_i,t})^2 \tag{3}$$

This algorithm can end op in local optima, so one should use many starting values to see which one results in the lowest sum of squared residuals. Another complication is that the number of groups is unknown. Sojli et al. (2018) therefore use the information criterion proposed by Bonhomme and Manresa (2015) and Su, Shi, and Phillips (2016) to determine the number of groups.

$$BIC(G) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\dot{y}_{it} - \dot{X}_{it}'\hat{\beta}_{\text{gfe}}^{(G)} - \hat{\theta}_{g_i,t}^{(G)})^2 + \hat{\sigma}^2 \frac{GT+K+N}{NT} \log(NT) \tag{4}$$

It is better to have G slightly too large than too small. If G is chosen too small, the parameter estimates are inconsistent. The GFE model groups firms based on the residuals of the time series. Firms with residuals that move together over time are placed in the same group. The parameter estimates closely match that of the infeasible IFE model, which assumes known groupings. Over 98 percent if firms are grouped correctly with as few as 10 observations per firm.

2.2 Fixed effects for binary dependent data

This section describes literature on fixed effects for binary dependent variables. Unfortunately I could not find an example where the authors did not impose the groupings themselves.

Most of the recent literature on fixed effects for binary dependent variables is not from the field of finance or economics, but from other social sciences. Existing literature agrees on a common assumption; asymptotics have to be in the groups and not the individual to avoid an incidental parameter problem, where the number of parameters to be estimated increases with N. This prevents the parameter estimates from converging to their true values. The most common models used for binary y_{it} with fixed effects are: logit models with grouping dummies, conditional logit models and linear probability models with fixed effects. I define p_{it} as $p(y_{it} = 1)$. All three of these models can be written in the following form:

$$p_{it} = H(X_{i,t}\beta + \theta_{g_i}) \tag{5}$$

where g_i is the group of firm i and θ_{g_i} are the fixed effects for group g_i.

Linear probability models

The simplest model for probabilities is the linear probability model. Equation H in equation **??** doesn't do anything in this case. p_{it} can be written as:

$$p_{it} = X_{g_i,t}\beta + \theta_{g_i} + \varepsilon. \tag{6}$$

This equation can be estimated by regressing y on X and the corresponding group dummies. A limitation of this class of models is that linearity is imposed on the probability function. Additionally, $\hat{p}$ is not bounded between zero and one.

Logit with dummies

The logistic model with dummies is an alternative to the linear probability model. Function H is now the logistic function and p_{it} can be written as:

$$p_{it} = \frac{1}{1 + \exp\left(-\alpha_i - X_{it}'\beta - \theta_{g_i}\right)}. \tag{7}$$

Now p_{it} is bounded between zero and one. A problem with this specification is that groups with no variation in y_{it} will have no influence on $\hat{\beta}$, as each group has its own intercept. This is intuitive, as we can choose θ_{g_i} sufficiently small (large) for groups where every observation is zero (one) to always get the right prediction of y, regardless of $\hat{\beta}$. As these groups might contain useful information on the DGP of y, this poses a serious problem.
Estimating LPMFE with OLS lets the information be included in β. This leads to a smaller $\hat{\beta}$ and smaller standard errors (Beck, 2018).

As groups with no variation will be more prevalent when p_{it} is low (high). This helps explain the findings of Timoneda (2021), who find that LPMFE outperforms logit with dummies when $y = 0$ (trivially $y = 1$) is rare.

Conditional logit

Conditional logit gives consistent estimates of β. But as the group effects are conditioned out, it provides no estimates of θ (Coupé, 2005). Conditional logit as described by Timoneda (2021), conditions on the known number of positives in a group and calculates the conditional probability for a vector y_g occurring as:

$$p(y_g|k_g) = \frac{exp(\sum_{t=1}^{T_g} y_{gt}X_{it}\beta)}{\sum_{d_g \in s_g} exp(\sum_{t=1}^{T_g} d_{it}X_{it}\beta)} \tag{8}$$

where T_g is the number of observation in a group, k_g is the number of successes (ones) in group g, d_g is a vector with k_g ones and $T_g - k_g$ zeros, s_g is the set of all possible combinations of drawing

k_g successes out of T_g and d_it are the observations in d_g. Estimation of the model comes with the assumption of no fixed effects when calculating predicted probabilities, which is not correct. The model overestimates the probabilities when X is at its smallest for rare events (Timoneda, 2021). This estimation method does not suffer from an incidental parameter problem. However, it cannot be used in combination with a GFE model, as it provides no estimates of θ. This is necessary as it is not feasible to compute all possible groupings; there are too many possible combinations. If we have a group of 200 observations with 100 successes there are 9.05e58 possible combinations.

2.3 The effects of G and p

It is noteworthy that adding a lot of parameters introduces a positive bias in ML estimation, but this bias is generally small (Timoneda, 2021). The bias is stronger for larger G, as more parameters have to be estimated. Timoneda shows that the bias is low for both logit models through simulations. For 2500 individuals, fifty groups and $p = 0.01$ the bias is eight percent for logit with dummies and two percent for conditional logit. For events that are less rare the bias is lower. As GFE models estimate a separate parameter for every group g at every time t, I suspect there might be a significant positive bias in the logit models. Beck (2018) finds that logit models perform worse compared to conditional logit model, when dealing with fixed effects. Especially for many smaller groups.

3 Methodology

In this section I describe the two fixed effects models that I use to estimate the simulations. Both models make use of a modified algorithm as proposed by Bonhomme and Manresa (2015). The groupings will never be recovered as well when y is binary, because y holds less information. We only know whether y is above or under a certain threshold, so there are often multiple plausible groups in which a firm can be placed. Because T is modest and y doesn't have to vary over time for a given firm, this poses a serious challenge.

3.1 GFE for linear probability models

Timoneda (2021) finds that LPMFE has its merits, despite its simplicity. Estimating the model is not different from estimating a regular regression, therefore it is easily implemented with group fixed effects. I modify the model to allow for time-varying group fixed effects. The linear probability model with group fixed effects (LPMGFE) model becomes:

$$p_{it} = X_{it}\beta + \theta_{g_i,t} + \varepsilon, \tag{9}$$

I use Algorithm 1 to estimate the groupings and parameters for the LPMFFE. The only difference is that I replace $\dot{y}$ with y. In other words: the binary variable is not demeaned.

3.2 Logit with dummies

Recall the logistic regression with dummies from section 2.2. Again I modify the model to allow for time-varying group fixed effects. p_{it} becomes:

$$p_{it} = \frac{1}{1 + \exp\left(-\alpha_i - \dot{X}_{it}'\beta - \theta_{g_i t}\right)}. \tag{10}$$

I modify the algorithm to maximize the following likelihood function, which assumes that y_{it} are all independently Bernoulli distributed:

$$L(\theta|y, X) = \prod_{i=1}^{N} \prod_{t=1}^{T} p_{it}^{y_{it}} * (1 - p_{it})^{(1-y_{it})}. \tag{11}$$

With this likelihood I formulate algorithm 2. This algorithm minimizes $-log(L)$ in steps 2 and 3 (maximizing the likelihood). For microeconomic or financial data the implicit assumption of independence often does not hold due to correlated sequential observations. Two examples: If we classify firms as dividend-payer or non-payer, the assumption does not hold because firms rarely stop paying dividends once they start doing so. This is seen as a bad sign by investors. y_t depends on y_{t-1} as well as X_t. If we classify individuals as employed or unemployed, it is obvious that sequential observations are correlated as well.

Adding a unique intercept α_i for each firm would result in an incidental parameter problem. $\dot{y}_{it}$ cannot be used here, as this variable would not be binary. Therefore I add $\bar{y}_{it}$ in the logistic function. Adding firm dummies instead would cause multicollineary on top of the incidental parameter problem, as the group dummy would be exactly equal to the sum of the firm dummies that correspond to the group. This can be resolved by setting one firm dummy to zero for each group, although this does change the interpretation of the parameters.

This model does not recover the original DGP, because there are infinitely many combinations of the parameters that would lead to these outcomes for y. Two examples are: adding a constant c to both α and q will lead to the same binary variable y. Scaling all parameters (and the idiosyncratic shocks) by a factor c would give the same outcomes for y, as the cutoff is based on a quantile of y. This complicates analysis on the accuracy of the parameter estimates, but we can verify the accuracy of the groupings. $\bar{y}_i$ serves as a base probability for a firm and clearly has a different interpretation than α_i. The estimated parameters β can be interpreted as the estimated change in probability of observing a success for a unit change in X.

Define $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$. I modify Algorithm 1 and obtain the following algorithm:

Algorithm 2

1. Let $g^{(0)}$ be an initial value of grouping. Set s=0.

2. For the given $g^{(s)}$, compute:

$$(\theta^{(s+1)}, \beta^{(s+1)}) = \arg\min_{\beta,\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} -y_{it} \log(\hat{p}_{it}) - (1 - y_{it}) \log(1 - \hat{p}_{it}),$$

   where

$$\hat{p}_{it} = \frac{1}{1 + \exp\left(\bar{y}_i + X_{it}'\beta + \theta_{g_i^{(s)},t}\right)}.$$

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

$$g_i^{(s+1)} = \arg\min_{g \in 1,...,G} \sum_{i=1}^{N} \sum_{t=1}^{T} -y_{it} \log(\hat{p}_{it}) - (1 - y_{it}) \log(1 - \hat{p}_{it}),$$

   where

$$\hat{p}_{it} = \frac{1}{1 + \exp\left(\bar{y}_i + X_{it}'\beta^{(s+1)} + \theta_{g_i,t}^{(s+1)}\right)}.$$

4. Set $s = s + 1$ and go to Step 2 (until numerical convergence)

I refer to this model as the logit model with group fixed effects (logitGFE). The parameters β can be interpreted as the change of the log-odds for a unit change in X, where the $log - odds$ is defined as $log(\frac{p}{1-p})$.

4 Simulation

This section describes the data generating processes used to test the accuracy of the the models described in section 3.

4.1 Continuous GFE model

To verify the accuracy of my implementation of the GFE model for continuous y, I generate 100 sets of data with 1000 firms and 10 observations per firm in the same way as Sojli et al. (2018). This data generating process (DGP) is designed to mimic common financial panel data, with a large panel size (N) and a modest time span (T). The notation follows Sojli et al. (2018).
First I generate $\tau_{l,g_i,t}$ from $G = 5$ independent normal distributions, with mean g and variance $2g$. Let X_{it} and $\theta_{g_i,t}$ both depend on the underlying component $\tau_{l,g_i,t}$ with the following structure:

$$X_{l,it} = c_\tau \tau_{l,g_i,t} + v_{it} \qquad \text{for} \quad l = 1, 2 \tag{12}$$

and

$$\theta_{g_i,t} = g_i^2(\tau_{1,g_i,t}\tau_{2,g_i,t})/c_\theta \tag{13}$$

The error terms v_{it} are the idiosyncratic shocks for each firm-year, drawn from the normal distribution with unit mean and variance. Intuitively, one can interpret $\tau_{l,g_i,t}$ as the group-specific driver of X_l, it, which influences both X and θ. I set $c_t = 0.5$ and $c_\theta = 0.15$ to replicate the DGP with moderate shocks from Sojli et al. (2018).

With the group component θ and X we generate the dependent variable y in the following manner:

$$y_{it} = \alpha_i + \theta_{g_i,t} + \sum_{l=1}^{2} X_{l,it}\beta_l + \varepsilon_{it}, \qquad i = 1, ..., N; \qquad t = i, ..., T; \varepsilon_{it} \sim \mathcal{N}(0, 1)\,, \tag{14}$$

where $\alpha_i = 1/T \sum_{t=1}^{T} X_{1,it}$ and $\beta = (1, 2)'$.

I then estimate the parameters for the simulated data with the GFE model described in Algorithm 1.

4.2 Binary GFE models

For the binary case, I generate new datasets in two ways: according to a Bernoulli distribution and in a linear fashion.

Linear DGP

I first generate a binary variable $y*$ by generating the continuous variable y as described in the previous section. To transform y into a binary variable, I select a quantile q from y. For all y larger than q, $y*$ is set to 1 and for all y smaller or equal, $y*$ is set to 0. I generate 20 datasets for $q = 50\%$.

Bernoulli DGP

I first generate α_i, X_{it} and $\theta_{g_i,t}$ as explained in section 4.1. Then I draw binary variables $y*$ according to a Bernoulli distribution with $p_{it} = \frac{1}{1+\exp{(y_{it}-\bar{y})}}$, where $y_{it} = \alpha_i + X_{it}\beta + \theta_{g_i,t}$ and $\bar{y} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(\alpha_i + X_{it}\beta + \theta_{g_i,t})$. y is demeaned so the probability p(y=1) will be close to 0.5, the distribution is positively skewed so this is not be exact.[1] Again, $N = 1000$ and $T = 10$.

[1] If we take $\bar{y}$ as the median of y instead, p is exactly 0.5.

Estimation

I estimate the datasets with logitGFE as well as LPMGFE. To determine the difference in the parameter estimates from determining the groupings; I estimate the models as well with infeasible logit and least squares models. Here the groupings don't have to be estimated, but only β and θ.

5 Results

This section describes the estimation results for the simulations.

5.1 Accuracy GFE implementation

Groupings

To determine the the accuracy with which the algorithms reconstruct the originally imposed groupings, I compute the average misclassification frequency (AMF) over all repetitions r, defined as:

$$\text{AMF} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{N}\sum_{i=1}^{N}1(\hat{g}_i^r \neq g_i^0), \tag{15}$$

where $\hat{g}_i^r$ is the estimated group of firm i in repetition r and g_i^0 is the true group of firm i in repetition r. The estimation of the GFE DGP with moderate shocks leads to an AMF of 0.0709. This is not as low as in Sojli et al. (2018), who find an AMF of 0.0124. The difference can likely be explained as the result of limited computing power. Table 1 displays the results from the simulation with y continuous.

Table 1: Results of GFE estimation.

	AMF	bias X1	bias X2	std X1	std X2	RMSE
GFE	0.0709	-0.0117	0.0229	0.2338	0.2349	0.3337

Parameters

To determine the accuracy of the models, I compute the bias, standard deviation and root mean squared error. The bias for the regular GFE model, defined as

$$\text{Bias}(\hat{\beta})\frac{1}{R}\sum_{r=1}^{R}\hat{\beta}_r - \beta^0, \tag{16}$$

is -0.0117 for X_1 and 0.0229 for X_2. To determine the uncertainty and accuracy of the parameter estimates I compute the empirical standard deviation

$$\mathrm{Std}(\hat{\beta}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\beta}^r - \bar{\hat{\beta}})^2} \tag{17}$$

and the root mean squared error (RMSE)

$$\mathrm{RMSE}(\hat{\beta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\|\hat{\beta}_r - \beta^0\|^2}, \tag{18}$$

where $\hat{\beta}_r$ is the parameter estimate of the r-th replication and β^0 is the true parameter value. The least squares GFE simulation results in a standard deviation of 0.2338 and 0.2349, for X_1 and X_2 respectively, and a RMSE of 0.3337.

5.2 Simulations with y binary

In this section I discuss the results from the logit model with dummies and the linear probability model with fixed effects. Table 2 displays the results. Bias and RMSE have no obvious meaning, due to different interpretation of the estimated parameters and are therefore not displayed. It is immediately clear that the LPMGFE model classifies firms better, with an AMF of 0.2573 for the linear DGP and 0.2637 for the Bernoulli DGP. The model differs less from its infeasible counterpart than the logitGFE model. The logitGFE model suffers more from the unknown groupings; it has a higher misclassification frequency and the parameters are about twice the size of their infeasible counterparts.

Table 2: Estimation results for 10 initial groupings over 20 repetitions.

DGP	Estimation	feasible	AMF	β_1	β_2	std(β_1)	std(β_2)
Linear	LPMGFE	yes	0.2573	0.0374	0.0760	0.0277	0.0295
		no	-	0.0580	0.0909	0.0191	0.0257
	logitGFE	yes	0.3302	1.9322	2.9614	0.6019	0.6355
		no	-	0.8048	1.2798	0.3073	0.2984
Bernoulli	LPMGFE	yes	0.2637	0.0287	0.0644	0.0199	0.0204
		no	-	0.0572	0.0858	0.0151	0.0230
	logitGFE	yes	0.3272	1.5384	2.3404	0.6810	0.6599
		no	-	0.7850	1.0620	0.1591	0.1874

As the linear probability model runs best with these settings, I rerun the simulations for LPMGFE model with a hundred initial groupings. I only rerun this model as it is almost an order of magnitude faster than the logisticGFE model. Again, y is generated as described in section 4.2, with $t = 10$, $N = 1000$, $p = 0.5$ and 20 repetitions. Table 3 displays the results. It is clear that the LMPGFE primarily benefited from the higher number of initial groupings when estimating the linear DGP. In this case the AMF drops by 28 percent. For the Bernoulli GDP the difference is smaller, with a reduction in the AMF of 1.5 percent. This is not unexpected as the linear model misspecifies the DGP in this case.

If one assumes that y is generated according to a linear probability model with time-varying group fixed effects, then one has the best chance of recovering most of the original groupings and should use the LPMGFE model. It is not evident that the assumption of linearity will hold in many cases.

Table 3: Estimation results for 100 initial groupings over 20 repetitions.

DGP	Estimation	feasible	AMF	β_1	β_2	std(β_1)	std(β_2)
Linear	LPMGFE	yes	0.1844	0.0342	0.0766	0.0112	0.0244
		no	-	0.0580	0.0909	0.0191	0.0257
Bernoulli	LPMGFE	yes	0.2598	0.0298	0.0543	0.0143	0.0179
		no	-	0.0572	0.0858	0.0151	0.0230

6 Conclusion

The GFE model can cluster firms very well based on their reaction to time-varying shocks. When extending the idea of the model to the binary case, the groupings cannot be recovered as well. The logitGFE model misclassifies firms more often than the LPMGFE model and its parameter estimates suffer. The LPMGFE model is impacted less severely by the unknown groupings. While not perfect, the LPMGFE model is a good starting point when incorporating fixed effects for binary dependent variables.

6.1 Discussion

A limitation is that all simulations, except for the one with continuous dependent variables, are done with only 20 repetitions and a limited number of initial groupings. It is possible that the models perform better when given more time and computing power.

Further research can focus on improvements in the current models, or try to find more ways to groups firms based on observable characteristics. As there is not as much information in a binary dependent variable, there likely is no a satisfying solution that can be applied in all cases of time-varying group heterogeneity for binary dependent variables. Further research can investigate how the models perform for different p and T. If successes (failures) are rare, I expect the linear probability model with fixed effects to have a greater advantage. For higher T all models should perform better.

References

Beck, N. (2018). Estimating grouped data models with a binary dependent variable and fixed effects: What are the issues. *arXiv preprint arXiv:1809.06505*.

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, *83*(3), 1147–1184.

Coupé, T. (2005). Bias in conditional and unconditional fixed effects logit estimation: A correction. *Political Analysis*, 292–295.

Sojli, E., Tham, W. W., & Wang, W. (2018). Time-varying group unobserved heterogeneity in finance. *Available at SSRN 3258048*.

Su, L., Shi, Z., & Phillips, P. C. (2016). Identifying latent structures in panel data. *Econometrica*, *84*(6), 2215–2264.

Timoneda, J. C. (2021). Estimating group fixed effects in panel data with a binary dependent variable: How the lpm outperforms logistic regression in rare events data. *Social Science Research*, *93*, 102486.