Market-wide Events and Time Fixed Effects*

Elvira Sojli†
University of New South Wales

Wing Wah Tham‡
University of New South Wales

Wendun Wang§
Erasmus University and Tinbergen Institute

Abstract

Market-wide events (e.g., financial crises) and regulatory changes empirically have group heterogenous impact on firm outcomes. Inappropriate modelling of the heterogeneity by existing methods such as time-fixed effect (assumes a homogenous response to shocks) and industry-year interacted fixed effect (assumes a heterogenous responses to shocks based on industry) is likely to result in biased estimates. This paper investigates the effect of heterogenous responses to common shocks for existing panel studies. We demonstrate theoretically and empirically that ignoring time-varying unobserved heterogeneity that is correlated with regressors in current empirical practices leads to biased estimates and standard errors. To overcome the bias, we propose the use of the "group fixed effect, GFE" class of models, which produce consistent estimates even under the two-way fixed effect and interacted fixed effect data generating process. We extend the GFE class of models to accommodate two-stage least square estimators. We study the finite sample properties of GFE through simulations and demonstrate its economic importance with two empirical applications. Finally, we provide researchers with guidance and user-written functions in statistical packages to overcome the limitations of existing approaches.

Keywords: Time-varying unobserved heterogeneity; clustering; common shocks; group fixed effects; fixed effects; heterogeneity bias.
JEL Classification: G10; G20; G14.

*We are very grateful to Renee Adams, Tomohiro Ando, Oleg Chuprinin, Sudipto Dasgupta, Kose John, Jean-Noel Barrot, Stefano Lovo, Alexi Ovtchinnikov, Jean Edouard Colliard,Thierry Foucault, Yeejin Jang, Ernst Maug, Michael McAleer, David Reeb, Chang Mo Kang, Kristle Romero-Cortes, Ioanid Rosu, Daniel Schmidt, Breno Schmidt, Rik Sen, BaoLian Wang, and Jason Zein for insightful and helpful discussions. We thank seminar participants at HEC Paris, the Econometric Institute of Rotterdam, UNSW and conference participants at the 25th International Panel Data Conference, the Econometric Society Asia Meeting 2019, CICF 2018, FIRN 2018, the International Workshop of Methods in International Finance Network 2018, and SETA 2018 for helpful comments.

†Address: School of Banking and Finance, UNSW Business School, UNSW Sydney, NSW 2052 Australia. Email: e.sojli@unsw.edu.au.

‡Address: School of Banking and Finance, UNSW Business School, UNSW Sydney, NSW 2052 Australia. Email: w.tham@unsw.edu.au.

§Address: Erasmus School of Economics, Erasmus University, PO Box 1738, Rotterdam, 3000DR, the Netherlands. Email: Wang@ese.eur.nl

1 Introduction

Market wide events like introduction of financial regulations (e.g. Sarbanes-Oxley Act, SOX) and financial crises affect firms' outcomes and policies. A salient empirical finding of the literature on market-wide events is that these shocks have heterogenous group effects on firm outcomes, i.e. firms in the same group respond similarly to the shocks, while the responses differ across groups.[1] Response heterogeneity across groups may be due to differences in firms' time-varying corporate culture, product markets, business strategies, and attitude towards risk and may be captured by observable and unobservable managerial and firm attributes (e.g., profitability, governance, geography, industry, credit rating, managerial qualities, etc.). Given the empirical findings several questions arise: how should one account for heterogenous effects of common shocks in panel regressions? Does ignoring the heterogeneity in econometric specifications affect slope estimates of interest, statistically and economically? What determines the degree of bias? How should one account for the endogenous grouping of firms with heterogenous responses in econometric models? In this paper, we provide answers to these questions through econometric derivations, simulations and empirical examples.

While most finance scholars regularly observe heterogenous responses to common (regulatory) shocks in financial markets, they routinely use time fixed effects with homogenous impact across firms, λ_t, to control for these shocks specified as:

$$y_{it} = \alpha_i + \lambda_t + X_{it}'\beta + \epsilon_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T,$$

where the individual and time fixed effects α_i and λ_t are allowed to be correlated with covariates X_{it}, but not with the idiosyncratic error ϵ_{it}. As an example, we survey recently published empirical papers in the top three finance journals for the use of fixed effects. In 2017 and 2018, there are 389 empirical papers using panel data in the Journal of Finance (76), the Review of Financial Studies (156), and the Journal of Financial Economics (157). There are 383 papers that use panel regressions and 6 that do not. Among those using panel regressions, 359 papers use fixed effect models: 95 papers only use one-way fixed effects and 264 use two-way (individual and time) fixed effect (TFE) models.

Only 81 papers employ an interacted fixed effects (IFE) model, relaxing the homogeneity assump-

[1] For example see Mitton (2002), Joh (2003), Lemmon and Lins (2003), Banerjee et al. (2015), Thakor (2015), Linck et al. (2009), Duchin et al. (2010a), Duchin et al. (2010b) among many others.

tion. Although interacted fixed effects models in finance mostly include time dummies interacted with industry or state dummies, the grouping of firms should depend on the context of the research question and the economic outcomes of interest. The assumption that heterogenous responses are driven by one covariate, either industry or state, is unlikely to be correct, because: (1) it is inconsistent with the empirical evidence that the group heterogeneity is related to observable and to unobservable corporate strategies, risk attitude, managerial and firm characteristics, and not (at least only) to industry or states;[2] (2) the pattern of group heterogeneity typically depends on the outcome variable of interest, and it is difficult for a researcher to know the exact grouping of the heterogeneity, ex ante;[3] (3) few categorizations of individuals and firms in finance are intrinsically exogenous (gender is one of the few cases). Categorization and grouping of economic agents and firms is probably endogenously determined through both observable and unobservable determinants.

We study the econometric and economic implications of heterogenous effects of market-wide shocks for the current econometric specifications, practices, and findings in the finance literature. We theoretically derive the bias in the coefficient estimates and standard errors of current practices under the data structure with time-varying group heterogeneity. We call this ***heterogeneity bias*** and find that heterogeneity bias can be economically substantial in estimates of TFE and IFE models. The degree of bias depends on the correlation between the grouped time fixed effect and the explanatory variables, as well as on the size of the heterogeneity. Particularly, the bias of coefficient estimates is high when heterogenous responses to market-wide shocks are large and highly correlated with X.[4]

[2]Two recent examples of the heterogeneous impact of common shocks are the implementation of the SOX Act and the 2008 financial crisis. The Securities and Exchange Commission SOX Act in 2002 affected a myriad of market participants, who are regularly expected to make firm-related decisions, e.g. firm managers and executive boards, and who affect firms and managers decisions, e.g. auditors and institutional investors. SOX was implemented to strengthen the independence of auditing firms, to improve the quality and transparency of financial statements and corporate disclosure, to enhance corporate governance, to improve the objectivity of research, and to strengthen the enforcement of the federal securities laws. Several recent papers show the heterogenous effect of SOX on corporate governance, due to the heterogeneity in observable and unobservable managers and firm characteristics, see Engel et al. (2007), Chhaochharia and Grinstein (2007), Heron and Lie (2007), Koh et al. (2008), Cohen et al. (2008), Linck et al. (2009), Bargeron et al. (2010), Carter et al. (2009), Duchin et al. (2010a), Duchin et al. (2010b), Brochet (2010), and Banerjee et al. (2015). Work on recent financial crises finds that observable firm characteristics (e.g., industry, profitability, and credit rating), as well as unobservable characteristics (e.g., managerial qualities and characteristics) affect firms' outcome variables in the cross-section (see Mitton, 2002; Joh, 2003; Lemmon and Lins, 2003; Campello et al., 2010, 2011; Thakor, 2015; Ho et al., 2016).

[3]In practice, the group membership that leads to heterogeneity is typically unknown, making such an interacted fixed effect estimator an "infeasible" model or impossible to specify correctly, especially when unobservables are driving the heterogenous responses. When the pre-specified group structure (e.g. industry or state) does not coincide with the underlying heterogeneity pattern, the coefficient estimates produced by IFE are biased, and therefore, IFE does not capture the heterogenous response to shocks.

[4]If the heterogeneous time fixed effect is independent from X, the slope coefficients can still be consistently estimated using standard two-way fixed effect models.

We overcome the problem of heterogeneity bias by introducing a new class of models, "group fixed effect" (GFE) models, which leaves the groups unrestricted and categorizes them *endogenously*. GFE models are analogous to the two-way fixed effect models but allow for heterogenous responses to market shocks, drawing insights from Bonhomme and Manresa (2015a). GFE allows the heterogeneity to be characterized by a latent group structure, which is endogenously recovered from the specified model and the data rather than exogenously imposed by researchers, such as industry grouping. The grouping of firms allows for a simple iterative procedure for the GFE estimation with desirable asymptotic and finite-sample properties (Bonhomme and Manresa, 2015a). We also propose a method to evaluate whether GFE is the appropriate model in comparison to TFE and IFE, based on a Hausman specification test. In addition, we extend GFE with a new methodology, instrumental variable "group fixed effect" (IV-GFE), to address heterogeneity bias as well as endogeneity concerns faced in most empirical research in finance.[5]

In general, GFE allows for heterogenous responses across groups to market-wide shocks through $\theta_{g_i,t}$:

$$y_{it} = \alpha_i + \theta_{g_i,t} + X'_{it}\beta + \epsilon_{it}, \quad g_i = 1, 2, ..., G,$$

where g_i is the group membership of firm i, G is the number of groups, and the group specific time effects $\theta_{g_i,t}$ can be regarded as group interacted with time dummies. Intuitively, GFE groups firms whose time profiles of outcomes conditional on covariates, i.e. residuals, are most similar. More specifically, it employs the "kmeans" method (Forgy, 1965; Steinley, 2006) that estimates group membership using least squares.

Asymptotically, the estimated group membership of each unit converges to the true population membership, as the time dimension increases. We examine the finite sample properties of the group membership estimator via Monte Carlo simulations under different data generating processes (DGPs) and shock characteristics. The group misclassification probability tends to zero at an exponential rate and each unit converges to the true population membership, as the time series increases.[6] Empirically, we show the reliability of the group estimator using the setting of natural disasters in (see Barrot and Sauvagnat, 2016), where we ex-ante know the observable determinants of group membership. Specifically, we back out group memberships from a GFE regression of sales

[5]IV-GFE is analogous to existing instrumental variable (IV) estimators such as two-stage least square (TSLS) and the generalized method of moments (GMM) estimators.

[6]The standard convergence rate of least square estimates of β in fixed effects models is $\sqrt{NT}$. The newly proposed GFE procedure with TSLS estimation of slope coefficients also leads to accurate estimates of group memberships, in the presence of endogenous variables.

growth on natural disasters. We observe that GFE groups cluster according to the location and magnitude of natural disasters, differently from TFE and IFE estimates.

We also conduct horse races via Monte Carlo simulations to evaluate the efficacy of the two-way fixed effect and interacted fixed effect models with respect to the GFE model, under different DGPs. The coefficient estimates produced by GFE are consistent, while those produced by TFE and IFE can be severely biased when time effects are heterogeneous and the heterogeneity pattern does not coincide with the specified grouping of IFE. These results show that GFE can capture heterogenous responses and overcome the risk of misspecifying the group pattern in IFE models. We also show that GFE provides consistent coefficient estimates not only under the assumption of time-varying unobserved heterogeneity, but it also produces consistent estimates, when units respond homogeneously (or heterogeneously based on industry) to market-wide shocks under the DGP of TFE and IFE. In other words, GFE is a ***robust method*** to incorporate time effects, providing consistent estimates no matter whether the time effect is heterogeneous or homogeneous.

Finally, we highlight the economic importance of GFE empirically by revisting the evidence in Sunder et al. (2017). We examine the relation between CEO attributes and corporate innovation using different estimators and find that the difference between GFE and competing methods is economically substantial. The results show that TFE and IFE can result in large biases in slope estimates by ignoring unobserved time-varying variables like corporate strategies and culture.

Overall, our work is closely related to Gormley and Matsa (2014), which provides practical guidance on how to handle unobserved (group) heterogeneity in empirical research. Complementing their work, our results challenge the standard assumptions of empirical estimation in the presence of unobserved group heterogeneity in empirical finance research, by showing the impact of heterogenous responses to market-wide events on empirical studies. We propose a new robust method to overcome the heterogeneity bias.

We also contribute to the econometric literature of panel group structure models by complementing the work of Bonhomme and Manresa (2015a) to jointly estimate coefficients and group membership using TSLS in the presence of endogenous regressors. Furthering the applicability of their work, we study GFE in general settings that are commonly encountered by finance empiricists and compare them to various popular techniques used in the finance literature. Currently, the literature does not inform us on the properties of GFE estimates under TFE and IFE DGPs and on the magnitude of the bias, when using TFE and IFE under time-varying heterogeneity. This is the first paper to explicitly derive the heterogeneity bias, which is determined by the shock size

and the correlation between heterogeneous responses and regressors. We show that GFE is a robust estimator for TFE and IFE DGPs both theoretically and in simulations. We also consider the case of time-varying group memberships and discuss under which situations GFE can capture the dynamics of the group membership structure. Finally, we propose a specification test to evaluate GFE against TFE and IFE. To encourage and facilitate the use of GFE estimators, we provide guidance on its use and STATA user-written functions.

Our paper is in line with the burgeoning work on econometric challenges faced by researchers in empirical methods. Bertrand et al. (2004), Petersen (2009), and Thompson (2011) discuss methods on appropriately computing standard errors in the presence of cross-sectional and time series dependence across residuals. Erickson and Whited (2012) and Koh et al. (2018) discuss issues on measurement error in investment regressions and missing data problem in R&D and patents.

The paper is organised as follows. The next section introduces the three main methods we investigate, TFE, IFE and GFE, derives the coefficient bias of different econometric models under a GFE DGP, and proposes the specification test. Section 3 conducts Monte Carlo simulations based on simulated data and the empirical distribution of financial data. Section 4 provides evidence on the economic importance of GFE by replicating current work in finance. Section 5 provides the two stage least square extensions and discusses issues related to standard errors. Section 6 concludes.

2 Unobserved Heterogeneity

In this section, we analyse issues related to unobserved heterogeneity. First, we discuss the most widely used empirical method to address unobserved heterogeneity, namely the two-way (individual and time) fixed effects model. We then derive the bias of TFE estimates when outcome variable is characterized by time-varying group heterogeneity, i.e. the DGP of GFE. Second, we discuss the interacted fixed effects estimate and derive its bias under the same DGP. Finally, we present the GFE model and discuss its properties.

2.1 Two-way fixed effects model

We consider the following two-way fixed effect model:

$$y_{it} = \alpha_i + \lambda_t + X_{it}'\beta + \epsilon_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T \tag{1}$$

where y_{it} is the dependent variable for individual i at time t, α_i models the individual fixed effects, λ_t models the time fixed effects, X_{it} is a $K \times 1$ vector of regressors, and ϵ_{it} is the error term. With strict exogeneity, it is typically assumed that

$$\begin{aligned}
&\epsilon_{it} \sim \text{i.i.d } (0, \sigma_\epsilon^2),\\
&\text{var}(X_{it}) = \sigma_X^2,\\
&\text{cov}(\epsilon_{it}, X_{it}) = 0,\\
&\text{cov}(\alpha_i, X_{it}) \neq 0,\ \text{cov}(\lambda_t, X_{it}) \neq 0,\\
&\text{cov}(\epsilon_{it}, \alpha_i) = 0,\ \text{cov}(\epsilon_{it}, \lambda_t) = 0.
\end{aligned}$$

These assumptions imply that the regressors, individual and time fixed effects are uncorrelated with errors. However, regressors can be arbitrarily correlated with individual and time fixed effects as in Gormley and Matsa (2014). As pointed out by Gormley and Matsa (2014), ignoring the nonzero covariance of the unobserved characteristics and the explanatory variable of interest results in inconsistent estimates. For notation convenience, we assume that the means of the fixed effects and the independent variables are zero without loss of generality.

To estimate this model, we define $y_i = (y_{i1}, \ldots, y_{iT})'$ and $X_i = (X_{i1}, \ldots, X_{iT})'$ for unit i. We define $y = (y_1, \ldots, y_N)'$ and $X = (X_1, \ldots, X_N)'$ across N cross-sectional units.[7] Two-way fixed effects estimators can be obtained by applying least square on the transformed data, data that is demeaned both in the cross-section and in the time series dimension. In particular, we define the transformation matrix M as:

$$M = I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T, \tag{2}$$

where I_N is an identity matrix of dimension N, I_T is an identity matrix of dimension T, J_T is a matrix of ones of dimension T, J_N is a matrix of ones of dimension N, $\bar{J}_N = J_N/N$, $\bar{J}_T = J_T/T$, and $\otimes$ is the Kronecker product. The transformed data $\widetilde{y} = My$ and $\widetilde{X} = MX$ have typical elements as $\widetilde{y}_{it} = (y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}_{\cdot\cdot})$ and $\widetilde{X}_{it} = (X_{it} - \bar{X}_{i\cdot} - \bar{X}_{\cdot t} + \bar{X}_{\cdot\cdot})$, respectively, where $\bar{y}_{i\cdot} = \sum_t y_{it}/T$, $\bar{y}_{\cdot t} = \sum_i y_{it}/N$, $\bar{y}_{\cdot\cdot} = \sum_i \sum_t y_{it}/NT$. Similar notations applies to X. The slope coefficient can then be estimated by applying least squares on the transformed data: $\widehat{\beta}_{\text{TFE}} = (X'MX)^{-1}X'My$.

In the two-way fixed effects model, the effect of regressors of interest β can be consistently

[7] A unit can be thought of as a firm or an individual.

estimated by within individual-time transformation, if the data generating process coincides with model (1). A key assumption of model (1) is that all individual units are exposed and respond homogeneously to the same time-varying shock, λ_t. This assumption may not hold in practice due to heterogeneity in individual units. In particular, when the dependent variable y_{it} is influenced by some shocks, units may respond to shocks differently. A group pattern of heterogeneity is often apparent, i.e. the individual's behavior and responses to shocks are similar within a group, even though there are differences across groups. In this case, imposing the homogeneous time fixed effects as in model (1) leads to biased estimates of the coefficients of interest, because the within transformation cannot completely eliminate the heterogeneous time fixed effects.

2.1.1 Bias in two-way fixed effects models

To illustrate the mechanism behind the bias and to quantify the size of the bias, we assume that there are G groups of units. Within the group, individual units have the same response to shocks, while the response differs across the groups. Then the data generating process can be written as:

$$y_{it} = \alpha_i^0 + \theta_{g_i^0,t}^0 + X_{it}'\beta^0 + \epsilon_{it}, \quad g_i^0 = 1, \ldots, G, \tag{3}$$

where g_i^0 is the group membership of unit i, and the superscript "0" denotes the true value of the parameter. For example, if unit i belongs to group 1, then $g_i^0 = 1$. In model (3), the outcome variable is affected by heterogeneous time-varying shocks $\theta_{g,t}^0$ for group g. Without loss of generality, we can decompose $\theta_{g,t}^0$ as:

$$\theta_{g,t}^0 = \bar{\theta}_t + (\theta_{g,t}^0 - \bar{\theta}_t) \quad \text{for} \quad g \in \{1, \ldots, G\},$$

where $\bar{\theta}_t$ can be interpreted as the common time fixed effects. A simple way to extract the common component of the time fixed effects is to use the arithmetic average, i.e. $\bar{\theta}_t = 1/N \sum_{i=1}^{N} \theta_{g_i,t}^0$.

The within individual-time transformation can only eliminate $\bar{\theta}_t$, while $\theta_{g,t}^0 - \bar{\theta}_t$ remains for each unit, which influences the coefficient estimates as omitted variables. To quantify this bias, we define $\widetilde{\theta}_{g_i^0,t} = \sum_{g=1}^{G} I(g_i^0 = g)(\theta_{g_i^0,t}^0 - \bar{\theta}_t)$, where $I(\cdot)$ is an indicator function.[8] Then the bias of the

[8] In a simple example of two groups ($G = 2$), we can express the group-wise time fixed effects as: $\theta_{1,t}^0 = \bar{\theta}_t + (\theta_{1,t}^0 - \bar{\theta}_t)/2$ and $\theta_{2,t}^0 = \bar{\theta}_t + (\theta_{2,t}^0 - \bar{\theta}_t)/2$. Then $\widetilde{\theta}_{1,t} = \theta_{1,t}^0 - \bar{\theta}_t$ for units in group 1 ($g_i^0 = 1$) and $\widetilde{\theta}_{2,t} = \theta_{2,t}^0 - \bar{\theta}_t$ for units in group 2 ($g_i^0 = 2$).

two-way fixed effects estimator can be obtained by:

$$\begin{aligned} b_{TFE} = E\left[\widehat{\beta}_{TFE} - \beta^0\right] &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}'\widetilde{X}_{it}\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}'\left(\widetilde{\epsilon}_{it}+\widetilde{\theta}_{g_i^0,t}\right)\right] \\ &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}'\widetilde{X}_{it}\right]^{-1}\underbrace{\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}'\underbrace{\widetilde{\theta}_{g_i^0,t}}_{\text{het. size}}\right]}_{\text{covariance}}, \end{aligned} \tag{4}$$

where $\widetilde{\epsilon}_{it}$ is the error term after applying transformation (2). The degree of bias depends on two components: on the covariance between the grouped time fixed effects and the regressors and on the relative size of the heterogeneity with respect to the variation of X_{it}. Particularly, the bias of coefficient estimates is high when the size of the heterogeneity is large and highly correlated with X_{it}. If the heterogeneous time fixed effects are independent from X_{it}, the slope coefficients can still be consistently estimated. However, existing empirical findings that firms have heterogeneous responses to market-wide events suggest that such bias is likely to exist. For example, Campello et al. (2010, 2011)'s findings that the 2008 financial crisis has different effects on firms due to unobservable characteristics, such as managerial liquidity and credit management qualities, implies that the heterogeneous time fixed effects are unlikely to be independent from the regressors, especially when some regressors are likely to be a function of managerial characteristics.

Table 1 provides a numerical example of how heterogeneity bias is caused. In this example, there are four individual units, and they are realizations of two population groups. Individuals 1 and 2 belong to the same group as their response to shocks is homogeneous $\theta_{1,t}^0$, while individuals 3 and 4 belong to another group, sharing a common $\theta_{2,t}^0$ that is different from $\theta_{1,t}^0$. The data is presented in columns (4), (5), and (6). If we implement the TFE transformation as in equation (2), i.e. first demean in the time series and cross-sectionally as in columns (7) and (8) and then add the total average given in column (9), the resulting data are given in column (11) of Table 1. It is clear that the transformed time fixed effect $\widetilde{\theta}_{g,t}$ is not a zero vector, suggesting that the time fixed effect is still not completely eliminated by the transformation. If we estimate the slope parameters simply using transformed $\widetilde{X}$ and $\widetilde{y}$, we fail to incorporate the effect of $\widetilde{\theta}_{g,t}$, which is non-zero and possibly correlated with $\widetilde{X}_{it}$, and thus the slope estimates suffer from heterogeneity bias.

Table 1
TFE transformation bias under GFE

The table presents a simple numerical example of how heterogeneity bias is caused by ignoring heterogeneity in responses to shocks, when using a two-way fixed effects model. Columns (1)-(3) present the basic data structure: (1) is the individual unique identification, (2) is the time period denomination, and (3) is the group identification. Columns (4)-(6) present the data: (4) is the observed value of the variable of interest, (5) is the individual fixed effect, (6) is the time effect. Column 7 presents the time-series average for each unit, column (8) presents the cross-sectional average for each time period, column (9) is the total average for the sample. Column (10) shows the demeaned cross-sectional fixed effect, $\tilde{\alpha}_i = \alpha_i - \sum_{t=1}^{T} \alpha_i - \sum_{i=1}^{N} \alpha_i + \sum_{t=1}^{T} \sum_{i=1}^{N} \alpha_i = 0$, and column (11) shows the demeaned time-series variable (11)=(5)+(6)-time series avg. (7)-cross sec. avg. (8)+total avg. (9), as in equation (2).

Unit & Time			Data						Transformed data	
Indiv. ID	Time	Group ID	Total	α_i	$\theta^0_{g,t}$	TS ave.	CS ave.	Total ave.	$\widetilde{\alpha}_i$	$\widetilde{\theta}_{g,t}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1	1	1	5	2	3	6	8	10	0	1
1	2	1	6	2	4	6	10	10	0	0
1	3	1	7	2	5	6	12	10	0	-1
2	1	1	7	4	3	8	8	10	0	1
2	2	1	8	4	4	8	10	10	0	0
2	3	1	9	4	5	8	12	10	0	-1
3	1	2	9	6	3	12	8	10	0	-1
3	2	2	12	6	6	12	10	10	0	0
3	3	2	15	6	9	12	12	10	0	1
4	1	2	11	8	3	14	8	10	0	-1
4	2	2	14	8	6	14	10	10	0	0
4	3	2	17	8	9	14	12	10	0	1

Note that if the heterogeneous time fixed effects only differ in a constant shift, then the within individual-time transformation can successfully eliminate both the individual and time fixed effects (see the Table A.1 in Appendix 2). This means that if the cross-sectional difference in a unit's response to shocks can be explained by the individual fixed effects, i.e., the cross-sectional difference of response is time-invariant, then the TFE coefficient estimates are still consistent. In practice, time-varying heterogeneity seems plausible, since individuals experience different types of shocks over time and the degree of heterogeneity in responses often depends on the features of shocks.

2.2 Interacted fixed effects model

A recent approach to control for time-varying heterogeneity is to include interactions of group and time fixed effects (see e.g. Gormley and Matsa, 2014). We follow Gormley and Matsa (2014) and refer to it as an interacted fixed effects model. This model requires the knowledge of the group structure, i.e. which group each unit belongs to. Typically, the grouping is based on observed characteristics, such as industry, firm size, etc. Given the group pattern, we can construct the (time-invariant) group dummies for each unit i as $GD_i = \{GD_{i1}, \ldots, GD_{iG}\}'$ with $GD_{ig} = 1$ if unit

i belongs to group g and zero otherwise. Let $TD_t = \{TD_{t1}, \ldots, TD_{tT}\}'$ be the (individual-invariant) time dummies for each period t with $TD_{ts} = 1$ if $t = s$ and zero otherwise. Then the interacted fixed effects model can be written as

$$y_{it} = \alpha_i + X_{it}'\beta + \sum_{g=1}^{G}\sum_{s=1}^{T}(GD_{ig} \times TD_{ts})\theta_{g,s} + \epsilon_{it}, \tag{5}$$

where $\theta_{g,s}$ is the heterogeneous time effect of group g at time s. For example, for industry-based grouping the TD is a year dummy and GD is a group dummy, then $\sum_{g=1}^{G}\sum_{s=1}^{T}(GD_{ig} \times TD_{ts})$ is an industry-time interaction dummy. If the postulated grouping is correct, we can obtain consistent estimates of homogeneous slope parameter β by applying least squares on the following within-transformation regression:

$$\dot{y}_{it} = \dot{X}_{it}'\beta + \sum_{g=1}^{G}\sum_{s=1}^{T}(GD_{ig} \times TD_{ts})\dot{\theta}_{g,s} + \dot{\epsilon}_{it},$$

where $\dot{y}_{it} = y_{it} - 1/T\sum_{t=1}^{T} y_{it}$, $\dot{X}_{it} = X_{it} - 1/T\sum_{t=1}^{T} X_{it}$, and $\dot{\epsilon}_{it} = \epsilon_{it} - 1/T\sum_{t=1}^{T} \epsilon_{it}$.

However, in practice, the group membership is typically unknown. The group structure may driven by unobserved characteristics or a mixture of observables and unobservables. Hence, the precise group dummies are difficult to know, and the estimator obtained from the "infeasible" model (5) is called an infeasible estimator. In real applications, when we misspecify the group structure, IFE leads to inconsistent slope coefficient estimates, if group-specific heterogeneity is correlated with regressors.

2.2.1 Bias in interacted fixed effects models

The bias of IFE estimates depends on how the IFE group structure is misspecified and there are infinitely many misspecifications. To derive the bias in this case, we assume that there are N_m units that are misclassified, i.e. $g_i^{IFE} \neq g_i^0$, where g_i^{IFE} is the postulated group membership of unit i based on some observables specified by the econometrician. Misclassification can be caused by, e.g. under-specification of the number of groups, incorrect choice of the grouping variables, or unavailability of the grouping variables. Misclassification can be regarded as a measurement error in the group dummies. In particular, the group dummies used in the IFE model (5) are now $GD_i = GD_i^0 + m_i$, where $GD_i^0 = \{GD_{i1}^0, \ldots, GD_{iG}^0\}'$ is the group dummy based on the true group membership g_i^0 and $m_i = \{m_{i1}, \ldots, m_{iG}\}$ is an error vector of unit i that is caused by possible deviation of the

postulated group membership from the true membership.[9] This error may often be correlated with X_{it} in practice, because the group pattern is likely to be correlated with other variables also included in the regression model. In this case, we can rewrite equation (5) with measurement errors as:

$$\begin{aligned} y_{it} &= \alpha_i + X_{it}'\beta + \sum_{g=1}^{G}\sum_{s=1}^{T}(GD_{ig} \times TD_{ts})\theta_{g,s} + \epsilon_{it}, \\ &= \alpha_i + X_{it}'\beta + \sum_{g=1}^{G}\sum_{s=1}^{T}(GD_{ig}^0 \times TD_{ts})\theta_{g,s} + \sum_{g=1}^{G}\sum_{s=1}^{T}(m_i \times TD_{ts})\theta_{g,s} + \epsilon_{it}. \end{aligned} \tag{6}$$

Obviously, the error term in equation (6) now contains two components, the idiosyncratic noise and the measurement error of the group memberships. We stack all explanatory variables and dummy interaction terms in a vector as $Z_{it} = \{X_{it}', GD_{i1}^0 TD_{t1}, \ldots, GD_{iG}^0 TD_{tT}\}'$ and let $\theta = \{\theta_{1,1}, \ldots, \theta_{1,T}, \theta_{2,1}, \ldots, \theta_{G,T}\}'$. Then we can obtain the bias of IFE estimators as:

$$E\begin{pmatrix} \widehat{\beta}_{IFE} - \beta^0 \\ \widehat{\theta}_{IFE} - \theta^0 \end{pmatrix} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T} Z_{it}' Z_{it}\right]^{-1} \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{s=1}^{S} Z_{it}'(m_i \times TD_{ts})\theta_{g,s}^0\right]. \tag{7}$$

If the difference between the postulated and the true group structure is uncorrelated with X_{it}, or the responses to shocks are zero for all groups at all times, we can still obtain unbiased and consistent slope coefficient estimates under IFE. However, neither of the two conditions are likely to be satisfied in practice. In financial studies, researchers typically classify firms based on one or two popular observables, e.g. the industry indicator and the size of a firm, whereas in many applications the true group structure is driven by a number of observable and even unobservable variables. These determinants of group structure are often closely related with covariates. Moreover, these determinants typically vary across applications, since the group pattern of heterogeneity in firms' responses to shocks depends on the features of shocks, the samples, and the variables of interest. The bias is large when the size of the heterogeneous response is large and when the degree of misspecification is large and highly correlated with explanatory variables.

2.3 Time-varying heterogeneity: Group fixed effects model

To identify the group structure, we consider a linear panel data model, where the additive time fixed effects have an unrestricted and flexible group pattern of heterogeneity, namely grouped fixed

[9] Since GD_i is a $G \times 1$ vector of dummy variables, m_i is also a vector of the same length whose elements only take the value -1, 0, or 1.

effects (GFE). The group pattern is unrestricted and flexible since it can depend on both observables and unobserables, which we do not need to specify and may vary across applications. This model class was proposed by Bonhomme and Manresa (2015a) as an effective way to avoid the issue of "incidental parameter" (Neyman and Scott, 1948), while still allowing for unobserved heterogeneity potentially correlated with regressors. The idea of "grouped fixed effects" fits the argument in the financial and accounting empirical literature that heterogeneous responses to market-wide shocks are due to unobserved time-varying heterogeneity, such as unobservable managerial characteristics and corporate culture.

2.3.1 GFE model and estimation method

The GFE model precisely coincides with the DGP in model (3) and can be written as

$$y_{it} = \alpha_i + \theta_{g_i,t} + X_{it}'\beta + \epsilon_{it}, \quad g_i = 1, \ldots, G. \tag{8}$$

This model differs from the standard fixed effects model, because it includes a time-varying group specific variable $\theta_{g_i,t}$, in addition to individual fixed effects α_i.[10] One may think of $\theta_{g_i,t}$ as interactions of group indicators and time dummies. The individual fixed effects capture the individual firm's characteristics, while the grouped-fixed effects parameter $\theta_{g_i,t}$ models how firms respond to time-varying common shocks. This response is common within a group but differs across groups. There are two types of parameters to estimate, the group membership parameter g_i for all unit $i = 1, \ldots, N$ and the standard regression parameters $\theta_{g_i,t}$, and β.

To identify the grouped number G, we need the group separation condition:

$$\text{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} (\theta_{g,t}^0 - \bar{\theta}_{g,t}^0 - \theta_{\widetilde{g},t}^0 + \bar{\theta}_{\widetilde{g},t}^0)^2 > 0, \quad \text{for} \quad g \neq \widetilde{g},$$

where $\bar{\theta}_{g,t}^0 = 1/T \sum_{t=1}^{T} \theta_{g,t}^0$ and $\theta_{g,t}^0$ denotes the true value of $\theta_{g,t}$. This implies that group heterogeneity exists after controlling for individual fixed effects, i.e. the group specific responses to shocks are time-varying. Time-varying heterogeneity is consistent with the empirical observations in the financial crises and market wide event literature.

To consistently estimate parameters in equation (8) without a priori knowledge of the group

[10] Bonhomme et al. (2017) extended this model by allowing the number of groups to increase to N, and thus incorporating individual heterogeneity. Alternatively, Bai (2009) proposes to model individual-specific response to shocks using an unobserved factor structure.

membership, one can employ the least square technique and jointly estimate the group and coefficient parameters by solving the following minimization problem:

$$Q_{NT} = \min_{\theta,\beta,g} \sum_{i=1}^{N} \sum_{t=1}^{T} (\dot{y}_{it} - \dot{X}'_{it}\beta - \theta_{g_i,t})^2. \tag{9}$$

Since the group membership parameters appear in the objective function in a nonlinear way, we cannot analytically solve this optimization problem. However, an exhaustive search of the optimal group partition is virtually infeasible due to a large number of possible partitions. To solve the optimization problem, Bonhomme and Manresa (2015a) suggest the following iterative procedure.

Algorithm 1

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

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

$$(\theta^{(s+1)}, \beta^{(s+1)}) = \arg\min_{\beta,\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} (\dot{y}_{it} - \dot{X}'_{it}\beta - \theta_{g_i^{(s)},t})^2.$$

3. Compute for all $i \in \{1, \ldots, N\}$:

$$g_i^{(s+1)} = \arg\min_{g\in\{1,\ldots,G\}} \sum_{t=1}^{T} (\dot{y}_{it} - \dot{X}'_{it}\beta^{(s+1)} - \theta_{g,t}^{(s+1)})^2.$$

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

This algorithm iterates between estimating the coefficient parameters and group membership parameters and can be viewed as an EM-type algorithm (see, for example, Dempster et al., 1977). Specifically, Step 2 estimates the coefficient parameters for a given group structure as in the usual least squares problem. Then Step 3 finds the optimal group in terms of minimum sum of squared residuals over time for each individual, based on the estimated coefficient parameters from the previous step. Put it differently, in step $s + 1$, firm i will be classified into group g if its time-series summation of squared residuals computed using the estimated coefficient parameters $\theta_{g,t}^{(s+1)}$ is less than that computed using $\theta_{g',t}^{(s+1)}$ for any $g' \neq g$. Since the objective function decreases at each iteration step, we can obtain the parameters that minimize equation (9) when the algorithm converges. Clearly, this algorithm depends on the chosen initial values $g^{(0)}$. Certain initial values may lead to

local optima of the least squares objective function. To avoid local optima, one should try a large number of initial values and select the one with the lowest sum of squared residuals.

2.3.2 Asymptotic properties of group membership and coefficient estimates

Bonhomme and Manresa (2015a) show that the group membership and coefficient estimates produced by this "grouped fixed effects" approach are consistent and have the "oracle" property. In particular, note that in step 3, the estimated g_i minimizes the time series summation of squared residuals. Therefore, one can show that the GFE estimate of group membership converges to the true population membership as the time series increases. Increases in the cross-sectional dimension (N) do not contribute to the convergence of group membership estimates asymptotically, but they may improve the membership estimates indirectly in finite samples by improving the estimation accuracy of coefficient parameters.

To precisely quantify the asymptotic properties of group membership estimates, we consider the probability of maximum degree of misclassification, i.e.

$$\text{Prob}\left(\sup_{i\in\{1,\ldots,N\}}|\widehat{g}_i - g_i^0| > 0\right),$$

where $\widehat{g}_i$ is the GFE estimate and g_i^0 is the true value of the group membership of individual i. Bonhomme and Manresa (2015a) show that the probability of maximum degree of misclassification converges to zero with the rate $\exp(T^{-\delta})$, where δ is an arbitrarily large positive number. This convergence rate is much faster than the usual convergence rate of least squares estimates $1/\sqrt{NT}$, and we often refer to it as "super-consistency". Our generated and real data based simulations in Section 3 show that the GFE group membership estimates are accurate in finite samples, even when the sample time series is moderately short.

Due to the super-consistent group membership estimates, the GFE estimates of coefficient parameters $\theta_{g,t}$ and β have the asymptotic "oracle" property that they are consistent and asymptotically equivalent to the infeasible estimate from model (5), as if we knew the true group membership. The asymptotic equivalence further implies that if a large number of time series observations is available, we can estimate standard errors using the sample analogue of the asymptotic robust variance, namely the robust variance of the least squares estimator for each group. To compute this variance,

let $\bar{\dot{x}}_{\widehat{g}_i,t}$ be the mean of $\dot{x}_{it}$ in group $\widehat{g}_i = g$, and define:

$$\Sigma_\beta = \text{plim}_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\dot{X}_{it} - \bar{\dot{X}}_{\widehat{g}_i,t})(\dot{X}_{it} - \bar{\dot{X}}_{\widehat{g}_i,t})',$$

$$\Omega_\beta = \lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[\epsilon_{it}\epsilon_{js}(\dot{X}_{it} - \bar{\dot{X}}_{\widehat{g}_i,t})(\dot{X}_{js} - \bar{\dot{X}}_{\widehat{g}_j,s})'\right].$$

Then, we can obtain the asymptotic variance of $\widehat{\beta}_{\text{gfe}}$ as:

$$var(\widehat{\beta}_{\text{gfe}}) = \frac{1}{NT} \Sigma_\beta^{-1} \Omega_\beta \Sigma_\beta^{-1}.$$

This variance is robust to heteroscedasticity and correlation across groups and time, but it does not take into account the uncertainty caused by estimating the group membership. Given that the group membership estimator converges very quickly as the number of time periods increases, the asymptotic variance can often be a good approximate of variance. However, if the panel is short with small T, one should account for errors caused by the group estimation, because misclassification may increase the small-sample dispersion of the estimator. A possible way to account for the uncertainty caused by the group estimation is to use a bootstrap standard error estimate based on re-sampling unit blocks of (y_i, X_i) from the complete sample. This is easy to implement but computationally more intense.[11]

2.3.3 Time-varying group membership

Although the group membership index g in model (8) does not explicitly depend on time, the GFE specification can implicitly model time-varying group memberships provided that the time variation of memberships and the number of individuals that change their memberships are both limited. Particularly, since classification is based on the entire time path of $\theta_{g,t}$, individuals that change their group memberships can always be segmented to form new groups, so that no membership change occurs in any group during the sample period.

For example, suppose two firms respond similarly to the first shock at time t_1 (in the same group at t_1), but these two firms respond differently to the second shock at time t_2. Then the GFE procedure will automatically segment these two firms into two groups, so that neither of these two

[11] Alternatively, one can also compute an analytical version of finite-sample standard errors using the formulas provided in Bonhomme and Manresa (2015b).

firms changes group memberships within the sample period.

2.3.4 Determining the number of groups

So far in this section, we have assumed that the number of groups G is known. However, the number of groups G is unknown a priori in practice. We first consider any issues that might arise due to the misspecification of the number of groups, and then discuss how to appropriately determine this number. When the postulated number of groups is less than the true number, the GFE estimator of slope coefficients $\widehat{\beta}_{gfe}$ is inconsistent, if the heterogeneous responses of misclassified units are correlated with regressors. However, if one over-specifies the number of groups, one can still achieve a consistent estimator of β, but probably with less efficiency (Liu et al., 2018; Bonhomme and Manresa, 2015b).

Using these properties, we can determine the number of groups by examining the values of $\widehat{\beta}_{gfe}$ for different choices of G. In particular, since $\widehat{\beta}_{gfe}$ remains consistent for an over-specified G but not for an under-specified G, we should expect that $\widehat{\beta}_{gfe}$ converges to a certain value and remains relatively stable as G increases. Thus, we can choose G, when $\widehat{\beta}_{gfe}$ is stable. To determine G, one can either use information criteria as suggested by Bonhomme and Manresa (2015a) and Su et al. (2016), i.e.

$$BIC(G) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(\dot{y}_{it} - \dot{X}_{it}'\widehat{\beta}_{gfe}^{(G)} - \widehat{\theta}_{g_i,t}^{(G)})^2 + \hat{\sigma}^2\frac{GT+K+N}{NT}\log(NT), \tag{10}$$

where $\widehat{\beta}_{gfe}^{(G)}$ and $\widehat{\theta}_{g_i,t}^{(G)}$ are the GFE estimators with G groups and $\hat{\sigma}^2$ is a consistent estimate of the variance of ϵ_{it}, e.g. the sum of squared residuals obtained from G_{max} groups.

GFE as a robust estimator

Overall, the results show that GFE is a robust method compared to TFE and IFE approaches. In particular, if the DGP is TFE, GFE can still produce consistent coefficient estimates, although it may be less efficient. This is a direct outcome of the consistency of GFE under overspecification of the number of groups. If industry is the true group structure (as specified in IFE models), GFE can consistently identify the industry group structure and produce consistent coefficient estimates. However, if a non-industry related group pattern of heterogeneity in responses exists, GFE still produces consistent estimates, but TFE and IFE coefficient estimates will be biased.

2.3.5 Specification test for group fixed effects

To evaluate which of the two specifications, two-way or group fixed effects, provides a better description of the data, we can perform specification tests. The null hypothesis is the two-way fixed effects model

$$y_{it} = \alpha_i + \lambda_t + X'_{it}\beta + \epsilon_{it},$$

and the alternative hypothesis is the group fixed effects model

$$y_{it} = \alpha_i + \theta_{g_i,t} + X'_{it}\beta + \epsilon_{it}.$$

The alternative is the encompassing general model that nest the null model by setting $\theta_{g_i,t}$ homogeneously across units at each time t.

The GFE estimator is consistent in both null and alternative models, but it is less efficient under the null, because it unnecessarily divides the sample into sub-groups and needs to estimate group memberships. The TFE estimator is consistent under the null, but not under the alternative model due to neglecting of heterogeneous time effects (see Section 2.2.1). These observations suggest that we can use a Hausman test statistic to evaluate the two methods. With a similar argument to Bai (2009), we can show that $\text{var}(\widehat{\beta}_{GFE} - \widehat{\beta}_{TFE}) = \text{var}(\widehat{\beta}_{GFE}) - \text{var}(\widehat{\beta}_{TFE})$, and thus the Hausman test statistic can be obtained by

$$H = (\widehat{\beta}_{GFE} - \widehat{\beta}_{TFE})' \left[\text{var}(\widehat{\beta}_{GFE}) - \text{var}(\widehat{\beta}_{TFE})\right]^{-1} (\widehat{\beta}_{GFE} - \widehat{\beta}_{TFE}) \sim \chi^2_K.$$

When the difference between the two estimates is statistically large, large χ^2 test statistic, it suggests that the TFE estimator is likely to be inconsistent, and one should use GFE for more reliable results. On the contrary, when the two estimates are not statistically different, it indicates that TFE yields consistent estimates.

The above discussion also applies to testing interacted versus group fixed effects, as the tradeoffs between consistency and efficiency are similar, depending on whether the interacted grouping is the correct specification. In particular, when firms respond to shocks differently according to their interacted groups, both interacted and group fixed effects estimators are consistent with the latter being less efficient since the group memberships need to be estimated. On the contrary, if interacted grouping is misspecified, the group fixed effects remain consistent, while interacted fixed effects are not. Thus, a similar Hausman test statistic can be used to compare between these two methods.

3 Simulon

In this section, we investigate the finite sample performance of grouped fixed effects estimates and compare them with two-way and interacted fixed effects estimates via Monte Carlo simulations. The purpose of this simulation study is twofold. First, we compare GFE with two widely used methods in finance applications and evaluate their relative advantages in different DGPs. We use data generated from an extensive set of simulation designs, including heterogeneous shocks that occur at different frequency, homogeneous shocks, shocks with different degrees of correlation with the regressors, and shocks of different sizes. We also consider the case when the number of groups is relatively large and the case when this number is misspecified. The second purpose of this simulation study is to examine the performance of GFE in typical finance data sets, especially the accuracy of group membership estimation. We conduct a real-data-based experiment that generates groups based on the natural disaster data set used in our empirical analysis.

3.1 Simulation with generated data

3.1.1 Data generating process

We generate the data from the following linear panel data model with time-varying and potentially heterogeneous shocks:

$$y_{it} = \alpha_i + \theta_{g_i,t} + \sum_{l=1}^{2} X_{l,it}\beta_l + \epsilon_{it}, \quad i = 1,\dots,N; \quad t = 1,\dots,T; \quad \epsilon_{it} \sim N(0,\sigma_\epsilon^2), \tag{11}$$

where $\alpha_i = 1/T\sum_{t=1}^{T} X_{1,it}$, $X_{it} = (X_{1,it}, X_{2,it})'$ is a vector of regressors with loadings $\beta = (\beta_1,\beta_2)' = (1,2)'$. The major difference between model (11) and standard panel data models lies in $\theta_{g_i,t}$ for $g = 1,\dots,G$ with $G = 5$ that captures the group-wise heterogenous response to time-varying shocks. To generate the correlation between $\theta_{g_i,t}$ and X_{it}, they are both driven by some group structure, i.e.

$$X_{l,it} = c_\tau \tau_{l,g_i,t} + \nu_{it}, \quad \nu_{it} \sim N(1,\sigma_\nu^2 = 5)$$

and

$$\theta_{g_i,t} = g_i^2(\tau_{1,g_i,t} + \tau_{2,g_i,t})/c_\theta, \quad \text{for } l = 1,2,$$

where $\tau_{l,g_i,t}$ is the group component drawn from $G(=5)$ different normal distributions with mean g and variance $2g$ for $g = 1,\dots,G$, and ν_{it} is the idiosyncratic component drawn from the normal

distribution with mean 1 and variance $\sigma_\nu^2 = 5$. One may think of $\tau_{l,g_i,t}$ as an unobserved group specific determinant of X, like corporate risk culture, which also drives the heterogeneous response to a shock. The constant c_τ reflects the importance of the group component in X and determines the degree of correlation between X and θ. The constant c_θ controls the size of shocks. To make the variation of error terms comparable to that of the regressors, we set $\sigma_\epsilon^2 = 5$.[12]

Since the nature and the characteristics of shocks are important for the consistency of each method, we consider three different types of shocks:

DGP1 (Frequent shocks): Shocks occur every time period t.

DGP2 (Sparse shocks): Shocks occur only 25% of the time.

DGP3 (Homogeneous shocks): Common shocks with homogeneous impact on all units.

DGP1 and DGP2 allow for heterogeneous responses to shocks. DGP2 contains less shocks than DGP1, and from theory we should expect the bias of TFE in DGP2 to be smaller than that in DGP1. DGP3 assumes that the effect of shocks is homogeneous across units, and therefore it is the same data generating process as the two-way fixed effect model with only individual and time fixed effects. Hence, GFE and TFE are expected to produce similar results under DGP3, see robust estimator discussion in 2.3.4.

Within each DGP, we also consider different choices of key parameters, i.e. c_τ, c_θ, and the number of groups G. These extensions help to better understand how estimation bias is affected by the correlation between regressors and heterogenous responses, the size of shocks, and the number of groups. In addition, we also consider the case of misspecifying G for GFE.

To mimic typical empirical data in corporate finance studies, we consider panels with large N and moderate T. In particular, we conduct simulations under three different sample sizes: ($N = 1000$, $T = 10$), ($N = 2000$, $T = 10$), and ($N = 1000$, $T = 20$). These sample sizes allow us to examine how the dimension of N and T affects the relative performance of various methods.

3.1.2 Method comparison and evaluation

We compare three methods: grouped fixed effects, two-way fixed effects, and interacted fixed effects. The major difference between GFE and IFE lies in that the grouping in IFE is specified exogenously

[12]Setting these variances to 5 is not essential. However, we need to ensure that the variation of the regressors and the error is of comparable size, to generate a reasonable signal to noise ratio as well as realistic shock sizes.

by researchers, while the group pattern in GFE is determined by the data. The group pattern produced by GFE is case specific and depends on the nature of the shock and outcome variable.

The validity of IFE obviously depends on the specification of groups and shocks, and we consider four scenarios for IFE. Scenario 1 correctly specifies the group memberships of all units and the time of all shocks (denoted as IFE_{gt}), which is impossible in practice. We refer to this as an infeasible estimator. Scenario 2 exactly specifies the groups but not the shocks (denoted as IFE_g). In IFE_g, When the shocks are misspecified we generate a time dummy every 1/3 of the time. Scenario 3 correctly specifies the shocks but not the groups (denoted as IFE_t). In IFE_t, when the groups are misspecified we assign 30% of units in one group to the closest neighboring groups. Scenario 4 misspecifies both groups and shocks (denoted as IFE_{null}).

We evaluate all methods based on the accuracy of grouping and the slope coefficient estimates. We measure the accuracy of grouping by the average of the misclassification frequency across replications, defined as $AMF = 1/R \sum_{r=1}^{R} MF_r$, where $R = 1000$ is the number of replications, and

$$MF_r = \frac{1}{N} \sum_{i=1}^{N} I(\widehat{g}_i \neq g_i^0).$$

We compute the bias, the estimated standard deviation, and the root mean square error (RMSE) as:

$$\text{Bias}(\widehat{\beta}) = \frac{1}{R} \sum_{r=1}^{R} \widehat{\beta}_r - \beta, \quad \text{Std}(\widehat{\beta}) = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} (\widehat{\beta}_r - \bar{\widehat{\beta}})^2}$$

and

$$\text{RMSE}(\widehat{\beta}) = \sqrt{\frac{1}{R} \sum_{r=1}^{R} \|\widehat{\beta}_r - \beta\|^2},$$

where $\widehat{\beta}_r$ is the estimator in the r-th replication and $\bar{\widehat{\beta}} = 1/R \sum_{r=1}^{R} \widehat{\beta}_r$ is the sample average across replications. We report the *empirical* standard deviation across replications, because it can capture the uncertainty caused by estimating the unknown group structure.[13,14]

[13]This standard deviation differs from the quantity $\frac{1}{R} \sum_{r=1}^{R} \widehat{\sigma}_{\widehat{\beta}}$, where $\widehat{\sigma}_{\widehat{\beta}}$ is the estimated asymptotic standard error of $\widehat{\beta}$, which does not take into account the grouping uncertainty. With a moderate time-series dimension, these two standard deviation estimates are equivalent, because the group estimation is super-consistent with the convergence rate of an arbitrarily large exponential order of T (see Theorem 2 of Bonhomme and Manresa (2015a)). However, they differ slightly from each other when T is small. Our unreported simulation results (available upon request) suggest that when T is larger than 10, the difference between these two estimates is small.

[14]Note that $\text{RMSE}(\widehat{\beta})$ is not precisely the square root of the summation of squared $\text{Bias}(\widehat{\beta})$ and $\text{Std}(\widehat{\beta})$, since $\text{Std}(\widehat{\beta})$ is the *estimated* standard deviation of $\widehat{\beta}$ and not the true one.

3.1.3 Results

In this section, we present the simulation results, which are in line with our theoretical discussion in Section 2. We first present the detailed results for DGP1 – DGP3 under a benchmark parametrization. Then, we consider alternative parametrizations in order to examine how the performance of different methods is affected by the size of the shock, the correlation between responses and regressors, and the number of groups. We also examine how GFE performs when the number of groups is misspecified.

Frequent shocks

Table 2 presents the bias and the mean square error of estimated coefficients under DGP1. We first consider the parameterization $c_\tau = 0.5$, $c_\theta = 15$, and $G = 5$ as a benchmark. This translates to moderate shocks, occurring 35% of the dependent variable time series, and the correlation between the time aggregated shocks ($\sum_{t=1}^{T} \theta_t$) and X is 0.3. We first focus on the results in Panel A of Table 2 with 1000 cross-sectional and 15 time-series observations. In this setting where units respond differently to shocks, TFE does not capture the heterogeneity and results in around 60% bias for β_1 (Bias=0.595 where β_1=1) and 30% bias for β_2 (Bias=0.599 for β_2=2). On the other hand, GFE successfully captures the heterogeneous response, and it correctly classifies more than 85% of units. The GFE coefficient estimate has a small bias (0.013).

IFE_{gt} is the best among all estimators as it correctly *specifies* (not estimates) all the groups and the shocks. IFE_{gt} is more accurate than GFE because there is no estimation of groups involved. However, it is very difficult, if not impossible, to always exactly specify all groups and shocks in practice. If there is any misspecification, IFE may perform differentely. To assess how misspecification influences the IFE estimator performance, we compare IFE_g, IFE_t, and IFE_{null}. Results in Panel A of Table 2 show that incorrectly specifying the shocks (IFE_g) leads to a bias of 0.5, and misspecification in grouping (IFE_t) causes a bias of 0.25. Naturally, misspecifying both shocks and groups (IFE_{null}) leads to an even larger bias.[15]

Comparing the standard deviation of different estimators, we find that GFE produces a slightly larger standard deviation than IFE_{gt}, suggesting that estimating the groups introduces little extra uncertainty, at least for sample sizes larger than 1,000 observations. The standard deviation of TFE

[15]The impact of misspecification obviously depends on the degree of misspecification. Unreported results (available upon demand from the authors) confirm that if we increase the proportion of individuals that are misclassified or decrease the number of specified shocks, the bias of IFE increases.

is almost twice as large as GFE, and that of misspecified IFE is also much larger than GFE.

The RMSE of GFE is only slightly higher than that of the infeasible estimator IFE_{gt}. However, the RMSE of TFE is almost 50 times larger than that of GFE. Incorrectly specifying the groups and shocks also dramatically decreases the performance of IFE (with RMSE 45 times larger than that of GFE), and the performance depends on the degree of misspecification.

Panels B and C of Table 2 examine how the sample size affects the performance of different methods. Increasing the sample size improves the accuracy of GFE and IFE_{gt}, but not the accuracy of TFE and other versions of IFE. For GFE, increasing the time dimension remarkably reduces the misclassification frequency and the bias, but increasing the individual dimension does not improve the classification, because only the time series variation contributes to the accurate classification of units. In fact, the convergence rate of group estimation is $1/T^{\delta}$, where δ is any arbitrary positive number, and N is not involved in the convergence rate. The standard deviation of GFE decreases by increasing N and T, confirming the convergence rate of GFE slope-coefficient estimators to be $1/\sqrt{NT}$. The accuracy of IFE_{gt} also improves by increasing any dimension of the sample, since it is asymptotically equivalent to GFE. On the contrary, the RMSEs of TFE and other versions of IFE do not decrease by increasing the sample size. The improvement in GFE and IFE_{gt} performance confirms the consistency of these two estimators.

Sparse shocks

Next, we study the case where shocks occur infrequently, as presented in Table 3. In this DGP, we add another IFE estimator, which correctly specifies groups but over-specifies the shocks by interacting groups with time dummies of each time period, denoted as IFE_{got}. Since the shocks remain the same size but appear less frequently, the aggregate heterogeneous response of the dependent variable to the shocks is of a smaller magnitude. Hence, although the estimators that omit these heterogeneous responses, i.e. TFE, IFE_{g}, IFE_{t}, and IFE_{null}, continue to be biased, the biases are all smaller than in the case of frequent shocks.

GFE continues to perform well in grouping with misclassification frequency less than 10% in all cases and its coefficient estimates remain highly accurate, although the standard deviations are slightly larger than in DGP1. This is because GFE unnecessarily models the shock at each time period while there are actually fewer shocks. This also explains the difference between GFE and IFE_{gt}, which exactly identifies both grouping and shocks. Similarly, we find that imposing more shocks (IFE_{got}) does not affect the consistency but only weakens the efficiency of IFE. Given

sufficient samples in our case, the efficiency loss in IFE_{got} is ignorable (compared to IFE_{gt}).[16]

Homogeneous shocks

The previous results have shown that when individuals respond differently to shocks, ignoring such heterogeneity can lead to severely misleading results for TFE and IFE estimation with misspecified groups. A natural question is, if the data generating process is of homogeneous shocks, i.e. the impacts of shocks are homogeneous across individuals, then how much inefficiency will be caused by assuming a relatively complicated model of grouped fixed effect?

To answer this question, we generate the data by model (11) but set $G = 1$. Table 4 compares various estimators in the homogeneous-shock case. We see that TFE and IFE_{gt} now perform equivalently well, because both models coincide with the DGP. As for GFE, we consider two cases with estimated number of groups $\hat{G} = 1$ and $\hat{G} = 4$.

When $\hat{G} = 1$, GFE boils down to TFE and IFE_{gt}, and it performs identically to the other two methods. When $\hat{G} = 4$, we over-specify the number of groups, which only marginally increases the RMSE. This confirms the theoretical result of Bonhomme and Manresa (2015a) that over-specifying the number of groups does not affect the consistency of the slope coefficient estimators.[17] In such large sample sizes, the potential efficiency loss is ignorable. We also observe that when the sample size increases, the RMSE decreases as expected.

Larger and more correlated shocks

Next we change the parametrization of the data to investigate how the size of the shocks and their correlation with regressors affect the performance of different methods. We first consider larger shocks ($c_\theta = 5$ and other parameters remaining the same) that is roughly 60% of the magnitude of the dependent variable on average. The results are in Panel A of Table 5 show that the biases of TFE and misspecified IFE (IFE_{g}, IFE_{t}, and IFE_{null}) all increase to a large extent, as expected. However, GFE is not affected by the size of shocks, and it remains as accurate as the infeasible estimator IFE_{gt}. The classification of GFE is even more accurate with less than 1% misclassification frequency, because the group separation is stronger.

Panel B of Table 5 examines the case when the correlation between shocks and regressors is high. To mimic this situation, we set $c_\tau = 1$ while other parameters keep the same, and this generates

[16] This result suggests that one could simply include time dummies for all periods, and not necessarily specify the time of shocks precisely.

[17] However, under-specifying G may lead to inconsistent estimates of group fixed effects.

a correlation of 0.5 and the size of shocks remains 35% of the dependent variable on average. We find that high correlation does not affect the grouping accuracy of GFE compared to the case of frequent shocks (Panel B of Table 2), but it increases the bias of all estimators including GFE. GFE remains the most accurate among all feasible estimators, even though its RMSE is twice as large as the infeasible estimator IFE_{gt}. The bias of GFE increases because the effect of misclassification increases with higher correlation. The RMSEs produced by TFE and IFE_{g} are both 100 times larger than those of IFE_{gt}.

Incorrect $\widehat{\mathsf{G}}$

From the experiment of DGP3, we have seen that overestimating the number of groups does not affect the consistency of GFE slope coefficient estimates. However, underestimation of G does affect consistency. This is because if the number of groups is under-specified, then some groups will contain heterogeneous individuals. Since individuals within a group are treated homogeneously, this leads to ignoring heterogeneity and a biased GFE estimator.

Table 6 presents the case where the true number of groups $\mathsf{G} = 4$ is incorrectly specified as $\widehat{\mathsf{G}} = 2$ in the GFE estimation. It shows that GFE is biased when G is under-specified, and this bias does not decrease as the sample size increases. However, we also find that the degree of bias and the RMSE of GFE are still much smaller than those of TFE. The direction of the bias of GFE estimates depends on the correlation between regressors and heterogenous responses. In this case, we do not report the misclassification frequency because this measure is difficult to define when the specified number of groups does not match the true number of groups.

Based on these results, a recommendation for best practice is to choose a larger G, if the number of groups is uncertain and the slope coefficients are the main variable of interest. In general, the cost of over-specification seems much smaller than under-specification and also than IFE, especially for large sample sizes.

More groups

Here we consider the situation where the data are generated by a large number of groups, i.e. $\mathsf{G} = 10$ and $\mathsf{G} = 20$, respectively. Enlarging the number of groups results in an even larger degree of heterogeneity. Therefore, we should expect the bias of TFE and misspecified IFE to increase,

while GFE is expected to remain consistent.[18] Table 7 shows the performance under $G = 10$ and $G = 20$. The RMSEs of TFE and misspecified IFE increase at least 7 times compared to those in DGP1, while the accuracy of GFE is unaffected.

3.2 Accuracy of the estimated group membership

The previous section has shown clear advantages of GFE over TFE and IFE models, when the group structure is not correctly specified using simulated data. In practice, we would also like to know how accurately the group memberships of individual units are estimated by GFE. In this section, we first evaluate the accuracy of the group estimator through an empirical application and supplement the first exercise with a simulation based on parameters that closely match the empirical data.

The design of a perfect setting for an empirical application to evaluate the group estimator is challenging, as it is nearly impossible to know ex-ante the correct endogenous groupings that are driven by both observable and unobservable determinants. As a result, we can never fully verify the accuracy of the grouping because of the unobservable determinants. However, if the grouping is predominantly driven by observable determinants and we know these determinants ex-ante, we can evaluate the GFE group estimator based on observables.

We study the performance of the GFE group estimator using the empirical application of the effect of the severity of major natural disasters on sales growth. The use of natural disasters as an observable determinant of firm grouping is ideal, as we can compare the GFE estimated group membership to the firms affected by the natural disasters. Moreover, the timing and geographical severity of the disasters is exogenous but correlated to the explanatory variables. The shortcoming of the exercise is that we ignore the role of unobservable characteristics in grouping. As a result, the GFE estimated grouping, which is based on observables and unobservables, will be positively but imperfectly related to groups based on only the location of natural disasters.

3.2.1 Sales growth and natural disasters

Section 2.3 discusses the estimation of the latent group membership and the "oracle" property of the GFE estimator, as both dimensions of the panel data increase. In this application, we study the determinants of firms' sales growth in the presence of major natural disasters. Barrot and Sauvagnat (2016) find that firms' sales growth decreases by 3.3 to 4.5 percent for three consecutive

[18] Bonhomme and Manresa (2015b) propose an alternative algorithm (variable neighborhood search) that potentially works more efficiently than the iterative algorithm for high-dimensional data.

quarters following a natural disaster. From Barrot and Sauvagnat (2016), we expect $\theta_{g_i,t}$ in the sales growth regression to be negative and to pick up some of these sizable geographical effects for the disaster-affected firms in each year, even in the presence of other unobservable determinants of heterogeneity (firm and managerial characteristics). This application provides some degree of validity check on the group membership estimator and focuses on how the estimated time effect corresponds to the severity of the impact on the work force in the affected firms.

We regress log sales growth on several firms characteristics and different fixed effects. We estimate the following regression:

$$\ln(\text{Sales growth})_{i,t} = \alpha_i + \theta_{g_{i,t}} + X'_{i,t-1}\beta + \epsilon_{i,t},\ g_i \in \{1, ..., G\} \tag{12}$$

where $\ln(\text{Sales growth})_{i,t}$ is the natural log of sales growth for firm i in year t. $X_{i,t-1}$ is a 6×1 vector of lagged variables that affect sales growth. We include the following commonly used explanatory variables in our regressions: natural log total assets (TA), Tobin's q (q), return on assets (ROA), capital expenditure (Capex), leverage (Leverage), and free cash flow (FCF). α_i is the firm fixed effect and $\theta_{g_{i,t}}$ is a time-varying group effect. For TFE, $\theta_{g_{i,t}} = \theta_t$ and for IFE, $\theta_{g_{i,t}}$ is industry interacted with year fixed effects. We use Compustat data for the firm characteristics and the natural disaster information from Barrot and Sauvagnat (2016). The sample period is 1987 to 2010.

We expect the estimates of θ_t from TFE, $\theta_{\text{industry}_i,t}$ from IFE, and $\theta_{g_i,t}$ from GFE to be negative for affected firms on the year of the disasters, as firms sales growth decrease due to disasters. Furthermore, we also expect the cross-sectional heterogeneity be negatively correlated with the share of total U.S. employment affected by the disasters, according to Barrot and Sauvagnat (2016). In other words, firms based in disaster states which have a higher percentage of employment affected by disasters should experience a larger drop in sales.

For exposition purposes, we first focus on 2004 as a year of very large natural disasters. Figure 1 depicts the estimates of the state aggregated effects for 2004. We average the estimates of θ across firms according to their headquarter states, as in Barrot and Sauvagnat (2016). Panel A of Figure 1 depicts the share of total U.S. employment affected by four major hurricanes (Charles, Fracis, Ivan and Jeanne), as computed from the County Business Pattern data, provided by the U.S. Census Bureau. Panel B of Figure 1 shows the GFE estimates, where $\theta_{g_i,t}$ is negative and is negatively correlated with states with larger disaster-affected work force. IFE estimates $\theta_{\text{industry}_i,t}$ in Panel C are also negative, but do not coincide with the dispersion of workforce affected in Panel A. Panel

D shows that while the TFE estimates are negative, they are uncorrelated with the share of total U.S. employment affected by disasters at the state level.

Using the sample from 1987-2010, Panel A of Table 8 reports the univariate regression estimates of TFE, IFE, and GFE on disaster-affected employment across all natural disasters for the sample period. Consistent with the above findings, disaster-affected employment is negatively related to GFE $\theta_{g_{i,t}}$. Based on the average drop in employment of 22 percent in major disaster-affected states, see Barrot and Sauvagnat (2016), the estimated drop of three to five percent in sales based on the GFE estimate is close to Barrot and Sauvagnat's reported estimate. On the other hand, TFE and IFE estimates suggest a positive but economically small increase in sales with an increase in affected employment. These results demonstrate the reliability of the GFE group estimator.

Panel B of Table 8 presents the coefficient estimates of the determinants of sales growth using two-way fixed effect, interacted fixed effect based on industry grouping, and grouped fixed effect based on $G = 10, 15, 30$ models.[19] In general, there are substantial differences between GFE, TFE and IFE estimates, confirmed by the Hausman test statistics. The test statistics suggest that TFE and IFE estimates are incorrect. Particularly, the coefficient estimates for total assets, capex, and cash flow are significant for all GFE specifications, but are larger for TFE and insignificant for IFE. The leverage effect is insignificant for TFE and negative for GFE. We also see economically significant differences across estimates for Tobin's Q. Overall, the results in this section show tht GFE accurately classifies firms into groups when a group structure exists. If the group structure is ignored, coefficient estimates are biased and inference incorrect. Later in the paper, we further demonstrate the economic importance of using GFE and how economic and statistical inferences change by replicating a recently published paper.

3.2.2 Simulation based on empirical parameters

The shortcoming of the above exercise is that we are never sure about the true grouping in the data, due to the unobserved determinants of group membership. To more critically evaluate the GFE group estimator, we perform Monte Carlo simulation based on empirical parameters from the observed data. For simulation convenience, we drop firms if any of the six explanatory variables

[19] To implement the GFE estimation, we first need to determine the number of groups G. We consider two evaluation criteria, the Bayesian Information Criterion (BIC) given by equation (10) and the convergence of slope coefficient estimates. The BIC serves as an effective criterion to tradeoff between the model fitness and the number of parameters (degree of complication of the model), and we select the number of groups that minimizes this criterion. Figure 2(a) depicts how the BIC varies for different numbers of groups in the sales growth regression. The value of BIC decreases dramatically when we increase the number of groups from 1 to 10, and reaches its minimum when G = 10.

are missing during the sample period 1987–2001 to construct a balanced panel.[20] This leads to a sample containing 207 firms with 15 years of observations. Data is from Compustat.

For a given number of groups, we generate the dependent variable by

$$y_{it} = \theta_{g_i,t} + \sum_{l=1}^{6} X_{l,it}\beta_l + \epsilon_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T; \quad \epsilon_{it} \ N(0, \sigma_\epsilon^2) \tag{13}$$

where X_{it} contains the six explanatory variables in regression 12. To specify the slope coefficients and group fixed effects, we first estimate model (13) on the empirical dataset using the GFE approach. Then, we fix the GFE estimates of β_l and $\theta_{g_i,t}$ as their true values, respectively. The true group memberships g_i for all i are also fixed to their corresponding GFE estimated memberships. Finally, we generate the error terms from i.i.d. normal distributions with zero mean and variance equal to the variance of the GFE residuals.

We evaluate the performance of GFE based on the accuracy of grouping and coefficient estimates. The accuracy of grouping is measured by the average misclassification frequency. For the coefficient estimates, we employ relative measures to facilitate comparison, slightly different from the absolute measures used in Section 3.1, since the magnitude of the true coefficients varies significantly across variables. In particular, we compute the absolute value of the relative bias (ARB) as

$$\text{ARB} = \left| \frac{\text{Bias}(\hat{\beta}_l)}{\beta_l^0} \right|,$$

where $\hat{\beta}_l$ and β_l^0 are the estimated and true value of β_l, respectively.

Table 9 presents the misclassification frequency, ARB, and overall RMSE. We find that when $T = 15$ and $G = 5$, GFE can correctly classify more than 90% of the units. The relative bias of GFE is much smaller than TFE and IFE for all variables, and the RMSE of GFE is almost half of that produced by TFE and IFE. Increasing the number of groups marginally affects the performance of GFE as expected, but the bias and RMSE of TFE and IFE increase sharply.

In summary, we demonstrate the effectiveness of the GFE group membership estimator through an empirical application where we recover the grouping based on ex-ante known and observable determinants. We further show the reliability of the group estimator through a simulation based on parameters drawn from the empirical distributions of the variables of interests.

[20] We use a shorter sample for this simulation to maximize the number of firms (cross-section) of the balanced sample and to have similar T to the baseline simulations in Section 3.1.2.

4 Economic importance of GFE

Section (3.2.1) examined the reliability of the group membership estimation and showed that economic inference can vary using different econometric approaches. In this section, we demonstrate the importance of GFE for economic and statistical inference through revisiting the study of pilot CEO's influence on corporate innovation, Sunder et al. (2017).

Sunder et al. (2017) argues that "CEOs who combine risk tolerance with a desire for new experiences achieve greater innovation success". The paper collects information about CEOs' pilot credentials using records from the Federal Aviation Administration (FAA) and studies the differences in innovation outcomes across firms with pilot and non-pilot CEOs. The paper finds that CEOs' hobby of flying airplanes is associated with significantly better innovation outcomes, measured by patents and citations. The paper argues that "sensation seeking combines risk taking with a desire to pursue novel experiences and has been associated with creativity." The econometric model in the paper controls for year and industry fixed effects, CEO and firm characteristics, including age, human capital proxied by CEO tenure and academic achievement, explicit risk-taking incentives, military experience, and overconfidence. Sunder et al. (2017) finds that firms led by CEOs with a hobby of flying small aircraft generates more patents and patents with greater impact. However, Sunder et al. (2017) cautions about the interpretation of their findings and acknowledges the important role of time-varying corporate risk preferences and the issue of endogenous matching between CEOs and firms.

For example, a group of firms with preferences for risk will select corporate leaders with similar risk preferences. As a result, studying the association between the risk taking behavior of a firm and an observable variable that measures manager's taste for risk might lead to misguided economic inference without accounting for firms' time-varying unobserved heterogeneity of such groups, i.e. their changing strategy on risk taking. In this section, we revisit and replicate the study of pilot CEO and the success of innovation investment by estimating various econometric models using TFE, IFE, and GFE and compare the resulting estimates.[21]

[21]For brevity, we will focus on the results for patent citation, see Column 6 of Table 4 in Sunder et al. (2017). We also study other outcome variables in their paper and reach similar conclusions (results are available from the authors upon demand).

4.1 Pilot CEO and corporate innovation

We model corporate innovation success as follows:

$$\ln(1+\text{citation})_{it} = \alpha_{ind} + \theta_{g_{i,t}} + X'_{i,t}\beta + \epsilon_{i,t},\ g_i \in \{1, ..., G\} \tag{14}$$

where $\ln(1+\text{citation})_{it}$ is the natural log of the number of raw citations multiplied by the weighting index of Hall et al. (2001) to all the patents applied for by firm i in year t. α_{ind} is the industry fixed effect and $\theta_{g_{i,t}}$ is a time-varying group effect. Following Sunder et al. (2017), X_{it} is a 17×1 vector of variables affecting patent citations: pilot CEO dummy, log total assets (TA), ratio of net property, plant, and equipment over the number of employees (PPE/EMP), stock returns, Tobin's Q, institutional holdings, CEO tenure, CEO stock and option delta and vega, CEO age, top university dummy, PhD, dummy for no school information, finance education, technical education, CEO military background dummy, CEO overconfidence. Table A.2 in the appendix provides all variable definitions. The sample period is 1993 to 2003.

The TFE estimator includes industry fixed effects and year dummies as in Sunder et al. (2017), $\theta_{g_{i,t}} = \theta_t$. IFE includes an industry fixed effect and $\theta_{g_{i,t}}$ is the industry FE interacted with year fixed effects. The data is from ExecuComp, BoardEx, Compustat, NBER patent database, and the U.S. Federal Aviation Administration airmen certification records. We exclude financial firms and regulated utilities and apply similar filters as in Sunder, Sunder and Zhang (2017).

To implement the GFE estimation, we first need to determine the number of groups G. We consider two evaluation criteria, the Bayesian Information Criterion (BIC) given by (10) and the convergence of slope coefficient estimates. The BIC serves as an effective criterion to tradeoff between the model fitness and the number of parameters (degree of complication of the model), and we select the number of groups that minimizes this criterion. Figure 2(b) depicts how the BIC varies for different numbers of groups in regression (14). The value of BIC decreases dramatically when we increase the number of groups from 1 to 10, and reaches its minimum when $G = 10$. From $G = 10$ onwards, increasing the number of groups does not significantly improve the model fitness, while introducing a much larger set of parameters to estimate. For robustness, we also examine the behaviour of slope coefficient estimates for G=15 and find little differences from G=10.

Table 10 presents the coefficient estimates of the determinants of patent citations obtained from two-way fixed effect, interacted fixed effect based on industry grouping, and grouped fixed effect regressions. Hausman tests strongly reject the null hypothesis that TFE, IFE and GFE produce

the same coefficient estimates. In general, we observe substantial differences in the estimated coefficients, which leads to different economic inferences between GFE and other estimators.

Consistent with Sunder et al. (2017), we find that pilot CEO is positively correlated with firms' total patent citations, when we account for industry and year fixed effect (TFE).[22] In particular, firms with pilot CEOs have an estimated 0.46 ($\exp^{0.38} - 1$) more citations that firms with non-pilot CEOs. We also find qualitative similar statistical relations between the covariates and patent citations. These relations remain similar and the pilot CEO remains positive and significant when we control for interacted fixed effects with industry-year interaction fixed effect.

Column 3 of Table 10 reports the estimated coefficients from GFE. Contrary to the TFE and IFE results, we find that there is statistically **no difference** in citations per year for firms with pilot CEOs compare to firms with non-pilot CEOs. This is consistent with the idea that a firm seeking to change its corporate culture or risk taking strategy will select a CEO with similar risk preferences. The inclusion of θ_{gt}, which might potentially capture the endogenous grouping of this changing culture, allows the model to compare pilot versus non-pilot CEOs across firms with similar time-varying risk culture or management strategy.

GFE's estimates for other CEO characteristics and compensation variables are mainly consistent with the literature's findings. We find that larger firms with higher capital intensity, higher Tobin's Q and lower institutional holdings have more cited patents. We find a positive relation between innovation and delta indicating that a strong CEO incentive alignment with shareholders generates better innovation outcomes. We also find that CEOs at the beginning of their tenure engage with riskier and more successful innovations. Education-related coefficients also suggest that education especially those related to finance and technical degrees play an important part in firms' innovation outcomes. We find that firms with military CEOs have slightly lower patent citations consistent with the literature's finding that military CEOs allocate less resources to research and development, see Benmelech and Frydman (2015). Contrary to Sunder et al. (2017) and Hirshleifer et al. (2012), GFE estimates report a positive but insignificant relation between overconfident CEO and firm's patent citations. Given that several papers report that overconfident CEOs react to regulatory shocks, such as SOX, differently from other CEOs, this heterogeneity might be captured by the GFE endogenous grouping of firms resulting in the finding that overconfident CEOs play a smaller role in innovation.[23]

[22] Our point estimate of 0.38 for TFE is slightly different from the Sunder et al. (2017) reported estimate of 0.61, see Table 4 Column 6.

[23] Several recent papers find that SOX has a different impact across firms with different managerial opportunistic

Overall, we demonstrate how the failure of TFE and IFE to appropriately account for time-varying group unobserved heterogeneity, possibly related to business strategy and corporate risk culture, can lead to incorrect economic inference and biased estimates, through a replication of a recently published work of the effects of managerial attributes on firm outcomes.

5 Endogeneity and standard errors

5.1 Endogeneity

Least square estimation (9) leads to inconsistent estimates of slope coefficients in the presence of endogenous explanatory variables, and instrument variable estimation is thus required. The GFE approach can be implemented with instrument variable estimation, e.g. two-stage least squares (TSLS) or generalized methods of moments (GMM).

Consider the structural equation of interest

$$y_{it} = \theta_{g_i,t} + X_{it}\beta + \epsilon_{it}, \tag{15}$$

where ϵ_{it} may be correlated with X_{it}, and thus X_{it} is endogeneous. The reduced form is a model of the conditional expectation of X_{it} given a $P \times 1$ vector of variables Z_{it}:

$$X_{it} = f(Z_{it}, \pi) + u_{it}, \tag{16}$$

where γ is an $M \times 1$ vector of parameters, $f(\cdot,\cdot)$ maps $\mathcal{R}^P \times \mathcal{R}^M$ to $\mathcal{R}^K$, and u_{it} is $K \times 1$ satisfying $E(u_{it}|Z_{it}) = 0$. This suggests that Z_{it} can serve as valid instruments for $K \times 1$ endogenous variables X_{it} with $R \geq K$. The reduced form (16) can be applied to widely used econometric models, such as linear or threshold regressions, etc. In the following analysis, we use linear pooled regressions as an example, and in this case we have:

$$X_{it} = \alpha + Z_{it}'\gamma + u_{it}. \tag{17}$$

Note that we do not allow for endogeneity in the group membership determination. Neither do we allow for a group pattern of heterogeneity in the reduced form model (16). Moreover, we require

behavior on firm disclosure, financial constraints, pre-cautionary savings, information costs, financial reporting, and managerial compensation practices (see e.g. Chhaochharia and Grinstein, 2007; Heron and Lie, 2007; Koh et al., 2008; Cohen et al., 2008; Carter et al., 2009; Duchin et al., 2010a,b; Brochet, 2010; Banerjee et al., 2015).

that the standard rank condition is satisfied to identify the structure parameters.

For this system, we can estimate the group membership, group fixed effects, and slope coefficients jointly by minimizing the following objective function as

$$(\tilde{\theta}, \tilde{\beta}, \tilde{g}) = \arg\min_{\theta,\beta,g} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \hat{X}_{it}'\beta - \theta_{g_i,t}\right)^2, \tag{18}$$

where $\hat{X}_{it}$ is the predicted value of X_{it} based on the first-stage regression model (17). To solve this optimization problem, one can again employ the iterative algorithm that iteratively estimates the group membership parameter g and slope and intercept parameters. In particular, we can modify Algorithm 1 as follow:

Algorithm 2

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

2. Compute: $\hat{X}_{it} = \hat{\alpha} + Z_{it}'\hat{\gamma}$, where $\hat{\alpha}$ and $\hat{\gamma}$ is the least square estimates from (16).

3. For the given $g^{(s)}$ and $\hat{X}_{it}$, compute:

$$(\theta^{(s+1)}, \beta^{(s+1)}) = \arg\min_{\beta,\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \hat{X}_{it}'\beta - \theta_{g_i,t}\right)^2,$$

   where $\hat{X}_{it}$ is the predicted value of X_{it} from the first stage regression of X_{it} on Z_{it}.

4. Compute for all $i \in \{1, \ldots, N\}$:

$$g_i^{(s+1)} = \arg\min_{g\in\{1,\ldots,G\}} \sum_{t=1}^{T} \left(y_{it} - \hat{X}_{it}'\beta^{(s+1)} - \theta_{g_i,t}^{(s+1)}\right)^2.$$

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

We show the effectiveness of this procedure through simulations. We generate the data from the linear panel data model with time-varying heterogeneous shocks as in our benchmark simulation design (11), except that now $X_{1,it}$ is endogenous. We generate $X_{1,it} = Z_{it} + \varepsilon_{it}/5$, where Z_{it} is independent of ε_{it} and generated by $Z_{it} = c_\tau \tau_{1,g_i,t} + \nu_{it}$ with the same parameterization as in Section 3. The remaining parameters are also set in the same way as in Section 3 and we focus on the cases of frequent but moderate shocks as in DGP1. Coefficients are estimated using TSLS.

We evaluate the performance of GFE estimators via the misclassification frequency, bias, and RMSE, and compare with the two-way fixed effect estimator. Table 11 shows that the misclassification frequency is slightly higher than in the cases of exogenous regressors (see Table 2). This could be due to an efficiency loss caused by the two-stage estimation. Nevertheless, GFE still correctly classifies more than 80% of units even in the smallest sample with $T = 15$. The misclustering rate decreases as T increases, confirming the consistency of group membership estimates, but it does not decrease as N increases, as expected. The accuracy of GFE coefficient estimates improves as either N or T increases. On the contrary, the estimates produced by TFE are severely biased. Increasing the sample size do not seem to help to reduce the bias and RMSE of the TFE estimates.

In principle, we can replace the two-stage least squares estimation by (optimal) GMM. In that case, the estimation of the optimal weight matrix requires a careful treatment since it depends on the group membership estimates.

5.2 Standard errors

Petersen (2009) highlights the importance of using correct methods to estimate standard errors. Here we discuss the standard error estimation in the presence of group fixed effects. Our discussion is twofold. First, we investigate the properties of standard error estimates produced by GFE. Next, we discuss how other standard error estimates perform, if the data generating process is characterized by group-specific heterogeneous responses to shocks.

There are two versions of standard errors for the GFE method, an asymptotic standard error and a finite sample standard error. The asymptotic standard error estimate is identical to White's (1984) robust standard error of OLS estimators under true group membership (see also Corollary 1 of Bonhomme and Manresa (2015a)). This estimate is often a good proxy because the group membership is estimated super-consistently, and the estimation errors caused by unknown grouping can be ignored. When the time dimension is particularly short, the finite sample standard error estimates need to employed to account for the group estimation error.[24] Since most financial data have moderate time series, our discussion here focus on asymptotic standard error estimates of GFE.

The GFE asymptotic standard error is unbiased in the presence of fixed firm and time effects, because α_i and $\theta_{g,t}$ control the firm and time effects, leaving an idiosyncratic error term. The GFE asymptotic standard error is also unbiased if the time effect is group-specific heterogeneous, i.e.,

[24]There are two ways to compute the finite sample standard error estimates. One can use the analytical formula proposed in Section 2.2.2 of Bonhomme and Manresa (2015a) or bootstrap the standard error estimate that is based by resampling unit specific blocks of observations from the original sample.

if the firm effect changes over time but in a group pattern. Such time-varying group effect is well captured by $\theta_{g,t}$, and thus again leaves only the idiosyncratic component in the error term.

Next, we discuss the properties of other standard error estimates in the presence of group-specific heterogeneous responses. The standard errors of OLS, FE, and TFE are all biased to different extents. The bias of FE standard errors has been demonstrated by equation (22) in Section A.1. Given the bias of FE standard errors, we can expect that the OLS standard errors are even more biased, because OLS fails to take into account both firm and group fixed effects. We show the bias of TFE standard errors using the case with only one explanatory variable. The within individual-time transformed variables are denoted by $\widetilde{X}_{it}$ and $\widetilde{\theta}_{g_i,t}$. Then the variance of the TFE estimator is

$$\begin{aligned}
var(\hat{\beta}_{TFE}) &= \left[\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}(\widetilde{\theta}_{g_i,t}+\widetilde{\epsilon}_{it})\right)^2\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}^2\right)^{-2}\right] \\
&= \left(\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}^2(\widetilde{\theta}_{g_i,t}+\widetilde{\epsilon}_{it})^2+\sum_{t=1}^{T}\sum_{i\neq j}\widetilde{X}_{it}\widetilde{X}_{jt}(\widetilde{\theta}_{g_i,t}+\widetilde{\epsilon}_{it})(\widetilde{\theta}_{g_j,t}+\widetilde{\epsilon}_{jt})\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{X}_{it}^2\right)^{-2} \\
&= \left[\sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}^2\widetilde{\epsilon}_{it}^2+C\right]\left(NT\sigma_X^2\right)^{-2} \\
&= \frac{\sigma_\epsilon^2}{NT\sigma_X^2}\left[1+NT\sigma_X^2 C\right], \qquad (19)
\end{aligned}$$

where $C = \sum_{t=1}^{T}\sum_{i=1}^{N}\widetilde{X}_{it}^2\widetilde{\theta}_{g_i,t}^2 + 2\widetilde{X}_{it}\widetilde{\theta}_{g_i,t}\widetilde{\epsilon}_{it} + \sum_{t=1}^{T}\sum_{i\neq j}\widetilde{X}_{it}\widetilde{X}_{jt}(\widetilde{\theta}_{g_i,t}+\widetilde{\epsilon}_{it})(\widetilde{\theta}_{g_j,t}+\widetilde{\epsilon}_{jt})$, which is nonzero as long as X_{it} and $\theta_{g_i,t}$ are correlated.

Table 12 presents the standard errors structure in the presence of group fixed effects. This structure resembles that of two-way fixed effects except that the firms in distinct groups are not correlated even at the same time point, because the time shocks impose heterogeneous effects across groups (assuming groups are independent). In the presence of group fixed effects, standard errors only clustered by firm are clearly biased because it imposes that residuals between firms are completely uncorrelated. In contrast, standard errors clustered by both firm and time are unbiased, but can be inefficient because they allow all firms (across groups) to be correlated at the same time.

6 Conclusions

Firms are often exposed to market-wide shocks such as regulatory changes and financial crises. Failing to incorporate these market-wide shocks can lead to inconsistent estimates of treatment effects

and can affect statistical inference. The most widely used approach to model market-wide shocks are time fixed effects models, which assume that all firms or individuals respond homogeneously to these common shocks. However, the homogeneity assumption is unlikely to be correct, because it is well-documented that regulatory changes like the Sarbanes-Oxley Act and financial crises have heterogenous impact on firms outcomes.

In this paper, we demonstrate how ignoring this heterogeneity results in biased slope coefficient estimates, when heterogeneous shocks are correlated with explanatory variables. More specifically, we formally quantify the bias of the estimates of existing methods in using OLS, two-way fixed effects and interacted industry-year fixed effects models.

We propose the use of the "grouped fixed effects" estimator to account for heterogenous responses to market-wide shocks. The group fixed effect estimators assume that there is an underlying latent group pattern for the population of firms. Within the group, firms respond similarly to each shock, while they respond differently across groups. One important advantage of this approach is that the group structure can be consistently estimated from the data, jointly with the regression parameters. Therefore, no a priori knowledge of grouping is required. This avoids possible estimate inconsistency caused by the misspecification of the group structure.

We theoretically compare the grouped fixed effect (GFE) estimators with popular alternatives, such as two-way fixed effects (TFE), interacted fixed effects (IFE) with artificially imposed group structure (e.g. industry-based grouping), and event studies. We show that two-way fixed effect estimates are inconsistent if firms respond differently to the shocks. On the contrary, grouped fixed effects always produce consistent slope coefficient estimates, when the number of groups is not under-specified. GFE is asymptotically equivalent to the interacted fixed effects model, if the imposed group structure in IFE happens to coincide with the data generating process. However, if the group structure of IFE is misspecified, IFE estimates are no longer consistent. We provide guidelines on how to appropriately correct for standard errors based on the group structure.

We conduct a large number of Monte Carlo simulation experiments, controlling for the size and the frequency of shocks, the degree of heterogeneity among units, the correlation between responses and regressors, the number of groups, etc. Simulation results confirm our theoretical argument that conventional methods are biased to different extents, depending on the feature of shocks, the correlation, and specification of groups, etc. On the contrary, GFE remains a consistent approach even when all firms respond homogeneously.

We demonstrate the economic importance of accounting for heterogeneity through an empirical

model of CEO attributes and corporate innovation. We find that accounting for time-varying heterogeneity plays an important role in practice. Our findings and the proposed GFE estimator are likely to be of increasing importance and of practical use to empirical researchers.

Figure 1
Firms' sales growth responses to natural disasters in 2004

Figure A shows the share of total U.S. employment affected by 4 natural disasters, Hurricane Charley, Hurricane Frances, Hurricane Ivan, and Hurricane Jeanne, with damages exceeding $1 billion in 2004. Employment numbers are computed from County Business Pattern data, publicly provided by the U.S. Census Bureau. Figures B, C, and D present the time effect estimates of the sales growth regression using GFE, TFE and IFE respectively for firms in the affected states in 2004, estimated for the sample period from 1987-2010 using data from Compustat. Figure B shows the estimate $\theta_{g_i,t}$ for GFE: $y_{it} = \alpha_i + \theta_{g_i,t} + X'_{it}\beta + \epsilon_{it}, \quad g_i = 1, 2, ...$ Figure C shows the estimate θ_t for TFE $y_{it} = \alpha_i + \theta_t + X'_{it}\beta + \epsilon_{it}$. Figure D shows the estimate $\theta_{industry,i,t}$ for IFE $y_{it} = \alpha_i + \theta_{industry,i,t} + X'_{it}\beta + \epsilon_{it}$. y_{it} is log annual sales changes and the vector of X_{it} consists of: natural log total assets (TA), Tobin's q, return on assets (ROA), capital expenditure (Capex), leverage (Leverage), and free cash flow (FCF). See Appendix Table A.2 for the description of variables.

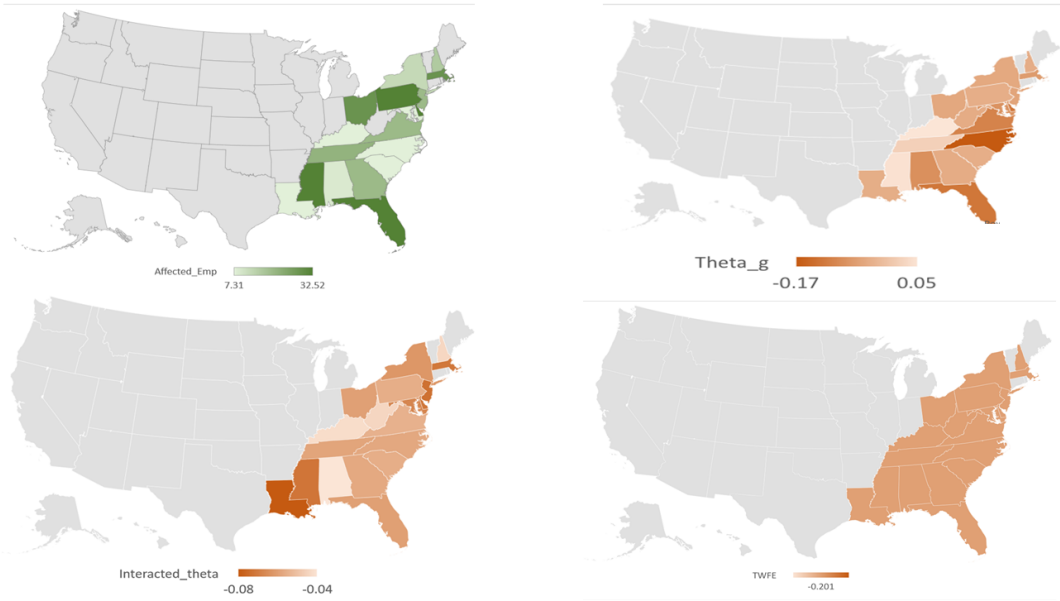
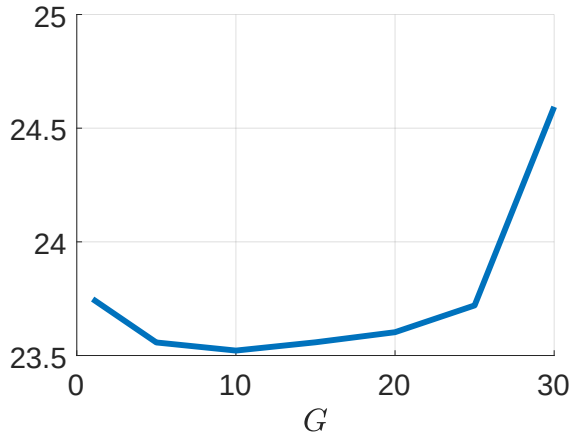
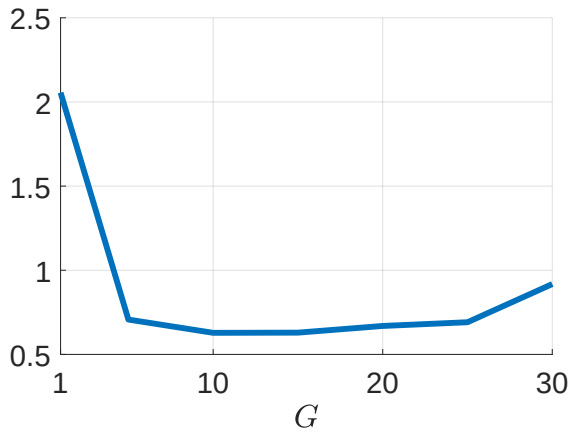


Figure 2
Bayesian information criterion and number of groups

This figure presents the Bayesian information criterion for different numbers of groups for the empirical examples. Figure A presents the BIC for the natural disaster regression $y_{it} = \alpha_i + \theta_{g_i,t} + X'_{it}\beta + \epsilon_{it}, \quad g_i = 1, 2, ...$ y_{it} is log annual sales changes and the vector of X_{it} consists of: natural log total assets, Tobin's q, return on assets, capital expenditure, leverage, and free cash flow. Figure B presents the BIC for the pilot CEO regression $\ln(1+\text{citation})_{it} = \alpha_{ind} + \theta_{g_i,t} + X'_{i,t}\beta + \epsilon_{i,t}$, $g_i \in \{1, ..., G\}$. $\ln(1+\text{citation})_{it}$ is the natural log of the number of raw citations multiplied by the weighting index of Hall et al. (2001) to all the patents applied for by firm i in year t, α_{ind} is the industry fixed effect and $\theta_{g_i,t}$ is a time-varying group effect, X_{it} is a 17×1 vector of control variables: pilot CEO dummy, log total assets, ratio of net property, plant, and equipment over the number of employees, stock returns, Tobin's Q, institutional holdings, CEO tenure, CEO stock and option delta and vega, CEO age, top university dummy, PhD, dummy for no school information, finance education, technical education, CEO military background dummy, CEO overconfidence. The x-axis is the number of groups, and the y-axis is the BIC. See Appendix Table A.2 for the description of variables.



(a) Natural disaster



(b) Pilot CEO

Table 2
Frequent moderate shocks

The table presents the results for simulations under DGP1 for different econometric specifications. We present the bias: $\text{Bias}(\widehat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}\widehat{\beta}_r - \beta$, the estimated standard deviation (Std.): $\text{Std}(\widehat{\beta}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\widehat{\beta}_r - \bar{\widehat{\beta}})^2}$, and the root mean square error (RMSE): $\text{RMSE}(\widehat{\beta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\|\widehat{\beta}_r - \beta\|^2}$ of estimated coefficients, where $\widehat{\beta}_r$ is the estimator in the r-th replication and $\bar{\widehat{\beta}} = 1/R\sum_{r=1}^{R}\widehat{\beta}_r$ is the sample average across replications. Misclassification frequency (Misclass. Freq.) is the average proportion of units that are misclassified by GFE across replications: $AMF = 1/R\sum_{r=1}^{R} MF_r$, where R is the number of replications and $MF_r = \frac{1}{N}\sum_{i=1}^{N} I(\widehat{g}_i \neq g_i^0)$. DGP1 is generated using the procedure described in Section 3, where shocks occur at every time t and heterogenous responses to shocks are allowed. The parameters in the DGP are set as $c_\tau = 0.5$, $c_\theta = 15$, and $G = 5$. c_τ is a constant which determines the degree of correlation between X and θ, c_θ is a constant which controls the size of shocks, G is the number of groups. The table reports the results using the following methods: grouped fixed effects (GFE), two-way fixed effects (TFE), and group-time interacted fixed effects (IFE). The IFE method includes 4 different specifications: IFE_{gt} exactly specifies the group memberships of all individuals and the time points of all shocks, IFE_g exactly specifies the groups but correctly specifies the shocks every 1/3 of the time, IFE_t correctly specifies the shocks but not the groups, and IFE_{null} misspecifies both groups and shocks. When the groups are misspecified, we assign 30% of individuals to the closest neighboring groups. The details of these methods are presented in Section 3.1.2. N is the number of units, and T is the number of time series observations. Panels A to C report results for three different sample sizes and T. Each simulation consists of 1000 replications.

		GFE	TFE	IFE_{gt}	IFE_g	IFE_t	IFE_{null}
Panel A. $N = 1000$, $T = 15$							
Misclass. Freq.		0.1491					
β_1	Bias	0.0139	0.5954	0.0002	0.5125	0.2548	0.5693
	Std.	0.0119	0.2057	0.0108	0.2203	0.1061	0.2148
β_2	Bias	0.0132	0.5991	−0.0006	0.5145	0.2590	0.5720
	Std.	0.0124	0.2102	0.0109	0.2206	0.1089	0.2133
β	RMSE	0.0183	0.8858	0.0144	0.9898	1.0035	0.8500
Panel B. $N = 1000$, $T = 30$							
Misclass. Freq.		0.0739					
β_1	Bias	0.0074	0.6107	0.0004	0.6336	0.2625	0.6577
	Std.	0.0074	0.1383	0.0072	0.1465	0.0733	0.1449
β_2	Bias	0.0071	0.6110	0.0002	0.6347	0.2591	0.6588
	Std.	0.0077	0.1381	0.0072	0.1453	0.0719	0.1433
β	RMSE	0.0150	0.8869	0.0102	0.9235	0.3831	0.9558
Panel C. $N = 2000$, $T = 15$							
Misclass. Freq.		0.1460					
β_1	Bias	0.0126	0.6045	0.0005	0.5208	0.2571	0.5778
	Std.	0.0093	0.2006	0.0077	0.1978	0.1055	0.1950
β_2	Bias	0.0122	0.6072	−0.0002	0.5217	0.2666	0.5806
	Std.	0.0089	0.2064	0.0072	0.2102	0.1088	0.2052
β	RMSE	0.0215	0.9008	0.0105	0.7929	0.3969	0.8664

Table 3
Sparse moderate shocks

The table presents the results for simulations under DGP2 for different econometric specifications. We present the bias: $\text{Bias}(\widehat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}\widehat{\beta}_r - \beta$, the estimated standard deviation (Std.): $\text{Std}(\widehat{\beta}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\widehat{\beta}_r - \bar{\widehat{\beta}})^2}$, and the root mean square error (RMSE): $\text{RMSE}(\widehat{\beta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\|\widehat{\beta}_r - \beta\|^2}$ of estimated coefficients, where $\widehat{\beta}_r$ is the estimator in the r-th replication and $\bar{\widehat{\beta}} = 1/R\sum_{r=1}^{R}\widehat{\beta}_r$ is the sample average across replications. Misclassification frequency (Misclass. Freq.) is the average proportion of units that are misclassified by GFE across replications: $AMF = 1/R\sum_{r=1}^{R}MF_r$, where R is the number of replications and $MF_r = \frac{1}{N}\sum_{i=1}^{N}I(\widehat{g}_i \neq g_i^0)$. DGP2 is generated using the procedure described in Section 3, where shocks occur 25% of the time and heterogenous responses to shocks are allowed. The parameters in the DGP are set as $c_\tau = 0.5$, $c_\theta = 15$, and $G = 5$. c_τ is a constant which determines the degree of correlation between X and θ, c_θ is a constant which controls the size of shocks, G is the number of groups. The table reports the results using the following methods: grouped fixed effects (GFE), two-way fixed effects (TFE), and group-time interacted fixed effects (IFE). The IFE method includes 5 different specifications: IFE_{gt} exactly specifies the group memberships of all individuals and the time points of all shocks, IFE_g exactly specifies the groups but correctly specifies the shocks every 1/3 of the time, IFE_{got} exactly specifies the groups but over-specifies the shocks at each time period, IFE_t correctly specifies the shocks but not the groups, and IFE_{null} misspecifies both groups and shocks. When the groups are misspecified, we assign 30% of individuals to the closest neighboring groups. The details of these methods are presented in Section 3.1.2. N is the number of units, and T is the number of time series observations. Panels A to C report results for different N and T. Each simulation consists of 1000 replications.

		GFE	TFE	IFE_{gt}	IFE_g	IFE_{got}	IFE_t	IFE_{null}
Panel A. $N = 1000$, $T = 15$								
Misclass. Freq.		0.0953						
β_1	Bias	0.0032	0.3584	−0.0002	0.2712	−0.0001	0.1340	0.3178
	Std.	0.0130	0.3537	0.0093	0.3835	0.0105	0.1328	0.3609
β_2	Bias	0.0035	0.3386	−0.0008	0.2410	−0.0005	0.1268	0.2897
	Std.	0.0161	0.3191	0.0086	0.3345	0.0100	0.1257	0.3159
β	RMSE	0.0212	0.6852	0.0127	0.6246	0.0144	0.2596	0.6438
Panel B. $N = 1000$, $T = 30$								
Misclass. Freq.		0.0482						
β_1	Bias	0.0018	0.4515	0.0003	0.4489	0.0001	0.1657	0.4744
	Std.	0.0077	0.2559	0.0062	0.3171	0.0068	0.0978	0.3068
β_2	Bias	0.0016	0.4450	0.0001	0.4425	0.0003	0.1618	0.4672
	Std.	0.0081	0.2639	0.0062	0.3214	0.0071	0.0973	0.3069
β	RMSE	0.0114	0.7326	0.0087	0.7751	0.0098	0.2695	0.7945
Panel C. $N = 2000$, $T = 15$								
Misclass. Freq.		0.0973						
β_1	Bias	0.0037	0.3636	0.0004	0.2754	0.0004	0.1280	0.3186
	Std.	0.0177	0.3438	0.0063	0.3622	0.0072	0.1240	0.3456
β_2	Bias	0.0016	0.3682	−0.0005	0.2866	−0.0005	0.1262	0.3316
	Std.	0.0125	0.3216	0.0063	0.3655	0.0071	0.1178	0.3381
β	RMSE	0.0220	0.6993	0.0089	0.6498	0.0101	0.2480	0.6669

Table 4
Homogeneous shocks

The table presents the results for simulations under DGP3 for different econometric specifications. We present the bias: $\text{Bias}(\hat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}\hat{\beta}_r - \beta$, the estimated standard deviation (Std.): $\text{Std}(\hat{\beta}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\beta}_r - \bar{\hat{\beta}})^2}$, and the root mean square error (RMSE): $\text{RMSE}(\hat{\beta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\|\hat{\beta}_r - \beta\|^2}$ of estimated coefficients, where $\hat{\beta}_r$ is the estimator in the r-th replication and $\bar{\hat{\beta}} = 1/R\sum_{r=1}^{R}\hat{\beta}_r$ is the sample average across replications. DGP3 is generated using the procedure described in Section 3, where shocks occur at every time t and there is a homogenous response to the shock. The parameters in the DGP are set as $c_\tau = 0.5$, $c_\theta = 15$, and $G = 1$. c_τ is a constant which determines the degree of correlation between X and θ, c_θ is a constant which controls the size of shocks, G is the number of groups. The table reports the results using the following methods: grouped fixed effects (GFE), two-way fixed effects (TFE), and group-time interacted fixed effects (IFE). GFE is estimated under two specifications of groups, $\widehat{G} = 1$ and $\widehat{G} = 4$. The IFE method includes 4 different specifications: IFE_{gt} exactly specifies the group memberships of all individuals and the time points of all shocks, IFE_g exactly specifies the groups but correctly specifies the shocks every 1/3 of the time, IFE_t correctly specifies the shocks but not the groups, and IFE_{null} misspecifies both groups and shocks. The details of these methods are presented in Section 3.1.2. N is the number of units, and T is the number of time series observations. Panels A to C report results for different N and T. Each simulation consists of 1000 replications.

		GFE ($\widehat{G}=1$)	GFE ($\widehat{G}=4$)	TFE	IFE_{gt}	IFE_g	IFE_t	IFE_{null}
				Panel A. $N=1000$, $T=15$				
β_1	Bias	0.0003	0.0007	0.0003	0.0003	0.0557	0.0003	0.0531
	Std.	0.0106	0.0095	0.0106	0.0106	0.0632	0.0106	0.0384
β_2	Bias	−0.0002	0.0000	−0.0002	−0.0002	0.0542	−0.0001	0.0494
	Std.	0.0110	0.0095	0.0110	0.0110	0.0633	0.0110	0.0381
β	RMSE	0.0152	0.0167	0.0152	0.0152	0.1184	0.0153	0.0905
				Panel B. $N=1000$, $T=30$				
β_1	Bias	0.0002	0.0003	0.0002	0.0002	0.0666	0.0002	0.0658
	Std.	0.0073	0.0069	0.0073	0.0073	0.0385	0.0073	0.0289
β_2	Bias	0.0001	0.0000	0.0001	0.0001	0.0667	0.0001	0.0669
	Std.	0.0071	0.0069	0.0071	0.0071	0.0402	0.0071	0.0275
β	RMSE	0.0102	0.0106	0.0102	0.0102	0.1089	0.0102	0.1006
				Panel C. $N=2000$, $T=15$				
β_1	Bias	0.0000	−0.0001	0.0000	0.0000	0.0566	0.0000	0.0527
	Std.	0.0071	0.0068	0.0071	0.0071	0.0627	0.0071	0.0362
β_2	Bias	0.0000	−0.0001	0.0000	0.0000	0.0535	0.0000	0.0490
	Std.	0.0074	0.0068	0.0074	0.0074	0.0623	0.0074	0.0344
β	RMSE	0.0102	0.0113	0.0102	0.0102	0.1177	0.0103	0.0875

Table 5
Large and highly correlated shocks

The table presents the results for simulations under DGP1 for different shock and correlation sizes and different econometric specifications. We present the bias: $\text{Bias}(\hat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}\hat{\beta}_r - \beta$, the estimated standard deviation (Std.): $\text{Std}(\hat{\beta}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\beta}_r - \bar{\hat{\beta}})^2}$, and the root mean square error (RMSE): $\text{RMSE}(\hat{\beta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\|\hat{\beta}_r - \beta\|^2}$ of estimated coefficients, where $\hat{\beta}_r$ is the estimator in the r-th replication and $\bar{\hat{\beta}} = 1/R\sum_{r=1}^{R}\hat{\beta}_r$ is the sample average across replications. Misclassification frequency (Misclass. Freq.) is the average proportion of units that are misclassified by GFE across replications: $AMF = 1/R\sum_{r=1}^{R} MF_r$, where R is the number of replications and $MF_r = \frac{1}{N}\sum_{i=1}^{N} I(\hat{g}_i \neq g_i^0)$. DGP1 is generated using the procedure described in Section 3, where shocks occur at every time t and heterogenous responses to shocks are allowed. The table reports the results using the following methods: grouped fixed effects (GFE), two-way fixed effects (TFE), and group-time interacted fixed effects (IFE). The IFE method includes 4 different specifications: IFE_{gt} exactly specifies the group memberships of all individuals and the time points of all shocks, IFE_g exactly specifies the groups but correctly specifies the shocks every 1/3 of the time, IFE_t correctly specifies the shocks but not the groups, and IFE_{null} misspecifies both groups and shocks. When the groups are misspecified, we assign 40% of individuals to the closest neighboring groups. The details of these methods are presented in Section 3.1.2. Panel A presents the results for $c_\tau = 0.5$, a larger time series shock $c_\theta = 5$, and $G = 5$, and Panel B presents the results for a larger correlation c_τ=1. c_τ is a constant which determines the degree of correlation between X and θ, c_θ is a constant which controls the size of shocks, G is the number of groups. We report the results for 1000 cross-sectional units (N), 30 time periods (T), and 5 groups (G). Each simulation consists of 1000 replications.

($N = 1000$, $T = 30$)		GFE	TFE	IFE_{gt}	IFE_g	IFE_t	IFE_{null}
Panel A. Large shocks: $c_\tau = 0.5$, $c_\theta = 5$, and $G = 5$							
Misclass. Freq.		0.0061					
β_1	Bias	0.0012	1.8553	−0.0002	1.9234	0.8076	1.9969
	Std.	0.0073	0.4343	0.0068	0.4613	0.2225	0.4565
β_2	Bias	0.0013	1.8431	0.0000	1.9104	0.7990	1.9844
	Std.	0.0076	0.4295	0.0073	0.4588	0.2189	0.4542
β	RMSE	0.0107	2.6855	0.0100	2.7878	1.1781	2.8878
Panel B. Highly correlated shocks: $c_\tau = 1$, $c_\theta = 15$, and $G = 5$							
Misclass. Freq.		0.1385					
β_1	Bias	0.0148	0.6809	−0.0004	0.6930	0.3723	0.7051
	Std.	0.0091	0.1079	0.0071	0.1174	0.0828	0.1119
β_2	Bias	0.0144	0.6833	−0.0004	0.6961	0.3745	0.7085
	Std.	0.0087	0.1101	0.0069	0.1161	0.0826	0.1125
β	RMSE	0.0242	0.9769	0.0099	0.9960	0.5409	1.0120

Table 6
Under-specify G

The table presents the results for simulations under DGP1 for an underspecified number of groups G for GFE and TFE. We present the bias: $\text{Bias}(\widehat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}\widehat{\beta}_r - \beta$, the estimated standard deviation (Std.): $\text{Std}(\widehat{\beta}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\widehat{\beta}_r - \bar{\widehat{\beta}})^2}$, and the root mean square error (RMSE): $\text{RMSE}(\widehat{\beta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\|\widehat{\beta}_r - \beta\|^2}$ of estimated coefficients, where $\widehat{\beta}_r$ is the estimator in the r-th replication and $\bar{\widehat{\beta}} = 1/R\sum_{r=1}^{R}\widehat{\beta}_r$ is the sample average across replications. DGP1 is generated using the procedure described in Section 3, where shocks occur at every time t and heterogenous responses to shocks are allowed. The parameters in the DGP are set as $c_\tau = 0.5$, $c_\theta = 15$, and $G = 4$. c_τ is a constant which determines the degree of correlation between X and θ, c_θ is a constant which controls the size of shocks, G is the number of groups. The table reports the results using grouped fixed effects (GFE) and two-way fixed effects (TFE). The estimates are generated using two groups, $\widehat{G} = 2$, where the true DGP of G=4. The details of these methods are presented in Section 3.1.2. N is the number of units, and T is the number of time series observations. Each simulation consists of 1000 replications.

		N = 1000, T = 15		N = 1000, T = 30		N = 2000, T = 15	
		GFE	TFE	GFE	TFE	GFE	TFE
β_1	Bias	0.0799	0.2772	0.0808	0.2848	0.0781	0.2833
	Std.	0.0397	0.1082	0.0277	0.0793	0.0365	0.1027
β_2	Bias	0.0778	0.2762	0.0819	0.2826	0.0777	0.2782
	Std.	0.0416	0.1061	0.0266	0.0764	0.0391	0.1117
β	RMSE	0.1255	0.4196	0.1213	0.4160	0.1224	0.4250

Table 7
More groups G

The table presents the results for simulations under DGP1 for two different number of groups G and different econometric specifications. We present the bias: $\text{Bias}(\hat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}\hat{\beta}_r - \beta$, the estimated standard deviation (Std.): $\text{Std}(\hat{\beta}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\beta}_r - \bar{\hat{\beta}})^2}$, and the root mean square error (RMSE): $\text{RMSE}(\hat{\beta}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\|\hat{\beta}_r - \beta\|^2}$ of estimated coefficients, where $\hat{\beta}_r$ is the estimator in the r-th replication and $\bar{\hat{\beta}} = 1/R\sum_{r=1}^{R}\hat{\beta}_r$ is the sample average across replications. Misclassification frequency (Misclass. Freq.) is the average proportion of units that are misclassified by GFE across replications: $AMF = 1/R\sum_{r=1}^{R} MF_r$, where R is the number of replications and $MF_r = \frac{1}{N}\sum_{i=1}^{N} I(\hat{g}_i \neq g_i^0)$. DGP1 is generated using the procedure described in Section 3, where shocks occur at every time t and heterogenous responses to shocks are allowed. The parameters in the DGP are set as $c_\tau = 0.5$, $c_\theta = 15$. c_τ is a constant which determines the degree of correlation between X and θ, c_θ is a constant which controls the size of shocks. We consider the number of groups $G = 10$ and $G = 20$ in DGP. The table reports the results using the following methods: grouped fixed effects (GFE), two-way fixed effects (TFE), and group-time interacted fixed effects (IFE). The IFE method includes 4 different specifications: IFE_{gt} exactly specifies the group memberships of all individuals and the time points of all shocks, IFE_g exactly specifies the groups but correctly specifies the shocks every 1/3 of the time, IFE_t correctly specifies the shocks but not the groups, and IFE_{null} misspecifies both groups and shocks. When the groups are misspecified, we assign 30% of individuals to the closest neighboring groups. The details of these methods are presented in Section 3.1.2. We report the results for 1000 cross-sectional units (N) and 30 time periods (T). Each simulation consists of 1000 replications.

		GFE	TFE	IFE_{gt}	IFE_g	IFE_t	IFE_{null}
				$G = 10$ $(N = 1000, T = 30)$			
Misclass. Freq.		0.0734					
β_1	Bias	0.0040	5.0593	−0.0002	4.9302	2.8958	5.0519
	Std.	0.0073	0.6131	0.0071	0.6419	0.4988	0.6227
β_2	Bias	0.0048	5.0731	0.0006	4.9358	2.8899	5.0528
	Std.	0.0079	0.5423	0.0077	0.5978	0.4745	0.5704
β	RMSE	0.0124	7.2112	0.0105	7.0312	4.1485	7.1947
				$G = 20$ $(N = 1000, T = 30)$			
Misclass. Freq.		0.0613					
β_1	Bias	0.0051	5.0402	0.0011	4.9088	2.8528	5.0238
	Std.	0.0078	0.6268	0.0075	0.6510	0.4893	0.6313
β_2	Bias	0.0039	5.0259	−0.0002	4.8785	2.8488	4.9996
	Std.	0.0076	0.5917	0.0075	0.6173	0.4883	0.6025
β	RMSE	0.0126	7.1697	0.0107	6.9785	4.0904	7.1411

Table 8
Sales growth and major natural disasters

Panel A reports results from univariate regressions of TFE, IFE and GFE estimated time effects on the percentage of employment affected by U.S. major disasters, $\theta = \alpha + \beta A.\mathsf{Employment}_{it} + \epsilon_{i,t}$, where i is the state and θ is the estimated year/state effect from the regressions below. $\theta_{g_i,t}$ is the GFE estimates: $y_{it} = \alpha_i + \theta_{g_i,t} + X'_{it}\beta + \epsilon_{it}, \quad g_i = 1, 2, ...$ θ_t is the TFE estimates from $y_{it} = \alpha_i + \theta_t + X'_{it}\beta + \epsilon_{it}$. $\theta_{industry,i,t}$ is the IFE estimates from $y_{it} = \alpha_i + \theta_{industry,i,t} + X'_{it}\beta + \epsilon_{it}$. y_{it} is log annual sales changes and the vector of X_{it} consists of: *TA* is log of total assets, *ROA* is return on assets, *Leverage* is long-term debt plus short-term debt scaled by total assets, *Capex* is capital expenditure scaled by total assets, Tobin's Q is the natural logarithm of the book value of debt plus the market value of equity scaled by the book value of total assets, *FCF* is free cash flow. Disaster affected employment numbers are computed from County Business Pattern data, publicly provided by the U.S. Census Bureau. All the rest of the variables are from Compustat and the sample period is 1987-2010. See Appendix Table A.2 for the description of variables. *Panel B* presents regressions of log annual sales changes on explanatory variables using different estimation models. TFE is the firm and year fixed effect estimator, IFE is the firm fixed effect with industry and year interaction estimator. GFE estimates are obtained based on $G = 10, 15, 30$, selected from a range of G based on various information criteria. For IFE, preliminary within-transformation is taken to integrate out the individual specific effect as in GFE. The industry IFE is based on 2-digit SIC code. Clustered standard errors are reported in parentheses except for TFE and IFE. For GFE, we present the asymptotic standard errors. ***, **, * denote significance at 1%, 5%, and 10% level, respectively. p-values of Hausman tests for TFE and IFE against GFE ($G = 10$) are presented.

Panel A. Relation of TFE, IFE, GFE, and disaster-affected employment

	TFE	IFE	GFE
Disaster-affected Employment	0.0001	0.0003	-0.0025
t-stats	2.13	3.86	-3.85
# Obs		66,185	

Panel B. Sales growth regression with TFE, IFE, and GFE

	TFE	IFE	GFE $G = 10$	GFE $G = 15$	GFE $G = 30$
TA	0.0232***	0.0045	0.0074***	0.0094***	0.0085***
	(0.0052)	(0.0043)	(0.0032)	(0.0029)	(0.0026)
ROA	-0.0001	0.0113	0.0101	0.0066	0.0149
	(0.0151)	(0.0127)	(0.0089)	(0.0088)	(0.0126)
Capex	5.60e-06***	3.08e-06	2.87e-06*	3.16e-06**	3.46e-06**
	(2.22e-06)	(1.92e-06)	(1.69e-06)	(1.60e-06)	(1.79e-06)
Leverage	-0.0146	-0.0215**	-0.0141*	-0.0165**	-0.0148*
	(0.0122)	(0.0105)	(0.0078)	(0.0073)	(0.0084)
Tobin's Q	0.0003***	0.0001*	0.0001***	0.0001***	0.0001***
	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
FCF	0.0010**	0.0000	0.0002***	0.0001***	0.0002***
	(0.0005)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
p-val of Hausman test	0.000	0.000	–	–	–
Obs	66,185	66,185	66,185	66,185	66,185

Table 9
Compustat data based simulation

This table presents the misclassification frequency, absolute value of relative bias for each regressor, and overall RMSE in the simulation based on the Compustat data for sales growth and major natural disasters. *TA* is log of total assets, *ROA* is return on assets, *Leverage* is long-term debt plus short-term debt scaled by total assets, *Capex* is capital expenditure scaled by total assets, and *FCF* is free cash flow. See more detailed variable description in the Appendix, Table A.2, for definition of all variables and of Tobin's Q. Each simulation consists of 1000 replications.

		$T = 15, G = 5$			$T = 15, G = 10$		
		GFE	TFE	IFE	GFE	TFE	IFE
Misclassification freq.		0.0565			0.1439		
|Relative bias|	TA	-0.0490	0.1680	0.1805	-0.0340	0.6930	0.6316
	ROA	-0.0075	-0.1037	0.8348	-0.1235	1.0103	3.9146
	Capex	0.0699	-0.0265	-0.0118	0.3648	2.0938	1.4327
	Leverage	0.0754	0.7029	0.9908	-0.5154	-9.4913	-16.519
	Tobin's Q	-3.6670	-11.741	-4.3948	-0.0592	-0.6009	-1.0451
	FCF	0.1121	0.6762	0.6768	-0.1762	-3.7295	-3.4110
Overall RMSE		0.0378	0.0775	0.0765	0.0528	0.1862	0.1791

Table 10
Pilot CEOs and patent citations

This table presents the results of the effect of pilot CEOs on patent citation counts. TFE is the industry and year fixed effect estimator, IFE is a firm fixed effect with industry and year interaction estimator. GFE estimates are obtained based on $G = 10, 15$, selected from a range of G from 1 to 30 based on various information criteria, and include industry fixed effects. For IFE, preliminary within-transformation is taken to integrate out the industry specific effect as in GFE. The industry IFE is based on two-digit SIC code. Following Sunder et al. (2017), the dependent variable is Log(1 + citation). Citation is the number of raw citations multiplied by the weighting index of Hall et al. (2001) to all the patents applied for during the year. Pilot CEO is an indicator variable equal to one if the CEO has been a pilot and zero otherwise. All regressors are lagged by one year. Variable definitions are provided in the Appendix, Table A.2. Following Sunder et al. (2017), the sample period is from 1993 to 2003. Standard errors are clustered at firm level for TFE and IFE. For GFE, the asymptotic standard errors are used for inference. t-statistics are reported in parentheses. ***, **, * denote significance at 1%, 5%, and 10% level, respectively. p-values of Hausman tests for TFE and IFE against GFE ($G = 10$) are presented.

	TFE	IFE	GFE $G = 10$	GFE $G = 15$
Pilot CEO	0.376*	0.203*	-0.027	-0.043
	(1.67)	(1.66)	(-0.44)	(-0.65)
TA	0.589***	0.611***	0.264***	0.335***
	(9.06)	(9.15)	(13.73)	(14.64)
PPE/EMP	0.000	-0.0002***	0.000	0.000
	(2.70)	(-3.64)	(1.07)	(-0.40)
Stock return	-0.098	-0.168***	0.001	-0.020
	(-1.33)	(-4.31)	(-0.01)	(-0.62)
Tobin's Q	0.156***	0.114***	0.020**	0.045***
	(3.04)	(8.38)	(2.01)	(3.48)
Institutional holdings	0.148	-0.061	-0.331***	-0.136
	(0.53)	(-0.43)	(-5.92)	(-1.43)
Log(1 + tenure)	-0.149***	0.0002	-0.079**	-0.104***
	(-2.22)	(0.00)	(-3.34)	(-3.45)
Log(1 + delta)	-0.001	0.035	0.029**	0.031***
	(-0.01)	(1.11)	(2.18)	(3.30)
Log(1 + vega)	0.145**	0.066	0.022	0.002
	(2.18)	(1.30)	(1.28)	(0.14)
Log(CEO age)	0.296	-0.562	0.213	0.400
	(0.45)	(-1.34)	(0.87)	(1.63)
Top university	-0.002	-0.037	-0.240***	-0.065
	(-0.01)	(-0.21)	(-3.96)	(-0.90)
Finance education	0.520***	0.243	1.448***	1.438***
	(3.02)	(0.40)	(3.21)	(3.05)
Technical education	1.253	0.960	1.795***	1.513***
	(1.00)	(1.61)	(10.02)	(8.67)
PhD	0.154	-0.042	0.090	0.201***
	(0.56)	(-0.26)	(1.08)	(2.72)
No school information	0.299*	0.044	0.028	0.086
	(1.84)	(0.60)	(0.54)	(1.59)
Military	0.004	-0.043***	-0.017***	-0.014***
	(0.62)	(-7.11)	(-7.51)	(-4.72)
Overconfidence	0.109	0.034	0.051	0.002
	(0.92)	(0.31)	(1.15)	(0.05)
p-val of Hausman test	0.000	0.000	–	–
Obs	2450	2450	2450	2450

Table 11
Frequent moderate shocks with endogenous regressors

The table presents the misclassification frequency, bias, and root mean square error of estimated coefficients for simulations under DGP1 with endogenous regressors. The data are generated in the same way as DGP1 except that X_{it} is correlated with the error term. Both GFE and TFE employ TSLS estimation with X_{it} instrumented by Z_{it}. Each simulation consists of 1000 replications.

		$N = 1000, T = 15$		$N = 1000, T = 30$		$N = 2000, T = 30$	
		GFE	TFE	GFE	TFE	GFE	TFE
Misclassification frequency		0.1877		0.1252		0.1232	
β_1	Bias	0.0215	0.6010	0.0093	0.6112	0.0101	0.6255
β_2	Bias	0.0232	0.6087	0.0105	0.6156	0.0099	0.6323
β	RMSE	0.0378	0.9028	0.0172	0.8561	0.0221	0.9063

Table 12
Residual cross product matrix: Group fixed effects

This table presents a sample covariance matrix of the residuals for a group fixed effect DGP. In the presence of group fixed effecs, residuals of the same firm across different years as well as residuals of the same year and the same group, may be correlated. However, residuals of the same year but different groups are assumed to be uncorrelated, and thus reported as zero in the matrix.

		Firm 1			Firm 2			Firm 3		
Group 1	Firm 1	ϵ_{11}^2	$\epsilon_{11}\epsilon_{12}$	$\epsilon_{11}\epsilon_{13}$	$\epsilon_{11}\epsilon_{21}$	0	0	0	0	0
		$\epsilon_{21}\epsilon_{11}$	ϵ_{12}^2	$\epsilon_{12}\epsilon_{13}$	0	$\epsilon_{12}\epsilon_{22}$	0	0	0	0
		$\epsilon_{13}\epsilon_{11}$	$\epsilon_{13}\epsilon_{12}$	ϵ_{13}^2	0	0	$\epsilon_{13}\epsilon_{23}$	0	0	0
	Firm 2	$\epsilon_{21}\epsilon_{11}$	0	0	ϵ_{21}^2	$\epsilon_{21}\epsilon_{22}$	$\epsilon_{21}\epsilon_{23}$	0	0	0
		0	$\epsilon_{22}\epsilon_{12}$	0	$\epsilon_{22}\epsilon_{21}$	ϵ_{22}^2	$\epsilon_{22}\epsilon_{23}$	0	0	0
		0	0	$\epsilon_{23}\epsilon_{13}$	$\epsilon_{23}\epsilon_{21}$	$\epsilon_{23}\epsilon_{22}$	ϵ_{23}^2	0	0	0
Group 2	Firm 3	0	0	0	0	0	0	ϵ_{31}^2	$\epsilon_{31}\epsilon_{32}$	$\epsilon_{31}\epsilon_{33}$
		0	0	0	0	0	0	$\epsilon_{32}\epsilon_{31}$	ϵ_{32}^2	$\epsilon_{32}\epsilon_{33}$
		0	0	0	0	0	0	$\epsilon_{33}\epsilon_{31}$	$\epsilon_{33}\epsilon_{32}$	ϵ_{33}^2

References

Bai, J. (2009). "Panel data models with interactive fixed effects." *Econometrica*, 77(4), 1229–1279.

Banerjee, S., M. H. Jenner, and V. Nanda (2015). "Restraining overconfident CEOs through improved governance: Evidence from the Sarbanes-Oxley Act." *Review of Financial Studies*, 28, 2812–2858.

Bargeron, L. L., K. M. Lehn, and C. J. Zutter (2010). "Sarbanes-oxley and corporate risk-taking." *Journal of Accounting and Economics*, 49(12), 34 – 52.

Barrot, J.-N. and J. Sauvagnat (2016). "Input specificity and the propagation of idiosyncratic shocks in production networks *." *The Quarterly Journal of Economics*, 131(3), 1543–1592.

Benmelech, E. and C. Frydman (2015). "Military ceos." *Journal of Financial Economics*, 117(1), 43–59.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics*, 119(1), 249.

Bonhomme, S., T. Lamadon, and E. Manresa (2017). "Discretizing unobserved heterogeneity." Working paper.

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

Bonhomme, S. and E. Manresa (2015b). "Supplement to 'grouped patterns of heterogeneity in panel data'." *Econometrica Supplemental Material*, 83.

Brochet, F. (2010). "Information content of insider trades before and after the Sarbanes-Oxley Act." *The Accounting Review*, 85(2), 419–446.

Campello, M., E. Giambona, J. R. Graham, and C. R. Harvey (2011). "Liquidity management and corporate investment during a financial crisis." *The Review of Financial Studies*, 24(6), 19–44.

Campello, M., J. R. Graham, and C. R. Harvey (2010). "The real effects of financial constraints: Evidence from a financial crisis." *Journal of Financial Economics*, 97(3), 470 – 487.

Carter, M. E., L. J. Lynch, and S. L. C. Zechman (2009). "Changes in bonus contracts in the post-sarbanes–oxley era." *Review of Accounting Studies*, 14(4), 480–506.

Chhaochharia, V. and Y. Grinstein (2007). "Corporate governance and firm value: The impact of the 2002 governance rules." *The Journal of Finance*, 62(4), 1789–1825.

Cohen, D. A., A. Dey, and T. Z. Lys (2008). "Real and accrual-based earnings management in the pre- and post-Sarbanes-Oxley periods." *The Accounting Review*, 83(3), 757–787.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

Duchin, R., J. G. Matsusaka, and O. Ozbas (2010a). "When are outside directors effective?" *Journal of Financial Economics*, 96(2), 195 – 214.

Duchin, R., O. Ozbas, and B. A. Sensoy (2010b). "Costly external finance, corporate investment, and the subprime mortgage credit crisis." *Journal of Financial Economics*, 97(3), 418 – 435.

Engel, E., R. M. Hayes, and X. Wang (2007). "The Sarbanes-Oxley Act and firms' going-private decisions." *Journal of Accounting and Economics*, 44(1&2), 116 – 145.

Erickson, T. and T. M. Whited (2012). "Treating measurement error in Tobin's q." *The Review of Financial Studies*, 25(4), 1286–1329.

Forgy, E. (1965). "Cluster analysis of multivariate data: Efficiency versus interpretability of classification." *Biometrics*, 21(3), 768–769.

Gormley, T. A. and D. A. Matsa (2014). "Common errors: How to (and not to) control for unobserved heterogeneity." *Review of Financial Studies*, 27, 617–661.

Hall, B., A. Jaffe, and M. Trajtenberg (2001). "The nber patent citations data file: Lessons, insights and methodological tools." Working paper.

Heron, R. A. and E. Lie (2007). "Does backdating explain the stock price pattern around executive stock option grants?" *Journal of Financial Economics*, 83(2), 271 – 295.

Hirshleifer, D., A. Low, and S. H. Teoh (2012). "Are overconfident CEOs better innovators?" *Journal of Finance*, 67(4), 1457–1498.

Ho, P.-H., C.-W. Huang, C.-Y. Lin, and J.-F. Yen (2016). "CEO overconfidence and financial crisis: Evidence from bank lending and leverage." *Journal of Financial Economics*, 120(1), 194 – 209.

Joh, S. W. (2003). "Corporate governance and firm profitability: Evidence from Korea before the economic crisis." *Journal of Financial Economics*, 68(2), 287 – 322.

Koh, K., D. A. Matsumoto, and S. Rajgopal (2008). "Meeting or beating analyst expectations in the post-scandals world: Changes in stock market rewards and managerial actions." *Contemporary Accounting Research*, 25, 1067–1098.

Koh, P.-S., D. Reeb, E. Sojli, W. W. Tham, and W. Wang (2018). "Unreported innovation." Working paper.

Lemmon, M. L. and K. V. Lins (2003). "Ownership structure, corporate governance, and firm value: Evidence from the East Asian financial crisis." *The Journal of Finance*, 58(4), 1445–1468.

Linck, J. S., J. M. Netter, and T. Yang (2009). "The effects and unintended consequences of the Sarbanes-Oxley Act on the supply and demand for directors." *The Review of Financial Studies*, 22(8), 3287–3328.

Liu, R., A. Schick, Z. Shang, Y. Zhang, and Q. Zhou (2018). "Identification and estimation in panel models with overspecified number of groups." Working paper.

Mitton, T. (2002). "A cross-firm analysis of the impact of corporate governance on the East Asian financial crisis." *Journal of Financial Economics*, 64(2), 215 – 241.

Neyman, J. and E. L. Scott (1948). "Consistent estimates based on partially consistent observations." *Econometrica*, 16, 1–32.

Petersen, M. A. (2009). "Estimating standard errors in finance panel data sets: Comparing approaches." *Review of Financial Studies*, 22, 435–480.

Steinley, D. (2006). "K-means clustering: A half-century synthesis." *British Journal of Mathematical and Statistical Psychology*, 59(1), 1–34.

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

Sunder, J., S. V. Sunder, and J. Zhang (2017). "Pilot CEOs and corporate innovation." *Journal of Financial Economics*, 123(1), 209–224.

Thakor, A. V. (2015). "The financial crisis of 2007-2009: Why did it happen and what did we learn?" *The Review of Corporate Finance Studies*, 4(2), 155–205.

Thompson, S. B. (2011). "Simple formulas for standard errors that cluster by both firm and time." *Journal of Financial Economics*, 99, 1–10.

A Appendix

A.1 Event study standard errors

Another popular method to control for heterogeneous responses in finance studies is event studies. There are two potential problems with event studies. First, event studies require that there are relatively few shocks over the entire time span, and the timing of the shocks needs to be precisely specified. However, in some cases, there may be many shocks over the whole sample period, or simply different time effects in each period. Then event studies are less reliable, since the time periods within each subsample is short. More importantly, if the timing of shocks is incorrectly identified or several shocks are ignored, event studies generally result in inconsistent estimates.

Second, the conventional standard errors provided by event studies are often inappropriate, because they do not take into account cross-sectional correlation. The group specific responses suggest that an unobserved group pattern of features exists, e.g. managerial and institutional characteristics, that drives heterogeneity in responses, and it is likely that these characteristics will also affect firm's responses to any kind of corporate related events. Therefore, we can expect unobserved group characteristics in both the independent variables and the residuals that are possibly changing over time. These time-varying group components cannot be canceled out by the fixed effects transformation.

To obtain appropriate standard error estimates in event studies, one should explicitly take into account the latent group pattern, and calculate the group-clustered standard error. Our GFE estimates of the group membership thus provide a natural and reliable estimate of the latent group pattern, which can be used to construct clustered standard errors.

The transformed residuals in the case of standard event studies still contain two components, a group-specific time-varying component $\gamma_{g_i,t}$ and an idiosyncratic component η_{it} for each individual unit, namely

$$\dot{\epsilon}_{it} = \gamma_{g_i,t} + \eta_{it}. \tag{20}$$

Also, the transformed independent variable $\dot{X}$ is driven by a group-specific time-varying component and an idiosyncratic component, and it can specified as

$$\dot{X}_{it} = \delta_{g_i,t} + \nu_{it}. \tag{21}$$

Here all four components $\gamma_{g_i,t}$, η_{it}, $\delta_{g_i,t}$, and ν_{it} are assumed to be independent with each other and across time, and they all have a zero mean and finite variance, σ_γ^2, σ_η^2, σ_δ^2, and σ_ν^2, respectively.

To see how the correlation causes bias in standard errors, we consider an example of a single regressor and thus X_{it} is a scalar. Due to the group-specific components, the residual and regressor of individual units (firms) are correlated with each other with the following correlation structure

$$\begin{aligned} \mathrm{corr}(\dot{X}_{it}, \dot{X}_{js}) &= 1 \quad \text{for } i = j \text{ and } t = s \\ &= \rho_x = \sigma_\delta^2/\sigma_{\dot{X}}^2 \quad \text{for } g_i = g_j \text{ and } t = s \\ &= 0 \quad \text{for all } g_i \neq g_j \text{ or } t \neq s, \end{aligned}$$

and

$$\begin{aligned} \mathrm{corr}(\dot{\epsilon}_{it}, \dot{\epsilon}_{js}) &= 1 \quad \text{for } i = j \text{ and } t = s \\ &= \rho_\epsilon = \sigma_\gamma^2/\sigma_\epsilon^2 \quad \text{for } g_i = g_j \text{ and } t = s \\ &= 0 \quad \text{for all } g_i \neq g_j \text{ or } t \neq s. \end{aligned}$$

Under such residual and regressor structure, the fixed effects coefficient estimate is still consistent, but its standard error estimate is downward biased due to the ignorance of the group correlation. In particular, we can obtain the asymptotic variance of the fixed effects coefficient estimate as

$$\begin{aligned} \mathrm{var}(\hat{\beta}_{FE}) &= \left[\left(\sum_{i=1}^{N} \sum_{t=1}^{T} \dot{X}_{it} \dot{\epsilon}_{it} \right)^2 \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \dot{X}_{it}^2 \right)^{-2} \right] \\ &= \left[\left(\sum_{i=1}^{N} \sum_{t=1}^{T} \dot{X}_{it}^2 \dot{\epsilon}_{it}^2 + \sum_{i,j \in g} \sum_{t=1}^{T} \dot{X}_{it} \dot{X}_{jt} \dot{\epsilon}_{it} \dot{\epsilon}_{jt} \right) \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \dot{X}_{it}^2 \right)^{-2} \right] \\ &= \left[NT\sigma_X^2 \sigma_\epsilon^2 + \sum_{g=1}^{G} N_g (N_g - 1) T \rho_X \rho_\epsilon \sigma_X^2 \sigma_\epsilon^2 \right] \left[NT\sigma_X^2 \right]^{-2} \\ &= \frac{\sigma_\epsilon^2}{NT\sigma_X^2} \left[1 + 1/N \sum_{g=1}^{G} N_g (N_g - 1) \rho_X \rho_\epsilon \right], \end{aligned} \tag{22}$$

where N_g is the number of individuals in the g-th group. Under a group structure, individuals within a group share the same component in both the independent variable and residuals, leading to a positive correlation. Therefore, the fixed effects standard error estimates underestimate the true standard errors.

Table A.1
Demonstrating example: Time-invariant heterogeneity

The table presents a simple numerical example of the two-way fixed effect absorbing the time invariant unobserved heterogeneity. Columns (1)-(3) present the basic data structure: (1) is the individual unique identification, (2) is the time period denomination, and (3) is the group identification. Columns (4)-(5) present the data: (4) is the individual fixed effect and (5) is the time effect. Column (6) shows the demeaned cross-sectional fixed effect and column (7) shows the demeaned time-series variable.

			Data			Transformed data	
Indiv. ID	Time	Group ID	Total	α_i	$\theta_{g,t}$	$\tilde{\alpha}_i$	$\tilde{\theta}_{g,t}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	1	1	5	2	3	0	0
1	2	1	6	2	4	0	0
1	3	1	7	2	5	0	0
2	1	1	7	4	3	0	0
2	2	1	8	4	4	0	0
2	3	1	9	4	5	0	0
3	1	2	10	6	4	0	0
3	2	2	11	6	5	0	0
3	3	2	12	6	6	0	0
4	1	2	12	8	4	0	0
4	2	2	13	8	5	0	0
4	3	2	14	8	6	0	0

Table A.2
Variable description

Variables	Description
Citation	The number of citations per firm summed across all patents applied for during the year. Each patent's number of citations is multiplied by the weighting index from Hall et al. (2001).
Employment affected by major disasters	The number of employees affected by natural disasters by U.S. state as computed by County Business Pattern data, publicly provided by the U.S. Census Bureau.
Pilot CEO	An indicator variable equal to one for CEOs with a pilot license and zero otherwise.
TA	Natural log of total assets in millions.
ROA	Return on assets.
Leverage	The long-term debt plus short-term debt divided by total assets.
Capex	The capital expenditure divided by total assets.
Cash flow	Lagged Operating Income before depreciation minus total income taxes, minus change in deferred taxes from the previous year to the current year based on minus gross interest expense minus preferred dividend requirement on cumulative preferred stock and dividends paid on non-cumulative preferred stock minus total dollar amount of dividends declared on common stock. Cash flow is divided by TA.
Tobin's Q	The market value of assets divided by the book value of assets where the market value of assets equals the book value of assets plus the market value of common equity less the sum of the book value of common equity and balance sheet deferred taxes.
QDum	1 if Tobin's Q is lower than 1; 0 if Tobin's Q is larger than or equal to 1.
FCF	Free cash flow = Cash flow multiply by QDum.
PPE/EMP	The ratio of net property, plant, and equipment over the number of employees.
Stock return	Firm buy-and-hold return over the fiscal year.
Institutional holdings	Percentage of shares held by financial institutions.
Tenure	CEO tenure in months.
Delta	Dollar change in CEO stock and option portfolio for a 1% change in stock price.
Vega	Dollar change in CEO option holdings for a 1% change in stock return volatility.
CEO age	CEO age in years.
Top university	An indicator variable equal to one if the CEO's undergraduate institution is listed as one of the top 50 schools ranked by U.S. News & World Report in any year during the period 1983 through 2007 and zero otherwise.
Finance education	An indicator variable equal to one if the CEO received a degree in accounting, finance, business (including MBA), or economics and zero otherwise.
Technical education	An indicator variable equal to one for CEOs with undergraduate or graduate degrees in engineering, physics, operations research, chemistry, mathematics, biology, pharmacy, or other applied science and zero otherwise.
PhD	An indicator variable equal to one for CEOs with a PhD and zero otherwise.
No school information	An indicator variable equal to one if we cannot identify the CEO's undergraduate school and zero otherwise.
Military	An indicator variable equal to one for CEOs with military background and zero otherwise.
Overconfidence	An indicator variable equal to one for all years after the CEO's options exceed 67% moneyness and zero otherwise.