CentER

Discussion Paper

No. 2010–110

ONE-STEP ROBUST ESTIMATION OF FIXED-EFFECTS PANEL DATA MODELS

By M. Aquaro, P. Čížek

September 2010

ISSN 0924-7815

TILBURG ◆ UNIVERSITY

One-step robust estimation of fixed-effects panel data models

M. Aquaro[*] and P. Čížek[†]

September 2010

Abstract

The panel-data regression models are frequently applied to micro-level data, which often suffer from data contamination, erroneous observations, or unobserved heterogeneity. Despite the adverse effects of outliers on classical estimation methods, there are only a few robust estimation methods available for fixed-effect panel data. Aiming at estimation under weak moment conditions, a new estimation approach based on two different data transformation is proposed. Considering several robust estimation methods applied on the transformed data, we derive the finite-sample, robust, and asymptotic properties of the proposed estimators including their breakdown points and asymptotic distribution. The finite-sample performance of the existing and proposed methods is compared by means of Monte Carlo simulations.

Keywords: breakdown point, fixed effects, panel data, robust estimation

JEL codes: C23

[*]Dept. of Econometrics & OR, Tilburg University, P.O.Box 90153, 5000LE Tilburg, The Netherlands. `Email:` M.Aquaro@uvt.nl.

[†]CentER, Dept. of Econometrics & OR, Tilburg University, P.O.Box 90153, 5000LE Tilburg, The Netherlands. `Email:` P.Cizek@uvt.nl.

1 Introduction

The panel-data regression models are increasingly popular in applications because each individual cross-sectional unit is observed over time, and consequently, the individual-specific heterogeneity can be accounted for. The majority of the regression methods used in linear panel-data models are based on linear estimators such as least squares (LS), and having unbounded normal equations, are very sensitive to data contamination and outliers (Ronchetti and Trojani, 2001). This sensitivity can be characterized by various measures of robustness such as the breakdown point, which measures the smallest contaminated fraction of a sample that can arbitrarily change the estimates (Genton and Lucas, 2003; Davies and Gather, 2005). Because the breakdown point of the linear estimators such as LS is asymptotically zero, many authors stressed the importance of robust and positive breakdown-point methods (e.g., Hampel et al., 1986; Simpson et al., 1992; Ronchetti and Trojani, 2001; Gervini and Yohai, 2002; Wagenvoort and Waldmann, 2002; Maronna et al., 2006; Čížek, 2008). This is even more important in the case of large panels, which can contain individuals with erroneous observations that are masked by the complex structure of the data.

Despite its relevance, the study of robust techniques for panel data seems to be rather limited. The works of Wagenvoort and Waldmann (2002) and Lucas et al. (2007) concentrate on the bounded-influence estimation of static and dynamic panel data models, respectively. Along with related quantile-regression estimation by Koenker (2004), these methods are generally locally robust, that is, their breakdown point can be arbitrarily close to zero for some kinds of data contamination. The positive breakdown-point methods were proposed only by Bramati and Croux (2007) and Dhaene and Zhu (2009), where the first concentrates on the static panel models and the latter on the dynamic panel models. Being interested in the static panel-data models here, Dhaene and Zhu (2009) aiming at dynamic models is not suitable, especially since it strictly relies on additional distributional assumptions (e.g., errors being normal or independent and identically distributed), which rule out heteroscedasticity and serial correlation of errors. On the other hand, the methods proposed by Bramati and Croux (2007) either are not equivariant with respect to various data transformations, for example rescaling of data, or have to explicitly estimate the fixed effects, causing bias due to the nonlinearity of the procedure if the number of periods is fixed (see Sections 2.2 and 4 for details). In both cases, the methods are consistent only if the number of time periods increases to infinity, which makes them unsuitable for short panels.

In this paper, we propose an alternative robust estimation approach for linear fixed-effect panel-data models that is equivariant with respect to standard data transformations, that is consistent for data observed in a (small) fixed number of time periods, and that, besides the standard identification assumptions, does not require any particular distributional assumptions (with the exception of the errors having a unimodal distribution). To achieve this, we employ two different data transformations and show that it is possible to apply standard robust estimators of linear regression to the transformed data. Because of the data transformations, the equivariance, robust, and asymptotic properties of the proposed estimators have to be established. All methods are shown to have a positive breakdown point equal to or converging to 1/4 and to have asymptotically a normal distribution. At the same time, Monte Carlo experiments indicate that the finite-sample performance of the proposed methods matches the standard within-group LS estimator and the robust properties thus do not adversely affect the precision of estimation.

The paper is organized as follows. After a survey of the existing fixed-effect panel-data estimators in Section 2, two data transformations and the corresponding robust estimators are proposed in Section 3, where their robust and asymptotic properties are also examined. The finite-sample properties are studied in Section 4. The proofs are given in the Appendix.

2 Panel data models

In this section, a brief account of some classical panel-data estimators is offered (Section 2.1), followed by the discussion of existing robust methods suitable for panel data (Section 2.2).

2.1 The fixed-effects model

A static linear fixed-effect panel-data model can be described by

$$y_{it} = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \qquad i = 1, \ldots, n, \qquad t = 1, \ldots, T, \tag{1}$$

where y_{it} denotes the dependent variable, $\boldsymbol{x}_{it} \in \mathbb{R}^p$ contains observable covariates, and the vector $\boldsymbol{\beta} \in \mathbb{R}^p$ represents the parameters of interest. The subscript i could refer to individuals, households, firms, or countries, whereas t indicates the periodicity. The unobservable terms consist of an unobservable individual-specific effect α_i and of the error term ε_{it}, which is assumed to have a zero mean, $\mathrm{E}\left(\varepsilon_{it}|\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}\right) = 0$, and to be independent across individuals; see Wooldridge (2002).

Without additional assumptions about the individial effects α_i and given a fixed number of observed time periods T, the estimation of $\boldsymbol{\beta}$ is straightforward only if α_i's are eliminated from the model equation. A standard procedure, based on the so-called within-group transformation, rules out the fixed effects by computing the time averages for each individual,

$$\bar{y}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \qquad \bar{\boldsymbol{x}}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}, \tag{2}$$

and then subtracting them from the original values: $\tilde{y}_{it} = y_{it} - \bar{y}_{i\cdot}$ and $\tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i\cdot}$. Model (1) then implies the linear relationship $\tilde{y}_{it} = \tilde{\boldsymbol{x}}_{it}^\top \boldsymbol{\beta} + \tilde{\varepsilon}_{it}$, which permits estimating the parameter vector $\boldsymbol{\beta}$ by the LS estimate $\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{LS,mean})}$. The within-group LS estimator is linear, which implies that it is equivariant with respect to scale, regression, and affine transformations: denoting the estimator explicitly as a function of data $\mathcal{T}^{\mathrm{LS}}(\{\boldsymbol{x}_{it}, y_{it}\}$, the scale, regression, and affine equivariance mean that $\mathcal{T}^{\mathrm{LS}}(\{\boldsymbol{x}_{it}, cy_{it}\}) = c\mathcal{T}^{\mathrm{LS}}(\{\boldsymbol{x}_{it}, y_{it}\})$, $\mathcal{T}^{\mathrm{LS}}(\{\boldsymbol{x}_{it}, y_{it} + \boldsymbol{x}_{it}^\top \boldsymbol{v}\}) = \mathcal{T}^{\mathrm{LS}}(\{\boldsymbol{x}_{it}, y_{it}\}) + \boldsymbol{v}$, and $\mathcal{T}^{\mathrm{LS}}(\{\boldsymbol{x}_{it}^\top \boldsymbol{A}, y_{it}\}) = \boldsymbol{A}^{-1}\mathcal{T}^{\mathrm{LS}}(\{\boldsymbol{x}_{it}, y_{it}\})$, respectively, for any $c \in \mathbb{R}, \boldsymbol{v} \in \mathbb{R}^p$, and $\boldsymbol{A} \in \mathbb{R}^{p\times p}$.

Unfortunately, the within-group LS estimator is very sensitive to erroneous observations and outliers as any linear regression LS method. To document this, let us introduce one of the global measures of robustness – the breakdown point. Informally, an estimator is said to break down when the procedure no longer conveys useful information on the data-generating mechanism (Genton and Lucas, 2003). In linear regression models, this general statement is equivalent to saying that the estimates can increase above any bound in the presence of data contamination. More formally, suppose we observe a random sample $Z = \{\boldsymbol{x}_{it}, y_{it}\}_{i=1,t=1}^{n,\ T}$ and let $\mathcal{T}$ be an estimator of the regression parameters estimating $\boldsymbol{\beta}$ by $\mathcal{T}(Z)$. The finite-sample breakdown point of $\mathcal{T}$ at the sample Z could be then be defined as the smallest fraction of data that can be modified so that the estimate increases above any bound (Rousseeuw and Leroy, 1987):

$$\varepsilon_{nT}^{*}\left(\mathcal{T}; Z\right) = \frac{1}{nT} \max_{m\geq 0} \left\{ m \mid \sup_{Z_m} \|\mathcal{T}(Z) - \mathcal{T}(Z_m)\| < \infty \right\}, \tag{3}$$

where the supremum is over all choices of Z_m consisting of $(nT - m)$ points from Z and m arbitrary points. The asymptotic breakdown point of $\mathcal{T}$ can be defined as the limit $\varepsilon^*(\mathcal{T}) = \lim_{n\to\infty} \varepsilon_{nT}^*(\mathcal{T}; Z)$, provided that this sample-independent limit exists. It can be at most 1/2 for regression equivariant estimators (cf. Davies and Gather, 2005). For the within-group LS estimator, which is scale, regression, and affine equivariant, the finite-sample breakdown point however does not exceed $1/nT$ and it converges to zero asymptotically.

2.2 Robust estimators for panel data

To the best of our knowledge, there are very few studies proposing robust estimators for panel data. Two of these, Koenker (2004) and Lucas et al. (2007), suggest estimators which are only locally robust, meaning that their breakdown points can be arbitrarily small for some data designs. Considering the globally robust estimators (i.e., having a positive breakdown point), the two existing contributions are Dhaene and Zhu (2009) and Bramati and Croux (2007). The first one proposes median-based estimators for dynamic fixed-effects models, which strictly require additional distributional assumptions such as errors being independent and identically distributed across all individuals and time periods and does not allow for heteroscedasticity and serial autocorrelation often encountered in static panel-data models. Thus, the only proposal generally applicable in static fixed-effect panel-data models stems from Bramati and Croux (2007), who adapt two existing high-breakdown point procedures and reach asymptotically a positive breakdown 1/4. We focus here on their within-group generalized M-estimator (WGM), since the other proposal – the MS-estimator of Maronna and Yohai (2000) – estimates the fixed-effects, and due to its two-step non-linear structure, would require a (non-existant) bias correction if the number of periods T is small.

The WGM estimator applies two robust estimators to centered data. As the mean used in the within-group transformation (2) is a non-robust location estimator, Bramati and Croux (2007) apply a robust regression estimator to data, where the individual fixed effects are eliminated by means of the median. Instead of transformation (2), variables are thus centered using the within-group medians:

$$\tilde{y}_{it} = y_{it} - \operatorname*{med}_t y_{it}, \qquad \tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it} - \operatorname*{med}_t \boldsymbol{x}_{it}. \tag{4}$$

After centering, a natural approach is to regress $\tilde{y}_{it}$ on $\tilde{\boldsymbol{x}}_{it}$ using a robust regression estimator. Bramati and Croux (2007) suggest to use first the least trimmed squares (LTS) estimator (Rousseeuw, 1984), which minimizes the sum of h_{nT} smallest residuals:

$$\hat{\boldsymbol{\beta}}_{nT}^{(\text{LTS, med},h_{nT})} = \operatorname*{arg\,min}_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{j=1}^{h_{nT}} r_{(j)}^{2,(\text{med})}(\boldsymbol{\beta}), \tag{5}$$

where h_{nT} is the trimming constant, $nT/2 < h_{nT} \leq nT$, and $r_{(j)}^{2,(\text{med})}(\boldsymbol{\beta})$ is the jth smallest order statistics of squared residuals $r_{it}^{2,(\text{med})}(\boldsymbol{\beta}) = (\tilde{y}_{it} - \tilde{\boldsymbol{x}}_{it}^\top \boldsymbol{\beta})^2$, $i = 1, \ldots, n$ and $t = 1, \ldots, T$. The trimming constant determines the number $nT - h_{nT}$ of observations, which are excluded from the objective function (5) and thus cannot directly influence the estimates. The LTS estimator attains maximum breakdown point when $h_{nT} = [nT/2] + [(p+1)/2]$ (Rousseeuw and Leroy, 1987), where $[x]$ denotes the integer part of x. The main disadvantage of this most robust choice of h_{nT} is the low relative efficiency of 8% for normal data.

Next, to improve this lack of efficiency, Bramati and Croux (2007) adopt the reweighted LS strategy using weights designed so that the breakdown point of the initial LTS estimator is preserved (Rousseeuw and Leroy, 1987). Let $\hat{\boldsymbol{\beta}}_{nT}^0$ and $\hat{\sigma}_{nT}^0$ be the regression and scale estimates obtained in the first estimation step using LTS (i.e., $(\hat{\sigma}_{nT}^0)^2$ is defined as a multiple of the LTS objective function at $\hat{\boldsymbol{\beta}}_{nT}^0$). The weighting scheme relies on two different kinds of weights. First, observations having large standardized residuals $r_{it}(\hat{\boldsymbol{\beta}}_{nT}^0)/\hat{\sigma}_{nT}^0$ are downweighted using residual weights $\hat{w}_{it}^r = w_c\{r_{it}(\hat{\boldsymbol{\beta}}_{nT}^0)/\hat{\sigma}_{nT}^0\}$, where Bramati and Croux (2007) use $w_c(u) = \{1-(u/c)^2\}^2 I(|u| \leq c)$

Table 1: The mean squared errors of the within-group LS and WGM estimates based on the mean and median transformations.

M		0		1		10	
# parameters		1	5	1	5	1	5
mean	LS	0.001	0.022	0.001	0.021	0.001	0.021
med	LS	0.001	0.017	0.003	0.065	0.002	4.272
	WGM	0.001	0.005	0.146	0.809	0.046	51.88

and $c = 4.685$. A further protection against observations with a high leverage is provided by the location weights indirectly proportional to the values of covariates: $\hat{w}_{it}^x = \min\{1, \sqrt{\chi^2_{p,0.975}}/RD_{it}\}$, where $\chi^2_{p,0.975}$ is the 97.5% quantile of the chi-square distribution with p degrees of freedom, $RD_{it} = [(\tilde{\boldsymbol{x}}_{it} - \hat{\boldsymbol{\mu}})^\top \hat{\boldsymbol{V}}^{-1}(\tilde{\boldsymbol{x}}_{it} - \hat{\boldsymbol{\mu}})]^{1/2}$ is a robust version of the Mahalanobis distance (Rousseeuw and Zomeren, 1990), and $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{V}}$ are robust estimates of the location and variance matrix of $\tilde{\boldsymbol{x}}_{it}$. The WGM estimator is then defined as the weighted LS (WLS) estimator for the median-transformed data

$$\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{WGM,\,med})} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n}\sum_{t=1}^{T} \hat{w}_{it}^r \hat{w}_{it}^x r_{it}^{2,(\mathrm{med})}(\boldsymbol{\beta}). \tag{6}$$

The complete WGM procedure can asymptotically achieve the breakdown point 1/4. On the other hand, WGM is neither regression nor affine equivariant and its asymptotic distribution (even for $T \to \infty$) has not been derived yet. The lack of equivariance properties comes from the nonlinearity of the median transformation and complicates the use of WGM in applications as we now demonstrate, at least if T is small. Consider the following linear panel-data model ($i = 1, \ldots, 100$; $t = 1, 2, 3$)

$$y_{it} = \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \tag{7}$$

where $\boldsymbol{x}_{it} \sim N(0,1)$, $\varepsilon_{it} \sim N(0,1)$, $\alpha_i \sim U(0,10)$, and $\boldsymbol{\beta} = -M \in \mathbb{R}$ or $\boldsymbol{\beta} =$

$(-M, 0, M, 0, -M)^\top \in \mathbb{R}^5$. Simulating the data 1000 times and estimating the model for $M = 0, 1$, and 10 by LS and WGM results in the mean squared errors in Table 1. Obviously, various levels of the multiplier M do not have any impact on the precision of the within-group LS estimator. Using LS and WGM after removing the individual effects by the median centering however leads to completely different results: the mean squared errors are substantially increasing with the magnitude of the regression coefficients, especially for the model with 5 variables.

3 New robust estimators for panel data

Because using the within-group transformation with LS is non-robust and using the (robust) median in place of the mean (Bramati and Croux, 2007) introduces inconsistency when the time dimension T is small and fixed, we now propose alternative robust estimators of $\boldsymbol{\beta}$ in (1) that do not rely on estimating the central tendency of fixed effects. This will be done in two steps. First, the elimination of unobserved individual effects will be addressed by considering other data transformations than the mean or median centering (Section 3.1). Second, in the light of recent contributions in robust statistical theory, LTS (Section 3.2) will be followed by new robust and efficient estimators adapted to the panel data setting (Section 3.3).

3.1 Data transformations

Since applying a robust estimate of location to centered data does not lead to a usable estimator (see Section 2.2), we focus on the first-difference and pairwise-difference transformations instead. The first-difference transformation is already well known in the literature (Wooldridge, 2002). Denoting the first-difference operator by Δ, the model (1) can be transformed to

$$\Delta y_{it} = y_{it} - y_{it-1} = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + \varepsilon_{it} - \boldsymbol{x}_{it-1}^{\top}\boldsymbol{\beta} - \varepsilon_{it-1} = \Delta\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + \Delta\varepsilon_{it}, \tag{8}$$

where $i = 1, \ldots, n$ and $t = 2, \ldots, T$ and where no fixed effects α_i appear. Under the strict-exogeneity assumption, $\boldsymbol{\beta}$ is consistently estimated by LS applied to (8). This alternative to the within-group estimator, which is the best linear unbiased estimator when error terms ε_{it} are uncorrelated, is preferable if error terms ε_{it} are serially correlated (see Wooldridge, 2002, for details).

Alternatively, one could try to obtain more accurate estimates than from (8) by eliminating individual effects by taking all pairwise differences within each individual. Inspired by Stromberg et al. (2000) and Honoré and Powell (2005), let us define the pairwise-difference transformation as $\Delta^s z_{it} = z_{it} - z_{it-s}$, where $s = 1, \ldots, t-1$, for any $t \in \{2, \ldots, T\}$ and $i \in \{1, \ldots, n\}$. Applied to model (1), the pairwise-difference transformation yields

$$\Delta^s y_{it} = y_{it} - y_{it-s} = (\boldsymbol{x}_{it} - \boldsymbol{x}_{it-s})^{\top}\boldsymbol{\beta} + \varepsilon_{it} - \varepsilon_{it-s} = \Delta^s\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + \Delta^s\varepsilon_{it}, \tag{9}$$

which removes the individual-specific variable α_i similarly to (8), but generates a larger sample size $nT(T-1)/2$ instead of $n(T-1)$ in (8) since differences for all $s = 1, \ldots, t-1$ are considered.

To handle all transformations in a unified way, let us now introduce a more general notation. Given the original data set $\{\boldsymbol{x}_{it}, y_{it}\}_{i=1,t=1}^{n,\ T}$, let $\{\tilde{\boldsymbol{x}}_{it}, \tilde{y}_{it}\}_{i=1,t=1}^{n,\ T^{(\mathfrak{T})}}$ be the data set created by one of the considered data transformations $\mathfrak{T}$, $\mathfrak{T} \in \{\text{med}, 1\Delta, P\Delta\}$, where med, 1Δ, and $P\Delta$ are shorthand symbols for the median-centering, first-difference, and pairwise-difference transformation and $T^{(\mathfrak{T})} = T, T-1$, and $T(T-1)/2$, respectively.

3.2 Initial robust estimator

Once the individual effects have been eliminated, it is of interest to find a proper robust estimator for $\boldsymbol{\beta}$ in (1). Similarly to Bramati and Croux (2007), we use initially the LTS estimator, which may be generally defined for the $\mathfrak{T}$-transformed data as

$$\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{LTS},\mathfrak{T},h_{nT})} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{j=1}^{h_{nT}} r_{(j)}^{2,(\mathfrak{T})}(\boldsymbol{\beta}), \tag{10}$$

where $r_{(j)}^{2,(\mathfrak{T})}(\boldsymbol{\beta})$ is the jth smallest order statistics of squared residuals, the (i,t)th residual equals $r_{it}^{(\mathfrak{T})}(\boldsymbol{\beta}) = \tilde{y}_{it} - \tilde{\boldsymbol{x}}_{it}^{\top}\boldsymbol{\beta}$, and h_{nT} is the trimming constant, $nT^{(\mathfrak{T})}/2 < h_{nT} \leq nT^{(\mathfrak{T})}$. We assume that the trimming constant is defined so that $h_{nT}/nT^{(\mathfrak{T})} \to \lambda \in \langle 1/2, 1\rangle$, and thus asymptotically, the $1-\lambda$ fraction of observations is eliminated from the objective function (10). To study the breakdown properties of the proposed LTS estimation under different transformations, let us make the following assumptions.

Assumption D

D1 Let $\{\boldsymbol{x}_{it}, y_{it}\}_{i=1,t=1}^{n,\ \ T}$ be a random sample generated according to model (1).

D2 The transformed data $\{\tilde{\boldsymbol{x}}_{it}, \tilde{y}_{it}\}_{i=1,t=1}^{n,\ \ T^{(\mathfrak{T})}}$ are almost surely in a general position for $nT^{(\mathfrak{T})} > 3(p+1)$, that is, any $p+1$ data points do not lie on the same hyperplane almost surely.

Contrary to the median centering, both the first-difference and pairwise-difference transformations are linear transformations of the data. Therefore, the LTS estimator applied to such transformed data does not lose its equivariance properties contrary to LTS applied to the median-transformed data in Bramati and Croux (2007).

Lemma 1 *Suppose that Assumption D1 holds. If $\mathfrak{T} \in \{1\Delta, P\Delta\}$, then the LTS estimator $\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{LTS},\mathfrak{T},h_{nT})}$ defined in (10) is scale, affine, and regression equivariant.*

Further, let now look at the breakdown properties of the LTS estimator.

Theorem 1 *Suppose that Assumption D hold. Let* $\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{LTS},\mathfrak{T},h_{nT})}$ *be the* LTS *estimator defined in* (10) *for* $h_{nT}/(nT^{(\mathfrak{T})}) \to \lambda$ *as* $nT^{(\mathfrak{T})} \to \infty$. *If* $h_{nT} \geq \underline{h}_{nT}^{T^{(\mathfrak{T})}} = [nT^{(\mathfrak{T})}/2] + [(p+1)/2] + 1$*, then it holds that*

$$\varepsilon_{nT}^{*}\left(\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{LTS},\mathfrak{T},h_{nT})}; \{\boldsymbol{x}_{it}, y_{it}\}_{i=1,t=1}^{n,\ \ T}\right) \geq \frac{nT^{(\mathfrak{T})} - h_{nT}}{2nT^{(\mathfrak{T})}} \cdot \kappa^{(\mathfrak{T})}(T), \tag{11}$$

where $\kappa^{(1\Delta)} = [2(T-1)]/[\min\{2, T-1\}T]$ *and* $\kappa^{(P\Delta)} = 1$. *The breakdown point of LTS tends asymptotically to* $\kappa^{(\mathfrak{T})}(T)(1-\lambda)/2$*, and in particular, to* $\kappa^{(\mathfrak{T})}(T)/4$ *for* $h_{nT} = \underline{h}_{nT}^{T^{(\mathfrak{T})}}$.

From the breakdown point of view, both proposed data transformations are asymptotically equivalent for $T = 2$ and for $T \to \infty$ as they yield the same maximum breakdown point 1/4 analogously to Bramati and Croux (2007). Whereas the pairwise differencing reaches this breakdown point for any number of time periods T, the first differencing has a smaller breakdown point equal to $(T-1)/(4T)$ for $T \geq 3$. Let us note that this disadvantage of the first differencing could be eliminated by considering only differences at even time periods (e.g., $\Delta y_2, \Delta y_4, \ldots$): the breakdown point would asymptotically equal 1/4 irrespective of T, but the precision of estimation would suffer.

3.3 Robust and efficient estimation

Since the LTS estimator with the maximum breakdown point achieves only 8% relative efficiency for normally distributed data, one-step estimators are often employed to improve the precision of estimation without substantially affecting the robust properties of estimation (see also Section 2.2).

To introduce the efficient one-step methods, suppose we have the transformed data $\{\tilde{\boldsymbol{x}}_{it}, \tilde{y}_{it}\}$ obtained by transformation $\mathfrak{T} \in \{\mathrm{med}, 1\Delta, P\Delta\}$ and a pair of initial robust estimators of the regression parameters $\hat{\boldsymbol{\beta}}^0_{nT}$ and residual scale $\hat{\sigma}^0_{nT}$ (e.g., the median absolute deviation). A classical example of a one-step augmentation procedure is the iteratively reweighted LS (IRLS) estimator proposed by Rousseeuw and Leroy (1987), which removes the observations having large absolute residuals according to some initial robust fit and then applies LS. Denoting the initial residuals $r^{(\mathfrak{T})}_{it}(\hat{\boldsymbol{\beta}}^0_{nT}) = \tilde{y}_{it} - \tilde{\boldsymbol{x}}^\top_{it}\hat{\boldsymbol{\beta}}^0_{nT}$, one can thus define weights

$$\hat{w}_{it}\left(\hat{\boldsymbol{\beta}}^0_{nT}, \hat{\sigma}^0_{nT}; v\right) = I\left(|r^{(\mathfrak{T})}_{it}(\hat{\boldsymbol{\beta}}^0_{nT})/\hat{\sigma}^0_{nT}| < v\right) \tag{12}$$

for a constant $v > 0$ (e.g., Gervini and Yohai (2002) suggest $v = 2.5$). The IRLS estimator then reads as follows:

$$\hat{\boldsymbol{\beta}}^{(\text{IRLS},\mathfrak{T})}_{nT} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n}\sum_{t=1}^{T^{(\mathfrak{T})}} \hat{w}_{it}\left(\hat{\boldsymbol{\beta}}^0_{nT}, \hat{\sigma}^0_{nT}; v\right) r^{2,(\mathfrak{T})}_{it}(\boldsymbol{\beta}). \tag{13}$$

A data-adaptive version of (13) designed to achieve efficiency for normally distributed data, the robust and efficient weighted least squares (REWLS) estimator, has been proposed by Gervini and Yohai (2002). A data-dependent cut-off point $\hat{v}_{nT}$ to define weights (12) is now determined by comparing two distribution functions, F^+ and F^+_0, where the former relates to the standardized absolute residuals $|r^{(\mathfrak{T})}_{it}(\hat{\boldsymbol{\beta}}^0_{nT})/\hat{\sigma}^0_{nT}|$ and the latter is the distribution function assumed for these standardized absolute residuals in the model (1). Since F^+ is usually unknown, it is estimated by the empirical distribution function F^+_{nT} of $|r^{(\mathfrak{T})}_{it}(\hat{\boldsymbol{\beta}}^0_{nT})/\hat{\sigma}^0_{nT}|$. The maximum discrepancy $\hat{d}_{nT}$ between F^+_{nT} and F^+_0 in the tail of the distributions can be then measured by

$$\hat{d}_{nT} = \sup_{v\geq\eta}\left\{\left[F^+_0(v) - F^+_{nT}(v)\right] \cdot I\left(F^+_0(v) - F^+_{nT}(v) \geq 0\right)\right\}, \tag{14}$$

where η is a large quantile of F^+_0, for example, $\eta = 2.5$ for Gaussian errors with

$F_0 \equiv N(0,1)$ (see Gervini and Yohai, 2002). The cutoff point $\hat{v}_{nT}$ is then defined as the $(1-\hat{d}_{nT})$th quantile of the distribution F^+_{nT}: $\hat{v}_{nT} = \min\left\{v \mid F^+_{nT}(v) \geq 1-d_0\right\}$. Finally, the REWLS estimator is obtained from (13) for $v = \hat{v}_{nT} \geq \eta$:

$$\hat{\boldsymbol{\beta}}_{nT}^{(\text{REWLS},\mathfrak{T})} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n} \sum_{t=1}^{T^{(\mathfrak{T})}} \hat{w}_{it}\left(\hat{\boldsymbol{\beta}}^0_{nT}, \hat{\sigma}^0_{nT}; \hat{v}_{nT}\right) r_{it}^{2,(\mathfrak{T})}(\boldsymbol{\beta}). \tag{15}$$

This method is proved to preserve the breakdown-point properties of the initial robust estimator and achieve the asymptotic efficiency for Gaussian errors.

An alternative to the traditional one-step estimators is the reweighted least trimmed squares (RLTS) estimator (Čížek, 2010). Similarly to Gervini and Yohai (2002), weights (12) are constructed using the data-dependent cutoff point $\hat{v}_{nT}$. The resulting weights are however used within the LTS estimator rather than LS. Since LTS requires only the total number h_{nT} of observations to be included in the objective function, the number of observations with non-zeros weights $\hat{w}_{it}(\cdot,\cdot;\hat{v}_{nT})$ has to be counted:

$$\hat{h}_{nT} = \sum_{i=1}^{n} \sum_{t=1}^{T^{(\mathfrak{T})}} I\left(\left|r_{it}^{(\mathfrak{T})}\left(\hat{\boldsymbol{\beta}}^0_{nT}\right)/\hat{\sigma}^0_{nT}\right| < \hat{v}_{nT}\right) = \sum_{i=1}^{n} \sum_{t=1}^{T^{(\mathfrak{T})}} \hat{w}_{it}\left(\hat{\boldsymbol{\beta}}^0_{nT}, \hat{\sigma}^0_{nT}; \hat{v}_{nT}\right) \tag{16}$$

The RLTS estimator is then simply defined as LTS using the data-dependent amount of trimming $\hat{h}_{nT}$ applied to the $\mathfrak{T}$-transformed panel data:

$$\hat{\boldsymbol{\beta}}_{nT}^{(\text{RLTS},\mathfrak{T})} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{j=1}^{\hat{h}_{nT}} r_{(j)}^{2,(\mathfrak{T})}(\boldsymbol{\beta}). \tag{17}$$

Similarly to REWLS, RLTS preserves the breakdown-point properties of the initial robust estimator. Additionally, RLTS is asymptotically independent of the initial estimator and achieves asymptotic efficiency when errors are normally distributed.

Let us now formally state the breakdown properties of these one-step estimators.

Theorem 2 *Assume that Assumption D holds and that the data have been transformed according to one of the $\mathfrak{T}$-transformations, $\mathfrak{T} \in \{1\Delta, P\Delta\}$. Further, let ε^{0*}_{nT} be the finite-sample breakdown point of the initial estimator $\hat{\boldsymbol{\beta}}^0_{nT}$ of the regression parameters with limit $\varepsilon^{0*} = \lim_{n\to\infty} \varepsilon^{0*}_{nT}$. Additionally, suppose $\hat{\sigma}^0_{nT} = \mathrm{MAD}_{i,t} r_{it}(\hat{\boldsymbol{\beta}}^0_{nT})/\Phi^{-1}(3/4)$ is the standardized median absolute deviation estimator and F_0 has a finite variance. Then it holds that $\varepsilon^*_{nT}(\hat{\boldsymbol{\beta}}^{(\mathrm{IRLS},\mathfrak{T})}_{nT}) \geq \varepsilon^{0*}_{nT}$, $\varepsilon^*_{nT}(\hat{\boldsymbol{\beta}}^{(\mathrm{REWLS},\mathfrak{T})}_{nT}) \geq \varepsilon^{0*}_{nT}$, and $\varepsilon^*_{nT}(\hat{\boldsymbol{\beta}}^{(\mathrm{RLTS},\mathfrak{T})}_{nT}) \geq \varepsilon^{0*}_{nT}$.*

Thus, we see that all one-step methods – WGM, REWLS and RLTS – have the same breakdown properties. Note that this holds even though IRLS does not use weights $\hat{w}^x_{it}$ in constrast to WGM. The different methods could differ though by the bias caused by outliers and in their finite-sample and asymptotical variances.

3.4 Asymptotic properties

The estimators introduced in the previous sections are applied to model (1) after the first-difference or pairwise-difference transformations, which lead to the serial correlation of the errors in (8) or (9), respectively. Almost all robust regression estimators are however asymptotically studied under the assumption of independent (and often identically distributed) errors, be it in the context of cross-sectional (Gervini and Yohai, 2002) or panel data (Lucas et al., 2007), or there are no asymptotic results available (Bramati and Croux, 2007; Dhaene and Zhu, 2009). This limits also the extent to which we can characterize the asymptotic distribution of the proposed estimators. In particular, the asymptotic distribution under the first- and pairwise-differences can be easily derived only for the initial LTS estimator and its reweighted form RLTS (with the notable exception of the estimation based only on the first differences taken at even time periods as mentioned in Section 3.2, which produces independent errors).

Now, the assumptions necessary to derive the asymptotic distribution of LTS and RLTS are presented. To this end, let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})^\top$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})^\top$, $\tilde{\boldsymbol{X}}_i =$

$(\tilde{\boldsymbol{x}}_{i1}, \ldots, \tilde{\boldsymbol{x}}_{iT^{(\mathfrak{T})}})^\top$, $\tilde{\boldsymbol{y}}_i = (\tilde{y}_{i1}, \ldots, \tilde{y}_{iT^{(\mathfrak{T})}})^\top$, and $\tilde{\varepsilon}_{it} = \tilde{y}_{it} - \tilde{\boldsymbol{x}}_{it}\boldsymbol{\beta}^0$ for all $i \in \mathbb{N}$ and $t = 1, \ldots, T^{(\mathfrak{T})}$, where $\boldsymbol{\beta}^0$ is the true parameter value in model (1). Further, let us recall that, in this context, $\lambda \in \langle 1/2, 1\rangle$ refers to the limits $\lim_{n\to\infty} h_{nT}/nT^{(\mathfrak{T})}$ or $\lim_{n\to\infty} \hat{h}_{nT}/nT^{(\mathfrak{T})}$, see (16), and that $T \geq 2$ is a fixed integer. The assumptions and the asymptotic distribution will be stated for symmetrically distributed errors for the sake of simplicity. A more general result can be found in Čížek (2010), where a detailed discussion of these assumptions can be found.

Assumption A

A1 Random vectors $\boldsymbol{y}_i$ and matrices $\boldsymbol{X}_i$ are independent and identically distributed for all $i \in \mathbb{N}$ and have finite second moments.

A2 Let $\{\varepsilon_{it}\}_{i\in\mathbb{N}}$ be a sequence of random variables with finite second moments and $\mathrm{E}(\varepsilon_{it}|\boldsymbol{X}_i) = 0$ for all $i \in \mathbb{N}$ and $t = 1, \ldots, T$. Further, the unconditional distribution function F of ε_{it} is assumed to be unimodal, absolutely continuous, and symmetrically distributed condionally on $\boldsymbol{X}_i$. Its density function has to be bounded and continuously differentiable.

A3 Let $Q(\lambda) = \mathrm{E}[\tilde{\boldsymbol{X}}_i^\top \operatorname{diag}(\{I[|F(\tilde{\varepsilon}_{it}) - F(-\tilde{\varepsilon}_{it} - 2C)| \leq \lambda]\}_{t=1}^{T^{(\mathfrak{T})}})\tilde{\boldsymbol{X}}_i]$ be a nonsingular matrix for any fixed $C \in \mathbb{R}$.

A4 Denoting G_β and g_β the unconditional cumulative distribution and density functions of $(\tilde{y}_{it} - \tilde{\boldsymbol{x}}_{it}^\top\boldsymbol{\beta})^2$, let $\sup_{\beta\in\mathbb{R}^p}\sup_{z>\alpha} g_\beta(z) < \infty$ for any $\alpha > 0$, and if $\lambda < 1$, that $\inf_{\beta\in\mathbb{R}^p}\inf_{z\in(-\delta,\delta)} g_\beta\left(G_\beta^{-1}(\lambda) + z\right) > 0$ for some $\delta > 0$.

Assumption A1 formulates standard conditions of the (uniform) central limit theorem: observed variables are independent across cross-sectional units and have finite second moments. Assumption A2 presents the assumptions on the error term ε_{it}, which is independent of explanatory variables and continuously distributed. Note that, in the most general case, only the second moments of the trimmed errors $\boldsymbol{e}_i(q_\lambda)$ defined below have to be finite (see Čížek, 2011). Next, Assumption A3 formulates an analog of the standard full-rank condition, and is actually equivalent to $\mathrm{E}(\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^\top) > 0$ if $\tilde{\varepsilon}_{it}$ is independent of $\boldsymbol{X}_i$. Finally, Assumption A4 formalizes the fact that the distribution of squared residuals should be absolutely continuous: its density should not approach ∞ at any point, which would correspond to the distribution becoming discontinuous at some point. If $\tilde{\varepsilon}_{it}$ is independent of $\tilde{\boldsymbol{x}}_{it}$, Assumption A4 is usually implied by F being absolutely continuous with a density function f positive, bounded and differentiable (Čížek, 2006).

Under Assumption A, Čížek (2010) derived the below stated result regarding the asymptotic distribution of LTS and RLTS. To formulate this result, the notation $q_\lambda = \sqrt{G^{-1}(\lambda)}$ is used, where $G \equiv G^0_\beta$ and G^{-1} represents the unconditional quantile function of $\tilde{\varepsilon}^2_{it}$. Additionally, one diagonal matrix and two vectors depending on q_λ are needed: $\boldsymbol{I}_i(q_\lambda) = \mathrm{diag}[\{I(\tilde{\varepsilon}_{it} \le q_\lambda)\}_{t=1}^{T^{(\mathfrak{T})}}]$, $\boldsymbol{e}_i(q_\lambda) = \boldsymbol{I}_i(q_\lambda)(\tilde{\varepsilon}_{i1}, \ldots, \tilde{\varepsilon}_{iT^{(\mathfrak{T})}})^\top$, and $\boldsymbol{f}_i(q_\lambda) = (f_{i1}(q_\lambda), \ldots, f_{iT^{(\mathfrak{T})}}(q_\lambda))^\top$, where f_{it} is the conditional distribution of $\tilde{\varepsilon}_{it}|\tilde{\boldsymbol{X}}_i$.

Theorem 3 *Let Assumption A hold. Next, let* $\Sigma(\lambda) = \mathrm{E}[\tilde{\boldsymbol{X}}_i^\top\{\boldsymbol{e}_i(q_\lambda)\boldsymbol{e}_i(q_\lambda)^\top\}\tilde{\boldsymbol{X}}_i]$, $Q(\lambda) = \mathrm{E}[\tilde{\boldsymbol{X}}_i^\top\boldsymbol{I}_i(q_\lambda)\tilde{\boldsymbol{X}}_i]$, $J(\lambda) = -\mathrm{E}[q_\lambda\tilde{\boldsymbol{X}}_i^\top\,\mathrm{diag}\{\boldsymbol{f}_i(-q_\lambda)+\boldsymbol{f}_i(q_\lambda)\}\tilde{\boldsymbol{X}}_i]$, *and* $Q(\lambda)+J(\lambda)$ *be a non-singular matrix. Then the (reweighted) LTS estimator* $\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{RLTS},\mathfrak{T})}$ *defined by trimming* $\hat{h}_{nT}$ *such that* $\lim_{n\to\infty}\hat{h}_{nT}/nT \to \lambda$ *for some* $\lambda \in \langle 1/2, 1)$ *is a* $\sqrt{n}$*-consistent and asymptotically normal,* $\sqrt{n}(\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{RLTS},\mathfrak{T})} - \beta^0) \xrightarrow{\mathcal{L}} N(0, V(\lambda))$ *as* $n \to \infty$*, where the asymptotic covariance matrix equals* $V(\lambda) = \{Q(\lambda)+J(\lambda)\}^{-1}\Sigma(\lambda)\{Q(\lambda)+J(\lambda)\}^{-1}$.

The theorem covers not only the reweighted, but also the initial LTS estimator for $\hat{h}_{nT} = h_{nT} = \mathrm{const}$. Consequently, the initial and reweighted LTS estimators are asymptotically normal. The estimation of their covariance matrix $V(\lambda)$ is discussed in detail by Čížek (2011).

4 Simulation study

This section contains a simulation study of the finite-sample properties of some proposed and existing panel-data estimators. The following simulations are meant to investigate the behavior of estimators when the sample dimensions vary (Section 4.1), when errors come from various error distributions (Section 4.2), and when different kinds of outlying observations are present (Section 4.3). The reference estimator is the within-group estimator $\hat{\boldsymbol{\beta}}_{nT}^{(\mathrm{LS,\,mean})}$. Other estimators under consideration are the LS, LTS with the maximum amount of trimming (see Theorem 1), WGM of Bramati and Croux (2007), IRLS, REWLS, and RLTS estimators subject to three data transformations $\mathfrak{T} \in \{\mathrm{med}, 1\Delta, P\Delta\}$. Let us recall that WGM and IRLS are both based on the same reweighted LS method, but differ by employed weights: WGM uses a continous weighting function and downweights observations with large covariates, whereas IRLS uses 0–1 weights determined only by absolute residuals similarly to REWLS and RLTS.

The data generating process is given by a static fixed-effect panel-data model

$$\begin{aligned} y_{it} &= \boldsymbol{x}_{it}^\top \boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \\ \alpha_i &= \sum_{t=1}^{T} \boldsymbol{x}_{it}^\top \boldsymbol{\gamma}/\sqrt{T} + \eta_i, \end{aligned} \qquad i = 1, \cdots, n, \quad t = 1, \cdots, T, \tag{18}$$

where the ε_{it}'s are independent and identically distributed according to some distribution H. The parameters of interest are chosen $\boldsymbol{\beta} = (1, 0, -1)^\top$. The unobservable individual effects α_i depend on $\eta_i \sim \mathrm{U}(0, 12)$ and on the covariates $\boldsymbol{x}_{it}$ through $\boldsymbol{\gamma} = (2, 2, 2)^\top$. Observable covariates $\boldsymbol{x}_{it}$'s are generated according to

$$x_{itk} \sim \begin{cases} \chi_2^2 - 2 & \text{if } k = 1, \\ \mathrm{N}(0, 1) & \text{if } k \geq 2, \end{cases} \tag{19}$$

Table 2: The mean squared errors of all estimators for normally distributed errors and various sample sizes.

n		50	100	200	400	50			
T		3				4	6	12	24
mean	LS	0.023	0.011	0.006	0.003	0.015	0.009	0.004	0.002
med	LS	0.044	0.032	0.026	0.022	0.018	0.012	0.006	0.002
	LTS	1.423	1.540	1.579	1.605	0.185	0.140	0.061	0.028
	WGM	0.511	0.465	0.441	0.433	0.045	0.032	0.014	0.006
	IRLS	0.612	0.572	0.549	0.541	0.050	0.031	0.012	0.006
	REWLS	0.581	0.530	0.502	0.493	0.044	0.025	0.008	0.003
	RLTS	0.309	0.252	0.221	0.211	0.033	0.019	0.007	0.003
1Δ	LS	0.029	0.014	0.007	0.004	0.020	0.013	0.006	0.003
	LTS	0.173	0.098	0.056	0.030	0.122	0.084	0.045	0.024
	WGM	0.042	0.019	0.009	0.005	0.028	0.017	0.008	0.004
	IRLS	0.043	0.018	0.008	0.004	0.027	0.016	0.008	0.003
	REWLS	0.042	0.017	0.008	0.004	0.026	0.015	0.007	0.003
	RLTS	0.038	0.016	0.007	0.004	0.024	0.014	0.007	0.003
$P\Delta$	LS	0.023	0.011	0.006	0.003	0.015	0.009	0.004	0.002
	LTS	0.139	0.076	0.042	0.023	0.084	0.043	0.014	0.005
	WGM	0.032	0.015	0.007	0.004	0.019	0.011	0.005	0.002
	IRLS	0.032	0.014	0.006	0.003	0.018	0.010	0.005	0.002
	REWLS	0.031	0.013	0.006	0.003	0.017	0.009	0.005	0.002
	RLTS	0.028	0.013	0.006	0.003	0.016	0.009	0.005	0.002

where x_{itk} denotes the kth component of $\boldsymbol{x}_{it}$, $k = 1, 2, 3$, χ^2_2 denotes the chi-squared distribution with 2 degrees of freedom, and $N(0, 1)$ represents the standard normal distribution.

Simulation experiments are conducted across different sample sizes nT, aiming at both short micro-panels and long macro-panels, with n and T ranging from $(n, T) = (100, 3)$ to $(n, T) = (50, 24)$. The performance of each estimator is evaluated using $S = 1000$ simulated samples and is measured by the mean squared error (MSE): $MSE = 1/S \sum_{s=1}^{S} \|\hat{\boldsymbol{\beta}}^s_{nT} - \boldsymbol{\beta}\|^2$, where $\hat{\boldsymbol{\beta}}^s_{nT}$, $s = 1, \ldots, S$, are the estimates for S simulated samples.

4.1 Sample sizes

The performance of the estimators is first evaluated for normal errors, $H \equiv N(0,1)$, at different sample sizes: for T fixed and n increasing and for T increasing while n is fixed. The simulation results are summarized in Table 2. The results for the median transformation confirm that the robust estimators based on this transformation are not consistent for a fixed number of time periods T, but are consistent if $T \to \infty$. Next, LTS performs much worse than LS for all transformations, while all one-step estimators (WGM, IRLS, REWLS, and RLTS) exhibit much smaller MSEs and can match the performance of LS if the sample size is sufficiently large. Finally, it is interesting to note that – while the within-group LS estimator outperforms the LS applied to first-differenced data (errors are iid) – the LS and one-step robust estimators applied to pairwise-differenced data can actually match the performance of the within-group LS estimator.

4.2 Different error distributions

In this subsection, three different distributions H of the error term ε_{it} in (18) are considered: the standard normal $N(0,1)$, the double exponential distribution $DExp(1)$ with rate 1, and the Student distribution t_3 with 3 degrees of freedom, see Table 3. The LS estimator is no longer optimal and is slightly outperformed by one-step robust estimators in the case of the double-exponential errors and more substantially in the case of the Student errors (the differences among WGM, IRLS, REWLS, and RLTS are practically negligible). Regarding the data transformations, the pairwise-differencing leads uniformly to the best results, and in combination with REWLS, is preferable to the within-group LS estimator.

Table 3: The mean squared errors of all estimators for errors from the standard normal, double exponential, and Student distributions.

Errors distr.		DExp(1)			N(0, 1)			t_3		
n		200	75	30	200	75	30	200	75	30
T		3	8	20	3	8	20	3	8	20
mean	LS	0.011	0.009	0.008	0.006	0.004	0.004	0.017	0.012	0.012
med	LS	0.032	0.011	0.009	0.026	0.007	0.005	0.039	0.014	0.013
	LTS	1.641	0.078	0.032	1.582	0.083	0.048	1.650	0.089	0.045
	WGM	0.506	0.023	0.012	0.441	0.018	0.010	0.514	0.025	0.013
	IRLS	0.630	0.023	0.012	0.551	0.016	0.009	0.636	0.024	0.013
	REWLS	0.580	0.018	0.009	0.504	0.011	0.006	0.586	0.018	0.010
	RLTS	0.233	0.014	0.009	0.222	0.009	0.005	0.241	0.014	0.009
1Δ	LS	0.014	0.013	0.012	0.007	0.006	0.006	0.022	0.018	0.018
	LTS	0.048	0.036	0.034	0.056	0.046	0.039	0.059	0.047	0.044
	WGM	0.013	0.011	0.010	0.010	0.008	0.007	0.015	0.012	0.012
	IRLS	0.014	0.012	0.011	0.009	0.007	0.006	0.016	0.012	0.013
	REWLS	0.013	0.012	0.010	0.008	0.007	0.006	0.015	0.012	0.012
	RLTS	0.014	0.012	0.011	0.008	0.007	0.006	0.015	0.012	0.012
$P\Delta$	LS	0.011	0.009	0.008	0.006	0.004	0.004	0.017	0.012	0.012
	LTS	0.035	0.014	0.008	0.042	0.018	0.010	0.044	0.018	0.011
	WGM	0.010	0.007	0.006	0.007	0.005	0.004	0.012	0.007	0.007
	IRLS	0.011	0.007	0.006	0.007	0.005	0.004	0.012	0.007	0.007
	REWLS	0.010	0.007	0.006	0.006	0.004	0.004	0.011	0.007	0.007
	RLTS	0.011	0.008	0.007	0.006	0.004	0.004	0.012	0.008	0.007

4.3 Outliers

The robust properties are now evaluated by including outliers in the data. Let m be the number of outliers and let I_m be the index set of contaminated observations. Contaminated values of the dependent variable $\breve{y}^r_{it} \sim U(-10, 30)$ or $\breve{y}^c_{it} \sim U(29, 30)$ and independent variables $\breve{x}_{itk} \sim N(6, 2)$, $(i, t) \in I_m$, $k = 1, 2, 3$, result in the following contamination schemes defined by the actual values of $(\boldsymbol{x}_{it}, y_{it})$ for $(i, t) \in I_m$. If $y_{it} = \breve{y}^r_{it}$ or $y_{it} = \boldsymbol{x}^\top_{it}\boldsymbol{\beta} + \alpha_i + \breve{y}^c_{it}$ for $(i, t) \in I_m$, we talk about the non-clustered and clustered outliers, respectively. On the other hand, if $\boldsymbol{x}_{it}$ is left unmodified or $\boldsymbol{x}_{it} = \breve{\boldsymbol{x}}_{it}$, the contamination schemes is said to contain vertical outliers (VO) or leverage points (LP), respectively. All non-contaminated data $(i, t) \notin I_m$ follow model (18)–(19) with $H \equiv N(0, 1)$. The sample size is fixed to $n = 70$ and $T = 3$ now and the number of outliers is set to $m = 10$ (5% contamination) and $m = 42$ (20% contamination).

The results summarized in Table 4 document that, even if only 5% observations are contaminated, LS can get extremely biased (actually more that the inconsistent estimators based on the median transformation). On the contrary, the proposed robust estimators are not substantially affected by any type of contamination. Similarly to experiments discussed in previous sections, there are no substantial differences among the one-step robust estimators, while the data transformation matters: the pairwise-differencing again outperforms the first-differencing.

5 Concluding remarks

The present study examines the parameter estimation in fixed-effects panel data models with a fixed number of time periods from the point of view of robust statistical procedures. To achieve consistent estimators, we privilege first-difference and propose pairwise-difference data transformations and then apply robust estimators: LTS fol-

Table 4: The mean squared errors of all estimators in the presence of 5% or 20% scattered and clustered outliers.

		Non-clustered outliers				Clustered outliers			
		VO		LP		VO		LP	
		5%	20%	5%	20%	5%	20%	5%	20%
mean	LS	0.164	0.673	2.627	5.146	0.703	2.317	6.069	8.010
med	LS	0.176	0.639	3.150	5.668	0.596	1.938	6.342	8.253
	LTS	1.519	1.613	1.476	1.382	1.507	1.558	1.557	1.510
	WGM	0.485	0.556	0.480	0.487	0.477	0.432	0.451	0.449
	IRLS	0.606	0.681	0.598	0.594	0.575	0.518	0.585	0.558
	REWLS	0.566	0.639	0.552	0.551	0.538	0.481	0.547	0.513
	RLTS	0.251	0.258	0.287	0.317	0.268	0.214	0.209	0.316
1Δ	LS	0.203	0.807	2.646	5.255	0.907	2.835	6.087	8.024
	LTS	0.137	0.098	0.128	0.117	0.124	0.068	0.120	0.059
	WGM	0.033	0.053	0.033	0.070	0.027	0.028	0.027	0.027
	IRLS	0.031	0.052	0.039	0.086	0.026	0.027	0.026	0.027
	REWLS	0.030	0.055	0.038	0.090	0.025	0.028	0.025	0.028
	RLTS	0.028	0.064	0.037	0.170	0.024	0.028	0.024	0.028
$P\Delta$	LS	0.165	0.673	2.625	5.149	0.705	2.332	6.068	8.009
	LTS	0.101	0.076	0.100	0.091	0.090	0.051	0.090	0.043
	WGM	0.024	0.042	0.025	0.058	0.021	0.021	0.021	0.021
	IRLS	0.022	0.041	0.031	0.074	0.019	0.021	0.020	0.021
	REWLS	0.021	0.044	0.031	0.078	0.019	0.022	0.019	0.021
	RLTS	0.020	0.050	0.030	0.156	0.018	0.022	0.019	0.021

lowed by various reweighted LS and LTS methods. For a given data transformation, all methods achieve the same breakdown point and have similar finite-sample performance; the asymptotic distribution could be however provided only in the case of LTS and RLTS. Comparing the two data transformations, the best robust properties (i.e., the breakdown point $1/4$ irrespective of the number of time periods T) and the best estimation results have been obtained for the new pairwise-difference transformation, which could motivate its further study in the context of panel data models.

APPENDIX: PROOFS

Proof of Lemma 1: Since the LTS estimator is regression, affine, and scale equivariant (Rousseeuw and Leroy, 1987, Lemma 3 in Chapter 3), we only have to verify that the data-transformations – the first- and pairwise-differencing – do not affect the regression, affine, and scale transformations. For any $s \in N$, this directly follows from $\Delta^s(cy_{it}) = c\Delta^s y_{it}$, $\Delta^s(y_{it} + \boldsymbol{x}_{it}^\top \boldsymbol{v}) = \Delta^s y_{it} + (\Delta^s \boldsymbol{x}_{it})^\top \boldsymbol{v}$, and $\Delta^s(\boldsymbol{x}_{it}^\top \boldsymbol{A}) = (\Delta^s \boldsymbol{x}_{it})^\top \boldsymbol{A}$. □

Proof of Theorem 1: Before applying the LTS estimator, data are subject to the differencing transformations (8) or (9), which generate $T^{(\mathfrak{T})} = n(T-1)$ or $T^{(\mathfrak{T})} = nT(T-1)/2$ transformed observations, respectively. With these transformations, the worst case scenario occurs when aberrant observations are located so that each single outlier contaminates always $\min\{2, T-1\}$ (first-differencing) or $T-1$ (pairwise-differencing) differentiated observations. Hence given m outliers in the original sample, the number of outliers after the first- and pairwise-differencing will be at most $\min\{2, T-1\}m$ and $(T-1)m$, respectively.

At the same time, the breakdown point of LTS with the trimming constant h_{nT} equals $(nT^{(\mathfrak{T})} - h_{nT})/[nT^{(\mathfrak{T})}]$ if $h_{nT} \geq (nT^{(\mathfrak{T})} + p + 1)/2$ (Vandev and Neykov, 1998). LTS thus breaks down only if the number of outliers exceeds $nT^{(\mathfrak{T})} - h_{nT}$. In the case of the first differences, this means that LTS breaks down if $\min\{2, T-1\}m > nT^{(\mathfrak{T})} - h_{nT}$, implying that the breakdown point of the proposed panel-data LTS estimator equals $(nT^{(\mathfrak{T})} - h_{nT})/[\min\{2, T-1\}nT] = \{(nT^{(\mathfrak{T})} - h_{nT})/[2nT^{(\mathfrak{T})}]\} \cdot \{(2(T-1))/(\min\{2, T-1\}T)\}$. In the case of the pairwise differences, LTS breaks down if $(T-1)m > nT^{(\mathfrak{T})} - h_{nT}$, implying that the breakdown point equals $(T^{(\mathfrak{T})} - h_n^{\mathfrak{T}})/[nT(T-1)] = (nT^{(\mathfrak{T})} - h_{nT})/(2nT^{(\mathfrak{T})})$. The last claim of the theorem follows from $\lim_{n\to\infty}(nT^{(\mathfrak{T})} - h_{nT})/(2nT^{(\mathfrak{T})}) = (1-\lambda)/2$. □

Proof of Theorem 2: The claim of the theorem is proved for REWLS by Gervini and Yohai (2002, Theorem 3.3) and for RLTS by Čížek (2010, Theorem 2). The result for the IRLS estimator also directly follows from Gervini and Yohai (2002, Theorem 3.3), since v_{nT} determined by REWLS is by definition always greater or equal to $v = \eta$ used in (12), see equation (14). □

References

Bramati, M. C. and C. Croux (2007). Robust estimators for the fixed effects panel data model. *Econometrics Journal 10*(3), 521–540.

Čížek, P. (2006). Least trimmed squares in nonlinear regression under dependence. *Journal of Statistical Planning and Inference 136*(11), 3967–3988.

Čížek, P. (2008). Robust and efficient adaptive estimation of binary-choice regression models. *Journal of the American Statistical Association 103*, 687–696.

Čížek, P. (2010). Reweighted least trimmed squares: an alternative to one-step estimators. Discussion Paper 91/2010, CentER, Tilburg University.

Čížek, P. (2011). Semiparametrically weighted robust estimation of regression models. *Computational Statistics and Data Analysis 55*, 774–788.

Davies, P. L. and U. Gather (2005). Breakdown and groups. *The Annals of Statistics 33*(3), 977–1035.

Dhaene, G. and Y. Zhu (2009). Median-based estimation of dynamic panel models with fixed effects. Working paper, Faculty of Business and Economics, Katholieke Universiteit Leuven.

Genton, M. G. and A. Lucas (2003). Comprehensive definitions of breakdown points for independent and dependent observations. *Journal Of The Royal Statistical Society Series B 65*(1), 81–94.

Gervini, D. and V. J. Yohai (2002). A class of robust and fully efficient regression estimators. *The Annals of Statistics 30*(2), 583–616.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust statistics : the approach based on influence functions*. John Wiley & Sons.

Honoré, B. E. and J. L. Powell (2005). Pairwise difference estimation of nonlinear models. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models. Essays in Honor of Thomas Rothenberg.*, Chapter 22, pp. 520–553. Cambridge University Press.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis 91*(1), 74–89.

Lucas, A., R. van Dijk, and T. Kloek (2007). Outlier robust gmm estimation of leverage determinants in linear dynamic panel data models. Unpublished manuscript.

Maronna, R. A., R. D. Martin, and V. J. Yohai (2006). *Robust Statistics*. John Wiley & Sons.

Maronna, R. A. and V. J. Yohai (2000). Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference 89*(1-2), 197–214.

Ronchetti, E. and F. Trojani (2001). Robust inference with gmm estimators. *Journal of Econometrics 101*(1), 37–69.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association 79*(388), 871–880.

Rousseeuw, P. J. and A. M. Leroy (1987). *Robust regression and outlier detection*. Wiley.

Rousseeuw, P. J. and B. C. v. Zomeren (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association 85*(411), 633–639.

Simpson, D. G., D. Ruppert, and R. J. Carroll (1992). On one-step gm estimates and stability of inferences in linear regression. *Journal of the American Statistical Association 87*(418), 439–450.

Stromberg, A. J., O. Hossjer, and D. M. Hawkins (2000). The least trimmed differences regression estimator and alternatives. *Journal of the American Statistical Association 95*(451), 853–864.

Vandev, D. L. and N. M. Neykov (1998). About regression estimators with high breakdown point. *Statistics 32*, 111–129.

Wagenvoort, R. and R. Waldmann (2002). On b-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics 106*(2), 297–324.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*, Volume 1 of *MIT Press Books*. The MIT Press.