Robust Estimators for the Fixed Effects Panel Data Model

Maria Caterina Bramati * Christophe Croux †

U.L. Bruxelles K.U. Leuven

Abstract

The presence of outlying observations in panel data can affect the classical estimates in a dramatic way. Nevertheless the common practice seems to disregard the problem. The aim of this work is to study robust regression techniques in the fixed effects linear panel data framework. Robustness of the procedures is investigated by means of breakdown point computations and simulation experiments. A distinction between outlying blocks and cells in a panel is made. To show the potential of robust panel data methods an empirical example on the response of the private sector behavior to fiscal policy is presented.

Key Words: Breakdown point, Fixed effects model, Linear Regression, Panel data, Robustness.

*We are grateful to the two anonymous referees for helpful comments and suggestions on earlier drafts.
Maria Caterina Bramati (corresponding author), ECARES, ISRO and Department of Mathematics, Université Libre de Bruxelles, Campus de la Plaine, C.P. 210 - Bd. du Triomphe, B-1050 Brussels, Belgium. Email: mbramati@ulb.ac.be

†Christophe Croux, Dept. of Applied Economics, KU Leuven, Naamsestraat 69, B-3000 Leuven, Belgium. Email: christophe.croux@econ.kuleuven.ac.be

1 Introduction

Panel data estimators can be strongly biased in the presence of outlying observations. Although most researchers are aware of this problem, little literature is existing on robust estimation of the parameters in a panel data model. In this paper, robust versions of the classical Within Group estimator are considered. The robustness of these estimators with respect to outliers will be investigated.

The presence of outliers can lead to erroneous estimates in regression models. Indeed, the classical Least Squares (LS) approach is known to be very sensitive to outliers. Moreover, outliers are not always detectable by looking at residuals from a Least Squares fit, since the latter suffers from the *masking effect*. Masking means here that outliers affect the LS estimator in such a way that outlier diagnostics based on LS are not capable of detecting them anymore. Note that also diagnostic measures like the Cook Distance suffer from the masking effect, as soon as multiple outliers are present. More robust alternatives to LS are the Least Absolute Deviation estimator and M-estimators. Unfortunately, these estimators are not robust with respect to leverage points, i.e. outliers in the space of the covariates. Thus, regression estimators having a high breakdown point, as *Least Trimmed Squares* (Rousseeuw, 1984) and S-estimators (Rousseeuw and Yohai, 1984) are needed. The breakdown point of an estimator is the highest fraction of outliers that an estimator can withstand, and it is one of the most popular measures of robustness. A classical textbook for robust regression analysis is Rousseeuw and Leroy (1987). Therein, many examples are given where masking occurs and where the LS estimator breaks down completely. For estimating regression models in presence of noisy datasets it is hence crucial to estimate the model parameters using robust estimators. Afterwards, outlying observations can be detected using robust diagnostic measures, as in Rousseeuw and van Zomeren (1990).

By the word *robust estimator* we mean that an estimator is not altered too much by removing or modifying a small percentage of the dataset. Formal measures of robustness, including the breakdown point, are given in Hampel et al (1986). An outlier can be seen as an observation being highly unlikely to be generated by the model being used. It is behaving differently from the vast majority of the data cloud. As pointed out in Zaman, Rousseeuw and Orhan (2001), still few literature is available on robust techniques applied to econometrics. These authors believe that the reluctance to apply robust regression techniques to real economic data sets is due to different factors, such as the belief that outliers can be identified simply by eye and the unfamiliarity with and unawareness of such techniques.

While the usefulness of robust estimators in the linear regression model is well established, not much effort has been given to the development of robust procedures for panel data models in econometrics. We believe that the problem of outliers is even more pertinent in the panel data context. Large panels of families, firms, countries, ... are likely to contain atypical observations or gross errors (as typing, recording or computation errors). Moreover visual inspection of panel data is less obvious than for cross-sectional data, certainly in the case of multiple regressors. Regarding to robustness with respect to outliers, some literature is available for SURE models (Peracchi, 1991), whereas for panel data models there is very few literature. Among the studies on estimators for panel data there are some which concern robustness with respect to heteroscedasticity and autocorrelation, as in Alvarez and Arellano (2004).

In this paper we stick to the simple fixed effects panel data model, and focus on robust alternatives to the Within Groups estimator. As measure of robustness we consider the breakdown point of the estimators. Since our main purpose is to build highly robust procedures, high breakdown point estimators for the fixed effects panel data model are developed. We will assume that no endogeneity problem is present, and will therefore not deal with robust instrumental variable estimation of the linear panel data model as in Wagenvoort and Waldmann (2002). One of the estimators studied in this paper is however similar as the Wagenvoort and Waldmann (2002) estimators. It will be compared with a version of the MS estimator of Maronna and Yohai (2000) well suited for application in the panel data setting.

Another way to robust estimation of panel data models would have been to apply the

Robust Generalized Method of Moments of Ronchetti and Trojani (2001). This general approach has not been applied to the panel data context yet, but it will not have a high breakdown point. Note that the robust estimators proposed for variance components (or random effects) models as in Rocke (1991), Fellner (1986) or Richardson and Welsh (1997) neither are robust in terms of a high breakdown point.

In Section 2 the definition of breakdown point of a regression estimator for the panel data model is given. In Section 3 robust estimators with high breakdown point are outlined for the fixed effects model parameters, hereby extending some known robust regression estimators. Expressions for breakdown points are given as well. Section 4 is devoted to a simulation study, while in Section 5 we apply the robust estimator to real macrodata. In particular, we replicate in a robust way the Giavazzi, Jappelli and Pagano (GJP, 2000) study on the response of the private sector behavior to fiscal policy. Section 6 concludes the work.

2 Outlier classification and definition of the breakdown point of panel data estimators

2.1 Classifying outliers

As first step we try to figure out how outliers arise in panel data model. In robust statistics the standard assumption is that the majority of data follows a certain specified distribution F, while a certain small percentage of the data takes values unlikely to come from this distribution F. The latter ones are then called outliers.

The main possible sources of bias of the estimates come from contamination in the error term (*vertical outliers*) and in the explanatory variables (*leverage points*). This results then in a couple (x_{it}, y_{it}) entering as an outlying observation in the stacked regression equation (3.2). In this work we also pay attention to the case of *concentrated outliers* as the situation in which most of outlying observations are likely to be concentrated in few time-series. Indeed, such a case is very frequent in economic data: as an example we could consider a macro panel in which for few countries many outliers are present (due to measurement errors, badly scaled data...), while other countries are not contaminated at all.

To have an idea of how these three kinds of outliers look like (vertical, leverage, block concentrated outliers), we have simulated a panel of $N = 3$ time-series, each of $T = 15$ observations in the case of one single covariate x_{it}. A scatterplot of the data cloud (x_{it}, y_{it}) is then presented with fitted regression lines $y = \alpha_i + \beta x$ added for every cross-section $i = 1, 2, 3$.

In Figure 1(a) some of the observations are outlying in the y direction, generating vertical outliers. In Figure 1(b) one sees that the *vertical outliers are concentrated* in one single time-series. We speak of block concentrated outliers. In Figure 1(c) one sees that leverage points are present in every time-series, while in Figure 1(d) one has *block concentrated leverage points*, which means that leverage points are concentrated within a time-series.

2.2 Breakdown point definition

In this subsection we are going to define the breakdown point of a regression estimator in the panel data context. Such a definition will be useful for the computation of the breakdown point of the estimators proposed in the next section. Generally spoken, the breakdown point of an estimator is defined as the smallest fraction of outlying observations that can cause a 'breakdown' of the estimator. The first to introduce this notion were Donoho and Huber (1983) as a measure of the sensitivity of an estimator to aberrant observations. Breakdown points of robust regression estimators have been computed, e.g. in Rousseeuw and Leroy (1987). Let $Z = \left\{(x_{11}^{(1)}, ..., x_{11}^{(K)}, y_{11}), ..., (x_{1T}^{(1)}, ..., x_{1T}^{(K)}, y_{1T}), ..., (x_{NT}^{(1)}, ..., x_{NT}^{(K)}, y_{NT})\right\}$ be the set of the NT data points $(x_{it}^{(1)}, ..., x_{it}^{(K)}, y_{it})$. Let

$$\tilde{Z}_M = \left\{(\tilde{x}_{it}^{(1)}, ..., \tilde{x}_{it}^{(K)}, \tilde{y}_{it}) \in \Re^{K+1} \mid \# \left\{(i,t);\ (\tilde{x}_{it}^{(1)}, ..., \tilde{x}_{it}^{(K)}, \tilde{y}_{it}) \neq (x_{it}^{(1)}, ..., x_{it}^{(K)}, y_{it})\right\} \leq M\right\}$$

be a contaminated set of NT data points, obtained by replacing at most M data points by arbitrary values. The symbol # is the usual cardinality sign. Let R be an estimator,

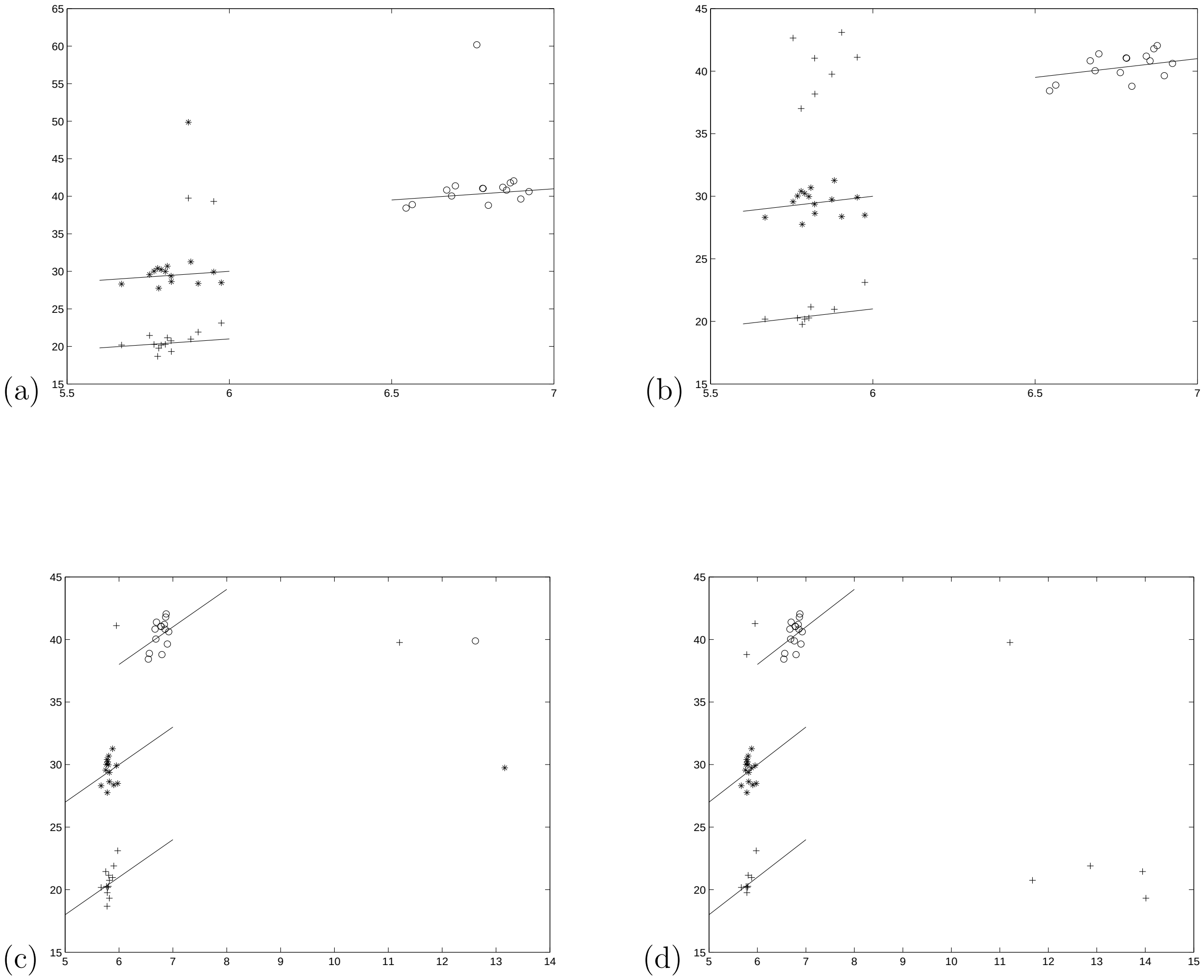

Figure 1: Scatterplot of y_{it} versus x_{it} for $1 \leq t \leq 15$ and $1 \leq i \leq 3$. Different time-series are indicated with different symbols. Four cases are considered: (a) only vertical outliers, (b) block concentrated vertical outliers, (c) leverage points, (d) block concentrated leverage points.

and $R(Z)$ its value at the sample Z. Typically, R could be an estimator for the regression parameter in a panel data model. The *cell* breakdown point of the estimator R at the sample Z is the smallest proportion of observations replaced by outliers which can cause the estimator R to take on values arbitrarily far from $R(Z)$. Formally,

$$\varepsilon^*_{NT}(R,Z) = \frac{1}{NT}\min\left\{M; \sup_{\tilde{Z}_M}\left\|R(Z) - R(\tilde{Z}_M)\right\| = \infty\right\} \tag{2.1}$$

The above definition is nothing else but the definition of breakdown point of a regression estimator following the general concept of breakdown point introduced by Donoho and Huber (1983), but now applied to panel data in stacked form. In this definition we do not take into account the presence of different groups or blocks in the data, and therefore one could say that it treats the outliers as *outlying cells* in the data matrix.

In what follows, we define a *block outlier* as a time-series in which there is at least one outlying observation. Of course, a block outlier may contain both horizontal and vertical outliers. Let

$$\tilde{Z}_B = \left\{(\tilde{x}^{(1)}_{it},...,\tilde{x}^{(K)}_{it},\tilde{y}_{it}) \in \Re^{K+1} \mid \#\left\{i;\ (\tilde{x}^{(1)}_{i.},...,\tilde{x}^{(K)}_{i.},\tilde{y}_{i.}) \neq (x^{(1)}_{i.},...,x^{(K)}_{i.},y_{i.})\right\} \leq B\right\}$$

be the set of all the NT data points which are contaminated by replacing at most B blocks by arbitrary blocks of the same size $(\tilde{x}^{(1)}_{i.},...,\tilde{x}^{(K)}_{i.},\tilde{y}_{i.})$, where $\tilde{y}_{i.}$ and $\tilde{x}^{(k)}_{i.}$ are $T \times 1$ vectors. The *block breakdown point* associated to the outlying blocks is the smallest proportion of contaminated time-series which can cause the breakdown of the estimator. Thus, considering the regression estimator R, the block breakdown point of the estimator R at the sample Z writes

$$\varepsilon^B_N(R,Z) = \frac{1}{N}\min\left\{B; \sup_{\tilde{Z}_B}\left\|R(Z) - R(\tilde{Z}_B)\right\| = \infty\right\}. \tag{2.2}$$

A similar notion of breakdown point was proposed by Müller and Uhlig (2001) but in the context of variance components estimation. Note that one could also think of a block outlier as a cross-section with at least one outlying observation. Treatment of outliers in the time domain is possible (see Franses et al (1999) and chapter 8 of Maronna, Martin and Yohai, 2006), but is beyond the scope of this paper, and less applicable for panels with a short time series dimension.

3 Robust Estimators for the FE panel data model

Consider the general formulation of the fixed effects linear panel data model.

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it} \quad i = 1, ..., N \quad t = 1, ..., T \tag{3.1}$$

where the i subscript denotes the cross-section dimension, whereas t denotes the time-series dimension. The $K \times 1$ column vector of explanatory variables is denoted by x_{it}, the regression parameter β is a $K \times 1$ vector, and the α_i are the unobservable time-invariant individual effects (which we consider to be fixed). Finally, the ε_{it} denote the disturbance terms, uncorrelated through time (i being fixed) and uncorrelated through cross-sections (t being fixed). In matrix notation, also called the *stacked* form representation, expression (3.1) becomes:

$$y = e_T \otimes \alpha + \mathbf{x}\beta + \varepsilon, \tag{3.2}$$

with $y = (y_{11}, y_{12}, ..., y_{NT})'$ an $NT \times 1$ vector and $\mathbf{x} = (x_{11}, x_{12}, ..., x_{NT})'$ an $NT \times K$ matrix. Furthermore, $\alpha = (\alpha_1, ..., \alpha_N)'$ is the $N \times 1$ vector of individual effects coefficients, e_T is a $T \times 1$ vector of ones, and $\otimes$ the Kronecker product.

The presence of outlying observations affects the classical Within Groups estimator. Our aim is to build up an estimation procedure which is less sensitive to the presence of aberrant observations. This study considers two approaches: the first one is based on a high breakdown Generalized M-estimator, similar to Hinloopen and Wagenvoort (1997). They propose a version of the Generalized M-estimator originally introduced by Simpson, Ruppert and Carroll (1992) with high breakdown point level. A difference in our approach is that we use the LTS estimator (Rousseeuw, 1984) as initial regression estimator and a multivariate S-estimator to downweigh leverage points. These estimators are more efficient and faster to compute than the ones used by Hinloopen and Wagenvoort (1997). Indeed, Rousseeuw and Van Driessen (2002) provide a fast algorithms for the LTS estimator (FAST-LTS) showing that it runs faster than all programs for least median of squares (LMS), which is the preliminary estimator in Hinloopen and Wagenvoort (1997). As for the multivariate S-estimator for location and scale, which we use instead of the Minimum Volume Ellipsoid (MVE) estimator applied in Hinloopen and Wagenvoort (1997). Croux and Haesbroeck (1999) present efficiencies for multivariate S-estimators and show that they are fairly high. Moreover, Davies (1992) shows that the convergence rate of the MVE estimator is slower than the asymptotically normal S-estimators. Given the smoothness of the objective function for the S-estimators, fast algorithms can be constructed as in Ruppert (1992), whereas the computation of the MVE estimator is much more demanding. Unfortunately, this estimator has the drawback of being not regression equivariant. Therefore, in the second approach we consider the MS estimator of Maronna and Yohai (2000) applied to the particular setting of the linear panel data model. Although the second estimator has nice statistical properties, among which the regession equivariance, it is difficult to distinguish between the two estimators on efficiency grounds based on theoretical results. Therefore, both estimators are considered in the simulation experiments of Section 4.

The classical Within Groups estimator is defined as follows. The data are centered within every time-series. So

$$\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} \tag{3.3}$$

and

$$\tilde{x}_{it} = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}. \tag{3.4}$$

Then it follows from (3.1) that

$$\tilde{y}_{it} = \beta'\tilde{x}_{it} + \text{error}_{it} \tag{3.5}$$

and it is observed that the fixed effects α_i have been eliminated by the centering operation. Regressing $\tilde{y}_{it}$ on $\tilde{x}_{it}$ by OLS results then in the Within Group estimator $\hat{\beta}_{WG}$. Econometricians are mainly interested in estimating β, but the fixed effects parameters can be estimated as well. We refer to Baltagi (2001) for a detailed treatment of the classical Within Groups estimator.

Before defining the robust versions of the within estimator, we introduce some equivariance properties that a panel data regression estimator should verify (following Rousseeuw and Leroy, 1987). Denote R for an estimator of the regression parameter β in (3.1). One has that a panel regression estimator R is *scale equivariant* if

$$R\left(\{(x_{it}, cy_{it}), i=1,...,N; t=1,...,T\}\right) = cR\left(\{(x_{it}, y_{it}), i=1,...,N; t=1,...,T\}\right),$$

for any scalar c. It is *regression equivariant* if

$$R\left(\{(x_{it}, y_{it}+x_{it}'\gamma), i=1,...,N; t=1,...,T\}\right) = R\left(\{(x_{it}, y_{it}), i=1,...,N; t=1,...,T\}\right)+\gamma,$$

where γ is a $K\times 1$ vector of constants and it is *affine equivariant* if

$$R\left(\{(Ax_{it}, y_{it}), i=1,...,N; t=1,...,T\}\right) = (A')^{-1}R\left(\{(\mathbf{x}_{it}, y_{it}), i=1,...,N; t=1,...,T\}\right),$$

where A is any $K\times K$ nonsingular matrix. The classical Within Groups estimator satisfies all 3 equivariance properties.

The philosophy underlying the two approaches we will consider is almost the same: in order to get a robust version of the *within groups estimator* we estimate robustly the center of each time-series in both the dependent and the explanatory variables, and subtract it from each observation in the block. Then a robust regression method to the centered data can be applied in order to obtain the robust estimate of the coefficients in (3.5). It seems natural to start with centering the variables (both dependent and independent) by the median instead of the mean, since the median is known to min-max robust (Huber 1981). So

$$\tilde{y}_{it} = y_{it} - \operatorname*{median}_t y_{it} \tag{3.6}$$

and

$$\tilde{x}_{it}^{(j)} = x_{it}^{(j)} - \operatorname*{median}_t x_{it}^{(j)}, \tag{3.7}$$

for $1 \le i \le N$, $1 \le t \le T$ and $1 \le k \le K$, where $x_{it}^{(j)}$ is the j-th explanatory variable measured at time t in the i-th time-series. Centering has a crucial advantage: it reduces the number of parameters enormously since it eliminates the fixed effects parameters. As a result, computation time for robust regression algorithms remains feasible. Although many progress has been made in developing fast algorithms for computing highly robust estimators, their computation is still cumbersome for very large numbers of explicative variables. As a consequence, we will disregard the approach proposed by Rousseeuw and Wagner (1994) being computationally infeasible in presence of many time-series.

After centering, the natural approach is to regress $\tilde{y}_{it}$ on $\tilde{x}_{it}$ but now using a robust regression estimator. A well known robust regression estimator is the *least trimmed squares* (LTS) estimator (Rousseeuw, 1984). Applied on centered data, it is defined as

$$\hat{\beta}_{LTS} = \arg\min_{\beta} \sum_{k=1}^{h} \left[(\tilde{y}_k - \tilde{\mathbf{x}}_k' \beta)^2 \right]_{k:NT} , \tag{3.8}$$

where

$$\left[(\tilde{y}_. - \tilde{\mathbf{x}}_. \beta)^2 \right]_{1:NT} \le \left[(\tilde{y}_. - \tilde{\mathbf{x}}_. \beta)^2 \right]_{2:NT} \le ... \le \left[(\tilde{y}_. - \tilde{\mathbf{x}}_. \beta)^2 \right]_{NT:NT}$$

are the ordered squared regression residuals. So the LTS minimizes the sum of the smallest h squared residuals. The value $1 \le h \le NT$ is a truncation value. A default choice is $h = [3NT/4]$, making it possible to cope with up to 25% of outliers. Although the LTS estimator is proved to be regression, scale and affine equivariant in a pure regression setting (Rousseeuw and Leroy, 1987), in our setup $\hat{\beta}_{LTS}$ is only scale equivariant, while regression and affine equivariance do not hold. This is due to the nonlinearity of the centering transformation by the median.

In case one is interested in having estimates for the fixed effects, they can easily be obtained as follows. Once that β is estimated, one gets:

$$\hat{\alpha}_i(\beta) = \operatorname*{median}_t (y_{it} - x_{it}' \beta) \qquad \text{for } i = 1, ..., N. \tag{3.9}$$

3.1 A high breakdown Generalized M-approach

This approach is basically an improvement of the Within Groups LTS estimator $\hat{\beta}_{LTS}$ in order to increase statistical efficiency, while maintaining enough robustness. Once LTS regression is performed on the centered data in stacked form, consider the residuals $r_{it} = \tilde{y}_{it} - \hat{\beta}'_{LTS}\tilde{x}_{it}$ and the robust scale estimate of the residuals $\hat{\sigma}_{LTS}$. The latter estimator is defined as

$$\hat{\sigma}^2_{LTS} = c_{LTS}\,\frac{1}{h}\sum_{k=1}^{h}(\tilde{y}_k - \tilde{x}'_k\hat{\beta}_{LTS})^2_{k:NT},$$

with c_{LTS} chosen to make $\hat{\sigma}^2_{LTS}$ a consistent estimator for $\sigma^2 = E[\varepsilon^2_{it}]$ at normal error distributions. Again the truncation value is set to $h = \lfloor 0.75NT \rfloor$, as a good compromise between robustness and efficiency. Then we set up a $NT \times NT$ diagonal matrix W_r to downweigh observations having large residuals with respect to the robust initial LTS fit (and thus not suffering from the masking effect). Its diagonal elements are of the form $\rho'(r_{it}/\hat{\sigma}_{LTS})/\,(r_{it}/\hat{\sigma}_{LTS})$ where the loss function ρ is taken to be the Tukey's biweight function (Beaton and Tukey, 1974):

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4} & \text{if } |x| \leq c \\ \frac{c^2}{6} & \text{if } |x| > c \end{cases}. \tag{3.10}$$

With this choice the diagonal elements of W_r turn out to be

$$(W_r)_{it} = \begin{cases} 0 & \text{if } \left|\frac{r_{it}}{\hat{\sigma}_{LTS}}\right| \geq c \\ \left(1 - (\frac{r_{it}}{c\hat{\sigma}_{LTS}})^2\right)^2 & \text{if } \left|\frac{r_{it}}{\hat{\sigma}_{LTS}}\right| < c \end{cases}, \tag{3.11}$$

where c is selected to obtain a good trade-off between the efficiency and robustness level. We selected $c = 4.685$ as in Wagenvoort and Waldmann (2002).

To have further protection against bad leverage points, outliers in the covariate space will be downweighed as well. Therefore we build a diagonal matrix W_x with weights decreasing to zero as the observations are getting further and further from the center of the data cloud. A robust version of the Mahalanobis distance, called a robust distance RD_{it}, is computed for every $\tilde{x}_{it}$:

$$RD_{it} = \sqrt{(\tilde{x}_{it} - \hat{\mu})\hat{V}^{-1}(\tilde{x}_{it} - \hat{\mu})'}, \tag{3.12}$$

for $t = 1, ..., T$ and $i = 1, ..., N$. In the above equation, $\hat{\mu}$ and $\hat{V}$ are robust location and covariance estimates of the centered explicative variables computed by applying the S-multivariate location and scale estimator, respectively (see Rousseeuw and Leroy 1987, pp. 263-264, Davies 1987). The S-estimator of multivariate location and scale has a high efficiency and is fast to compute (see Ruppert 1992 for computational aspects). The diagonal elements of the $NT \times NT$ weighing matrix W_x are then defined as

$$(W_x)_{it} = \min\left(1, \frac{\sqrt{\chi^2_{K,0.975}}}{RD_{it}}\right), \tag{3.13}$$

where $\chi^2_{K,0.975}$ is the upper 97.5% quantile of a chi-squared distribution with K degrees of freedom.

At this point we are ready to perform a weighed least squares estimation in order to obtain an estimate of β as

$$\hat{\beta}_{WGM} = (\tilde{X}' W_x W_r \tilde{X})^{-1} \tilde{X}' W_x W_r \tilde{Y} \tag{3.14}$$

Expression (3.14) is a version of the *high breakdown point generalized M*-estimator constructed by Hinloopen and Wagenvoort (1997) in the regression context. We will simply call it the Within GM estimator. Matlab programs to compute $\hat{\beta}_{LTS}$ and $\hat{\beta}_{WGM}$ are available from the authors. Note that the choices for the LTS as an initial estimator, the use of the Tukey biweight function, and the use of leverage weights as in (3.13) are fairly standard in the practice of robust statistics.

3.2 Within MS (WMS) estimator

This is basically a special case of the MS regression estimator proposed by Maronna and Yohai (2000) for a robust regression with both continuous and categorical explanatory variables. In contrast to the Within GM estimator considered before, it is also affine and regression equivariant. The idea underlying this procedure is to use alternately M-estimators for the categorical variables and S-estimators of regression for the continuous ones. The

M-estimator is very fast to compute, but is not robust with respect to leverage points. Since categorical variables (and certainly dummy variables) do not contain major leverage points, we can fast and safely compute M-estimators of regression on them. The M-estimator serves to remove the effect of categorical variables from the continuous ones. The S-estimator of regression is highly robust, also with respect to leverage points, and will then be applied to the continuous variables.

Suppose for a moment that the fixed effects α_i's are known. Following the definition of Rousseeuw and Yohai (1984), an S-estimate of regression is defined as the minimization of an M-estimator of scale computed from the regression residuals $r_{it}(\alpha, \beta) = y_{it} - x_{it}'\beta - \alpha_i$. An M-estimator of scale S is defined as the solution of the following equation in s:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_S(\frac{r_{it}(\alpha,\beta)}{s}) = b \tag{3.15}$$

where $\rho_S(\cdot)$ is an even, symmetric and continuously differentiable loss function with $\rho_S(0) = 0$. The constant b is put equal to $E_\Phi[\rho_S(\varepsilon)]$, for Φ the standard normal distribution, to get consistent estimates of the regression scale parameter. The S-estimator of regression is then defined as

$$\hat{\beta}_S(\alpha) = \arg\min_{\beta} S(r_1(\alpha,\beta), ..., r_{NT}(\alpha,\beta)). \tag{3.16}$$

S-estimators of regression have a high breakdown point and are much more efficient than the LTS estimator. As loss function we took the Tukey Biweight loss function, having a derivative defined in (3.11), with the constant c (which represents a tuning constant to achieve the desired value of the breakdown point, see Rousseeuw and Leroy, 1987) selected to have an overall 25% breakdown point for the regression estimator. If the minimization in (3.16) is over a low dimensional parameter β, then its computation is fast using for example the algorithm of Ruppert (1992). This is the reason why we do not minimize simultaneously over α and β. Indeed, α can easily contain hundreds of parameters, making it computationally impossible to compute the S-estimator.

On the other hand, suppose that one knew β. Then $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_N)'$ is obtained as an

M-estimator of regression:

$$\hat{\alpha}(\beta) = \arg\min_{\alpha} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_M(r_{it}(\alpha, \beta)), \tag{3.17}$$

from which it follows directly that

$$\hat{\alpha}_i(\beta) = \arg\min_{\alpha_i} \sum_{t=1}^{T} \rho_M(y_{it} - \mathbf{x}'_{it}\beta - \alpha_i), \quad \text{for } i = 1, ..., N. \tag{3.18}$$

Since typically there are many fixed effects, it is important to have a fast way to estimate them. Our proposal is to take $\rho_M(\cdot) = |\cdot|$ the absolute value loss-function, yielding an explicit formula for (3.18):

$$\hat{\alpha}_i(\beta) = \operatorname*{median}_t(y_{it} - x'_{it}\beta) \quad i = 1, ..., N. \tag{3.19}$$

The WMS estimator $\hat{\beta}_{WMS}$ for a linear panel data model is then defined as

$$\hat{\beta}_{WMS} = \arg\min_{\beta} S(r_1(\hat{\alpha}(\beta), \beta), ..., r_{NT}(\hat{\alpha}(\beta), \beta)), \tag{3.20}$$

with

$$r_{it}(\hat{\alpha}(\beta), \beta) = y_{it} - x'_{it}\beta - \operatorname*{median}_t(y_{it} - x'_{it}\beta).$$

The MS estimator was proposed for regression with both continuous and categorical variables. As we saw, it can be applied elegantly to a panel data context. In contrast to $\hat{\beta}_{WGM}$, the estimator is also regression and affine equivariant. The reason for this maintenance of equivariance properties is that the WMS estimator is defined by minimizing jointly over α and β. In this sense, the WMS estimator is the robust counterpart of the Least Squares Dummy variables representation of the Within Group estimator.

It is important to note that $\hat{\alpha}(\beta)$ can be computed directly from (3.19), which will simplify the algorithm as proposed by Maronna and Yohai (2000). This algorithm can then be summarized as follows:

- Take a subsample $\mathcal{I}$ of size K of the set $\{(x_{it} - \text{median}_t\, x_{it}, y_{it} - \text{median}_t\, y_{it}) | 1 \leq i \leq N, 1 \leq t \leq T\}$. Then there exists a $\hat{\beta}_{\mathcal{I}}$ fitting perfectly the observations of this subsample. The quality of the fit of $\hat{\beta}_{\mathcal{I}}$ to the whole data set is measured by

$$s_{\mathcal{I}} = S(r_1(\hat{\alpha}(\hat{\beta}_{\mathcal{I}}), \hat{\beta}_{\mathcal{I}}), ..., r_{NT}(\hat{\alpha}(\hat{\beta}_{\mathcal{I}}), \hat{\beta}_{\mathcal{I}})).$$

We will generate randomly $N_{samp} = 500$ subsamples $\mathcal{I}$, and select the one having the minimal value for $s_{\mathcal{I}}$. The number of generated subsamples N_{samp} is set according to a given probability that at least one of the subsamples is outlier-free. Given the percentage of outliers ε in the data, as the number of observations increases the probability of having at least one 'clean' K-subsample among the N_{samp} subsamples tends to $1 - (1 - (1 - \varepsilon)^K)^{N_{samp}} > 0$.

Denote $\hat{\beta}^0$ the fit corresponding to this optimal value. It is believed to be already close to the global solution $\hat{\beta}_{WMS}$.

- Given the initial estimate $\hat{\beta}^0$, an iterative algorithm is started to get closer to the minimum of (3.20). At the $(k+1)$th iteration step we set

$$\hat{\beta}^{(k+1)} = \arg\min_{\beta} S(r_1(\hat{\alpha}(\hat{\beta}^{(k)}), \beta), ..., r_{NT}(\hat{\alpha}(\hat{\beta}^{(k)}), \beta)). \quad (3.21)$$

The first order condition associated to (3.21) is

$$\sum_{i=1}^{N} \sum_{t=1}^{T} w_{it} \, x_{it}(y_{it} - \beta x_{it} - \hat{\alpha}_i(\hat{\beta}^{(k)})) = 0 \quad (3.22)$$

where $w_{it} = W\left(r_{it}(\alpha(\hat{\beta}^{(k)}), \beta)\right)$ are weights, for the weighing function $W(r) = \rho'_S(r)/r$. Since the β needed to compute the weights is unknown, we approximate w_{it} by $W(r_{it}(\alpha(\hat{\beta}^{(k)}), \hat{\beta}^{(k)})$. Then equation (3.22) is linear and can be solved directly yielding $\hat{\beta}^{(k+1)}$. Afterwards $\hat{\alpha}(\hat{\beta}^{(k+1)})$ is computed as in (3.19), the weights are updated, and a next step in the iteration procedure can be taken. Maronna and Yohai (2000) suggest to iterate a fixed number of times ($M = 20$) and to choose the $\hat{\beta}^{(k)}$ which produces the minimum value of the objective function in (3.20).

The above algorithm has been implemented and runs quite fast. It is available from the authors.

3.3 Breakdown point of WGM and WMS estimates

In this section we will derive the breakdown point of the robust estimators for the fixed effects linear panel data model considered before. We will show that the cell breakdown point of the WGM estimator is essentially determined by the breakdown point of the initial Least Trimmed Squares regression estimator, while the cell breakdown point of the WMS estimator depends on the S-estimator used. Recall that in both the WGM-estimator and in the M-step of the WMS-estimator data are centered by the median. Besides computational advantages, the median has also the property of having the highest possible value for the breakdown point: at a univariate data set $Z = \{z_1, \ldots, z_T\}$ we have

$$\varepsilon^*_T(\text{median}, Z) = \frac{\lfloor (T+1)/2 \rfloor}{T} \approx 0.5$$

where $\lfloor \cdot \rfloor$ stands for the integer part.

The worst case scenario is that for a number of time-series exactly $\lfloor (T+1)/2 \rfloor$ observations are corrupted, implying that after centering by the median all observations belonging to that block will be contaminated. In the proposition below the number k will stand for the maximum number of time-series where exactly half of the observations are corrupted. (All proofs can be found in the Appendix.)

Proposition 1 *Let Z be a panel data set of N blocks over T time periods. If the breakdown point of the auxiliary LTS or S estimator is given by $m/(NT)$, then*

$$\varepsilon^*_{NT}(\hat{\beta}_{WMS}, Z) = \varepsilon^*_{NT}(\hat{\beta}_{WGM}, Z) = \frac{k\lfloor (T+1)/2 \rfloor + \min(m - kT, \lfloor (T+1)/2 \rfloor)}{NT} \qquad (3.23)$$

with

$$k = \lfloor \frac{m}{\lfloor \frac{T+1}{2} \rfloor} \rfloor.$$

As an immediate corollary, if the breakdown point of the LTS or S-estimator tends to α^*, then the cell breakdown point of the WGM or WMS estimator tends to $\alpha^* \frac{\lfloor (T+1)/2 \rfloor}{T}$ for N tending to infinity. In addition, if also T tends to infinity, then the cell breakdown point tends to $\alpha^*/2$. Using a maximal breakdown point LTS or S-estimator, for which $\alpha^* = 50\%$, yields then a 25% cell breakdown point for the regression estimator of the panel data model (for both N and T tending to infinity). Indeed, the Robust Within Groups estimators break down when more than half of the transformed observations (i.e. in deviation of the medians) are outliers. If half of the original observations in one time series are outliers, then the median for that time series breaks down. Thus, one may corrupt $N/2$ medians with only $NT/4$ observations. If $N/2$ medians are corrupted, then 50% of the transformed observations are corrupted. Hence, the breakdown point of the Robust Within Groups estimator is 25%.

Note that in the above theorem the worst scenario does not correspond to a random distribution of the outliers over the data set. If one requires random positioning of the outliers, which we believe to be an unrealistic scenario, then the breakdown points would be significantly higher.

Now, we move on to the *block breakdown point* computation for the estimators WGM and WMS estimator of β. By definition of block breakdown point, we are going to look for the minimal number of blocks a time series which, being contaminated, cause the breakdown of the slope estimate β. We get:

Proposition 2 *Let Z be a panel data set of N blocks over T time periods. If the breakdown point of the auxiliary LTS or S-estimator is given by $m/(NT)$, then*

$$\varepsilon_N^B(\hat{\beta}_{WMS}, Z) = \varepsilon_N^B(\hat{\beta}_{WGM}, Z) = \frac{\lceil m/T \rceil}{N}, \tag{3.24}$$

where $\lceil \cdot \rceil$ is the ceiling operator. As an immediate corollary we have that if m/NT tends to α^* (breakdown point of the LTS or S-estimator), then the breakdown point of the WGM or WMS estimator will tend to α^* as N goes to infinity. So the block breakdown point equals the breakdown point of the auxiliary regression estimator.

4 Simulation study

The asymptotic normality of the WMS estimator was discussed in Maronna and Yohai (2000). Also for the WGM estimator we expect that the asypmptotic normality still holds, following the lines of Wagenvoort and Waldmann (2002), tough we do not provide a formal proof. The efficiency of both procedures will be compared by means of a simulation experiment.

In order to study the performance of the methods previously described, we have run a simulation study. The dependent variable is generated according to model (3.1), with $\varepsilon_{it} \sim N(0,1)$, $\alpha_i \sim U(0,20)$, and the vector of the slope coefficients β set equal to a vector of zeros. The explanatory variables are generated from a multivariate standard normal distribution.

In the next step contamination is led on y's only (vertical outliers) and afterwards on both y and the explicative variables hereby introducing bad leverage points. Contamination is generated in two different ways: (a) either completely randomly over all observations (*random contamination*), hereby creating outlying cells (b) or concentrating the contamination in a number of blocks such that half of the observations in affected time series are contaminated (*concentrated contamination*), hereby creating outlying blocks. Outliers generated by *random contamination* are either vertical outliers or leverage points, whereas in the case of *concentrated contamination* they are either block concentrated vertical outliers or block concentrated leverage points. Vertical outliers are obtained by adding to the y's originally generated an additional term $\sim N(50,1)$. Bad leverage points are obtained by replacing x-values corresponding to the observations already contaminated in the y-direction, by points coming from a K-variate normal distribution $N(10 \times \mathbf{1}, I)$, where $\mathbf{1}$ is a $K \times 1$ vector of ones. The percentages of contamination considered are 5% and 10%. We choose $K = 1$ and $N = 100$, and two different values for T are taken to check the performance of the estimators in the case of small time series dimension ($T = 4$), and in case of large panels ($T = 20$).

For each of the $M = 1000$ replications we estimate the coefficient β of the fixed effects model applying the classical within estimator $\hat{\beta}_{WG}$, the robust version $\hat{\beta}_{WGM}$ using the high breakdown-GM estimator, and $\hat{\beta}_{WMS}$. Then we compute the Mean Squared Error (MSE):

$$MSE = \frac{1}{M}\sum_{j=1}^{M} ||\widehat{\beta}^{(j)} - \beta||^2, \tag{4.1}$$

where $\widehat{\beta}^{(j)}$ is the slope estimate in the j-th replication. Results are reported in Tables 1 and 2.

We observe that in the normal situation without errors the efficiency of the two robust estimators is very close to that of the Within Group estimator. This classical estimator becomes very bad when there are vertical outliers and even worse in presence of bad leverage points, as indicated by the huge MSE values. On the other hand, WGM and WMS estimator yield good and stable results over all sampling schemes considered. Comparing the performance of the robust estimators, we observe that WMS and WGM estimates yield similar outcomes. In presence of bad leverage points, the WMS estimator gives slightly better results, while the WGM can cope better with concentrated outliers. On the whole we cannot clearly distinguish, on the basis of this simulation experiment, between the performance of the WGM and the WMS estimator. It has become clear, however, that both methods yield a large gain in MSE with respect to the classical procedure in presence of outliers, and also only very small efficiency loss in absence of outliers.

5 Empirical Illustration

In this section we compare the Within Groups, Within GM and WMS fixed effects estimators by means of a real macroeconomic application. The idea is to replicate, using a dataset from the World Bank, the study led by Giavazzi, Jappelli and Pagano (2000) on the response of the private sector to fiscal policy. The basic model they estimate is

$$\frac{S_{it}}{Y_{it}^*} = \alpha_i + \beta_1 \frac{S_{it-1}}{Y_{it-1}^*} + \beta_2 \frac{Y_{it} - Y_{it}^*}{Y_{it}^*} + \beta_3 DR_{it-1} + \beta_4 \frac{T_{it}}{Y_{it}^*} + \beta_5 \frac{G_{it}}{Y_{it}^*} + \varepsilon_{it} \tag{5.1}$$

Table 1: MSE of the WG, WGM and WMS estimator under several sampling schemes for $N = 100$, $T = 4$ and levels of contamination 5% and 10%.

	WG	WGM	WMS	WG	WGM	WMS
no outliers	0.004	0.003	0.004	0.004	0.003	0.004
Sampling Scheme	5% contamination			10% contamination		
vertical outliers	0.395	0.004	0.004	0.936	0.004	0.004
leverage points	16.933	0.004	0.003	20.260	0.006	0.005
concentrated vertical outliers	0.273	0.004	0.010	0.640	0.004	0.009
concentrated leverage points	14.818	0.003	0.004	18.885	0.003	0.004

Table 2: MSE of the WG, WGM and WMS estimator under several sampling schemes for $N = 100$, $T = 20$ and levels of contamination 5% and 10%.

	WG	WGM	WMS	WG	WGM	WMS
no outliers	0.0005	0.0005	0.0007	0.0005	0.0005	0.0007
Sampling Scheme	5% contamination			10% contamination		
vertical outliers	0.0700	0.0004	0.0007	0.1094	0.0006	0.0008
leverage points	17.076	0.0007	0.0007	20.2340	0.0016	0.0009
concentrated vertical outliers	0.0366	0.0006	0.0011	0.0690	0.0005	0.0021
concentrated leverage points	13.175	0.0004	0.0008	17.6250	0.0005	0.0008

with the national saving rate S_t as a fraction of potential output Y_t^* as dependent variable and the lagged national saving rate, the output gap $(Y_t - Y_t^*)/Y_t^*$, the dependency ratio DR_t (proportion of the population less than 15 years old and older than 65), net taxes T_t and the government consumption G_t (scaled on potential output) as regressors. Potential output is computed by fitting on the log output a quadratic trend, as in GJP(2000).

The dataset they use is the World Saving Data Base which contains yearly national income and fiscal variables for a group of 150 industrial and developing countries from 1960 to 1995. As GJP(2000) did, we dropped from the panel countries for which the most crucial variables were missing (18 countries), but kept all other countries. Furthermore, OECD countries are dropped, leaving us with 108 observations. Although this dataset is quite accurate, the authors proceed to a preliminary cleaning of the data, selecting 101 developing countries out of 108 and restricting the sample period considered to 1970-1994. As such, a restricted dataset is obtained.

In our panel, we have some more countries with respect to the ones considered by GJP (2000) (i.e. Angola, Burundi, Guinea-B, Sudan, Solomon Islands, Somalia and Sao Tome), which we call the *extended dataset* in what follows. Moreover, for many countries we kept the time range as large as possible, considering all the available data points[1]. This extended dataset is not subject to preliminary, rather subjective data cleaning. Then we performed the classical and the robust estimates which gave us the results reported in Table 3. Computation of the associated standard errors is outlined in the Appendix and are reported as well (between parenthesis) in the Table.

In the first column of Table 3 are given the results applying the classical fixed effects estimator with the selection of countries as in GJP (2000). In the second column are given the coefficient estimates obtained applying the classical within groups estimator to the extended dataset. Then robust techniques are applied on the extended dataset: in column 3 are

[1]GJP (2000) selected shorter time intervals than those available for countries like United Arab Emirates, Bahrain, Buhtan, Kuwait, Gambia, Lesotho, Mali, Mozambique, Uganda and Vanuatu.

Table 3: Estimation results of model (5.1) using the classical WG and the robust estimators WGM and WMS on the extended dataset, together with the classical estimator on the restricted data set as reported by GJP(2000).

	GJP(2000)	WG	WGM	WMS
β_1	0.459	−1.082	0.528	0.556
	(0.019)	(0.001)	(0.000)	(0.000)
β_2	0.300	−1.932	0.140	0.139
	(0.017)	(0.002)	(0.000)	(0.000)
β_3	−0.228	−2.850	−0.211	−0.355
	(0.064)	(1.690)	(0.005)	(0.008)
β_4	0.257	0.062	0.153	0.1011
	(0.029)	(0.000)	(0.000)	(0.000)
β_5	−0.493	17.907	−0.372	−0.419
	(0.039)	(0.101)	(0.002)	(0.005)

given the WGM procedure estimates, while in the last column the WMS estimates are given. Comparing the results, we observe that the classical estimates based on the extended dataset take values which are clearly far from GJP (2000) evidence and from the economic theory. Note that the huge differences between the estimates reported by GJP (2000) and the classical, non robust, estimates of β_1, β_2 and β_5. On the other hand, robust estimates seem to confirm the results of the author, producing estimates very close to the ones obtained using the "*cleaned*" dataset. Instead of a preliminary cleaning of the data, being subjective and having the risk of not detecting all outliers, a robust analysis could thus be performed directly on the complete data set.

As a byproduct of the robust analysis, it is instructive to look at the weights given by the robust procedures to each cell (x_{it}, y_{it}). In Table 4 (Appendix) we report the average weights over the time span 1970-1994 assigned to a country. A weight close to one means that the data for that country follow the model closely over the time span and hence are not downweighed. Smaller weights for a country imply that data for that country are believed to be outlying, and hence being downweighed. For example Lesotho has been attributed a low weight, both by WGM and WMS, while a country like India is almost not downweighed. Many of the countries which were dropped by GJP (2000), like Angola, Somalia and Solomon Islands, received a low weight by the robust estimators, as it is the case for many countries, like Bhutan, United Arab Emirates, Gambia and Kuwait, for which the available time span was not fully used by GJP (2000). A few countries, like Guyana and Nicaragua, are in the restricted data set, but still get rather low weight.

Note that an initial data cleaning corresponds to give weights one or zero to a country, where a zero weight means that the country is dropped from the study. A robust approach allows for a more careful, data-driven weighing of the observations.

6 Conclusions

In this paper we have studied the problem of outliers in panel data, focusing on fixed effects models with the basic assumptions of homoscedasticity and no autocorrelation of the errors. Similar to Wagenvoort and Waldmann (2002) we developed a Within GM estimator. Then, we applied the MS estimator of Maronna and Yohai (2000) to the panel data context. Simulations show the high robustness of both procedures. Breakdown point considerations have also been made. An example of the potentialities of robust regression procedures in panels is given in the empirical application. There we observed that robust techniques do not require a preliminary subjective cleaning of the data and still produce reasonable parameter estimates even when rough errors occur in the data.

We feel that it is important that applied econometricians are aware of the risk of outlying observations and the usefulness of robust methods, since classical inference based on a data set containing outliers can yield completely erroneous results. In this paper two robust estimators are studied in more detail. The robustness of both procedures is similar, as indicated by their equal breakdown point and the results of the simulation study. The MS approach is computationally slightly more demanding, but offers the advantage of being regression and scale equivariant. The WGM is not regression equivariant, but it can be useful as an initial estimator for WMS, instead of using time-consuming subsampling algorithms.

7 Appendix

Proof of Proposition 1. In the computation of the cell breakdown point of the slope coefficient, we need to consider 2 possible reasons why data points are contaminated. They can be outlying either because of the centering (median breaks down) or because the observation is outlying itself before centering. The first case occurs when half of the observations in the block is outlying which implies that after centering all the data belonging to that block are corrupted. Of course we want that the sum of the outliers determined by both situations does not exceed m, the maximal number of outliers that the auxiliary regression estimator can withstand. Let $k = \lfloor \frac{m}{\lfloor \frac{T+1}{2} \rfloor} \rfloor$, where $\lfloor \cdot \rfloor$ stands for integer part. We need to corrupt at least k blocks in a concentrated way, by which we mean contaminating $\lfloor (T+1)/2 \rfloor$ of its points. Then we still need to have at least $l = m - kT$ extra outliers in the panel. But if $l > \lfloor (T+1)/2 \rfloor$, then we contaminate only $\lfloor (T+1)/2 \rfloor$ outliers within the same block, yielding T extra outliers after centering, causing breakdown of the regression estimator. Therefore we may conclude that (3.23) holds. □

Proof of Proposition 2. By definition of block breakdown point, we are going to look for the smallest number of blocks which, being contaminated, cause the breakdown of the slope estimates. Thus, if k blocks are contaminated, this yields in the worst case kT outliers. Then, to cause the break of the β estimates we require that

$$kT \geq m$$

that is

$$k \geq \frac{m}{T}.$$

So it suffices to take $k = \lceil m/T \rceil$ which implies that the block breakdown point is given by (3.24). □

Computation of the standard errors of the robust estimators

For the computation of the (asymptotic) standard errors we use the formula:

$$\hat{\text{Cov}}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{x}'D_1\mathbf{x})^{-1}(\mathbf{x}'D_2\mathbf{x})(\mathbf{x}'D_1\mathbf{x})^{-1}. \tag{7.1}$$

For the MS approach Maronna and Yohai (2002) showed that the estimator is asymptotically normal. The first order condition (3.22) for the slope estimator $\hat{\beta}_{WMS}$ is the same as for an M-estimator. Hence, the formulas for asymptotic variance of M-estimators can be used (cfr. Huber 1981): D_1 is a $NT \times NT$ diagonal matrix with diagonal elements defined as

$$(D_1)_{it} = W(\frac{r_{it}}{\hat{\sigma}}) + W'(\frac{r_{it}}{\hat{\sigma}})(\frac{r_{it}}{\hat{\sigma}}) = \rho_S''(\frac{r_{it}}{\hat{\sigma}}) \qquad i = 1, ..., N \quad t = 1, \ldots, T$$

with the ρ_S function defined in (3.10) and the weighing function $W(u) = \rho_S'(u)/u$. The $NT \times NT$ diagonal matrix D_2 has elements

$$(D_2)_{it} = W^2(\frac{r_{it}}{\hat{\sigma}})(\frac{r_{it}}{\hat{\sigma}})^2 \qquad i = 1, ..., N \quad t = 1, \ldots, T.$$

Here $\hat{\sigma}$ is the S-scale estimate computed from the residuals $r_{it} = y_{it} - \hat{\beta}_{WMS}' x_{it} - \hat{\alpha}_i(\hat{\beta}_{WMS})$.

In the WGM approach formula (7.1) still applies, with diagonal matrix D_1 defined as

$$(D_1)_{it} = (W_x)_{it}\ \rho''(\frac{r_i}{\hat{\sigma}}) \qquad i = 1, ..., N \quad t = 1, \ldots, T$$

while D_2 has elements

$$(D_2)_{it} = (W_x)_{it}\ W^2(\frac{r_{it}}{\hat{\sigma}})(\frac{r_{it}}{\hat{\sigma}})^2 \qquad i = 1, ..., N \quad t = 1, \ldots, T.$$

As in section 3.1, the ρ function is defined by (3.10), $W(u) = \rho'(u)/u$, W_x is the diagonal weighing matrix defined in (3.13), and $\hat{\sigma} = \hat{\sigma}_{LTS}$. These standard errors are robust with respect to heteroscedastic error terms, and correspond to White standard errors in the classical case.

Table 4: Weights (averaged over the time span) assigned to countries by the WGM and WMS procedures.

Country	WGM	WMS	Country	WGM	WMS	Country	WGM	WMS
Angola	0.73	0.63	Cyprus	0.98	0.95	Kiribati	0.76	0.77
UArbEmir	0.66	0.56	Dominica	0.97	0.95	Korea	0.97	0.94
Argentina	0.83	0.72	DominicR	0.93	0.81	Kuwait	0.65	0.52
Antigua	0.98	0.96	Ecuador	0.97	0.90	Liberia	0.92	0.83
Burundi	0.95	0.87	Egypt, A	0.92	0.81	Sri Lanka	0.98	0.95
Benin	0.89	0.79	Ethiopia	0.97	0.93	Lesotho	0.54	0.36
Burkina	0.93	0.84	Fiji	0.95	0.88	Morocco	0.98	0.94
Bangladesh	0.95	0.87	Gabon	0.68	0.56	Madagascar	0.97	0.93
Bahrain	0.84	0.66	Ghana	0.98	0.94	Mexico	0.92	0.80
Bahamas	0.95	0.89	Guinea	0.99	0.97	Mali	0.93	0.83
Belize	0.85	0.67	Gambia	0.58	0.43	Malta	0.95	0.89
Bolivia	0.66	0.60	Guinea-B	0.81	0.60	Mozambique	0.80	0.66
Brazil	0.75	0.61	Grenada	1.00	0.99	Mauritan	0.83	0.65
Barbados	0.96	0.91	Guatemala	0.98	0.96	Mauritius	0.95	0.91
Bhutan	0.56	0.46	Guyana	0.50	0.35	Malawi	0.94	0.87
Botswana	0.76	0.60	Hong Kong	0.97	0.93	Malaysia	0.95	0.88
CAfricRp	0.92	0.80	Honduras	0.96	0.88	Namibia	0.86	0.75
Chile	0.76	0.62	Haiti	0.97	0.92	Niger	0.88	0.73
China	0.99	0.96	Indonesia	0.95	0.89	Nigeria	0.91	0.78
C d'Ivoire	0.93	0.85	India	0.99	0.98	Nicaragua	0.18	0.53
Cameroon	0.95	0.88	Iran, Is	0.90	0.82	Nepal	0.99	0.97
Congo	0.80	0.65	Israel	0.70	0.65	Oman	0.75	0.51
Colombia	0.98	0.95	Jamaica	0.90	0.78	Pakistan	0.95	0.90
Comoros	0.92	0.81	Jordan	0.94	0.85	Panama	0.83	0.69
Costa Rica	0.97	0.94	Kenya	0.96	0.90	Peru	0.76	0.67

Country	WGM	WMS	Country	WGM	WMS
Philippines	0.97	0.90	Togo	0.87	0.77
PapuaNGu	0.91	0.77	Thailand	0.98	0.96
Paraguay	0.95	0.87	Tonga	0.88	0.71
Rwanda	0.92	0.83	Trinidad	0.87	0.76
Sudan	0.94	0.85	Tunisia	0.97	0.92
Senegal	0.96	0.90	Turkey	0.98	0.95
Singapore	0.96	0.92	Taiwan	0.99	0.98
Solomon	0.65	0.58	Tanzania	0.90	0.80
Sierra Leone	0.96	0.90	Uganda	0.88	0.81
El Salvador	0.95	0.89	Uruguay	0.89	0.85
Somalia	0.63	0.62	Venezuela	0.85	0.71
Sao Tome	0.76	0.70	Vanuatu	0.76	0.56
Suriname	0.72	0.56	S Africa	0.97	0.94
Swazilan	0.87	0.71	Zaire	0.97	0.91
Seychelles	0.72	0.57	Zambia	0.71	0.58
Syria	0.87	0.72	Zimbabwe	0.93	0.84
Chad	0.63	0.47			

8 References

Alvarez, J., Arellano, M. (2004). Robust likelihood estimation of dynamic panel data models. *CEMFI Working Paper* No. 0421.

Baltagi, B.H. (2001). *Econometric Analysis of Panel Data.* 2nd ed. New York, John Wiley.

Beaton, A.E., Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectro-scopic data. *Technometrics* **16**, 147-185.

Croux, C. and Haesbroeck, G. (1999). Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator. *Journal of Multivariate Analysis*, **71**, 161-190.

Davies, P. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics* **15**, 1269-1292.

Davies, P. (1992). The asymptotics of Rousseeuw's minimum voulme ellipsoid estimator. *Annals of Statistics* **4**, 1828-1843.

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point, in *A Festschrift for Erich L. Lehmann*, edited by Bickel, P.J., Doksum, K.A., Hodges Jr., J.L., Wadsworth, 157-184.

Fellner, W.H. (1986). Robust estimation of variance components. *Technometrics* **28**, 51-60.

Franses, P.H., Kloek, T., Lucas, A. (1999). Outlier robust analysis of long-run marketing effects for weekly scanning data. *Journal of Econometrics* **89**, 293-315.

Giavazzi, F., Jappelli, T. and Pagano, M. (2000). Searching for non-linear effects of fiscal policy: evidence from industrial and developing countries. *European Economic Review* **44**, 1259-1289.

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

Hinloopen, J. and Wagenvoort, J.L.M. (1997). On the computation and efficiency of a HBP-GM estimator: some simulation results. *Computational Statistics and Data Analysis* **25**, 1-15.

Huber, P.J. (1981). *Robust Statistics.* New York, John Wiley.

Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics. Theory and Methods.* New York, John Wiley.

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

Müller, C. and Uhlig, S. (2001). Estimation of variance components with high breakdown point and high efficiency. *Biometrika* **88**, 353-366.

Peracchi, F. (1991). Bounded-influence estimators for the SURE model. *Journal of Econometrics* **48**, 119-134.

Rocke, D.M. (1991). Robustness and balance in the mixed model. *Biometrics* **47**, 303-309.

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

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

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

Rousseeuw, P.J. and Van Driessen, K. (2002), Computing LTS Regression for Large Data Sets. *Estadistica*, **54**, 163-190.

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

Rousseeuw, P.J. and Yohai, V.J. (1984) Robust regression by means of S-estimators, in *Robust and Nonlinear Time Series Analysis*, edited by J. Franke, W. Härdle and R.D. Martin, Lecture Notes in Statistics No 26, Springer Verlag, New York, 256-272.

Rousseeuw, P.J. and Wagner, J. (1994). Robust regression with a distributed intercept using least median of squares. *Computational Statistics and Data Analysis* **17**, 65-76.

Ruppert, D. (1992). Computing S-estimators for regression and multivariate location and dispersion. *Journal of Computational Graphical Statistics* **1**, 253-270.

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

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

Zaman, A., Rousseeuw, P.J. and Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economics Letters* **71**, 1-8.