*Pre-testing for Unobserved Cluster Effects and Inference in Panel Data Sets**

Ercio Muñoz† Wim Vijverberg‡

This draft: December 23, 2020

Abstract

This paper addresses the question of how to estimate the standard errors in panel data when unobserved cluster effects are potentially present. We analyze the performance of statistical inference regarding the parameters of a panel data model when it is first subjected to a pretest for the presence of individual and/or time unobserved cluster effects. Using Monte Carlo simulations we compare the performance of six proposed diagnostics that make use of statistical tests available in the literature such as Lagrange multiplier, Likelihood ratios, and F tests. We find that these six pretest estimators are a viable alternative to estimate panel data models with unobserved cluster effects, in the sense that they achieve empirical sizes very close to the ones obtained using an estimator of the variance as if we knew the true data generating process. These results are robust, at least in the context of our simulations, to the presence of temporary clusters effects, and non-normality of the disturbance, as well as non-normality of the regressor. We provide several empirical examples to illustrate the importance of our findings.

JEL-Codes: C01; G00.

Keywords: *Monte Carlo, Unobserved cluster effects, Hypothesis testing, Panel data, Pre-test estimator.*

*This research was supported, in part, under National Science Foundation Grants CNS-0958379, CNS-0855217, ACI-1126113 and the City University of New York High Performance Computing Center at the College of Staten Island. We are thankful for comments and suggestions received at EEA 2020, WEAI 2020, SEA 2020, Stata Conference 2020, and internal seminars at the Graduate Center.

†CUNY Graduate Center; email: emunozsaavedra@gc.cuny.edu.

‡CUNY Graduate Center; email: WVijverberg@gc.cuny.edu.

1 Introduction

Moulton (1990) shows the consequences of ignoring cluster structures or cluster heterogeneities: inefficient parameter estimates, incorrect standard errors, and erroneous statistical inference.

Petersen (2009), considering typical finance panel data sets of firms over time, provide intuition as to why the different approaches to compute standard errors sometimes give different answers, and gives researchers guidance for their use. He provides a description of the different approaches used in papers in finance published in top journals between 2001-2004. These methods include Fama-MacBeth, which was developed to account for the correlation between observations on different firms in the same year, not to account for the correlation between observations on the same firm in different years. Newey and West (1987)'s estimator, designed to account for a serial correlation of unknown form in the residuals within cluster. White (1980)' standard errors robust to heteroskedasticity. Standard errors clustered either for firm, year (see Arellano (1987); Liang and Zeger (1986); Rogers (1993); White (1984)) or both following the method proposed by Thompson (2011), which is unbiased as long as there are enough number of firms and years.

Adjusting the standard errors either for firm or time is not valid when the disturbances are correlated across both firms and time. On the other hand, Thompson (2011) shows that when sample sizes are small, the more robust standard errors lead us to find statistical significance even when it does not exist. He argues that double-clustering is likely to be most helpful in data sets with the following characteristics: the regression errors include significant time and firm components, the regressors themselves include significant firm and time components, and the number of firms and time is not too different. In cases where N or T is small, double-clustering may cause more harm than good because the variance of the estimated standard error increases, and we reject the null hypothesis using a t-test more often.

In this paper we analyze the performance of statistical inference regarding the parameters of a panel data model when this is first subjected to a diagnostic for the presence of individual and/or time unobserved cluster effects. Using simulations we compare the performance of six different proposed pretests that are based on combinations of simple Lagrange multiplier, Likelihood ratio, and F tests for each or both unobserved cluster effects. Our results suggest that pre-testing for the presence of unobserved cluster effects is a viable alternative to estimate panel data models when the true model is unknown, in the sense that they achieve empirical sizes very close to the ones obtained using an estimator of the variance as if we knew the true data generating process. We do not find significant differences between the alternatives, which suggests that researchers could pick the simplest diagnostic in terms of computation. As robustness check, we find that these results hold when the cluster effects are temporary instead of fixed, when the disturbance is non-normally distributed, and when the regressor is non-normally distributed.

This paper adds to the literature of pretesting in econometrics. Previous studies that have evaluated pretest estimators in the context of panel data include Ziemer and Wetzstein (1983) on the pooling problem, Baltagi and Li (1997) on the estimation of error component models with autocorrelated disturbances, Baltagi, Bresson, and Pirotte (2003b) on a possibly misspecified two-way error component model and Guggenberger (2010) together with Kabaila, Mainzer, and Farchione (2015) on the Hausman test used as a pretest of the random effects specification. In this literature, the paper that it is most closely related to ours is Baltagi, Bresson, and Pirotte (2003a), which evaluates a pretest estimator in terms of MSE performance with a focus on model misspecification.

In addition, it adds to the recent literature about how and when to cluster standard errors in economics. Examples of recent surveys are Cameron and Miller (2015) and Mackinnon (2018). Bertrand, Duflo, and Mullainathan (2004) show how inference with difference in difference estimators is affected by serial correlation of the outcomes influencing importantly the practice of statistical inference in this context. More recently, Abadie, Athey, Imbens, and Wooldridge (2017) argue that clustering is a sampling design issue (e.g. the sample follows a two-stage sampling process where we randomly draw a subset of clusters from a population of clusters and then we randomly select units from those previously selected clusters) or an experimental design issue (e.g. treatment assignment is correlated within clusters). In contrast, Mackinnon and Webb (2019) argue that the previous conclusions depend critically on the assumption that the sample is large relative to a finite population in which the researcher is interested. However, under a meta-population framework, i.e. a view where researchers are interested not in the actual population but in a meta-population from which they imagine the former was drawn, the finite-population arguments of Abadie et al. (2017) do not apply and cluster-robust inference can be done in the usual way. Our paper fits within the meta-population setting, in the sense that we assume that the potential clusters were drawn from a meta-population of clusters.

We also add to the finance literature by taking one step further in the analysis of Petersen (2009) by suggesting a pretest estimator to compute the standard errors and showing that it is feasible to do inference without distorting the size. This paper also relates to Thompson (2011), who provided a formula to compute two-way clustered standard errors in finance panel data sets.

This paper is organized as follows. Section 2 describes the approaches to estimate the standard errors in finance panel data sets. Section 3 describes the pretest estimator, the hypothesis to be tested, and a brief description of the diagnostic tests considered. Section 4 describes the Monte Carlo set up and reports the simulation results. Section 5 provides an empirical example to show the relevance of our results. Some final remarks are then provided in Section 6.

2 Estimating the Variance

Consider the standard regression for a balanced panel data set:

$$y_{it} = X_{it}\beta + \epsilon_{it} \qquad i = 1, ..., N \quad t = 1, ..., T \tag{1}$$

where there are observations on entities i across years t. Stacking the y_{it}, X_{it} and ϵ_{it} over the NT observations then yields the model:

$$y = X\beta + \epsilon \tag{2}$$

The $NT \times K$ matrix X and the NT-dimensional vector ϵ are assumed to be independent of each other, and to have a zero mean and finite variance.

The OLS estimator for the K-dimensional vector β is:

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\epsilon \tag{3}$$

In general, the variance matrix conditional on X is:

$$V[\hat{\beta}] = (X'X)^{-1}B(X'X)^{-1} \tag{4}$$

with

$$B = X'V[\epsilon|X]X \tag{5}$$

Under the assumption of homoskedasticity and zero correlation of any kind (over time or across entities) in the disturbances ($V[\epsilon|X] = \sigma^2 I_{NT}$) we have:

$$B = \sigma^2 X'X \tag{6}$$

and the variance can be estimated using:

$$\hat{V}[\hat{\beta}] = \frac{e'e}{NT - K}(X'X)^{-1} \tag{7}$$

where $e = y - X\hat{\beta}$. This is the standard OLS formula that is correct when the disturbances are independent and identically distributed (Greene (2012)).

Relaxing the homoskedasticity assumption, allowing instead $V[\epsilon|X] = \sigma^2\Omega$ where Ω is a diagonal $NT \times NT$ matrix, we have that B becomes:

$$B = \sigma^2 X'\Omega X \tag{8}$$

which can be estimated using the White heteroskedasticity consistent estimator:

$$\hat{V}_{HC}[\hat{\beta}] = \frac{1}{NT}(\frac{1}{NT}X'X)^{-1}(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2 X_{it}X_{it}')(\frac{1}{NT}X'X)^{-1} \tag{9}$$

2.1 Cluster robust variance estimator (One-Way clustering)

By relaxing the assumption of independent disturbances, we can first assume that the data have an unobserved cluster effect by entity that is fixed. We can stack all observations in the i^{th} cluster and Model 1 can be written as:

$$y_i = X_i\beta + \epsilon_i \qquad i = 1, ..., N \tag{10}$$

where y_i and ϵ_i are $N_i \times 1$ vectors, X_i is an $N_i \times K$ matrix, and there are N_i observations in cluster i. Further stacking y_i, X_i and ϵ_i over the N clusters then yields the model:

$$y = X\beta + \epsilon \tag{11}$$

The OLS estimator now is:

$$\hat{\beta} = (X'X)^{-1}X'y = (\sum_{i=1}^{N} X_i'X_i)^{-1}\sum_{i=1}^{N} X_i'y_i \tag{12}$$

The variance matrix conditional on X is:

$$V[\hat{\beta}] = (X'X)^{-1}B(X'X)^{-1} \tag{13}$$

with

$$B = X'V[\epsilon|X]X \tag{14}$$

Given the error independence across clusters, $V[\epsilon|X]$ has a block-diagonal structure, and this equation simplifies to:

$$B_{CR} = \sum_{i=1}^{N} X_i' E[\epsilon_i \epsilon_i' | X_i] X_i \tag{15}$$

The cluster-robust estimator of the variance matrix (CRVE) of the OLS estimator is the sandwich estimator:

$$\hat{V}_{CR_i}[\hat{\beta}] = (X'X)^{-1} \hat{B}_{CR_i} (X'X)^{-1} \tag{16}$$

where

$$\hat{B}_{CR_i} = \sum_{i=1}^{N} X_i' e_i e_i' X_i \tag{17}$$

where X_i and e_i are T stacked observations of X_{it} and e_{it} in the i^{th} cluster. This accounts for arbitrary forms of serial correlation for each entity over time.[1]

Similarly, if we assume that the data have an unobserved cluster effect by time, the cluster-robust variance estimator becomes:

$$\hat{V}_{CR_t}[\hat{\beta}] = (X'X)^{-1} (\sum_{t=1}^{T} X_t' e_t e_t' X_t)(X'X)^{-1} \tag{18}$$

where X_t and e_t are N stacked observations of X_{it} and e_{it} in the t^{th} cluster. This accounts for arbitrary forms of correlation among the entities at any given point in time.

Finite sample modifications of Equation (16) and (18) are typically used to reduce downward bias of the estimator due to a finite numbers of clusters (Cameron and Miller (2015)). In this paper, we use the finite-sample adjustment $\frac{G}{G-1}\frac{NT-1}{NT-K}$, where we substitute G by N or T respectively in the case of clustered by entity or clustered by time.

Inference can be done on the basis of critical values drawn from a standard normal or alternatively a t distribution with *G-1* degrees of freedom as first suggested by Bester,

[1]This formula is attributed to White (1984), Liang and Zeger (1986), and Arellano (1987).

Conley, and Hansen (2011), which is the approach we follow.

In addition, we could use the Fama-MacBeth approach (see Fama and Macbeth (1973)) for the case of arbitrary correlation across entities and the Newey-West approach for the case of autocorrelation over time. Both methods are also used in Finance literature as reported in Petersen (2009). However, we choose to focus on the already described approaches to save space and because they are the methods more heavily used in the recent literature.

2.2 Cluster robust variance estimator (Two-Way clustering)

Cameron, Gelbach, and Miller (2011) and Thompson (2011) provide a variance estimator that enables cluster-robust inference when there is two-way or multi-way clustering that is non nested:

$$\hat{V}_{CR_{it}}[\hat{\beta}] = \hat{V}_{CR_i}[\hat{\beta}] + \hat{V}_{CR_t}[\hat{\beta}] - \hat{V}_{CR_{i\cap t}}[\hat{\beta}] \tag{19}$$

where each variance component is adjusted with the finite-sample factor previously mentioned.[2]

In this case, inference could be based on the OLS slope coefficient with critical values based on a standard normal distribution as in Petersen (2009) or alternatively we can use $t_{0.025;min(N,T)-1}$ critical value. We follow the latter approach.

3 Pretest of Unobserved Cluster Effects

3.1 Pretest estimator of the variance

Suppose we have the following model:

$$y_{it} = x'_{it}\beta + \epsilon_{it} \qquad i = 1, ..., N \quad t = 1, ..., T \tag{20}$$

[2]Alternatively, we can substitute G by $Min(N, T)$ and use the same factor in the three components (Cameron et al., 2011).

where

$$\epsilon_{it} = \gamma_i + \delta_t + \eta_{it} \tag{21}$$

and

$$x_{it} = \mu_i + \zeta_t + \nu_{it} \tag{22}$$

where the components γ_i, δ_t, μ_i and ζ_t are mean zero random variables with variances σ^2_γ, σ^2_δ, σ^2_μ and σ^2_ζ, respectively.

A pretest estimator for the variance of β is given by:

$$\hat{V}_{pt}[\hat{\beta}] = D_I \hat{V}_{HC}[\hat{\beta}] + D_{II} \hat{V}_{CR_i}[\hat{\beta}] + D_{III} \hat{V}_{CR_t}[\hat{\beta}] + D_{IV} \hat{V}_{CR_{it}}[\hat{\beta}] \tag{23}$$

where $\hat{V}_{HC}$, $\hat{V}_{CR_i}$, $\hat{V}_{CR_t}$, and $\hat{V}_{CR_{it}}$ corresponds to the variance estimators defined in the previous section; and D_i is an indicator variable equal to 1 when the outcome of a particular diagnostic favors the estimation of the variance according to case i, with the cases defined as Case I: $\sigma^2_\gamma = 0$ and $\sigma^2_\delta = 0$; Case II: $\sigma^2_\gamma = 0$ and $\sigma^2_\delta > 0$; Case III: $\sigma^2_\gamma > 0$ and $\sigma^2_\delta = 0$; and Case IV: $\sigma^2_\gamma > 0$ and $\sigma^2_\delta > 0$.

3.2 Pretesting for the presence of the unobserved cluster effects

To decide between the four cases, we evaluate six pretest alternatives, which can be represented with a decision tree as it is done in Figures 1, 2, and 3. They consist of a sequential set of statistical tests, in which each test focuses on a particular component with different assumptions regarding the remaining component. At each node of the pretest we evaluate the existence of a particular cluster effect or both, denote $H_0^{\gamma=0|\delta\geq 0}$ as a shorthand notation of the null hypothesis $H_0 : \sigma^2_\gamma = 0$ (allowing $\sigma^2_\delta \geq 0$); $H_0^{\gamma=0|\delta=0}$ for the null hypothesis $H_0 : \sigma^2_\gamma = 0$ (assuming $\sigma^2_\delta = 0$); $H_0^{\delta=0|\gamma\geq 0}$ for the null hypothesis $H_0 : \sigma^2_\delta = 0$ (allowing $\sigma^2_\gamma \geq 0$); $H_0^{\delta=0|\gamma=0}$ for the null hypothesis $H_0 : \sigma^2_\delta = 0$ (assuming $\sigma^2_\gamma = 0$); and $H_0^{\delta=0\&\gamma=0}$ for the null hypothesis $H_0 : \sigma^2_\delta = 0$ and $\sigma^2_\gamma = 0$. In the next subsection we describe the tests applied for each node, which corresponds to particular forms of the Lagrange multiplier, Likelihood ratio, and F tests.

These six pretests can be classified into three groups In the first two alternatives (Figure 1), we start testing one component assuming the absence of the remaining component and then continue testing one by one assuming the outcome of the test in the previous node. In the second group (Figure 2) we start assuming the existence of both errors components and then continue testing one by one assuming the outcome according to the test done in the previous node. Finally, in the third group (Figure 3 option e) starts assuming the existence of the error components but it does not impose the outcome of the previous node into the next one, which makes the initial cluster being tested irrelevant (i.e. whether you first test firm cluster effects or time cluster effects). In the case of option f, we start with a test with the null hypothesis of both effects being non-existent as done in Baltagi et al. (2003a).

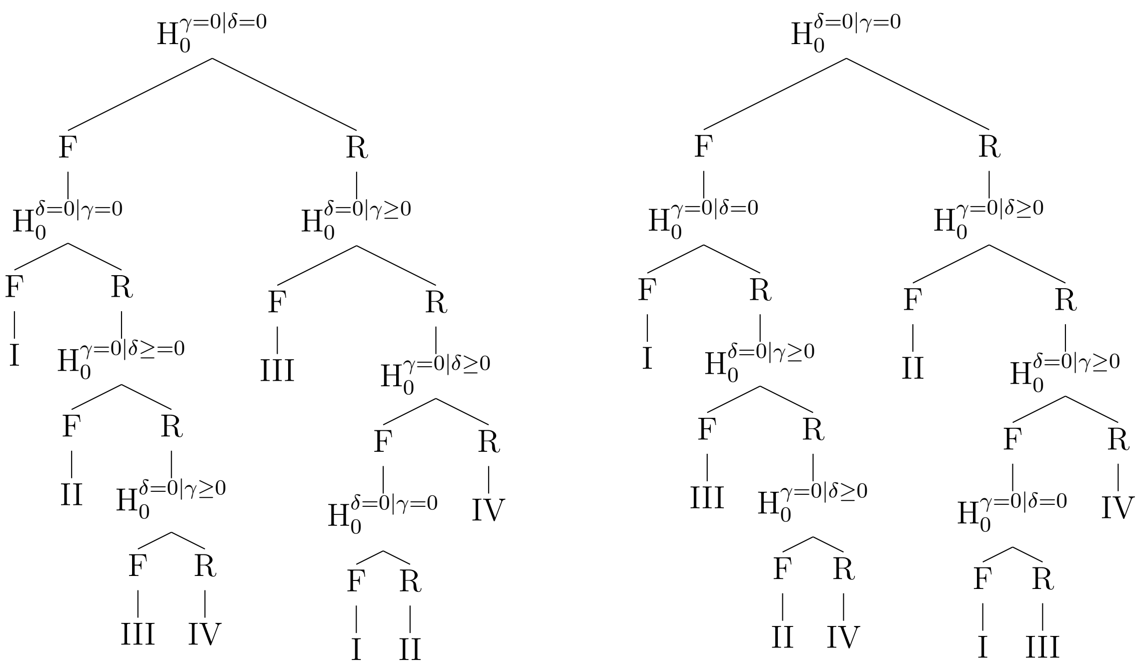
To test the null hypothesis that corresponds to each node in the diagrams, we use a set of tests available in the literature that we describe in the next subsections. They are part of three different types of tests: the Lagrange multiplier, Likelihood ratio, and F tests.

3.2.1 Lagrange multiplier tests

For the null hypotheses $H_0^{\gamma=0|\delta=0}$ and $H_0^{\delta=0|\gamma=0}$ we use a Standarized Lagrange Multiplier test given by:

$$SLM = \frac{(d - E[d])}{\sqrt{Var(d)}} \tag{24}$$

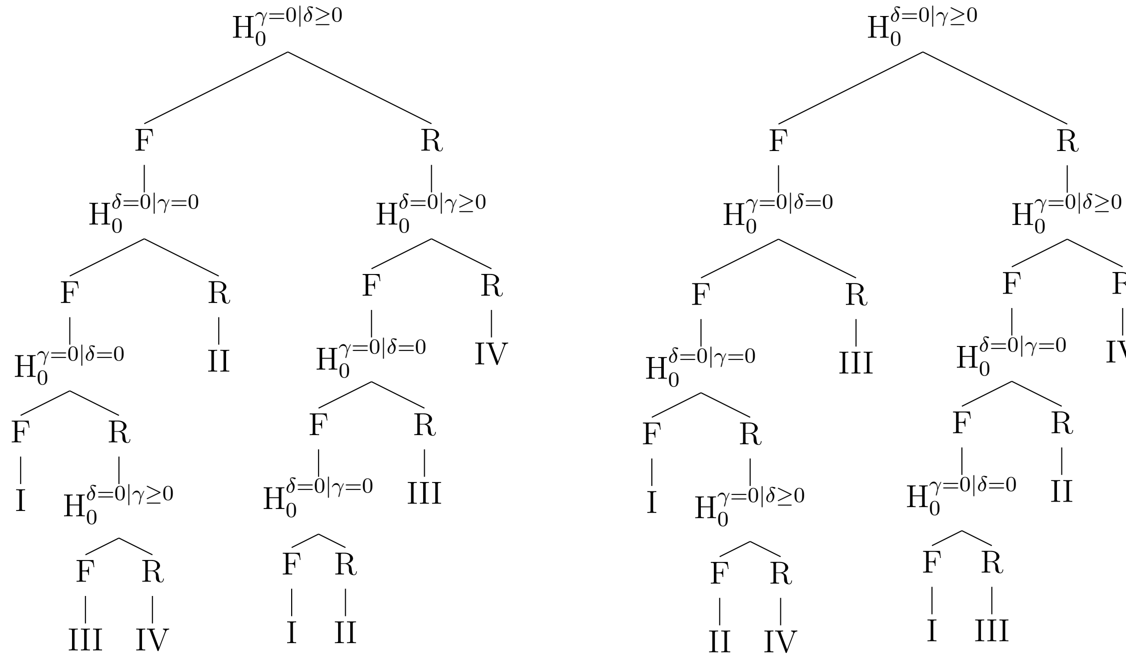
where $d = \sum_{g=1}^{G}(\sum_{i=1}^{N_g} \hat{u}_{gi})^2 / \sum_{g=1}^{G}\sum_{i=1}^{N_g} \hat{u}_{gi}^2$, $E[d] = tr(D'MD)/n$, $D = diag(\iota_g)$, ι_g is a column of ones of length N_g, $M = I - W(W'W)^{-1}W'$, $W = (XZ)$ with X and Z matrices that stack all observations, $n = N - K_x - K_z$, $Var(d) = 2\{ntr(D'MD)^2 - [tr(D'MD)]^2\}/n^2(n+2)$, and G the number of clusters of the component being tested. Under the null hypothesis, the limiting distribution of SLM is standard normal. This test is based on the work of Honda



Option a) Option b)

Notes:
Case I is $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 = 0$
Case II is $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 > 0$
Case III is $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 = 0$
Case IV is $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 > 0$
$H_0^{\gamma=0|\delta\geq0}$ is a test for the null $H_0 : \sigma_\gamma^2 = 0$ (allowing $\sigma_\delta^2 \geq 0$)
$H_0^{\gamma=0|\delta=0}$ is a test for the null $H_0 : \sigma_\gamma^2 = 0$ (assuming $\sigma_\delta^2 = 0$)
$H_0^{\delta=0|\gamma\geq0}$ is a test for the null $H_0 : \sigma_\delta^2 = 0$ (allowing $\sigma_\gamma^2 \geq 0$)
$H_0^{\delta=0|\gamma=0}$ is a test for the null $H_0 : \sigma_\delta^2 = 0$ (assuming $\sigma_\gamma^2 = 0$)
R denotes rejection of the null hypothesis and F denotes fail to reject

Figure 1: Diagram of the pre-test procedure, specific to general.



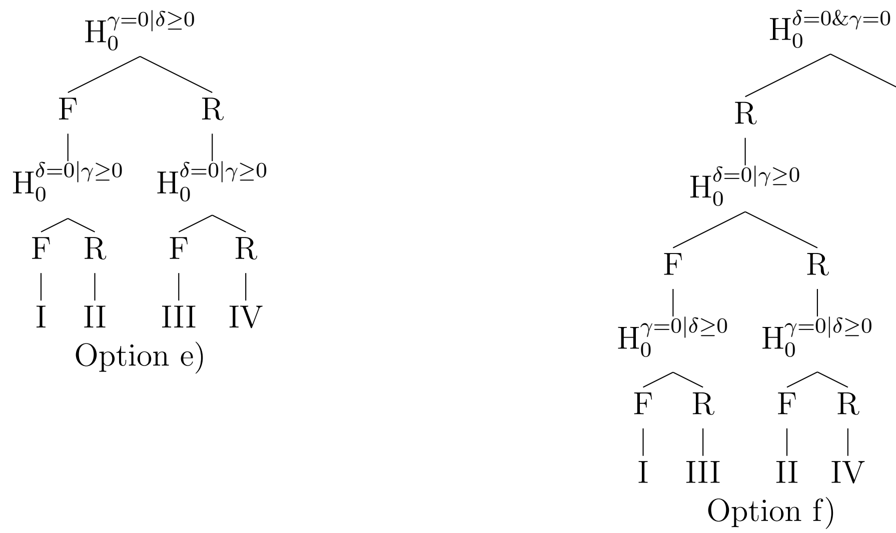
Option c) Option d)

Notes:
Case I is $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 = 0$
Case II is $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 > 0$
Case III is $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 = 0$
Case IV is $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 > 0$
$H_0^{\gamma=0|\delta\geq 0}$ is a test for the null $H_0 : \sigma_\gamma^2 = 0$ (allowing $\sigma_\delta^2 \geq 0$)
$H_0^{\gamma=0|\delta=0}$ is a test for the null $H_0 : \sigma_\gamma^2 = 0$ (assuming $\sigma_\delta^2 = 0$)
$H_0^{\delta=0|\gamma\geq 0}$ is a test for the null $H_0 : \sigma_\delta^2 = 0$ (allowing $\sigma_\gamma^2 \geq 0$)
$H_0^{\delta=0|\gamma=0}$ is a test for the null $H_0 : \sigma_\delta^2 = 0$ (assuming $\sigma_\gamma^2 = 0$)
R denotes rejection of the null hypothesis and F denotes fail to reject

Figure 2: Diagram of the pre-test procedure, general to specific.

(1985) and Moulton and Randolph (1989), and has been shown to dominate the Lagrange Multiplier test (proposed by Breusch and Pagan (1980) and modified for unbalanced clusters by Baltagi and Li (1990)) in Ma and Vijverberg (2010).

For the null hypotheses $H_0^{\gamma=0|\delta\geq 0}$ and $H_0^{\delta=0|\gamma\geq 0}$, we use the conditional LM test proposed by Baltagi, Chang, and Li (1992). They propose testing the individual effects conditional on the time-specific effects. The corresponding LM test for testing $H_0^{\gamma=0|\delta\geq 0} : \sigma_\gamma^2 = 0$ (allowing $\sigma_\delta^2 > 0$) is given by:



Notes:
Case I is $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 = 0$
Case II is $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 > 0$
Case III is $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 = 0$
Case IV is $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 > 0$
$H_0^{\gamma=0|\delta\geq 0}$ is a test for the null $H_0 : \sigma_\gamma^2 = 0$ (allowing $\sigma_\delta^2 \geq 0$)
$H_0^{\delta=0|\gamma\geq 0}$ is a test for the null $H_0 : \sigma_\delta^2 = 0$ (allowing $\sigma_\gamma^2 \geq 0$)
$H_0^{\delta=0\&\gamma=0}$ is a test for the null $H_0 : \sigma_\delta^2 = 0$ and $\sigma_\gamma^2 = 0$
R denotes rejection of the null hypothesis and F denotes fail to reject.

Figure 3: Diagram of the pre-test procedure.

$$LM_\gamma = \frac{\sqrt{2}\tilde{\sigma}_2^2\tilde{\sigma}_\eta^2}{\sqrt{T(T-1)[\tilde{\sigma}_\eta^4 + (N-1)\tilde{\sigma}_2^4]}}\tilde{D}_\gamma \tag{25}$$

where

$$\tilde{D}_\gamma = \frac{T}{2}\{\frac{1}{\tilde{\sigma}_2^2}[\frac{\tilde{u}'(\bar{J}_N \otimes \bar{J}_T)\tilde{u}}{\tilde{\sigma}_2^2} - 1] + \frac{(N-1)}{\tilde{\sigma}_\eta^2}[\frac{\tilde{u}'(E_N \otimes \bar{J}_T)\tilde{u}}{(N-1)\tilde{\sigma}_\eta^2} - 1]\} \tag{26}$$

with $\tilde{\sigma}_2^2 = \tilde{u}'(\bar{J}_N \otimes I_T)\tilde{u}/T$, $\tilde{\sigma}_\eta^2 = \tilde{u}'(E_N \otimes I_T)\tilde{u}/T(N-1)$, $\bar{J}_N$ is a $N \times N$ matrix of ones divided by N and E_N is defined as $I_N - \bar{J}_N$. This test statistic LM_γ is asymptotically distributed $N(0,1)$ under $H_0^{\gamma=0|\delta\geq 0}$. The estimated disturbances $\tilde{u}$ represent the one-way GLS residuals using the MLEs $\tilde{\sigma}_\eta^2$ and $\tilde{\sigma}_2^2$.

On the other hand, the LM statistic for testing $H_0^{\delta=0|\gamma\geq 0} : \sigma_\delta^2 = 0$ (allowing $\sigma_\gamma^2 > 0$) is given by:

$$LM_\delta = \frac{\sqrt{2}\tilde{\sigma}_1^2\tilde{\sigma}_\eta^2}{\sqrt{N(N-1)[\tilde{\sigma}_\eta^4 + (T-1)\tilde{\sigma}_1^4]}}\tilde{D}_\delta \tag{27}$$

where

$$\tilde{D}_\delta = \frac{N}{2}\{\frac{1}{\tilde{\sigma}_1^2}[\frac{\tilde{u}'(\bar{J}_N \otimes \bar{J}_T)\tilde{u}}{\tilde{\sigma}_1^2} - 1] + \frac{(T-1)}{\tilde{\sigma}_\eta^2}[\frac{\tilde{u}'(\bar{J}_N \otimes E_T)\tilde{u}}{(T-1)\tilde{\sigma}_\eta^2} - 1]\} \tag{28}$$

with $\tilde{\sigma}_1^2 = \tilde{u}'(I_N \otimes \bar{J}_T)\tilde{u}/N$ and $\tilde{\sigma}_\eta^2 = \tilde{u}'(I_N \otimes E_T)\tilde{u}/N(T-1)$. This test statistic LM_δ is asymptotically distributed as $N(0,1)$ under $H_0^{\delta=0|\gamma\geq 0}$.

3.2.2 Likelihood Ratio test

An one-sided likelihood ratio (LR) test has the form:

$$LR = -2(lnL(res) - lnL(unres)) \tag{29}$$

where $lnL(res)$ denotes log of the restricted maximum likelihood value (under the null hypothesis) and $lnL(unres)$ denotes log of the unrestricted maximum likelihood value. Under the null $H_0^{\gamma=0|\delta=0}$ and $H_0^{\delta=0|\gamma=0}$, the restricted model corresponds to a pooled OLS, while the unrestricted model corresponds to a one-way individual random effects and one-way time random effects respectively, and $LR \sim (\frac{1}{2})\chi^2(0) + (\frac{1}{2})\chi^2(1)$. Under the null hypotheses $H_0^{\gamma=0|\delta\geq 0}$ and $H_0^{\delta=0|\gamma\geq 0}$, the unrestricted model corresponds to a two-way random effects model, and the restricted is a one-way entity random effects model and a one-way time random effects model respectively, and $LR \sim (\frac{1}{2})\chi^2(0) + (\frac{1}{2})\chi^2(1)$ in both cases (Baltagi (2013)).

In contrast to the Lagrange multiplier tests, as can be inferred from equation 29, the Likelihood ratio test requires the estimation of both the model under the null hypothesis and the model under the alternative hypothesis. For some particular null hypothesis, this translates to the estimation of a two-way random effects model and a one-way random effects model using maximum likelihood.

3.2.3 F test

As described in Greene (2012), an F statistic for the two-way error component model has the general form:

$$F = \frac{(R^2 - R_*^2)/J}{(1 - R^2)/(n - K)} \tag{30}$$

Under the null hypothesis, this statistic has a central F distribution with J (number of restrictions) and n-K degrees of freedom (number of observations minus number of parameters). For $H_0^{\gamma=0|\delta=0}$: $J = N - 1$, $n - K = NT - N - 1$, R^2 is computed estimating a model with entity fixed effects, and R_*^2 computed with pooled OLS model. For $H_0^{\delta=0|\gamma=0}$: $J = T - 1$, $n - K = NT - T - 1$, R^2 is computed estimating a model with time fixed effects, and R_*^2 computed with pooled OLS model. For $H_0^{\gamma=0|\delta\geq 0}$: $J = N - 1$, $n - K = NT - N - T$, R^2 is computed estimating a two-way fixed effects model, and R_*^2 computed with only time effects. For $H_0^{\delta=0|\gamma\geq 0}$: $J = T - 1$, $n - K = NT - N - T$, R^2 is computed estimating a two-way fixed effects model, and R_*^2 computed with only fixed effects.

Similar to the case of the Likelihood ratio, the F test requires the estimation of two models. However, the required least square estimator is usually simpler than the equivalent maximum likelihood estimator.

3.2.4 Gourieroux, Holly, and Monfort Test

Following Gourieroux, Holly, and Monfort (1982), from now on GHM, Baltagi et al. (1992) proposed the following test for the null hypothesis $H_0^{\gamma=0\&\delta=0}$:

$$\begin{aligned} \chi_m^2 &= \{A^2 + B^2 \quad if \quad A > 0, B > 0 \\ &= \{A^2 \quad if \quad A > 0, B \leq 0 \\ &= \{B^2 \quad if \quad A \leq 0, B > 0 \\ &= \{0 \quad if \quad A \leq 0, B \leq 0 \end{aligned} \tag{31}$$

where χ^2_m denotes the mixed χ^2 distribution. Under the null hypothesis, we have that:

$$\chi^2_m \sim .25\chi^2(0) + .5\chi^2(1) + .25\chi^2(2) \tag{32}$$

where $\chi^2(0)$ equals zero with probability one.

4 Monte Carlo Study

In this section we perform an extensive set of Monte Carlo experiments. The objectives of these experiments are to evaluate the performance in terms of empirical size of doing statistical inference with panel data after applying the previously discussed pretests. We compare rejection rates of a true null hypothesis obtained with a variance estimator decided ex-ante (i.e. White heteroskedasticity robust, allowing clustering by entity, allowing clustering by time, and allowing two-way clustering) against the ones obtained by using a variance estimator dictated by a data-driven diagnostic (i.e. a pretest). We first present a set of baseline simulations, which addresses the possibility of having a data generating process with disturbances that are correlated across one dimension (either entity or time), correlation across both dimensions, and finally disturbances without any within cluster correlation. We then also analyze several different robustness checks, in which we vary particular aspects of the baseline set up.

4.1 Baseline Monte Carlo design

We simulate a panel data set similar to what it is used in Petersen (2009), pretest for entity (denoted by i and assumed to be a firm component) and/or time (denoted by t and assumed to be a year component) unobserved cluster effects, estimate the slope coefficient and its standard error according to the result of the pretest, and finally we perform a t-test on the slope β of the model with the null hypothesis of β equal to the true value set in the data generating process. We consider three cases derived from the following model (described previously in section 3) with just one exogenous regressor:

$$y_{it} = \beta x_{it} + \epsilon_{it} \tag{33}$$

where:

$$\epsilon_{it} = \gamma_i + \delta_t + \eta_{it}$$

and

$$x_{it} = \mu_i + \zeta_t + \nu_{it}$$

First, we consider the case where there are cluster effects only in one dimension. We consider a data set with 250 firms and 10 years. Given that our panel data set has a relatively small dimension (T) and a large one (N), we consider two scenarios: i) we generate a data set with a firm component and no time component (i.e. $\sigma^2_\gamma \geq 0$, $\sigma^2_\mu \geq 0$, $\sigma^2_\delta = 0$, and $\sigma^2_\zeta = 0$), which corresponds to having clusters by the largest dimension in terms of the number of clusters; and ii) we simulate a panel data set with a time component and no firm component (i.e. $\sigma^2_\gamma = 0$, $\sigma^2_\mu = 0$, $\sigma^2_\delta \geq 0$, and $\sigma^2_\zeta \geq 0$), which corresponds to having clusters by the smallest dimension in regards to the number of clusters.[3] Across simulations, the true slope β is equal to 1, the standard deviation of the independent variable x_{it} and the disturbance ϵ_{it} are both assumed to be constant at 1 and 2, respectively. We run different sets of simulations where we modify the fraction of the variance in the independent variable, and separately, in the disturbance that is due to the clusters. This fraction ranges from 0% to 75% in 25% increments, although we omit 0% in the disturbance as we analyze the case without cluster effects separately.

Second, a panel data set with both firm and time components (i.e. $\sigma^2_\gamma > 0$, $\sigma^2_\mu > 0$, $\sigma^2_\delta > 0$, and $\sigma^2_\zeta > 0$). We vary the number of firms N from 5 to 500 (5, 10, 20, 50, 125, 250, 500) keeping the total number of observations constant at $NT = 2,500$. We assume

[3] If we were to use a panel data set with equal number of entities and time periods, then the results of having clusters in either of those dimensions would be symmetrical.

that one-third of the variability of the disturbance and the independent variable is due to the firm effect and one-third of the variability is due to the time effect.

Third, a panel data similar to the previous one in terms of size, in which we assume that there are no firm or year effects (i.e. $\sigma_\delta^2 = 0$, $\sigma_\zeta^2 = 0$, $\sigma_\gamma^2 = 0$, and $\sigma_\mu^2 = 0$).

4.2 Results

Figure 4a reports the percentage of rejections of the null hypothesis $H_0 : \beta = 1$ over 5,000 simulations using the different approaches when the data generating process (DGP) contains only unobserved clusters by firm (the dimension with largest number of clusters).[4] The marker labeled *CR-F* (the circle) corresponds to standard errors computed allowing clustering by firm in every simulation, which in this case corresponds to the right approach if the DGP were known and delivers rejection rates close to the nominal size of the test (one percent). Using standard errors that allow two-way clustering (*CR-FY*) results in under and over rejection of the null hypothesis depending of the exercise although never too severe. Specifically, the empirical size shows over rejection when the variance explained by the firm clusters in the regressor is smaller relative to the variance explained by the firm clusters in the disturbance and under rejection when the opposite happens. Clustering by year (*CR-Y*) gives results similar to the right model when there is no cluster effect in the regressor X (see the bottom rows that start with zero in Figure 4a). However, when there is a cluster effect in the regressor, it over rejects the null hypothesis with percentages that increase when the share of variance in the disturbance explained by the cluster effects increase and reaches almost fourty percent in the last row, which contain the highest shares. Similarly, using White robust standard errors (*Rob*) give results that also remain similar to the right model when there is no cluster effect in the regressor but over rejects the null hypothesis when both the regressor and the disturbance contain cluster effects by firm. In the case of the pretest estimators, most of them are able to perform as well as the estimator that allows clustering

[4]Table 2 in the Appendix reports these results with more detail.

by the right dimension according to the true data generating process. The level of deviation is never more than 0.06 percentage points relative to the benchmark (*CR-F*).

To get some insights about the different performance between the pretest estimators, Table 6 in the Appendix reports the distribution of chosen models by each pretest in the different simulation set ups. Despite similar performance in terms of empirical size, the three different diagnostic tests (diagnostic that uses Lagrange multiplier tests, Likelihood ratio tests, or F tests) do not always choose models in the same proportions. For example, the pretest that uses Likelihood ratios tend to choose model 3 (model with an unobserved firm cluster) slightly more often than the alternatives.

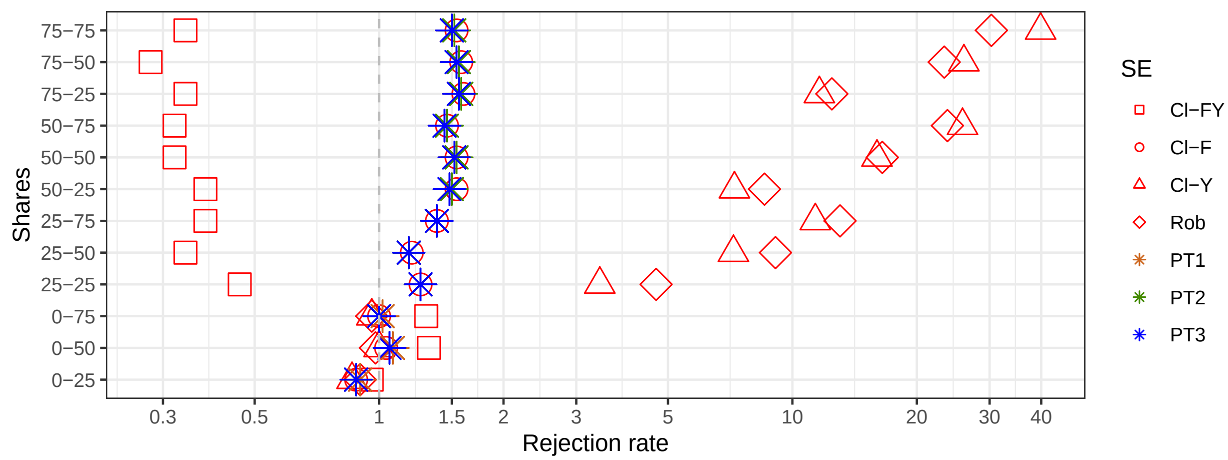
Figure 4b reports the percentage of rejections over 5,000 simulations using different approaches when the data generating process contains only unobserved clusters by year.[5] Given the small number of clusters (10 years), even clustering by year (see *CR-Y* in Figure 4b) produces over rejection of the null hypothesis that in one case goes even slightly above four percent. Two-way clustering by both firm and year performs very close to the case where we compute standard errors always clustering by year. Clustering standard errors by the wrong dimension (see *CR-F*) or using only heteroskedasticity-robust standard errors (see *Rob*) increases severely the percentage of over rejection, with some cases going even above eighty percent. The pretest estimators perform as well as clustering by year or both dimensions, reaching similar empirical sizes.

Table 7 in the Appendix reports the distribution of chosen models by each alternative pretest in the different simulation set ups. *PT-2* (which uses LR diagnostics) chooses the right model slightly more often than the alternatives *PT-1* and *PT-3* even though they achieve similar rates of rejection. This is because they tend to choose model 4 (clusters by time and entity) when they do not pick model 2 (the correct model according to the DGP), which does not severely distorts inference as we saw above (two-way clustering does not perform significantly worst than clustering by time).

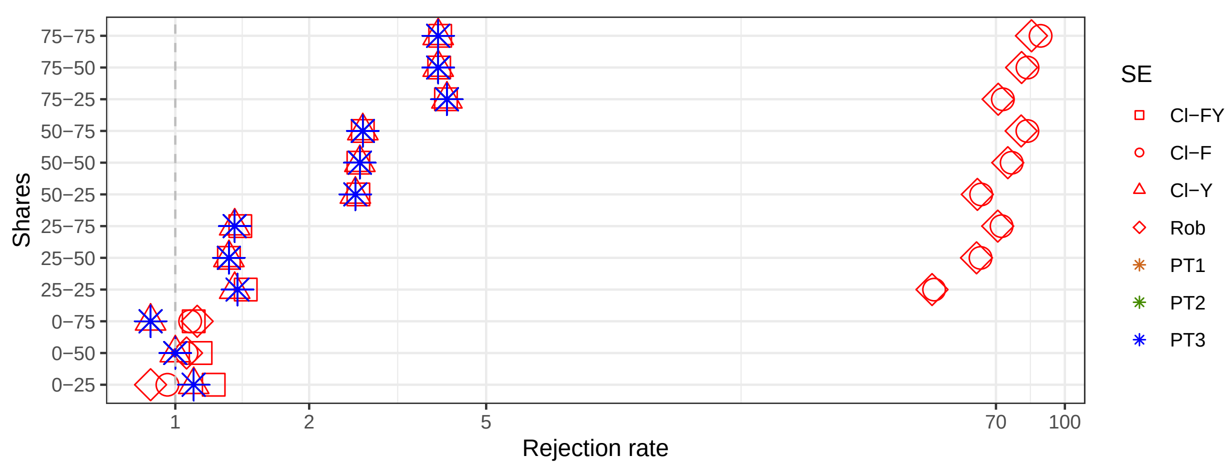
[5]Table 3 in the Appendix reports these results with more detail.

Figure 4: Empirical size at $\alpha = 0.01$, by fraction of variance explained by the cluster effects

(a) Panel data with cluster effects by firm (N=250)



(b) Panel data with cluster effects by time (T=10)



Notes:
The graph shows the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The y axis contains contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i in panel 4a and δ_t and ζ_t in panel 4b.
Cl-FY uses standard errors (s.e.) clustered by firm and year, *Cl-F* uses s.e. clustered by firm, *Cl-Y* uses s.e. clustered by year, and *Rob* uses heteroskedascity-robust s.e.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests, *PT-2* using LR tests, and *PT-3* using F tests.
To avoid cluttering the legend and considering that they provide very similar results, the six different implementations described in Figure 1, Figure 2, and Figure 3 are all plotted with a marker and color according to whether they use *PT-1*, *PT-2*, or *PT-3*.

Figure 5 reports the percentage of rejections over 5,000 simulations using the different approaches when the DGP contains both entity and time unobserved cluster effects.[6] Using White robust standard error (*Rob*) or clustering by just one of the dimensions (either *Cl-F* or *Cl-Y*) produces an over rejection of the null hypothesis higher than using two-way clustered standard errors (*Cl-FY*) in all the cases, although smaller when clustering by the smallest dimension in cases with just 5 or 10 clusters (i.e. clustering by firm with just 5 or 10 firms or clustering by year with just 5 or 10 years). The pre-test estimators perform as well as the two-way clustered standard errors in all the cases.

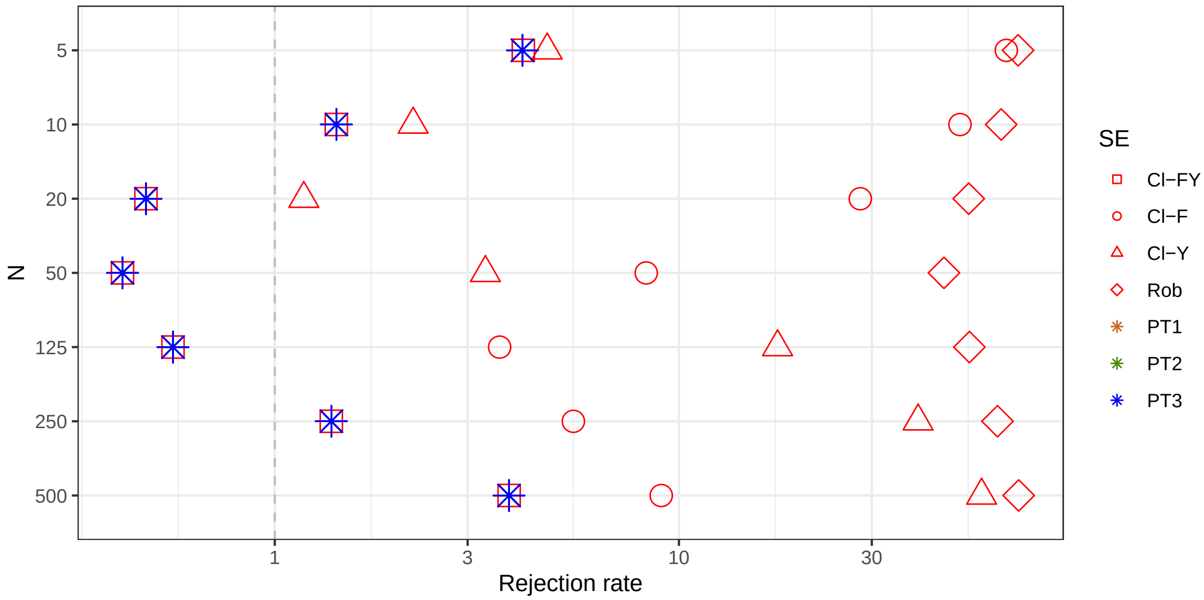
Table 8 in the Appendix reports the distribution of chosen models by each alternative pretest in the different set ups. In this case, the pretest always identify the right model with the exception of the simulation with 500 firms where Likelihood ratios identified the right model in 99.9% of the time.

Finally, Figure 6 reports the percentage of rejections over 5,000 simulations using the different approaches when the model does not contain any unobserved cluster effects.[7] In this case, clustering standard errors by both dimensions (*Cl-FY*) leads to over rejection in most of the cases but especially when one dimension is relatively small (see rows 5, 10, 250, 500). When the dimensions are more symmetric, then multi-way clustering leads to under rejection of the null hypothesis (see row 50) and rates of rejection slightly smaller than White robust standard errors (*Rob*). The pretest estimator performs as well as the White robust estimator in general with some deviations in the case of the pretest that uses Lagrange Multiplier tests (see *PT-1*). Clustering only by one dimension over/under rejects the null hypothesis depending on the number of clusters.

Table 9 in the Appendix reports the distribution of the models chosen by each pretest in the different cases. In general, the pretest estimators that use likelihood ratio diagnostics chooses the right model more often while the lagrange multiplier fails in favor of model 2

[6]Table 4 in the Appendix reports these results with more detail.
[7]Table 5 in the Appendix reports these results with more detail.

Figure 5: Empirical size at $\alpha = 0.01$, by number of firms with a DGP using two-way cluster effects

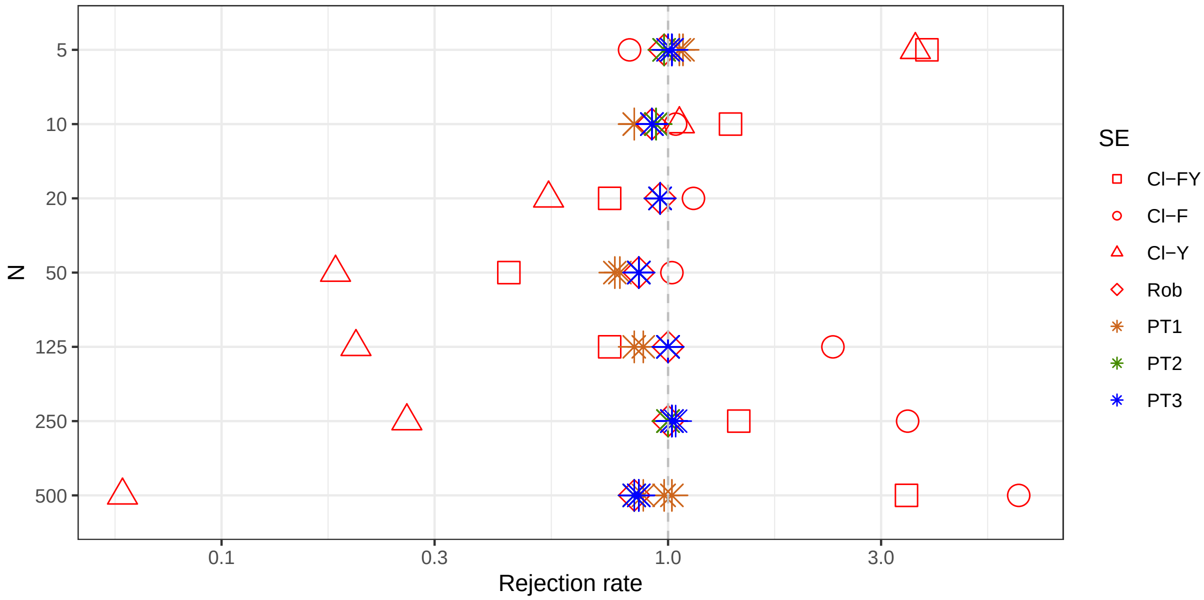


Notes:
The graph shows the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The y axis contains the number of firms (keeping N*T=2500).
Cl-FY uses standard errors (s.e.) clustered by firm and year, *Cl-F* uses s.e. clustered by firm, *Cl-Y* uses s.e. clustered by year, and *Rob* uses heteroskedascity-robust s.e.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests, *PT-2* using LR tests, and *PT-3* using F tests.
To avoid cluttering the legend and considering that they provided very similar results, the six different implementations described in Figure 1, Figure 2, and Figure 3 are all plotted with a marker and color according to whether they use *PT-1*, *PT-2*, or *PT-3*.

and 3 more often.

To sum up, the baseline simulations allow us to state the following results: i) In the presence of any cluster effect, the use of White heteroskedasticity robust standard errors leads to severe over rejection of the null hypothesis, ii) In the presence of cluster effects in one dimension, the use of standard errors clustered by the wrong dimension can lead to substantial over rejection of the null hypothesis when the regressor also contain cluster effects, iii) When the DGP contain two-way clustering of disturbances, standard errors that do not accomodate this correlation tend to over reject the null hypothesis, iv) If the DGP does not have any kind of within cluster disturbances correlation, allowing for clustering of

Figure 6: Empirical size at $\alpha = 0.01$, by number of firms



Notes:
The graph shows the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The y axis contains the number of firms (keeping N*T=2500).
Cl-FY uses standard errors (s.e.) clustered by firm and year, *Cl-F* uses s.e. clustered by firm, *Cl-Y* uses s.e. clustered by year, and *Rob* uses heteroskedascity-robust s.e.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests, *PT-2* using LR tests, and *PT-3* using F tests.
To avoid cluttering the legend and considering that they provided very similar results, the six different implementations described in Figure 1, Figure 2, and Figure 3 are all plotted with a marker and color according to whether they use *PT-1*, *PT-2*, or *PT-3*.

any type distorts the rejection rates although not severely, v) Standard errors that allow for two-way clustering tend to perform relatively well in this baseline set up, vi) Standard errors that follow a pretesting procedure tend to perform as well as the preferred method if we knew the true DGP, and vii) All the options of pretest give similar results, although the use of Likelihood ratio tend to choose the right model slightly more often.

4.3 Robustness checks

In this section we consider alternative scenarios in the Monte Carlo design to assess the robustness of our results.

4.3.1 Non-normal disturbances

Our results are robust to non-normality of the disturbance. We run our main simulations (except for the experiment without cluster effects) assuming that the cluster effects are drawn from a chi-square distribution while the remainder component is normally distributed. For this exercise we keep the size of the standard deviation for the regressor and disturbance in 1 and 2 respectively, as it is done in our main simulations. Figure 7 reports the percentage of rejections of the null hypothesis over 5,000 simulations using different approaches when the GDP contains only unobserved clusters by firm in 7a, and only clusters by time in 7b.[8] We find qualitatively similar results, in general the pretest estimators achieve similar empirical size than using a cluster-robust variance estimator according to the true data generating process. However, relative to our baseline set up with the existence of cluster effects by firm (see Figure 4a), the deviation in the rejection rate when using two-way clustered standard errors is more pronounced, and the pretesting approach closely matches the rejection rate achieved by the benchmark around 1%.

4.3.2 Non-normal regressor

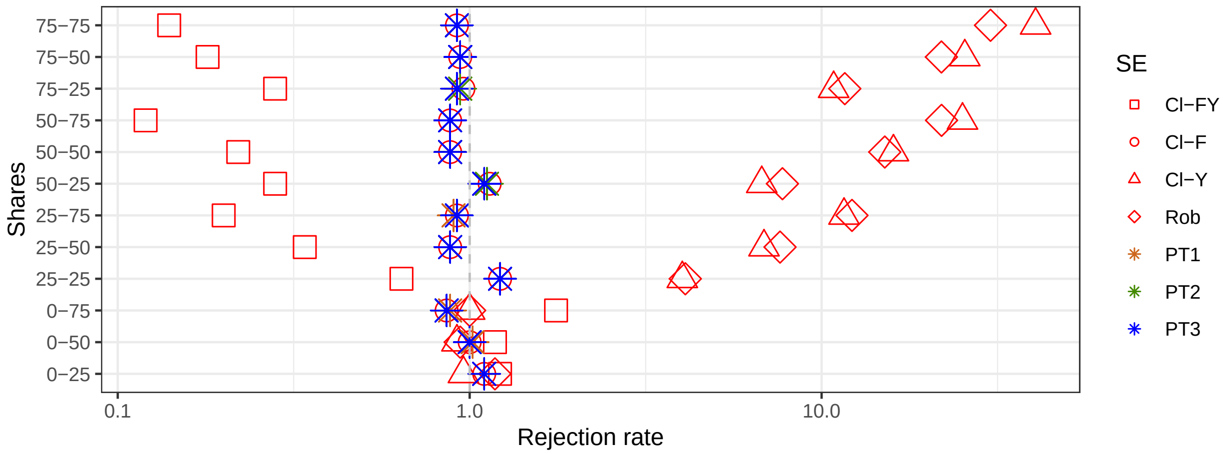
Our results are also robust to the use of an exogenous regressor X that is distributed log-normal. We run the same set of simulations as in our main exercises using the exponential of the exogenous regressor X. However, we set the variance of the regressor to be 0.5 when there is one cluster component and 0.48 when there are clusters in both dimensions. This set up is similar to a set of simulations done in Mackinnon, Nielsen, and Webb (2019) in the context of estimation of standard errors using Wild cluster bootstrap.

The results are reported in the Appendix (see Table 17, 18, 19 and 20). We find qualitatively similar results, the pretest estimators achieve similar empirical size than using a cluster-robust variance estimator according to the true data generating process.

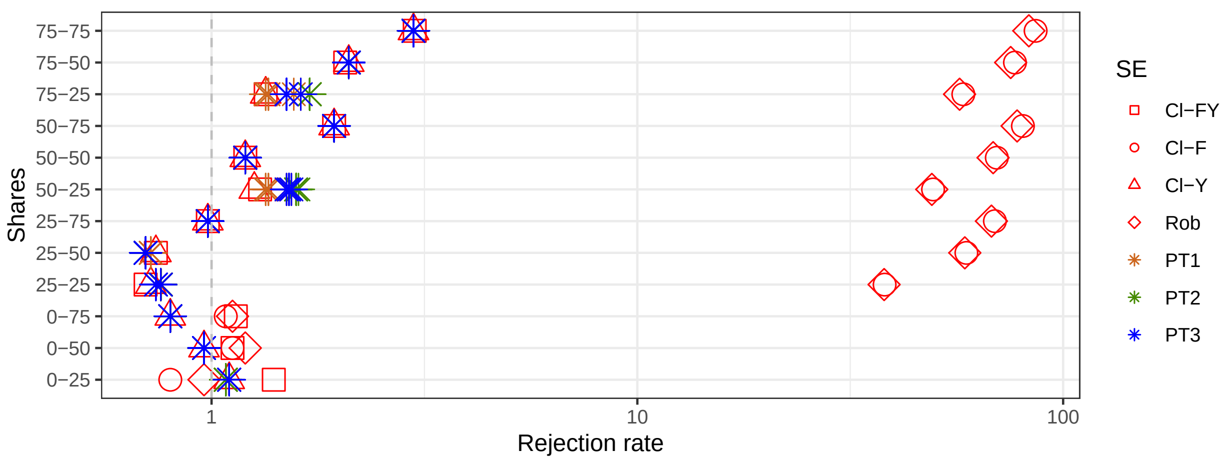
[8]Detailed tables are reported in the Appendix (see Table 14, 15, and 16)

Figure 7: Empirical size at $\alpha = 0.01$, by fraction of variance explained by the cluster effects

(a) Panel data with cluster effects by firm (N=250)



(b) Panel data with cluster effects by time (T=10)



Notes:
The graph shows the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The y axis contains contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i in panel 7a and δ_t and ζ_t in panel 7b.
Cl-FY uses standard errors (s.e.) clustered by firm and year, *Cl-F* uses s.e. clustered by firm, *Cl-Y* uses s.e. clustered by year, and *Rob* uses heteroskedascity-robust s.e.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests, *PT-2* using LR tests, and *PT-3* using F tests.
To avoid cluttering the legend and considering that they provide very similar results, the six different implementations described in Figure 1, Figure 2, and Figure 3 are all plotted with a marker and color according to whether they use *PT-1*, *PT-2*, or *PT-3*.

Similar to the case of non-normal disturbances reviewed in the previous subsection, we find a more pronounced deviation in the rejection rate (relative to the baseline set up) when there are cluster effects by firm (see Figure 6).

4.3.3 The case of temporary firm effects

The pretest performance depends on the capacity of the individual tests to identify the unobserved cluster effects. This may be more difficult when the error components are not fixed and they vanish over time. In this set of simulations, we explore the performance of the pretest estimator in this context.

We follow the data structure of Petersen (2009) with a simulation that include both a permanent component (a fixed firm effect) and a temporary component (nonfixed firm effect) that is assumed to be a first-order autoregressive process. We construct the nonfirm effect share of the disturbance (η_{it}) as:

$$\begin{aligned} \eta_{it} &= \xi_{it} \quad if \quad t=1 \\ &= \phi\eta_{it-1} + \sqrt{1-\phi^2}\xi_{it} \quad if \quad t>1 \end{aligned} \tag{34}$$

where ϕ is the first-order auto correlation between η_{it} and η_{it-1}, and the correlation between η_{it} and η_{it-k} is ϕ^k.

We construct the independent variable with the same logic using:

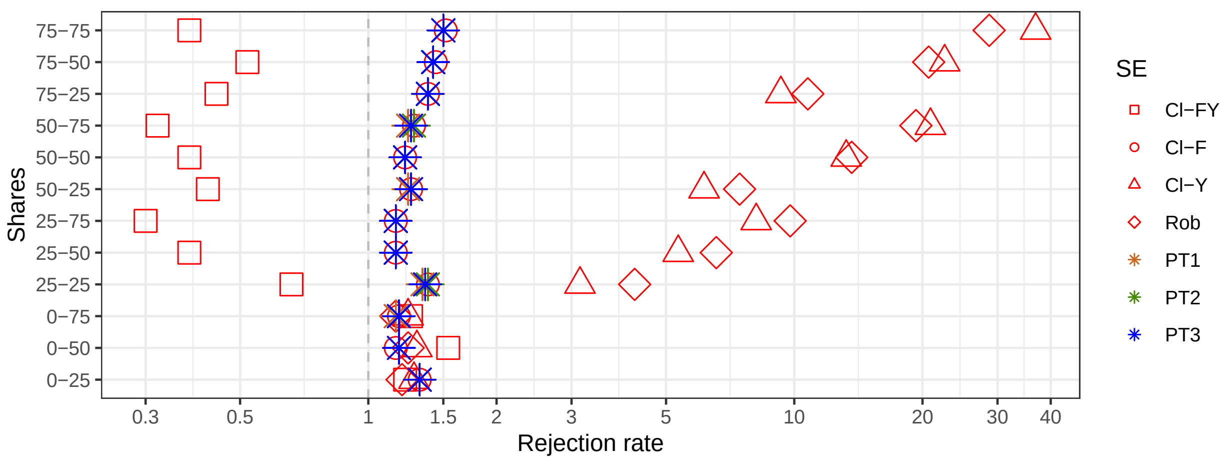
$$\begin{aligned} \nu_{it} &= \kappa_{it} \quad if \quad t=1 \\ &= \phi\nu_{it-1} + \sqrt{1-\phi^2}\kappa_{it} \quad if \quad t>1 \end{aligned} \tag{35}$$

where ξ_{it} and κ_{it} are both uncorrelated.

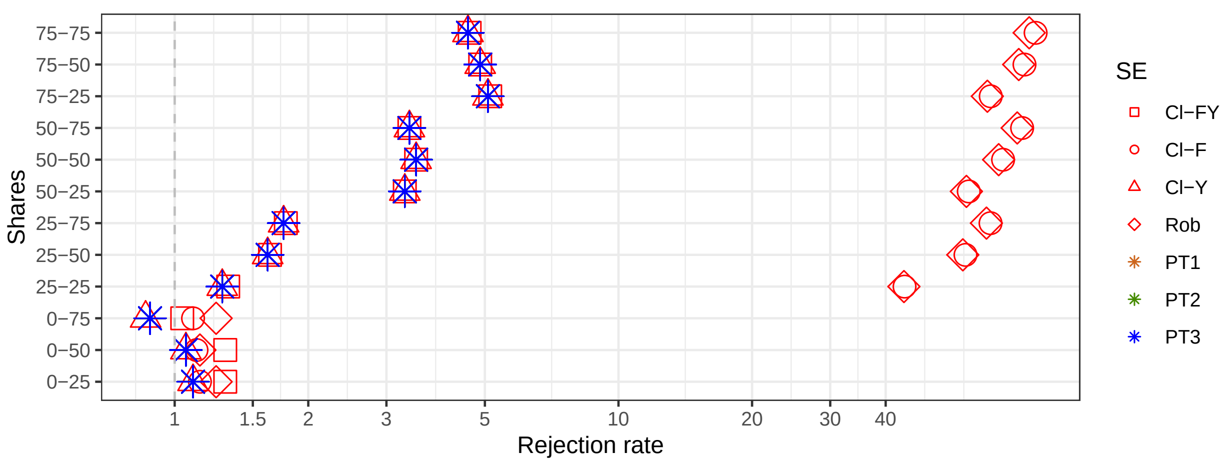
We generate a panel data set with 250 firms and 10 years in which we try 3 different combinations of ϕ, and the share of the variance of ϵ_{it} and x_{it} explained by the fixed firm component.

Figure 8: Empirical size at $\alpha = 0.01$, by fraction of variance explained by the cluster effects

(a) Panel data with cluster effects by firm (N=250)



(b) Panel data with cluster effects by time (T=10)



Notes:
The graph shows the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The y axis contains contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i in panel 8a and δ_t and ζ_t in panel 8b.
Cl-FY uses standard errors (s.e.) clustered by firm and year, *Cl-F* uses s.e. clustered by firm, *Cl-Y* uses s.e. clustered by year, and *Rob* uses heteroskedascity-robust s.e.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests, *PT-2* using LR tests, and *PT-3* using F tests.
To avoid cluttering the legend and considering that they provide very similar results, the six different implementations described in Figure 1, Figure 2, and Figure 3 are all plotted with a marker and color according to whether they use *PT-1*, *PT-2*, or *PT-3*.

Figure 9 reports the results of the simulations.[9] We find that our results are still valid in the case of a temporary firm effect.

Applying two-way clustered standard errors when there is only a temporary cluster effect by firm produces under rejection the null hypothesis. On the contrary, clustering by year or using White robust standard errors generates rejection rates that can go close to 40% in one of the set ups. However, the pretest estimators achieves rejection rates that in some cases are even closer to the nominal size than when we compute the standard errors clustering by firm.

When we consider a time cluster effect plus a temporary cluster effect by firm, we get some under rejection of the null hypothesis with the two-way clustered standard errors, which is followed by the pretest alternatives. With all the other approaches we get important over rejection.

5 An Empirical Application in Finance

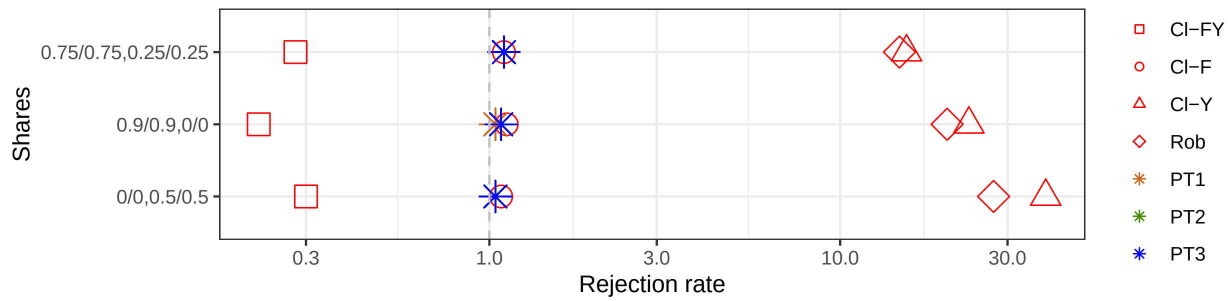
In this section we analyze an empirical example of the impact of different approaches to compute the standard errors and the pretest application. We replicate a panel data regression that has been studied in Moskowitz, Ooi, and Pedersen (2012), and recently revisited in Huang, Li, Wang, and Zhou (2020).[10] The latter paper assesses the existence of "time series momentum" (i.e. whether past returns predict future returns) in equity index, currency, commodity, and bond futures using 55 liquid instruments for a sample period from January 1985 to December 2015. Its empirical analysis uses univariate time series regression, and pooled regression, which it is the approach we follow.

[9]Table 21 and 22 in the Appendix contain detailed results of the simulations.

[10]The data is available at http://jfe.rochester.edu/Huang_Li_Wang_Zhou_data.zip.

Figure 9: Empirical size at $\alpha = 0.01$, by fraction of variance explained by the cluster effects

(a) DGP with temporary cluster effects by firm



(b) DGP with temporary cluster effects by firm and cluster effects by time

Notes:
The graph shows the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The y axis contains the values of ϕ and the share of the variance of ϵ_{it} and x_{it} explained by the cluster components.
Cl-FY uses standard errors (s.e.) clustered by firm and year, *Cl-F* uses s.e. clustered by firm, *Cl-Y* uses s.e. clustered by year, and *Rob* uses heteroskedascity-robust s.e.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests, *PT-2* using LR tests, and *PT-3* using F tests.
To avoid cluttering the legend and considering that they provide very similar results, the six different implementations described in Figure 1, Figure 2, and Figure 3 are all plotted with a marker and color according to whether they use *PT-1*, *PT-2*, or *PT-3*.

We estimate the following specification:[11]

$$r_t^s/\sigma_{t-1}^s = \alpha + \beta_h r_{t-h}^s/\sigma_{t-h-1}^s + \epsilon_t^s \tag{36}$$

where r_t^s is the excess return for instrument s in month t, h is the lag considered (h=1,..,12), and σ_{t-1}^s defined as ex ante volatility:

$$(\sigma_t^s)^2 = 261\sum_{i=0}^{\infty}(1-\delta)\delta^i(r_{t-1-i}^s - \bar{r}_t^s)^2 \tag{37}$$

where the scalar 261 scales the variance to be annual, the weights $(1-\delta)\delta^i$ add up to one, and $\bar{r}_t^s$ is the exponentially weighted average return computed similarly.

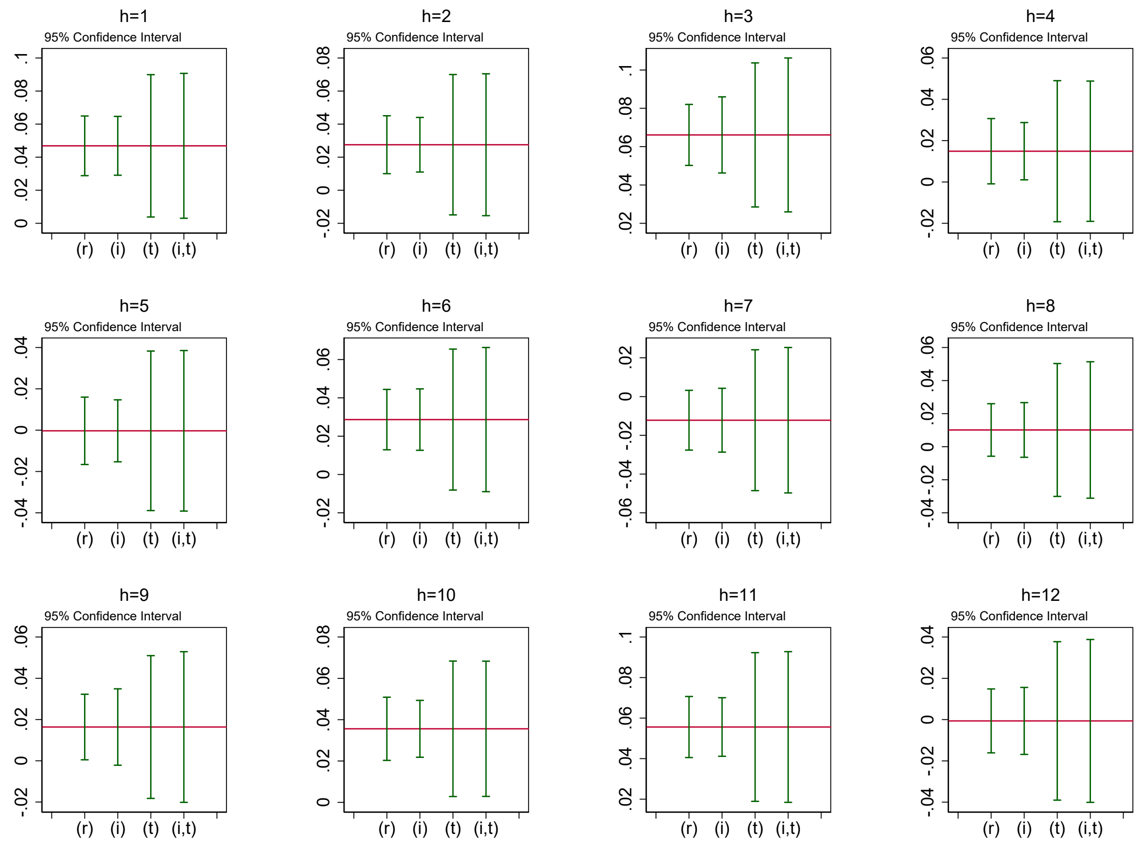
Figure 10 reports 95% confidence intervals for the twelve lags considered using different assumptions regarding the disturbances. There are important differences across variance estimators, especially between the White robust estimator or clustering by stock versus clustering by month or two-way clustered standard errors. Computing the standard errors using two-way clustering in this example, does not significantly increase the size of the confidence interval relative to the use of standard errors clustered by month.

Table 1 reports the pretest using F tests. We find that it favors the use of two-way cluster standard errors over the one-way clustered by time as done in Moskowitz et al. (2012) and Huang et al. (2020). Nonetheless, for this particular application both alternatives make little difference regarding the rejection of the null as previously pointed out.

The similarity of the confidence intervals obtained using standard errors clustered by time relative to those using two-way clustering is not a general result, as evidenced in the Monte Carlo simulations and the additional economic applications that we include in the Appendix.

[11]We restrict the beginning of the sample period to January 1993, so the panel becomes balanced. The results are qualitatively similar to the original length.

Figure 10: 95% Confidence intervals



6 Final Remarks

This paper analyzes the statistical inference regarding the parameters of a panel data model when it is first subjected to a pretest for the presence of individual and/or time unobserved cluster effects. Using simulations we compare the performance of six proposed diagnostics that use statistical tests available in the literature: Lagrange Multiplier, Likelihood ratios, and F tests. We find that these pretest estimators are a viable alternative to estimate panel data models with unobserved cluster effects. The empirical size of the t-test after pretesting remains close to the size of the test using a correction for the standard errors in line with true data generating process. These results are robust to the presence of temporary clusters effects, and non-normality of the disturbance, as well as non-normality of the regressor.

Table 1: Pretesting using F tests

	F_1	p-value	F_1	p-value
h=1	4.87	0.00	1.87	0.00
h=2	4.88	0.00	1.94	0.00
h=3	4.80	0.00	1.83	0.00
h=4	4.88	0.00	1.95	0.00
h=5	4.87	0.00	2.00	0.00
h=6	4.85	0.00	1.84	0.00
h=7	4.87	0.00	1.98	0.00
h=8	4.88	0.00	1.85	0.00
h=9	4.89	0.00	1.84	0.00
h=10	4.86	0.00	1.77	0.00
h=11	4.87	0.00	1.72	0.00
h=12	4.82	0.00	2.02	0.00

Notes: F_1 corresponds to a F statistic for the null hypothesis that all month fixed effects are zero, the associated p-values are reported next to them, and F_2 corresponds to a F statistic for the null hypothesis that all asset fixed effects are zero, the associated p-values are reported next to them.

References

Abadie, A., Athey, S., Imbens, G., & Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? *NBER Working Paper*, *24003*.

Acemoglu, D., Johnson, S., Robinson, J. A., & Yared, P. (2008). Income and Democracy. *American Economic Review*, *98*(3), 808–842.

Arellano, M. (1987). Computing Robust Standard Errors for Within-groups Estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–434.

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data* (5th ed.). Wiley.

Baltagi, B. H., Bresson, G., & Pirotte, A. (2003a). *A Comparative Study of Pure and Pretest Estimators for a Possibly Misspecified Two-Way Error Component Model* (Vol. 17; T. Fomby & R. Carter Hill, Eds.). Emerald Group Publishing Limited, Bingley.

Baltagi, B. H., Bresson, G., & Pirotte, A. (2003b). Fixed Effects, Random Effects or HausmanTaylor? A Pretest Estimator. *Economic Letters*, *79*, 361–369.

Baltagi, B. H., Chang, Y.-J., & Li, Q. (1992). Monte Carlo Results on Several New and Existing Tests for the Error Component Model. *Journal of Econometrics*, *54*, 95–120.

Baltagi, B. H., & Li, Q. (1990). A Lagrange Multiplier Test for the Error Components Model with Incomplete Panels. *Econometric Reviews*, *9*(1), 103–107.

Baltagi, B. H., & Li, Q. (1997). Monte Carlo Results on Pure and Pretest Estimators of an Error Component

Model with Autocorrelated Disturbances. *Annales D'Economie et de Statistique*, *48*, 69–82.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275.

Bester, A., Conley, T., & Hansen, C. (2011). Inference with Dependent Data using Cluster Covariance Estimators. *Journal of Econometrics*, *165*(2), 137–151. Retrieved from `http://dx.doi.org/10.1016/j.jeconom.2011.01.007`

Bonhomme, S., & Manresa, E. (2015). Grouped Patterns of Heterogeneity in Panel Data. *Econometrica*, *83*(3), 1147–1184.

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies*, *47*(1), 239–253.

Cameron, C., Gelbach, J., & Miller, D. (2011). Robust Inference with Multi-Way Clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Cameron, C., & Miller, D. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, *50*, 317–372.

Fama, E. F., & Macbeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *The Journal of Political Economy*, *81*(3), 607–636.

Gourieroux, C., Holly, A., & Monfort, A. (1982). Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica*, *50*(1), 63–80.

Gourio, F., Messer, T., & Siemer, M. (2016). Firm Entry and Macroeconomic Dynamics: A State-Level Analysis. *American Economic Review: Papers & Proceedings*, *106*(5), 214–218.

Greene, W. (2012). *Econometric Analysis* (7th ed.). Pearson.

Guggenberger, P. (2010). The Impact of a Hausman Pretest on the Size of a Hypothesis Test: The Panel Data Case. *Journal of Econometrics*, *156*, 337–343.

Honda, Y. (1985). Testing the Error Components Model with Non-Normal Disturbances. *The Review of Economic Studies*, *52*, 681–690.

Huang, D., Li, J., Wang, L., & Zhou, G. (2020). Time Series Momentum: Is it there? *Journal of Financial Economics*, *135*(3), 774–794.

Kabaila, P., Mainzer, R., & Farchione, D. (2015). The Impact of a Hausman Pretest, applied to Panel Data, on the Coverage Probability of Confidence Intervals. *Economics Letters*, *131*, 12–15.

Kim, J., & Wang, L. (2019). Hidden Group Patterns in Democracy Developments: Bayesian Inference for Grouped Heterogeneity. *Journal of Applied Econometrics*, *34*(6), 1016–1025.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal Data Analysis using Generalized Linear Models. *Biometrika*, *73*(1), 13–22.

Ma, J., & Vijverberg, W. P. M. (2010). Five Diagnostic Tests for Unobserved Cluster Effects. *Communications in Statistics - Simulation and Computation*, *39*, 1212–1227.

Mackinnon, J. (2018). How Cluster-Robust Inference is Changing Applied Econometrics. *mimeo*, 1–19.

Mackinnon, J., Nielsen, M., & Webb, M. (2019). Wild Bootstrap and Asymptotic Inference with Multiway Clustering. *Journal of Business & Economic Statistics*, *Published*.

Mackinnon, J., & Webb, M. D. (2019). When and How to Deal With Clustered Errors in Regression Models. *QED Working Paper*(1421).

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time Series Momentum. *Journal of Financial Economics*, *104*, 228–250.

Moulton, B. R. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units. *The Review of Economics and Statistics*, *72*(2), 334–338.

Moulton, B. R., & Randolph, W. C. (1989). Alternative Tests of the Error Components Model. *Econometrica*, *57*(3), 685–693.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708.

Petersen, M. A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies*, *22*(1), 435–480.

Rogers, W. (1993). Regression Standard Errors in Clustered Samples. *Stata Technical Bulletin*, *13*, 19–23.

Roodman, D., MacKinnon, J., Nielsen, M., & Webb, M. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *Stata Journal*, *19*(1), 4–60.

Thompson, S. B. (2011). Simple Formulas for Standard Errors that Cluster by Both Firm and Time. *Journal of Financial Economics*, *99*, 1–10.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817–838.

White, H. (1984). *Asymptotic Theory for Econometricians.* San Diego: Academic Press.

Ziemer, R. F., & Wetzstein, M. E. (1983). A Stein-Rule Method for Pooling Data. *Economic Letters*, *11*, 137–143.

Appendix

Table 2: Empirical size at $\alpha = 0.01$, by fraction of firm component

Shares	Cl.FY	Cl.F	Cl.Y	Rob	PT.1a	PT.1b	PT.1c	PT.1d	PT.1e	PT.1f	PT.2a	PT.2b	PT.2c	PT.2d	PT.2e	PT.2f	PT.3a	PT.3b	PT.3c	PT.3d	PT.3e	PT.3f
0-25	0.96	0.88	0.86	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
0-50	1.32	1.04	1.00	0.98	1.08	1.08	1.08	1.08	1.08	1.08	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06
0-75	1.30	1.00	0.96	0.96	1.02	1.02	1.02	1.02	1.02	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
25-25	0.46	1.26	3.42	4.68	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26
25-50	0.34	1.20	7.20	9.10	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18
25-75	0.38	1.38	11.36	13.04	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38
50-25	0.38	1.54	7.24	8.56	1.48	1.48	1.48	1.48	1.48	1.48	1.50	1.50	1.50	1.50	1.50	1.50	1.48	1.48	1.48	1.48	1.48	1.48
50-50	0.32	1.54	16.02	16.50	1.52	1.52	1.52	1.52	1.52	1.52	1.54	1.54	1.54	1.54	1.54	1.54	1.52	1.52	1.52	1.52	1.52	1.52
50-75	0.32	1.46	25.80	23.72	1.44	1.44	1.44	1.44	1.44	1.44	1.46	1.46	1.46	1.46	1.46	1.46	1.44	1.44	1.44	1.44	1.44	1.44
75-25	0.34	1.60	11.62	12.46	1.56	1.56	1.56	1.56	1.56	1.56	1.58	1.58	1.58	1.58	1.58	1.58	1.56	1.56	1.56	1.56	1.56	1.56
75-50	0.28	1.58	26.00	23.30	1.54	1.54	1.54	1.54	1.54	1.54	1.56	1.56	1.56	1.56	1.56	1.56	1.54	1.54	1.54	1.54	1.54	1.54
75-75	0.34	1.54	39.86	30.30	1.50	1.50	1.50	1.50	1.50	1.50	1.52	1.52	1.52	1.52	1.52	1.52	1.50	1.50	1.50	1.50	1.50	1.50

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i respectively.
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 3: Empirical size at $\alpha = 0.01$, by fraction of time component

Shares	Cl.FY	Cl.F	Cl.Y	Rob	PT.1a	PT.1b	PT.1c	PT.1d	PT.1e	PT.1f	PT.2a	PT.2b	PT.2c	PT.2d	PT.2e	PT.2f	PT.3a	PT.3b	PT.3c	PT.3d	PT.3e	PT.3f
0-25	1.22	0.96	1.10	0.88	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10
0-50	1.14	1.06	1.00	1.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0-75	1.10	1.08	0.88	1.12	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
25-25	1.44	50.82	1.36	50.28	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38
25-50	1.32	64.62	1.32	63.30	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32
25-75	1.40	72.06	1.36	70.66	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36	1.36
50-25	2.58	64.84	2.54	63.58	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54	2.54
50-50	2.58	75.92	2.60	74.44	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60
50-75	2.64	82.36	2.64	79.76	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.64
75-25	4.06	72.54	4.08	70.82	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08
75-50	3.92	82.42	3.90	79.94	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90
75-75	3.94	88.20	3.90	84.14	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the fraction of the variance in x_{it} and ϵ_{it} that is due to δ_t and ζ_t respectively.
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 4: Empirical size at $\alpha = 0.01$, by number of firms

Cl-FY	Cl-F	Cl-Y	Rob	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
3.80	9.04	56.08	69.28	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80	3.80
1.38	5.48	39.06	61.40	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38
0.56	3.60	17.54	52.32	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56
0.42	8.30	3.32	45.26	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
0.48	28.10	1.18	52.12	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
1.42	49.58	2.20	62.70	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42
4.12	64.56	4.72	69.04	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10	4.10

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The first column contains the number of firms (keeping N*T=2500).
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 5: Empirical size at $\alpha = 0.01$, by number of firms when there are no clusters

Cl-FY	Cl-F	Cl-Y	Rob	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
3.42	6.10	0.06	0.84	1.02	0.98	0.98	1.02	0.88	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.86	0.86	0.86	0.86	0.84	0.84
1.44	3.44	0.26	1.00	1.02	1.00	1.00	1.02	1.02	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.02	1.02	1.02	1.04	1.02
0.74	2.34	0.20	1.00	0.88	0.84	0.84	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.44	1.02	0.18	0.86	0.78	0.76	0.76	0.78	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
0.74	1.14	0.54	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
1.38	1.04	1.06	0.92	0.84	0.84	0.84	0.84	0.92	0.92	0.94	0.94	0.94	0.94	0.94	0.94	0.92	0.92	0.92	0.92	0.92	0.92
3.80	0.82	3.58	0.98	1.06	1.08	1.08	1.06	1.02	1.00	0.98	0.98	0.98	0.98	0.98	0.98	1.02	1.02	1.02	1.02	1.00	1.00

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The first column contains the number of firms (keeping N*T=2500).
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* to Figure 3.

A.1 Selected models

Table 6: Simulation with a firm effect - Selected models

	Variance	Model	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
1	0-0	1	77.72	77.72	77.72	77.72	97.52	98.5	98.7	98.7	98.7	98.7	98.7	99.04	98.1	98.1	98.1	98.1	98.04	98.54
2	0-0	2	9.44	10.3	10.3	9.44	1.56	1.02	0.52	0.52	0.52	0.52	0.52	0.52	1.06	1.06	1.06	1.06	1.1	1
3	0-0	3	12.84	11.98	11.98	12.84	0.92	0.48	0.78	0.78	0.78	0.78	0.78	0.44	0.84	0.84	0.84	0.84	0.86	0.46
4	0-0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0-25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0-25	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0-25	3	98.54	98.54	98.54	98.54	98.54	98.54	99.5	99.5	99.5	99.5	99.5	99.5	99.18	99.18	99.18	99.18	99.18	99.18
8	0-25	4	1.46	1.46	1.46	1.46	1.46	1.46	0.5	0.5	0.5	0.5	0.5	0.5	0.82	0.82	0.82	0.82	0.82	0.82
9	0-50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0-50	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0-50	3	98.56	98.56	98.56	98.56	98.56	98.56	99.48	99.48	99.48	99.48	99.48	99.48	99.18	99.18	99.18	99.18	99.18	99.18
12	0-50	4	1.44	1.44	1.44	1.44	1.44	1.44	0.52	0.52	0.52	0.52	0.52	0.52	0.82	0.82	0.82	0.82	0.82	0.82
13	0-75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0-75	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0-75	3	98.56	98.56	98.56	98.56	98.56	98.56	99.46	99.46	99.46	99.46	99.46	99.46	99.18	99.18	99.18	99.18	99.18	99.18
16	0-75	4	1.44	1.44	1.44	1.44	1.44	1.44	0.54	0.54	0.54	0.54	0.54	0.54	0.82	0.82	0.82	0.82	0.82	0.82
17	25-0	1	77.28	77.28	77.28	77.28	97.74	98.74	98.92	98.92	98.92	98.92	98.84	99.2	98.28	98.28	98.28	98.28	98.26	98.82
18	25-0	2	10.3	11.42	11.42	10.3	1.36	0.8	0.32	0.32	0.32	0.32	0.36	0.36	0.78	0.78	0.78	0.78	0.82	0.72
19	25-0	3	12.4	11.28	11.28	12.4	0.88	0.44	0.76	0.76	0.76	0.76	0.8	0.44	0.94	0.94	0.94	0.94	0.92	0.46
20	25-0	4	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0	0	0	0	0	0	0	0	0	0
21	25-25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	25-25	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	25-25	3	98.56	98.56	98.56	98.56	98.56	98.56	99.4	99.4	99.4	99.4	99.4	99.4	99.08	99.08	99.08	99.08	99.08	99.08
24	25-25	4	1.44	1.44	1.44	1.44	1.44	1.44	0.6	0.6	0.6	0.6	0.6	0.6	0.92	0.92	0.92	0.92	0.92	0.92
25	25-50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	25-50	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	25-50	3	98.56	98.56	98.56	98.56	98.56	98.56	99.36	99.36	99.36	99.36	99.36	99.36	99.08	99.08	99.08	99.08	99.08	99.08
28	25-50	4	1.44	1.44	1.44	1.44	1.44	1.44	0.64	0.64	0.64	0.64	0.64	0.64	0.92	0.92	0.92	0.92	0.92	0.92
29	25-75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	25-75	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	25-75	3	98.56	98.56	98.56	98.56	98.56	98.56	99.36	99.36	99.36	99.36	99.36	99.36	99.08	99.08	99.08	99.08	99.08	99.08
32	25-75	4	1.44	1.44	1.44	1.44	1.44	1.44	0.64	0.64	0.64	0.64	0.64	0.64	0.92	0.92	0.92	0.92	0.92	0.92
33	50-0	1	77.42	77.42	77.42	77.42	97.7	98.76	98.88	98.88	98.88	98.88	98.82	99.22	98.32	98.32	98.32	98.32	98.3	98.86
34	50-0	2	10.32	11.42	11.42	10.32	1.4	0.8	0.32	0.32	0.32	0.32	0.36	0.36	0.8	0.8	0.8	0.8	0.82	0.7
35	50-0	3	12.24	11.14	11.14	12.24	0.88	0.42	0.8	0.8	0.8	0.8	0.82	0.42	0.88	0.88	0.88	0.88	0.88	0.44
36	50-0	4	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0	0	0	0	0	0	0	0	0	0
37	50-25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	50-25	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	50-25	3	98.56	98.56	98.56	98.56	98.56	98.56	99.4	99.4	99.4	99.4	99.4	99.4	99.08	99.08	99.08	99.08	99.08	99.08
40	50-25	4	1.44	1.44	1.44	1.44	1.44	1.44	0.6	0.6	0.6	0.6	0.6	0.6	0.92	0.92	0.92	0.92	0.92	0.92
41	50-50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42	50-50	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	50-50	3	98.56	98.56	98.56	98.56	98.56	98.56	99.38	99.38	99.38	99.38	99.38	99.38	99.08	99.08	99.08	99.08	99.08	99.08
44	50-50	4	1.44	1.44	1.44	1.44	1.44	1.44	0.62	0.62	0.62	0.62	0.62	0.62	0.92	0.92	0.92	0.92	0.92	0.92
45	50-75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	50-75	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47	50-75	3	98.56	98.56	98.56	98.56	98.56	98.56	99.36	99.36	99.36	99.36	99.36	99.36	99.08	99.08	99.08	99.08	99.08	99.08
48	50-75	4	1.44	1.44	1.44	1.44	1.44	1.44	0.64	0.64	0.64	0.64	0.64	0.64	0.92	0.92	0.92	0.92	0.92	0.92
49	75-0	1	77.6	77.6	77.6	77.6	97.72	98.78	98.92	98.92	98.92	98.92	98.86	99.2	98.3	98.3	98.3	98.3	98.3	98.86
50	75-0	2	10.36	11.5	11.5	10.36	1.4	0.78	0.32	0.32	0.32	0.32	0.36	0.36	0.82	0.82	0.82	0.82	0.82	0.7
51	75-0	3	12.02	10.88	10.88	12.02	0.86	0.42	0.76	0.76	0.76	0.76	0.78	0.44	0.88	0.88	0.88	0.88	0.88	0.44
52	75-0	4	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0	0	0	0	0	0	0	0	0	0
53	75-25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54	75-25	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
55	75-25	3	98.56	98.56	98.56	98.56	98.56	98.56	99.4	99.4	99.4	99.4	99.4	99.4	99.08	99.08	99.08	99.08	99.08	99.08
56	75-25	4	1.44	1.44	1.44	1.44	1.44	1.44	0.6	0.6	0.6	0.6	0.6	0.6	0.92	0.92	0.92	0.92	0.92	0.92
57	75-50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	75-50	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59	75-50	3	98.56	98.56	98.56	98.56	98.56	98.56	99.38	99.38	99.38	99.38	99.38	99.38	99.08	99.08	99.08	99.08	99.08	99.08
60	75-50	4	1.44	1.44	1.44	1.44	1.44	1.44	0.62	0.62	0.62	0.62	0.62	0.62	0.92	0.92	0.92	0.92	0.92	0.92
61	75-75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	75-75	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63	75-75	3	98.56	98.56	98.56	98.56	98.56	98.56	99.36	99.36	99.36	99.36	99.36	99.36	99.08	99.08	99.08	99.08	99.08	99.08
64	75-75	4	1.44	1.44	1.44	1.44	1.44	1.44	0.64	0.64	0.64	0.64	0.64	0.64	0.92	0.92	0.92	0.92	0.92	0.92

The chosen model corresponds to Case 1: $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 = 0$; Case 2: $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 > 0$; Case 3: $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 = 0$; and Case 4: $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 > 0$.

Table 7: Simulation with a time effect - Selected models

	Variance	Model	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
1	0-0	1	77.72	77.72	77.72	77.72	97.52	98.5	98.7	98.7	98.7	98.7	98.7	99.04	98.1	98.1	98.1	98.1	98.04	98.54
2	0-0	2	9.44	10.3	10.3	9.44	1.56	1.02	0.52	0.52	0.52	0.52	0.52	0.52	1.06	1.06	1.06	1.06	1.1	1
3	0-0	3	12.84	11.98	11.98	12.84	0.92	0.48	0.78	0.78	0.78	0.78	0.78	0.44	0.84	0.84	0.84	0.84	0.86	0.46
4	0-0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0-25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0-25	2	98.98	98.98	98.98	98.98	98.98	98.98	99.24	99.24	99.24	99.24	99.24	99.24	99.2	99.2	99.2	99.2	99.2	99.2
7	0-25	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0-25	4	1.02	1.02	1.02	1.02	1.02	1.02	0.76	0.76	0.76	0.76	0.76	0.76	0.8	0.8	0.8	0.8	0.8	0.8
9	0-50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0-50	2	98.98	98.98	98.98	98.98	98.98	98.98	99.24	99.24	99.24	99.24	99.24	99.24	99.2	99.2	99.2	99.2	99.2	99.2
11	0-50	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0-50	4	1.02	1.02	1.02	1.02	1.02	1.02	0.76	0.76	0.76	0.76	0.76	0.76	0.8	0.8	0.8	0.8	0.8	0.8
13	0-75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0-75	2	98.98	98.98	98.98	98.98	98.98	98.98	99.24	99.24	99.24	99.24	99.24	99.24	99.2	99.2	99.2	99.2	99.2	99.2
15	0-75	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0-75	4	1.02	1.02	1.02	1.02	1.02	1.02	0.76	0.76	0.76	0.76	0.76	0.76	0.8	0.8	0.8	0.8	0.8	0.8
17	25-0	1	77.58	77.58	77.58	77.58	97.94	98.96	99.1	99.1	99.1	99.1	99.06	99.38	98.36	98.36	98.36	98.36	98.42	99
18	25-0	2	9.4	10.56	10.56	9.4	1.12	0.62	0.18	0.18	0.18	0.18	0.2	0.2	0.8	0.8	0.8	0.8	0.74	0.58
19	25-0	3	12.98	11.82	11.82	12.98	0.9	0.38	0.7	0.7	0.7	0.7	0.72	0.4	0.82	0.82	0.82	0.82	0.82	0.4
20	25-0	4	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
21	25-25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	25-25	2	99.08	99.08	99.08	99.08	99.08	99.08	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24
23	25-25	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	25-25	4	0.92	0.92	0.92	0.92	0.92	0.92	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
25	25-50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	25-50	2	99.08	99.08	99.08	99.08	99.08	99.08	99.26	99.26	99.26	99.26	99.26	99.26	99.24	99.24	99.24	99.24	99.24	99.24
27	25-50	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	25-50	4	0.92	0.92	0.92	0.92	0.92	0.92	0.74	0.74	0.74	0.74	0.74	0.74	0.76	0.76	0.76	0.76	0.76	0.76
29	25-75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	25-75	2	99.08	99.08	99.08	99.08	99.08	99.08	99.28	99.28	99.28	99.28	99.28	99.28	99.24	99.24	99.24	99.24	99.24	99.24
31	25-75	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	25-75	4	0.92	0.92	0.92	0.92	0.92	0.92	0.72	0.72	0.72	0.72	0.72	0.72	0.76	0.76	0.76	0.76	0.76	0.76
33	50-0	1	78.26	78.26	78.26	78.26	98.16	99.06	99.1	99.1	99.1	99.1	99.08	99.4	98.42	98.42	98.42	98.42	98.5	99.1
34	50-0	2	8.7	9.72	9.72	8.7	0.86	0.52	0.18	0.18	0.18	0.18	0.18	0.18	0.72	0.72	0.72	0.72	0.66	0.48
35	50-0	3	13	11.98	11.98	13	0.94	0.38	0.7	0.7	0.7	0.7	0.72	0.4	0.82	0.82	0.82	0.82	0.8	0.38
36	50-0	4	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.04	0.04	0.04	0.04	0.04	0.04
37	50-25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	50-25	2	99.08	99.08	99.08	99.08	99.08	99.08	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24
39	50-25	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	50-25	4	0.92	0.92	0.92	0.92	0.92	0.92	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
41	50-50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42	50-50	2	99.08	99.08	99.08	99.08	99.08	99.08	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24
43	50-50	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	50-50	4	0.92	0.92	0.92	0.92	0.92	0.92	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
45	50-75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	50-75	2	99.08	99.08	99.08	99.08	99.08	99.08	99.26	99.26	99.26	99.26	99.26	99.26	99.24	99.24	99.24	99.24	99.24	99.24
47	50-75	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	50-75	4	0.92	0.92	0.92	0.92	0.92	0.92	0.74	0.74	0.74	0.74	0.74	0.74	0.76	0.76	0.76	0.76	0.76	0.76
49	75-0	1	78.8	78.8	78.8	78.8	98.28	99.1	99.1	99.1	99.1	99.1	99.08	99.4	98.36	98.36	98.36	98.36	98.36	99.12
50	75-0	2	8.12	9.12	9.12	8.12	0.74	0.48	0.18	0.18	0.18	0.18	0.18	0.18	0.76	0.76	0.76	0.76	0.8	0.46
51	75-0	3	13.04	12.04	12.04	13.04	0.94	0.38	0.7	0.7	0.7	0.7	0.72	0.4	0.84	0.84	0.84	0.84	0.8	0.38
52	75-0	4	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.04	0.04	0.04	0.04	0.04	0.04
53	75-25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54	75-25	2	99.08	99.08	99.08	99.08	99.08	99.08	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24
55	75-25	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56	75-25	4	0.92	0.92	0.92	0.92	0.92	0.92	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
57	75-50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	75-50	2	99.08	99.08	99.08	99.08	99.08	99.08	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24
59	75-50	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	75-50	4	0.92	0.92	0.92	0.92	0.92	0.92	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
61	75-75	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	75-75	2	99.08	99.08	99.08	99.08	99.08	99.08	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24	99.24
63	75-75	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
64	75-75	4	0.92	0.92	0.92	0.92	0.92	0.92	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76

The chosen model corresponds to Case 1: $\sigma^2_\gamma = 0$ and $\sigma^2_\delta = 0$; Case 2: $\sigma^2_\gamma = 0$ and $\sigma^2_\delta > 0$; Case 3: $\sigma^2_\gamma > 0$ and $\sigma^2_\delta = 0$; and Case 4: $\sigma^2_\gamma > 0$ and $\sigma^2_\delta > 0$.

Table 8: Simulation with both firm and time effect - Selected models

	Firms	Model	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
1	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	5	4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
5	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	10	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	10	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	10	4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
9	20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	20	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	20	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	20	4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
13	50	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	50	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	50	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	50	4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
17	125	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	125	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	125	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	125	4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
21	250	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	250	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	250	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	250	4	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
25	500	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	500	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	500	3	0.02	0.02	0.02	0.02	0.02	0.02	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02
28	500	4	99.98	99.98	99.98	99.98	99.98	99.98	99.96	99.96	99.96	99.96	99.96	99.96	99.98	99.98	99.98	99.98	99.98	99.98

The chosen model corresponds to Case 1: $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 = 0$; Case 2: $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 > 0$; Case 3: $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 = 0$; and Case 4: $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 > 0$.

Table 9: Simulation without firm or time effect - Selected models

	Firms	Model	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
1	5	1	78.96	78.96	78.96	78.96	98.48	99.12	99.22	99.22	99.22	99.22	99.2	99.44	98.52	98.52	98.52	98.52	98.6	99.2
2	5	2	11.92	12.92	12.92	11.92	0.58	0.4	0.62	0.62	0.62	0.62	0.64	0.4	0.74	0.74	0.74	0.74	0.76	0.4
3	5	3	9.1	8.1	8.1	9.1	0.92	0.46	0.16	0.16	0.16	0.16	0.16	0.16	0.72	0.72	0.72	0.72	0.62	0.38
4	5	4	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0	0	0	0	0.02	0.02	0.02	0.02	0.02	0.02
5	10	1	78.64	78.64	78.64	78.64	98.18	99.02	99.16	99.16	99.16	99.16	99.14	99.44	98.28	98.28	98.28	98.28	98.24	99.1
6	10	2	11.28	12.3	12.3	11.28	0.78	0.38	0.66	0.66	0.66	0.66	0.68	0.38	0.82	0.82	0.82	0.82	0.82	0.38
7	10	3	10.08	9.06	9.06	10.08	1.04	0.6	0.18	0.18	0.18	0.18	0.18	0.18	0.9	0.9	0.9	0.9	0.94	0.52
8	10	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	20	1	77.74	77.74	77.74	77.74	97.58	98.4	98.74	98.74	98.74	98.74	98.68	98.78	97.76	97.76	97.76	97.76	97.78	98.42
10	20	2	11.4	12.34	12.34	11.4	1.24	0.8	0.82	0.82	0.82	0.82	0.84	0.74	1.18	1.18	1.18	1.18	1.16	0.78
11	20	3	10.86	9.92	9.92	10.86	1.18	0.8	0.44	0.44	0.44	0.44	0.48	0.48	1.06	1.06	1.06	1.06	1.06	0.8
12	20	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	50	1	76.76	76.76	76.76	76.76	97.98	98.88	98.9	98.9	98.9	98.9	98.88	98.94	98.2	98.2	98.2	98.2	98.32	98.9
14	50	2	10.9	11.94	11.94	10.9	1.12	0.52	0.5	0.5	0.5	0.5	0.5	0.46	0.88	0.88	0.88	0.88	0.82	0.5
15	50	3	12.34	11.3	11.3	12.34	0.9	0.6	0.6	0.6	0.6	0.6	0.62	0.6	0.92	0.92	0.92	0.92	0.86	0.6
16	50	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	125	1	76.86	76.86	76.86	76.86	97.38	98.64	98.76	98.76	98.76	98.76	98.74	99.02	97.92	97.92	97.92	97.92	97.8	98.68
18	125	2	10.34	11.58	11.58	10.34	1.3	0.76	0.44	0.44	0.44	0.44	0.46	0.46	1.02	1.02	1.02	1.02	1.04	0.76
19	125	3	12.76	11.52	11.52	12.76	1.28	0.56	0.78	0.78	0.78	0.78	0.78	0.5	1.04	1.04	1.04	1.04	1.14	0.54
20	125	4	0.04	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
21	250	1	78.5	78.5	78.5	78.5	97.78	98.66	98.72	98.72	98.72	98.72	98.7	99.08	97.98	97.98	97.98	97.98	98	98.74
22	250	2	8.68	9.56	9.56	8.68	1.16	0.78	0.4	0.4	0.4	0.4	0.42	0.42	1	1	1	1	1	0.7
23	250	3	12.8	11.92	11.92	12.8	1.04	0.54	0.88	0.88	0.88	0.88	0.88	0.5	1	1	1	1	0.98	0.54
24	250	4	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0	0	0	0	0.02	0.02	0.02	0.02	0.02	0.02
25	500	1	79.08	79.08	79.08	79.08	97.92	98.76	98.82	98.82	98.82	98.82	98.74	99.32	97.74	97.74	97.74	97.74	97.98	98.82
26	500	2	7.82	8.62	8.62	7.82	1.04	0.72	0.18	0.18	0.18	0.18	0.18	0.18	1.12	1.12	1.12	1.12	0.86	0.66
27	500	3	13.08	12.28	12.28	13.08	1.02	0.5	1	1	1	1	1.08	0.5	1.12	1.12	1.12	1.12	1.14	0.5
28	500	4	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0	0	0	0	0.02	0.02	0.02	0.02	0.02	0.02

The chosen model corresponds to Case 1: $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 = 0$; Case 2: $\sigma_\gamma^2 = 0$ and $\sigma_\delta^2 > 0$; Case 3: $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 = 0$; and Case 4: $\sigma_\gamma^2 > 0$ and $\sigma_\delta^2 > 0$.

A.2 Rejection rates of individual tests

Table 10: Simulation with a firm effect - rejection rates for individual tests

	LM				LR				F				GHM
	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0,\delta=0}$
0-25	100.00	2.28	100.00	1.46	100.00	0.04	100.00	0.50	100.00	0.04	100.00	0.82	100.00
0-50	100.00	0.06	100.00	1.44	100.00	0.00	100.00	0.52	100.00	0.00	100.00	0.82	100.00
0-75	100.00	0.00	100.00	1.44	100.00	0.00	100.00	0.54	100.00	0.00	100.00	0.82	100.00
25-25	100.00	2.44	100.00	1.44	100.00	0.00	100.00	0.60	100.00	0.02	100.00	0.92	100.00
25-50	100.00	0.04	100.00	1.44	100.00	0.00	100.00	0.64	100.00	0.00	100.00	0.92	100.00
25-75	100.00	0.00	100.00	1.44	100.00	0.00	100.00	0.64	100.00	0.00	100.00	0.92	100.00
50-25	100.00	2.42	100.00	1.44	100.00	0.00	100.00	0.60	100.00	0.02	100.00	0.92	100.00
50-50	100.00	0.04	100.00	1.44	100.00	0.00	100.00	0.62	100.00	0.00	100.00	0.92	100.00
50-75	100.00	0.00	100.00	1.44	100.00	0.00	100.00	0.64	100.00	0.00	100.00	0.92	100.00
75-25	100.00	2.42	100.00	1.44	100.00	0.00	100.00	0.60	100.00	0.02	100.00	0.92	100.00
75-50	100.00	0.04	100.00	1.44	100.00	0.00	100.00	0.62	100.00	0.00	100.00	0.92	100.00
75-75	100.00	0.00	100.00	1.44	100.00	0.00	100.00	0.64	100.00	0.00	100.00	0.92	100.00

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i respectively.

Table 11: Simulation with a time effect - rejection rates for individual tests

	LM				LR				F				GHM
	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0,\delta=0}$
0-25	0.18	100.00	1.02	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.80	100.00	100.00
0-50	0.00	100.00	1.02	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.80	100.00	100.00
0-75	0.00	100.00	1.02	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.80	100.00	100.00
25-25	0.14	100.00	0.92	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.76	100.00	100.00
25-50	0.00	100.00	0.92	100.00	0.00	100.00	0.74	100.00	0.00	100.00	0.76	100.00	100.00
25-75	0.00	100.00	0.92	100.00	0.00	100.00	0.72	100.00	0.00	100.00	0.76	100.00	100.00
50-25	0.16	100.00	0.92	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.76	100.00	100.00
50-50	0.00	100.00	0.92	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.76	100.00	100.00
50-75	0.00	100.00	0.92	100.00	0.00	100.00	0.74	100.00	0.00	100.00	0.76	100.00	100.00
75-25	0.18	100.00	0.92	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.76	100.00	100.00
75-50	0.00	100.00	0.92	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.76	100.00	100.00
75-75	0.00	100.00	0.92	100.00	0.00	100.00	0.76	100.00	0.00	100.00	0.76	100.00	100.00

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the fraction of the variance in x_{it} and ϵ_{it} that is due to δ_t and ζ_t respectively.

Table 12: Simulation with both firm and time effects - individual tests

	LM				LR				F				GHM
	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0,\delta=0}$
5	100.00	99.96	100.00	100.00	99.98	99.84	100.00	100.00	99.98	99.86	100.00	100.00	100.00
10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
20	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
50	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
125	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
250	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
500	99.94	99.98	100.00	99.98	99.84	99.88	100.00	99.96	99.86	99.92	100.00	99.98	100.00

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The first column contains the number of firms (keeping N*T=2500).

Table 13: Simulation without both firm and time effects - rejection rates for individual tests

	LM				LR				F				GHM
	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0\|\delta=0}$	$H_0^{\delta=0\|\gamma=0}$	$H_0^{\gamma=0\|\delta>0}$	$H_0^{\delta=0\|\gamma>0}$	$H_0^{\gamma=0,\delta=0}$
5	9.28	13.00	0.94	0.60	0.16	0.62	0.16	0.64	0.74	0.76	0.64	0.78	0.88
10	10.12	12.46	1.04	0.78	0.18	0.66	0.18	0.68	0.90	0.82	0.94	0.82	1.10
20	11.00	12.50	1.18	1.24	0.44	0.82	0.48	0.84	1.08	1.18	1.06	1.16	1.62
50	12.46	12.00	0.90	1.12	0.60	0.50	0.62	0.50	0.92	0.88	0.86	0.82	1.16
125	12.98	11.72	1.32	1.34	0.80	0.44	0.80	0.48	1.06	1.04	1.16	1.06	1.46
250	12.92	9.72	1.06	1.18	0.88	0.40	0.88	0.42	1.02	1.02	1.00	1.02	1.34
500	13.24	8.74	1.04	1.06	1.00	0.18	1.08	0.18	1.12	1.14	1.16	0.88	1.24

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The first column contains the number of firms (keeping N*T=2500).

A.2 Simulation with non-normality of the disturbance

Table 14: Simulation with a firm cluster effect - non-normal disturbances

Shares	Cl.FY	Cl.F	Cl.Y	Rob	PT.1a	PT.1b	PT.1c	PT.1d	PT.1e	PT.1f	PT.2a	PT.2b	PT.2c	PT.2d	PT.2e	PT.2f	PT.3a	PT.3b	PT.3c	PT.3d	PT.3e	PT.3f
0-25	1.22	1.10	0.96	1.18	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10
0-50	1.18	1.00	0.92	0.94	1.02	1.02	1.02	1.02	1.02	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0-75	1.76	0.86	1.00	1.00	0.88	0.88	0.88	0.88	0.88	0.88	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
25-25	0.64	1.22	4.02	4.10	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22
25-50	0.34	0.88	6.86	7.62	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
25-75	0.20	0.92	11.58	12.20	0.90	0.90	0.90	0.90	0.90	0.90	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
50-25	0.28	1.14	6.76	7.74	1.10	1.10	1.10	1.10	1.10	1.10	1.12	1.12	1.12	1.12	1.12	1.12	1.10	1.10	1.10	1.10	1.10	1.10
50-50	0.22	0.88	16.00	15.12	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
50-75	0.12	0.88	25.14	21.92	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
75-25	0.28	0.96	10.82	11.64	0.92	0.92	0.92	0.92	0.92	0.92	0.94	0.94	0.94	0.94	0.94	0.94	0.92	0.92	0.92	0.92	0.92	0.92
75-50	0.18	0.94	25.52	21.90	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
75-75	0.14	0.92	40.56	30.20	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i respectively.
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 15: Simulation with a time cluster effect - non-normal disturbances

Shares	Cl.FY	Cl.F	Cl.Y	Rob	PT.1a	PT.1b	PT.1c	PT.1d	PT.1e	PT.1f	PT.2a	PT.2b	PT.2c	PT.2d	PT.2e	PT.2f	PT.3a	PT.3b	PT.3c	PT.3d	PT.3e	PT.3f
0-25	1.40	0.80	1.10	0.96	1.08	1.08	1.08	1.08	1.10	1.10	1.08	1.08	1.08	1.08	1.08	1.08	1.10	1.10	1.10	1.10	1.10	1.10
0-50	1.12	1.12	0.96	1.20	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
0-75	1.14	1.08	0.80	1.12	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
25-25	0.70	38.08	0.72	37.98	0.76	0.76	0.76	0.76	0.74	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.74	0.74	0.74	0.74	0.76	0.76
25-50	0.74	59.26	0.74	58.76	0.72	0.72	0.72	0.72	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
25-75	0.98	69.14	0.98	67.86	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
50-25	1.30	49.48	1.26	49.16	1.36	1.34	1.34	1.36	1.50	1.54	1.58	1.58	1.58	1.58	1.58	1.60	1.52	1.52	1.52	1.52	1.50	1.54
50-50	1.20	69.86	1.20	68.48	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20
50-75	1.94	80.44	1.94	77.96	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94	1.94
75-25	1.34	58.28	1.34	57.12	1.36	1.34	1.34	1.36	1.56	1.62	1.70	1.70	1.70	1.70	1.70	1.70	1.50	1.50	1.50	1.50	1.50	1.62
75-50	2.06	77.06	2.10	75.30	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10
75-75	3.00	86.16	2.98	83.08	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98	2.98

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i respectively.
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 16: Simulation with firm and time effects - non-normal disturbances

N	Cl.FY	Cl.F	Cl.Y	Rob	PT.1a	PT.1b	PT.1c	PT.1d	PT.1e	PT.1f	PT.2a	PT.2b	PT.2c	PT.2d	PT.2e	PT.2f	PT.3a	PT.3b	PT.3c	PT.3d	PT.3e	PT.3f
5.00	2.64	8.48	39.92	53.50	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52
10.00	0.62	3.94	30.98	53.84	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62
20.00	0.18	4.40	17.68	48.32	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
50.00	0.14	10.64	5.00	44.78	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
125.00	0.24	26.40	1.68	48.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
250.00	0.58	41.02	1.92	54.00	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58
500.00	2.74	47.58	4.76	53.82	2.70	2.70	2.70	2.70	2.70	2.70	2.72	2.72	2.72	2.72	2.72	2.72	2.70	2.70	2.70	2.70	2.70	2.70

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The first column contains the number of firms (keeping N*T=2500).
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

A.3 Simulation with non-normality of the exogenous regressor

Table 17: Simulation with a firm cluster effect - non-normal regressor

Shares	Cl-FY	Cl-F	Cl-Y	Rob	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
0-25	1.22	1.32	1.28	1.2	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32
0-50	1.54	1.16	1.3	1.24	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18
0-75	1.26	1.18	1.24	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18
25-25	0.66	1.38	3.14	4.22	1.34	1.34	1.34	1.34	1.34	1.34	1.38	1.38	1.38	1.38	1.38	1.38	1.36	1.36	1.36	1.36	1.36	1.36
25-50	0.38	1.16	5.34	6.56	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16
25-75	0.3	1.16	8.14	9.78	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16
50-25	0.42	1.26	6.14	7.44	1.24	1.24	1.24	1.24	1.24	1.24	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26
50-50	0.38	1.22	13.24	13.64	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22
50-75	0.32	1.28	20.88	19.3	1.24	1.24	1.24	1.24	1.24	1.24	1.28	1.28	1.28	1.28	1.28	1.28	1.26	1.26	1.26	1.26	1.26	1.26
75-25	0.44	1.38	9.3	10.76	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38
75-50	0.52	1.44	22.54	20.68	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42
75-75	0.38	1.52	36.86	28.68	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i respectively.
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 18: Simulation with a time cluster effect - non-normal regressor

Shares	Cl-FY	Cl-F	Cl-Y	Rob	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
0-25	1.3	1.14	1.1	1.24	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
0-50	1.3	1.12	1.06	1.14	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06
0-75	1.04	1.1	0.86	1.24	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
25-25	1.32	44.12	1.28	43.98	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28
25-50	1.64	60.52	1.62	59.7	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62
25-75	1.78	68.92	1.76	67.48	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76	1.76
50-25	3.3	61.56	3.3	60.82	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3	3.3
50-50	3.5	73.56	3.5	71.9	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
50-75	3.38	81.2	3.38	79.16	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38	3.38
75-25	5.14	69	5.08	67.8	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08
75-50	4.88	82.2	4.88	79.8	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88	4.88
75-75	4.62	87.08	4.58	84.24	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58	4.58

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the fraction of the variance in x_{it} and ϵ_{it} that is due to γ_i and μ_i respectively.
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 19: Simulation with firm and time effects - non-normal regressor

N	Cl-FY	Cl-F	Cl-Y	Rob	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
5	4.56	10.46	52.34	66.08	4.56	4.56	4.56	4.56	4.56	4.56	4.58	4.58	4.58	4.58	4.58	4.58	4.56	4.56	4.56	4.56	4.56	4.56
10	2.04	7.14	37.3	58.9	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04	2.04
20	0.86	5.48	17.06	48.72	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86
50	0.54	9.24	3.9	41.86	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54
125	0.92	26.36	2.2	48.7	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
250	2.06	46.44	2.76	58.46	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06	2.06
500	4.5	59.56	5.32	65.18	4.5	4.5	4.5	4.5	4.5	4.5	4.54	4.54	4.54	4.54	4.54	4.54	4.5	4.5	4.5	4.5	4.5	4.5

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The first column contains the number of firms (keeping N*T=2500).
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 20: Simulation without cluster effects - non-normal regressor

N	Cl-FY	Cl-F	Cl-Y	Rob	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
5	5.28	8.78	0.2	1.28	1.84	1.74	1.74	1.84	1.34	1.34	1.3	1.3	1.3	1.3	1.3	1.3	1.32	1.32	1.32	1.32	1.32	1.32
10	2.64	4.62	0.1	0.74	0.96	0.9	0.9	0.96	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
20	1.5	3.1	0.28	1.28	1.26	1.22	1.22	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26	1.26
50	1	2.04	0.34	1.16	1.12	1.06	1.06	1.12	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16	1.16
125	1.5	1.22	0.7	0.94	0.82	0.8	0.8	0.82	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
250	2.32	1.02	1.84	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
500	5.04	1.32	4.52	1.24	1.32	1.34	1.34	1.32	1.26	1.26	1.24	1.24	1.24	1.24	1.24	1.24	1.26	1.26	1.26	1.26	1.26	1.26

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
The first column contains the number of firms (keeping N*T=2500).
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

A.4 Simulation with temporary firm cluster effect

Table 21: Simulation with a temporary firm effect

	Cl-FY	Cl-F	Cl-Y	Rob	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
0/0,.5/.5	0.28	1.10	15.46	14.76	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10
.9/.9,0/0	0.30	1.08	38.56	27.38	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04
.75/.75,.25/.25	0.22	1.12	23.28	20.18	1.04	1.04	1.04	1.04	1.04	1.04	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08	1.08

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the values of ϕ and the share of the variance of ϵ_{it} and x_{it} explained by the fixed firm component.
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

Table 22: Simulation with a time effect and a temporary firm effect

	Cl-FY	Cl-F	Cl-Y	Rob	PT-1a	PT-1b	PT-1c	PT-1d	PT-1e	PT-1f	PT-2a	PT-2b	PT-2c	PT-2d	PT-2e	PT-2f	PT-3a	PT-3b	PT-3c	PT-3d	PT-3e	PT-3f
0/0,.5/.5	0.40	8.14	3.68	45.64	0.40	0.40	0.40	0.40	0.40	0.40	0.48	0.48	0.48	0.48	0.48	0.48	0.40	0.40	0.40	0.40	0.40	0.40
.9/.9,.5/.5	0.40	5.70	6.96	54.22	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
.75/.75,.5/.5	0.42	6.94	4.72	47.14	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42

Notes:
Cells denote the percentage of rejections of $H_0 : \beta = 1$ at a nominal significance level (alpha) of 1% out of 5,000 simulations.
All the simulations are done with N=250 and T=10.
The first column contains the values of ϕ and the share of the variance of ϵ_{it} and x_{it} explained by the fixed firm component.
All the simulations are done with N=50 and T=50.
Cl-FY uses standard errors clustered by firm and year.
Cl-F uses standard errors clustered by firm.
Cl-Y uses standard errors clustered by year.
Rob uses heteroskedascity-robust standard errors.
PT-1 refers to pre-test estimator using Lagrange Multiplier tests.
PT-2 refers to pre-test estimator using LR tests.
PT-3 refers to pre-test estimator using F tests.
a and *b* refers to the implementation according to Figure 1, *c* and *d* to Figure 2, and *e* and *f* to Figure 3.

7 Applications in economics

7.1 Application 1

Gourio, Messer, and Siemer (2016) estimates impulse responses to entry shocks on an annual panel of US states over the period 1982-2014 using local projections (see **?**). We successfully replicate the main table of the paper (see Table 23).[12]

Table 23: Effect of a shock to entry of firms four years later

	1	2	3	4	5
GDP	.23 (4.78)	.21 (3.92)	.16 (2.56)	.3 (6.34)	.12 (3.52)
TFP	.09 (3.12)	.09 (3.31)	.05 (1.74)	.21 (5.04)	.06 (2.45)
Population	.07 (3.77)	.07 (3.26)	.06 (2.52)	.11 (4.6)	–
NFP	.06 (2.4)	.07 (2.4)	.04 (1.46)	.09 (3.86)	.01 (.68)
Number of Firms	.2 (3.61)	.18 (3.52)	.17 (3.35)	.28 (3.93)	.09 (2.34)
Number of Firms age 1	.52 (5.33)	.44 (4.52)	.4 (3.29)	.67 (6.23)	.31 (4.6)
Number of exiting firms	.33 (4.77)	.31 (4.43)	.19 (1.71)	.51 (3.42)	.23 (5.28)

Specification: $y_{i,t+4} = \alpha_i + \delta_t + \gamma s_{it} + x'_{i,t}\beta + \epsilon_{i,t}$
where $y_{i,t+4}$ is the log of the outcome variable specified on the left of the table, α_i is a state fixed effect, δ_t is a time fixed effect, s_{it} is the log change in the number of startups in state i between *t-1* and t, and $x'_{i,t}$ a vector of controls (which always includes $y_{i,t}$, $y_{i,t-1}$, and $y_{i,t-2}$).
The table report estimates of γ for different outcome variables and different specifications: column 1 is baseline with data over the period 1982-2014, column 2 sample without the Great Recession, column 3 including pre-1982 data, column 4 only lagged dependent variable as a control, and column 5 includes future population growth as a control. Standard errors are two-way clustered and t-stats are reported in parenthesis.

We compute confidence intervals for each coefficient reported in the baseline estimate of Table 23 according to different assumptions regarding the disturbances.[13] Figure 11 reports the results. We find important variation across methods but in general with confidence intervals computed using two-way clustering being wider than the competing approaches.

We conduct our pretest strategy using F tests with this data set according to the strategy lined out in Figure 3 as option e) and find that it supports the approach taken by the authors

[12]The data is publicly available at https://www.openicpsr.org/openicpsr/project/113458/version/V1/view.
[13]We compute the confidence interval that allows for two-way clustering using the Stata command reghdfe, which uses a finite-sample adjustment according to the smallest number of clusters (in this case, years) while Gourio et al. (2016) uses cluster2 with different finite-sample adjustment, so our standard errors differ slightly.

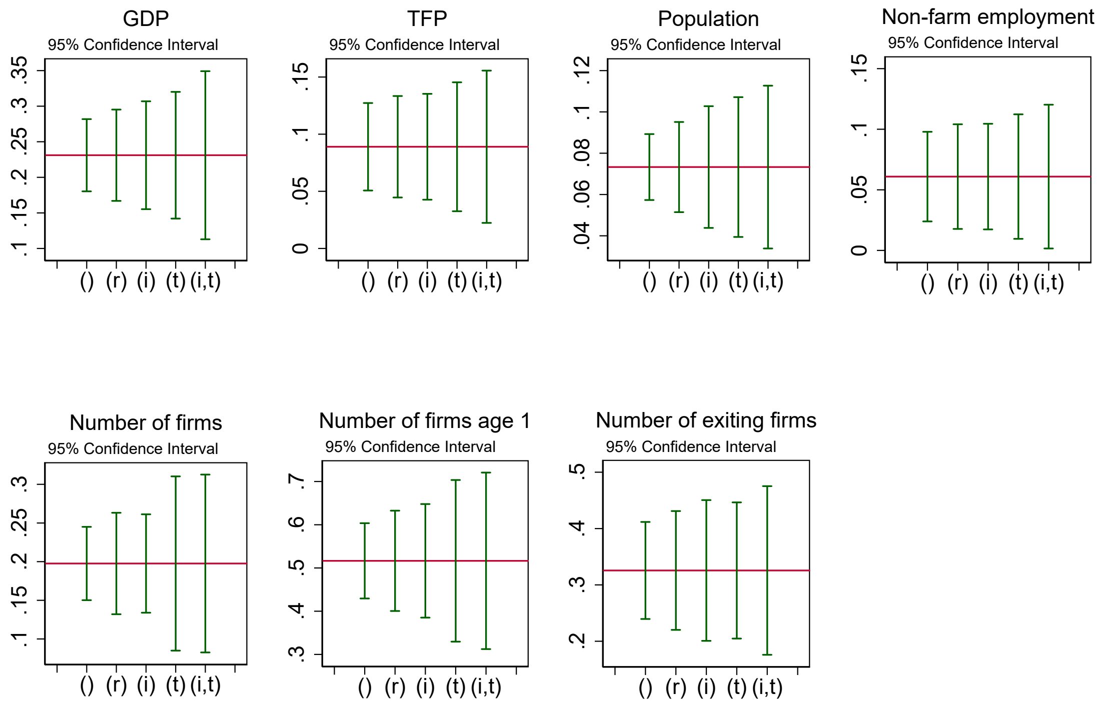


Figure 11: Confidence intervals for baseline estimates in Gourio et al. (2016)
Notes: *()*, *(r)*, *(i)*, *(t)*, *(i,t)* denotes respectively standard errors computed assuming iid disturbances, using heteroskedasticity robust s.e., clustering by state, clustering by year, and two-way clustering. The red line marks the value of the estimated coefficient.

in the original paper, i.e. two-way clustering by state and year (See Table 24).

Table 24: Pretesting using F tests and option e)

	F_1	p-value	F_2	p-value	Result
GDP	10.65	0.00	17.24	0.00	Two-way
TFP	9.05	0.00	9.66	0.00	Two-way
Population	33.66	0.00	15.31	0.00	Two-way
Non-farm employment	16.13	0.00	34.82	0.00	Two-way
Number of firms	15.86	0.00	27.13	0.00	Two-way
Number of firms age 1	17.79	0.00	53.21	0.00	Two-way
Number of exiting firms	31.20	0.00	25.18	0.00	Two-way

Notes: F_1 corresponds to a F statistic for the null hypothesis that all state fixed effects are zero, the associated p-values are reported next to them, and F_2 corresponds to a F statistic for the null hypothesis that all year fixed effects are zero, the associated p-values are reported next to them.

7.2 Application 2

Acemoglu, Johnson, Robinson, and Yared (2008) study the relationship between income per capita and democracy arguing that the strong cross-country correlation disappears after controlling for factors that simultaneously affect both variables. We replicate a regression provided as robustness to their main results, which has also been used in Kim and Wang (2019), and Bonhomme and Manresa (2015) to illustrate novel econometric techniques.[14]

We estimate a fixed effects OLS regression with balanced panel data that uses 5-year averaged data from 1970 to 2000 for 90 countries with the following simple econometric specification:

$$d_{it} = \alpha d_{it-1} + \gamma y_{it-1} + \mu_t + \delta_i + \epsilon_{it} \tag{38}$$

where d_{it} is the democracy score of country i in period t, y_{it-1} the lagged value of log income per capita, μ_t are time fixed effects, and δ_i country fixed effects.

Figure 12 reports the estimated γ together with confidence intervals computed with different assumptions.

Table 25 reports our pretest approach using F tests, which favors two-way clustering of the standard errors. This contrasts with clustering by country as done in the paper but the results remain consistent in the sense that the coefficient is not statistically significant.

Table 25: Pretesting using F tests and option e)

	F_1	p-value	F_2	p-value	Result
Statistics	10.27	0.00	2.32	0.00	Two-way

Notes: F_1 corresponds to a F statistic for the null hypothesis that all country fixed effects are zero, the associated p-values are reported next to them, and F_2 corresponds to a F statistic for the null hypothesis that all year fixed effects are zero, the associated p-values are reported next to them.

[14] The data is available online at http://qed.econ.queensu.ca/jae/2019-v34.6/kim-le/.

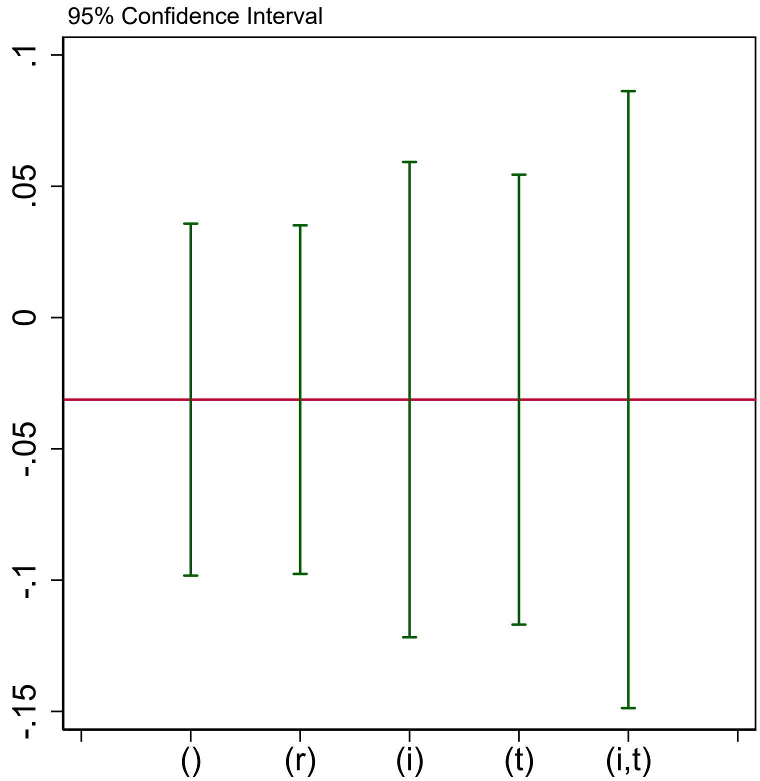


Figure 12: Confidence intervals for estimates for lagged gdp
Notes: *()*, *(r)*, *(i)*, *(t)*, *(i,t)* denotes respectively standard errors computed assuming iid disturbances, using heteroskedasticity robust s.e., clustering by state, clustering by year, and two-way clustering. The red line marks the value of the estimated coefficient.

7.3 Application 3

We use the data set from the Ph.D. dissertation of Y. Grunfeld (Univ. of Chicago, 1958), which has been used to illustrate the empirical use of panel data methods among others by Stata user manual, Greene (2012), and Baltagi (2013).[15] We estimate the following specification:

$$y_{it} = \alpha + \beta_1 v_{it} + \beta_2 k_{it} + \epsilon_{it} \tag{39}$$

where y_{it} corresponds to investment by firm i in year t, v_{it} the real value of the firm, k_{it} the real value of firm's capital stock. The data set spans 10 firms and 20 years. Figure 13 reports the confidence intervals constructed according to different methods.

[15]Data available at http://www.stata-press.com/data/r13/grunfeld.

We run our pretest strategy using F tests with this data set according to the strategy lined out in Figure 3 as option e) and find that it supports standard errors clustered by firm (See Table 26).

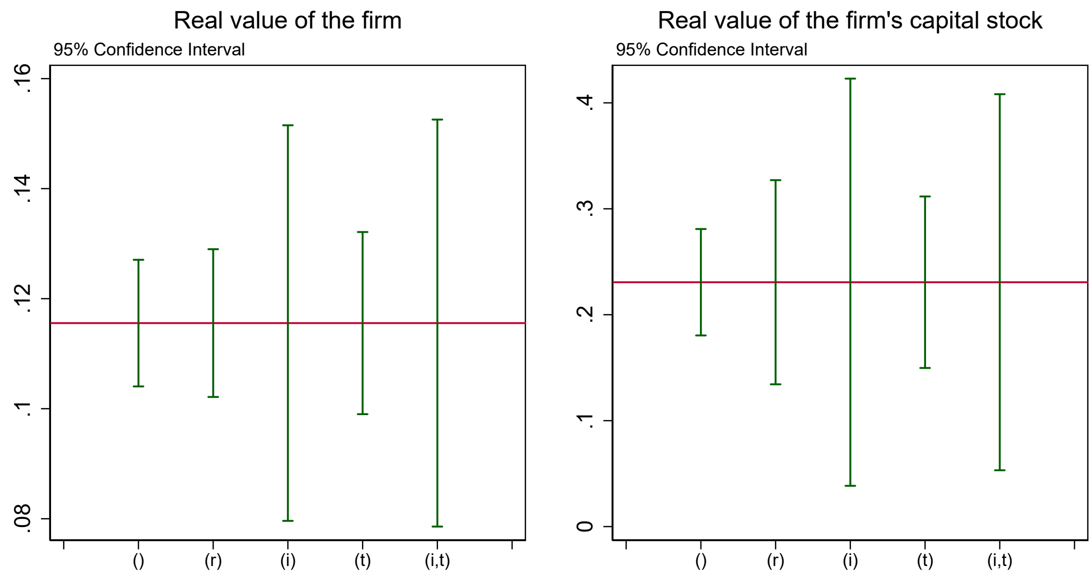


Figure 13: Confidence intervals for estimates in Grunfeld data set
Notes: *()*, *(r)*, *(i)*, *(t)*, *(i,t)* denotes respectively standard errors computed assuming iid disturbances, using heteroskedasticity robust s.e., clustering by state, clustering by year, and two-way clustering. The red line marks the value of the estimated coefficient.

Table 26: Pretesting using F tests and option e)

	F_1	p-value	F_2	p-value	Result
Statistics	1.40	0.13	52.36	0.00	Cluster by firm

Notes: F_1 corresponds to a F statistic for the null hypothesis that all firm fixed effects are zero, the associated p-values are reported next to them, and F_2 corresponds to a F statistic for the null hypothesis that all year fixed effects are zero, the associated p-values are reported next to them.

7.4 Application 4

Our paper have focused on the use of panel data, however we may also have to decide how to computer standard errors in the context of cross-sectional data.[16] Here we ilustrate the use of our pretest approach with the F test on a cross section. We use a data set that comes from a women sample of the National Longitudinal Survey, which has been used to illustrate econometrics methods for example in Roodman, MacKinnon, Nielsen, and Webb (2019). We estimate the following specification:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \epsilon_i \tag{40}$$

where y_i corresponds to wage, x_{1i} number of hours worked, x_{2i} total work experience, x_{3i} a dummy variable about black race , x_{4i} is a dummy variable indicating graduate status, and finally x_{5i} a dummy variable indicating whether lives in South. We pretest the model to decide how to cluster using industry and occupation.

Figure 14 reports the confidence intervals for the coefficients associated with these five regressors. Table 27 reports our pretest strategy, which favors two-way clustering by industry and occupation.

Table 27: Pretesting using F tests and option e)

	F_1	p-value	F_2	p-value	Result
Statistics	12.47	0.00	8.67	0.00	Two-way

[16] Data available at http://www.stata-press.com/data/r8/nlsw88.dta.

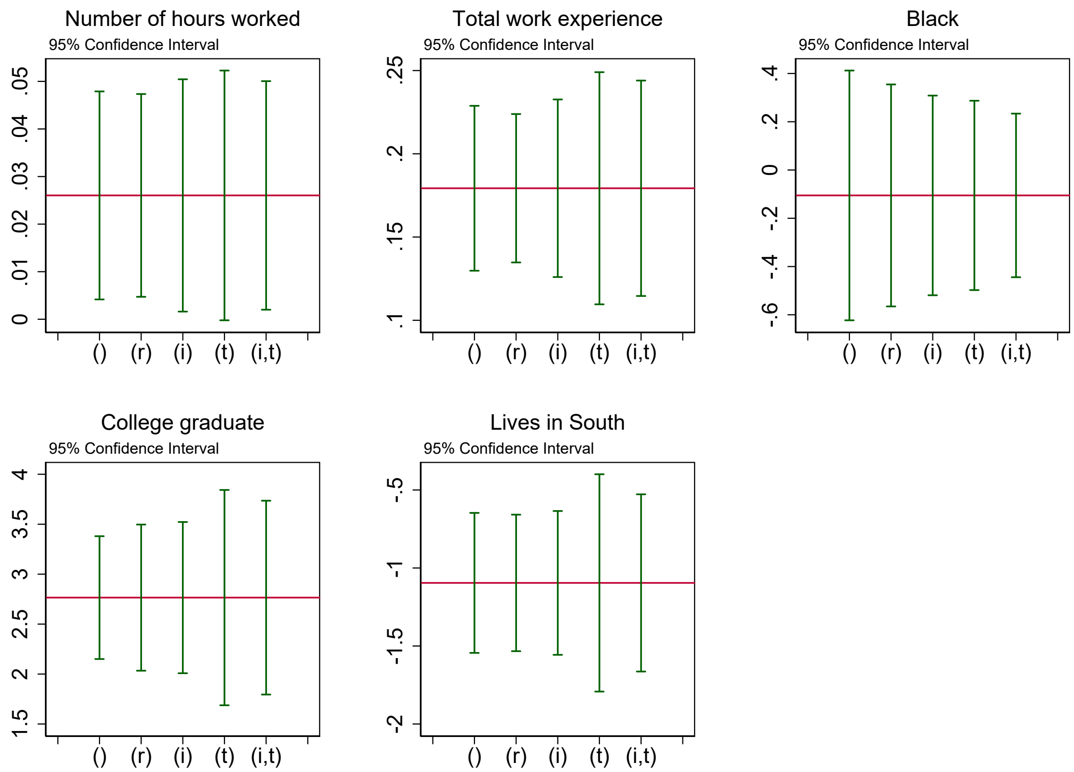


Figure 14: Confidence intervals for estimates in nlsw88 data set

Notes: *()*, *(r)*, *(i)*, *(t)*, *(i,t)* denotes respectively standard errors computed assuming iid disturbances, using heteroskedasticity robust s.e., clustering by state, clustering by year, and two-way clustering. The red line marks the value of the estimated coefficient.