Depth-Weighted Means of Noisy Data: An Application to Estimating the Average Effect in Heterogeneous Panels*

Yoonseok Lee†
Syracuse University

Donggyu Sul‡
University of Texas at Dallas

August 2022

Abstract

We study the depth-weighted L-type location estimator of multivariate data when the observations are measured with noise. Under a drifting asymptotic framework, we show that the depth-weighted mean estimators with noisy data are still consistent and asymptotically mean-zero Gaussian under mild conditions. We apply the results to longitudinal data models of heterogeneous agents and develop the depth-weighted mean-group estimator of a vector of random coefficients, which estimates the multivariate average effect in heterogeneous panels. As an empirical illustration, we examine the relative purchasing power parity.

Keywords: Depth, Noisy data, Longitudinal data, Heterogeneous panel, Random coefficient, Average effect.

*The authors thank to Robert Serfling and participants at numerous seminars and conference presentations for helpful comments. Lee acknowledges support from Appleby-Mosher Research Fund, Maxwell School, Syracuse University.

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

The depth of multivariate data is commonly used to measure how close a given observation vector is toward the center of the underlying joint distribution, and hence it leads to a center-outward ordering of each observation. Examples of data depth measures include the half-space depth (Tukey (1975)), the simplicial depth (Liu (1990)), the projection depth (Liu (1992); Zuo and Serfling (2000); Zuo (2003)), and Mahalanobis depth (Liu and Singh (1993)), to name a few. As a robust location estimator of multivariate observations, Stahel (1981), Donoho (1982), Liu (1990), and Liu et al. (1999) for instance, consider using the depth to construct weighted means, which enjoys good efficiency and robustness properties. Zuo et al. (2004) study the asymptotic behavior of the general form of the depth-weighted L-type location estimators.

These studies presume that we observe the true multivariate data of interest and estimate its depth-weighted mean. In this paper, we instead assume the situation that we cannot observe the true data of interest but the observations measured with errors. The main interest of this paper is to use such noisy observations to estimate the depth-weighted mean of the latent variables (i.e., the data without noise) and study its limiting properties. To this end, we consider a drifting asymptotic framework to ensure a meaningful bias-variance trade-off in the limit, where the noise vanishes at a certain rate as the sample size increases though it presents for any fixed sample size. Under this framework, we extend the asymptotic results of Zuo et al. (2004) to the depth-weighted L-type location estimator of noisy data. We show that, though such a local deviation or noise in general yields non-zero bias in the limit expressions of the empirical distribution and its linear statistical functions, it is not the case for the depth-weighted L-type location estimator when the noise vanishes at (or faster than) the square-root of the sample size. This reveals the robustness property of the depth-weighted mean estimator in the view of local misspecification of the underlying distribution.

A motivating example of observations measured with noise that satisfies the drifting asymptotics is a collection of consistent estimators. More precisely, for any parameter vector of each individual agent in a heterogeneous longitudinal model, such as the heterogeneous treatment effect, we can regard its consistent estimator as the observation with vanishing noise, whereas the true parameter vector value is the latent observation without noise. In this context, as an application, we consider model averaging in heterogeneous longitudinal

data regression models and develop the depth-weighted mean-group (DWMG) estimator of a vector of random coefficients. Under a certain rate condition between the magnitude of the noise and the sample size, it estimates multivariate average effects in heterogeneous longitudinal data models consistently and also robustly toward outlying individuals or erroneous reports. In this regard, it extends the common average estimators in heterogeneous panel data regression models such as Swamy (1970), Pesaran and Smith (1995), Pesaran (2006), and Hsiao and Pesaran (2008). When the partial effect is indeed homogeneous, furthermore, the depth-weighted mean-group estimator is shown to be consistent to the true effect. The simulation results show that this new estimator can achieve a good balance between robustness and efficiency.

The rest of the paper is organized as follows. In Section 2, we study the empirical distribution of multivariate noisy data and establish a asymptotic representation of the depth-weighted mean estimator of the noisy observations. In Section 3, we apply the new estimator in the context of longitudinal data regression and develop the depth-weighted mean-group estimator. In Section 4, we present the finite sample efficiency of the depth-weighted mean-group estimator in simulations and examine relative purchasing power parity as an empirical illustration. We conclude in Section 5. We collect all the proofs and computational details in the Appendix.

2 Depth-Weighted Mean Estimator of Noisy Data

2.1 Multivariate noisy samples

We let $X_i \in \mathcal{X} \subset \mathbb{R}^k$ for $k \geq 1$ be a random sample from a distribution F. As a robust measure of the central tendency of multivariate observations, Liu (1990), Liu et al. (1999), and Zuo et al. (2004) study a weighted mean whose weighting scheme is determined by the statistical depth function. More precisely, the depth-weighted mean of X_i is defined as

$$L(F) = \frac{\int xW(\mathcal{D}(x,F))F(dx)}{\int W(\mathcal{D}(x,F))F(dx)}, \tag{1}$$

where $\mathcal{D}(x,F) \in [0,1]$ is a depth function and $W(\cdot)$ is a non-negative weight function satisfying $\int W(\mathcal{D}(x,F))F(dx) > 0$ and $\int ||x||W(\mathcal{D}(x,F))F(dx) < \infty$.

This definition of the centrality in (1) is known to be robust to outlying observations in X_i. It is also general enough to encompass the popular centrality measures. For example, if the moment $\mathbb{E}[X_i]$ exists, then $L(F) = \mathbb{E}[X_i]$ when $W(\cdot) = 1$. If we define $W(\cdot)$ as a trimming form, then $L(F)$ becomes the trimmed (depth-weighted) mean. When the density function of X_i is symmetric about its mode, $L(F)$ is the median in the multivariate sense.

When X_i is observable, we can readily estimate $L(F)$ by $L(F_n)$ as in the aforementioned studies (e.g., equation (2.1) of Zuo et al. (2004)), where F_n is the empirical distribution of X_i given as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq x\}$$

for $x \in \mathcal{X}$ and $1\{\cdot\}$ is the binary indicator. However, we suppose that we cannot directly observe X_i but we observe data measured with errors. More precisely, for each $i = 1, \ldots, n$, we suppose multivariate noisy observations $X_i^e \in \mathbb{R}^k$ such that

$$X_i^e = X_i + e_i, \tag{2}$$

where e_i is a vector of noise. We assume $\mathbb{E}[e_i] = 0$ and $\mathbb{E}[e_i e_i^\top] = O(m_i^{-1})$ for some $m_i \to \infty$ as $n \to \infty$, and hence the noise vanishes as the sample size increases, while it presents for any fixed sample size.

Natural examples of such noisy observations X_i^e are generated or predicted variables by projection; and the estimators of heterogeneous parameters in longitudinal data models and multilevel models (see Section 3). For the latter example, it should be noted that m_i is not necessarily the sample size used to estimate X_i, though it is typically the case for the M-estimators, such as the maximum likelihood and the least squares estimators. For instance, if X_i^e is an estimator obtained by maximum score estimation, $m_i = T_i^{2/3}$, where T_i is the sample size in estimating X_i for each i. If X_i^e is a point of a local constant nonparametric estimator, $m_i = T_i^{4/(4+k)}$ with the optimal bandwidth choice, where we allow for local-to-zero mean of e_i.

Since we do not observe X_i, the empirical distribution F_n in (2) is no longer feasible, and hence we cannot use $L(F_n)$ to estimate $L(F)$. Instead, we consider the empirical distribution

of the noisy sample X_i^e given as

$$F_n^e(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i^e \leq x\}$$

and estimate $L(F)$ by the feasible depth-weighted mean estimator:

$$L(F_n^e) = \frac{\sum_{i=1}^{n} X_i^e W(\mathcal{D}(X_i^e, F_n^e))}{\sum_{i=1}^{n} W(\mathcal{D}(X_i^e, F_n^e))}, \tag{3}$$

which is a simple plug-in estimator. Note that the depth-weighted mean estimator $L(F_n^e)$ is still well defined even when $X_i = X_j = X$ for all i, j. In this case, the variation in the noisy observations is solely from the noise e_i and the data depth of X_i^e is based on the distribution of the noise e_i. The depth-weighted mean naturally estimates X; see Section 3.2 for the details.

2.2 Asymptotics of depth-weighted mean estimator

To study the asymptotic properties of the depth-weighted mean estimator based on the noisy data, $L(F_n^e)$ in (3), we first want to see how the local misspecification affects the limit of F_n^e. To this end, we assume the followings. Without loss of generality, we let $m_i = m$ for all i.

(A1) *$e_i = \varepsilon_i/\sqrt{m}$, where ε_i is a mean-zero $k \times 1$ random vector independent across i with finite fourth moment and $\Omega = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \mathbb{E}[\varepsilon_i \varepsilon_i^\top] < \infty$.*

(A2) *X_i is a random sample from a distribution F and independent of ε_i.*

(A3) *F is twice continuously differentiable over $\mathcal{X}$ with uniformly bounded derivatives.*

(A4) *$\sqrt{n}/m \to \phi$ for some constant $0 \leq \phi < \infty$ as $n, m \to \infty$.*

In Assumption (A1), the noise e_i can be non-identically distributed. This will allow for a heteroskedastic error term in the longitudinal data regression models, when we regard X_i^e as a vector of estimators of heterogeneous parameters as in Section 3. In (A2), X_i can be from a fat-tailed distribution and hence its moments do not necessarily exist. (A4) ensures a meaningful bias-variance trade-off in the limit under our drifting asymptotic framework.

The following lemma summarizes the limiting properties of $F_n^e(\cdot)$ when $n, m \to \infty$. The proof is in the Appendix A.1.

Lemma 1 *As $n, m \to \infty$, we have*

(i) $\sup_{x \in \mathcal{X}} |F_n^e(x) - F(x)| = O_p(n^{-1/2} + m^{-1})$ *under Assumptions (A1)-(A3);*

(ii) for any $x \in \mathcal{X}$, under Assumptions (A1)-(A4),

$$\sqrt{n}\left(F_n^e(x) - F(x) - \frac{1}{2m} tr[\ddot{F}(x)\Omega]\right) \to_d \mathcal{N}(0, F(x)[1 - F(x)]), \tag{4}$$

where $\ddot{F}(x)$ is the Hessian matrix of $F(x)$.

Lemma 1 shows that the local deviation yields non-zero asymptotic bias of order m^{-1}, which is resulted from the deviation of the probability limit of $F_n^e(x)$, say $\mathbb{P}\{X_i^e \leq x\}$, from $F(x) = \mathbb{P}\{X_i \leq x\}$. For this reason, we need both large n and m to achieve (uniform) consistency. However, neither the stochastic order of the local deviation of $\mathbb{P}\{X_i^e \leq x\}$ from $\mathbb{P}\{X_i \leq x\}$ nor the limiting ratio $\lim_{n,m\to\infty} \sqrt{n}/m$ is important for consistency. On the other hand, the relative size of m to n becomes important in deriving the limiting distribution in (4). It can be readily seen from $\sqrt{n}(F_n^e(x) - F(x)) = \sqrt{n}(F_n(x) - F(x)) + \sqrt{n}(F_n^e(x) - F_n(x))$, where the first term satisfies the standard Functional Central Limit Theorem and the second term is verified to be $O_p(\sqrt{n}/m)$ that contributes to the non-zero mean in the limiting distribution unless $\sqrt{n}/m \to 0$. When $\sqrt{n}/m \to \phi$ for some non-zero constant $\phi < \infty$, we need to correct the bias to have a mean-zero Gaussian in the limit. This is quite common in the panel survey data when the number of survey waves m is small relative to the number of surveyees n. It should be noted that, however, the bias term in (4) can be ignored at some x with $\ddot{F}(x) = 0$. For instance, if x is a (local) optimizer of the density function of X_i, then $\ddot{F}(x) = 0$ and hence the bias becomes zero.

From Lemma 1, we can derive the limiting distribution of an estimator $\widehat{\theta} = \theta(F_n^e)$, where $\theta(\cdot)$ is a regular function that is Hadamard differentiable at F with a bounded derivative. We can readily see that the bias in (4) can results in asymptotic bias in the limiting distribution of $\widehat{\theta}$ in general. For instance, we consider a linear statistical function $\theta_0 = \theta(F) = \int q(x)F(dx)$ for some smooth $q(\cdot)$ and its estimator based on a noisy sample $\widehat{\theta} = \theta(F_n^e) = \int q(x)F_n^e(dx)$. If $q(\cdot)$ is second-order continuously differentiable with satisfying $\mathbb{E}[q(X_i)q(X_i)^\top] < \infty$ and

$\mathbb{E}[\ddot{q}(X_i)] < \infty$, where $\ddot{q}(\cdot)$ is the Hessian matrix of $q(\cdot)$, Lemma 1 yields that

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}-\theta_0\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{q(X_i)-\mathbb{E}[q(X_i)]\}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{q(X_i^e)-q(X_i)\} \\ &\to_d \mathcal{N}\left(\phi\Psi, \mathbb{E}[(q(X_i)-\mathbb{E}[q(X_i)])(q(X_i)-\mathbb{E}[q(X_i)])^\top]\right) \qquad (5)\end{aligned}$$

as $n, m \to \infty$, with $\Psi = tr[\mathbb{E}[\ddot{q}(X_i)]\Omega]/2 < \infty$. It is important to note that, if $q(x)$ is linear and hence $\ddot{q}(x) = 0$ for any $x \in \mathcal{X}$, we have $\Psi = 0$ regardless of the value of $\phi = \lim_{n,m\to\infty}\sqrt{n}/m$. Apparently, when $m \geq n$ and hence $\phi = 0$, the limiting distribution of $\widehat{\theta}$ is naturally centered at the origin. Otherwise, we can correct the bias Ψ when a consistent estimator of Ω is available, or correct the leading bias term in (5) using jackknife estimation as in Hahn and Newey (2004).

Now we further assume the following conditions, which are similar to the conditions in Zuo et al. (2004). We define $\mathcal{X}_{d_0} = \{x : \mathcal{D}(x, F) \geq d_0\}$ for some $d_0 \geq 0$.

(A5) *$W(\cdot) \in [0, \infty)$ is continuously differentiable with a bounded derivative $\dot{W}(\cdot)$; $W(d) = 0$ for $d \in [0, d_0\kappa]$ with some $\kappa > 1$.*

(A6) *$\int W(\mathcal{D}(x, F))F(dx) > 0$; $\int\{||x||W(\mathcal{D}(x, F))\}^2F(dx)$ and $\int\{||x||\dot{W}(\mathcal{D}(x, F))\}^2F(dx)$ are bounded.*

(A7) *$\sup_{x\in\mathcal{X}}|\mathcal{D}(x, F_n^e)-\mathcal{D}(x, F)| = O_p(n^{-1/2}+m^{-1})$ and $\sup_{x\in\mathcal{X}_{d_0}}||x||\mathcal{D}(x, F_n^e)-\mathcal{D}(x, F)| = O_p(n^{-1/2}+m^{-1})$.*

(A8) *$\mathcal{D}(x, \cdot)$ is Fréchet differentiable at F with respect to the supremum metric, whose derivative is bounded.*

Assumptions (A5) and (A6) are the same as those in Zuo et al. (2004), except for the last condition in (A6), and they are well satisfied for popular depth functions. Assumption (A5) supposes a sufficiently smooth weight function $W(\cdot)$ that is zero in the neighborhood of the origin. Assumption (A6) ensures that the depth-weighted mean $L(F)$ in (1) is well-defined. The last condition in (A6) could impose some restrictions on the choice of $W(\cdot)$.

Under the condition in Assumption (A4), Assumption (A7) implies $\sup_{x\in\mathcal{X}}\sqrt{n}|\mathcal{D}(x, F_n^e)-\mathcal{D}(x, F)| = O_p(1)$ and $\sup_{x\in\mathcal{X}_{d_0}}\sqrt{n}||x||\mathcal{D}(x, F_n^e)-\mathcal{D}(x, F)| = O_p(1)$, which are similar to the conditions in Zuo et al. (2004). For some depth functions, this assumption can be verified

under proper conditions on the location and the scale parameter estimators.[1] By Lemmas A and B in Section 6.2.2 of Serfling (1980) or Proposition 2.19 of Huber and Ronchetti (2009), the Fréchet differentiability of $\mathcal{D}(x,\cdot)$ in Assumption (A8) yields that there exists a bounded function $h(x,y)$ such that

$$\sqrt{n}(\mathcal{D}(x,F_n^e)-\mathcal{D}(x,F)) = \int h(x,y)\sqrt{n}(F_n^e-F)(dy) + o_p\left(\sup_{x\in\mathcal{X}}|F_n^e(x)-F(x)|\right) \tag{6}$$

uniformly on $\mathcal{X}_{d_0}$. Some examples of the form of $h(\cdot,\cdot)$ in (6) are given in the Appendix A.2.

The following theorem gives asymptotics of the depth-weighted mean estimator $L(F_n^e)$ in (3). It shows that $L(F_n^e)$ is $\sqrt{n}$-consistent to $L(F)$ and it is asymptotically mean-zero Gaussian as $n,m\to\infty$, which corresponds to the standard results with uncontaminated data (e.g., Maronna and Yohai (1995); Zuo et al. (2004)). Therefore, unlike the cases in (5), it does not suffer from the asymptotic bias problem. The proof is in the Appendix A.1.

Theorem 1 *Suppose Assumptions (A1)-(A8) hold and* $\sup_{x\in\mathcal{X}}|h(x,y_1)-h(x,y_2)| = O(||y_1-y_2||)$ *for any* $y_1,y_2\in\mathcal{X}$. *As* $n,m\to\infty$, $L(F_n^e)-L(F)=O_p(n^{-1/2})$ *and*

$$\sqrt{n}(L(F_n^e)-L(F)) \to_d \mathcal{N}(0,V_F), \tag{7}$$

where $V_F=\int K_F^0(x)K_F^0(x)^\top F(dx)$ *with*

$$K_F^0(x) = \frac{\int (y-L(F))\dot{W}(\mathcal{D}(y,F))h^0(y,x)F(dy) + (x-L(F))W(\mathcal{D}(x,F))}{\int W(\mathcal{D}(y,F))F(dy)} \tag{8}$$

and $h^0(y,x)=h(y,x)-\int h(y,x')F(dx')$.

The additional condition on $h(x,y)$ can be readily verified for the Mahalanobis depth and the projection depth. In Theorem 1, though it seems like that we only need $n\to\infty$ in the expression (7), we use $n,m\to\infty$ asymptotics to obtain the limiting distribution of the depth-weighted mean estimator $L(F_n^e)$. This is because

$$\sqrt{n}(L(F_n^e)-L(F)) = \sqrt{n}(L(F_n)-L(F)) + \sqrt{n}(L(F_n^e)-L(F_n)) \equiv N_{n1}+N_{n2},$$

[1] For example, see Donoho and Gasko (1992) for the halfspace depth; Liu (1990) and Dümbgen (1992) for the simplicial depth; Liu and Singh (1993) for the majority and Mahalanobis depths; Zuo and Serfling (2000) and Zuo (2003) for the projection depth.

in which $N_{n1} \to_d \mathcal{N}(0, V_F)$ as $n \to \infty$ from the standard results (e.g., Theorems 2.1 and 3.1 of Zuo et al. (2004)), whereas $N_{n2} = o_p(1)$ only when both $n, m \to \infty$. In other words, to have the limiting expression correspond to the case with observed X_i without noise, it is required that both $n, m \to \infty$ and $\lim_{n,m\to\infty} \sqrt{n}/m$ does not diverge as in Assumption (A4). It is however important to note that, unlike the result in (5), the mean-zero limiting distribution of $L(F_n^e)$ is obtained even when $\sqrt{n}/m$ does not go to zero. Therefore, the depth-weighted mean estimator is asymptotically robust to a local deviation from the underlying distribution, provided the degree of local deviation is controlled by the condition $\lim_{n,m\to\infty} \sqrt{n}/m < \infty$.

For the asymptotic variance V_F in Theorem 1, since $L(F_n^e) - L(F) = o_p(1)$, $X_i^e - X_i = o_p(1)$ for all i, and $\sup_{x\in\mathcal{X}} |\mathcal{D}(x, F_n^e) - \mathcal{D}(x, F)| = o_p(1)$ from Assumption (A7) for large n and m, we can estimate it as

$$\widehat{V}_F = \frac{1}{n} \sum_{i=1}^{n} \widehat{K}_F^0(X_i^e) \widehat{K}_F^0(X_i^e)^\top, \tag{9}$$

where

$$\widehat{K}_F^0(X_i^e) = \frac{n^{-1} \sum_{j=1}^{n} (X_j^e - L(F_n^e)) \dot{W}(\mathcal{D}(X_j^e, F_n^e)) \widehat{h}^0(X_j^e, X_i^e) + (X_i^e - L(F_n^e)) W(\mathcal{D}(X_i^e, F_n^e))}{n^{-1} \sum_{j=1}^{n} W(\mathcal{D}(X_j^e, F_n^e))} \tag{10}$$

with data depth $\mathcal{D}(X_i^e, F_n^e)$ of the noisy observations and

$$\widehat{h}^0(X_j^e, X_i^e) = \widehat{h}(X_j^e, X_i^e) - \frac{1}{n} \sum_{\ell=1}^{n} \widehat{h}(X_j^e, X_\ell^e)$$

for some consistent estimator $\widehat{h}(\cdot, \cdot)$ of $h(\cdot, \cdot)$. Though the expression of $\widehat{K}_F^0(X_i^e)$ in (10) appears complicated, it can be readily obtained in practice since $\widehat{h}(\cdot, \cdot)$ is mostly based on some sample analogues. Some examples and detailed steps of calculating $\widehat{h}(\cdot, \cdot)$ are in the Appendix A.2.

3 Depth-Weighted Mean-Group Estimator

3.1 Heterogeneous case

As an application, we consider model averaging in a heterogeneous longitudinal data regression model and develop the depth-weighted mean-group estimator. More precisely, we

suppose a longitudinal data model with potentially heterogeneous multivariate partial effects (across agents in panel data, or across groups in clustered data) and estimate the average effect using the depth-weighted mean estimator. For instance, we consider panel data regression with random coefficients given as

$$y_{it} = z_{it}^{\top}\beta_i + \lambda_i^{\top} f_t + u_{it} \tag{11}$$

for agent $i = 1, \ldots, n$ and time $t = 1, \ldots, m$, where the slope parameters $\beta_i \in \mathcal{B} \subset \mathbb{R}^k$ are potentially heterogeneous across i. z_{it} is a $k \times 1$ vector of exogenous regressors or treatment variables of interest, λ_i is a vector of factor loadings, and f_t is a vector of common factors.[2] When $\lambda_i = (\alpha_i, 1)^{\top}$ and $f_t = (1, \tau_t)^{\top}$, (11) becomes the two-way fixed effects regression model with random coefficients. z_{it} can be discrete as in the difference-in-difference analysis.

The parameter vector β_i can be either heterogeneous or homogeneous (i.e., $\beta_i = \beta$ a.s. for all i), which is unknown. Individual responses to some treatment are often very heterogeneous, and they are described as the heterogeneous coefficients β_i in the regression (11), which can include some outlying or erroneous responses to the treatment. In such cases, pooled estimation with imposing the homogeneity restriction on the coefficients β_i can result in biased estimators of the average effect.

As a robust measure of the average effect, we consider the depth-weighted mean of the multivariate random coefficients β_i defined as

$$\overline{\beta} = \frac{\int bW(\mathcal{D}(b, F))F(db)}{\int W(\mathcal{D}(b, F))F(db)} \tag{12}$$

as in (1), where F is the joint distribution function of β_i and $W(\cdot)$ is some non-negative weight function. The depth function $\mathcal{D}(b, F) \in [0, 1]$ is defined on F and hence it measures how much each heterogeneous parameter vector β_i is distant from the center of its distribution. If $\mathbb{E}[\beta_i]$ exists, then $\overline{\beta} = \mathbb{E}[\beta_i]$ when $W(\cdot) = 1$. Under the homogeneous parameter setup (i.e., $\beta_i = \beta$ a.s. for all i), β_i has a point mass only at β and hence $\overline{\beta} = \beta$.

We develop a generalized mean-effect estimator using the idea of the depth-weighted mean estimator in (3). Whether β_i is heterogeneous or not, we can obtain the individual-specific parameter estimator $\widehat{\beta}_i$ for each i. We define the *depth-weighted mean-group (DWMG)*

[2] We suppose the number of factors is given and fixed. When the common factors are not observed, we project on the estimated factors as in Pesaran (2006) and Bai (2009).

estimator as

$$\widehat{\beta}_{DW} = \frac{\sum_{i=1}^{n} \widehat{\beta}_i W(\mathcal{D}(\widehat{\beta}_i, F_n^e))}{\sum_{i=1}^{n} W(\mathcal{D}(\widehat{\beta}_i, F_n^e))} \tag{13}$$

as in (3), where $F_n^e(b) = n^{-1} \sum_{i=1}^{n} 1\{\widehat{\beta}_i \leq b\}$ for $b \in \mathcal{B}$ in this case. $\widehat{\beta}_{DW}$ is an L-estimator of generated variables $\widehat{\beta}_i$, in which the order statistic is based on the data depth of each $\widehat{\beta}_i$. It is hence more general than the mean-group estimator (e.g., Pesaran and Smith (1995)), which is an equally-weighted average of $\widehat{\beta}_i$.

Assumptions (A1) and (A2) are mild for this regression setup. In particular, Assumption (A1) is rewritten as

$$\widehat{\beta}_i - \beta_i = O_p(m^{-1/2}), \tag{14}$$

and it holds for the M-estimator or the GMM estimator $\widehat{\beta}_i$, when β_i is independent of $\{z_{it}, \lambda_i, f_t, u_{it}\}$ for all i and t, like the standard assumption of the random coefficient panel data regression models (e.g., Hsiao and Pesaran (2008)). Hence, we can see $\widehat{\beta}_i$ as a noisy observation of β_i, which respectively correspond to X_i^e and X_i in (2). Though we consider the linear regression model (11) with exogenous z_{it} mainly for the sake of presentation simplicity, we can obtain the DWMG estimator in (13) for more general cases such as nonlinear models and dynamic models with sequentially exogenous regressors, as long as we have an estimator $\widehat{\beta}_i$ satisfying (14).[3] We also allow for (conditionally) heteroskedastic regression error u_{it}. It is important to note that Assumption (A2) does not require the moments of β_i exist, and hence it is more general than the conditions of the typical random coefficient models (e.g., Swamy (1970)).

The asymptotic properties of $\widehat{\beta}_{DW}$ can be readily obtained from Theorem 1.

Corollary 1 *Suppose $\beta_i \sim$ iid F. Under the same conditions as Theorem 1, $\widehat{\beta}_{DW} \to_p \overline{\beta}$ and $\sqrt{n}(\widehat{\beta}_{DW} - \overline{\beta}) \to_d \mathcal{N}(0, V_F)$ as $n, m \to \infty$, where V_F is given in (8).*

A (joint) confidence interval of $\widehat{\beta}_{DW}$ can be obtained using the normal approximation in Corollary 1, where V_F is estimated as (9). In practice, using the fact that $\widehat{\beta}_{DW}$ is a weighted

[3] For instance, we can consider M-estimation given by

$$\min_{\{\beta_i, \lambda_i, f_t\}_{i=1}^n} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi\left(y_{it}, z_{it}; \beta_i, \lambda_i, f_t\right)$$

for some known objective function ψ. Examples include panel linear regression models (e.g., Pesaran (2006)), panel binary choice models (e.g., Boneva and Linton (2016)), and panel quantile regression models (e.g., Harding et al. (2020)) with individual specific intercepts and potentially heterogeneous slope parameters.

average in the form of $\widehat{\beta}_{DW} = \sum_{i=1}^{n} \omega_i \widehat{\beta}_i$, where $\omega_i = W(\mathcal{D}(\widehat{\beta}_i, F_n^e)) / \sum_{j=1}^{n} W(\mathcal{D}(\widehat{\beta}_j, F_n^e))$, one can estimate V_F as

$$\widehat{V}_F = n \sum_{i=1}^{n} \omega_i^2 \widehat{\sigma}_i^2 \tag{15}$$

conditional on the weights ω_i, where $\widehat{\sigma}_i^2$ is a consistent estimator of $var(\widehat{\beta}_i)$. Though this approach ignores the randomness in the weight ω_i, simulation study in Section 4 shows that inference based on this approach works well in finite samples.

3.2 Homogeneous case

We can still use the DWMG estimator (13) even when the slope parameters in (11) are homogeneous in the true data generating model: $\beta_i = \beta$ almost surely for all i. In this case, the depth-weighted mean $\overline{\beta}$ in (12) is the same as the true slope parameter β and the rate of convergence of the DWMG estimator $\widehat{\beta}_{DW}$ can be improved to $(nm)^{-1/2}$, which is quite natural since we take average over i and t at the same time.

For the homogeneous case, however, the meaning of the data depth becomes different from the heterogeneous case. Since $\beta_i = \beta$ a.s. for all i, we cannot define the depth of β_i and the data depth of $\widehat{\beta}_i$ no longer estimates the depth of $\beta_i = \beta$. Instead, the heterogeneity of $\widehat{\beta}_i$ is now solely from the estimation error and the depth based on $\widehat{\beta}_i$ describes that of the scaled estimation error

$$\widehat{\xi}_i = m^{1/2}(\widehat{\beta}_i - \beta). \tag{16}$$

To study the statistical properties of the DWMG estimator $\widehat{\beta}_{DW}$ in this case, we first suppose the following condition that replaces (A1).

(A1′) *There exists a mean-zero random vector ξ_i satisfying $\widehat{\xi}_i - \xi_i = o_p(1)$ as $m \to \infty$ and independent across i; ξ_i has finite fourth moment and $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbb{E}[\xi_i \xi_i^\top] < \infty$.*

We let $\mathcal{C} \subset \mathbb{R}^k$ be the support of ξ_i and $G(r) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbb{P}\{\xi_i \le r\}$ for any $r \in \mathcal{C}$. When ξ_i is identically distributed, G is simply the distribution function of ξ_i. We also define $G_n(r) = n^{-1} \sum_{i=1}^{n} 1\{\xi_i \le r\}$ and $G_n^e = n^{-1} \sum_{i=1}^{n} 1\{\widehat{\xi}_i \le r\}$. Assumptions (A3) and (A6)-(A7) are rewritten as follows by replacing F with G.

(A3′) *G is twice continuously differentiable over $\mathcal{C}$ with uniformly bounded derivatives.*

(A6′) $\int W(\mathcal{D}(r,G))G(dr) > 0$; $\int\{||r||W(\mathcal{D}(r,G))\}^2 G(dr)$ *and* $\int\{||r||\dot{W}(\mathcal{D}(r,G))\}^2 G(dr)$ *are bounded.*

(A7′) $\sup_{r\in\mathcal{C}}\sqrt{n}|\mathcal{D}(x,G_n^e)-\mathcal{D}(x,G)| = O_p(1)$ *and* $\sup_{r\in\mathcal{C}_{d_0}}||r||\sqrt{n}|\mathcal{D}(x,G_n^e)-\mathcal{D}(x,G)| = O_p(1)$, *where* $\mathcal{C}_{d_0}=\{r:\mathcal{D}(r,G)\geq d_0\}$ *for some* $d_0\geq 0$.

Then, as Corollary 1, we can obtain the asymptotic properties of $\widehat{\beta}_{DW}$ under the homogeneous parameter case as follows.

Corollary 2 *Suppose* $\beta_i=\beta$ *a.s. for all* i. *Under the same conditions as Corollary 1 using (A1′), (A3′), (A6′), and (A7′) instead of the corresponding assumptions,* $\widehat{\beta}_{DW}\to_p\beta$ *and* $\sqrt{nm}(\widehat{\beta}_{DW}-\beta)\to_d\mathcal{N}(0,V_G)$ *as* $n,m\to\infty$, *where* $V_G=\int K_G^0(r)K_G^0(r)^\top G(dr)$ *with*

$$K_G^0(r)=\frac{\int s\dot{W}(\mathcal{D}(s,G))\eta^0(s,r)G(ds)+rW(\mathcal{D}(r,G))}{\int W(\mathcal{D}(s,G))G(ds)},$$

$\eta^0(s,r)=\eta(s,r)-\int\eta(s,r')G(dr')$, *and* $\eta(s,r)$ *is a bounded function satisfying* $\sqrt{n}(\mathcal{D}(s,G_n^e)-\mathcal{D}(s,G))=\int\eta(s,r)\sqrt{n}(G_n^e-G)(dr)+o_p(1)$ *uniformly on* $\mathcal{C}_{d_0}$.

We let $\widetilde{\xi}_i=m^{1/2}(\widehat{\beta}_i-\widehat{\beta}_{DW})$. Then it holds that $\widetilde{\xi}_i=m^{1/2}(\widehat{\beta}_i-\beta)-m^{1/2}(\widehat{\beta}_{DW}-\beta)=\widehat{\xi}_i+o_p(1)$ since $\widehat{\beta}_{DW}-\beta=O_p((nm)^{-1/2})$. Hence, the asymptotic variance V_G can be estimated as $\widehat{V}_G=n^{-1}\sum_{i=1}^n\widehat{K}_G^0(\widetilde{\xi}_i)\widehat{K}_G^0(\widetilde{\xi}_i)^\top$, where

$$\widehat{K}_G^0(\widetilde{\xi}_i)=\frac{n^{-1}\sum_{j=1}^n\widetilde{\xi}_j\dot{W}(\mathcal{D}(\widetilde{\xi}_j,\widetilde{G}_n))\widehat{\eta}^0(\widetilde{\xi}_j,\widetilde{\xi}_i)+\widetilde{\xi}_iW(\mathcal{D}(\widetilde{\xi}_i,\widetilde{G}_n))}{n^{-1}\sum_{j=1}^nW(\mathcal{D}(\widetilde{\xi}_j,\widetilde{G}_n))},$$

and $\widehat{\eta}^0(\widetilde{\xi}_j,\widetilde{\xi}_i)=\widehat{\eta}(\widetilde{\xi}_j,\widetilde{\xi}_i)-n^{-1}\sum_{\ell=1}^n\widehat{\eta}(\widetilde{\xi}_j,\widetilde{\xi}_\ell)$. Here, $\widetilde{G}_n(r)=n^{-1}\sum_{i=1}^n 1\{\widetilde{\xi}_i\leq r\}$ and $\widehat{\eta}(\cdot,\cdot)$ is the sample analogue of $\eta(\cdot,\cdot)$. Alternatively, because of the affine-invariance property of the depth function, we can simply let $\widehat{V}_G=n^{-1}\sum_{i=1}^n(m^{1/2}\widehat{K}_F^0(\widehat{\beta}_i))(m^{1/2}\widehat{K}_F^0(\widehat{\beta}_i))^\top$, where $\widehat{K}_F^0(\widehat{\beta}_i)$ is from (10).

Remark 1 (Robustness) For the homogeneous case, since the weights in $\widehat{\beta}_{DW}$ depend on the depth of the scaled estimation error $\widehat{\xi}_i$, we can expect its robustness against outlying $\widehat{\beta}_i$'s. For instance, in the simple one-way fixed effects regression model, $y_{it}=z_{it}^\top\beta_i+\alpha_i+u_{it}$, we have $\widehat{\xi}_i=(m^{-1}\sum_{t=1}^m z_{it}^0z_{it}^{0\top})^{-1}(m^{-1/2}\sum_{t=1}^m z_{it}^0u_{it}^0)$, where $z_{it}^0=z_{it}-m^{-1}\sum_{s=1}^m z_{is}$,

and the typical sources of outlying behaviors include: when z_{it} has little time-variation for some i and hence $m^{-1}\sum_{t=1}^{m} z_{it}^0 z_{it}^{0\top}$ is near singular; when several individual i's have measurement errors in z_{it} resulting in non-zero $\mathbb{E}[z_{it}^0 u_{it}^0]$; when $var(z_{it}^0 u_{it}^0)$ is very large for some i under heteroskedasticity. Existence of such outlying individuals is very likely to yield heterogeneity in the individual-specific estimates $\widehat{\beta}_i$. In such cases, the standard maximum likelihood (which corresponds to the within estimator or the least squares dummy variable estimator) or the mean-group estimators may not be even consistent. It is also possible that a poolability test rejects the null of homogeneous model, not because the true parameters β_i's are indeed heterogeneous, but because some $\widehat{\beta}_i$'s impose serious estimation errors that causes the test to conclude an incorrect result. The DWMG estimator $\widehat{\beta}_{DW}$ uses the resulting heterogeneity in the individual-specific estimates $\widehat{\beta}_i$ directly to construct a robust estimator without identifying the source of the heterogeneity.

Remark 2 (Mixture of heterogeneous and homogeneous parameters) Suppose $\beta_i = (\beta_{Ii}^\top, \beta_{IIi}^\top)^\top \in \mathbb{R}^{k_I + k_{II}}$, where $\beta_{Ii} \neq \beta_{Ij}$ for some $i \neq j$ but $\beta_{IIi} = \beta_{II}$ for all i. Also let $\overline{\beta} = (\overline{\beta}_I^\top, \beta_{II}^\top)^\top$ be the depth-weighted mean of β_i. In this case, Corollaries 1 and 2 imply that

$$\sqrt{n}\left(\widehat{\beta}_{DW} - \overline{\beta}\right) = \sqrt{n}\begin{pmatrix} \widehat{\beta}_{I,DW} - \overline{\beta}_I \\ \widehat{\beta}_{II,DW} - \beta_{II} \end{pmatrix} \to_d \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_I & 0 \\ 0 & 0 \end{pmatrix}\right)$$

as $n, m \to \infty$, where $V_I > 0$ is defined as in Corollary 1 based on the marginal distribution of β_{Ii}. It suggests that we only need to consider the depth of the heterogeneous component β_{Ii} and their depth-weighted estimator $\widehat{\beta}_{I,DW}$ for inferences of $\overline{\beta}_I$. In fact, if we denote the joint distribution of $\beta_i = (\beta_{Ii}^\top, \beta_{IIi}^\top)^\top$ as F and the marginals as (F_I, F_{II}), we can readily verify that the following two estimators are both consistent to $\overline{\beta}_I$:

$$\widehat{\beta}_{I,DW} = \frac{\sum_{i=1}^n \widehat{\beta}_{Ii} W(\mathcal{D}(\widehat{\beta}_i, F_n^e))}{\sum_{i=1}^n W(\mathcal{D}(\widehat{\beta}_i, F_n^e))} \quad \text{and} \quad \widetilde{\beta}_{I,DW} = \frac{\sum_{i=1}^n \widetilde{\beta}_{Ii} W(\mathcal{D}(\widetilde{\beta}_{Ii}, F_{In}^e))}{\sum_{i=1}^n W(\mathcal{D}(\widetilde{\beta}_{Ii}, F_{In}^e))},$$

where $\widehat{\beta}_i = (\widehat{\beta}_{Ii}^\top, \widehat{\beta}_{IIi}^\top)^\top$ is an estimator using all the regressors and $\widetilde{\beta}_{Ii}$ is an estimator using only the regressors with heterogeneous slopes β_{Ii}. An interesting example is when $\beta_{IIi} = 0$ for all i, which is the case that the corresponding regressors are irrelevant and hence redundant. In this case, whether we use the joint depth or marginal depth, we have the same limit of the depth-weighted estimator.

4 Numerical Illustrations

4.1 Simulations

We examine the finite sample efficiency of the depth-weighted mean estimator. In particular, we consider the DWMG estimators in the two-way fixed effects regression model:

$$y_{it} = z_{it}^{\top}\beta_i + \alpha_i + \tau_t + u_{it} \tag{17}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, m$, with $z_{it} \in \mathbb{R}^2$. The individual effects are generated as $\alpha_i = \underline{\alpha}_i + m^{-1}\sum_{t=1}^{m}(z_{1,it} + z_{2,it})$ with $\underline{\alpha}_i \sim iid\mathcal{U}[0,1]$ and the time effects as $\tau_t = \underline{\tau}_t + n^{-1}\sum_{i=1}^{n}(z_{1,it} + z_{2,it})$ with $\underline{\tau}_t \sim iid\mathcal{U}[0,1]$. The regressors $z_{it} = (z_{1,it}, z_{2,it})^{\top}$ and the error term u_{it} are uncorrelated and respectively generated as

$$z_{it} \sim iid\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right) \quad \text{and} \quad u_{it} \sim iid\mathcal{N}(0,1),$$

where $\rho = 0.1$.[4]

In this setup, β_i corresponds to the latent observations without noise X_i in (2); and the individual time series estimate $\widehat{\beta}_i$ corresponds to the observed noisy data $X_i^e = X_i + e_i$ in (2), where e_i is the estimation error satisfying Assumption (A1). We examine the following two data generating processes F:

- DGP1: $\beta_i = (\beta_{1,i}, \beta_{2,i})^{\top}$ are randomly generated from $\mathcal{N}((1,1)^{\top}, I_2)$, where I_2 is the identify matrix of rank 2.
- DGP2: $\beta_i = (\beta_{1,i}, 1)^{\top}$ for all i, where $\beta_{1,i}$ is randomly generated from $\mathcal{N}(1,1)$.

DGP 1 assumes the fully heterogeneous slopes and DGP 2 assumes a mixture of homogeneous and heterogeneous slopes. For each data generating process, we also consider contaminated β_i similarly as in the simulation design of Zuo et al. (2004). More precisely, we suppose there are outlying individuals whose slope parameters β_i are contaminated and generated from $(1-\epsilon)F + \epsilon\mathcal{N}((10,10)^{\top}, 5^2 I_2)$ with $\epsilon = 0\%$, 5%, and 10% (i.e., ϵ% of β_i in the sample

[4] We also considered a heteroskedastic error with $\mathbb{E}[u_{it}^2] = \sigma_i^2$ and $\sigma_i^2 \sim iid\mathcal{X}_1^2$. However, the simluation results remain very similar and hence are not reported.

are randomly generated from $\mathcal{N}((10,10)^\top, 5^2 I_2))$, where F is the distribution of β_i either in DGP1 or in DGP2.

We compare the sample mean and the four types of depth-weighted means. $\mathrm{DW_P}$ is based on the projection depth and $\mathrm{DW_M}$ is based on the Mahalanobis depth. For each case, we consider two types of weight functions: $W(d) = d$ and

$$W(d) = \frac{\exp(-3\left(1 - d/\overline{D}\right)^2) - \exp(-3)}{1 - \exp(-3)} 1\{d < \overline{D}\} + 1\{d \geq \overline{D}\} \tag{18}$$

as in Zuo et al. (2004), where $\overline{D}$ is the sample median of $\mathcal{D}(\widehat{\beta}_i, F_n^e)$. We indicate each weighting type using superscript 0 and W, like $\mathrm{DW_P^0}$ and $\mathrm{DW_P^W}$. We investigate the sample relative efficiency by comparing the empirical mean squared errors (MSE). In particular, Tables 1 and 2 report the ratio obtained by dividing the empirical MSE of the sample mean by that of each estimator. The simulation results are based on 2000 iterations for different combinations of sample sizes $n = (100, 200)$ and $m = (5, 10)$. We choose small m values here, which show the cases with heavy noise.

Note that when we consider the noisy observations $\widehat{\beta}_i$, the sample mean $n^{-1}\sum_{i=1}^n \widehat{\beta}_i$ corresponds to the mean-group estimator. We also include the ML estimator in comparison, which corresponds to the two-way fixed effect estimator or the least squares (two-way) dummy variables estimator: $\widehat{\beta}_{ML} = (\sum_{i=1}^n \sum_{t=1}^m z_{it}^* z_{it}^{*\top})^{-1} \sum_{i=1}^n \sum_{t=1}^m z_{it}^* y_{it}^*$, where $z_{it}^* = z_{it}^0 - n^{-1}\sum_{j=1}^n z_{jt}^0$ with $z_{it}^0 = z_{it} - m^{-1}\sum_{s=1}^m z_{is}$. This estimator presumes the homogeneous slope parameter (i.e., $\beta_i = \beta_j$ for all i, j) and hence is the optimal one for the homogeneous panel regression with a homoskedastic error.

Table 1 reports the empirical MSE ratios of the sample mean to the DWMG estimators in DGP1, where both elements in β_i are fully heterogeneous. Hence, the value greater than unity implies that the DWMG estimator has smaller MSE than the sample mean; the relative efficiency of the DWMG estimator improves as the ratio gets large. The left panel of the table is based on the latent values of β_i, which corresponds to the case that Zuo et al. (2004) considered. Note that the values on the left panel do not depend on m by construction, because β_i are not estimated. The right panel of the table is based on the individual time series estimates $\widehat{\beta}_i$ and hence the noisy observations that we are interested in.

We summarize the main findings as follows. First, the overall performance between these two panels are quite similar, and the values on the right panel get closer to those on the

Table 1: Mean Square Error Ratio to Mean: Heterogeneous Case

			data without noise $\beta_i = (\beta_{1,i}, \beta_{2,i})^\top$				*noisy data* $\widehat{\beta}_i$				
ϵ	n	m	$\mathrm{DW}_{\mathrm{P}}^{0}$	$\mathrm{DW}_{\mathrm{P}}^{W}$	$\mathrm{DW}_{\mathrm{M}}^{0}$	$\mathrm{DW}_{\mathrm{M}}^{W}$	$\mathrm{DW}_{\mathrm{P}}^{0}$	$\mathrm{DW}_{\mathrm{P}}^{W}$	$\mathrm{DW}_{\mathrm{M}}^{0}$	$\mathrm{DW}_{\mathrm{M}}^{W}$	ML
0%	100	5	0.86	0.91	0.90	0.86	1.04	1.11	1.09	1.07	0.99
	100	10	0.86	0.91	0.90	0.86	0.88	0.93	0.92	0.88	0.79
	200	5	0.89	0.93	0.91	0.87	1.06	1.12	1.09	1.09	1.02
	200	10	0.89	0.93	0.91	0.87	0.90	0.94	0.92	0.88	0.82
5%	100	5	13.40	18.10	12.60	17.80	8.18	10.57	7.60	10.19	0.86
	100	10	13.40	18.10	12.60	17.80	11.63	15.69	10.82	15.45	0.90
	200	5	19.23	30.57	17.20	30.34	11.92	17.66	10.51	17.07	0.90
	200	10	19.23	30.57	17.20	30.34	16.44	25.56	14.52	25.26	0.94
10%	100	5	23.74	46.42	20.71	46.87	13.74	23.82	12.07	23.53	0.91
	100	10	23.74	46.42	20.71	46.87	19.22	36.52	17.06	37.40	0.94
	200	5	27.61	65.05	23.76	67.40	16.97	34.71	14.64	35.17	0.93
	200	10	27.61	65.05	23.76	67.40	22.97	51.55	19.83	53.49	0.97

Note: $\mathrm{DW}_{\mathrm{P}}^{0}$ and $\mathrm{DW}_{\mathrm{P}}^{W}$ are the DWMG estimators based on the projection depth, using the weight function $W(d) = d$ and (18), respectively; $\mathrm{DW}_{\mathrm{M}}^{0}$ and $\mathrm{DW}_{\mathrm{M}}^{W}$ are based on the Mahalanobis depth, using the weight function $W(d) = d$ and (18), respectively. ML corresponds to the least squares (two-way) dummy variables estimator. Each value is the ratio obtained by dividing the empirical MSE of the sample mean by that of each estimator.

Table 2: Mean Square Error Ratio to Mean: Mixture of Hetero and Homogeneous Case

			data without noise $\beta_i = (\beta_{1,i}, 1)^\top$				*noisy data* $\widehat{\beta}_i$				
ϵ	n	m	$\mathrm{DW}_{\mathrm{P}}^{0}$	$\mathrm{DW}_{\mathrm{P}}^{W}$	$\mathrm{DW}_{\mathrm{M}}^{0}$	$\mathrm{DW}_{\mathrm{M}}^{W}$	$\mathrm{DW}_{\mathrm{P}}^{0}$	$\mathrm{DW}_{\mathrm{P}}^{W}$	$\mathrm{DW}_{\mathrm{M}}^{0}$	$\mathrm{DW}_{\mathrm{M}}^{W}$	ML
0%	100	5	0.83	0.85	0.93	0.89	1.25	1.31	1.27	1.26	1.27
	100	10	0.83	0.85	0.93	0.89	0.91	0.95	0.94	0.90	0.84
	200	5	0.83	0.85	0.92	0.87	1.23	1.30	1.26	1.26	1.31
	200	10	0.83	0.85	0.92	0.87	0.90	0.93	0.93	0.88	0.90
5%	100	5	34.46	36.70	30.65	34.55	13.06	16.74	11.59	15.09	0.86
	100	10	34.46	36.70	30.65	34.55	22.46	28.53	20.57	26.20	0.90
	200	5	62.15	67.08	45.56	57.66	19.72	28.91	16.66	25.61	0.90
	200	10	62.15	67.08	45.56	57.66	34.54	49.09	30.18	44.33	0.95
10%	100	5	110.3	128.3	42.61	83.09	22.27	39.70	17.18	33.12	0.92
	100	10	110.3	128.3	42.61	83.09	41.62	76.17	29.20	60.28	0.94
	200	5	183.2	227.6	49.60	115.4	29.26	62.69	21.49	50.28	0.93
	200	10	183.2	227.6	49.60	115.4	56.57	124.3	35.50	87.49	0.97

Note: $\mathrm{DW}_{\mathrm{P}}^{0}$ and $\mathrm{DW}_{\mathrm{P}}^{W}$ are the DWMG estimators based on the projection depth, using the weight function $W(d) = d$ and (18), respectively; $\mathrm{DW}_{\mathrm{M}}^{0}$ and $\mathrm{DW}_{\mathrm{M}}^{W}$ are based on the Mahalanobis depth, using the weight function $W(d) = d$ and (18), respectively. ML corresponds to the least squares (two-way) dummy variables estimator. Each value is the ratio obtained by dividing the empirical MSE of the sample mean by that of each estimator.

left panel as (m, n) increases, which supports the main theorem. Second, when there is no contamination (i.e., $\epsilon = 0\%$), though the depth-weighted mean estimator is less efficient with β_i as expected, it is not the always the case with the noisy data $\widehat{\beta}_i$. Importantly, when the noise is large (i.e., when m is small), the depth-weighted mean estimator can be more efficient than the sample mean. Third, it is evident that the depth-weighted mean estimators outperform the sample mean when there is contamination; the gain in relative efficiency improves as more samples are contaminated (i.e., ϵ increases). Fourth, the difference in the relative efficiency of the depth-weighted mean estimators attribute the choice of the weight function $W(\cdot)$; imposing low weights on outlying observations improves the efficiency.

Table 2 reports the empirical relative efficiency in DGP2, where one element in β_i is homogeneous. In this case, the relative efficiency improvement is more dramatic than the fully heterogeneous case in Table 1, but the overall patterns remain the same. Without contamination, the ML also performs well, and it is because the model gets closer to the homogeneous panel regression, under which the ML is the optimal.

4.2 Empirical Illustration

We investigate the cross-sectional heterogeneity in deviations from the law of one price and illustrate how the conventional panel pooled estimator can be misled by a small number of outlying individuals in the heterogeneous environment. In particular, we consider the relative purchasing power parity (PPP) model, which predicts that the change in the exchange rate of two countries is determined by the difference in price level changes. We estimate the relative PPP parameters of various currencies to the U.S. Dollar (USD) and examine if there exists any evidence supporting the relative PPP hypothesis on average.

To this end, we consider the following factor augmented regression as in Greenaway-McGrevy et al. (2018):

$$\Delta_{it} = \beta_i(\pi_{it} - \pi_{0t}) + \lambda_i^\top f_t + e_{it} \tag{19}$$

for currency (or country) $i = 1, \ldots, n$ and time $t = 1, \ldots, m$, where Δ_{it} is the monthly depreciation rate of the ith currency against the USD, π_{it} is the monthly inflation rate in the country i, and π_{0t} is the monthly inflation rate in the U.S. We include 5 common factors $f_t = (1, f_{\Delta,t}^\top, f_{\pi,t}^\top)^\top \in \mathbb{R}^{1+2+2}$, where $f_{\Delta,t}$ and $f_{\pi,t}$ collect two factors respectively from Δ_{it} and π_{it}, so that the cross-sectional dependence among the nominal and relative inflation rates are

well controlled. The number of factors are selected using Bai and Ng (2002)'s IC_2 criterion with the maximum number of 8 for each variable. We use a monthly panel data set of 27 bilateral spot exchange rates and consumer price indices (CPI) from 1999.M1 to 2015.M6. The source of the data is Global Insight (GI) at Information Handling Service (IHS). The list of the currencies are in the note of Table 3.

We expect $\beta_i = 1$ if the relative PPP strictly holds for the ith currency in the short run; $\beta_i = 0$ if there is no relationship between the depreciation rate and the relative inflation rate, which strongly rejects the relative PPP hypothesis. Knowing that the strict relative PPP rarely holds, however, one is typically interested in the hull hypothesis $\beta_i = 0$ for each i. Table 3 reports the factor augmented least squares estimate $\widehat{\beta}_i$ for each currency, its standard error, and the two t-statistics for the null of $\beta_i = 0$ and $\beta_i = 1$. The standard error of each $\widehat{\beta}_i$ is calculated by the Newey-West robust estimator with the lag length selection of $\lfloor m^{1/3} \rfloor$. At the individual level, the null of $\beta_i = 1$ is very often rejected: 13 out of 27 cases are rejected at the 5% level. Meanwhile, the null of $\beta_i = 0$ is hardly rejected: only 4 out of 27 cases are rejected at the 5% level. Also note that the sign of $\widehat{\beta}_i$ is negative in 12 out of 27 cases and positive in 15 cases.[5]

We now consider pooled estimation of the relative PPP parameter, which tells if the world economy on average supports the relative PPP to the USD. Studies using cross-country panel data often report such pooled estimates and conclude mixed results. We conjecture that one reason of such mixed results is the severe heterogeneity among the currencies and the conventional pooled estimator could yield misleading results because of some outlying currencies. Table 4 reports several pooled relative PPP estimates. The ML estimator (or the pooled least square estimator) presumes $\beta_i = \beta$ for all i and estimate β;[6] the mean-group (MG) estimator is the equal-weight average of $\widehat{\beta}_i$'s, which estimates the mean of β_i if it exists. We compare them with the DWMG estimators based on the Mahalanobis ($\mathrm{DW_M}$) and the projection ($\mathrm{DW_P}$) depths, where we consider the same weight function $W(d) = d$ and (18) as in the simulations in the previous subsection. The standard errors of the DWMG estimators are calculated as in (15) to construct the t-statistics.[7] In Table 3, we also report

[5] In fact, Engel et al. (2015) and Greenaway-McGrevy et al. (2018) found that a small number of common factors explain more than 50% of the real exchange rate variation and the relative prices or inflations do not provide any meaningful information to predict the future exchange rates.

[6] However, all the panel homogeneity tests we consider strongly reject the homogeneity.

[7] For this panel model, one could instead use the bootstrap method, such as the cross-sectional residual-based bootstrap procedure proposed by Lee et al. (2019).

Table 3: Relative Purchaing Power Parity Estimation of Individual Countries

	individual PPP				weight in %		
	$\widehat{\beta}_i$	s.e.	t-stat $\beta_i = 0$	t-stat $\beta_i = 1$	ML	DW_M^0	DW_P^0
AUS	-1.40	1.08	-1.30	-2.22*	0.68	1.25	1.36
NZ	-1.10	1.64	-0.67	-1.28	0.58	1.66	1.62
HUN	-0.71	0.58	-1.22	-2.95*	2.87	2.46	2.16
SIN	-0.34	0.27	-1.26	-4.96*	2.97	3.57	3.13
GBR	-0.27	0.72	-0.37	-1.76	0.70	3.83	3.44
JPN	-0.26	0.49	-0.54	-2.57*	1.41	3.85	3.48
TWN	-0.22	0.13	-1.71	-9.38*	8.83	4.00	3.70
SWE	-0.19	0.63	-0.30	-1.89*	1.02	4.10	3.86
KOR	-0.14	0.61	-0.23	-1.87*	1.58	4.27	4.17
THA	-0.07	0.29	-0.25	-3.69*	3.32	4.48	4.68
ROM	-0.05	0.36	-0.14	-2.92*	9.99	4.55	4.87
MEX	-0.03	0.60	-0.04	-1.72	1.28	4.62	5.12
PHI	0.04	0.35	0.11	-2.74*	2.67	4.79	5.87
EURO	0.13	0.94	0.13	-0.93	0.34	4.98	7.39
IND	0.17	0.21	0.80	-3.95*	7.49	5.04	6.65
COL	0.17	1.00	0.17	-0.83	1.35	5.05	6.59
CZE	0.18	0.78	0.23	-1.05	1.98	5.06	6.40
SUI	0.29	0.88	0.33	-0.81	0.64	5.13	5.02
POL	0.39	0.95	0.41	-0.64	1.51	5.08	4.20
NOR	0.40	0.58	0.69	-1.03	2.71	5.07	4.13
BRA	0.62	1.34	0.47	-0.28	3.54	4.58	3.02
RSA	0.92	0.88	1.04	-0.09	2.92	3.59	2.24
CAN	1.13	0.64	1.77	0.20	1.36	2.90	1.88
TUR	1.17	0.57	2.08*	0.30	27.38	2.78	1.83
ICE	1.58	0.62	2.54*	0.94	5.39	1.82	1.41
ISR	2.25	0.47	4.82*	2.66*	2.84	0.99	1.03
CHI	3.19	0.68	4.72*	3.22*	2.65	0.51	0.75

Note: * denotes significant at 5% from each t-test. The countries in the table are ordered by the individual relative PPP estimates to the U.S. Dollar. The currencies are (in alphabetical order) of Australia (AUS), Brazil (BRA), Canada (CAN), Chile (CHI), Columbia (COL), the Czech Republic (CZE), the Euro (EUR), Hungary (HUN), Iceland (ICE), India (IND), Israel (ISR), Japan (JPN), Korea (KOR), Mexico (MEX), Norway (NOR), New Zealand (NZL), the Philippines (PHI), Poland (POL), Romania (ROM), Singapore (SIN), South Africa (RSA), Sweden (SWE), Switzerland (SUI), Taiwan (TWN), Thailand (THA), Türkiye (TUR), and the U.K. (GBR).

Table 4: Pooled Relative Purchaing Power Parity Estimation

	ML	MG	$\mathrm{DW}_{\mathrm{P}}^{0}$	$\mathrm{DW}_{\mathrm{P}}^{W}$	$\mathrm{DW}_{\mathrm{M}}^{0}$	$\mathrm{DW}_{\mathrm{M}}^{W}$
$\widehat{\beta}$	0.574	0.291	0.135	0.112	0.168	0.156
s.e.	0.225	0.184	0.154	0.145	0.150	0.145
t-stat ($\overline{\beta}=0$)	2.56	1.58	0.88	0.77	1.12	1.07
t-stat ($\overline{\beta}=1$)	-1.89	-3.84	-5.63	-6.12	-5.56	-5.81

the calculated weights used for ML, $\mathrm{DW}_{\mathrm{M}}^{0}$, and $\mathrm{DW}_{\mathrm{P}}^{0}$.[8]

Before we read Table 4, we first want to show a brief simulation results on the inference based on the DWMG estimators. To this end, we consider the same model (17) with DGP1 as in the previous section with $z_{it} \in \mathbb{R}^1$. Table 5 shows that small contamination in the sample (5% in this case) can heavily affect inferences for ML and MG. On the other hand, DWMG estimators are relative stable though they tend to over reject when the sample gets contaminated. A notable case is when $(n, m) = (25, 200)$, which is a similar sample size of our empirical example in this section. When some contamination presents, ML severely over rejects (0.91), whereas MG under rejects (0.02). The DWMG estimators show rejection probabilities close to the nominal size of 5%.

Now, in Table 4, ML is 0.574 and significantly different from zero, which is rather puzzling since the null of $\beta_i = 0$ was hardly rejected at the individual level in Table 3. We can see that the weight used for ML in Türkiye (TUR) is extremely high (27%), which is resulted from high fluctuation of its inflation rate, and hence its large $\widehat{\beta}_i$ value (1.17) is heavily counted.[9] For MG, it is affected by Chile (CHI) the most, which has the largest $\widehat{\beta}_i$ value (3.19). In comparison, the DWMG estimators properly weight such currencies and we expect that the depth-weighted estimators DW_{M} and DW_{P} better estimate the average effect in this highly heterogeneous environment. Furthermore, from our observations in Table 5, the t-statistcs (for $\overline{\beta} = 0$) of ML and MG are likely over-estimated and under-estimated, respectively, and hence it is hard to make any statistical conclusion using them. Based on these findings, we conclude that the mixed evidence on the relative PPP based on the ML or MG estimators would be because of over-weighting on a few outliers with highly unstable

[8] Recall that we can write $\widehat{\beta}_{ML} = (\sum_{i=1}^{n}\sum_{t=1}^{m} z_{it}^{*} z_{it}^{*\top})^{-1} \sum_{i=1}^{n}\sum_{t=1}^{m} z_{it}^{*} y_{it}^{*} = \sum_{i=1}^{n} \omega_i \widehat{\beta}_i$, where $\widehat{\beta}_i = (\sum_{t=1}^{m} z_{it}^{*} z_{it}^{*\top})^{-1} \sum_{t=1}^{m} z_{it}^{*} y_{it}^{*}$ and hence we define the weight as $\omega_i = (\sum_{i=1}^{n}\sum_{t=1}^{m} z_{it}^{*} z_{it}^{*\top})^{-1} \sum_{t=1}^{m} z_{it}^{*} z_{it}^{*\top}$.

[9] For the sample excluding TUR, ML drops from 0.574 to 0.348; for the sample excluding CHI, MG drops from 0.291 to 0.180.

Table 5: Rejection Probabilities

ϵ	n	m	ML	MG	$\mathrm{DW}_{\mathrm{P}}^{0}$	$\mathrm{DW}_{\mathrm{P}}^{W}$	$\mathrm{DW}_{\mathrm{M}}^{0}$	$\mathrm{DW}_{\mathrm{M}}^{W}$
0%	25	10	0.32	0.06	0.08	0.08	0.07	0.08
	25	50	0.61	0.07	0.07	0.06	0.05	0.05
	25	100	0.73	0.07	0.05	0.06	0.04	0.04
	25	200	0.79	0.07	0.05	0.06	0.04	0.05
	100	10	0.28	0.05	0.08	0.09	0.05	0.07
	100	50	0.59	0.05	0.06	0.06	0.04	0.05
	100	100	0.69	0.05	0.05	0.06	0.04	0.05
	100	200	0.76	0.05	0.05	0.05	0.04	0.05
5%	25	10	0.43	0.02	0.07	0.07	0.08	0.06
	25	50	0.79	0.02	0.05	0.05	0.05	0.04
	25	100	0.88	0.02	0.05	0.05	0.05	0.04
	25	200	0.91	0.02	0.05	0.05	0.05	0.04
	100	10	0.94	0.45	0.15	0.12	0.29	0.18
	100	50	1.00	0.47	0.13	0.09	0.24	0.14
	100	100	1.00	0.49	0.13	0.09	0.24	0.13
	100	200	1.00	0.46	0.11	0.09	0.23	0.13

Note: The estimators are the same as in Tables 1 and 2; MG is the mean-group estimator, which is the equally-weighted average of $\widehat{\beta}_i$'s. ϵ is the level of contamination as in the previous subsection. Each value shows the rejection probability of the t-test at the 5% significance level based on 2000 simulations.

domestic economies. In comparison, depth-weighted estimators automatically impose very low weights on such outlying currencies; they are not significantly different from zero even at the 10% level, which concludes weak evidence for the relative PPP to the USD on average.

5 Concluding Remarks

In this paper, we study the depth weighted mean estimator when we use noisy observations. We employ a drifting asymptotic framework, where the noise e_i is local-to-zero such that $||e_i||^2 = O_p(m^{-1})$ and $\sqrt{n}/m$ does not diverge. We show that the depth weighted mean estimator still follows the standard asymptotic results as Zuo et al. (2004) that use the latent variables without noise. We apply this idea to develop the depth-weighted mean-group estimator, which estimates a centrality of potentially heterogeneous multi-dimensional parameters in longitudinal data models. The new estimator shows promising finite sample performance whether the true coefficient is heterogeneous or homogeneous, in that it is robust to outlying behaviors of heterogeneous agents in longitudinal data. This estimator can be hence useful to estimate average partial or treatment effects in practice when the effects are more likely to be heterogeneous across agents.

The weight function $W(\cdot)$ considered in this paper is smooth and it does not completely eliminate the outliers. Similar to Zuo (2006), we can consider the ϖ-trimmed depth-weighted mean estimator,

$$L_{\varpi}\left(F_n^e\right) = \frac{\sum_{i=1}^n X_i^e W(\mathcal{D}(X_i^e, F_n^e)) 1\{\mathcal{D}(X_i^e, F_n^e) \geq \varpi\}}{\sum_{i=1}^n W(\mathcal{D}(X_i^e, F_n^e)) 1\{\mathcal{D}(X_i^e, F_n^e) \geq \varpi\}}$$

for some $\varpi \in (0, \max_{1 \leq i \leq n} \mathcal{D}(X_i^e, F_n^e))$. When $\varpi = 0$, it is simply $L\left(F_n^e\right)$; when $W(\cdot) = 1$, it is the ϖ-trimmed mean where trimming is based on the data depth. In a similar vein, Lee and Sul (2022a) considers trimmed mean-group estimator for panel data regression and Lee and Sul (2022b) applies trimmed depth-weighted mean to develop a robust forecasting combination.

A Appendix

A.1 Proofs

Proof of Lemma 1 We denote F_{X,e_i}, F, and F_{e_i} be the joint distribution of (X_i, e_i), the marginal distribution of X_i, and the marginal distribution of e_i, respectively. The noise e_i can be non-identically distributed over i and hence we write the marginal distribution F_{e_i} instead of F_e. For each i, the distribution function of X_i^e can be expressed as

$$\begin{aligned}
\mathbb{P}\left\{X_i^e \leq x\right\} &= \int\int 1\left\{v_1 + v_2 \leq x\right\} F_{X|e_i}(dv_1|v_2) F_{e_i}(dv_2) \\
&= \int \left[\int 1\left\{v_1 \leq x - v_2\right\} F(dv_1)\right] F_{e_i}(dv_2) \\
&= \int F\left(x - v_2\right) F_{e_i}(dv_2), \qquad \text{(A.1)}
\end{aligned}$$

where the second equality is because e_i is assumed independent of X_i in Assumption (A2). Now, for a given $x \in \mathcal{X}$, we write

$$\begin{aligned}
F_n^e(x) - F(x) &= \frac{1}{n}\sum_{i=1}^{n}\left(1\{X_i^e \leq x\} - F(x)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\left(1\{X_i^e \leq x\} - 1\{X_i \leq x\}\right) - \left(\mathbb{E}\left[1\left\{X_i^e \leq x\right\}\right] - F(x)\right)\right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left(1\{X_i \leq x\} - F(x)\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\int F\left(x - y_2\right) F_{e_i}(dy_2) - F(x)\right) \\
&\equiv A_{n,1}(x) + A_{n,2}(x) + A_{n,3}(x),
\end{aligned}$$

where the expression of $A_{n,3}(x)$ is from (A.1).

First, since the binary indicator function is Donsker and $||X_i^e - X_i|| \to_p 0$ as $m \to \infty$ for each i from Assumption (A1), $\sup_{x\in\mathcal{X}} |A_{n,1}(x)| = o_p(n^{-1/2})$ as $n, m \to \infty$ by Theorem 2.1 of van der Vaart and Wellner (2007). Second, $\sup_{x\in\mathcal{X}} |A_{n,2}(x)| = O_p(n^{-1/2})$ as $n \to \infty$ by the standard results (e.g., Dvoretzky et al. (1956); Kiefer (1961)). Finally, we denote $\dot{F}(x) = \nabla F(x)$ and $\ddot{F}(x) = \nabla^2 F(x)$. From Assumptions (A1) and (A3), $F\left(x - e_i\right) - F(x) = -e_i^\top \dot{F}(x) + (1/2) e_i^\top \ddot{F}(x_i^*) e_i = -e_i^\top \dot{F}(x) + (1/2) e_i^\top \ddot{F}(x) e_i + o_p(m^{-1})$ for some x_i^* between $x - e_i$ and x, where $\sup_{x\in\mathcal{X}} ||\ddot{F}(x_i^*) - \ddot{F}(x)|| = o_p(1)$. It follows that

$$A_{n,3}(x) = \frac{1}{n}\sum_{i=1}^{n}\int \left\{F\left(x - v\right) - F(x)\right\} F_{e_i}(dv)$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\int\left\{-v^\top\dot{F}(x)+\frac{1}{2}v^\top\ddot{F}(x)v+o_p\left(m^{-1}\right)\right\}F_{e_i}(dv)\\
&= \frac{1}{2n}\sum_{i=1}^{n}tr[\ddot{F}(x)\mathbb{E}\left[e_ie_i^\top\right]]+o_p\left(\frac{1}{m}\right)\\
&= \frac{1}{2m}tr\left[\ddot{F}(x)\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\varepsilon_i\varepsilon_i^\top\right]\right]+o_p\left(\frac{1}{m}\right), \qquad \text{(A.2)}
\end{aligned}
$$

where $\mathbb{E}[e_i]=0$. Since $tr[\ddot{F}(x)n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\varepsilon_i\varepsilon_i^\top\right]]\to tr[\ddot{F}(x)\Omega]$ as $n\to\infty$, which is assumed to be bounded uniformly in $x\in\mathcal{X}$, $\sup_{x\in\mathcal{X}}|A_{n,3}(x)|=O_p(m^{-1})$. The uniform consistency follows by combining these three results. The asymptotic normality is also verified since $\sqrt{n}A_{n,2}(x)\to_d\mathcal{N}(0,F(x)(1-F(x))$ as $n\to\infty$ by FCLT (e.g., p.57 in Serfling (1980)) and $\sqrt{n}A_{n,3}(x)=(\sqrt{n}/m)tr[\ddot{F}(x)\Omega]/2+o_p(\sqrt{n}/m)$ from (A.2). ■

Proof of Theorem 1 We prove the second result; the proof for the consistency is similar and omitted. For further details, see Massé (2004, Proposition 3.1). We let

$$
\nu_n^e(\cdot)=\sqrt{n}(F_n^e(\cdot)-F(\cdot)) \quad \text{and} \quad H_n^e(\cdot)=\sqrt{n}(\mathcal{D}(\cdot,F_n^e)-\mathcal{D}(\cdot,F)).
$$

We first observe that

$$
\begin{aligned}
Q_n &\equiv \sqrt{n}\int(x-L(F))W(\mathcal{D}(x,F_n^e))F_n^e(dx)-\sqrt{n}\int(x-L(F))W(\mathcal{D}(x,F))F(dx)\\
&= \int(x-L(F))\dot{W}(\delta_n^e(x))H_n^e(x)F_n^e(dx)+\int(x-L(F))W(\mathcal{D}(x,F))\nu_n^e(dx)\\
&\equiv Q_{n,1}+Q_{n,2} \qquad \text{(A.3)}
\end{aligned}
$$

for some $\delta_n^e(x)$ between $\mathcal{D}(x,F_n^e)$ and $\mathcal{D}(x,F)$, where $\sup_{x\in\mathcal{X}}|\delta_n^e(x)-\mathcal{D}(x,F)|=O_p(n^{-1/2}+m^{-1})$. Note that

$$
\begin{aligned}
Q_{n,1} &= \int(x-L(F))\dot{W}(\delta_n^e(x))H_n^e(x)F_n(dx)\\
&\quad +\sqrt{n}\int(x-L(F))\dot{W}(\delta_n^e(x))\left\{\mathcal{D}(x,F_n^e)-\mathcal{D}(x,F)\right\}\left(F_n^e-F_n\right)(dx)\\
&\equiv Q_{n,11}+Q_{n,12},
\end{aligned}
$$

where

$$
Q_{n,11} = \int(x-L(F))\dot{W}(\mathcal{D}(x,F))H_n^e(x)F(dx)+o_p(1)
$$

$$
\begin{aligned}
&= \int_{\mathcal{X}_{d_0}} (x - L(F))\dot{W}(\mathcal{D}(x,F)) \left\{ \int h(x,y)\nu_n^e(dy) \right\} F(dx) + o_p(1) \\
&= \iint (y - L(F))\dot{W}(\mathcal{D}(y,F))h(y,x)F(dy)\nu_n^e(dx) + o_p(1)
\end{aligned}
$$

using the same argument as the proof of Theorem 2.1 in Zuo et al. (2004) under Assumptions (A1)-(A8). However,

$$
\begin{aligned}
\|Q_{n,12}\| &\leq \sup_{x\in\mathcal{X}} \left| \dot{W}(\delta_n^e(x)) \right| \sup_{x\in\mathcal{X}} |\mathcal{D}(x,F_n^e) - \mathcal{D}(x,F)| \left\| \sqrt{n} \int (x - L(F)) \left(F_n^e - F_n\right)(dx) \right\| \\
&= O_p\left(n^{-1/2} + m^{-1}\right) \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \{X_i^e - X_i\} \right\| = o_p(1),
\end{aligned}
$$

since $X_i^e - X_i = m^{-1/2}\varepsilon_i$. Therefore, by putting this expression into (A.3), we have

$$
Q_n = \int \left\{ \int (y - L(F))\dot{W}(\mathcal{D}(y,F))h(y,x)F(dy) + xW(\mathcal{D}(x,F)) \right\} \nu_n^e(dx) + o_p(1). \quad \text{(A.4)}
$$

Likewise, we can also show that

$$
\int W(\mathcal{D}(x,F_n^e))F_n^e(dx) = \int W(\mathcal{D}(x,F))F(dx) + o_p(1). \quad \text{(A.5)}
$$

From (A.4) and (A.5), therefore,

$$
\begin{aligned}
\sqrt{n}(L(F_n^e) - L(F)) &= \frac{Q_n}{\int W(\mathcal{D}(x,F))F(dx)} + o_p(1) \\
&= \frac{\int \Lambda(x,F)\sqrt{n}(F_n - F)(dx)}{\int W(\mathcal{D}(x,F))F(dx)} + \frac{\int \Lambda(x,F)\sqrt{n}(F_n^e - F_n)(dx)}{\int W(\mathcal{D}(x,F))F(dx)} + o_p(1) \\
&\equiv \Phi_{n,1} + \Phi_{n,2} + o_p(1), \quad \text{(A.6)}
\end{aligned}
$$

where $\Lambda(x,F) = \int (y - L(F))\dot{W}(\mathcal{D}(y,F))h(y,x)F(dy) + (x - L(F))W(\mathcal{D}(x,F))$ for $x \in \mathcal{X}$.

The first term $\Phi_{n,1}$ in (A.6) satisfies

$$
\Phi_{n,1} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(K_F(X_i) - \mathbb{E}[K_F(X_i)]\right) \to_d \mathcal{N}\left(0, \mathbb{E}[K_F^0(X_i)K_F^0(X_i)^\top]\right)
$$

from Theorem A of Serfling (1980; p.226), where $K_F(x) = \Lambda(x,F)/\int W(\mathcal{D}(y,F))F(dy)$. Since $\int (x - L(F))W(\mathcal{D}(x,F))F(dx)/\int W(\mathcal{D}(x,F))F(dx) = 0$ by construction,

$$
\begin{aligned}
K_F^0(x) &= K_F(x) - \mathbb{E}[K_F(X_i)] \\
&= \frac{\int (y - L(F))\dot{W}(\mathcal{D}(y,F))h^0(y,x)F(dy) + (x - L(F))W(\mathcal{D}(x,F))}{\int W(\mathcal{D}(y,F))F(dy)}
\end{aligned}
$$

with $h^0(y,x) = h(y,x) - \int h(y,x')F(dx')$. For the second term $\Phi_{n,2}$ in (A.6), we note that

$$\begin{aligned} & \left\| \int (x - L(F))W(\mathcal{D}(x,F))\sqrt{n}(F_n^e - F_n)(dx) \right\| \\ \leq & \sup_{x\in\mathcal{X}} |W(\mathcal{D}(x,F))| \left\| \int (x - L(F))\sqrt{n}(F_n^e - F_n)(dx) \right\| \\ = & C \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{X_i^e - X_i\} \right\| = O_p\left(m^{-1/2}\right) \end{aligned}$$

for some $C < \infty$ and

$$\begin{aligned} & \mathbb{E} \left\| \iint (y - L(F))\dot{W}(\mathcal{D}(y,F))h(y,x)F(dy)\sqrt{n}(F_n^e - F_n)(dx) \right\|^2 \\ \leq & \sup_{y\in\mathcal{X}} \mathbb{E}\left|h(y,X_i^e) - h(y,X_i)\right|^2 \int \|y - L(F)\|^2 \dot{W}(\mathcal{D}(y,F))^2 F(dy) = O\left(m^{-1}\right) \end{aligned}$$

from Assumptions (A6) and the condition $\sup_{y\in\mathcal{X}} \mathbb{E}\left|h(y,x_1) - h(y,x_2)\right| = O(||x_1 - x_2||)$. It follows that $\Phi_{n,2} \to_p 0$ as $n, m \to \infty$, which yields the desired result. ■

Proof of Corollary 2 By the affine invariance property of the statistical depth function (e.g., Zuo and Serfling (2000)), we have

$$\mathcal{D}(\widehat{\beta}_i, F_n^e) = \mathcal{D}(\widehat{\xi}_i, G_n^e) \tag{A.7}$$

from (16). Therefore, we can write

$$\widehat{\beta}_{DW} - \beta = \frac{\int (b-\beta)W(\mathcal{D}(b,F_n^e))F_n^e(db)}{\int W(\mathcal{D}(b,F_n^e))F_n^e(db)} = \frac{m^{-1/2}\int rW(\mathcal{D}(r,G_n^e))G_n^e(dr)}{\int W(\mathcal{D}(r,G_n^e))G_n^e(dr)} \tag{A.8}$$

from (A.7) and by the change of variables with $r = m^{1/2}(b-\beta)$.

We let $\gamma_n^e(\cdot) = \sqrt{n}(G_n^e(\cdot) - G(\cdot))$ and $M_n^e(\cdot) = \sqrt{n}(\mathcal{D}(\cdot, G_n^e) - \mathcal{D}(\cdot, G))$. Similarly to the proof of Theorem 1, we have

$$\begin{aligned} & \sqrt{n}\int rW(\mathcal{D}(r,G_n^e))G_n^e(dr) \qquad (A.9) \\ = & \int \left\{ \int s\dot{W}(\mathcal{D}(s,G))\eta(s,r)G(ds) + rW(\mathcal{D}(r,G)) \right\} \gamma_n^e(dr) + o_p(1) \end{aligned}$$

and

$$\int W(\mathcal{D}(r,G_n^e))G_n^e(dr) = \int W(\mathcal{D}(r,G))G(dr) + o_p(1). \tag{A.10}$$

By combining (A.8), (A.9), and (A.10), we thus have

$$
\begin{aligned}
& \sqrt{nm}(\widehat{\beta}_{DW} - \beta) \\
= & \frac{\sqrt{n}\int rW(\mathcal{D}(r, G_n^e))G_n^e(dr)}{\int W(\mathcal{D}(r, G_n^e))G_n^e(dr)} \\
= & \frac{\int \left\{ \int s\dot{W}(\mathcal{D}(s, G))\eta(s, r)G(ds) + rW(\mathcal{D}(r, G)) \right\} \sqrt{n}(G_n^e - G)(dr)}{\int W(\mathcal{D}(r, G))G(dr)} + o_p(1).
\end{aligned}
$$

The desired result follows using the same argument as the proof of Theorem 1 because

$$
\sqrt{nm}(\widehat{\beta}_{DW} - \beta) \to_d \mathcal{N}\left(0, \mathbb{E}[K_G^0(\xi_i)K_G^0(\xi_i)^\top]\right)
$$

as $n, m \to \infty$, where

$$
K_G^0(r) = \frac{\int s\dot{W}(\mathcal{D}(s, G))\eta^0(s, r)G(dr) + rW(\mathcal{D}(r, G))}{\int W(\mathcal{D}(s, G))G(ds)}.
$$

■

A.2 Examples of depth and $h(\cdot, \cdot)$ estimation

$h(\cdot, \cdot)$ in (6) is determined based on the choice of a specific statistical depth function $\mathcal{D}(\cdot, \cdot)$. For the Mahalanobis and the projection depth functions, $h(\cdot, \cdot)$ in (6) corresponds to what are given in Zuo et al. (2004). We summarize their specific forms and the estimators $\widehat{h}(\cdot, \cdot)$. We also summarize computation steps how to estimate $\mathcal{D}(\cdot, \cdot)$ for each case.

Mahalanobis depth The Mahalanobis depth is defined as

$$
\mathcal{D}^M(x; F) = \frac{1}{1 + (x - \mu(F))^\top \Sigma(F)^{-1} (x - \mu(F))},
$$

where $\mu(F)$ and $\Sigma(F)$ are some location and scale parameters of F. Suppose $\mu(F)$ and $\Sigma(F)$ satisfy $\sqrt{n}(\mu(F_n^e) - \mu(F)) = n^{-1/2}\sum_{i=1}^n h_1(X_i^e) + o_p(1)$ and $\sqrt{n}(\Sigma(F_n^e) - \Sigma(F)) = n^{-1/2}\sum_{i=1}^n h_2(X_i^e) + o_p(1)$ for some $h_1(x)$ and $h_2(x)$ with $\int h_1(x)F(dx) = \int h_2(x)F(dx) = 0$.

For instance, we have $h_1(x) = x - \mu(F)$ for the mean and $h_2(x) = (x - \mu(F))(x - \mu(F))^\top - \Sigma(F)$ for the variance. Then, as in Example 2.3 of Zuo et al. (2004), we have

$$
h(y, x) = \frac{2(y - \mu(F))^\top \Sigma(F)^{-1} h_1(x) + (y - \mu(F))^\top \Sigma(F)^{-1} h_2(x) \Sigma(F)^{-1} (y - \mu(F))}{(1 + (y - \mu(F))^\top \Sigma(F)^{-1} (y - \mu(F)))^2}.
$$

For estimation, we let the sample mean $\mu(F_n^e) = \widehat{\mu}^e = n^{-1}\sum_{i=1}^n X_i^e$ and the sample variance $\Sigma(F_n^e) = \widehat{\Sigma}^e = (n-k)^{-1}\sum_{i=1}^n (X_i^e - \widehat{\mu}^e)(X_i^e - \widehat{\mu}^e)^\top$ of X_i^e. Then, the sample Mahalanobis depth of X_i^e is obtained as

$$\mathcal{D}^M(X_i^e, F_n^e) = \frac{1}{1 + (X_i^e - \widehat{\mu}^e)^\top (\widehat{\Sigma}^e)^{-1}(X_i^e - \widehat{\mu}^e)}.$$

Furthermore, for each $i, j = 1, \ldots, n$, we can obtain

$$\widehat{h}(X_j^e, X_i^e) = \frac{2\widetilde{X}_j^{e\top}(\widehat{\Sigma}^e)^{-1}\widetilde{X}_i^e + \widetilde{X}_j^{e\top}\widehat{\Sigma}^{-1}(\widetilde{X}_i^e\widetilde{X}_i^{e\top} - \widehat{\Sigma}^e)(\widehat{\Sigma}^e)^{-1}\widetilde{X}_j^e}{[1 + \widetilde{X}_j^{e\top}(\widehat{\Sigma}^e)^{-1}\widetilde{X}_j^e]^2},$$

where $\widetilde{X}_i^e = X_i^e - \mu(F_n^e)$ for each i.

Projection depth The projection depth is defined as

$$\mathcal{D}^P(x; F) = \frac{1}{1 + \sup_{v:||v||=1}\{|v^\top x - \mu(F; v)| / \sigma(F; v)\}},$$

where v is a $k \times 1$ nonrandom vector with $||v|| = 1$ and $\mu(F; v)$ and $\sigma(F; v)$ are some location and scale parameters of the distribution of $v^\top X_i$. When $v^\top x - \mu(F_{v^\top X}) = 0$ and $\sigma(F_{v^\top X}) = 0$, we let $\sup_{v:||v||=1}\{\left|v^\top x - \mu(F_{v^\top X})\right| / \sigma(F_{v^\top X})\} = 0$ and hence $\mathcal{D}^p(x; F) = 1$. Suppose $\mu(F; v)$ and $\sigma(F; v)$ satisfy $\sqrt{n}(\mu(F_n^e; v) - \mu(F; v)) = n^{-1/2}\sum_{i=1}^n h_1(X_i^e; v) + o_p(1)$ and $\sqrt{n}(\Sigma(F_n^e; v) - \Sigma(F; v)) = n^{-1/2}\sum_{i=1}^n h_2(X_i^e; v) + o_p(1)$ uniformly in v for some $h_1(b; v)$ and $h_2(b; v)$, where they satisfy $\mathbb{E}[h_j(X_i^e; v)] = 0$, $\mathbb{E}[\sup_{||v||=1} h_j^2(X_i^e; v)] < \infty$, and $\mathbb{E}[\sup_{||v_1 - v_2|| \le \delta, ||v_1|| = ||v_2|| = 1} |h_j(X_i^e; v_1) - h_j(X_i^e; v_2)|^2] \to 0$ as $\delta \to 0$ for $j = 1, 2$.

For instance, for the median $\mu(F; v)$ and the median absolute deviation (MAD) $\sigma(F; v)$, we have

$$\begin{aligned} h_1(x; v) &= \frac{\sqrt{v^\top \Sigma(F) v}}{p(0)}\left(\frac{1}{2} - 1\left\{v^\top(x - L(F)) \le 0\right\}\right), \\ h_2(x; v) &= \frac{\sqrt{v^\top \Sigma(F) v}}{2p(\varphi_0)}\left(\frac{1}{2} - 1\left\{|v^\top(x - L(F))| \le \varphi_0\sqrt{v^\top \Sigma(F) v}\right\}\right). \end{aligned}$$

from Lemma 3.2 of Zuo et al. (2004), where $\Sigma(F)$ is some positive definite matrix such that $(v^\top \Sigma(F) v)^{-1/2} v^\top (X_i^e - L(F))$ is a univariate symmetric variable; $p(\cdot)$ is the density function of $(v^\top \Sigma(F) v)^{-1/2} v^\top (X_i^e - L(F))$ and φ_0 is its MAD, satisfying $p(0) > 0$ and $p(\varphi_0) > 0$.

Then, from Theorem 3.1 of Zuo et al. (2004), we have

$$h\left(y,x\right)=\frac{h_1(x;v^*(y))+\mathcal{O}^p(y,F)h_2(x;v^*(y))}{\sigma(F;v^*(y))\left(1+\mathcal{O}^p(y,F)\right)^2},$$

where $\mathcal{O}^p(y,F)=\sup_{v:||v||=1}\{\left|v^\top y-\mu(F;v)\right|/\sigma(F;v)\}$ and $v^*(y)$ is such that $\mathcal{O}^p(y,F)=\left|v^*(y)^\top y-\mu(F;v^*(y))\right|/\sigma(F;v^*(y))$.

For estimation of the sample projection depth $\mathcal{D}^P(X_i^e,F_n^e)$ and $\widehat{h}(X_j^e,X_i^e)$ for each $i,j=1,\ldots,n$, we take the following steps:

1. We generate v_r from a k-dimensional multivariate standard normal for $r=1,\ldots,R$. For each r, redefine v_r as $v_r/(v_r^\top v_r)^{1/2}$ so that $\|v_r\|=1$. Recall that since the standard normal density function is rotationally symmetric, standard-normal-distributed random coordinates yields a uniform distribution of directions and hence it generates random points on the surface of the unit circle.

2. For each $r=1,\ldots,R$, we let $\mu(F_n^e;v)=\mathsf{med}_{1\le\ell\le n}\{v_r^\top X_\ell^e\}$ be the sample median and $\sigma(F_n^e;v)=\mathsf{MAD}_{1\le\ell\le n}\{v_r^\top X_\ell^e\}=\mathsf{med}_{1\le\ell\le n}\{|v_r^\top X_\ell^e-\mathsf{med}_{1\le\ell'\le n}\{v_r^\top X_{\ell'}^e\}|\}$ be the sample MAD of $v_r^\top X_i^e$.

3. For each $i=1,\ldots,n$, we find $v^*(i)$ such that

$$v^*(i)=\underset{v_r:1\le r\le R}{\arg\max}\frac{\left|v_r^\top X_i^e-\mu(F_n^e;v)\right|}{\sigma(F_n^e;v)}.$$

   Then, the sample depth of X_i^e is defined as

$$\mathcal{D}^P(X_i^e,F_n^e)=\frac{1}{1+\mathcal{O}^p(X_i^e,F_n^e)}\quad\text{with}\quad\mathcal{O}^p(X_i^e,F_n^e)=\frac{\left|v^*(i)^\top X_i^e-\mu(F_n^e;v^*(i))\right|}{\sigma(F_n^e;v^*(i))}.$$

4. For each $j,\ell=1,\ldots,n$, we let

$$\zeta_\ell(j)=\left[v^*(j)^\top\widehat{\Sigma}^e v^*(j)\right]^{-1/2}v^*(j)^\top\left(X_\ell^e-L(F_n^e)\right),$$

   where $\widehat{\Sigma}^e=\Sigma(F_n^e)$ is the sample variance of X_i^e as defined in the Mahalanobis depth above. Furthermore, we define

$$\begin{aligned}\widehat{\varphi}_0(j)&=\mathsf{MAD}_{1\le\ell\le n}\{\zeta_\ell(j)\}\\ \widehat{p}(j;x)&=\frac{1}{n\varrho}\sum_{\ell=1}^n\Upsilon\left(\frac{\zeta_\ell(j)-x}{\varrho}\right)\end{aligned}$$

for some kernel function $\Upsilon(\cdot)$ and a bandwidth ϱ that satisfies the conventional conditions for the consistent kernel density estimator.

5. Then, for each $i, j = 1, \ldots, n$, we can obtain

$$\widehat{h}(X_j^e, X_i^e) = \frac{\widehat{h}_1(X_i^e; v^*(j)) + \mathcal{O}^p(X_j^e, F_n^e)\widehat{h}_2(X_i^e; v^*(j))}{\sigma(F_n^e; v^*(j))\left(1 + \mathcal{O}^p(X_j^e, F_n^e)\right)^2},$$

where

$$\begin{aligned}
\widehat{h}_1(X_i^e; v^*(j)) &= \frac{\sqrt{v^*(j)^\top \widehat{\Sigma}^e v^*(j)}}{\widehat{p}(j;0)}\left(\frac{1}{2} - 1\left\{v^*(j)^\top (X_i^e - L(F_n^e)) \leq 0\right\}\right) \\
\widehat{h}_2(X_i^e; v^*(j)) &= \frac{\sqrt{v^*(j)^\top \widehat{\Sigma}^e v^*(j)}}{2\widehat{p}(j;\widehat{\varphi}_0(j))}\left(\frac{1}{2} - 1\left\{|\zeta_i(j)| \leq \widehat{\varphi}_0(j)\right\}\right).
\end{aligned}$$

References

Bai, J. (2009). Panel data models with interactive fixed effects, *Econometrica*, 77, 1229-1279.

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 181-221.

Boneva, L. and O. Linton (2016). A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance, *Journal of Applied Econometrics*, 32, 1226-1243.

Donoho, D.L. (1982). Breakdown properties of multivariate location estimators, Qualifying paper, Harvard University.

Donoho, D.L. and M. Gasko (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness, *Annals of Statistics*, 20, 1803-1827.

Dümbgen, L. (1992). Limit theorems for the simplicial depth, *Statistics and Probability Letters*, 14, 119-128.

Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *Annals of Mathematical Statistics*, 27, 642-669.

Engel, C., N.C. Mark, and K.D. West (2015). Factor Model Forecasts of Exchange Rates, *Econometric Reviews*, 34, 32-55.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models, *Econometrica*, 72, 1295-1319.

Harding, M., C. Lamarche, and M.H. Pesaran (2020). Common correlated effects estimation of heterogeneous dynamic panel quantile regression models, *Journal of Applied Econometrics*, Forthcoming.

Hsiao, C. and M.H. Pesaran (2008). Random Coefficients Models, in (eds.) L. Matayas and P. Sevestre, *The Econometrics of Panel Data*, 3rd ed., Kluwer, 187-216.

Huber, P.J. and E.M. Ronchetti (2009). *Robust Statistics*, 2nd ed., Wiley.

Kiefer, J. (1961). On large deviations of the empiric D.F. of vector chance variables and a law of the iterated logarithm, *Pacific Journal of Mathematics*, 11, 649-660.

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 (2022a). Trimmed mean group estimation, in (eds.) A. Chudik, C. Hsiao, and A. Timmermann, *Essays in Honor of M. Hashem Pesaran: Panel Modeling, Micro Applications, and Econometric Methodology (Advances in Econometrics, Vol. 43B)*, Emerald Publishing Limited, Bingley, 177-202.

Lee, Y. and D. Sul (2022b). Depth-Weighted Forecast Combination: Application to COVID-19 Cases, *Advances in Econometrics*, forthcoming.

Liu, R.Y. (1990). On a notation of data depth based on random simplices, *Annals of Statistics*, 18, 405-414.

Liu, R.Y. (1992). Data depth and multivariate rank tests, in (ed.) Y. Dodge, *L1-Statistical Analysis and Related Methods*, North-Holland, Amsterdam, 279-294.

Liu, R.Y., J.M. Parelius, and K. Singh (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion), *Annals of Statistics*, 27, 783-858.

Liu, R.Y. and K. Singh (1993). A quality index based on data depth and multivariate rank tests, *Journal of the American Statistical Association*, 88 252-260.

Maronna, R.A. and V.J. Yohai (1995). The behavior of the Stahel–Donoho robust multivariate estimator, *Journal of the American Statistical Association*, 90, 330-341.

Massé, J.-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean, *Bernoulli*, 10, 397-419.

Pesaran, M.H. (2006). Estimation and inference in large heterogenous panels with a multi-factor 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.

Ryan Greenaway-McGrevy, R, N.C. Mark, D. Sul, and J.-L. Wu (2018). Identifying Exchange Rates Common Factors, *International Economic Review*, 59, 2193-2218.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Stahel, W.A. (1981). Robust estimation: Infinitesimal optimality and covariance matrix estimators, Ph.D. thesis, ETH, Zurich (in German).

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

Tukey, J.W. (1975). Mathematics and the picturing data, *Proceedings of the International Congress of Mathematicians*, 2, Canadian Mathematical Congress, Montreal, 523-531.

van der Vaart, A.W. and J.A. Wellner (2007). Empirical processes indexed by estimated functions. *Asymptotics: Particles, Processes and Inverse Problems.* Institute of Mathematical Statistics Lecture Notes - Monograph Series 55, 234-252.

Zuo, Y. (2003). Projection-based depth functions and associated medians, *Annals of Statistics*, 31, 1460-1490.

Zuo, Y. (2006). Multidimensional trimming based on projection depth, *Annals of Statistics*, 34, 2211-2251.

Zuo, Y., H. Cui, and X. He (2004). On the Stahel-Donoho estimator and depth-weighted mean of multivariate data, *Annals of Statistics*, 32, 167-188.

Zuo, Y. and R. Serfling (2000). General notions of statistical depth function, *Annals of Statistics*, 28, 461-482.