Trimmed Mean Group Estimation

Yoonseok Lee*
Syracuse University

Donggyu Sul†
University of Texas at Dallas

December 2020

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.

Keywords: Trimmed mean group estimator, Robust estimator, Heterogeneous panel, Random coefficient, Two-way fixed effects, Common correlated effects
JEL Classifications: C23, C33

* *Address*: Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244. *E-mail*: `ylee41@maxwell.syr.edu`

† *Address*: Department of Economics, University of Texas at Dallas, 800 W. Campbell Road, Richardson, TX 75080. *E-mail*: `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

$$y_{it} = a_i + F_t + x_{it}'\beta_i + u_{it} \tag{1}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where 'i' stands for the ith individual and 't' stand for the tth 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 $\lambda_i' F_t$ can be considered. The regression coefficient β_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 β_i is heterogeneous, we suppose that

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

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

$$\widehat{\beta} = \sum_{i=1}^{n} w_i \widehat{\beta}_i \tag{3}$$

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. $\widehat{\beta}_i$ is typically the least squares estimator of β_i for each i:

$$\widehat{\beta}_i = \widehat{\Sigma}_i^{-1} \frac{1}{T} \sum_{t=1}^{T} \dot{x}_{it} \dot{y}_{it},$$

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

$$\widehat{\Sigma}_i = \frac{1}{T} \sum_{t=1}^{T} \dot{x}_{it} \dot{x}_{it}'$$

with

$$\dot{x}_{it} = \tilde{x}_{it} - \frac{1}{n} \sum_{j=1}^{n} \tilde{x}_{jt} \quad \text{and} \quad \tilde{x}_{it} = x_{it} - \frac{1}{T} \sum_{s=1}^{T} x_{is}. \tag{4}$$

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

$$w_{MG,i} = \frac{1}{n} \tag{5}$$

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

$$w_{SB,i} = \left(\sum_{j=1}^{n} \left\{ \widehat{\sigma}_{uj}^2 \widehat{\Sigma}_j^{-1}/T + \widehat{\Omega} \right\}^{-1} \right)^{-1} \left\{ \widehat{\sigma}_{ui}^2 \widehat{\Sigma}_i^{-1}/T + \widehat{\Omega} \right\}^{-1} \tag{6}$$

in (3) using some consistent estimators $\widehat{\sigma}_{ui}^2$ and $\widehat{\Omega}$ 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

$$\widehat{\beta}_{FE} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \dot{x}_{it} \dot{x}_{it}' \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \dot{x}_{it} \dot{y}_{it}$$

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

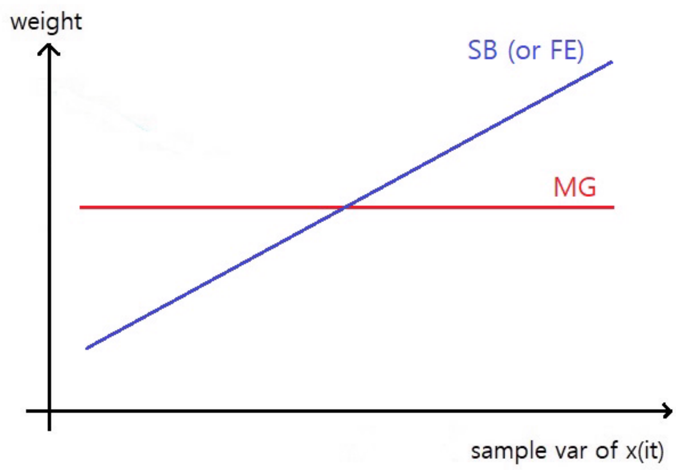

Figure 1: Comparison of the weights

weights

$$w_{FE,i} = \left(\sum_{j=1}^{n} \widehat{\Sigma}_j \right)^{-1} \widehat{\Sigma}_i, \tag{7}$$

which is proportional to the inverse of $var(\widehat{\beta}_i)$ for each i under homogeneity (e.g., Sul, 2016). In this case, this weight is optimal in the sense that $\widehat{\beta}_{FE}$ is the best linear unbiased estimator of β.

2.2 Trimming based on the variances of regressors

Comparing the weights (5), (6), and (7), we can see that $\widehat{\beta}_{MG}$ puts equal weights for all $\widehat{\beta}_i$; but both $\widehat{\beta}_{SB}$ and $\widehat{\beta}_{FE}$ put unequal weights over $\widehat{\beta}_i$, where the weights are proportional to $\widehat{\Sigma}_i$, the sample variance of x_{it} of each i. Note that these weights are basically proportional to the inverse of each individual variance of $\widehat{\beta}_i$, and hence $\widehat{\beta}_{SB}$ and $\widehat{\beta}_{FE}$ are to be more efficient than $\widehat{\beta}_{MG}$. 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}, $\widehat{\Sigma}_i$, when β_i is a scalar.

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

In fact, it is a common practice in panel data analysis that ith observations with large time series variation of x_{it} are considered as outlying individuals and dropped to get more

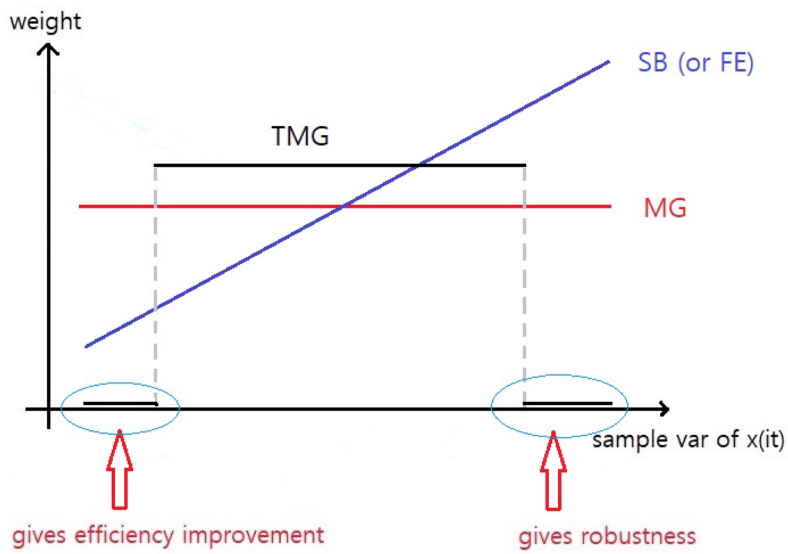

Figure 2: Weight of the trimmed MG estimator

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 $\widehat{\Sigma}_i$ are *extremely large*. In particular, we take an equally weighted average over $\widehat{\beta}_i$'s in (3), but the weights take a hard-threshold trimming based on the sample variance of x_{it} of each i, $\widehat{\Sigma}_i$. 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 ith observations whose sample variances of x_{it} are *extremely small* (i.e., $\widehat{\beta}_i$'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 $\widehat{\Sigma}_i$.

In sum, we consider the double-sided trimming scheme that drops ith 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:

$$\Delta s_{it} = a_i + \lambda_i' F_t + \beta_i \left(\pi_{it} - \pi_t^* \right) + u_{it},$$

where s_{it} and π_{it} are the foreign exchange rate and the inflation rate of the ith foreign country. π_t^* stands for the domestic (U.S. in this example) inflation rate. As Greenaway-McGrevy, 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 $\widehat{\beta}_i$ ($n = 27$) 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 β_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 $\widehat{\beta}_{FE}$ assigns a huge weight on the point estimate of Turkey, which leads $\widehat{\beta}_{FE}$ 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-McGrevy, Han, and Sul (2012).

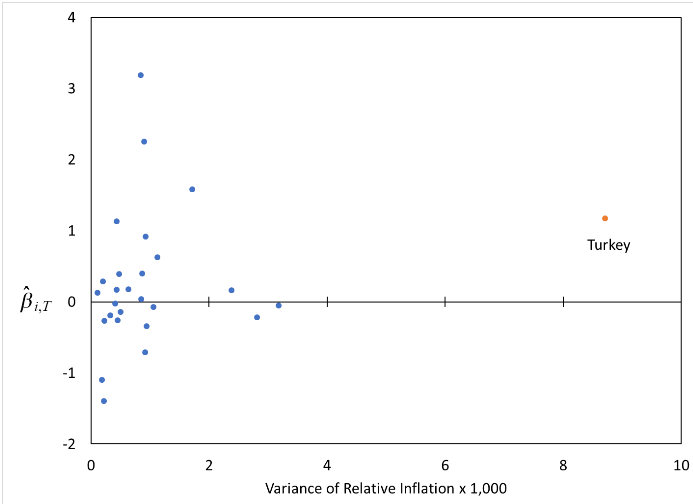

Figure 3: PPP estimates versus variances of relative inflation

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:

$$\Delta \ln E_{it} = a_i + F_t + \beta_i \Delta \ln Y_{it} + u_{it},$$

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

Figure 4 shows the MPC estimates $\widehat{\beta}_i$ 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 $\widehat{\beta}_i$ of which the variances of the regressor are near zero, we can obtain a more efficient estimator of the mean of individual MPC, say $\mathbb{E}[\beta_i]$ in this example.

2.4 Trimming weights

We can construct the trimming weights based on the sample covariance of the regressors, $\widehat{\Sigma}_i$, in practice. When x_{it} is univariate (i.e., $m = 1$), we can simply sort $\widehat{\Sigma}_i$ and decide the trimming

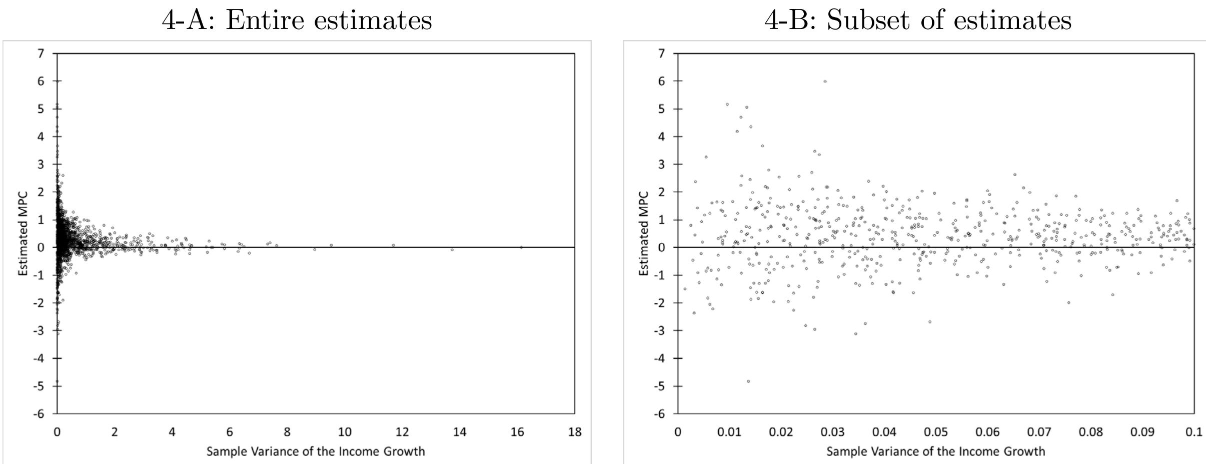

Figure 4: MPC estimates versus variances of income growth

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 jth regression, we only report the estimates of the jth element of the TMG estimator that corresponds to the slope coefficient of the jth 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 $\widehat{\sigma}_{j,i}^2$ be the sample variance of the jth element of the agent i's regressors for $j = 1, \ldots, m$. We conduct marginal trimming using $\widehat{\sigma}_{j,i}^2$ 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 $\widehat{\Sigma}_i$ (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 $\widehat{v}_i = (\widehat{\sigma}_{1,i}^2, \widehat{\sigma}_{2,i}^2, \ldots, \widehat{\sigma}_{m,i}^2)'$ defined as

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

where $\widehat{\mu}_v$ and $\widehat{\Lambda}_v$ are the sample mean and sample variance of $\widehat{v}_i$. By construction, $\widehat{D}\,(v) \in [0, 1]$ and it is close to zero if $v = \widehat{v}_i$ is either extremely small or large (i.e., for outlying $\widehat{v}_i$ 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 ith observations with $\widehat{D}(\widehat{v}_i) < \underline{d}$ for some threshold $0 < \underline{d} < 1$. 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 $\widehat{\Sigma}_i$ in the trimming methods, there are cases when we need to consider the off-diagonal terms of $\widehat{\Sigma}_i$, 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 ith individuals when their $\widehat{\Sigma}_i$ 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 $(1/n)\sum_{i=1}^{n} x_{it}$ should be modified by the trimming step because we will drop some i observations. In particular, we subtract the trimmed mean $(1/n_{\mathcal{G}})\sum_{i\in\mathcal{G}_n} x_{it}$, where $\mathcal{G}_n$ is the set of individuals that are not trimmed in the sample and $n_{\mathcal{G}}$ is the cardinality of $\mathcal{G}_n$. We define the Wallace-Hussain transformed x_{it} as

$$\dot{x}_{it}^{\tau} = \tilde{x}_{it} - \frac{1}{n_{\mathcal{G}}}\sum_{j\in\mathcal{G}_n} \tilde{x}_{jt},$$

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 $\widehat{\beta}_i^{\tau}$ by regressing $\dot{y}_{it}^{\tau}$ on $\dot{x}_{it}^{\tau}$. Then, we define the TMG estimator as

$$\widehat{\beta}_{TMG} = \frac{1}{n_{\mathcal{G}}} \sum_{i \in \mathcal{G}_n} \widehat{\beta}_i^{\tau} = \sum_{i=1}^{n} w_i^{\tau} \widehat{\beta}_i^{\tau} \quad \text{with} \quad w_i^{\tau} = \frac{1}{n_{\mathcal{G}}} 1\{i \in \mathcal{G}_n\}, \tag{8}$$

where $1\{\cdot\}$ is the binary indicator. Recall that $n_{\mathcal{G}} = \sum_{i=1}^{n} 1\{i \in \mathcal{G}_n\}$ and $\mathcal{G}_n$ 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_t in the heterogeneous two-way FE regression (1), it does not yields the desired regression equation given as

$$\dot{y}_{it}^{\tau} = \dot{x}_{it}^{\tau\prime} \beta_i + \dot{u}_{it}^{\tau}.$$

Instead, it results in a transformed regression

$$\dot{y}_{it}^{\tau} = \tilde{x}_{it}' \beta_i - \frac{1}{n_{\mathcal{G}}} \sum_{j \in \mathcal{G}_n} \tilde{x}_{jt}' \beta_j + \dot{u}_{it}^{\tau} = \dot{x}_{it}^{\tau\prime} \beta_i + \dot{e}_{it}^{\tau}, \tag{9}$$

where

$$\dot{e}_{it}^{\tau} = \tilde{\xi}_{it} + \dot{u}_{it}^{\tau} \quad \text{with} \quad \tilde{\xi}_{it} = \frac{1}{n_{\mathcal{G}}} \sum_{j \in \mathcal{G}_n} \tilde{x}_{jt}' (\beta_i - \beta_j). \tag{10}$$

Because of $\tilde{\xi}_{it}$ term, therefore, the least squares estimator $\widehat{\beta}_i^{\tau}$ by regressing $\dot{y}_{it}^{\tau}$ on $\dot{x}_{it}^{\tau}$ for each i does not necessarily yield a consistent estimator of β_i. In other words, unlike the homogeneous two-way FE model, $\widehat{\beta}_i^{\tau} - \beta_i = O_p(T^{-1/2})$ 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 β_i satisfies (2). Also let $n_{\mathcal{G}} \to \infty$ as $n \to \infty$ satisfying $\phi = \lim_{n\to\infty} n/n_{\mathcal{G}}$ 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.

(2006),[3] *we have* $\sqrt{n}(\widehat{\beta}_{TMG} - \beta) \to_d \mathcal{N}(0, V_{TMG})$ *as* $n, T \to \infty$*, where* $V_{TMG} = \phi\Omega$*.*

Proof of Theorem 1 Note that $T^{-1}\sum_{t=1}^{T} \dot{x}_{it}^{\tau}\dot{x}_{it}^{\tau\prime} \to_p \Sigma_i^{\tau} > 0$ for each i as $T \to \infty$. Since $\widehat{\beta}_i^{\tau} - \beta_i = (T^{-1}\sum_{t=1}^{T} \dot{x}_{it}^{\tau}\dot{x}_{it}^{\tau\prime})^{-1}(T^{-1}\sum_{t=1}^{T} \dot{x}_{it}^{\tau}\xi_{it} + T^{-1}\sum_{t=1}^{T} \dot{x}_{it}^{\tau}\dot{u}_{it})$, (9) and (10) yields

$$\begin{aligned} \sqrt{n_{\mathcal{G}}}(\widehat{\beta}_{TMG} - \beta) &= \frac{1}{\sqrt{n_{\mathcal{G}}}}\sum_{i\in\mathcal{G}_n}(\beta_i - \beta) + \frac{1}{\sqrt{n_{\mathcal{G}}}}\sum_{i\in\mathcal{G}_n}(\widehat{\beta}_i^{\tau} - \beta_i) \qquad (11) \\ &= \frac{1}{\sqrt{n_{\mathcal{G}}}}\sum_{i\in\mathcal{G}_n}(\beta_i - \beta) + \frac{1}{\sqrt{n_{\mathcal{G}}}}\sum_{i\in\mathcal{G}_n}(\Sigma_i^{\tau})^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}^{\tau}\dot{u}_{it}\right) \\ &\quad + \frac{1}{\sqrt{n_{\mathcal{G}}}}\sum_{i\in\mathcal{G}_n}(\Sigma_i^{\tau})^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}^{\tau}\xi_{it}\right) + o_p(1) \\ &\equiv M_{1n} + M_{2n} + M_{3n} + o_p(1), \end{aligned}$$

where $M_{1n} \to_d \mathcal{N}(0, \Omega)$ as $n, T \to \infty$ 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

$$\begin{aligned} \frac{1}{\sqrt{n_{\mathcal{G}}}T}\sum_{i\in\mathcal{G}_n}\sum_{t=1}^{T}\dot{x}_{it}^{\tau}\xi_{it} &= \frac{1}{\sqrt{n_{\mathcal{G}}}T}\sum_{i\in\mathcal{G}_n}\sum_{t=1}^{T}\left(\tilde{x}_{it} - \frac{1}{n_{\mathcal{G}}}\sum_{k\in\mathcal{G}_n}\tilde{x}_{kt}\right)\left(\frac{1}{n_{\mathcal{G}}}\sum_{j\in\mathcal{G}_n}\tilde{x}_{jt}'(\beta_i - \beta_j)\right) \\ &= \frac{1}{\sqrt{n_{\mathcal{G}}}}\sum_{i\in\mathcal{G}_n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}\overline{\tilde{x}}_t'\right)\left(\beta_i - \overline{\beta}\right), \end{aligned}$$

where $\overline{\tilde{x}}_t = n_{\mathcal{G}}^{-1}\sum_{j\in\mathcal{G}_n}\tilde{x}_{jt}$ and $\overline{\beta} = n_{\mathcal{G}}^{-1}\sum_{j\in\mathcal{G}_n}\beta_j$. We thus write

$$M_{3n} = \frac{1}{\sqrt{n_{\mathcal{G}}}}\sum_{i\in\mathcal{G}_n}(\Sigma_i^{\tau})^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}\overline{\tilde{x}}_t'\right)\left(\beta_i - \overline{\beta}\right),$$

which satisfies $\mathbb{E}[M_{3n}] = 0$ since $\mathbb{E}[\beta_i - \overline{\beta}|x_i] = 0$. Moreover, $\mathbb{E}[(\beta_i - \overline{\beta})(\beta_i - \overline{\beta})'|x_i] = (1-(1/n))\Omega$, which we denote Ω_n. Apparently, $\Omega_n \to \Omega < \infty$ as $n \to \infty$. Under cross-sectional independence,

[3] In this two-way FE case, in particular, we set $d_t = \gamma_i = \Gamma_i = 1$ in the multifactor error structure in (13) and (14). Hence, $\mathbb{E}[u_{it}|x_{it}, F_t, \alpha_i, \beta_i] = 0$ for all i and t; and $\{x_{it}, u_{it}, \beta_i, \alpha_i\}$ 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, $T^{-1}\sum_{t=1}^{T} \dot{x}_{it}^{\tau}\dot{x}_{it}^{\tau\prime} \to_p \Sigma_i^{\tau} > 0$ for each i as $T \to \infty$. However, the cross-sectional independence assumption is to simplify the proof. We can relax the cross-sectional independence assumption of x_{it} by imposing a common factor structure like (14) as in the following section.

we have

$$\begin{aligned}
\mathbb{E}[M_{3n}M_{3n}'] &= \frac{1}{n_{\mathcal{G}}}\sum_{i\in\mathcal{G}_n}(\Sigma_i^{\tau})^{-1}\,\mathbb{E}\left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{x}_{it}\overline{\tilde{x}}_t'\mathbb{E}\left[\left(\beta_i-\overline{\beta}\right)\left(\beta_i-\overline{\beta}\right)'|x_i\right]\overline{\tilde{x}}_s\tilde{x}_{is}'\right](\Sigma_i^{\tau})^{-1} \\
&= \frac{1}{n_{\mathcal{G}}}\sum_{i\in\mathcal{G}_n}(\Sigma_i^{\tau})^{-1}\,\mathbb{E}\left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\tilde{x}_{it}\overline{\tilde{x}}_t'\Omega_n^{1/2}\right)\left(\tilde{x}_{is}\overline{\tilde{x}}_s'\Omega_n^{1/2}\right)'\right](\Sigma_i^{\tau})^{-1} \\
&= O\left(\frac{1}{T}+\frac{1}{n_{\mathcal{G}}^2}\right)
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\tilde{x}_{it}\overline{\tilde{x}}_t'\Omega_n^{1/2}\right)\left(\tilde{x}_{is}\overline{\tilde{x}}_s'\Omega_n^{1/2}\right)'\right] \\
=\ & \frac{1}{T}\mathbb{E}\left[\left(\tilde{x}_{it}\overline{\tilde{x}}_t'\Omega_n^{1/2}\right)\left(\tilde{x}_{it}\overline{\tilde{x}}_t'\Omega_n^{1/2}\right)'\right]+\frac{1}{T}\sum_{s=1}^{T-1}\left(1-\frac{s}{T}\right)Cov\left[\tilde{x}_{i1}\overline{\tilde{x}}_1'\Omega_n^{1/2},\tilde{x}_{is}\overline{\tilde{x}}_s'\Omega_n^{1/2}\right] \\
&+\frac{1}{T}\sum_{s=1}^{T-1}\left(1-\frac{s}{T}\right)\mathbb{E}\left[\tilde{x}_{i1}\overline{\tilde{x}}_1'\Omega_n^{1/2}\right]\mathbb{E}\left[\tilde{x}_{is}\overline{\tilde{x}}_s'\Omega_n^{1/2}\right] \\
=\ & O\left(\frac{1}{T}+\frac{1}{n_{\mathcal{G}}^2}\right)
\end{aligned}$$

for $\mathbb{E}[\tilde{x}_{i1}\overline{\tilde{x}}_1'\Omega_n^{1/2}] = n_{\mathcal{G}}^{-1}\sum_{j\in\mathcal{G}_n}\mathbb{E}[\tilde{x}_{i1}'\tilde{x}_{j1}]\Omega_n^{1/2} = O(1/n_{\mathcal{G}})$. Therefore, $M_{3n} = O_p(T^{-1/2}+n_{\mathcal{G}}^{-1})$, and the desired result follows as $n,T\to\infty$ by pre-multiplying $\sqrt{n/n_{\mathcal{G}}}$ to (11). □

Since $\phi \geq 1$, V_{TMG} cannot be smaller than the asymptotic variance of the standard MG estimator without trimming, which is Ω. Note that when the trimmed sample size is fixed (i.e., it does not depend on n), ϕ 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 $\widehat{\Sigma}_i$. It only depends on the reduction of the sample size from trimming. This is because we consider exogenous trimming; the efficiency gain from trimming the individual samples with small $\widehat{\Sigma}_i$ should be understood as the finite sample property.

The asymptotic variance can be consistently estimated by the sample covariance of $\widehat{\beta}_i^{\tau}$ as

$$\widehat{V}_{TMG} = \frac{n}{n_{\mathcal{G}}^2}\sum_{i\in\mathcal{G}_n}(\widehat{\beta}_i^{\tau}-\widehat{\beta}_{TMG})(\widehat{\beta}_i^{\tau}-\widehat{\beta}_{TMG})' \tag{12}$$

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

3.2 Trimmed 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

$$y_{it} = a_i' d_t + \gamma_i' f_t + x_{it}' \beta_i + u_{it}, \tag{13}$$

$$x_{it} = A_i' d_t + \Gamma_i' f_t + \epsilon_{it} \tag{14}$$

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

$$y_{it} = a_i' d_t + \delta_i' \overline{z}_t + x_{it}' \beta_i + u_{it}$$

for each i, where $\overline{z}_t = \sum_{i=1}^n \pi_i z_{it}$ with $z_{it} = (y_{it}, x_{it}')'$ for some weights $\{\pi_i\}$ satisfying Assumption 5 of Pesaran (2006). For our case, however, $\overline{z}_t$ 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

$$\overline{z}_t^{\tau} = \sum_{i=1}^n \pi_i^{\tau} z_{it},$$

where π_i^{τ} now imposes the same trimming scheme as w_i^{τ} in (8) but still satisfies Assumption 5 of Pesaran (2006). The trimmed CCE estimator $\widehat{\beta}_{CCE,i}^{\tau}$ for individual slope parameter β_i is then obtained from the least squares of

$$y_{it} = a_i' d_t + \delta_i' \overline{z}_t^{\tau} + x_{it}' \beta_i + u_{it},$$

which uses $\overline{z}_t^{\tau}$ instead of $\overline{z}_t$. From Theorem 1 of Pesaran (2006), we still have

$$\widehat{\beta}_{CCE,i}^{\tau} - \beta_i = O_p\left(T^{-1/2}\right) \tag{15}$$

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

$$\widehat{\beta}_{TCCEMG} = \frac{1}{n_{\mathcal{G}}} \sum_{i \in \mathcal{G}_n} \widehat{\beta}^{\tau}_{CCE,i} = \sum_{i=1}^{n} w_i^{\tau} \widehat{\beta}^{\tau}_{CCE,i},$$

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

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

Proof of Theorem 2 Note that

$$\begin{aligned} \sqrt{n}(\widehat{\beta}_{TCCEMG} - \beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (n w_i^{\tau})(\widehat{\beta}^{\tau}_{CCE,i} - \beta_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (n w_i^{\tau})(\beta_i - \beta) \\ &= O_p\left(\frac{1}{\sqrt{T}}\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (n w_i^{\tau})(\beta_i - \beta) \end{aligned}$$

from Theorem 1 of Pesaran (2006). The result follows immediately since β_i satisfies (2), where w_i^{τ} only depends on x_i, and $(1/n)\sum_{i=1}^{n}(n w_i^{\tau})^2 = (1/n)\sum_{i=1}^{n}(n 1\{i \in \mathcal{G}_n\}/n_{\mathcal{G}})^2 = n/n_{\mathcal{G}} \to \phi$ as $n \to \infty$. $\square$

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

$$\frac{n}{n_{\mathcal{G}}^2} \sum_{i \in \mathcal{G}_n} (\widehat{\beta}^{\tau}_{CCE,i} - \widehat{\beta}_{TCCEMG})(\widehat{\beta}^{\tau}_{CCE,i} - \widehat{\beta}_{TCCEMG})'.$$

We now denote an "induced" order statistic $\{\widehat{\beta}^{\tau}_{CCE,[i]}\}$, 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 $\{\widehat{\beta}^{\tau}_{CCE,[i]}\}$ itself (e.g., the ranking of $\widehat{\beta}^{\tau}_{CCE,i}$ is the same as the ranking of $\widehat{\Sigma}_i$ in our case). In such cases, we can define the TCCEMG estimator whose trimming scheme is directly from the order statistic of $\{\widehat{\beta}^{\tau}_{CCE,[i]}\}$.

For instance, for the scalar x_{it} case ($m = 1$), we can consider the TCCEMG estimator in the form of the sample trimmed mean defined as

$$\widehat{\beta}^*_{TCCEMG} = \frac{1}{\lfloor np_u \rfloor - \lfloor np_\ell \rfloor} \sum_{i=\lfloor np_\ell \rfloor + 1}^{\lfloor np_u \rfloor} \widehat{\beta}^{\tau}_{CCE,[i]}, \tag{16}$$

where $0 < p_\ell < p_u < 1$ are some fixed numbers denoting the lower and upper trimmed proportions, and $\lfloor c \rfloor$ 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 β_i. Then, we can obtain the limiting distribution of $\widehat{\beta}_{TCCEMG}$ as in the following theorem. We let $F_\tau(\cdot)$ be the cdf of trimmed β_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 β_i satisfies (2) with continuous ε_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 $\sqrt{n}(\widehat{\beta}^*_{TCCEMG} - \beta_\tau) \to_d \mathcal{N}(0, (p_u - p_\ell)^{-2} v_\tau)$ as $n, T \to \infty$ and $\sqrt{T}/n \to 0$, 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

$$\begin{aligned} \sqrt{n}(\widehat{\beta}^*_{TCCEMG} - \beta_\tau) &= \frac{\sqrt{n}}{\sqrt{\lfloor np_u \rfloor - \lfloor np_\ell \rfloor}} \times \frac{1}{\sqrt{\lfloor np_u \rfloor - \lfloor np_\ell \rfloor}} \sum_{i=\lfloor np_\ell \rfloor + 1}^{\lfloor np_u \rfloor} (\widehat{\beta}_{CCE,[i]} - \beta_{[i]}) \\ &\quad + \frac{\sqrt{n}}{\lfloor np_u \rfloor - \lfloor np_\ell \rfloor} \sum_{i=\lfloor np_\ell \rfloor + 1}^{\lfloor np_u \rfloor} \left(\beta_{[i]} - \beta_\tau\right) \\ &= O_p\left(\frac{1}{\sqrt{T}}\right) + \frac{\sqrt{n}}{\lfloor np_u \rfloor - \lfloor np_\ell \rfloor} \sum_{i=\lfloor np_\ell \rfloor + 1}^{\lfloor np_u \rfloor} \left(\beta_{[i]} - \beta_\tau\right) \end{aligned}$$

from Theorem 1 of Pesaran (2006). The result follows from the Theorem of Stigler (1973) with $A = B = 0$ in their expression. $\square$

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

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

$$\widehat{v}_\tau = \frac{1}{p_u - p_\ell} \sum_{i=\lfloor np_\ell \rfloor+1}^{\lfloor np_u \rfloor} \left(\widehat{\beta}_{CCE,[i]} - \widehat{\beta}^*_{TCCEMG} \right)^2$$

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

4 Monte Carlo Simulation

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

$$y_{it} = \beta_{1i} x_{1,it} + \beta_{2i} x_{2,it} + u_{it}, \tag{17}$$

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 $\mathcal{N}(0, \sigma_{ji}^2)$ where $\sigma_{ji}^2 \sim \chi_1^2$ 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

$$y_{it} = a_i F_t + \beta_{1i} x_{1,it} + \beta_{2i} x_{2,it} + u_{it} \tag{18}$$

and

$$x_{1,it} = x^o_{1,it} + \lambda_{1i} F_t \ \text{ and } \ x_{2,it} = x^o_{2,it} + \lambda_{2i} F_t,$$

where $\lambda_{2i} = \lambda_{1i} + \eta_i$ with both η_i and λ_{1i} being generated from $\mathcal{U}[1, 2]$. In addition, $x^o_{j,it} = \rho x^o_{j,it-1} + \epsilon_{j,it}$ for $j = 1, 2$ and $u_{it} = \rho u_{it-1} + v_{it}$, where ρ, $\epsilon_{j,it}$, 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 β_i case, we set $\beta_{j,i} = 1$ for all i and $j = 1, 2$. Meanwhile, for the heterogeneous β_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 $\widehat{\beta}_{1,i}$ and $\widehat{\beta}_{2,i}$, 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.[5] 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 χ^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

[5] 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.

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 $\lambda_i' F_t$ 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. β_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 β_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 $(\beta_{1,i}, \beta_{2,i}, \beta_{3,i})$ and the variance of the regressors using the test proposed by Sul (2016) and Campello, Galvao, and Juhl (2019),[6] which is asymptotically distributed as χ^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 β_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 β_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 β_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

[6] Both test are for the null hypothesis of $cov(\beta_i, \Sigma_i) = 0$. They require strict exogeneity, which holds in our example.

Table 7: Determinants of property crime

Two-way FE Estimation

	FE	MG	DTMG			XTMG		
			20%	10%	5%	20%	10%	5%
$\Delta \ln P_{i,t-1}$	0.027	0.014	-0.018	-0.003	-0.006	0.003	-0.001	-0.006
	(1.10)	(0.39)	(-0.66)	(-0.10)	(-0.23)	(0.09)	(-0.04)	(-0.23)
$\Delta \ln U_{i,t-1}$	-0.017	-0.014	-0.005	-0.001	-0.004	-0.003	-0.002	-0.002
	(-1.52)	(-1.31)	(-0.90)	(-0.29)	(-0.86)	(-0.46)	(-0.28)	(-0.44)
$\Delta \ln B_{i,t-1}$	-0.489**	-0.269*	-0.167	-0.190	-0.202*	-0.203	-0.210*	-0.192
	(-3.96)	(-1.80)	(-1.22)	(-1.51)	(-1.65)	(-1.45)	(-1.68)	(-1.54)

CCE Estimation

	pool	MG	DTMG			XTMG		
			20%	10%	5%	20%	10%	5%
$\Delta \ln P_{i,t-1}$	0.019	0.001	-0.011	-0.004	-0.001	-0.010	0.010	0.001
	(0.70)	(0.02)	(-0.25)	(-0.10)	(-0.03)	(-0.23)	(-0.24)	(0.02)
$\Delta \ln U_{i,t-1}$	-0.015	-0.014	-0.012	-0.011	-0.018	-0.015	-0.010	-0.012
	(-1.21)	(-1.22)	(-1.04)	(-1.04)	(-1.64)	(-1.18)	(-0.86)	(-1.06)
$\Delta \ln B_{i,t-1}$	-0.593**	-0.315	-0.224	-0.269	-0.329	-0.269	-0.308	-0.294
	(-4.36)	(-1.39)	(-0.85)	(-1.12)	(-1.42)	(-1.19)	(-1.33)	(-1.29)

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.

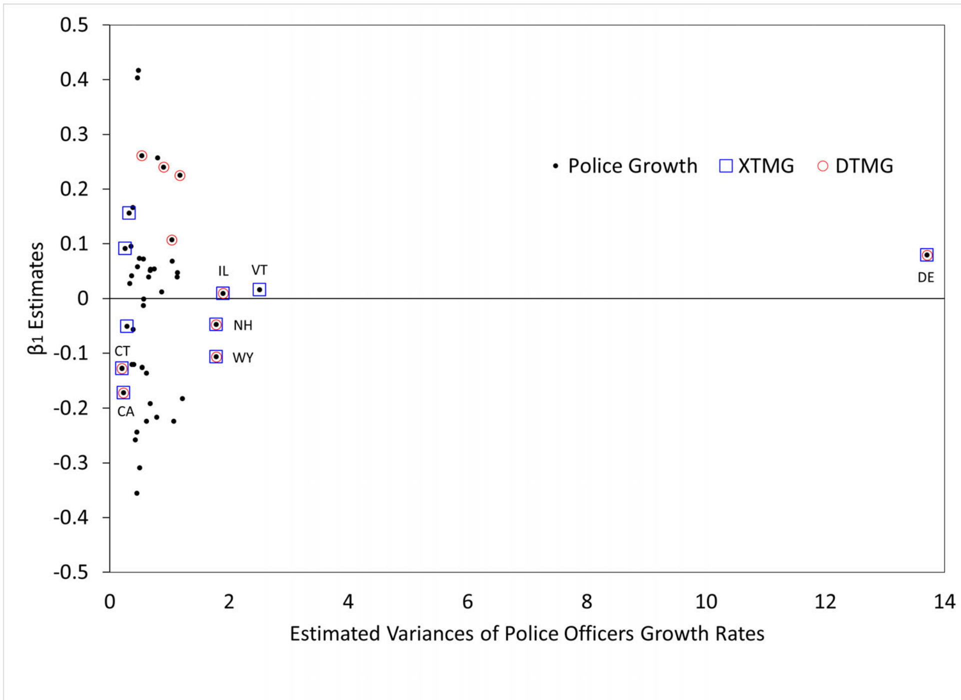

Figure 5: Relationship between $\widehat{\beta}_{1,i}$ and $\widehat{var}\left(\Delta \ln P_{it-1}\right)$

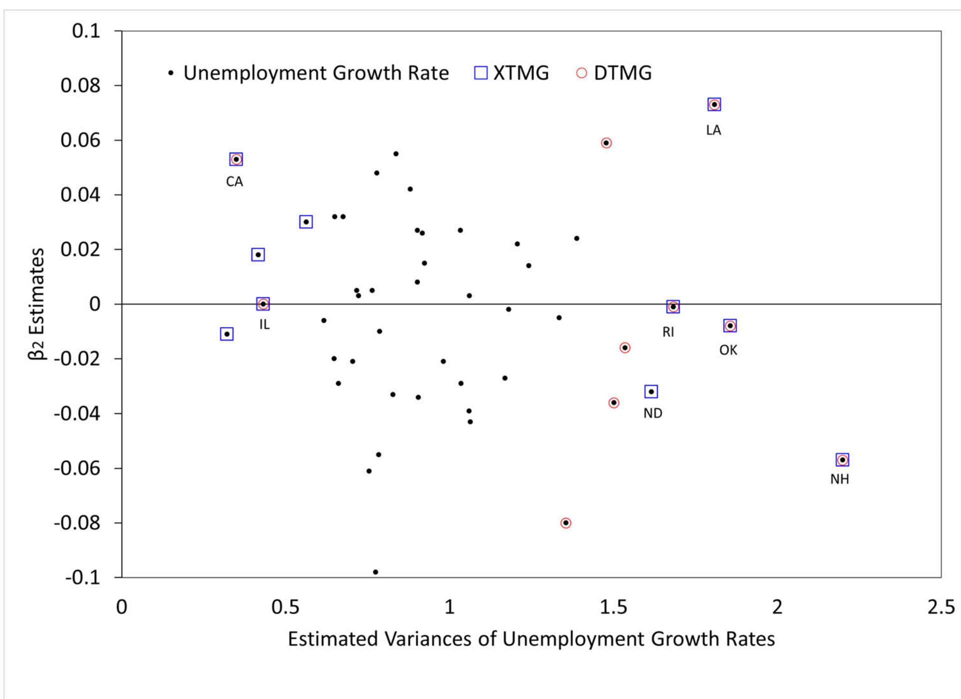

Figure 6: Relationship between $\widehat{\beta}_{2,i}$ and $\widehat{var}\left(\Delta \ln U_{it-1}\right)$

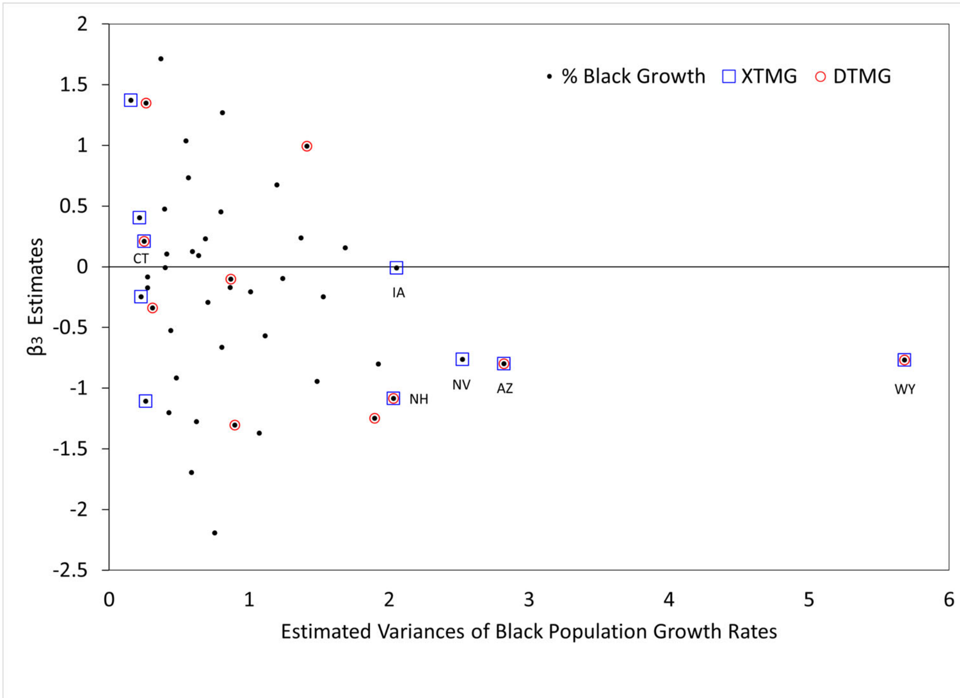

Figure 7: Relationship between $\widehat{\beta}_{3,i}$ and $\widehat{var}\left(\Delta \ln B_{it-1}\right)$

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 $\widehat{\beta}_i$ could be more desirable though its asymptotic analysis is not straightforward. We leave these important topics for future challenges.

References

Baltagi, B.H. and J.M. Griffin, 1997, Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline, *Journal of Econometrics*, 77, 303–327.

Baltagi, B.H., J.M. Griffin, and W. Xiong, 2000, To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand, *Review of Economics and Statistics*, 82, 117–126.

Baltagi, B.H., G. Bresson, and A. Pirotte, 2008, To pool or not to pool?, in L. Mátyás, P. Sevestre (eds.), *The Econometrics of Panel Data*, Springer Berlin Heidelberg, 517–546.

Bonhomme, S., and E. Manresa, 2015, Grouped patterns of heterogeneity in panel data, *Econometrica*, 83, 1147–1184.

Campello, M., Galvao, A. F., and Juhl, T., 2019, Testing for slope heterogeneity bias in panel data models. *Journal of Business & Economic Statistics*, 37, 749-760.

Chudik, A. and M.H. Pesaran, 2015, Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors, *Journal of Econometrics*, 188, 393–420.

Greenaway-McGrevy, R., Han, C., and Sul, D. 2012, Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics*, 169, 48-53.

Greenaway-McGrevy, R., N.C. Mark, D. Sul, and J.-L. Wu, 2018, Identifying exchange rate common factors, *International Economic Review*, 59, 2193–2218.

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.

Hsiao, C., M.H. Pesaran, and A. Tahmiscioglu, 1999, Bayes estimation of short-run coefficients in dynamic panel data models, in C. Hsiao, K. Lahiri, L.-F. Lee, M.H. Pesaran (eds.), *Analysis of panels and limited dependent variables: In honour of G.S. Maddala*, Cambridge University Press, 268–296.

Lee, Y., D. Mukherjee, and A. Ullah, 2019, Nonparametric estimation of the marginal effect in fixed-effect panel data models, *Journal of Multivariate Analysis*, 171, 53–67.

Lee, Y. and D. Sul. 2019, Depth-weigthed estimation of panel data models, *mimeo*, Syracuse University.

Levitt, S.D., 1997, Using electoral cycles in police hiring to estimate the effects of police on crime, *American Economic Review*, 87, 270–290.

Maddala, G.S., R.P. Trost, H. Li, and F. Joutz, 1997, Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators, *Journal of Business and Economic Statistics*, 15, 90–100.

Pesaran, M.H., 2006, Estimation and inference in large heterogenous panels with a multifactor error structure, *Econometrica*, 74, 967–1012.

Pesaran, M.H. and R. Smith, 1995, Estimating long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics*, 68, 79–113.

Pesaran, M.H., R. Smith, and K.S. Im, 1996, Dynamic Linear Models for Heterogeneous Panels, in L. Mátyás and P. Sevestre (eds.), *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, Dordrecht: Kluwer Academic Publishers, 145–195.

Pesaran, M.H., Y. Shin, and R. Smith, 1999, Pooled mean group estimation of dynamic heterogeneous panels, *Journal of the American Statistical Association*, 94, 621–634.

Pesaran, M.H. and T. Yamagata. 2008, Testing slope homogeneity in large panels, *Journal of Econometrics*, 142, 50–93.

Smith, A.F.M., 1973, Bayes estimates in one-way and two-way models, *Biometrika*, 60, 319–329.

Stigler, S.M., 1973, The asymptotic distribution of the trimmed mean, *Annals of Statistics*, 1, 472-477.

Swamy, P.A.V.B., 1970, Efficient inference in a random coefficient regression model, *Econometrica*, 38, 311–323.

Su, L., Z. Shi, and P.C.B. Phillips, 2016, Identifying latent structures in panel data, *Econometrica*, 84, 2215–2264.

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

Sul, D. 2019, *Panel data econometrics: Common factor analysis for empirical researchers*, Routledge.

Wallace, T.D. and A. Hussain, 1969, The use of error components models in combining cross section with time series data, *Econometrica*, 37, 55–72.

Wooldridge, J.M., 2005, Fixed effects and related estimators for correlated random-coefficient and treatment effect panel data models, *Review of Economics and Statistics*, 87, 385–390.

Table 1: MSE comparisons under the absence of any outliers

		$\beta_{1,i}=1$ & $\beta_{2,i}=1$ for all i				$\beta_{1,i}\sim\mathcal{N}(1,1)$ & $\beta_{2,i}\sim\mathcal{N}(1,1)$			
n	T	FE	MG	DTMG	XTMG	FE	MG	DTMG	XTMG
50	5	0.560	15.75	26.77	15.65	9.195	20.14	29.34	19.10
100	5	0.263	11.79	19.482	7.516	4.736	13.23	21.88	8.792
200	5	0.132	8.252	14.502	3.465	2.378	9.562	15.811	4.274
500	5	0.052	5.230	9.127	1.314	0.989	5.958	9.755	1.575
50	10	0.246	2.440	4.050	2.495	7.197	5.475	7.482	5.426
100	10	0.115	1.833	3.042	1.315	3.794	3.266	4.636	2.637
200	10	0.058	1.322	2.216	0.653	1.919	2.038	2.971	1.333
500	10	0.022	0.870	1.455	0.257	0.776	1.144	1.761	0.515
50	25	0.090	0.709	1.109	0.694	6.479	3.249	4.017	3.332
100	25	0.044	0.510	0.824	0.365	3.326	1.753	2.305	1.662
200	25	0.021	0.381	0.599	0.195	1.634	0.994	1.330	0.827
500	25	0.008	0.253	0.401	0.080	0.675	0.497	0.700	0.333
50	50	0.042	0.309	0.486	0.307	6.136	2.731	3.315	2.933
100	50	0.021	0.231	0.367	0.171	3.183	1.418	1.796	1.449
200	50	0.010	0.174	0.275	0.089	1.550	0.722	0.959	0.715
500	50	0.004	0.115	0.183	0.037	0.623	0.327	0.436	0.286
50	100	0.022	0.153	0.232	0.152	5.778	2.548	3.103	2.808
100	100	0.011	0.113	0.176	0.082	3.025	1.270	1.582	1.376
200	100	0.005	0.081	0.130	0.042	1.494	0.627	0.799	0.668
500	100	0.002	0.055	0.086	0.017	0.626	0.268	0.350	0.270

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%)

		$\beta_{1,i} = 1$ & $\beta_{2,i} = 1$ for all i				$\beta_{1,i} \sim \mathcal{N}(1,1)$ & $\beta_{2,i} \sim \mathcal{N}(1,1)$			
n	T	FE	MG	DTMG	XTMG	FE	MG	DTMG	XTMG
50	5	0.084	0.044	0.045	0.047	0.120	0.062	0.057	0.058
100	5	0.066	0.044	0.041	0.047	0.084	0.056	0.052	0.057
200	5	0.064	0.047	0.045	0.049	0.070	0.051	0.048	0.051
500	5	0.054	0.043	0.046	0.047	0.058	0.050	0.048	0.050
50	10	0.081	0.042	0.045	0.050	0.106	0.088	0.079	0.071
100	10	0.064	0.045	0.045	0.052	0.077	0.071	0.061	0.057
200	10	0.058	0.045	0.046	0.050	0.064	0.061	0.054	0.057
500	10	0.052	0.045	0.047	0.047	0.056	0.055	0.050	0.051
50	25	0.077	0.044	0.046	0.051	0.098	0.100	0.087	0.080
100	25	0.062	0.043	0.042	0.048	0.077	0.082	0.077	0.066
200	25	0.056	0.042	0.043	0.051	0.063	0.071	0.064	0.058
500	25	0.051	0.045	0.045	0.051	0.060	0.060	0.059	0.054
50	50	0.074	0.042	0.043	0.049	0.097	0.106	0.096	0.085
100	50	0.061	0.041	0.039	0.048	0.079	0.087	0.083	0.064
200	50	0.058	0.043	0.042	0.048	0.063	0.070	0.069	0.058
500	50	0.055	0.048	0.046	0.052	0.055	0.061	0.056	0.053
50	100	0.075	0.042	0.040	0.052	0.094	0.114	0.111	0.092
100	100	0.068	0.041	0.042	0.048	0.077	0.094	0.089	0.070
200	100	0.057	0.041	0.045	0.050	0.062	0.071	0.067	0.059
500	100	0.054	0.045	0.046	0.047	0.058	0.066	0.060	0.056

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

		$\beta_{j,i} = 1$ for all i except $i = n-1, n$ $\sigma^2_{j,n-1} = \sigma^2_{j,n} = 25$; $\beta_{j,n-1} = \beta_{j,n} = 5$				$\beta_{j,i} \sim \mathcal{N}(1,1)$ except $i = n-1, n$ $\sigma^2_{j,n-1} = \sigma^2_{j,n} = 25$; $\beta_{j,n-1} = \beta_{j,n} = 5$			
n	T	FE	MG	DTMG	XTMG	FE	MG	DTMG	XTMG
50	5	402.9	64.73	23.98	14.15	406.2	68.28	28.32	18.25
100	5	191.2	21.87	17.71	9.161	194.2	23.93	20.85	8.513
200	5	74.43	10.28	13.90	3.467	76.00	11.26	15.15	4.098
500	5	16.48	5.465	8.865	1.313	17.25	6.063	9.366	1.544
50	10	413.7	43.23	3.577	2.203	417.8	46.16	7.008	5.053
100	10	189.0	10.58	2.806	1.263	191.3	12.03	4.420	2.526
200	10	69.44	3.022	2.063	0.635	71.59	3.732	2.957	1.280
500	10	15.24	1.078	1.421	0.260	15.79	1.301	1.690	0.507
50	25	423.4	38.30	0.988	0.610	424.3	40.32	3.947	3.290
100	25	186.7	8.334	0.762	0.358	189.7	9.324	2.211	1.629
200	25	68.0	1.861	0.555	0.182	69.08	2.425	1.264	0.807
500	25	14.04	0.403	0.384	0.077	14.38	0.614	0.667	0.319
50	50	425.6	36.53	0.439	0.275	428.0	38.90	3.322	2.904
100	50	187.2	7.744	0.330	0.163	188.5	8.837	1.714	1.417
200	50	67.00	1.592	0.260	0.084	68.03	2.124	0.943	0.723
500	50	13.66	0.256	0.172	0.035	14.26	0.479	0.441	0.281
50	100	426.7	36.08	0.207	0.130	429.0	37.94	3.085	2.758
100	100	187.3	7.535	0.163	0.077	188.0	8.389	1.523	1.320
200	100	66.41	1.459	0.120	0.040	67.35	1.973	0.804	0.678
500	100	13.60	0.196	0.084	0.017	14.10	0.406	0.346	0.267

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%)

		$\beta_{j,i} = 1$ for all i except $i = n-1, n$ $\sigma^2_{j,n-1} = \sigma^2_{j,n} = 25$; $\beta_{j,n-1} = \beta_{j,n} = 5$				$\beta_{j,i} \sim \mathcal{N}(1,1)$ except $i = n-1, n$ $\sigma^2_{j,n-1} = \sigma^2_{j,n} = 25$; $\beta_{j,n-1} = \beta_{j,n} = 5$			
n	T	FE	MG	DTMG	XTMG	FE	MG	DTMG	XTMG
50	5	0.709	0.434	0.045	0.051	0.689	0.395	0.059	0.060
100	5	0.387	0.223	0.044	0.047	0.417	0.210	0.051	0.051
200	5	0.091	0.107	0.048	0.047	0.166	0.107	0.046	0.052
500	5	0.001	0.062	0.040	0.048	0.063	0.062	0.049	0.052
50	10	0.867	0.778	0.043	0.053	0.817	0.669	0.078	0.071
100	10	0.491	0.493	0.042	0.051	0.501	0.386	0.062	0.058
200	10	0.053	0.214	0.044	0.050	0.174	0.182	0.059	0.054
500	10	0.000	0.079	0.045	0.050	0.060	0.079	0.053	0.050
50	25	0.979	0.970	0.045	0.050	0.919	0.830	0.095	0.081
100	25	0.638	0.797	0.046	0.052	0.572	0.528	0.075	0.062
200	25	0.015	0.416	0.046	0.046	0.180	0.256	0.064	0.058
500	25	0.000	0.116	0.042	0.049	0.062	0.097	0.059	0.052
50	50	0.997	0.998	0.042	0.049	0.954	0.890	0.106	0.088
100	50	0.745	0.950	0.041	0.049	0.615	0.605	0.079	0.067
200	50	0.004	0.637	0.043	0.048	0.185	0.304	0.069	0.059
500	50	0.000	0.180	0.046	0.050	0.073	0.115	0.057	0.050
50	100	1.000	1.000	0.045	0.048	0.965	0.911	0.112	0.091
100	100	0.829	0.996	0.043	0.045	0.631	0.641	0.084	0.070
200	100	0.001	0.848	0.042	0.046	0.183	0.330	0.073	0.060
500	100	0.000	0.285	0.046	0.051	0.069	0.127	0.063	0.054

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

		$\beta_{j,i} = 1$ for all i except $i = n-1, n$ $\sigma^2_{j,n-1} = \sigma^2_{j,n} = 25$; $\beta_{j,n-1} = \beta_{j,n} = 5$				$\beta_{j,i} \sim \mathcal{N}(1,1)$ except for $i = n-1, n$ $\sigma^2_{j,n-1} = \sigma^2_{j,n} = 25$; $\beta_{j,n-1} = \beta_{j,n} = 5$			
n	T	pool	MG	DTMG	XTMG	pool	MG	DTMG	XTMG
50	10	217.3	14.67	12.782	15.08	233.2	16.56	16.40	15.10
50	25	214.1	7.185	2.190	2.013	229.0	9.045	4.868	4.698
50	50	214.6	6.178	0.932	0.757	233.2	8.259	3.600	3.429
50	100	214.4	5.894	0.433	0.357	231.8	7.624	2.952	2.917
100	10	86.52	8.189	9.109	8.469	99.84	8.860	10.30	9.780
100	25	81.00	2.891	1.569	1.354	94.96	3.778	2.780	2.684
100	50	79.74	2.275	0.620	0.552	93.90	3.221	1.985	1.871
100	100	79.05	2.044	0.308	0.263	92.63	2.948	1.598	1.594
200	10	31.78	5.230	6.987	6.560	39.42	5.974	9.567	6.190
200	25	27.70	1.360	1.047	0.902	36.19	1.928	1.919	1.771
200	50	27.08	0.946	0.603	0.428	35.11	1.418	1.135	1.084
200	100	26.60	0.760	0.207	0.185	34.70	1.221	0.863	0.832
500	10	7.655	2.853	4.391	3.558	10.84	3.088	4.449	3.931
500	25	6.423	0.612	0.680	0.585	9.343	0.775	0.903	0.822
500	50	6.207	0.332	0.269	0.234	8.999	0.519	0.525	0.491
500	100	6.093	0.231	0.127	0.111	8.872	0.422	0.391	0.372

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%)

		$\beta_{j,i} = 1$ for all i except $i = n-1, n$ $\sigma^2_{j,n-1} = \sigma^2_{j,n} = 25$; $\beta_{j,n-1} = \beta_{j,n} = 5$				$\beta_{j,i} \sim \mathcal{N}(1,1)$ except $i = n-1, n$ $\sigma^2_{j,n-1} = \sigma^2_{j,n} = 25$; $\beta_{j,n-1} = \beta_{j,n} = 5$			
n	T	pool	MG	DTMG	XTMG	pool	MG	DTMG	XTMG
50	10	0.502	0.124	0.043	0.050	0.524	0.113	0.054	0.061
50	25	0.618	0.256	0.044	0.041	0.611	0.168	0.058	0.061
50	50	0.703	0.357	0.042	0.043	0.655	0.197	0.059	0.058
50	100	0.759	0.457	0.043	0.044	0.662	0.192	0.063	0.065
100	10	0.178	0.087	0.042	0.044	0.281	0.080	0.048	0.052
100	25	0.110	0.203	0.041	0.042	0.291	0.126	0.052	0.056
100	50	0.067	0.324	0.042	0.042	0.287	0.154	0.058	0.059
100	100	0.036	0.459	0.043	0.042	0.283	0.166	0.058	0.059
200	10	0.015	0.066	0.045	0.045	0.120	0.066	0.048	0.048
200	25	0.000	0.139	0.046	0.043	0.128	0.098	0.051	0.052
200	50	0.000	0.240	0.049	0.046	0.123	0.122	0.056	0.060
200	100	0.000	0.375	0.042	0.045	0.129	0.133	0.055	0.056
500	10	0.002	0.053	0.051	0.044	0.065	0.051	0.047	0.048
500	25	0.000	0.084	0.044	0.049	0.081	0.070	0.052	0.049
500	50	0.000	0.125	0.048	0.047	0.079	0.083	0.049	0.053
500	100	0.000	0.209	0.045	0.046	0.079	0.097	0.057	0.053

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.