Trimmed Mean Group Estimation

Yoonseok Lee and Donggyu Sul

Paper No. 237
February 2021
CENTER FOR POLICY RESEARCH – Spring 2021

Leonard M. Lopoo, Director
Professor of Public Administration and International Affairs (PAIA)

Associate Directors
Margaret Austin
Associate Director, Budget and Administration

John Yinger
Trustee Professor of Economics (ECON) and Public Administration and International Affairs (PAIA)
Associate Director, Center for Policy Research

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON
Robert Bifulco, PAIA
Leonard Burman, PAIA
Carmen Carrión-Flores, ECON
Alfonso Flores-Lagunes, ECON
Sarah Hamersma, PAIA
Madonna Harrington Meyer, SOC
Colleen Heflin, PAIA
William Horrace, ECON
Yilin Hou, PAIA
Hugo Jales, ECON

Jeffrey Kubik, ECON
Yoonseok Lee, ECON
Amy Lutz, SOC
Yingyi Ma, SOC
Katherine Michelmore, PAIA
Jerry Miner, ECON
Shannon Monnat, SOC
Jan Ondrich, ECON
David Popp, PAIA
Stuart Rosenthal, ECON
Michah Rothbart, PAIA

Alexander Rothenberg, ECON
Rebecca Schewe, SOC
Amy Ellen Schwartz, PAIA/ECON
Ying Shi, PAIA
Saba Siddiki, PAIA
Perry Singleton, ECON
Yulong Wang, ECON
Peter Wilcoxen, PAIA
Maria Zhu, ECON

GRADUATE ASSOCIATES

Rhea Acuña, PAIA
Graham Ambrose, PAIA
Mariah Brennan, SOC. SCI.
Ziqiao Chen, PAIA
Yoon Jung Choi, PAIA
Dahae Choo, ECON
Stephanie Coffey, PAIA
William Clay Fannin, PAIA
Giuseppe Germinario, ECON
Myriam Gregoire-Zawilski, PAIA
Jeehee Han, PAIA

Mary Helander, SOC. SCI.
Amra Kandic, SOC
Sujung Lee, SOC
Mattie Mackenzie-Liu, PAIA
Maeve Maloney, ECON
Austin McNeill Brown, SOC. SCI.
Qasim Mehdi, PAIA
Nicholas Osterling, PAIA
Claire Pendergrast, SOC
Lauryn Quick, PAIA
Krushna Ranaware, SOC

Sarah Reilly, SOC
Christopher Rick, PAIA
Spencer Shanholtz, PAIA
Sarah Souders, PAIA
Joaquin Urrego, ECON
Yao Wang, ECON
Yi Yang, ECON
Xiaoyan Zhang, Human Dev.
Bo Zheng, PAIA
Dongmei Zuo, SOC. SCI.

STAFF

Ute Brady, Postdoctoral Scholar
Willy Chen, Research Associate
Katrina Fiacchi, Administrative Specialist
Michelle Kincaid, Senior Associate, Maxwell X Lab

Emily Minnoe, Administrative Assistant
Candi Patterson, Computer Consultant
Samantha Trajkovski, Postdoctoral Scholar
Laura Walsh, Administrative Assistant
Abstract

This paper develops robust panel estimation in the form of trimmed mean group estimation for potentially heterogenous panel regression models. It trims outlying individuals of which the sample variances of regressors are either extremely small or large. The limiting distribution of the trimmed estimator can be obtained in a similar way to the standard mean group estimator, provided the random coefficients are conditionally homoskedastic. We consider two trimming methods. The first one is based on the order statistic of the sample variance of each regressor. The second one is based on the Mahalanobis depth of the sample variances of regressors. We apply them to the mean group estimation of the two-way fixed effects model with potentially heterogeneous slope parameters and to the common correlated effects regression, and we derive limiting distribution of each estimator. As an empirical illustration, we consider the effect of police on property crime rates using the U.S. state-level panel data.

JEL No.: C23, C33

Keywords: Trimmed Mean Group Estimator, Robust Estimator, Heterogeneous Panel, Random Coefficient, Two-Way Fixed Effects, Common Correlated Effects

Authors: Yoonseok Lee, Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244, ylee41@maxwell.syr.edu; Donggyu Sul, Department of Economics, University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080, d.sul@utdallas.edu
1 Introduction

Though it is popular in practice to impose homogeneity in panel data regression, the homogeneity restriction is often rejected (e.g., Baltagi, Bresson, and Pirotte, 2008). Even under the presence of heterogeneous slope coefficients, pooling the observations is widely believed to be innocuous. As Baltagi and Griffin (1997) and Woodridge (2005) point out, the standard fixed-effects (FE) estimators consistently estimate the mean of the heterogeneous slope coefficients. One may use the group-heterogeneity approach by Bonhomme and Manresa (2015) or Su, Shi, and Phillips (2016), which assumes homogenous slope coefficients within a group but heterogeneous across groups. However, with a large cross section size, it is still unknown whether the true slope coefficients are homogenous even within each group.

Meanwhile, under the assumption of heterogeneous panels, Pesaran and Smith (1995) and Pesaran, Shin, and Smith (1999) proposed the mean group (MG) estimator, which is the cross-sectional average of individual-specific time series least squares (LS) estimators. In particular, the MG estimator is the most preferable when estimating the long-run effects in dynamic panel models (e.g., Pesaran, Smith, and Im, 1996). Maddala, Trost, Li and Joutz (1997) considered a model average and shrinkage estimator by combining the individual time series LS estimators as well as the FE estimator. In practice, however, the pooled or two-way FE estimator has been more popular than the MG estimator, mostly because the MG estimator is known to be less efficient. Baltagi, Griffin, and Xiong (2000) empirically demonstrate that the pooling method leads to more accurate forecasts than the MG estimator.

The purpose of this paper is to revisit a salient feature of the MG estimator. We emphasize the importance of investigating the individual time series estimators in details when pooling or averaging. In particular, we develop the trimmed MG estimator, where we trim individual observations whose time series sample variances of regressors are either extremely large or small. The trimming instruments are not the estimates of the individual slope coefficients themselves, but the time series variances of the regressors. Therefore, it naturally reflects the standard error of each time series estimator or the interval estimates. Once we have the sample proportion of trimming, the limiting distribution of the trimmed MG estimator can be obtained in a similar way to the standard MG estimator, provided the (potentially) heterogeneous slope coefficients are conditionally homoskedastic.

To obtain more robust estimation results, researchers often drop cross-sectional units in a panel data set, whose time series variances are unusually large compared to the rest of the units. Trimming individual observations with a large sample variance of the regressor (say right-tail trim) in our case can be understood in a similar vein. However, dropping such observations will result in efficiency loss of the MG or any pooled estimators. Trimming individual observations with a small sample variance of the regressor (say left-tail trim) is basically the same as dropping the individual time series estimators whose standard errors are large. Therefore, the left-tail
trim will offset efficiency loss from the right-tail trim and result in a robust MG estimator with little efficiency loss in finite samples.

In Section 2, we motivate the trimmed MG estimation by comparing popular panel estimators in the context of the weighted average (or weighted MG) estimator. We also discuss trimming methods in practice and provide some economic examples. Section 3 formally develops the trimmed MG estimators in the two-way FE model and the common correlated effects (CCE) regression and studies their asymptotic properties. Section 4 reports the results of Monte Carlo studies and Section 5 presents an empirical illustration of the effect of police on property crime rates across 48 contiguous states of the U.S. Section 6 concludes.

2 Motivation

2.1 Weighted mean group estimation

We consider a panel regression model given by

\[ y_{it} = a_i + F_t + x_{it}'\beta_i + u_{it} \] (1)

for \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \), where ‘i’ stands for the \( i \)th individual and ‘t’ stand for the \( t \)th time. \( x_{it} \) is an \( m \times 1 \) vector of exogenous regressors of interest. \( a_i \) and \( F_t \) are individual and time fixed effects, respectively. Instead of a time effect \( F_t \), a factor augmented term \( \lambda_t F_t \) can be considered. The regression coefficient \( \beta_i \) can be either homogeneous or heterogenous across \( i \), which is unknown.

We are interested in estimating the mean of the individual specific slope coefficients, \( \beta = \mathbb{E}[\beta_i] \). When \( \beta_i \) is heterogeneous, we suppose that

\[ \beta_i = \beta + \varepsilon_i \quad \text{with} \quad \varepsilon_i|x_i \sim iid(0, \Omega) \] (2)

for some \( m \times m \) matrix \( 0 < \Omega < \infty \), where \( x_i = (x_{i1}, \ldots, x_{iT})' \). When \( \beta_i \) is indeed homogeneous, \( \beta \) corresponds to the true slope parameter value. We estimate \( \beta \) in the form of a weighted average (or weighted mean-group) estimator given by

\[ \hat{\beta} = \sum_{i=1}^{n} w_i \hat{\beta}_i \] (3)

for some non-negative weights \( \{w_i\} \) such that \( \sum_{i=1}^{n} w_i = 1 \). The weight \( w_i \) can be either a scalar
or a matrix. \( \hat{\beta}_i \) is typically the least squares estimator of \( \beta_i \) for each \( i \):

\[
\hat{\beta}_i = \hat{\Sigma}_i^{-1} \frac{1}{T} \sum_{t=1}^{T} \hat{x}_{it} \hat{y}_{it},
\]

where we denote the sample variance of the regressor in (1) as

\[
\hat{\Sigma}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{x}_{it} \hat{x}_{it}'
\]

with

\[
\hat{x}_{it} = \bar{x}_{it} - \frac{1}{n} \sum_{j=1}^{n} \hat{x}_{jt} \quad \text{and} \quad \bar{x}_{it} = x_{it} - \frac{1}{T} \sum_{s=1}^{T} x_{is}.
\]

(4)

Popular examples include the mean group (MG) estimator \( \hat{\beta}_{MG} \) by Pesaran and Smith (1995) and Pesaran, Shin, and Smith (1999), which uses equal (scalar) weights as

\[
w_{MG,i} = \frac{1}{n}
\]

in (3). The Bayes (SB) estimator \( \hat{\beta}_{SB} \) by Swamy (1970) and Smith (1973) is also a weighted average estimator with \( w_i \) as the inverse of a consistent estimator of \( \text{var}(\beta_i) \) in the context of random effect models. In particular, when \( u_{it} \sim iid \mathcal{N}(0, \sigma^2_{ui}) \) for each \( i \), we define the weights as

\[
w_{SB,i} = \left( \sum_{j=1}^{n} \left\{ \hat{\sigma}^2_{uj} \hat{\Sigma}_j^{-1} / T + \hat{\Omega} \right\}^{-1} \right)^{-1} \left\{ \hat{\sigma}^2_{ui} \hat{\Sigma}_i^{-1} / T + \hat{\Omega} \right\}^{-1}
\]

(6)

in (3) using some consistent estimators \( \hat{\sigma}^2_{ui} \) and \( \hat{\Omega} \) in (2). These two estimators are known to be asymptotically equivalent under \( n, T \to \infty \) when \( \sqrt{n}/T \to 0 \) (e.g., Hsiao, Pesaran, and Tahmiscioglu, 1999).

When slope homogeneity tests or panel poolability tests (e.g., Pesaran and Yamagata, 2008) support \( \beta_i = \beta \) for all \( i \), the two-way fixed-effects (FE) estimator

\[
\hat{\beta}_{FE} = \left( \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{x}_{it} \hat{x}_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{x}_{it} \hat{y}_{it}
\]

is often used, which is consistent and the most efficient under the spherical error variance structure. In fact, it can be also expressed in the form of the average estimator in (3) with the
weights
\[ w_{FE,i} = \left( \sum_{j=1}^{n} \hat{\Sigma}_j \right)^{-1} \hat{\Sigma}_i, \quad (7) \]
which is proportional to the inverse of \( \text{var}(\hat{\beta}_i) \) for each \( i \) under homogeneity (e.g., Sul, 2016). In this case, this weight is optimal in the sense that \( \hat{\beta}_{FE} \) is the best linear unbiased estimator of \( \beta \).

2.2 Trimming based on the variances of regressors
Comparing the weights (5), (6), and (7), we can see that \( \hat{\beta}_{MG} \) puts equal weights for all \( \hat{\beta}_i \); but both \( \hat{\beta}_{SB} \) and \( \hat{\beta}_{FE} \) put unequal weights over \( \hat{\beta}_i \), where the weights are proportional to \( \hat{\Sigma}_i \), the sample variance of \( x_{it} \) of each \( i \). Note that these weights are basically proportional to the inverse of each individual variance of \( \hat{\beta}_i \), and hence \( \hat{\beta}_{SB} \) and \( \hat{\beta}_{FE} \) are to be more efficient than \( \hat{\beta}_{MG} \). See Figure 1 for comparison, which depicts the weights of MG and SB (or FE) estimators as functions of the sample variance of \( x_{it}, \hat{\Sigma}_i \), when \( \beta_i \) is a scalar.

Because of its lower efficiency, the MG estimator \( \hat{\beta}_{MG} \) is not popular in practice. However, since \( \hat{\beta}_{MG} \) does not use the weights based on \( \hat{\Sigma}_i \), it is more robust toward extreme values or behaviors of \( x_{it} \). In contrast, when \( \hat{\Sigma}_i \) is extremely large for some individual \( i \) because of some outlying observations in \( x_{it} \), both \( \hat{\beta}_{SB} \) and \( \hat{\beta}_{FE} \) are heavily influenced by such an individual. In some extreme cases, it can even result in inconsistency of \( \hat{\beta}_{SB} \) and \( \hat{\beta}_{FE} \).

In fact, it is a common practice in panel data analysis that \( \text{ith} \) observations with large time series variation of \( x_{it} \) are considered as outlying individuals and dropped to get more
robust estimators of the key structural parameters. For example, consider an uncovered interest parity regression, where the dependent variable is a depreciation rate (or a growth rate of a foreign exchange rate) and an independent variable is the difference between domestic and foreign interest rates. In this context, empirical researchers often exclude foreign countries that experienced currency or financial crises, and their interest or exchange rates have large time series fluctuations (e.g., a sudden increase during the crisis). Furthermore, a sudden change in an independent variable during a particular time period could make the time series more persistent, which even leads to nonstationarity. Unless all other variables are nonstationary and they are cointegrated, it can yield spurious regression results.

Based on such observations, we propose a simple robust estimator, the trimmed mean group (TMG) estimator, where we still use the equal scalar weights but drop individuals whose $\Sigma_i$ are extremely large. In particular, we take an equally weighted average over $\hat{\beta}_i$'s in (3), but the weights take a hard-threshold trimming based on the sample variance of $x_{it}$ of each $i$, $\Sigma_i$. However, such an estimator could suffer from big efficiency loss compared with the standard MG, SB, or FE estimators, though it achieves higher robustness. As a way of improving the efficiency, we also drop $i$th observations whose sample variances of $x_{it}$ are extremely small (i.e., $\hat{\beta}_i$'s with large standard errors). The resulting weighting scheme of the TMG that we develop in this paper is depicted in Figure 2, as a function of $\hat{\Sigma}_i$.

In sum, we consider the double-sided trimming scheme that drops $i$th observations whose sample variances of $x_{it}$ are either extremely large or small. Trimming individual observations with large sample variances of $x_{it}$ (say, right-tail trim) will give robustness toward some outlying
individuals. On the other hand, trimming individual observations with small sample variances of \( x_{it} \) (say, left-tail trim) will improve efficiency by offsetting some efficiency loss from dropping individual observations with large sample variances of \( x_{it} \) in the right-tail trim, which is in the same (but more extreme) spirit as the SB or FE estimators as we illustrated in Figure 2.

**Remark** Instead of the hard-threshold trimming, we could consider a trapezoid type weight. We could also apply our trimming scheme on the SB estimator instead of the MG estimator, where the “SB” line in Figure 2 is forced to be zero when it is near the origin or extremely large. For the latter case, however, the weights are no longer scalar values and we need to iterate estimation to obtain such (infeasible) GLS weights, whose finite sample property does not necessarily dominate its OLS counterpart—TMG estimator we propose. We hence focus on the equal (scalar) weight with trimming in this paper.

### 2.3 Examples

We present some empirical evidence where the trimming idea could improve the results. The first example shows the case when large sample variance of the regressor matters. Lee and Sul (2019) consider the following relative Purchasing Power Parity (PPP) regression:

\[
\Delta s_{it} = a_i + X_i F_i + \beta_i (\pi_{it} - \pi_i^*) + u_{it},
\]

where \( s_{it} \) and \( \pi_{it} \) are the foreign exchange rate and the inflation rate of the \( i \)th foreign country. \( \pi_i^* \) stands for the domestic (U.S. in this example) inflation rate. As Greenaway-McGregory, Mark, Sul, and Wu (2018) showed, the relative PPP in the idiosyncratic level does not hold even in the short run. In other words, the idiosyncratic inflation differential is independent of the idiosyncratic change of the foreign exchange rate.

Figure 3 plots 27 point estimates \( \hat{\beta}_i \) \((n = 27)\) in Lee and Sul (2019) against the variances of inflation rates.\(^1\) The total number of time series sample is \( T = 202 \) (from 1999.M2 to 2015.M11). The point estimate of \( \beta_i \) of Turkey is near unity, but the inflation differential between Turkey and the U.S. is quite huge. Due to several currency crises in Turkey, the inflation rate in Turkey has widely fluctuated, which results in a large time series variance of the inflation differential. Usually the relative PPP has been investigated in countries with stable inflation, and hence researchers may want to exclude Turkey in their sample. In fact, as Lee and Sul (2019) point out, the FE estimator \( \hat{\beta}_{FE} \) assigns a huge weight on the point estimate of Turkey, which leads \( \hat{\beta}_{FE} \) to be biased. However, if we exclude individuals whose time series variances are extremely large, we can get a more robust FE estimator.

---

\(^1\)For the factor-augmented regression, we use the method by Greenaway-McGregory, Han, and Sul (2012).
The second example shows the opposite case when a small sample variance of the regressor matters. We estimate an individual marginal propensity to consume (MPC) across 1,875 households \((n = 1875)\) in South Korea during the period from 1999 to 2003 \((T = 5)\). The data source is from the Korea Labor Income Panel Study (KLIPS). To estimate the MPC, we run the following two-way FE regression with potentially heterogeneous slope parameters:

\[
\Delta \ln E_{it} = a_i + F_i + \beta_i \Delta \ln Y_{it} + u_{it},
\]

where \(E_{it}\) and \(Y_{it}\) are the annual expenditure and wage income of the \(i\)th household at year \(t\). Theoretically, the MPC is supposed to be between zero and unity.

Figure 4 shows the MPC estimates \(\hat{\beta}_i\) against the sample variance of the regressor \(\Delta \ln Y_{it}\). Figure 4-A plots all the estimates. Evidently the MPC estimates widely fluctuate, especially when the time series variances of \(\Delta \ln Y_{it}\) are small. Figure 4-B magnifies the area of Figure 4-A where the time series variances of \(\Delta \ln Y_{it}\) are near zero. It clearly shows that the point estimates of the MPC becomes unreasonably larger as the variances of the regressor get smaller. In fact, many estimates are larger than unity and even negative. If we exclude \(\hat{\beta}_i\) of which the variances of the regressor are near zero, we can obtain a more efficient estimator of the mean of individual MPC, say \(E[\beta_i]\) in this example.

### 2.4 Trimming weights

We can construct the trimming weights based on the sample covariance of the regressors, \(\hat{\Sigma}_i\), in practice. When \(x_{it}\) is univariate (i.e., \(m = 1\)), we can simply sort \(\hat{\Sigma}_i\) and decide the trimming
point. When $x_{it}$ is multivariate (i.e., $m \geq 2$), however, such ordering is not straightforward. In such cases, we can consider either a marginal trimming scheme or a joint (or balanced) trimming scheme.

For the marginal trimming approach, we pick one regressor and determine trimmed cross-sectional units based on the order statistics of the sample variance of the regressor picked. In this case, a researcher chooses the key regressor as a trimming instrument or the one with the largest variation. We can also apply this marginal trimming idea on each regressor $j = 1, \ldots, m$ and run the regression $m$ times using each trimmed sample in turn. From each $j$th regression, we only report the estimates of the $j$th element of the TMG estimator that corresponds to the slope coefficient of the $j$th regressor.

The joint trimming approach can be done using the intersection of trimmed sets from the aforementioned marginal trimming for all $j = 1, \ldots, m$. More precisely, we let $\hat{\sigma}_{j,i}^2$ be the sample variance of the $j$th element of the agent $i$'s regressors for $j = 1, \ldots, m$. We conduct marginal trimming using $\hat{\sigma}_{j,i}^2$ for each $j$ and obtain the joint trimmed set from the intersection of all the trimmed units over $j = 1, \ldots, m$.

Alternatively, we can use the data depth of $\hat{\Sigma}_i$ (e.g., Lee and Sul, 2019), based on which we conduct trimming using the depth-induced statistic. For instance, we form a contour plot over the $m$-dimensional space based on the sample Mahalanobis depth of $\hat{\nu}_i = (\hat{\sigma}_{1,i}^2, \hat{\sigma}_{2,i}^2, \ldots, \hat{\sigma}_{m,i}^2)'$ defined as

$$\hat{D} (v) = [1 + (v - \hat{\mu}_v)' \hat{\Lambda}_v^{-1} (v - \hat{\mu}_v)]^{-1},$$

where $\hat{\mu}_v$ and $\hat{\Lambda}_v$ are the sample mean and sample variance of $\hat{\nu}_i$. By construction, $\hat{D} (v) \in [0, 1]$ and it is close to zero if $v = \hat{\nu}_i$ is either extremely small or large (i.e., for outlying $\hat{\nu}_i$ from the center of its distribution). Using this data depth, we can construct multivariate quantile
contours and determine the trimming contour line for any dimension of $x_{it}$. In this case, we define the TMG estimator where we trim $i$th observations with $\tilde{D}(\tilde{\gamma}_i) < d$ for some threshold $0 < d < 1$. Note that this depth-based approach automatically trims outliers on both sides, possibly asymmetrically. In addition, the Mahalanobis depth is affine invariant, and hence the TMG estimator based on the depth-based weights has the invariance property to non-singular linear transformations like the typical least squares estimator.

Though we focus on the diagonal elements of $\tilde{\Sigma}_i$ in the trimming methods, there are cases when we need to consider the off-diagonal terms of $\tilde{\Sigma}_i$, such as the regressors of the individual $i$ impose near multicollinearity (including the cases with nearly time-invariant regressors in the FE model) or cross-product terms are sources of outlyingness. In such cases, we can also consider the determinant of $x_i$ or particular covariances when ranking the individuals, in addition to the variance.

3 Trimmed Mean Group Estimation

3.1 Trimmed MG estimator for two-way FE models

The trimming scheme described in the previous section drops the entire history of the $i$th individuals when their $\tilde{\Sigma}_i$ are either extremely large or small. Therefore, in the one-way fixed-effect (FE) model, $y_{it} = a_i + x_{it}'\beta_i + u_{it}$, subtracting individual mean (i.e., within transformation) for each $i$ is not affected from the trimming step. Hence, it does not matter whether we trim before or after the within transformation in this case.

In contrast, the TMG estimator for the two-way FE regression in (1)

$$y_{it} = a_i + F_t + x_{it}'\beta_i + u_{it}$$

needs caution. This is because when we take the Wallace-Hussain transformation (Wallace and Hussain, 1969) as in (4), subtracting $\frac{1}{n} \sum_{t=1}^{n} x_{it}$ should be modified by the trimming step because we will drop some $i$ observations. In particular, we subtract the trimmed mean $(1/n_G) \sum_{i \in G_n} x_{it}$, where $G_n$ is the set of individuals that are not trimmed in the sample and $n_G$ is the cardinality of $G_n$. We define the Wallace-Hussain transformed $x_{it}$ as

$$\tilde{x}_{it} = \tilde{x}_{it} - \frac{1}{n_G} \sum_{j \in G_n} \tilde{x}_{jt},$$

where $\tilde{x}_{it} = x_{it} - T^{-1} \sum_{s=1}^{T} x_{is}$, instead of (4). So, in practice, we trim the individual observations first and then take the Wallace-Hussain transformation.
Whether it is the one-way or two-way FE model, once we conduct the proper demeaning using the trimmed sample as explained above, we obtain the individual trimmed least squares estimator \( \hat{\beta}_i^T \) by regressing \( \hat{y}_{it}^T \) on \( \hat{x}_{it}^T \). Then, we define the TMG estimator as

\[
\hat{\beta}_{TMG} = \frac{1}{n_G} \sum_{i \in G_n} \hat{\beta}_i^T = \sum_{i=1}^{n} w_i^T \hat{\beta}_i^T \text{ with } w_i^T = \frac{1}{n_G} 1 \{ i \in G_n \},
\]

where \( 1 \{ \cdot \} \) is the binary indicator. Recall that \( n_G = \sum_{i=1}^{n} 1 \{ i \in G_n \} \) and \( G_n \) is the set of individuals that are not trimmed in the sample.

It is important to note that, though the Wallace-Hussain transformation in (4) eliminates fixed effects \( a_i \) and \( F_i \) in the heterogeneous two-way FE regression (1), it does not yields the desired regression equation given as

\[
\hat{y}_{it}^T = \hat{x}_{it}^T \beta_i + \hat{u}_{it}^T.
\]

Instead, it results in a transformed regression

\[
\hat{y}_{it}^T = \hat{x}_{it}^T \beta_i - \frac{1}{n_G} \sum_{j \in G_n} \hat{x}_{jt}^T \beta_j + \hat{u}_{it}^T = \hat{x}_{it}^T \beta_i + \hat{\epsilon}_{it}^T,
\]

where

\[
\hat{\epsilon}_{it}^T = \hat{\xi}_{it} + \hat{u}_{it}^T \text{ with } \hat{\xi}_{it} = \frac{1}{n_G} \sum_{j \in G_n} \hat{x}_{jt}^T (\beta_i - \beta_j).
\]

Because of \( \hat{\xi}_{it} \) term, therefore, the least squares estimator \( \hat{\beta}_i^T \) by regressing \( \hat{y}_{it}^T \) on \( \hat{x}_{it}^T \) for each \( i \) does not necessarily yield a consistent estimator of \( \beta_i \). In other words, unlike the homogeneous two-way FE model, \( \hat{\beta}_i - \beta_i = O_p(T^{-1/2}) \) no longer holds in this case. Note that this result is not because of the trimming, and the same issue applies to the standard MG estimator for the heterogeneous two-way FE regression. However, it can be readily verified that the MG estimator is still consistent and achieves the asymptotic normality. The same results extend to the case of TMG estimator in (8) as summarized in the following theorem. It, hence, implies that whether the true model is homogeneous or heterogeneous, we can use the Wallace-Hussain transformation for the (trimmed) MG estimation of two-way FE regression models.\(^2\)

**Theorem 1** Suppose the random coefficient \( \beta_i \) satisfies (2). Also let \( n_G \to \infty \) as \( n \to \infty \) satisfying \( \phi = \lim_{n \to \infty} n/n_G \) and \( \phi \in [1, \infty) \). Under the same condition in Theorem 1 of Pesaran

\(^2\)A similar result was also found in Lee, Mukherjee, and Ullah (2019) in the context of a partially linear two-way FE regression, where the linearized form can be seen as a heterogeneous panel model.
we have \(\sqrt{n}(\beta_{TMG} - \beta) \rightarrow_d \mathcal{N}(0, V_{TMG})\) as \(n, T \rightarrow \infty\), where \(V_{TMG} = \phi \Omega\).

**Proof of Theorem 1** Note that \(T^{-1} \sum_{t=1}^{T} \tilde{x}_{it}^{\top} \tilde{x}_{it}^{\top} \rightarrow_P \Sigma_{it} > 0\) for each \(i\) as \(T \rightarrow \infty\). Since 
\[
\tilde{\beta}_{it}^{\top} - \beta_{it} = (T^{-1} \sum_{t=1}^{T} \tilde{x}_{it}^{\top} \tilde{x}_{it}^{\top})^{-1} (T^{-1} \sum_{t=1}^{T} \tilde{x}_{it}^{\top} \tilde{x}_{it}^{\top} + T^{-1} \sum_{t=1}^{T} \tilde{x}_{it}^{\top} \tilde{u}_{it}),
\]
(9) and (10) yields 
\[
\sqrt{n_{G}}(\beta_{TMG} - \beta) = \frac{1}{\sqrt{n_{G}}} \sum_{i \in G} (\beta_{i} - \beta) + \frac{1}{\sqrt{n_{G}}} \sum_{i \in G} (\beta_{i}^{\top} - \beta_{i})
\]
\[
= \frac{1}{\sqrt{n_{G}}} \sum_{i \in G} (\beta_{i} - \beta) + \frac{1}{\sqrt{n_{G}}} \sum_{i \in G} (\Sigma_{ii})^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} \tilde{x}_{it}^{\top} \tilde{u}_{it} \right)
\]
\[
= \frac{1}{\sqrt{n_{G}}} \sum_{i \in G} (\Sigma_{ii})^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} \tilde{x}_{it}^{\top} \tilde{x}_{it} \right) + o_{P}(1)
\]
\[
\equiv M_{1n} + M_{2n} + M_{3n} + o_{P}(1),
\]
where \(M_{1n} \rightarrow_d \mathcal{N}(0, \Omega)\) as \(n, T \rightarrow \infty\) by the CLT, and \(M_{2n} = O_{P}(T^{-1/2})\) similarly to Theorem 1 of Pesaran (2006) since \(x_{it}\) is exogenous. For \(M_{3n}\), we can verify that 
\[
\frac{1}{\sqrt{n_{G}}} \sum_{i \in G} \sum_{t=1}^{T} \tilde{x}_{it}^{\top} \tilde{x}_{it} = \frac{1}{\sqrt{n_{G}}} \sum_{i \in G} \sum_{t=1}^{T} \left( \tilde{x}_{it}^{\top} - \frac{1}{n_{G}} \sum_{k \in \mathcal{G}_{n}} \tilde{x}_{kt} \right) \left( \frac{1}{n_{G}} \sum_{j \in \mathcal{G}_{n}} \tilde{x}_{jt}^{\top} (\beta_{i} - \beta_{j}) \right)
\]
\[
= \frac{1}{\sqrt{n_{G}}} \sum_{i \in G} \left( \frac{1}{T} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it} \right) (\beta_{i} - \beta),
\]
where \(\bar{x}_{i} = n_{G}^{-1} \sum_{j \in \mathcal{G}_{n}} \tilde{x}_{jt}\) and \(\bar{\beta} = n_{G}^{-1} \sum_{j \in \mathcal{G}_{n}} \beta_{j}\). We thus write 
\[
M_{3n} = \frac{1}{\sqrt{n_{G}}} \sum_{i \in G} (\Sigma_{ii})^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{i} \right) (\beta_{i} - \bar{\beta}),
\]
which satisfies \(\mathbb{E}[M_{3n}] = 0\) since \(\mathbb{E}[\beta_{i} - \bar{\beta} x_{i}] = 0\). Moreover, \(\mathbb{E}[(\beta_{i} - \bar{\beta})(\beta_{i} - \bar{\beta})^{\top} | x_{i}] = (1-(1/n)) \Omega\), which we denote \(\Omega_{n}\). Apparently, \(\Omega_{n} \rightarrow \Omega < \infty\) as \(n \rightarrow \infty\). Under cross-sectional independence,

\[\text{In this two-way FE case, in particular, we set} \ d_{t} = \gamma_{i} = \Gamma_{i} = 1 \text{in the multifactor error structure in (13) and (14). Hence,} \ \mathbb{E}[u_{it} | x_{it}, \Gamma_{it}, \alpha_{it}, \beta_{i}] = 0 \text{for all} \ i \text{and} \ t \text{; and} \ \{x_{it}, u_{it}, \beta_{i}, \alpha_{i}\} \text{are cross-sectionally independent, have bounded fourth moments, and are stationary and mixing over time with a proper mixing condition yielding the CLT. In addition,} \ T^{-1} \sum_{t=1}^{T} \tilde{x}_{it}^{\top} \tilde{x}_{it}^{\top} \rightarrow_P \Sigma_{it} > 0 \text{for each} \ i \text{as} \ T \rightarrow \infty. \text{However, the cross-sectional independence assumption is to simplify the proof. We can relax the cross-sectional independence assumption of} \ x_{it} \text{by imposing a common factor structure like (14) as in the following section.}\]

11
we have

\[
\mathbb{E}[M_{3n}M'_{3n}] = \frac{1}{n_g} \sum_{i \in G_n} (\Sigma_i')^{-1} \mathbb{E} \left[ \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \bar{x}_{it} \bar{x}_{is}' \mathbb{E} \left[ (\beta_i - \hat{\beta}) (\beta_i - \hat{\beta})' | x_i \right] \bar{x}_{is} \bar{x}_{is}' \right] (\Sigma_i')^{-1}
\]

\[
= \frac{1}{n_g} \sum_{i \in G_n} (\Sigma_i')^{-1} \mathbb{E} \left[ \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left( \bar{x}_{it} \bar{x}_{i}^{1/2} \right) \left( \bar{x}_{is} \bar{x}_{is}' \right)' \right] (\Sigma_i')^{-1}
\]

\[
= O \left( \frac{1}{T} + \frac{1}{n_g^2} \right)
\]

because under the stationarity with a proper mixing condition of \( x_{it} \),

\[
\mathbb{E} \left[ \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left( \bar{x}_{it} \bar{x}_{i}^{1/2} \right) \left( \bar{x}_{is} \bar{x}_{is}' \right)' \right]
\]

\[
= \frac{1}{T} \mathbb{E} \left[ \left( \bar{x}_{it} \bar{x}_{i}^{1/2} \right) \left( \bar{x}_{i}^{1/2} \right)' \right] + \frac{1}{T} \sum_{s=1}^{T-1} \left( 1 - \frac{s}{T} \right) \text{Cov} \left[ \bar{x}_{i} \bar{x}_{i}^{1/2}, \bar{x}_{is} \bar{x}_{is}' \right]
\]

\[
+ \frac{1}{T} \sum_{s=1}^{T-1} \left( 1 - \frac{s}{T} \right) \mathbb{E} \left[ \bar{x}_{i} \bar{x}_{i}^{1/2} \right] \mathbb{E} \left[ \bar{x}_{is} \bar{x}_{is}' \right]
\]

\[
= O \left( \frac{1}{T} + \frac{1}{n_g^2} \right)
\]

for \( \mathbb{E}[\bar{x}_{i} \bar{x}_{i}^{1/2}] = n_g^{-1} \sum_{j \in G_n} \mathbb{E}[\bar{x}'_i \bar{x}_j] \Omega_n^{1/2} = O(1/n_g) \). Therefore, \( M_{3n} = O_p(T^{-1/2} + n_g^{-1}) \), and the desired result follows as \( n, T \to \infty \) by pre-multiplying \( \sqrt{n}/n_g \) to (11). \( \square \)

Since \( \phi \geq 1 \), \( V_{TMG} \) cannot be smaller than the asymptotic variance of the standard MG estimator without trimming, which is \( \Omega \). Note that when the trimmed sample size is fixed (i.e., it does not depend on \( n \)), \( \phi \) is simply 1 and \( V_{TMG} \) reaches to the asymptotic variance of the standard MG estimator. It is worthy to note that, in the limit, the efficiency loss of the TMG estimator does not depend on the specific trimming scheme, whether to trim individual samples with extremely large or small \( \hat{\Sigma}_i \). It only depends on the reduction of the sample size from trimming. This is because we consider endogenous trimming; the efficiency gain from trimming the individual samples with small \( \hat{\Sigma}_i \) should be understood as the finite sample property.

The asymptotic variance can be consistently estimated by the sample covariance of \( \hat{\beta}_i' \) as

\[
\hat{V}_{TMG} = \frac{n}{n_g} \sum_{i \in G_n} (\hat{\beta}_i' - \hat{\beta}_{TMG})(\hat{\beta}_i' - \hat{\beta}_{TMG})'
\]

using the same argument in Section 8.2.2 of Pesaran, Smith and Im (1996).
3.2 Trimming CCEMG estimator

The two-way FE estimator in the previous section becomes inconsistent when a factor augmented term is included in the regression model. The pooled common correlated effects (CCE) estimator or the CCE mean-group (CCEMG) estimator by Pesaran (2006) and Chudik and Pesaran (2015) can be employed in this case. In particular, we consider a panel regression model with a multifactor error structure given by

\[
y_{it} = a'_i d_t + \gamma'_i f_t + x'_i \beta_i + u_{it}, \quad (13)
\]

\[
x_{it} = A'_i d_t + \Gamma'_i f_t + \epsilon_{it} \quad (14)
\]

for \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \), where \( d_t \) is the vector of observed factors and \( f_t \) is the vector of latent factors.

The original CCE estimator obtains the individual slope parameter estimates from the least squares of

\[
y_{it} = a'_i d_t + \delta'_i \pi_t + x'_i \beta_i + u_{it}
\]

for each \( i \), where \( \pi_t = \sum_{i=1}^n \pi_i z_{it} \) with \( z_{it} = (y_{it}, x'_{it})' \) for some weights \( \{\pi_i\} \) satisfying Assumption 5 of Pesaran (2006). For our case, however, \( \pi_t \) is to be affected by the trimming step as in the two-way FE regression; it would not even be a consistent estimator for the latent factor particularly when \( z_{it} \) includes extreme outliers. In this case, we let

\[
\tilde{\pi}_t = \sum_{i=1}^n \pi_i^T z_{it},
\]

where \( \pi_i^T \) now imposes the same trimming scheme as \( w^T_t \) in (8) but still satisfies Assumption 5 of Pesaran (2006). The trimmed CCE estimator \( \hat{\beta}_{CCE,i}^T \) for individual slope parameter \( \beta_i \) is then obtained from the least squares of

\[
y_{it} = a'_i d_t + \delta'_i \tilde{\pi}_t + x'_i \beta_i + u_{it},
\]

which uses \( \tilde{\pi}_t \) instead of \( \pi_t \). From Theorem 1 of Pesaran (2006), we still have

\[
\hat{\beta}_{CCE,i}^T - \beta_i = O_p \left( T^{-1/2} \right) \quad (15)
\]

for each \( i = 1, \ldots, n \), provided \( \sqrt{T}/n \to 0 \) as \( n, T \to \infty \). We define the trimmed CCE mean-
group (TCCEMG) estimator as

$$\hat{\beta}_{TCCEMG} = \frac{1}{n_G} \sum_{i \in G_n} \hat{\beta}_{CCE,i} = \sum_{i=1}^{n} w_i^T \tilde{\beta}_{CCE,i},$$

where $w_i^T$ is the same trimming weight as in (8). Similar to Theorem 1, We derive the limiting distribution of $\hat{\beta}_{TCCEMG}$ as follows.

**Theorem 2** Suppose the random coefficient $\beta_i$ satisfies (2). Also let $n_G \to \infty$ as $n \to \infty$ satisfying $\phi = \lim_{n \to \infty} n/n_G$ and $\phi \in [1, \infty)$. Under the same condition in Theorem 1 of Pesaran (2006), we have $\sqrt{n}(\hat{\beta}_{TCCEMG} - \beta) \to_d N(0, \phi \Omega)$ as $n, T \to \infty$ and $\sqrt{T}/n \to 0$.

**Proof of Theorem 2** Note that

$$\sqrt{n}(\hat{\beta}_{TCCEMG} - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (nw_i^T)(\hat{\beta}_{CCE,i} - \beta_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (nw_i^T)(\beta_i - \beta)$$

$$= O_p \left( \frac{1}{\sqrt{T}} \right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (nw_i^T)(\beta_i - \beta)$$

from Theorem 1 of Pesaran (2006). The result follows immediately since $\beta_i$ satisfies (2), where $w_i^T$ only depends on $x_i$, and $(1/n) \sum_{i=1}^{n} (nw_i^T)^2 = (1/n) \sum_{i=1}^{n} (n1 \{i \in G_n\}/n_G)^2 = n/n_G \to \phi$ as $n \to \infty$. □

We can readily estimate the asymptotic variance of $\hat{\beta}_{TCCEMG}$ in Theorem 2 as the sample covariance of $\hat{\beta}_{CCE,i}$ as in (12):

$$\frac{n}{n_G^2} \sum_{i \in G_n} (\hat{\beta}_{CCE,i} - \hat{\beta}_{TCCEMG})(\hat{\beta}_{CCE,i} - \hat{\beta}_{TCCEMG})'.$$

We now denote an “induced” order statistic $\{\hat{\beta}_{CCE,[i]}^T\}$, where $[i]$ are reordered based on the given trimming scheme. As a special case, this induced order statistic could correspond to the order statistic of $\{\hat{\beta}_{CCE,[i]}^T\}$ itself (e.g., the ranking of $\hat{\beta}_{CCE,i}$ is the same as the ranking of $\hat{\Sigma}_i$ in our case). In such cases, we can define the TCCEMG estimator whose trimming scheme is directly from the order statistic of $\{\hat{\beta}_{CCE,[i]}^T\}$.

For instance, for the scalar $x_{it}$ case ($m = 1$), we can consider the TCCEMG estimator in the

14
form of the sample trimmed mean defined as

\[ \tilde{\beta}_{TCCEMG} = \frac{1}{[np_u] - [np_\ell]} \sum_{i=[np_\ell]+1}^{[np_u]} \tilde{\beta}_{CCE,|i|}, \]

where 0 < \( p_\ell < p_u < 1 \) are some fixed numbers denoting the lower and upper trimmed proportions, and \( [c] \) denotes the largest integer that does not exceed the constant \( c \). Similarly as Stigler (1973), we let

\[ \ell = \sup \{ b : F(b) \leq p_\ell \} \quad \text{and} \quad u = \inf \{ b : F(b) \geq p_u \}, \]

where \( F(\cdot) \) is the cdf of \( \beta_i \). Then, we can obtain the limiting distribution of \( \tilde{\beta}_{TCCEMG} \) as in the following theorem. We let \( F_\tau(\cdot) \) be the cdf of trimmed \( \beta_i \), which is defined as

\[ F_\tau(b) = \begin{cases} 
0 & \text{if } b < \ell \\
F(b)/(p_u - p_\ell) & \text{if } \ell \leq b \leq u \\
1 & \text{if } b > u,
\end{cases} \]

and

\[ \beta_\tau = \int_{-\infty}^{\infty} b F_\tau(db), \quad s_\tau^2 = \int_{-\infty}^{\infty} (b - \beta_\tau)^2 F_\tau(db). \]

**Theorem 3** Suppose the random coefficient \( \beta_i \) satisfies (2) with continuous \( \varepsilon_i \), and the trimming scheme is based on the order statistic of \( \{\beta_i\} \). When \( \beta_i \in \mathbb{R} \), under the same condition in Theorem 1 of Pesaran (2006) and Theorem of Stigler (1973), we have \( \sqrt{n}(\tilde{\beta}_{TCCEMG} - \beta_\tau) \rightarrow_d \mathcal{N}(0, (p_u - p_\ell)^{-2}v_\tau) \) as \( n, T \rightarrow \infty \) and \( \sqrt{T}/n \rightarrow 0 \), where \( v_\tau = (p_u - p_\ell)s_\tau^2 + p_\ell (1 - p_\ell) (\ell - \beta_\tau)^2 + p_u (1 - p_u) (u - \beta_\tau)^2 - 2p_\ell (1 - p_u) (\ell - \beta_\tau) (u - \beta_\tau) \).

**Proof of Theorem 3** Note that

\[
\sqrt{n}(\tilde{\beta}_{TCCEMG} - \beta_\tau) = \frac{\sqrt{n}}{\sqrt{[np_u] - [np_\ell]}} \times \frac{1}{\sqrt{[np_u] - [np_\ell]}} \sum_{i=[np_\ell]+1}^{[np_u]} (\tilde{\beta}_{CCE,|i|} - \beta_{|i|}) \\
\quad + \frac{\sqrt{n}}{[np_u] - [np_\ell]} \sum_{i=[np_\ell]+1}^{[np_u]} (\beta_{|i|} - \beta_\tau) \\
\quad = O_p \left( \frac{1}{\sqrt{T}} \right) + \frac{\sqrt{n}}{[np_u] - [np_\ell]} \sum_{i=[np_\ell]+1}^{[np_u]} (\beta_{|i|} - \beta_\tau)
\]
from Theorem 1 of Pesaran (2006). The result follows from the Theorem of Stigler (1973) with \( A = B = 0 \) in their expression. 

When \( \beta_i \) is symmetrically distributed about the origin and \( p_\ell = 1 - p_u \) with \( 1/2 < p_u < 1 \), we have \( \beta_r = \beta = 0 \). Furthermore, in this case, the asymptotic variance can be simplified as \( \nu_r = (2p_u - 1)^{-1}s_r^2 \).

Given \((p_\ell, p_u)\), since \( \hat{\beta}_{CCE,i}^T \) is consistent to \( \beta_i \) with large \( T \) as in (15), the values \( \ell \) and \( u \) can be obtained as the \( [np_\ell] \)th and the \( [np_u] \)th elements in the ordered statistic \( \{ \beta_{CCE,[i]} \} \), respectively. Therefore, we can estimate the asymptotic variance in Theorem 3 as

\[
\hat{\nu}_r = \frac{1}{p_u - p_\ell} \sum_{i = [np_\ell] + 1}^{[np_u]} \left( \hat{\beta}_{CCE,[i]} - \hat{\beta}_{TCCEMG}^* \right)^2
\]

with \( \hat{\beta}_{TCCEMG}^* \) given in (16).

4 Monte Carlo Simulation

We consider two data generating processes (DGPs). The first DGP is given by

\[
y_{it} = \beta_1 x_{1,it} + \beta_2 x_{2,it} + u_{it},
\]

(17)

where \( x_{j,it} = \rho x_{j,it-1} + \epsilon_{j,it} \) with \( j = 1,2 \) and \( u_{it} = \rho u_{it-1} + v_{it} \). The innovation \( \epsilon_{j,it} \) are generated from \( N(0, \sigma^2_{j,i}) \) where \( \sigma^2_{j,i} \sim \chi^2_1 \) for \( j = 1,2 \). Meanwhile, we generate \( v_{it} \) from the standard normal. We consider two values of \( \rho = 0, 0.8 \), but only report the case with \( \rho = 0 \) here.\(^4\) There is little difference between the two cases except for the absolute magnitude of the mean square error (MSE). The size of tests and relative MSEs are almost identical.

The second DGP includes a common factor, which is given as

\[
y_{it} = a_i F_t + \beta_1 x_{1,it} + \beta_2 x_{2,it} + u_{it}
\]

(18)

and

\[
x_{1,it} = x_{1,it}^0 + \lambda_1 F_t \quad \text{and} \quad x_{2,it} = x_{2,it}^0 + \lambda_2 F_t,
\]

where \( \lambda_{2i} = \lambda_{1i} + \eta_i \) with both \( \eta_i \) and \( \lambda_{1i} \) being generated from \( U[1,2] \). In addition, \( x_{j,it}^0 = \rho x_{j,it-1} + \epsilon_{j,it} \) for \( j = 1,2 \) and \( u_{it} = \rho u_{it-1} + v_{it} \), where \( \rho, \epsilon_{j,it}, \) and \( v_{it} \) are same as in the first DGP above. \( F_t = \rho F_{t-1} + e_t \), where \( e_t \) is generated from the standard normal. For the

\(^4\)When \( \rho = 0.8 \), we discard the first 100 observations to avoid the effect of the initial condition.
homogeneous $\beta_i$ case, we set $\beta_{j,i} = 1$ for all $i$ and $j = 1, 2$. Meanwhile, for the heterogeneous $\beta_i$ case, we set $\beta_{j,i} \sim \mathcal{N}(1, 1)$ for each $j = 1, 2$.

For both DGPs, we also consider the case with outliers by letting the variance of each regressor as $\sigma^2_{j,i} = 25$ and the slope parameter values as $\beta_{j,i} = 5$ for $i = n - 1$, $n$ for both $j = 1, 2$, implying that the last two cross-sectional units are outliers. We consider sample sizes of $n = 50, 100, 200, 500$ and $T = 5, 10, 25, 50, 100$. All simulation results are based on 5000 iterations. Tables are collected at the end of this paper.

Table 1 reports the MSE values for four estimators in the first DGP in (17) under the absence of any outliers: the two-way FE estimator; the MG estimator; the TMG estimator (DTMG) with a joint trimming method based on the Mahalanobis depth of the sample variance of $(x_{1,it}, x_{2,it})'$, and the TMG estimator (XTMG) with a marginal trimming method based on individually trimmed sets using each sample variance of $x_{j,it}$ for $j = 1, 2$. The reported MSE values in Table 1 are the averages over those of $\hat{\beta}_{1,i}$ and $\hat{\beta}_{2,i}$, and they are all multiplied by 100. We trim 20% of $n$: for the depth-based trimming, it drops the 20% of the cross-section samples with the smallest depth; for the marginal trimming, it drops 10% of the cross-section samples from the bottom and the top respectively. Since the DGP does not have outliers in this case, we want to see whether or not the 20% of trimming leads to noticeable efficiency loss in finite samples. The first four columns show the case of homogenous coefficients, and the next four columns report the case of heterogeneous coefficients. Evidently, for all cases of $n$ and $T$, the FE estimator produces the minimum MSE since we purposely design the DGP in this way.\footnote{If we set $\sigma^2_{j,i} = 1$ for all $i$ and $j$, but generate $v_{it}$ with widely heterogeneous variances, then the FE estimator is no longer efficient.} As there are no outliers, the MSE of the DTMG estimator is generally larger than that of the MG estimator for all cases. The variances of all $x_{j,it-1}$ are generated from the $\chi^2_1$ distribution so that only individuals in the right side of the distribution are excluded from the Mahalanobis-depth-based trimming, which leads to inefficient estimation under the absence of any outliers. Meanwhile, the MSE of the XTMG is smaller than that of the MG estimator since the XTMG trims out individuals both in left and right tails.

Table 2 shows the average size of the $t$-test of each estimator in the first DGP in (17). For each $j = 1, 2$, we construct the $t$-ratio for the null hypothesis of $H_0 : \mathbb{E}[\beta_{j,i}] = 1$ and take the average of the rejection rates over $j = 1, 2$. The nominal size is 5%. With a small $n$ like $n = 50$, the FE estimator shows a mild upward size distortion in the case of homogenous coefficients. As $n$ increases, the rejection frequencies with the FE estimator approaches the nominal size very quickly. Meanwhile, with heterogeneous coefficients, the rejection rate with the FE estimator is slightly higher than that with homogenous coefficients. However, the difference between the two reduces quickly as $T$ increases. The MG estimator suffers little size distortion except for small $n$. Nonetheless, all trimmed estimators perform better compared to the MG estimator.

Table 3 provides the MSE of each estimator in the first DGP in (17) under the presence of two
outliers on the right tail of the distribution as described above. Since only the last two individuals are outliers, the FE estimator becomes biased in finite \( n \). But as \( n \to \infty \), the bias approaches zero. More importantly, the FE estimator becomes more biased than the MG estimator since the FE estimator assigns higher weights on the last two individuals as \( \sigma^2_{j,n-1} = \sigma^2_{j,n} = 25 \). On the other hand, the DTMG and XTMG estimators exclude outlying individuals whose regressors’ sample variances are relatively larger than the rest of the individuals in this case.

Table 4 reports the size of the \( t \)-test of each estimator in the first DGP in (17) under the presence of two outliers. Evidently the FE and MG estimators suffer from serious size distortion when \( n \) is small. As \( n \) increases, however, the influence of the outliers becomes localized, so that the sizes of both estimators become milder. Meanwhile, the sizes of both trimmed estimators DTMG and XTMG are about the nominal size. Only when \( n \) is small, both trimmed estimators show mild size distortions, but as \( n \) increases, these size distortions disappear very quickly.

Tables 5 and 6 show the MSEs and sizes of the \( t \)-test of the four CCE estimators in the second DGP in (18). The first two estimators are the pooled CCE (pool) and CCEMG (MG), which do not trim out any individuals. The last two estimators are trimmed CCEMG estimators, whose trimming schemes are the same as those of DTMG and XTMG, respectively. The overall results in Tables 5 and 6 are quite similar to those in Tables 3 and 4.

5 Empirical Illustration: Effect of Police on Crime

As an empirical illustration, we consider the effect of police on crime using the following two-way FE regression:

\[
\Delta \ln C_{it} = a_i + F_t + \beta_1 \Delta \ln P_{it-1} + \beta_2 \Delta \ln U_{it-1} + \beta_3 \Delta \ln B_{it-1} + u_{it}, \tag{19}
\]

where \( C_{it} \) is the number of reported property crimes per capita, \( P_{it} \) is the number of police officers, \( U_{it} \) is the unemployment rate, and \( B_{it} \) is the percentage of black population in state \( i \) and year \( t \). This is similar to Levitt (1997) but we exclude other control variables (i.e., public welfare spending, percentage of female-headed households, and percentage of ages between 15 and 24 years old) because of the limited data availability. We include the pre-determined \( \Delta \ln P_{it-1} \) to minimize any simultaneity. We also take first-difference for all variables because they are either \( I(1) \) or near \( I(1) \) processes but do not impose cointegrating relations.

The annual property crimes and the number of police officers across 48 contiguous states from years 1970 to 2013 are collected from the FBI Uniform Crime Reports. Unemployment rates and the percentage of black population are collected from the Bureau of Economic Analysis and the Census Bureau, respectively. This regression was also used by Han, Kwak and Sul (2019) for violent crime; they examine whether \( F_t \) or \( \chi_i F_t \) should be in (19) and report that the two-way FE
estimation is good enough. Sul (2019) shows that the property crime rates across 48 contiguous U.S. states have a single common factor with a homogeneous factor loading, hence including the time effect $F_t$ is sufficient. $\beta_1$ is the main interest and it describes the average marginal effect from idiosyncratic increases in the sworn officers on idiosyncratic growth rates of the property crime rates, after controlling for the common dynamics $F_t$.

Table 7 reports the estimation results of the two-way FE estimation and the CCE estimation. The numbers in the parentheses are $t$-ratios, which are constructed using a panel robust covariance estimator. In the first two columns, the FE estimate of $\beta_1$ is positive though not significantly different from zero. Meanwhile, the property crime rates are decreasing as unemployment rates increase, but this relationship is not statistically significant. The percentage of black population influences negatively on the property crime rates and it is statistically significant, which is a puzzling finding. Furthermore, the MG estimation gives very similar results.

We next consider the TMG results based three different trimming ratios, 20%, 10% and 5%. As in the previous section, for the depth-based trimming, we drop 20%, 10%, and 5% of the cross-section samples with the smallest depth, respectively; for the marginal trimming, we drop 10%, 5%, and 2.5% of the cross-section samples from each side of the tail of the distribution, respectively. We first test for the independence between $(\beta_{1,t}, \beta_{2,t}, \beta_{3,t})$ and the variance of the regressors using the test proposed by Sul (2016) and Campello, Galvao, and Juhl (2019), which is asymptotically distributed as $\chi^2_3$. The test statistic is 6.978, which is smaller than the 5% critical value of 7.81. Therefore, the time series variance of each regressor can be used as a trimming instrument in this illustration. All the TMG estimates of $\beta_1$ are not significantly different from zero regardless of the trimming fractions, though they are negative, which implies that an exogenous increase in the number of officers does not reduce property crime rates if other things are equal. Similarly, all the TMG estimates of $\beta_2$ are negative and not significantly different from zero, even though the point estimates are slightly different depending on the choice of trimming instruments. Lastly, the TMG estimates of $\beta_3$ are not significantly different from zero at the 5% level regardless of the trimming methods and threshold values. However, for the two-way FE case, we find that some TMG estimates are significantly different from zero at the 10% level. This result shows a weak evidence of the effect of police on property crime and the TMG estimation method yields a robust finding even in a simple regression form.

To better understand how the trimming affects the estimation results, we plot the relation between the variance of each regressor and its corresponding slope parameter estimates in Figures 5, 6, and 7. They are based on the two-way FE estimation; the empty squared ones are outliers based on the marginal variance of the regressor in XTMG estimation and the empty circled ones are outliers identified by the Mahalanobis depth of the variances in DTMG estimation, both based on 20% trimming. We find that Delaware and Wyoming show extremely large

---

6Both test are for the null hypothesis of $cov(\beta_1, \Sigma) = 0$. They require strict exogeneity, which holds in our example.
Table 7: Determinants of property crime

**Two-way FE Estimation**

<table>
<thead>
<tr>
<th></th>
<th>FE</th>
<th>MG</th>
<th>DTMG</th>
<th>XTMG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td>10%</td>
<td>5%</td>
<td>20%</td>
</tr>
<tr>
<td>( \Delta \ln P_{i,t-1} )</td>
<td>0.027</td>
<td>0.014</td>
<td>-0.018</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(1.10)</td>
<td>(0.39)</td>
<td>(-0.66)</td>
<td>(-0.10)</td>
</tr>
<tr>
<td>( \Delta \ln U_{i,t-1} )</td>
<td>-0.017</td>
<td>-0.014</td>
<td>-0.005</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(-1.52)</td>
<td>(-1.31)</td>
<td>(-0.90)</td>
<td>(-0.29)</td>
</tr>
<tr>
<td>( \Delta \ln B_{i,t-1} )</td>
<td>-0.489**</td>
<td>-0.269*</td>
<td>-0.167</td>
<td>-0.190</td>
</tr>
<tr>
<td></td>
<td>(-3.96)</td>
<td>(-1.80)</td>
<td>(-1.22)</td>
<td>(-1.51)</td>
</tr>
</tbody>
</table>

**CCE Estimation**

<table>
<thead>
<tr>
<th></th>
<th>pool</th>
<th>MG</th>
<th>DTMG</th>
<th>XTMG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td>10%</td>
<td>5%</td>
<td>20%</td>
</tr>
<tr>
<td>( \Delta \ln P_{i,t-1} )</td>
<td>0.019</td>
<td>0.001</td>
<td>-0.011</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.02)</td>
<td>(-0.25)</td>
<td>(-0.10)</td>
</tr>
<tr>
<td>( \Delta \ln U_{i,t-1} )</td>
<td>-0.015</td>
<td>-0.014</td>
<td>-0.012</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(-1.21)</td>
<td>(-1.22)</td>
<td>(-1.04)</td>
<td>(-1.04)</td>
</tr>
<tr>
<td>( \Delta \ln B_{i,t-1} )</td>
<td>-0.593**</td>
<td>-0.315</td>
<td>-0.224</td>
<td>-0.269</td>
</tr>
<tr>
<td></td>
<td>(-4.36)</td>
<td>(-1.39)</td>
<td>(-0.85)</td>
<td>(-1.12)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are t-ratio using a panel robust covariance estimator; * and ** are significant at 10% and 5%, respectively. For Two-way FE, “FE” is the fixed effect, “MG” is the mean group, “DTMG” is the Mahalanobis-depth-based trimmed MG, and “XTMG” is the marginally trimmed MG estimators. For CCE, “pool” is the pooled CCE, “MG” is the CCEMG, “DTMG” is the Mahalanobis-depth-based trimmed CCEMG, and “XTMG” is the marginally trimmed CCEMG estimators. % for DTMG and XTMG stands for the trimming fraction.

variances of \( \Delta \ln P_{it-1} \) and \( \Delta \ln B_{it-1} \), respectively; New Hampshire and Wyoming show very large variances for most of the cases and hence trimmed; California and Connecticut show very small variances for most of the cases and hence also trimmed. Note that the Mahalanobis depth considers variances of three regressors simultaneously, thus joint trimming based on the Mahalanobis depth can drop states whose marginal variances are not extreme. However, both joint and marginal trimmings overlap most of the cases, especially for extreme outliers. Note that trimming based on the variance of \( \Delta \ln B \) appears to have the most impact on the \( \mathbb{E}[\beta_i] \) estimate. This is because most of the trimmed states’ point estimates under this scheme are negative as shown in Figure 7, whereas they are spread symmetrically about zero for the other cases as shown in Figures 5 and 6.
Figure 5: Relationship between $\hat{\beta}_{1,i}$ and $\text{var} (\Delta \ln P_{it-1})$

Figure 6: Relationship between $\hat{\beta}_{2,i}$ and $\text{var} (\Delta \ln U_{it-1})$


6 Concluding Remarks

This paper shows the importance of individual-specific time series estimation and a way to average them for robust panel data analysis. Since pooled estimators, including the two-way FE or pooled CCE, assign heavier weights on the individuals as the corresponding regressor’s variance gets larger, they are sensitive to outlying observations in the sample variance of the regressor. The MG estimators, without considering such outliers, may not be fully robust either. To obtain more robust estimators without much sacrifice of efficiency, this paper proposes a trimmed MG estimator, where we trim individual observations of which the sample variances of regressors are outlying (i.e., either extremely small or large).

Though this paper focuses on static panel regression, the idea can be extended to dynamic regression. In addition, we suppose the trimming thresholds are given in this paper, but we could pick the thresholds by optimizing some objective function such as the higher-order MSE of the TMG estimator. Finally, an endogenous trimming directly based on the order statistic of $\hat{\beta}_i$ could be more desirable though its asymptotic analysis is not straightforward. We leave these important topics for future challenges.
References


Han, M., J.H. Kwak, and D. Sul, 2019, Two-way fixed effects versus panel factor augmented estimators: Asymptotic comparison among pre-testing procedures, *mimeo*, University of Texas at Dallas.


Sul, D. 2016, Pooling is sometimes harmful, *mimeo*, University of Texas at Dallas.


Table 1: MSE comparisons under the absence of any outliers

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>$\beta_{1,i} = 1 &amp; \beta_{2,i} = 1$ for all $i$</th>
<th>$\beta_{1,i} \sim N(1,1) &amp; \beta_{2,i} \sim N(1,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>MG</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.560</td>
<td>15.75</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>0.263</td>
<td>11.79</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>0.052</td>
<td>5.230</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.246</td>
<td>2.440</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.115</td>
<td>1.833</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>0.058</td>
<td>1.322</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>0.022</td>
<td>0.870</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.090</td>
<td>0.709</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.044</td>
<td>0.510</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.021</td>
<td>0.381</td>
</tr>
<tr>
<td>500</td>
<td>25</td>
<td>0.008</td>
<td>0.253</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.042</td>
<td>0.309</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.021</td>
<td>0.231</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.010</td>
<td>0.174</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>0.004</td>
<td>0.115</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.022</td>
<td>0.153</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.011</td>
<td>0.113</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.005</td>
<td>0.081</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>0.002</td>
<td>0.055</td>
</tr>
</tbody>
</table>

Note: “FE” is the fixed effect, “MG” is the mean group, “DTMG” is the Mahalanobis-depth-based trimmed MG, and “XTMG” is the marginally trimmed MG estimators.
Table 2: Sizes of tests under the absence of any outliers
(Nominal size: 5%)

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>FE</th>
<th>MG</th>
<th>DTMG</th>
<th>XTMG</th>
<th>FE</th>
<th>MG</th>
<th>DTMG</th>
<th>XTMG</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>0.084</td>
<td>0.044</td>
<td>0.045</td>
<td>0.047</td>
<td>0.120</td>
<td>0.062</td>
<td>0.057</td>
<td>0.058</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>0.066</td>
<td>0.044</td>
<td>0.041</td>
<td>0.047</td>
<td>0.084</td>
<td>0.056</td>
<td>0.052</td>
<td>0.057</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>0.064</td>
<td>0.047</td>
<td>0.045</td>
<td>0.049</td>
<td>0.070</td>
<td>0.051</td>
<td>0.048</td>
<td>0.051</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>0.054</td>
<td>0.043</td>
<td>0.046</td>
<td>0.047</td>
<td>0.058</td>
<td>0.050</td>
<td>0.048</td>
<td>0.050</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.081</td>
<td>0.042</td>
<td>0.045</td>
<td>0.050</td>
<td>0.106</td>
<td>0.088</td>
<td>0.079</td>
<td>0.071</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.064</td>
<td>0.045</td>
<td>0.045</td>
<td>0.052</td>
<td>0.077</td>
<td>0.071</td>
<td>0.061</td>
<td>0.057</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>0.058</td>
<td>0.045</td>
<td>0.046</td>
<td>0.050</td>
<td>0.064</td>
<td>0.061</td>
<td>0.054</td>
<td>0.057</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>0.052</td>
<td>0.045</td>
<td>0.047</td>
<td>0.047</td>
<td>0.056</td>
<td>0.055</td>
<td>0.050</td>
<td>0.051</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.077</td>
<td>0.044</td>
<td>0.046</td>
<td>0.051</td>
<td>0.098</td>
<td>0.100</td>
<td>0.087</td>
<td>0.080</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.062</td>
<td>0.043</td>
<td>0.042</td>
<td>0.048</td>
<td>0.077</td>
<td>0.082</td>
<td>0.077</td>
<td>0.066</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.056</td>
<td>0.042</td>
<td>0.043</td>
<td>0.051</td>
<td>0.063</td>
<td>0.071</td>
<td>0.064</td>
<td>0.058</td>
</tr>
<tr>
<td>500</td>
<td>25</td>
<td>0.051</td>
<td>0.045</td>
<td>0.045</td>
<td>0.051</td>
<td>0.060</td>
<td>0.060</td>
<td>0.059</td>
<td>0.054</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.074</td>
<td>0.042</td>
<td>0.043</td>
<td>0.049</td>
<td>0.097</td>
<td>0.106</td>
<td>0.096</td>
<td>0.085</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.061</td>
<td>0.041</td>
<td>0.039</td>
<td>0.048</td>
<td>0.079</td>
<td>0.087</td>
<td>0.083</td>
<td>0.064</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.058</td>
<td>0.043</td>
<td>0.042</td>
<td>0.048</td>
<td>0.063</td>
<td>0.070</td>
<td>0.069</td>
<td>0.058</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>0.055</td>
<td>0.048</td>
<td>0.046</td>
<td>0.052</td>
<td>0.055</td>
<td>0.061</td>
<td>0.056</td>
<td>0.053</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.075</td>
<td>0.042</td>
<td>0.040</td>
<td>0.052</td>
<td>0.094</td>
<td>0.114</td>
<td>0.111</td>
<td>0.092</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.068</td>
<td>0.041</td>
<td>0.042</td>
<td>0.048</td>
<td>0.077</td>
<td>0.094</td>
<td>0.089</td>
<td>0.070</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.057</td>
<td>0.041</td>
<td>0.045</td>
<td>0.050</td>
<td>0.062</td>
<td>0.071</td>
<td>0.067</td>
<td>0.059</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>0.054</td>
<td>0.045</td>
<td>0.046</td>
<td>0.047</td>
<td>0.058</td>
<td>0.066</td>
<td>0.060</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Note: “FE” is the fixed effect, “MG” is the mean group, “DTMG” is the Mahalanobis-depth-based trimmed MG, and “XTMG” is the marginally trimmed MG estimators.
Table 3: MSE comparisons under the presence of two outliers

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>( \beta_{j,i} = 1 ) for all ( i ) except ( i = n - 1, n )</th>
<th>( \beta_{j,i} \sim \mathcal{N}(1, 1) ) except ( i = n - 1, n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>\begin{tabular}{c</td>
<td>c</td>
</tr>
<tr>
<td>FE</td>
<td>MG</td>
<td>DTMG</td>
<td>XTMG</td>
</tr>
<tr>
<td>-----</td>
<td>----</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>64.73</td>
<td>23.98</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>21.87</td>
<td>17.71</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>10.28</td>
<td>13.90</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>5.465</td>
<td>8.865</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>43.23</td>
<td>3.577</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>10.58</td>
<td>2.806</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>3.022</td>
<td>2.063</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>1.078</td>
<td>1.421</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>38.30</td>
<td>0.988</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>8.334</td>
<td>0.762</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>1.861</td>
<td>0.555</td>
</tr>
<tr>
<td>500</td>
<td>25</td>
<td>0.403</td>
<td>0.384</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>36.53</td>
<td>0.439</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>7.744</td>
<td>0.330</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>1.592</td>
<td>0.260</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>0.256</td>
<td>0.172</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>36.08</td>
<td>0.207</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>7.535</td>
<td>0.163</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>1.459</td>
<td>0.120</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>0.196</td>
<td>0.084</td>
</tr>
</tbody>
</table>

Note: “FE” is the fixed effect, “MG” is the mean group, “DTMG” is the Mahalanobis-depth-based trimmed MG, and “XTMG” is the marginally trimmed MG estimators.
Table 4: Sizes of tests under the presence of two outliers
(Nominal size: 5%)

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>FE</th>
<th>MG</th>
<th>DTMG</th>
<th>XTMG</th>
<th>FE</th>
<th>MG</th>
<th>DTMG</th>
<th>XTMG</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5</td>
<td>0.709</td>
<td>0.434</td>
<td>0.045</td>
<td>0.051</td>
<td>0.689</td>
<td>0.395</td>
<td>0.059</td>
<td>0.060</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>0.387</td>
<td>0.223</td>
<td>0.044</td>
<td>0.047</td>
<td>0.417</td>
<td>0.210</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>0.091</td>
<td>0.107</td>
<td>0.048</td>
<td>0.047</td>
<td>0.166</td>
<td>0.107</td>
<td>0.046</td>
<td>0.052</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>0.001</td>
<td>0.062</td>
<td>0.040</td>
<td>0.048</td>
<td>0.063</td>
<td>0.062</td>
<td>0.049</td>
<td>0.052</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.867</td>
<td>0.778</td>
<td>0.043</td>
<td>0.053</td>
<td>0.817</td>
<td>0.669</td>
<td>0.078</td>
<td>0.071</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.491</td>
<td>0.493</td>
<td>0.042</td>
<td>0.051</td>
<td>0.501</td>
<td>0.386</td>
<td>0.062</td>
<td>0.058</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>0.053</td>
<td>0.214</td>
<td>0.044</td>
<td>0.050</td>
<td>0.174</td>
<td>0.182</td>
<td>0.059</td>
<td>0.054</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>0.000</td>
<td>0.079</td>
<td>0.045</td>
<td>0.050</td>
<td>0.060</td>
<td>0.079</td>
<td>0.053</td>
<td>0.050</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.979</td>
<td>0.970</td>
<td>0.045</td>
<td>0.050</td>
<td>0.919</td>
<td>0.830</td>
<td>0.095</td>
<td>0.081</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.638</td>
<td>0.797</td>
<td>0.046</td>
<td>0.052</td>
<td>0.572</td>
<td>0.528</td>
<td>0.075</td>
<td>0.062</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.015</td>
<td>0.416</td>
<td>0.046</td>
<td>0.046</td>
<td>0.180</td>
<td>0.256</td>
<td>0.064</td>
<td>0.058</td>
</tr>
<tr>
<td>500</td>
<td>25</td>
<td>0.000</td>
<td>0.116</td>
<td>0.042</td>
<td>0.049</td>
<td>0.062</td>
<td>0.097</td>
<td>0.059</td>
<td>0.052</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.997</td>
<td>0.998</td>
<td>0.042</td>
<td>0.049</td>
<td>0.954</td>
<td>0.890</td>
<td>0.106</td>
<td>0.088</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.745</td>
<td>0.950</td>
<td>0.041</td>
<td>0.049</td>
<td>0.615</td>
<td>0.605</td>
<td>0.079</td>
<td>0.067</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.004</td>
<td>0.637</td>
<td>0.043</td>
<td>0.048</td>
<td>0.185</td>
<td>0.304</td>
<td>0.069</td>
<td>0.059</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>0.000</td>
<td>0.180</td>
<td>0.046</td>
<td>0.050</td>
<td>0.073</td>
<td>0.115</td>
<td>0.057</td>
<td>0.050</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
<td>0.045</td>
<td>0.048</td>
<td>0.965</td>
<td>0.911</td>
<td>0.112</td>
<td>0.091</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.829</td>
<td>0.996</td>
<td>0.043</td>
<td>0.045</td>
<td>0.631</td>
<td>0.641</td>
<td>0.084</td>
<td>0.070</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.001</td>
<td>0.848</td>
<td>0.042</td>
<td>0.046</td>
<td>0.183</td>
<td>0.330</td>
<td>0.073</td>
<td>0.060</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>0.000</td>
<td>0.285</td>
<td>0.046</td>
<td>0.051</td>
<td>0.069</td>
<td>0.127</td>
<td>0.063</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Note: “FE” is the fixed effect, “MG” is the mean group, “DTMG” is the Mahalanobis-depth-based trimmed MG, and “XTMG” is the marginally trimmed MG estimators.
Table 5: MSE comparisons among CCEs under the presence of two outliers

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>( \beta_{j:i} = 1 ) for all ( i = n - 1, n )</th>
<th>( \beta_{j:i} \sim \mathcal{N}(1,1) ) for all ( i = n - 1, n )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \sigma^2_{j:n-1} = \sigma^2_{j,n} = 25 ); ( \beta_{j,n-1} = \beta_{j,n} = 5 )</td>
<td>( \sigma^2_{j,n-1} = \sigma^2_{j,n} = 25 ); ( \beta_{j,n-1} = \beta_{j,n} = 5 )</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>217.3</td>
<td>14.67</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>214.1</td>
<td>7.185</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>214.6</td>
<td>6.178</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>214.4</td>
<td>5.894</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>86.52</td>
<td>8.189</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>81.00</td>
<td>2.891</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>79.74</td>
<td>2.275</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>79.05</td>
<td>2.044</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>27.70</td>
<td>1.360</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>27.08</td>
<td>0.946</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>26.60</td>
<td>0.760</td>
</tr>
<tr>
<td>500</td>
<td>25</td>
<td>6.423</td>
<td>0.612</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>6.207</td>
<td>0.332</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>6.093</td>
<td>0.231</td>
</tr>
</tbody>
</table>

Note: “pool” is the pooled CCE, “MG” is the CCEMG, “DTMG” is the Mahalanobis-depth-based trimmed CCEMG, and “XTMG” is the marginally trimmed CCEMG estimators.
Table 6: Sizes of CCE tests under the presence of two outliers
(Nominal size: 5%)

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>pool</th>
<th>MG</th>
<th>DTMG</th>
<th>XTMG</th>
<th>pool</th>
<th>MG</th>
<th>DTMG</th>
<th>XTMG</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>10</td>
<td>0.502</td>
<td>0.124</td>
<td>0.043</td>
<td>0.050</td>
<td>0.524</td>
<td>0.113</td>
<td>0.054</td>
<td>0.061</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.618</td>
<td>0.256</td>
<td>0.044</td>
<td>0.041</td>
<td>0.611</td>
<td>0.168</td>
<td>0.058</td>
<td>0.061</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.703</td>
<td>0.357</td>
<td>0.042</td>
<td>0.043</td>
<td>0.655</td>
<td>0.197</td>
<td>0.059</td>
<td>0.058</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.759</td>
<td>0.457</td>
<td>0.043</td>
<td>0.044</td>
<td>0.662</td>
<td>0.192</td>
<td>0.063</td>
<td>0.065</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.178</td>
<td>0.087</td>
<td>0.042</td>
<td>0.044</td>
<td>0.281</td>
<td>0.080</td>
<td>0.048</td>
<td>0.052</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.110</td>
<td>0.203</td>
<td>0.041</td>
<td>0.042</td>
<td>0.291</td>
<td>0.126</td>
<td>0.052</td>
<td>0.056</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.067</td>
<td>0.324</td>
<td>0.042</td>
<td>0.042</td>
<td>0.287</td>
<td>0.154</td>
<td>0.058</td>
<td>0.059</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0.036</td>
<td>0.459</td>
<td>0.043</td>
<td>0.042</td>
<td>0.283</td>
<td>0.166</td>
<td>0.058</td>
<td>0.059</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>0.015</td>
<td>0.066</td>
<td>0.045</td>
<td>0.045</td>
<td>0.120</td>
<td>0.066</td>
<td>0.048</td>
<td>0.048</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.000</td>
<td>0.139</td>
<td>0.046</td>
<td>0.043</td>
<td>0.128</td>
<td>0.098</td>
<td>0.051</td>
<td>0.052</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.000</td>
<td>0.240</td>
<td>0.049</td>
<td>0.046</td>
<td>0.123</td>
<td>0.122</td>
<td>0.056</td>
<td>0.060</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.000</td>
<td>0.375</td>
<td>0.042</td>
<td>0.045</td>
<td>0.129</td>
<td>0.133</td>
<td>0.055</td>
<td>0.056</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>0.002</td>
<td>0.053</td>
<td>0.051</td>
<td>0.044</td>
<td>0.065</td>
<td>0.051</td>
<td>0.047</td>
<td>0.048</td>
</tr>
<tr>
<td>500</td>
<td>25</td>
<td>0.000</td>
<td>0.084</td>
<td>0.044</td>
<td>0.049</td>
<td>0.081</td>
<td>0.070</td>
<td>0.052</td>
<td>0.049</td>
</tr>
<tr>
<td>500</td>
<td>50</td>
<td>0.000</td>
<td>0.125</td>
<td>0.048</td>
<td>0.047</td>
<td>0.079</td>
<td>0.083</td>
<td>0.049</td>
<td>0.053</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>0.000</td>
<td>0.209</td>
<td>0.045</td>
<td>0.046</td>
<td>0.079</td>
<td>0.097</td>
<td>0.057</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Note: “pool” is the pooled CCE, “MG” is the CCEMG, “DTMG” is the Mahalanobis-depth-based trimmed CCEMG, and “XTMG” is the marginally trimmed CCEMG estimators.