CrossMark

Robust estimation and moment selection in dynamic fixed-effects panel data models

P. Čížek[1] · M. Aquaro[2]

Abstract Considering linear dynamic panel data models with fixed effects, existing outlier–robust estimators based on the median ratio of two consecutive pairs of first-differenced data are extended to higher-order differencing. The estimation procedure is thus based on many pairwise differences and their ratios and is designed to combine high precision and good robust properties. In particular, the proposed two-step GMM estimator based on the corresponding moment equations relies on an innovative weighting scheme reflecting both the variance and bias of those moment equations, where the bias is assumed to stem from data contamination. To estimate the bias, the influence function is derived and evaluated. The robust properties of the estimator are characterized both under contamination by independent additive outliers and the patches of additive outliers. The proposed estimator is additionally compared with existing methods by means of Monte Carlo simulations.

Keywords Dynamic panel data · Generalized method of moments · Influence function · Pairwise differences

✉ P. Čížek
P.Cizek@tilburguniversity.edu

[1] CentER, Department of Econometrics and OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

[2] Joint Research Centre, European Commission, 21027 Ispra, Italy

1 Introduction

Dynamic panel data models with fixed effects have been used in many empirical applications in economics; see Bun and Sarafidis (2015) and Harris et al. (2008) for an overview of the methodology and applications. Despite the complex data structure of dynamic panels, a vast majority of literature focuses on the models assuming that data are free of influential observations or outliers. This is often not the case in reality (Janz 2002; Verardi and Wagner 2011; Zaman et al. 2001), and procedures robust to outliers are thus very important in the case of panel data, where erroneous observations can be easily masked by the complex data structure.

The robust methods for panel data have been studied only to a limited extent till now. There are some methods available for static models (e.g., Bramati and Croux 2007; Aquaro and Čížek 2013) and just a handful for the dynamic models. Locally robust estimation procedures have been proposed by Lucas et al. (2007), based on the generalized method of moment estimator with a bounded influence function, and by Galvao (2011), using quantile regression techniques. On the other hand, Dhaene and Zhu (2017) and Aquaro and Čížek (2014) propose globally robust estimators that are based on the median ratios of the first differences of the dependent variable and of the first- or higher-order differences of the lagged dependent variable [note that previously studied median-unbiased estimation such as Cermeño (1999) was based on the least squares method and was thus not robust to outliers]. The main shortcomings of these methods follow from the use of a fixed number of the differences and their ratios. On the one hand, using just the first differences as in Dhaene and Zhu (2017) can be beneficial for the robustness of the estimator, but it results in a lower precision of estimates. On the other hand, Aquaro and Čížek (2014) employ multiple differences of the explanatory variables to improve the precision of estimation, but it leads to a high sensitivity to sequences of outliers. Additionally, estimation using higher-order differences of the dependent variables has not been explored in neither case.

Our aim is to extend these median-based estimators of Dhaene and Zhu (2017) and Aquaro and Čížek (2014) by employing multiple pairwise difference transformations in such a way that the resulting estimator is robust and also exhibits good finite-sample performance in data without outliers. The use of higher-order differences of the dependent variable is not new (see Aquaro and Čížek 2013), but presents two big challenges when applied in dynamic models. In particular, higher-order differences have not been previously used since (1) they can result in a substantial increase in bias in the presence of particular types of outliers and (2) their number grows quadratically with the number of time periods, which can lead to additional biases due to weak identification or outliers. We address this by proposing a data-driven weighting and selection of the median ratios of differenced data since the traditional strategy used in the robust statistics—using an initial robust estimator to detect outlying observations, and after removing them, applying an efficient non-robust estimator (c.f., Gervini and Yohai 2002)—is not feasible in this context. Even in the case of using the first differences only, removing a single observation means that the observation and its two or three following data points (depending on the actual estimation method) cannot be used in estimation. Especially in short panels with less than five time periods, removing a single observation for a given individual thus means that no observations

of that individual can be used in estimation and the problem gets worse if higher-order differences are used.

In this paper, we generalize the estimation method of Dhaene and Zhu (2017) to a combination of the pth and sth order differences, $p, s \in \mathbb{N}$, and combine multiple pairwise differences by means of the generalized method of moments (GMM). To account for the shortcomings of the current methods and to extend the analysis of Aquaro and Čížek (2014), we first analyze the robustness of the median-based moment conditions, derive their influence functions, and quantify the bias caused by data contamination. Subsequently, we use the maximum bias and propose a two-step GMM estimator, which weights the (median-based) moment conditions both by their variance and bias; this guarantees that imprecise or biased moment conditions get low weights in estimation. Finally, as the number of applicable moment conditions grows quadratically with the number of time periods, a suitable number of moment conditions for the underlying data generating process needs to be selected using a robust version of moment selection procedure of Hall et al. (2007).

In the rest of the paper, the new estimator is introduced first in Sect. 2. Its robust properties are studied in Sect. 3 and are used to define the data-dependent GMM weights. The existing and proposed methods are then compared by means of Monte Carlo simulations in Sect. 4 and the proofs can be found in the "Appendix".

2 Median-based estimation of dynamic panel models

The dynamic panel data model (Sect. 2.1) and its median-based estimation (Sect. 2.2) will be now discussed. Later, the two-step GMM estimation procedure (Sect. 2.3) and the moment selection method (Sect. 2.4) will be introduced.

2.1 Dynamic panel data model

Consider the simple dynamic panel data model ($i = 1, \ldots, n; t = 1, \ldots, T; T \geq 3$)

$$y_{it} = \alpha y_{it-1} + \eta_i + \varepsilon_{it}, \tag{1}$$

where y_{it} denotes the response variable, η_i is the unobservable fixed effect, and ε_{it} represents the idiosyncratic error. Parameter $|\alpha| < 1$ so that this data generating process can be stationary. The number T of time periods is fixed, which implies that fixed or stochastic effects η_i are nuisance parameters and cannot be consistently estimated. Finally note that the extension of the discussed estimators to a model with exogenous covariates is straightforward (see Dhaene and Zhu 2017, Section 4.1).

As in Aquaro and Čížek (2014) and similarly to Cermeño (1999) and Han et al. (2014), we will consider model (1) under the following assumptions:

A.1 Errors ε_{it} are independent across $i = 1, \ldots, n$ and $t = 1, \ldots, T$ and possess finite second moments. Errors $\{\varepsilon_{it}\}_{t=1}^{T}$ are also independent of fixed effects η_i.

A.2 The sequences $\{y_{it}\}_{t=1}^{T}$ are time stationary for all $i = 1, \ldots, n$. In particular, the first and second moments of y_{it} conditional on η_i exist and do not depend of time.

A.3 Errors $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 > 0$, for all $i = 1, \ldots, n$ and $t = 1, \ldots, T$.

Except of the independence in Assumption A.1, there are no assumptions are made about the unobservable fixed effects η_i. Although we impose rather strict Assumptions A.1 and A.3 on idiosyncratic errors, they can be relaxed. The errors ε_{it} do not have to follow the same distribution across cross-sectional units i, allowing for heteroscedasticity. Additionally, the consistency of the estimators introduced below requires that the joint distributions of errors $\{\varepsilon_{it}\}_{t=1}^{T}$ are elliptically contoured, making the normality Assumption A.3 sufficient, but not necessary (see Dhaene and Zhu 2017, Section 4.2). On the other hand, the violation of the time-homoscedasticity in Assumption A.3 leads to the inconsistency of the discussed estimators. If $\varepsilon_{it} \sim N(0, \sigma_{\epsilon t}^2)$ for $t = 1, \ldots, T$, the model equation (1) has to be therefore rescaled by unknown standard deviations $\sigma_{\epsilon t}$, which can be treated as unknown parameters and estimated along with α by GMM. Finally, the stationarity Assumption A.2 is used not only by the proposed estimators, but also by frequently applied GMM estimators such as Blundell and Bond (1998) and it is implied by the assumptions of Han et al. (2014) if $|\alpha| < 1$.

2.2 Median-based moment conditions

To generalize the estimator by Dhaene and Zhu (2017), let Δ^s denote the sth difference operator, that is, $\Delta^s \upsilon_t := \upsilon_t - \upsilon_{t-s}$ (cf. Abrevaya 2000; Aquaro and Čížek 2013). Given model (1), stationarity Assumption A.2 implies for any integers $s, q, p \in \mathbb{N}$ that

$$E(\Delta^s y_{it} | \Delta^p y_{it-q}) = r_j \Delta^p y_{it-q}, \tag{2}$$

where the triplet $j = (s, q, p)$ and $r_j = \text{cov}(\Delta^s y_{it}, \Delta^p y_{it-q})/\text{var}(\Delta^p y_{it-q})$ are independent of i and t, $\max\{s, p+q\} < T$. Consequently, the variables $\Delta^s y_{it} - r_j \Delta^p y_{it-q}$ and $\Delta^p y_{it-q}$ are uncorrelated, and by Assumption A.3, independent and symmetrically distributed around zero. Hence, $E[\text{sgn}(\Delta^s y_{it} - r_j \Delta^p y_{it-q}) \, \text{sgn}(\Delta^p y_{it-q})] = 0$ and $E\left[\text{sgn}\left(\Delta^s y_{it}/\Delta^p y_{it-q} - r_j\right)\right] = 0$. The estimation of r_j can be therefore based on the sample analog of this moment condition:

$$\hat{r}_{nj} = \text{med}\left\{\frac{\Delta^s y_{it}}{\Delta^p y_{it-q}}; \ t = p + q + 1, \ldots, T; \ i = 1, \ldots, n\right\}. \tag{3}$$

To relate r_j to the autoregressive coefficient α in (1), Aquaro and Čížek (2014) derived under Assumptions A.1 and A.2 that the correlation coefficient r_j satisfies the moment condition

$$g_j(\alpha) = 2(1 - \alpha^p)r_j - \alpha^q + \alpha^{q+p} + \alpha^{|s-q|} - \alpha^{|s-p-q|} = 0. \tag{4}$$

If $s = q = p = 1$, (4) defines Dhaene and Zhu (2017)'s estimator: $\alpha \in (-1, 1)$ is identified by $g_{111}(\alpha) = (1 - \alpha)(2r_{111} + 1 - \alpha) = 0$. Dhaene and Zhu's (DZ) estimator $\hat{\alpha}_n$ therefore simply equals to $2\hat{r}_{n111} + 1$ and it was proved to be consistent

and asymptotically normal. Aquaro and Čížek (2014)'s estimator (AC-DZ) of α uses $s = q = 1$ and p being odd, $p < T - 1$. They do not use differences with $s > 1$ due to their robustness properties: while they seem robust to sequences of outliers, they can lead to large biases if outliers occur at random times.

2.3 Two-step GMM estimation

To increase the precision and robustness of the estimation, we propose to extend the (AC-)DZ estimator by allowing for multiple differences with $s = q \geq 1$ and $p \geq 1$. We consider only $s = q$ as the moment conditions (4) do not allow distinguishing outlying and regular observations for $s \neq q$ as shown in Aquaro and Čížek (2014). For $s = q$, (4) simplifies after dividing by $1 - \alpha^p$ and accordingly redefining $g_j(\alpha)$ to

$$g_j(\alpha) = 2r_j + 1 - \alpha^s = 0. \tag{5}$$

The full set of moment conditions in (5) can be then written as

$$\boldsymbol{g}(\alpha) = \boldsymbol{0}, \tag{6}$$

where $\boldsymbol{g}(\alpha) = \{g_j(\alpha)\}_{j \in \mathcal{J}}$ and a fixed finite set $\mathcal{J}$ contains all triplets $\boldsymbol{j} = (s, q, p)$ that are considered in estimation. The DZ estimator then corresponds to the special case $\mathcal{J} = \{(1, 1, 1)\}$ and the AC-DZ relies on a set $\mathcal{J} = \{(1, 1, p): 1 \leq p < T - 1 \text{ odd}\}$. Here we consider all combinations with any $s = q$ odd and p odd, $\mathcal{J} \subseteq \mathcal{J}_o = \{(s, s, p): s \in \mathbb{N} \text{ odd}, p \in \mathbb{N} \text{ odd}, 1 \leq s + p < T\}$, as the single moment conditions do not identify uniquely α for even values of s or p and this could negatively affect the bias caused by contamination.

Given the system of equations in (6), the parameter α can be estimated by the GMM procedure. This GMM estimator is referred to here as the pairwise-difference DZ (PD-DZ) estimator and is defined by

$$\hat{\alpha}_n = \arg \min_{c \in (-1,1)} \boldsymbol{g}_n(c)' \boldsymbol{A}_n \boldsymbol{g}_n(c), \tag{7}$$

where $\boldsymbol{g}_n(c) = (g_{nj}(c))_{j \in \mathcal{J}}$ is the sample analog of $\boldsymbol{g}(\alpha)$ and corresponds to (5) with r_j being replaced by $\hat{r}_{nj}$ defined in (3).

The weighting matrix $\boldsymbol{A}_n$ can be initially chosen as in Aquaro and Čížek (2014) proportional to the number of observations available for the estimation of each moment equation: $\boldsymbol{A}_n = \boldsymbol{A} = \text{diag}\{(T-p-s)/T\}$. The traditional variance-minimizing choice of the GMM weighting matrix $\boldsymbol{A}_n$ however equals the inverse of the variance matrix $\boldsymbol{V}_n$ of the moment conditions $\boldsymbol{g}_n(\alpha)$, which converges to the asymptotic variance matrix $\boldsymbol{V}$ of the moment conditions (5); see the Appendix for the asymptotic distribution of $\hat{\alpha}_n$ and the asymptotic variance matrix $\boldsymbol{V}$ previously obtained by Aquaro and Čížek (2014).

On the other hand, we aim to account also for the presence of outlying observations that can substantially bias the estimates. Since simply removing outliers would result in a substantial data loss as explained in the introduction, we propose instead to use

the moment conditions (5) and minimize the mean squared error (MSE) of estimates instead of the asymptotic variance. First, let us denote the MSE of $\boldsymbol{g}_n(\alpha)$ by $\boldsymbol{W}_n$,

$$\boldsymbol{W}_n = MSE\{\boldsymbol{g}_n(\alpha)\} = Bias\{\boldsymbol{g}_n(\alpha)\}Bias\{\boldsymbol{g}_n(\alpha)\}' + Var\{\boldsymbol{g}_n(\alpha)\} = \boldsymbol{b}_n\boldsymbol{b}_n' + \boldsymbol{V}_n.$$

Given a weighting matrix $\boldsymbol{A}_n$ and the asymptotic linearity of $\hat{\alpha}_n$ (see Aquaro and Čížek 2014, the proof of Theorem 1)

$$\hat{\alpha}_n - \alpha = (\boldsymbol{d}'\boldsymbol{A}_n\boldsymbol{d})^{-1}\boldsymbol{d}'\boldsymbol{A}_n\boldsymbol{g}_n(\alpha) + o_p(1) \qquad (8)$$

as $n \to \infty$, it immediately follows that the MSE of $\hat{\alpha}_n$ equals

$$(\boldsymbol{d}'\boldsymbol{A}_n\boldsymbol{d})^{-1}\boldsymbol{d}'\boldsymbol{A}_n\boldsymbol{W}_n\boldsymbol{A}_n\boldsymbol{d}(\boldsymbol{d}'\boldsymbol{A}_n\boldsymbol{d})^{-1} + o_p(1),$$

which is (asymptotically) minimized by choosing $\boldsymbol{A}_n = \boldsymbol{W}_n^{-1}$ (Hansen 1982, Theorem 3.2).

Next, to create a feasible procedure, both the variance and squared bias matrices have to be estimated. The estimation thus proceeds in two steps: first, the (AC-)DZ estimator is applied to obtain an initial parameter estimates; then—after estimating the bias $\boldsymbol{b}_n$ and variance $\boldsymbol{V}_n$ of moment conditions—the GMM estimator with all applicable pairwise differences is evaluated using an estimate of the weighting matrix $\boldsymbol{A}_n = [\boldsymbol{b}_n\boldsymbol{b}_n' + \boldsymbol{V}_n]^{-1}$. On the one hand, the estimate $\hat{\boldsymbol{V}}_n$ of $\boldsymbol{V}_n$ can be directly obtained from Theorem 5 in the "Appendix" using initial estimates of r_j and α because both the responses y_{it} as well as estimates $\hat{\alpha}_n$ are continuously distributed with bounded densities due to the stationarity Assumptions A.2 and A.3. On the other hand, estimating $\boldsymbol{b}_n$ by $\hat{\boldsymbol{b}}_n$ requires first studying the biases of median-based moment conditions and constructing a feasible estimate thereof in Sect. 3. Using estimates $\hat{\boldsymbol{V}}_n$ and $\hat{\boldsymbol{b}}_n$ to construct $\hat{\boldsymbol{W}}_n = \hat{\boldsymbol{b}}_n\hat{\boldsymbol{b}}_n' + \hat{\boldsymbol{V}}_n$ and $\hat{\boldsymbol{A}}_n = \hat{\boldsymbol{W}}_n^{-1}$ then leads to the proposed second-step GMM estimator

$$\hat{\alpha}_n = \underset{c\in(-1,1)}{\arg\min} \boldsymbol{g}_n(c)'\hat{\boldsymbol{A}}_n\boldsymbol{g}_n(c) = \underset{c\in(-1,1)}{\arg\min} \boldsymbol{g}_n(c)'[\hat{\boldsymbol{b}}_n\hat{\boldsymbol{b}}_n' + \hat{\boldsymbol{V}}_n]^{-1}\boldsymbol{g}_n(c). \qquad (9)$$

2.4 Robust moment selection

The proposed two-step GMM estimator is based on the moment conditions (5), and given that we consider only odd s and p, their number equals approximately $T(T-1)/8$ and grows quadratically with the number of time periods. Although the extra moment conditions based on higher-order differences might improve precision of estimation for larger values of $|\alpha|$, their usefulness is rather limited if α is close to zero. At the same time, a large number of moment conditions might increase estimation bias due to outliers. More specifically, Aquaro and Čížek (2014) showed for α close to 0 that the original moment condition of the DZ estimator $s = q = p = 1$ is least sensitive to random outliers, for instance; including higher-order moment

conditions then just increases bias, does not improve the variance, and is thus harmful.

To account for this, we propose to select the moment conditions used in estimation by a robust analog of a moment selection criterion (e.g., see Cheng and Liao 2015, for an overview). Since all moments are valid and no weak instruments are involved, the information content of the moment equations and their number have to be balanced as in Hall et al. (2007), whose approach to moment selection in the presence of nearly redundant moment conditions can be adapted to robust estimation. They propose the so-called relevant moment selection criterion (RMSC) that—for a given set of moment conditions defined by triplets $\mathcal{J}$ in our case—equals

$$RMSC(\mathcal{J}) = \ln(|\hat{V}_{n,\mathcal{J}}|) + \kappa(|\mathcal{J}|, n).$$

Matrix $\hat{V}_{n,\mathcal{J}}$ represents an estimate of the variance matrix $V_{\mathcal{J}}$ of moment conditions (6) defined by triplets $\mathcal{J}$ and $\kappa(\cdot, \cdot)$ is a deterministic penalty term depending on the number $|\mathcal{J}|$ of triplets (or moment conditions) and on the sample size n used for estimating the elements of V_n (see Theorem 5). To select relevant moment conditions, this criterion has to be minimized:

$$\hat{\mathcal{J}} = \arg \min_{\mathcal{J} \subseteq \mathcal{J}_o} RMSC(\mathcal{J}).$$

Two examples of the penalization term used by Hall et al. (2007) are the Bayesian information criterion (BIC) with $\kappa(c, n) = (c - K) \cdot \ln(\sqrt{n})/\sqrt{n}$ and the Hannan–Quinn information criterion (HQIC) with $\kappa(c, n) = (c - K) \cdot \kappa_c \ln(\ln(\sqrt{n}))/\sqrt{n}$, where the number of estimated parameters $K = 1$ in model (1) and constant $\kappa_c > 2$.

As in Sect. 2.3, the proposed robust estimator (9) should minimize the MSE error rather than just the variance of the estimates. We therefore suggest to use the relevant robust moment selection criterion (RRMSC),

$$RRMSC(\mathcal{J}) = \ln(|\hat{W}_{n,\mathcal{J}}|) + \kappa(|\mathcal{J}|, n), \tag{10}$$

which is based on the determinant of an estimate $\hat{W}_n$ of the MSE matrix W_n rather than on the variance matrix estimate $\hat{V}_n$ of the moment conditions. The relevant robust moment conditions are then obtained by minimizing

$$\hat{\mathcal{J}} = \arg \min_{\mathcal{J} \subseteq \mathcal{J}_o} RRMSC(\mathcal{J}).$$

3 Robustness properties

There are many measures of robustness that are related to the bias of an estimator, or more typically, the worst-case bias of an estimator due to an unknown form of outlier contamination. In this section, various kinds of contamination are introduced and some relevant measures of robustness are defined (Sect. 3.1). Using these measures,

we characterize the robustness of moment conditions (5) in Sect. 3.2 and the robustness of the GMM estimator (7) in Sect. 3.3. Next, we use these results to estimate of the bias of the moment conditions (5) as discussed in Sect. 3.4. Finally, the whole estimation procedure is summarized in Sect. 3.5.

3.1 Measures of robustness

Given that the analyzed data from model (1) are dependent, the effect of outliers can depend on their structure. Therefore, we first describe the considered contamination schemes and then the relevant measures of robustness.

More formally, let $\mathcal{Z}$ be the set of all possible samples $Z = \{z_{it}\}$ of size (n, T) following model (1) and let $Z_\epsilon = \{z_{it}^\epsilon\}$ be a contaminating sample of size (n, T) following a fixed data-generating process, where the index ϵ of Z_ϵ indicates the probability that an observations in Z_ϵ is different from zero. The observed contaminated sample is $Z + Z_\epsilon = \{z_{it} + z_{it}^\epsilon\}_{i=1, t=1}^{n, \ T}$. Similarly to Dhaene and Zhu (2017), we consider the contamination by independent additive outliers following a degenerate distribution with the point mass at ζ,

$$
\begin{aligned}
Z_{\epsilon, \zeta}^1 = \{z_{it}^\epsilon\}_{i=1, t=1}^{n, \ T} &= \{\zeta \cdot I(v_{it}^\epsilon = 1)\}_{i=1, t=1}^{n, \ T}, \\
P(v_{it}^\epsilon = 1) &= \epsilon, \quad P(v_{it}^\epsilon = 0) = 1 - \epsilon,
\end{aligned}
\tag{11}
$$

and by patches of k additive outliers,

$$
Z_{\epsilon, \zeta}^2 = \{z_{it}^\epsilon\}_{i=1, t=1}^{n, \ T} = \left\{\zeta \cdot I\left(v_{it}^\epsilon = 1 \text{ or } \ldots \text{ or } v_{it-k+1}^\epsilon = 1\right)\right\}_{i=1, \ t=1}^{n, \ T},
\tag{12}
$$

where v_{it}^ϵ follows the Bernoulli distribution with the parameter $\tilde{\epsilon}$ such that $(1 - \tilde{\epsilon})^k = 1 - \epsilon$. Additionally, a third contamination scheme $Z_{\epsilon, \zeta}^3 = \{z_{it}^\epsilon\}_{i=1, t=1}^{n, \ T}$ is considered, where

$$
z_{it}^\epsilon = \begin{cases} a_{it-l}(-1)^l & \text{if the smallest index } l \geq 0 \text{ with } v_{it-l}^\epsilon = 1 \text{ satisfies } l \leq k - 1, \\ 0 & \text{otherwise,} \end{cases}
\tag{13}
$$

where $\Pr(a_{it-l} = \zeta) = 1/2$ and $\Pr(a_{it-l} = -\zeta) = 1/2$ and where v_{it}^ϵ is defined as in $Z_{\epsilon, \zeta}^2$. Note that (12) and (13) are special cases of a more general type of contamination $Z_{\epsilon, \zeta}^4 = \{z_{it}^\epsilon\}_{i=1, t=1}^{n, \ T}$, where

$$
z_{it}^\epsilon = \begin{cases} a_{it-l}\rho^l & \text{if the smallest index } l \geq 0 \text{ with } v_{it-l}^\epsilon = 1 \text{ satisfies } l \leq k - 1, \\ 0 & \text{otherwise,} \end{cases}
\tag{14}
$$

and $-1 \leq \rho \leq 1$. Note that this general type of contamination closely corresponds to the contamination by innovation outliers for large k and $\rho = \alpha$ and it is therefore important to study. As we can however conjecture from Dhaene and Zhu (2017)'s results for $s = p = 1$ that the contamination scheme $Z_{\epsilon, \zeta}^4$ biases estimates towards ρ

for $\zeta \rightarrow +\infty$ and ρ is unknown in practice, we are not analysing this most general case with $\rho \in [-1, 1]$. Instead, we concentrate on the most extreme cases of $\rho = 1$ and $\rho = -1$ as they can arguably bias the estimate most. Hence, the contamination schemes $Z_{\epsilon,\zeta}^{1}$, $Z_{\epsilon,\zeta}^{2}$, and $Z_{\epsilon,\zeta}^{3}$ bias the DZ estimates of α towards 0, 1, and -1, respectively—see Sect. 3.2 and Dhaene and Zhu (2017).

Given the contamination schemes, one of the traditional measures of the global robustness of an estimator is the breakdown point. It can be defined as the smallest fraction of the data that can be changed in such a way that the estimator will not reflect any information concerning the remaining (non-contaminated) observations. Aquaro and Čížek (2014) derived the breakdown points of the estimators $\hat{r}_j$, $j \in \mathcal{J}$, for contamination schemes $Z_{\epsilon,\zeta}^{1}$, $Z_{\epsilon,\zeta}^{2}$, and $Z_{\epsilon,\zeta}^{3}$, and under some regularity conditions, proved that the breakdown point of the GMM estimator (7) equals the breakdown point of the DZ estimator $\hat{r}_{(1,1,1)}$ if $(1, 1, 1) \in \mathcal{J}$. While such results characterize the global robustness of the PD-DZ estimators, they are not informative about the size of the bias caused by outliers.

We therefore base the estimation of the bias due to contamination on the influence function. It is a traditional measure of local robustness and can be defined as follows. Let $\mathcal{T}(Z + Z_{\epsilon})$ denote a generic estimator of an unknown parameter θ based on a contaminated sample $Z + Z_{\epsilon} = \{z_{it} + z_{it}^{\epsilon}\}_{i=1, t=1}^{n, T}$, where Z and Z_{ϵ} have been defined at the beginning of Sect. 3. As the definition is asymptotic, let $\mathcal{T}(\theta, \zeta, \epsilon, T)$ be the probability limit of $\mathcal{T}(Z + Z_{\epsilon})$ when T is fixed and $n \rightarrow \infty$. Note that $\mathcal{T}(\theta, \zeta, \epsilon, T)$ depends on the unknown parameter θ describing the data generating process, on the fraction ϵ of data contamination, on the non-zero value ζ characterizing the outliers, and on the number of time periods T. Assume $\mathcal{T}$ is consistent under non-contaminated data, that is, $\mathcal{T}(\theta, \zeta, 0, T) = \theta$. The influence function (IF) of estimator $\mathcal{T}$ at data generating process Z due to contamination Z_{ϵ} is defined as

$$\text{IF}\left(\mathcal{T}; \theta, \zeta, T\right) := \lim_{\epsilon \rightarrow 0} \frac{\mathcal{T}(\theta, \zeta, \epsilon, T) - \theta}{\epsilon} = \left. \frac{\partial \, \text{bias}(\mathcal{T}; \theta, \zeta, \epsilon, T)}{\partial \epsilon} \right|_{\epsilon=0}, \quad (15)$$

where the equality follows by the definition of asymptotic bias of $\mathcal{T}$ due to the data contamination Z_{ϵ}, $\text{bias}(\mathcal{T}; \theta, \zeta, \epsilon, T) := \mathcal{T}(\theta, \zeta, \epsilon, T) - \theta$. (If IF does not depend on the number T of time periods, T can be omitted from its arguments.)

Clearly, the knowledge of the influence function allows us to approximate the bias of an estimator $\mathcal{T}$ at $Z + Z_{\epsilon}$ by $\epsilon \cdot \text{IF}(\mathcal{T}; \theta, \zeta, T)$. Although such an approximation is often valid only for small values of $\epsilon > 0$ (e.g., in the linear regression model, where the bias can get infinite), it is relevant in a much wider range of contamination levels ϵ in model (1) given that the parameter space $(-1, 1)$ is bounded and so is the bias (the dependence of the bias on the contamination level ϵ has been studied by Dhaene and Zhu 2017).

The disadvantage of approximating bias by $\epsilon \cdot \text{IF}(\mathcal{T}; \theta, \zeta, T)$ is that it depends on the unknown magnitude ζ of outliers. We therefore suggest to evaluate the supremum of the influence function, the gross error sensitivity (GES)

$$\text{GES}(\mathcal{T}; \theta, T) = \sup_{\zeta} |\text{IF}(\mathcal{T}; \theta, \zeta, T)|, \tag{16}$$

and approximate the worst-case bias by $\epsilon \cdot \text{GES}(\mathcal{T}; \theta, T)$. For the PD-DZ estimator and the corresponding moment conditions, IF and GES are derived in the following Sects. 3.2 and 3.3, where $\mathcal{T}$ will equal to $\hat{\alpha}$ and $\hat{r}_j$, respectively (without the subscript n since the IF and GES definitions depend only on the probability limits of the estimators).

3.2 Influence function

The GMM estimator (7) is based on moment conditions depending on the data only by means of the medians r_j. We therefore derive first the influence functions of the estimators $\hat{r}_j$ and then combine them to derive the influence function of the GMM estimator. Building on Dhaene and Zhu (2017, Theorems 3.2 and 3.7), the IFs of $\hat{r}_j$ in model (1) under contamination schemes $Z^1_{\epsilon,\zeta}$, $Z^2_{\epsilon,\zeta}$, and $Z^3_{\epsilon,\zeta}$ are derived in the following Theorems 1–3. Only the point-mass distribution G_ζ with the mass at $\zeta \in R$ is considered. In all theorems, Φ denotes the cumulative distribution function of the standard normal distribution $N(0, 1)$.

Theorem 1 *Let Assumptions* A.1–A.3 *hold and* $j \in \mathcal{J}_o$. *Then it holds in model* (1) *under the independent-additive-outlier contamination* $Z^1_{\epsilon,\zeta}$ *with point-mass distribution at* $\zeta \neq 0$ *that*

$$\text{IF}(\hat{r}_j; \alpha, \zeta) = -\pi \sqrt{\frac{1 - \alpha^s}{1 - \alpha^p} - \frac{1}{4}(1 - \alpha^s)^2}$$

$$\times \left[\Phi\left(\frac{\zeta(1 + \alpha^s)/2}{\sqrt{2\frac{\sigma_\varepsilon^2}{1-\alpha^2}\left(1 - \alpha^s - \frac{(1-\alpha^s)^2}{4}(1 - \alpha^p)\right)}} \right) \right.$$

$$\left. - \Phi\left(\frac{-\zeta(1 - \alpha^s)/2}{\sqrt{2\frac{\sigma_\varepsilon^2}{1-\alpha^2}\left(1 - \alpha^s - \frac{(1-\alpha^s)^2}{4}(1 - \alpha^p)\right)}} \right) \right]$$

$$\times \left[\Phi\left(\frac{\zeta}{\sqrt{2\sigma_\varepsilon^2 \frac{1-\alpha^p}{1-\alpha^2}}} \right) - \Phi\left(-\frac{\zeta}{\sqrt{2\sigma_\varepsilon^2 \frac{1-\alpha^p}{1-\alpha^2}}} \right) \right]. \tag{17}$$

Theorem 2 *Let Assumptions* A.1–A.3 *hold and* $j \in \mathcal{J}_o$. *Then it holds in model* (1) *under the patched-additive-outlier contamination* $Z^2_{\epsilon,\zeta}$ *with point-mass distribution at* $\zeta \neq 0$ *and patch length* $k \geq 2$ *that*

$$\text{IF}(\hat{r}_j; \alpha, \zeta) = -\frac{\pi}{k} \sqrt{\frac{1-\alpha^s}{1-\alpha^p} - \frac{(1-\alpha^s)^2}{4}}$$
$$\times \left[\mathfrak{p}'_C(0) \left(C(r_j; \zeta, 0) - \frac{1}{2} \right) + \mathfrak{p}'_D(0) \left(D(r_j; \zeta, 0) - \frac{1}{2} \right) \right], \tag{18}$$

where $\mathfrak{p}'_C(0)$, $\mathfrak{p}'_D(0)$, $C(r_j; \zeta, 0)$, and $D(r_j; \zeta, 0)$ are defined in (52), (53), (56), and (57), respectively.

Theorem 3 *Let Assumptions* A.1–A.3 *hold and* $j \in \mathcal{J}_o$. *Then it holds in model* (1) *under the patched-additive-outlier contamination* $Z^3_{\epsilon,\zeta}$ *with point-mass distribution at* $\zeta \neq 0$ *and patch length* $k \geq 2$ *that*

$$\text{IF}(\hat{r}_j; \alpha, \zeta) = -\frac{\pi}{k} \sqrt{\frac{1-\alpha^s}{1-\alpha^p} - \frac{(1-\alpha^s)^2}{4}}$$
$$\times \left[\mathfrak{p}'_C \mathcal{C} \left(\frac{1}{2} \right) + \mathfrak{p}'_D \mathcal{D} \left(\frac{1}{2} \right) + \mathfrak{p}'_E \mathcal{E} \left(\frac{1}{2} \right) + \mathfrak{p}'_G \mathcal{G} \left(\frac{1}{2} \right) + \mathfrak{p}'_I \mathcal{I} \left(\frac{1}{2} \right) \right] \tag{19}$$

where $\mathfrak{p}'_L$, $L \in \{C, D, E, G, I\}$, are defined in Eqs. (73), (74), (75), (77), (79), $\mathcal{L}(1/2) = L(r_j; \zeta, 0) - 1/2$ for $\mathcal{L} \in \{\mathcal{C}, \mathcal{D}, \mathcal{E}, \mathcal{G}, \mathcal{I}\}$ and $L \in \{C, D, E, G, I\}$, and $L(r_j; \zeta, 0)$ for $L \in \{C, D, E, G, I\}$ are defined in Eqs. (82)–(86) in "Appendix A.3".

The influence functions reported in Theorems 1–3 are complicated objects both due to their algebraic forms and their dependence on the unknown parameter values α and ζ. As ζ is generally unknown, we characterize the worst-case scenario by means of the gross error sensitivity: recall that $\text{GES}(\hat{r}_j; \alpha) = \sup_\zeta \left| \text{IF}(\hat{r}_j; \alpha, \zeta) \right|$ by Eq. (16). Inspection of the influence functions and their elements in Theorems 1–3 reveals though that the largest effect can be attributed to outliers with magnitude $|\zeta| \to +\infty$ (possibly with an exception of term $\mathcal{E}(1/2)$ in Theorem 3).

Given the results in Theorems 1–3, we thus have to compute the GES of estimators $\hat{r}_j$ numerically for each $j = (s, s, p) \in \mathcal{J}_o$ and $\alpha \in (-1, 1)$. Although this might be relatively demanding if T is large and a dense grid for α is used, note that the GES values are asymptotic and independent of a particular data set. They have to be therefore evaluated just once and then used repeatedly during any application of the proposed PD-DZ estimator. We computed the GES of $\hat{r}_j$ for $j \in \{(s, s, p); s = 1, 3, 5, 7$ and $p = 1, 3, 5, 7, 9, 11\}$ with the variance σ^2_ε set equal to one without loss of generality. The results corresponding to Theorems 1–3 are depicted on Figs. 1, 2 and 3. Irrespective of the contamination scheme, most GES curves display typically higher sensitivity to outliers for $|\alpha|$ close to one than for values of the autoregressive parameter around zero. One can also see that the DZ estimator corresponding to $s = 1$ and $p = 1$ is indeed biased towards 0, 1, and -1 for the contamination schemes $Z^1_{\epsilon,\zeta}$, $Z^2_{\epsilon,\zeta}$, and $Z^3_{\epsilon,\zeta}$, respectively. Concerning the higher-order differences we propose to add to the (AC-)DZ methods, Fig. 1 documents they do exhibit high sensitivity to independent
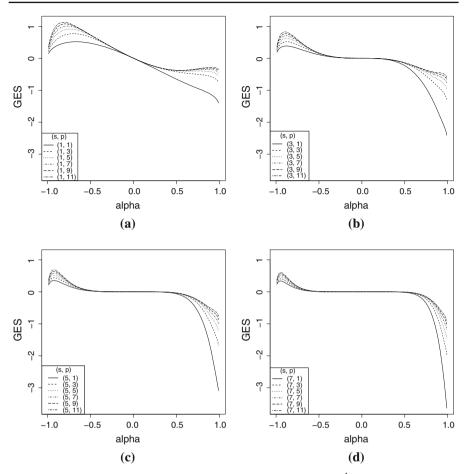
Fig. 1 Gross-error sensitivity of $\hat{r}_j$, $j = (s, s, p) \in \mathcal{J}_o$, under contamination $Z^1_{\epsilon,\zeta}$ by independent additive outliers. **a** $s = 1$, **b** $s = 3$, **c** $s = 5$, **d** $s = 7$

outliers. On the other hand, their sensitivity to the patches of outliers on Fig. 2, for instance, decreases with an increasing s and becomes very low (relative to $s = 1$ and $p \geq 1$) if s is larger than the patch length k, for example, $s = 7 > k = 6$.

3.3 Robust properties of the GMM estimator $\hat{\alpha}_n$

Given the results of the previous sections, we will now analyze the robust properties of the general GMM estimator $\hat{\alpha}$ based on moment equations (6) for $j = (s, s, p) \in \mathcal{J}_o$. The results are stated first for the initial PD-DZ estimator (7) with a deterministic weight matrix and later for the second step of the PD-DZ estimator (9). Since the weight matrix and the bias in particular can be estimated in different ways, we consider in the latter case a weight matrix as a general function of the parameter α and the considered fraction ϵ of outliers.
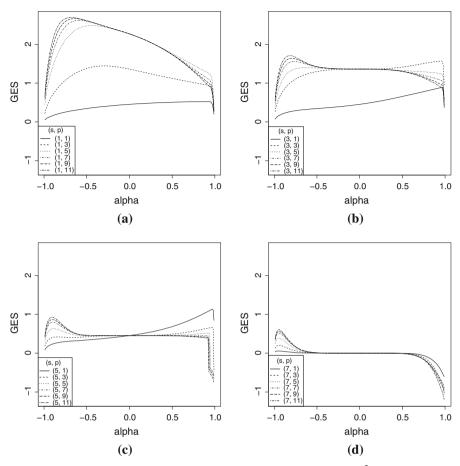
Fig. 2 Gross-error sensitivity of $\hat{r}_j$, $j = (s, s, p) \in \mathcal{J}_o$, under contamination $Z_{\epsilon,\zeta}^2$ by patch additive outliers, length of the path $k = 6$. **a** $s = 1$, **b** $s = 3$, **c** $s = 5$, **d** $s = 7$

Theorem 4 *Consider a particular additive outlier contamination Z_ϵ occurring with probability ϵ, where $0 < \epsilon < 1$. Further, let $\mathcal{J} \subseteq \mathcal{J}_o$.*

First, assume that $A_n = A^0$ is a positive definite diagonal matrix. Then the influence function of the GMM estimator $\hat{\alpha}^0$ using moment conditions indexed by $\mathcal{J}$ is given by

$$\mathrm{IF}(\hat{\alpha}^0; \alpha, \zeta) = -(d' A^0 d)^{-1} d' A^0 \psi, \tag{20}$$

where d is defined in Theorem 5 and ψ is the $|\mathcal{J}| \times 1$ vector of the influence functions of each single $\hat{r}_j$, $\psi = \big(\mathrm{IF}(\hat{r}_j; \alpha, \zeta) \big)_{j \in \mathcal{J}}$.

Next, assume that $A_n = A(\hat{\alpha}_n^0, \epsilon)$ is a positive definite matrix function of the initial estimate $\hat{\alpha}_n^0$ based on the deterministic weight matrix A^0. If $A_n = A(\hat{\alpha}_n^0, \epsilon) \to A = A(\alpha, \epsilon)$ has a finite probability limit and bounded influence function as $n \to \infty$, then the influence function of $\hat{\alpha}$ using moment conditions indexed by $\mathcal{J}$ is again given by
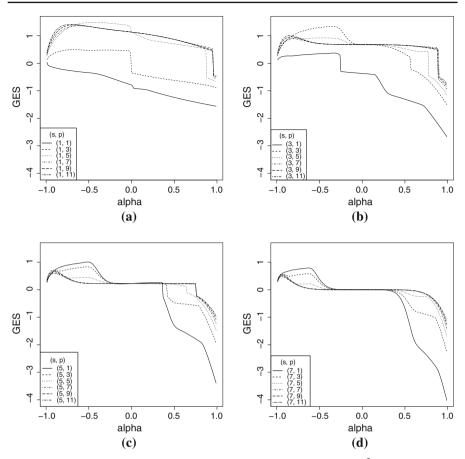
Fig. 3 Gross-error sensitivity of $\hat{r}_j$, $j = (s, s, p) \in \mathcal{J}_o$, under contamination $Z^3_{\epsilon,\zeta}$ by patch additive outliers, length of the path $k = 6$. **a** $s = 1$, **b** $s = 3$, **c** $s = 5$, **d** $s = 7$

$$\text{IF}(\hat{\alpha}; \alpha, \zeta) = -(\boldsymbol{d}'\boldsymbol{A}\boldsymbol{d})^{-1}\boldsymbol{d}'\boldsymbol{A}\boldsymbol{\psi}. \tag{21}$$

Contrary to the breakdown point of Aquaro and Čížek (2014) mentioned earlier, the bias of the proposed PD-DZ estimators is a linear combination of the biases of the individual moment conditions depending on $\hat{r}_j$. To minimize the influence of outliers on the estimator, one could theoretically select the moment condition with the smallest IF value, which could however result in a poor estimation if the moment condition is not very informative of the parameter α. As suggested in Sect. 2.3, we aim to minimize the MSE of the estimates and thus downweight the individual moment conditions if their biases or variances are large. Obviously, this will also lead to lower effects of biased or imprecise moment conditions on the IF in Theorem 4. To quantify the maximum influence of generally unknown outliers on the estimate, the GES function of the GMM estimator, that is, the supremum of IF in (21) with respect to ζ can be used again.

3.4 Estimating the bias

The IF and GES derived in Sect. 3.2 characterize only the derivative of the bias caused by outlier contamination. We will refer to them in the case of contamination schemes $Z^1_{\epsilon,\zeta}$, $Z^2_{\epsilon,\zeta}$, and $Z^3_{\epsilon,\zeta}$ by IF^c_k and GES^c_k, $c = 1, 2, 3$, respectively, where k denotes the number of consecutive outliers (patch length) in schemes $Z^2_{\epsilon,\zeta}$, and $Z^3_{\epsilon,\zeta}$. Whenever the sequence of consecutive outliers is mentioned in this section, we understand by that a sequence of observations y_{it}, $t = t_1, \ldots, t_2$, that can all be considered outliers.

To approximate $\boldsymbol{b}_n = Bias\{\boldsymbol{g}_n(\alpha)\}$ introduced in Sect. 2.3, we therefore need to estimate the type and amount of outliers in a given sample. Assuming that the consecutive outliers form sequences of length k and the fraction of such outliers in data is denoted ϵ_k, the bias can be approximated using the ϵ_k-multiple of $|\mathrm{IF}^1_1|$ or GES^1_1 if $k = 1$ and of $\max\{|\mathrm{IF}^2_k|, |\mathrm{IF}^3_k|\}$ or $\max\{\mathrm{GES}^2_k, \mathrm{GES}^3_k\}$ if $k > 1$ since we cannot reliably distinguish contamination $Z^2_{\epsilon,\zeta}$ and $Z^3_{\epsilon,\zeta}$. Given that the outlier locations cannot be reliably computed either, GES is preferred for estimating the bias due to contamination.

We therefore suggest to compute the bias vector $\boldsymbol{b}_n$ in the following way, provided that the estimates $\hat{\epsilon}_k$ of the fractions of outliers forming sequences or patches of length k are available:

$$\hat{\boldsymbol{b}}_n = \left\{ \max_{k=1,\ldots,T} \left[\hat{\epsilon}_k \cdot \max_c \mathrm{GES}^c_k(\hat{r}_j; \hat{\alpha}^0_n) \right] \right\}_{j \in \mathcal{J}}, \tag{22}$$

where $\hat{\alpha}^0_n$ is an initial estimate of the parameter α and the inner maximum is taken over $c \in \{1\}$ for $k = 1$ and $c \in \{2, 3\}$ for $k > 1$. Note that if outliers (or particular types of outliers) are not present, $\hat{\epsilon}_k = 0$ and the corresponding bias term is zero.

To estimate $\hat{\epsilon}_k$, an initial estimate $\hat{\alpha}^0_n$ is needed. Once it is obtained by the DZ or AC-DZ estimator, the regression residuals $\hat{\varepsilon}_{it}$ can be constructed, for example, by $\hat{u}_{it} = y_{it} - \hat{\alpha}^0_n y_{it-1}$ and $\hat{\varepsilon}_{it} = \hat{u}_{it} - \mathrm{med}_{t=2,\ldots,T} \hat{u}_{it}$ for any $i = 1, \ldots, n$ and $t = 2, \ldots, T$; the median $\mathrm{med}_{t=2,\ldots,T} \hat{u}_{it}$ is used here as an estimate of the individual effect η_i similarly to Bramati and Croux (2007). Having estimated residuals $\hat{\varepsilon}_{it}$, the outliers are detected and the fractions ϵ_k of outliers in data forming the patches or sequences of k consecutive outliers are computed. We consider as outliers all observations with $|\hat{\varepsilon}_{it}| > \gamma \hat{\sigma}_\varepsilon$, where $\hat{\sigma}_\varepsilon$ estimates the standard deviation of ε_{it}, for example, by the median absolute deviation $\hat{\sigma}_\varepsilon = \mathrm{MAD}(\hat{\varepsilon}_{it})/\Phi^{-1}(3/4)$, and γ is a cut-off point. Although one typically uses a fixed cut-off point such as $\gamma = 2.5$, it can be chosen in a data-adaptive way by determining the fraction of residuals compatible with the normal distribution function of errors, for instance. This approach pioneered by Gervini and Yohai (2002) determines the cut-off point as the quantile of the empirical distribution function F^+_n of $|\hat{\varepsilon}_{it}|/\hat{\sigma}_\varepsilon$:

$$\hat{\gamma}_n = \min \left\{ t \colon F^+_n(t) \geq 1 - d_n \right\} \tag{23}$$

for

$$d_n = \sup_{t \geq 2.5} \max \left\{ 0, F^+_0(t) - F^+_n(t) \right\},$$

where $F_0^+(t) = \Phi(t) - \Phi(-t), t \geq 0$, denotes the distribution function of $|V|$, $V \sim N(0, 1)$.

3.5 Algorithm

The whole procedure of the bias estimation, and subsequently, the proposed GMM estimation with the robust moment selection can be summarized as follows.

1. Obtain an initial estimate $\hat{\alpha}_n^0$ by DZ or AC-DZ estimator.
2. Compute residuals $\hat{u}_{it} = y_{it} - \hat{\alpha}_n^0 y_{it-1}$ and $\hat{\varepsilon}_{it} = \hat{u}_{it} - \text{med}_{t=2,\dots,T} \hat{u}_{it}$ and estimate their standard deviation $\hat{\sigma}_\varepsilon$.
3. Using the data-adaptive cut-off point (23), determine the fractions $\hat{\epsilon}_k$ of outliers present in the data in the forms of outlier sequences of length k.
4. Approximate the bias $\boldsymbol{b}_n$ due to outliers by $\hat{\boldsymbol{b}}_n$ using (22) and estimate the variance matrix $\boldsymbol{V}_n$ in Theorem 5 by $\hat{\boldsymbol{V}}_n$ for all moment conditions (5) defined for indices $\boldsymbol{j} \in \mathcal{J}_o$.
5. For all $\boldsymbol{j} = (s, s, p) \in \mathcal{J}_o$,
    (a) set $\mathcal{J} = \{(k, k, l) : 1 \leq k \leq s \text{ is odd}, 1 \leq l \leq p \text{ is odd}\}$;
    (b) compute the GMM estimate $\hat{\alpha}_{n,\mathcal{J}}$ defined in (9) using the moment conditions selected by $\mathcal{J}$ and the weighting matrix defined as the inverse of the corresponding submatrix of $\hat{\boldsymbol{W}}_n = \hat{\boldsymbol{b}}_n \hat{\boldsymbol{b}}_n' + \hat{\boldsymbol{V}}_n$;
    (c) evaluate the criterion $RRMSC(\mathcal{J})$ defined in (10).
6. Select the set of moment conditions by

$$\hat{\mathcal{J}} = \arg \min_{\mathcal{J} \subseteq \mathcal{J}_o} RRMSC(\mathcal{J}).$$

7. The final estimate equals $\hat{\alpha}_{n,\hat{\mathcal{J}}}$.

Let us note that the algorithm in step 5 does not evaluate the GMM estimates for all subsets of indices $\mathcal{J} \subseteq \mathcal{J}_o$ and the corresponding moment conditions as that would be very time-consuming. It is therefore suggested to limit the number of $\mathcal{J}_o$ subsets and one possible proposal, which always includes the DZ condition in the estimation, is described in point 5 of the algorithm. If an extensive evaluation of many GMM estimators has to be avoided, it is possible to opt for a simple selection between the DZ, AC-DZ, and PD-DZ estimator, where PD-DZ uses all moment conditions defined by $\mathcal{J}_o$.

4 Monte Carlo simulation

In this section, we evaluate the finite sample performance of the proposed and existing estimators by Monte Carlo simulations to see whether the proposed method can weight the moment conditions so that it picks and mimics the performance of the better estimator (e.g., out of those with fixed sets of moment conditions such as DZ and AC-DZ) for each considered data generating process.

Let $\{y_{it}\}$ follow model (1). We generate $T + 100$ observations for each i and discard the first 100 observations to reduce the effect of the initial observations and to achieve stationarity. We consider cases with $\alpha = 0.1, 0.5, 0.9$, $n = 25, 50, 100, 200$, $T = 6, 12$, $\eta_i \sim N(0, \sigma_\eta^2)$, and $\varepsilon_{it} \sim N(0, 1)$. If data contamination is present, it follows the contamination schemes (11) and (12) for $\epsilon = 0.20$. More specifically, $Z_{\epsilon,\zeta}^1$ and $Z_{\epsilon,\zeta}^2$ used with $p = 3$ are both based on ζ drawn for each outlier or patch of outliers randomly from $U(10, 90)$; $U(\cdot, \cdot)$ denotes here the uniform distribution. The extreme values of outliers are chosen as they are supposed to have the largest influence on the estimates—cf. Theorem 1, for instance. Note that we have also considered mixes of two contamination schemes, for example, mixing equally independent additive outliers and patches of outliers, but the results are not reported as they are just convex combinations of the corresponding results obtained with only the first and only the second contamination schemes.

All estimators are compared by means of the mean bias and the root mean squared error (RMSE) evaluated using 1000 replications. The included estimators are chosen as follows. The non-robust estimators are represented by the Arellano–Bond (AB) two-step GMM estimator[1] (Arellano and Bond 1991), the system Blundell and Bond (BB) estimator[2] (Blundell and Bond 1998), and the X-differencing (XD) estimator (Han et al. 2014). The globally robust estimators are represented by the original DZ and AC-DZ estimators and by the proposed PD-DZ estimator. For the latter, we consider two different moment selection criteria *RRMSC*: BIC and HQIC with $\kappa_c = 2.1$ introduced in Sect. 2.4.

Considering the clean data first (see Table 1), most estimators exhibit small RMSEs except of the AB estimator that is usually strongly negatively biased if α is close to 1. The BB estimator performs well under these circumstances as expected, but is outperformed by the XD estimation. Regarding the robust estimators, the results are closer to each other for $T = 6$ than for $T = 12$ since there are only three possible moment conditions (5) if $T = 6$. The DZ estimator based on the first moment condition only is lacking behind AC-DZ and PD-DZ when α is not close to zero and additional higher-order moment conditions thus improve estimation. The results for AC-DZ and PD-DZ are rather similar in most situations, with PD-DZ becoming relatively more precise as n increases due to less noisy moment selection. Overall, adding moment conditions improves performance of AC-DZ and PD-DZ relative to DZ; the performance of PD-DZ is worse than that of the AB and BB estimators for $\alpha = 0.1$, matches them for $\alpha = 0.5$, and outperforms them for $\alpha = 0.9$.

Next, the two different data contaminations schemes are considered: independent additive outliers and the patches of additive outliers. Considering the independent additive outliers first (see Table 2), which generally bias estimates toward zero and lead thus to larger biases especially for values of α close to 1, AB, BB, and XD are

---

[1] The (optimal) inverse weight matrix, which is used here, is $\sum_i Z_i^{AB'} H Z_i^{AB}$, where Z_i^{AB} is the matrix of instruments per individual and H is a $(T-1) \times (T-1)$ tridiagonal matrix with 2 in the main diagonal, -1 in the first two sub-diagonals, and zeros elsewhere (see Arellano and Bond 1991, p. 279).

[2] The inverse weight matrix is $\sum_i Z_i^{BB'} G Z_i^{BB}$, where Z_i^{BB} is the matrix of instruments per individual and G is a partitioned matrix, $G = \text{diag}(H, I)$, where H is as in Arellano–Bond and I is the identity matrix [see Kiviet 2007, Eq. (38)].

Table 1 RMSE for all estimators in model with $\varepsilon_{it} \sim N(0, 1)$ and $\eta_i \sim N(0, 1)$ under different sample sizes

RMSE		RRMSC	$T = 6$				$T = 12$			
α	n		25	50	100	200	25	50	100	200
0.1	XD		0.120	0.083	0.060	0.042	0.068	0.048	0.034	0.023
	AB		0.160	0.117	0.082	0.057	0.098	0.065	0.045	0.030
	BB		0.143	0.105	0.074	0.054	0.101	0.069	0.048	0.032
	DZ		0.255	0.188	0.125	0.094	0.164	0.118	0.081	0.059
	AC-DZ		0.247	0.177	0.125	0.090	0.145	0.106	0.076	0.051
	PD-DZ	BIC	0.258	0.183	0.125	0.090	0.155	0.108	0.071	0.050
	PD-DZ	HQIC	0.251	0.179	0.124	0.089	0.152	0.100	0.069	0.050
0.5	XD		0.135	0.093	0.066	0.047	0.065	0.047	0.034	0.023
	AB		0.256	0.193	0.127	0.094	0.131	0.089	0.059	0.041
	BB		0.163	0.127	0.095	0.069	0.119	0.083	0.058	0.042
	DZ		0.286	0.207	0.145	0.099	0.186	0.128	0.091	0.063
	AC-DZ		0.226	0.167	0.115	0.080	0.120	0.086	0.061	0.044
	PD-DZ	BIC	0.238	0.176	0.121	0.083	0.130	0.091	0.064	0.044
	PD-DZ	HQIC	0.240	0.176	0.122	0.081	0.123	0.087	0.061	0.044
0.9	XD		0.139	0.097	0.070	0.050	0.061	0.042	0.030	0.021
	AB		0.693	0.612	0.523	0.431	0.322	0.275	0.230	0.162
	BB		0.096	0.086	0.083	0.068	0.058	0.056	0.051	0.044
	DZ		0.292	0.219	0.153	0.106	0.197	0.139	0.095	0.067
	AC-DZ		0.172	0.127	0.096	0.074	0.087	0.064	0.047	0.033
	PD-DZ	BIC	0.184	0.132	0.098	0.078	0.089	0.065	0.050	0.035
	PD-DZ	HQIC	0.195	0.136	0.101	0.074	0.090	0.067	0.050	0.033

strongly biased in all cases as expected. In the case of robust estimators, the negative biases of DZ, AC-DZ, and PD-DZ are rather small, although increasing with α. As AC-DZ outperforms DZ in this case, PD-DZ should and does exhibit performance more similar to AC-DZ than to DZ; PD-DZ even outperforms AC-DZ for $\alpha = 0.9$ or the largest sample size. This confirms the functionality of the weighting as the inclusion of higher-order differences with $s > 1$ in PD-DZ could lead to large biases due to independent additive outliers especially for $\alpha = 0.9$, see Fig. 1.

On the other hand, the higher-order differences with $s > 1$ should provide benefits when the data are contaminated by the patches of additive outliers, see Table 3. This type of contamination leads again to substantially biased non-robust estimates by XD, AB, and BB, In the case of the robust estimators, the patches of outliers tend to bias them toward 1 and have thus largest effect for α close to 0. Hence, the biases of and more generally differences among the robust estimators are smallest for $\alpha = 0.9$. For smaller values of α, DZ outperforms AC-DZ, in particular for $\alpha = 0.1$, as the patches of outliers have a larger impact on the higher-order differences of AC-DZ— see Fig. 2a. Thus, the proposed PD-DZ should and does perform similarly to DZ and actually outperforms it in most situations most cases with $\alpha \leq 0.5$, which again

Table 2 Biases and RMSE for all estimators in data with $\varepsilon_{it} \sim N(0,1)$, $\eta_i \sim N(0,1)$, and 20% contamination by independent additive outliers under different sample sizes

T		RRMSC	Bias				RMSE			
α	n		6	6	12	12	6	6	12	12
			50	200	50	200	50	200	50	200
0.1	XD		−0.107	−0.099	−0.101	−0.101	0.129	0.105	0.111	0.103
	AB		−0.096	−0.067	−0.103	−0.091	0.122	0.087	0.115	0.095
	BB		−0.088	−0.069	−0.125	−0.100	0.122	0.081	0.137	0.104
	DZ		−0.019	−0.015	−0.019	−0.018	0.290	0.143	0.181	0.092
	AC-DZ		−0.013	−0.011	−0.020	−0.017	0.269	0.134	0.178	0.088
	PD-DZ	BIC	0.002	−0.010	−0.014	−0.012	0.282	0.139	0.165	0.080
	PD-DZ	HQIC	−0.003	−0.012	−0.017	−0.017	0.286	0.140	0.163	0.079
0.5	XD		−0.502	−0.501	−0.501	−0.498	0.507	0.502	0.503	0.499
	AB		−0.479	−0.470	−0.516	−0.491	0.485	0.471	0.518	0.492
	BB		−0.482	−0.472	−0.528	−0.499	0.489	0.474	0.531	0.500
	DZ		−0.091	−0.074	−0.083	−0.087	0.307	0.166	0.206	0.131
	AD-DZ		−0.080	−0.055	−0.066	−0.063	0.259	0.139	0.159	0.103
	PD-DZ	BIC	−0.087	−0.048	−0.043	−0.058	0.260	0.148	0.174	0.106
	PD-DZ	HQIC	−0.084	−0.057	−0.057	−0.063	0.259	0.146	0.168	0.104
0.9	XD		−0.901	−0.898	−0.897	−0.898	0.904	0.899	0.898	0.898
	AB		−0.894	−0.883	−0.906	−0.895	0.898	0.884	0.907	0.896
	BB		−0.887	−0.874	−0.925	−0.902	0.891	0.875	0.927	0.902
	DZ		−0.210	−0.160	−0.175	−0.168	0.325	0.211	0.246	0.192
	AC-DZ		−0.180	−0.127	−0.115	−0.100	0.264	0.164	0.149	0.112
	PD-DZ	BIC	−0.163	−0.106	−0.088	−0.071	0.259	0.148	0.126	0.086
	PD-DZ	HQIC	−0.161	−0.104	−0.083	−0.071	0.257	0.147	0.123	0.086

Table 3 Biases and RMSE for all estimators in data with $\varepsilon_{it} \sim N(0, 1)$, $\eta_i \sim N(0, 1)$, and 20% contamination by the patches of 3 additive outliers under different sample sizes

α	n	RRMSC	Bias				RMSE			
			6 50	6 200	12 50	12 200	6 50	6 200	12 50	12 200
0.1	XD		0.751	0.753	0.593	0.592	0.753	0.753	0.594	0.592
	AB		0.565	0.571	0.552	0.566	0.569	0.574	0.543	0.565
	BB		0.577	0.579	0.459	0.461	0.585	0.581	0.462	0.462
	DZ		0.438	0.448	0.442	0.440	0.502	0.466	0.472	0.449
	AC-DZ		0.537	0.583	0.640	0.672	0.571	0.591	0.649	0.674
	PD-DZ	BIC	0.295	0.298	0.391	0.416	0.347	0.306	0.413	0.423
	PD-DZ	HQIC	0.299	0.301	0.409	0.425	0.345	0.310	0.420	0.427
0.5	XD		0.354	0.354	0.192	0.193	0.357	0.355	0.195	0.193
	AB		0.162	0.169	0.168	0.162	0.209	0.199	0.171	0.169
	BB		0.181	0.180	0.060	0.061	0.205	0.186	0.082	0.067
	DZ		0.345	0.391	0.377	0.399	0.378	0.397	0.388	0.401
	AC-DZ		0.369	0.409	0.402	0.422	0.384	0.412	0.405	0.423
	PD-DZ	BIC	0.369	0.415	0.167	0.096	0.384	0.416	0.197	0.109
	PD-DZ	HQIC	0.371	0.410	0.120	0.088	0.387	0.412	0.156	0.101
0.9	XD		−0.046	−0.045	−0.207	−0.206	0.067	0.051	0.210	0.207
	AB		−0.229	−0.211	−0.328	−0.316	0.257	0.235	0.316	0.303
	BB		−0.207	−0.203	−0.330	−0.325	0.228	0.208	0.334	0.326
	DZ		0.082	0.090	0.088	0.090	0.100	0.092	0.091	0.090
	AC-DZ		0.073	0.083	0.077	0.081	0.081	0.084	0.079	0.081
	PD-DZ	BIC	0.029	−0.098	−0.076	−0.087	0.079	0.114	0.129	0.103
	PD-DZ	HQIC	0.028	−0.091	−0.066	−0.084	0.077	0.108	0.125	0.102

confirms that the proposed weighting scheme is able to choose moment conditions that are less affected by the outliers. Note that the largest difference between DZ and PD-DZ is observed for $T = 12$ and $\alpha = 0.5$ as the higher-order differences can be used only if the number T of time periods is sufficiently large and they have a reasonable precision only if α is not close to zero.

5 Concluding remarks

In this paper, we propose an extension of the median-based robust estimator for dynamic panel data model of Dhaene and Zhu (2017) by means of multiple pairwise differences. The newly proposed GMM estimation procedure that uses weights accounting both for the variance and outlier-related bias of the moment conditions is combined with the moment selection method. As a result, the estimator performs well in non-contaminated data as well as in data containing both independent outliers and patches of outliers.

A Appendix

The outlier contamination schemes $Z^1_{\epsilon,\zeta}$, $Z^2_{\epsilon,\zeta}$, and $Z^3_{\epsilon,\zeta}$ are generally described by the contamination fraction ϵ and the magnitude of outliers ζ (recall that only the point-mass distribution G_ζ is considered here). Therefore, we will denote the non-contaminated sample observations following model (1) by y_{it} and the contaminated sample observations by $y^{\zeta,\epsilon}_{it}$. By definition of $Z^1_{\epsilon,\zeta}$, $Z^2_{\epsilon,\zeta}$, and $Z^3_{\epsilon,\zeta}$, the difference $w_{it} = y^{\zeta,\epsilon}_{it} - y_{it}$ can only equal $-\zeta$, 0, or ζ.

In order to prove the theorems concerning the influence function of $\hat{\alpha}$, it is useful to derive first the asymptotic bias of $\hat{r}_j$ as an estimator of r_j. Similarly to Sect. 3.1, it is defined as

$$\text{bias}\left(\hat{r}_j; r_j, \zeta, \epsilon\right) := \text{plim}_{n \to \infty} \hat{r}_j(r_j, \zeta, \epsilon) - r_j, \tag{24}$$

where plim denotes the probability limit operator. Let $b := b(r_j, \zeta, \epsilon)$ be a short-hand notation for (24). Then, b solves the following equation:

$$\Pr\left(\frac{\Delta^s y^{\zeta,\epsilon}_{it}}{\Delta^p y^{\zeta,\epsilon}_{it-s}} - r_j \leq b\right) = \Pr\left(\frac{\Delta^s y^{\zeta,\epsilon}_{it} - r_j \Delta^p y^{\zeta,\epsilon}_{it-s}}{\Delta^p y^{\zeta,\epsilon}_{it-s}} \leq b\right) = \frac{1}{2}. \tag{25}$$

Since r_j is considered only for $j = (s, s, p) \in \mathcal{J}_o$, where both s and p are odd, $r_j = -(1 - \alpha^s)/2$. This mapping of α to $r_j = -(1 - \alpha^s)/2$ has the same important properties for $s = 1$ and any odd $s > 1$: it maps interval $(-1, 0)$ to $(-1, -1/2)$ and interval $(0, 1)$ to $(-1/2, 0)$, it is continuous, and it is strictly increasing on $(-1, 1)$.

One can thus follow the proofs in Dhaene and Zhu (2017, Theorems 3.5 and 3.8) and apply them not only to the case of $s = p = 1$, but any odd s and p with only two adjustments: (i) the variables $\Delta^s y_{it} - r_j \Delta^p y_{it-s}$ and $\Delta^p y_{it-s}$ have to be standardized (Dhaene and Zhu 2017, equation (A.3)) and their variances generally depend on the values of s and p and (ii) in the case of patches of outliers, the probability that a patch contaminates the ratio $\Delta^s y_{it}/\Delta^p y_{it-s}$ needs to be generalized.

As for (i), note that, by Eq. (2), the variables $\Delta^s y_{it} - r_j \Delta^p y_{it-s}$ and $\Delta^p y_{it-s}$ are uncorrelated, and by Assumption A.3, they are independent and normally distributed around zero. Additionally, the stationarity Assumptions A.1 and A.2 imply that, after substituting from the model equation, $\operatorname{cov}(y_{it}, \eta) = \alpha \operatorname{cov}(y_{it-1}, \eta_i) + \operatorname{var}(\eta_i)$,

$$
\begin{aligned}
\operatorname{var}(y_{it}) &= \alpha^2 \operatorname{var}(y_{it-1}) + \operatorname{var}(\eta_i) + 2\alpha \operatorname{cov}(y_{it-1}, \eta_i) + \operatorname{var}(\varepsilon_{it}) \\
&= \alpha^2 \operatorname{var}(y_{it-1}) + \operatorname{cov}(y_{it}, \eta) + \alpha \operatorname{cov}(y_{it-1}, \eta_i) + \operatorname{var}(\varepsilon_{it}),
\end{aligned}
$$

and subsequently, $(1 - \alpha^2) \operatorname{var}(y_{it}) = (1 + \alpha) \operatorname{cov}(y_{it}, \eta) + \operatorname{var}(\varepsilon_{it})$ and $\operatorname{var}(y_{it}) - \operatorname{cov}(y_{it}, \eta)/(1-\alpha) = \operatorname{var}(\varepsilon_{it})/(1-\alpha^2)$. From this result and Aquaro and Čížek (2014, Equation (24)), we can thus conclude that

$$
\begin{pmatrix} \Delta^s y_{it} - r_j \Delta^p y_{it-s} \\ \Delta^p y_{it-s} \end{pmatrix} \sim \mathrm{N} \left[\mathbf{0}, \frac{2\sigma_\varepsilon^2}{1-\alpha^2} \begin{pmatrix} 1 - \alpha^s - r_j^2(1 - \alpha^p) & 0 \\ 0 & 1 - \alpha^p \end{pmatrix} \right] \quad (26)
$$

(the diagonal structure of the covariance matrix can be also seen from Equation 2.2 that implies $\operatorname{cov}(\Delta^s y_{it}, \Delta^p y_{it-s}) = r_j \operatorname{var}(\Delta^p y_{it-s})$).

Based on these observations, Aquaro and Čížek (2014, Theorem 1) derived the asymptotic distribution of the PD-DZ estimator defined by Eq. (7), which is presented here for the case of the triplet sets $\mathcal{J} \subseteq \mathcal{J}_o$.

Theorem 5 *Suppose that Assumptions* A.1–A.3 *hold and that* $A_n \to A > 0$ *in probability as* $n \to \infty$. *Let* $(1, 1, 1) \in \mathcal{J} \subseteq \mathcal{J}_o$ *and* $\boldsymbol{d} = \partial \boldsymbol{g}(\alpha)/\partial \alpha$, *where* α *represents the true parameter value.*

Then for a fixed T and $n \to \infty$, $\hat{\alpha}_n$ is consistent and asymptotically normal,

$$
\sqrt{n}(\hat{\alpha}_n - \alpha) \to \mathrm{N} \left(0, (\boldsymbol{d}' \boldsymbol{A} \boldsymbol{d})^{-1} \boldsymbol{d}' \boldsymbol{A} \boldsymbol{V} \boldsymbol{A} \boldsymbol{d} (\boldsymbol{d}' \boldsymbol{A} \boldsymbol{d})^{-1} \right), \quad (27)
$$

where $\boldsymbol{d} = \partial \boldsymbol{g}(\alpha)/\partial \alpha = \{-s\alpha^{s-1}\}_{j=(s,s,p)\in\mathcal{J}}$ *and* $\boldsymbol{V}$ *is has a typical element with indices* $\boldsymbol{j} = (s, s, p) \in \mathcal{J}$, $\boldsymbol{j}' = (s', s', p') \in \mathcal{J}$ *defined by*

$$
\frac{\pi^2 \sqrt{\left\{1 - \alpha^s - \frac{1}{4}(1 - \alpha^s)^2(1 - \alpha^p)^2\right\} \left\{1 - \alpha^{s'} - \frac{1}{4}(1 - \alpha^{s'})^2(1 - \alpha^{p'})^2\right\}}}{\sqrt{[T - s - p][T - s' - p']}}
$$

$$
\times \mathrm{E}\left[\left(\sum_{t=s+p+1}^{T} \operatorname{sgn}\left(\Delta^s y_{it} - r_j \Delta^p y_{it-s}\right) \operatorname{sgn}(\Delta^p y_{it-s}) \right) \right.
$$
$$
\left. \left(\sum_{t=s'+p'+1}^{T} \operatorname{sgn}(\Delta^{s'} y_{it} - r_{j'} \Delta^{p'} y_{it-s'}) \operatorname{sgn}(\Delta^{p'} y_{it-s'}) \right) \right].
$$

A.1 Independent additive outlier contamination $Z^1_{\epsilon,\zeta}$

Under independent additive outlier contamination $Z^1_{\epsilon,\zeta}$, Eq. (25) can be written as

$$\Pr\left(\frac{\Delta^s y_{it}^{\zeta,\epsilon} - r_j \Delta^p y_{it-s}^{\zeta,\epsilon}}{\Delta^p y_{it-s}^{\zeta,\epsilon}} \le b\right) = \Pr\left(\frac{u_{itj} + \Delta^s w_{it} - r_j \Delta^p w_{it-s}}{\Delta^p y_{it-s} + \Delta^p w_{it-s}} \le b\right)$$

$$= \Pr\left[f(\boldsymbol{w}_{it}) \le b\right] = \frac{1}{2}, \qquad (28)$$

where residual $u_{itj} = \Delta^s y_{it} - r_j \Delta^p y_{it-s}$, $w_{it} \in \{0, \zeta\}$, $\boldsymbol{w}_{it} = (w_{it}, w_{it-s}, w_{it-s-p})'$ is a random vector, and $f(\boldsymbol{w}_{it})$ is a random scalar. Let $\Omega_{\boldsymbol{w}_{it}}$ be the set of the eight possible outcomes of $\boldsymbol{w}_{it}$, that is,

$$\Omega_{\boldsymbol{w}_{it}} := \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \zeta \end{pmatrix}, \cdots, \begin{pmatrix} \zeta \\ \zeta \\ \zeta \end{pmatrix} \right\}, \qquad (29)$$

where the number of elements is $\#\Omega_{\boldsymbol{w}_{it}} = 8$. To simplify the notation, let us refer to (29) as Ω_{it}, and denote each of its element as $\boldsymbol{\omega}_{itj}$, $j = 1, \ldots, 8$. Then it holds

$$\Pr\left[f(\boldsymbol{w}_{it}) \le b\right] = \Pr\left[(f(\boldsymbol{w}_{it}) \le b) \cap \left(\bigcup_{j=1}^{8} \boldsymbol{w}_{it} = \boldsymbol{\omega}_{itj}\right)\right]$$

$$= \sum_{j=1}^{8} \Pr\left[(f(\boldsymbol{w}_{it}) \le b) \cap (\boldsymbol{w}_{it} = \boldsymbol{\omega}_{itj})\right]$$

$$= \sum_{j=1}^{8} \Pr\left[f(\boldsymbol{w}_{it}) \le b \,\middle|\, \boldsymbol{w}_{it} = \boldsymbol{\omega}_{itj}\right] \Pr\left(\boldsymbol{w}_{it} = \boldsymbol{\omega}_{itj}\right)$$

$$= \sum_{j=1}^{8} \Pr\left[f(\boldsymbol{\omega}_{itj}) \le b\right] \Pr\left(\boldsymbol{w}_{it} = \boldsymbol{\omega}_{itj}\right). \qquad (30)$$

Note that $\Pr(\boldsymbol{w}_{it} = \boldsymbol{\omega}_{itj}) = \Pr(\boldsymbol{w}_{it} = \boldsymbol{\omega}_{itj'})$ for some j and j' because the data contamination $Z^1_{\epsilon,\zeta}$ is characterized by outliers occurring independently from each other. For instance, $\Pr[(\zeta, 0, 0)'] = \Pr[(0, \zeta, 0)'] = \Pr[(0, 0, \zeta)'] = (1-\epsilon)^2\epsilon$. Moreover, $f[(0, 0, 0)'] = f[(\zeta, \zeta, \zeta)']$. Therefore, Eq. (28) can be decomposed as

$$\Pr\left[f(\boldsymbol{w}_{it}) \le b\right] = \left[(1-\epsilon)^3 + \epsilon^3\right] A + (1-\epsilon)^2\epsilon B + (1-\epsilon)\epsilon^2 C = \frac{1}{2}, \quad (31)$$

where A, B, and C are defined for r_j, ζ, and b as follows:

$$
\begin{aligned}
A(r_j, b) &:= \Pr\left(\frac{u_{itj}}{\Delta^p y_{it-s}} \le b\right), \\
B(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + \zeta}{\Delta^p y_{it-s}} \le b\right) + \Pr\left(\frac{u_{itj} - \zeta(1 + r_j)}{\Delta^p y_{it-s} + \zeta} \le b\right) \\
&\quad + \Pr\left(\frac{u_{itj} + \zeta r_j}{\Delta^p y_{it-s} - \zeta} \le b\right), \\
C(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} - \zeta r_j}{\Delta^p y_{it-s} + \zeta} \le b\right) + \Pr\left(\frac{u_{itj} + \zeta(1 + r_j)}{\Delta^p y_{it-s} - \zeta} \le b\right) \\
&\quad + \Pr\left(\frac{u_{itj} - \zeta}{\Delta^p y_{it-s}} \le b\right).
\end{aligned}
\tag{32}
$$

These probabilities are all of the form

$$
L(k, l, b) = \Pr\left(\frac{u_{itj} + k}{\Delta^p y_{it-s} - l} \le b\right)
\tag{33}
$$

for given k, l, and b, and they can be conveniently standardized by using (26) as follows:

$$
L(k, l, b) = \Pr\left(\frac{X + k'}{Y - l'} \le b'\right),
\tag{34}
$$

where X and Y are independent N(0, 1) variables and

$$
k' := \frac{k}{\sigma_u}, \quad l' := \frac{l}{\sigma_{\Delta^p}}, \quad b' := \sigma^* b,
\tag{35}
$$

and

$$
\sigma^* := \frac{\sigma_{\Delta^p}}{\sigma_u} = \sqrt{\frac{1 - \alpha^p}{1 - \alpha^s - (1 - \alpha^s)^2 (1 - \alpha^p)/4}},
\tag{36}
$$

where $\sigma_u := \sqrt{\operatorname{var}(u_{itj})}$ and $\sigma_{\Delta^p} := \sqrt{\operatorname{var}(\Delta^p y_{it-s})}$ can be found in (26). Finally, note that $L(k, l, b) = L(-k, -l, b)$, hence $B = C$ and (28) becomes

$$
A + \epsilon(1 - \epsilon)(B - 3A) = \frac{1}{2}.
\tag{37}
$$

Proof of Theorem 1 As in Dhaene and Zhu (2017, proof of Theorem 3.2), it follows from the definition of influence function that

$$
\operatorname{IF}(\hat{r}_j; r_j, \zeta) := \left.\frac{\partial \operatorname{bias}(\hat{r}_j; r_j, \zeta)}{\partial \epsilon}\right|_{\epsilon=0} = \frac{3A(r_j, 0) - B(r_j, \zeta, 0)}{A'_b(r_j, 0)},
\tag{38}
$$

where the equality follows from the implicit function theorem applied to (37) and where

$$A_b'(r_j, 0) := \left. \frac{\partial A(r_j, b)}{\partial b} \right|_{\epsilon = 0}. \tag{39}$$

As in Dhaene and Zhu (2017, Equation (A.4)),

$$A(r_j, b) = \Pr\left(\frac{X}{Y} \leq \sigma^* b\right) = \frac{1}{2} + \frac{1}{\pi} \arctan \sigma^* b, \tag{40}$$

where σ^* is defined in (36) and $X, Y \sim \mathrm{N}(0, 1)$. Hence, $A(r_j, 0) = 1/2$ and

$$A_b'(r_j, 0) = \frac{1}{\pi \sigma^*} = \frac{1}{\pi} \sqrt{\frac{1 - \alpha^p}{1 - \alpha^s - r_j^2(1 - \alpha^p)}} \tag{41}$$

(recall that $r_j = (1 - \alpha^s)/2$). Next, Dhaene and Zhu (2017, Lemma A.1) implies that, for $X, Z \sim \mathrm{N}(0, 1)$ and constants c, c', c'', $P\{(X + c)/Z \leq 0\} = 1/2$ and $P\{(X + c')/(Z - c) \leq 0\} + P\{(X + c'')/(Z - c) \leq 0\} = 1 + [\Phi(c') - \Phi(-c'')][\Phi(c) - \Phi(-c)]$. Hence, the definition of $B(r_j, \zeta, b)$ and the standardization (34) imply

$$B(r_j, \zeta, 0) = \frac{3}{2} + \left[\Phi\left(\frac{\zeta(1 + r_j)}{\sigma_u}\right) - \Phi\left(-\frac{\zeta r_j}{\sigma_u}\right)\right] \times \left[\Phi\left(\frac{\zeta}{\sigma_{\Delta^p}}\right) - \Phi\left(-\frac{\zeta}{\sigma_{\Delta^p}}\right)\right]. \tag{42}$$

Substituting for $\sigma_u := \sqrt{\mathrm{var}(u_{itj})}$ and $\sigma_{\Delta^p} := \sqrt{\mathrm{var}(\Delta^p y_{it-s})}$ from (26) and $r_j = -(1 - \alpha^s)/2$ into (42) and for terms $A(r_j, 0)$, $B(r_j, \zeta, 0)$, and $A_b'(r_j, 0)$ in (38) completes the proof. □

A.2 Patch additive outlier contamination $Z_{\epsilon, \zeta}^2$

As in "Appendix A.1", it is useful to derive first the asymptotic bias of $\hat{r}_j$ under the outlier contamination $Z_{\epsilon, \zeta}^2$ as defined in (12). This is given by $b := b(r_j, \zeta, \epsilon, k)$ solving the equation

$$\Pr\left(\frac{\Delta^s y_{it}^{\zeta, \epsilon} - r_j \Delta^p y_{it-s}^{\zeta, \epsilon}}{\Delta^p y_{it-s}^{\zeta, \epsilon}} \leq b\right) = \Pr\left(\frac{u_{itj} + \Delta^s w_{it} - r_j \Delta^p w_{it-s}}{\Delta^p y_{it-s} + \Delta^p w_{it-s}} \leq b\right)$$

$$= \mathfrak{p}_A A + \mathfrak{p}_B B + \mathfrak{p}_C C + \mathfrak{p}_D D = \frac{1}{2}, \tag{43}$$

where the notation is defined below. Note that the decomposition in the second equality follows along the same lines as in "Appendix A.1", in particular Eq. (30). In this case, the only difference is that outliers no longer occur independently but in patches. The number of elements of Ω_{it} increases to $\#\Omega_{it} = 13$ as now, if we observe multiple outliers, we shall distinguish the event of the outliers belonging to the same patch from the event of these outliers belonging to different patches. For instance, $(0, \zeta, \zeta)'$

may be that result of one patch only, $(0, \zeta_1, \zeta_1)'$, or of two patches, $(0, \zeta_2, \zeta_1)'$, where the subscript of ζ indicates the patch. Recalling that $(1 - \tilde{\epsilon})^k = 1 - \epsilon$,

$$
\mathfrak{p}_B := \Pr\left[\begin{pmatrix} \zeta \\ 0 \\ 0 \end{pmatrix} \cup \begin{pmatrix} 0 \\ \zeta \\ \zeta \end{pmatrix} \right] = \Pr\begin{pmatrix} \zeta_1 \\ 0 \\ 0 \end{pmatrix} + \Pr\begin{pmatrix} 0 \\ \zeta_1 \\ \zeta_1 \end{pmatrix} + \Pr\begin{pmatrix} 0 \\ \zeta_2 \\ \zeta_1 \end{pmatrix}
$$

$$
= (1 - \tilde{\epsilon})^{k+\min\{p,k\}} \cdot \left(1 - (1 - \tilde{\epsilon})^{\min\{s,k\}}\right)
$$
$$
+ \left(1 - (1 - \tilde{\epsilon})^{\max\{0, s+k-\max\{s+p,k\}\}}\right) \cdot (1 - \tilde{\epsilon})^k
$$
$$
+ \left(1 - (1 - \tilde{\epsilon})^{(p+k-\max\{p,k\})}\right) \cdot \left(1 - (1 - \tilde{\epsilon})^{\max\{0, s+\min\{p,k\}-\max\{s,k\}\}}\right)
$$
$$
\cdot (1 - \tilde{\epsilon})^k , \tag{44}
$$

$$
\mathfrak{p}_C := \Pr\left[\begin{pmatrix} 0 \\ 0 \\ \zeta \end{pmatrix} \cup \begin{pmatrix} \zeta \\ \zeta \\ 0 \end{pmatrix} \right] = \Pr\begin{pmatrix} 0 \\ 0 \\ \zeta_1 \end{pmatrix} + \Pr\begin{pmatrix} \zeta_1 \\ \zeta_1 \\ 0 \end{pmatrix} + \Pr\begin{pmatrix} \zeta_2 \\ \zeta_1 \\ 0 \end{pmatrix}
$$

$$
= \left(1 - (1 - \tilde{\epsilon})^{(p+k-\max\{p,k\})}\right) \cdot (1 - \tilde{\epsilon})^{k+\min\{s,k\}}
$$
$$
+ (1 - \tilde{\epsilon})^k \cdot \left(1 - (1 - \tilde{\epsilon})^{\max\{0, \min\{s+p,k\}-s\}}\right)
$$
$$
+ (1 - \tilde{\epsilon})^k \cdot \left(1 - (1 - \tilde{\epsilon})^{\max\{0, s+\min\{p,k\}-\max\{s,k\}\}}\right) \cdot \left(1 - (1 - \tilde{\epsilon})^{\min\{s,k\}}\right), \tag{45}
$$

$$
\mathfrak{p}_D := \Pr\left[\begin{pmatrix} \zeta \\ 0 \\ \zeta \end{pmatrix} \cup \begin{pmatrix} 0 \\ \zeta \\ 0 \end{pmatrix} \right] = \Pr\begin{pmatrix} \zeta_2 \\ 0 \\ \zeta_1 \end{pmatrix} + \Pr\begin{pmatrix} 0 \\ \zeta_1 \\ 0 \end{pmatrix}
$$

$$
= \left(1 - (1 - \tilde{\epsilon})^{(p+k-\max\{p,k\})}\right) \cdot (1 - \tilde{\epsilon})^k \cdot \left(1 - (1 - \tilde{\epsilon})^{\min\{s,k\}}\right) \tag{46}
$$
$$
+ (1 - \tilde{\epsilon})^{2k} \cdot \left(1 - (1 - \tilde{\epsilon})^{\max\{0, s+\min\{p,k\}-\max\{s,k\}\}}\right),
$$

and $\mathfrak{p}_A = 1 - \mathfrak{p}_B - \mathfrak{p}_C - \mathfrak{p}_D$. Next, the terms A, B, C, D are defined for r_j, ζ, and b as follows:

$$
A(r_j, b) := \Pr\left(\frac{u_{itj}}{\Delta^p y_{it-s}} \le b \right),
$$
$$
B(r_j, \zeta, b) := \Pr\left(\frac{u_{itj} + \zeta}{\Delta^p y_{it-s}} \le b \right),
$$
$$
C(r_j, \zeta, b) := \Pr\left(\frac{u_{itj} + \zeta r_j}{\Delta^p y_{it-s} - \zeta} \le b \right), \tag{47}
$$
$$
D(r_j, \zeta, b) := \Pr\left(\frac{u_{itj} + \zeta(1 + r_j)}{\Delta^p y_{it-s} - \zeta} \le b \right),
$$

where the symmetry $L(k, l, b) = L(-k, -l, b)$ has been used, recall Eq. (33).

Proof of Theorem 2 By the definition of influence function in (15),

$$\text{IF}(\hat{r}_j; r_j, \zeta) = \left. \frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \epsilon} \right|_{\epsilon=0}, \tag{48}$$

where b denotes the bias of $\hat{r}_j$. Given that $(1 - \tilde{\epsilon})^k = 1 - \epsilon$, it holds

$$\frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \epsilon} = \frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \tilde{\epsilon}} \frac{\partial \tilde{\epsilon}}{\partial \epsilon} = \frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \tilde{\epsilon}} \frac{1}{k(1 - \tilde{\epsilon})^{k-1}}. \tag{49}$$

The derivative in (49) can obtained by applying the implicit function theorem to (43),

$$\left. \frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \tilde{\epsilon}} \right|_{\epsilon=0} = -\frac{\sum_{j\in\{B,C,D\}} \mathfrak{p}'_j(0) j(r_j, \zeta, 0) + \mathfrak{p}'_A(0) A(r_j, 0)}{A'_b(r_j, 0)}, \tag{50}$$

where $A'_b(r_j, 0)$ is the same as in (39) and where $\mathfrak{p}'_j$, $j \in \{B, C, D\}$, denote the derivatives of $\mathfrak{p}_j$ in Eqs. (44)–(46) with respect to $\tilde{\epsilon}$, that is,

$$\mathfrak{p}'_B(0) := \left. \frac{\partial \mathfrak{p}_B(\tilde{\epsilon}; s, p, k)}{\partial \tilde{\epsilon}} \right|_{\epsilon=0}$$
$$= \min\{s, k\} + \max\left\{0, s + k - \max\{s + p, k\}\right\}, \tag{51}$$

$$\mathfrak{p}'_C(0) = (p + k - \max\{p, k\}) + \max\left\{0, \min\{s + p, k\} - s\right\}, \tag{52}$$

$$\mathfrak{p}'_D(0) = \max\left\{0, s + \min\{p, k\} - \max\{s, k\}\right\}, \tag{53}$$

and

$$\mathfrak{p}'_A(0) = -\left[\mathfrak{p}'_B(0) + \mathfrak{p}'_C(0) + \mathfrak{p}'_D(0)\right]. \tag{54}$$

As in "Appendix A.1", $A(r_j; 0) = 1/2$. Further, it follows from Dhaene and Zhu (2017, Lemma A.1) that, for $X, Z \sim N(0, 1)$ and constants c, c', $P\{(X+c')/(Z-c) \le 0\} = \Phi(-c')\Phi(-c) + \Phi(c')\Phi(c)$. Hence, the definition (47) and the standardization (34)–(36) imply

$$B(r_j; \zeta, 0) = \frac{1}{2}, \tag{55}$$

$$C(r_j; \zeta, 0) = \Phi\left(-\frac{r_j\zeta}{\sigma_u}\right) \Phi\left(-\frac{\zeta}{\sigma_{\Delta^p}}\right) + \Phi\left(\frac{r_j\zeta}{\sigma_u}\right) \Phi\left(\frac{\zeta}{\sigma_{\Delta^p}}\right), \tag{56}$$

$$D(r_j; \zeta, 0) = \Phi\left(-\frac{(1 + r_j)\zeta}{\sigma_u}\right) \Phi\left(-\frac{\zeta}{\sigma_{\Delta^p}}\right) + \Phi\left(\frac{(1 + r_j)\zeta}{\sigma_u}\right) \Phi\left(\frac{\zeta}{\sigma_{\Delta^p}}\right), \tag{57}$$

where $\sigma_u := \sqrt{\text{var}(u_{itj})}$ and $\sigma_{\Delta^p} := \sqrt{\text{var}(\Delta^p y_{it-s})}$ are given in (26) and $r_j = -(1 - \alpha^s)/2$. Substituting (49)–(57) in (48) completes the proof. ☐

Table 4 Configurations of patch outliers and their probabilities

$(\lvert\zeta\rvert_1,\ldots,0,\ldots,0)'$	$(1-\tilde{\epsilon})^{k+\min\{p,k\}}\cdot(1-(1-\tilde{\epsilon})^{\min\{s,k\}})$
$(0,\ldots,0,\ldots\lvert\zeta\rvert_1)'$	$(1-(1-\tilde{\epsilon})^{(p+k-\max\{p,k\})})\cdot(1-\tilde{\epsilon})^{k+\min\{s,k\}}$
$(0,\ldots,\lvert\zeta\rvert_1,\ldots,0)'$	$(1-\tilde{\epsilon})^{2k}\cdot(1-(1-\tilde{\epsilon})^{\max\{0,s+\min\{p,k\}-\max\{s,k\}\}})$
$(\lvert\zeta\rvert_1,\ldots,\lvert\zeta\rvert_1,\ldots,0)'$	$(1-\tilde{\epsilon})^{k}\cdot(1-(1-\tilde{\epsilon})^{\max\{0,\min\{s+p,k\}-s\}})$
$(\lvert\zeta\rvert_2,\ldots,\lvert\zeta\rvert_1,\ldots,0)'$	$(1-\tilde{\epsilon})^{k}\cdot(1-(1-\tilde{\epsilon})^{\max\{0,s+\min\{p,k\}-\max\{s,k\}\}})\cdot(1-(1-\tilde{\epsilon})^{\min\{s,k\}})$
$(0,\ldots,\lvert\zeta\rvert_1,\ldots,\lvert\zeta\rvert_1)'$	$(1-(1-\tilde{\epsilon})^{\max\{0,s+k-\max\{s+p,k\}\}})\cdot(1-\tilde{\epsilon})^{k}$
$(0,\ldots,\lvert\zeta\rvert_2,\ldots,\lvert\zeta\rvert_1)'$	$(1-(1-\tilde{\epsilon})^{(p+k-\max\{p,k\})})$ $\cdot(1-(1-\tilde{\epsilon})^{\max\{0,s+\min\{p,k\}-\max\{s,k\}\}})\cdot(1-\tilde{\epsilon})^{k}$
$(\lvert\zeta\rvert_2,\ldots,0,\ldots,\lvert\zeta\rvert_1)'$	$(1-(1-\tilde{\epsilon})^{(p+k-\max\{p,k\})})\cdot(1-\tilde{\epsilon})^{k}\cdot(1-(1-\tilde{\epsilon})^{\min\{s,k\}})$
$(\lvert\zeta\rvert_1,\ldots,\lvert\zeta\rvert_1,\ldots,\lvert\zeta\rvert_1)'$	$(1-(1-\tilde{\epsilon})^{\max\{0,k-s-p\}})$
$(\lvert\zeta\rvert_2,\ldots,\lvert\zeta\rvert_1,\ldots,\lvert\zeta\rvert_1)'$	$(1-(1-\tilde{\epsilon})^{\max\{0,k-p\}})\cdot(1-(1-\tilde{\epsilon})^{\min\{s,k\}})$
$(\lvert\zeta\rvert_2,\ldots,\lvert\zeta\rvert_2,\ldots,\lvert\zeta\rvert_1)'$	$(1-(1-\tilde{\epsilon})^{k})\cdot(1-(1-\tilde{\epsilon})^{\max\{0,\min\{s+p,k\}-s\}})$
$(\lvert\zeta\rvert_3,\ldots,\lvert\zeta\rvert_2,\ldots,\lvert\zeta\rvert_1)'$	$(1-(1-\tilde{\epsilon})^{k})\cdot(1-(1-\tilde{\epsilon})^{\max\{0,s+\min\{p,k\}-\max\{s,k\}\}})$ $\cdot(1-(1-\tilde{\epsilon})^{\min\{s,k\}})$

A.3 Patch additive outlier contamination $Z^3_{\epsilon,\zeta}$

This case is a generalization of the $Z^2_{\epsilon,\zeta}$ contamination. The proof structure is very similar to the one in "Appendices A.1 and A.2", although the algebra is a bit more lengthy. As before, it is useful to derive first the bias of $\hat{r}_j$ under the outlier contamination $Z^3_{\epsilon,\zeta}$ as defined in (13). This is given by $b := b(r_j, \zeta, \epsilon, k)$ solving the equation

$$
\Pr\left(\frac{\Delta^s y_{it}^{\zeta,\epsilon} - r_j \Delta^p y_{it-s}^{\zeta,\epsilon}}{\Delta^p y_{it-s}^{\zeta,\epsilon}} \le b\right) = \Pr\left(\frac{u_{itj} + \Delta^s w_{it} - r_j \Delta^p w_{it-s}}{\Delta^p y_{it-s} + \Delta^p w_{it-s}} \le b\right)
$$
$$
= \mathfrak{p}_A A + \mathfrak{p}_B B + \mathfrak{p}_C C + \mathfrak{p}_D D + \mathfrak{p}_E E + \mathfrak{p}_F F
$$
$$
+ \mathfrak{p}_G G + \mathfrak{p}_H H + \mathfrak{p}_I I + \mathfrak{p}_J J = \frac{1}{2}, \qquad (58)
$$

where the notation is explained below. Note that the set Ω_{it} in (29) is different than it was for previous types of contaminations as now outliers can be either negative or positive multiple of ζ. Also recall that $(1-\tilde{\epsilon})^k = 1 - \epsilon$.

By using the results in Table 4, we have that

$$
\mathfrak{p}_B := \Pr\left[\begin{pmatrix}\zeta\\0\\0\end{pmatrix}\cup\begin{pmatrix}-\zeta\\0\\0\end{pmatrix}\cup\begin{pmatrix}0\\\zeta\\\zeta\end{pmatrix}\cup\begin{pmatrix}0\\-\zeta\\-\zeta\end{pmatrix}\right]
$$
$$
= \frac{1}{2}\Pr\begin{pmatrix}\zeta_1\\0\\0\end{pmatrix} + \frac{1}{2}\Pr\begin{pmatrix}-\zeta_1\\0\\0\end{pmatrix} + \frac{1}{4}\Pr\begin{pmatrix}0\\\zeta_2\\\zeta_1\end{pmatrix} + \frac{1}{4}\Pr\begin{pmatrix}0\\-\zeta_2\\-\zeta_1\end{pmatrix}
$$

$$= \Pr \begin{pmatrix} |\zeta|_1 \\ 0 \\ 0 \end{pmatrix} + \frac{1}{2} \Pr \begin{pmatrix} 0 \\ |\zeta|_2 \\ |\zeta|_1 \end{pmatrix}, \tag{59}$$

$$\mathfrak{p}_C := \Pr \left[\begin{pmatrix} 0 \\ 0 \\ \zeta \end{pmatrix} \cup \begin{pmatrix} 0 \\ 0 \\ -\zeta \end{pmatrix} \cup \begin{pmatrix} \zeta \\ \zeta \\ 0 \end{pmatrix} \cup \begin{pmatrix} -\zeta \\ -\zeta \\ 0 \end{pmatrix} \right]$$

$$= \Pr \begin{pmatrix} 0 \\ 0 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} 0 \\ 0 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_2 \\ \zeta_1 \\ 0 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_2 \\ -\zeta_1 \\ 0 \end{pmatrix}$$

$$= \Pr \begin{pmatrix} 0 \\ 0 \\ |\zeta|_1 \end{pmatrix} + \frac{1}{2} \Pr \begin{pmatrix} |\zeta|_2 \\ |\zeta|_1 \\ 0 \end{pmatrix}, \tag{60}$$

$$\mathfrak{p}_D := \Pr \left[\begin{pmatrix} \zeta \\ 0 \\ \zeta \end{pmatrix} \cup \begin{pmatrix} -\zeta \\ 0 \\ -\zeta \end{pmatrix} \cup \begin{pmatrix} 0 \\ \zeta \\ 0 \end{pmatrix} \cup \begin{pmatrix} 0 \\ -\zeta \\ 0 \end{pmatrix} \right]$$

$$= \Pr \begin{pmatrix} \zeta_2 \\ 0 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_2 \\ 0 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} 0 \\ \zeta_1 \\ 0 \end{pmatrix} + \Pr \begin{pmatrix} 0 \\ -\zeta_1 \\ 0 \end{pmatrix}$$

$$= \frac{1}{2} \Pr \begin{pmatrix} |\zeta|_2 \\ 0 \\ |\zeta|_1 \end{pmatrix} + \Pr \begin{pmatrix} 0 \\ |\zeta|_1 \\ 0 \end{pmatrix}, \tag{61}$$

$$\mathfrak{p}_E := \Pr \left[\begin{pmatrix} 0 \\ -\zeta \\ \zeta \end{pmatrix} \cup \begin{pmatrix} 0 \\ \zeta \\ -\zeta \end{pmatrix} \right]$$

$$= \Pr \begin{pmatrix} 0 \\ -\zeta_1 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} 0 \\ -\zeta_2 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} 0 \\ \zeta_1 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} 0 \\ \zeta_2 \\ -\zeta_1 \end{pmatrix}$$

$$= \Pr \begin{pmatrix} 0 \\ |\zeta|_1 \\ |\zeta|_1 \end{pmatrix} + \frac{1}{2} \Pr \begin{pmatrix} 0 \\ |\zeta|_2 \\ |\zeta|_1 \end{pmatrix}, \tag{62}$$

$$\mathfrak{p}_F := \Pr \left[\begin{pmatrix} -\zeta \\ 0 \\ \zeta \end{pmatrix} \cup \begin{pmatrix} \zeta \\ 0 \\ -\zeta \end{pmatrix} \right] = \Pr \begin{pmatrix} -\zeta_2 \\ 0 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_2 \\ 0 \\ -\zeta_1 \end{pmatrix}$$

$$= \frac{1}{2} \Pr \begin{pmatrix} |\zeta|_2 \\ 0 \\ |\zeta|_1 \end{pmatrix}, \tag{63}$$

$$\mathfrak{p}_G := \Pr \left[\begin{pmatrix} \zeta \\ -\zeta \\ 0 \end{pmatrix} \cup \begin{pmatrix} -\zeta \\ \zeta \\ 0 \end{pmatrix} \right]$$

$$= \Pr \begin{pmatrix} \zeta_1 \\ -\zeta_1 \\ 0 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_2 \\ -\zeta_1 \\ 0 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_1 \\ \zeta_1 \\ 0 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_2 \\ \zeta_1 \\ 0 \end{pmatrix}$$

$$= \Pr \begin{pmatrix} |\zeta|_1 \\ |\zeta|_1 \\ 0 \end{pmatrix} + \frac{1}{2} \Pr \begin{pmatrix} |\zeta|_2 \\ |\zeta|_1 \\ 0 \end{pmatrix}, \tag{64}$$

$$\mathfrak{p}_H := \Pr \left[\begin{pmatrix} -\zeta \\ -\zeta \\ \zeta \end{pmatrix} \cup \begin{pmatrix} \zeta \\ \zeta \\ -\zeta \end{pmatrix} \right]$$

$$= \Pr \begin{pmatrix} -\zeta_2 \\ -\zeta_1 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_3 \\ -\zeta_2 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_2 \\ \zeta_1 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_3 \\ \zeta_2 \\ -\zeta_1 \end{pmatrix}$$

$$= \frac{1}{2} \Pr \begin{pmatrix} |\zeta|_2 \\ |\zeta|_1 \\ |\zeta|_1 \end{pmatrix} + \frac{1}{4} \Pr \begin{pmatrix} |\zeta|_3 \\ |\zeta|_2 \\ |\zeta|_1 \end{pmatrix}, \tag{65}$$

$$\mathfrak{p}_I := \Pr \left[\begin{pmatrix} \zeta \\ -\zeta \\ \zeta \end{pmatrix} \cup \begin{pmatrix} -\zeta \\ \zeta \\ -\zeta \end{pmatrix} \right]$$

$$= \Pr \begin{pmatrix} \zeta_1 \\ -\zeta_1 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_2 \\ -\zeta_1 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_2 \\ -\zeta_2 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_3 \\ -\zeta_2 \\ \zeta_1 \end{pmatrix}$$

$$+ \Pr \begin{pmatrix} -\zeta_1 \\ \zeta_1 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_2 \\ \zeta_1 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_2 \\ \zeta_2 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_3 \\ \zeta_2 \\ -\zeta_1 \end{pmatrix}$$

$$= \Pr \begin{pmatrix} |\zeta|_1 \\ |\zeta|_1 \\ |\zeta|_1 \end{pmatrix} + \frac{1}{2} \Pr \begin{pmatrix} |\zeta|_2 \\ |\zeta|_1 \\ |\zeta|_1 \end{pmatrix} + \frac{1}{2} \Pr \begin{pmatrix} |\zeta|_2 \\ |\zeta|_2 \\ |\zeta|_1 \end{pmatrix} + \frac{1}{4} \Pr \begin{pmatrix} |\zeta|_3 \\ |\zeta|_2 \\ |\zeta|_1 \end{pmatrix}, \tag{66}$$

$$\mathfrak{p}_J := \Pr \left[\begin{pmatrix} \zeta \\ -\zeta \\ -\zeta \end{pmatrix} \cup \begin{pmatrix} -\zeta \\ \zeta \\ \zeta \end{pmatrix} \right]$$

$$= \Pr \begin{pmatrix} \zeta_2 \\ -\zeta_2 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} \zeta_3 \\ -\zeta_2 \\ -\zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_2 \\ \zeta_2 \\ \zeta_1 \end{pmatrix} + \Pr \begin{pmatrix} -\zeta_3 \\ \zeta_2 \\ \zeta_1 \end{pmatrix}$$

$$= \frac{1}{2} \Pr \begin{pmatrix} |\zeta|_2 \\ |\zeta|_2 \\ |\zeta|_1 \end{pmatrix} + \frac{1}{4} \Pr \begin{pmatrix} |\zeta|_3 \\ |\zeta|_2 \\ |\zeta|_1 \end{pmatrix}, \tag{67}$$

and

$$\mathfrak{p}_A = 1 - \sum_{j \in \mathcal{I} \setminus \{A\}} \mathfrak{p}_j, \tag{68}$$

where $\mathcal{I} := \{A, B, C, D, E, F, G, H, I, J\}$. Moreover,

$$
\begin{aligned}
A(r_j, b) &:= \Pr\left(\frac{u_{itj}}{\Delta^p y_{it-s}} \le b\right), \\
B(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + \zeta}{\Delta^p y_{it-s}} \le b\right), \\
C(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + \zeta r_j}{\Delta^p y_{it-s} - \zeta} \le b\right), \\
D(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + \zeta(1 + r_j)}{\Delta^p y_{it-s} - \zeta} \le b\right), \\
E(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + \zeta(1 + 2r_j)}{\Delta^p y_{it-s} - 2\zeta} \le b\right), \\
F(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + \zeta(r_j - 1)}{\Delta^p y_{it-s} - \zeta} \le b\right), \\
G(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + \zeta(2 + r_j)}{\Delta^p y_{it-s} - \zeta} \le b\right), \\
H(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + 2\zeta r_j}{\Delta^p y_{it-s} - 2\zeta} \le b\right), \\
I(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + 2\zeta(1 + r_j)}{\Delta^p y_{it-s} - 2\zeta} \le b\right), \\
J(r_j, \zeta, b) &:= \Pr\left(\frac{u_{itj} + 2\zeta}{\Delta^p y_{it-s}} \le b\right),
\end{aligned}
\tag{69}
$$

where the symmetry $L(k, l, b) = L(-k, -l, b)$ has been used, recall Eq. (33).

Proof of Theorem 3 Denote

$$
\mathfrak{p}'_j(0) := \left.\frac{\partial \mathfrak{p}_j(\tilde{\epsilon}; s, p, k)}{\partial \tilde{\epsilon}}\right|_{\tilde{\epsilon}=0}, \quad j \in \mathcal{I} := \{A, B, C, D, E, F, G, H, I, J\},
$$

where $\mathfrak{p}_j(\cdot)$, $j \in \mathcal{I}$, are defined in (59)–(68). Given that $(1 - \tilde{\epsilon})^k = 1 - \epsilon$, it holds

$$
\begin{aligned}
\mathrm{IF}(\hat{r}_j; r_j, \zeta) &= \frac{\partial \operatorname{bias}(\hat{r}_j; r_j, \zeta)}{\partial \epsilon} = \frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \epsilon} = \frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \tilde{\epsilon}} \frac{\partial \tilde{\epsilon}}{\partial \epsilon} \\
&= \frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \tilde{\epsilon}} \frac{1}{k(1 - \tilde{\epsilon})^{k-1}}.
\end{aligned}
\tag{70}
$$

Differentiating (58) with respect to ϵ and evaluating it at $\epsilon = 0$ yields

$$
\left.\frac{\partial b(r_j, \zeta, \epsilon, k)}{\partial \epsilon}\right|_{\epsilon=0} = -\frac{\sum_{j \in \mathcal{I} \setminus \{A\}} \mathfrak{p}'_j(0) j(r_j, \zeta, 0) - A(r_j, 0) \sum_{j \in \mathcal{I} \setminus \{A\}} \mathfrak{p}'_j(0)}{A'_b(r_j, 0)},
\tag{71}
$$

where $A'_b(r_j, 0)$ is defined in (39) and where (see results in Table 4)

$$\mathfrak{p}'_B(0) = \min\{s, k\}, \tag{72}$$

$$\mathfrak{p}'_C(0) = p + k - \max\{p, k\}, \tag{73}$$

$$\mathfrak{p}'_D(0) = \max\left\{0, s + \min\{p, k\} - \max\{s, k\}\right\}, \tag{74}$$

$$\mathfrak{p}'_E(0) = \max\left\{0, s + k - \max\{s + p, k\}\right\}, \tag{75}$$

$$\mathfrak{p}'_F(0) = 0, \tag{76}$$

$$\mathfrak{p}'_G(0) = \max\left\{0, \min\{s + p, k\} - s\right\}, \tag{77}$$

$$\mathfrak{p}'_H(0) = 0, \tag{78}$$

$$\mathfrak{p}'_I(0) = \max\{0, k - s - p\}, \tag{79}$$

$$\mathfrak{p}'_J(0) = 0. \tag{80}$$

As in "Appendix A.1", $A(r_j; 0) = 1/2$. Further, it follows from Dhaene and Zhu (2017, Lemma A.1) that, for $X, Z \sim N(0, 1)$ and constants c, c', $P\{(X + c')/(Z - c) \leq 0\} = \Phi(-c')\Phi(-c) + \Phi(c')\Phi(c)$. Hence, the definition (69) and the standardization (34)–(36) imply

$$B(r_j; \zeta, 0) = \frac{1}{2}, \tag{81}$$

$$C(r_j; \zeta, 0) = \Phi\left(-\frac{r_j\zeta}{\sigma_u}\right)\Phi\left(-\frac{\zeta}{\sigma_{\Delta^p}}\right) + \Phi\left(\frac{r_j\zeta}{\sigma_u}\right)\Phi\left(\frac{\zeta}{\sigma_{\Delta^p}}\right), \tag{82}$$

$$D(r_j; \zeta, 0) = \Phi\left(-\frac{(1 + r_j)\zeta}{\sigma_u}\right)\Phi\left(-\frac{\zeta}{\sigma_{\Delta^p}}\right) + \Phi\left(\frac{(1 + r_j)\zeta}{\sigma_u}\right)\Phi\left(\frac{\zeta}{\sigma_{\Delta^p}}\right), \tag{83}$$

$$E(r_j; \zeta, 0) = \Phi\left(-\frac{(1 + 2r_j)\zeta}{\sigma_u}\right)\Phi\left(-\frac{2\zeta}{\sigma_{\Delta^p}}\right) + \Phi\left(\frac{(1 + 2r_j)\zeta}{\sigma_u}\right)\Phi\left(\frac{2\zeta}{\sigma_{\Delta^p}}\right), \tag{84}$$

$$G(r_j; \zeta, 0) = \Phi\left(-\frac{(2 + r_j)\zeta}{\sigma_u}\right)\Phi\left(-\frac{\zeta}{\sigma_{\Delta^p}}\right) + \Phi\left(\frac{(2 + r_j)\zeta}{\sigma_u}\right)\Phi\left(\frac{\zeta}{\sigma_{\Delta^p}}\right), \tag{85}$$

$$I(r_j; \zeta, 0) = \Phi\left(-\frac{2(1 + r_j)\zeta}{\sigma_u}\right)\Phi\left(-\frac{2\zeta}{\sigma_{\Delta^p}}\right) + \Phi\left(\frac{2(1 + r_j)\zeta}{\sigma_u}\right)\Phi\left(\frac{2\zeta}{\sigma_{\Delta^p}}\right), \tag{86}$$

where $\sigma_u := \sqrt{\mathrm{var}(u_{itj})}$ and $\sigma_{\Delta^p} := \sqrt{\mathrm{var}(\Delta^p y_{it-s})}$ are given in (26) and $r_j = -(1 - \alpha^s)/2$. Substituting (71)–(86) in (70) completes the proof. $\qquad\square$

A.4 General results

Proof of Theorem 4 Given a non-stochastic weighting matrix A^0, the proof follows directly from Eq. (8). The estimator $\hat{\alpha}^0$ is defined by the solution of the sample analogs

of Eq. (5), which are deterministic functions of $\hat{r}_j$. Thus the influence function of $\hat{\alpha}^0$ is fully determined by the influence functions of each $\hat{r}_j$ being an element of $\boldsymbol{g}(\alpha)$:

$$\text{IF}(\hat{\alpha}^0; \alpha, \zeta) = -(\boldsymbol{d}'A^0\boldsymbol{d})^{-1}\boldsymbol{d}'A^0\boldsymbol{\psi}, \tag{87}$$

where $\boldsymbol{\psi} := \left(\text{IF}(\hat{r}_j; r_j, \zeta)\right)_{j \in \mathcal{J}_o}$ is a $\#\mathcal{J}_o \times 1$ vector whose elements $\text{IF}(\hat{r}_j; r_j, \zeta)$, $j \in \mathcal{J}_o$, are derived for each considered data contamination $Z^1_{\epsilon,\zeta}$, $Z^2_{\epsilon,\zeta}$, and $Z^3_{\epsilon,\zeta}$ in Theorem 1, 2, and 3, respectively.

If the weight matrix $A_n = A(\hat{\alpha}^0_n, \epsilon)$, then it follows by the same argument as above ($\boldsymbol{d}$ is still deterministic) and the matrix differentiation rules that

$$\text{IF}(\hat{\alpha}; \alpha, \zeta) = -(\boldsymbol{d}'A\boldsymbol{d})^{-1}(\boldsymbol{d}'A\boldsymbol{\psi} + \boldsymbol{d}'\dot{A}\boldsymbol{g}(\alpha)) - (\boldsymbol{d}'A\boldsymbol{d})^{-1}(\boldsymbol{d}'\dot{A}\boldsymbol{d})(\boldsymbol{d}'A\boldsymbol{d})^{-1}\boldsymbol{d}'A\boldsymbol{g}(\alpha), \tag{88}$$

where $\dot{A} = \text{IF}(A_n; A, \zeta)$. Since this influence function is bounded, the result follows from the asymptotic validity of the moment conditions $\boldsymbol{g}(\alpha) = 0$. $\qquad\square$

References

Abrevaya J (2000) Rank estimation of a generalized fixed-effects regression model. J Econ 95(1):1–23
Aquaro M, Čížek P (2013) One-step robust estimation of fixed-effects panel data models. Comput Stat Data Anal 57(1):536–548
Aquaro M, Čížek P (2014) Robust estimation of dynamic fixed-effects panel data models. Stat Pap 55:169–186
Arellano M, Bond S (1991) Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. Rev Econ Stud 58(2):277–297
Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. J Econ 87(1):115–143
Bramati MC, Croux C (2007) Robust estimators for the fixed effects panel data model. Econ J 10(3):521–540
Bun MJG, Sarafidis V (2015) Dynamic panel data models. In: Baltagi BH (ed) The Oxford handbook of panel data. The Oxford University Press, Oxford, pp 76–110
Cermeño R (1999) Median-unbiased estimation in fixed-effects dynamic panels. Annales dconomie et de Statistique 55–56:351–368
Cheng X, Liao Z (2015) Select the valid and relevant moments: an information-based LASSO for GMM with many moments. J Econ 186(2):443–464
Dhaene G, Zhu Y (2017) Median-based estimation of dynamic panel models with fixed effects. Comput Stat Data Anal 113:398–423
Galvao AF Jr (2011) Quantile regression for dynamic panel data with fixed effects. J Econ 164(1):142–157
Gervini D, Yohai VJ (2002) A class of robust and fully efficient regression estimators. Ann Stat 30(2):583–616
Hall AR, Inoue A, Jana K, Shin C (2007) Information in generalized method of moments estimation and entropy-based moment selection. J Econ 138:488–512
Han C, Phillips PCB, Sul D (2014) X-differencing and dynamic panel model estimation. Econ Theory 30:201–251
Hansen LP (1982) Large sample properties of generalized method of moments estimators. Econometrica 50(4):1029–1054
Harris MN, Mátyás L, Sevestre P (2008) Dynamic models for short panels, chapter 8. In: Mátyás L, Sevestre P (eds) The econometrics of panel data. Springer, Berlin, pp 249–278
Janz N (2002) Outlier robust estimation of an Euler equation investment model with German firm level panel data. In: Klein I, Mittnik S (eds) Contributions to modern econometrics. Dynamic modeling and econometrics in economics and finance, vol 4. Springer, Boston

Kiviet JF (2007) Judging contending estimators by simulation: tournaments in dynamic panel data models, chapter 11. In: Phillips G, Tzavalis E (eds) The refinement of econometric estimation and test procedures. Cambridge University Press, Cambridge, pp 282–318

Lucas A, van Dijk R, Kloek T (2007) Outlier robust GMM estimation of leverage determinants in linear dynamic panel data models. https://doi.org/10.2139/ssrn.20611

Verardi V, Wagner J (2011) Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium. J Econ Stat 231(4):546–557

Zaman A, Rousseeuw PJ, Orhan M (2001) Econometric applications of high-breakdown robust regression techniques. Econ Lett 71(1):1–8