The Theil-Sen Estimators In Linear Regression

Hanxiang Peng

Department of Mathematics
University of Mississippi

September 15, 2008

Outline

The Theil-Sen Estimator In Simple Linear Regression

- Simple Linear Regression

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where x_i's are nonrandom covariates, ϵ_i's are IID random errors with common cdf F, and α, β are parameters.

The Theil-Sen Estimator In Simple Linear Regression

- Simple Linear Regression

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where x_i's are nonrandom covariates, ϵ_i's are IID random errors with common cdf F, and α, β are parameters.

- Theil-Sen Estimator (Theil(1950), Sen(1968, JASA)):

$$\hat{\beta}_n = \mathrm{Median}\left\{\frac{Y_i - Y_j}{x_i - x_j} : x_i \neq x_j, i < j = 1, \cdots, n\right\}$$

The Theil-Sen Estimator In Simple Linear Regression

- Simple Linear Regression

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where x_i's are nonrandom covariates, ϵ_i's are IID random errors with common cdf F, and α, β are parameters.

- Theil-Sen Estimator (Theil(1950), Sen(1968, JASA)):

$$\hat{\beta}_n = \mathrm{Median}\left\{\frac{Y_i - Y_j}{x_i - x_j} : x_i \neq x_j, i < j = 1, \cdots, n\right\}$$

- Robustness: Breaking-down Point $BP = 0.293$–Global and Bounded Influence Function–Local.

The Theil-Sen Estimator In Simple Linear Regression

- Simple Linear Regression

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where x_i's are nonrandom covariates, ϵ_i's are IID random errors with common cdf F, and α, β are parameters.

- Theil-Sen Estimator (Theil(1950), Sen(1968, JASA)):

$$\hat{\beta}_n = \mathrm{Median}\left\{\frac{Y_i - Y_j}{x_i - x_j} : x_i \neq x_j, i < j = 1, \cdots, n\right\}$$

- Robustness: Breaking-down Point $BP = 0.293$–Global and Bounded Influence Function–Local.
- Compete favorably with LSE (Wilcox, 1998).

Strong Consistency (Peng, Wang & Wang, 2008, JSPI)

Theorem 1 Suppose non-random covariates $x_1, \cdots, x_n$ satisfy

$$\frac{n^{-1}\log n}{\bar{a}_n^2} = o(1), \quad \text{where } \bar{a}_n = \binom{n}{2}^{-1} \sum_{i<j} \mathbf{1}[x_i \neq x_j].$$

Strong Consistency (Peng, Wang & Wang, 2008, JSPI)

Theorem 1 Suppose non-random covariates $x_1, \cdots, x_n$ satisfy

$$\frac{n^{-1}\log n}{\bar{a}_n^2} = o(1), \quad \text{where } \bar{a}_n = \binom{n}{2}^{-1} \sum_{i<j} \mathbf{1}[x_i \neq x_j].$$

(i) If F is discontinuous, then

$$\mathbb{P}(\omega : \hat{\beta}_n \neq \beta_0 \text{ happens only finite many times}) = 1.$$

Strong Consistency (Peng, Wang & Wang, 2008, JSPI)

Theorem 1 Suppose non-random covariates $x_1, \cdots, x_n$ satisfy

$$\frac{n^{-1}\log n}{\bar{a}_n^2} = o(1), \quad \text{where } \bar{a}_n = \binom{n}{2}^{-1} \sum_{i<j} \mathbf{1}[x_i \neq x_j].$$

(i) If F is discontinuous, then

$$\mathbb{P}(\omega : \hat{\beta}_n \neq \beta_0 \text{ happens only finite many times}) = 1.$$

Hence, the TSE is *strongly consistent*: $\mathbb{P}(\lim_{n\to\infty} \hat{\beta}_n = \beta_0) = 1$.

Strong Consistency (Peng, Wang & Wang, 2008, JSPI)

Theorem 1 Suppose non-random covariates $x_1, \cdots, x_n$ satisfy

$$\frac{n^{-1}\log n}{\bar{a}_n^2} = o(1), \quad \text{where } \bar{a}_n = \binom{n}{2}^{-1} \sum_{i<j} \mathbf{1}[x_i \neq x_j].$$

(i) If F is discontinuous, then

$$\mathbb{P}(\omega : \hat{\beta}_n \neq \beta_0 \text{ happens only finite many times}) = 1.$$

Hence, the TSE is *strongly consistent*: $\mathbb{P}(\lim_{n\to\infty} \hat{\beta}_n = \beta_0) = 1$.

(ii) If F is continuous and

$$\liminf_{n\to\infty} \{|x_i - x_j| : x_i \neq x_j : i < j\} > 0,$$

then the TSE is strongly consistent.

Asymptotic Distribution(Peng, Wang & Wang, 2008, JSPI)

Theorem 2 Case I. Suppose F is discontinuous. Then

$$\mathbb{P}(n^{\nu}(\hat{\beta}_n - \beta_0) \to 0) = 1, \quad \nu \geq 0.$$

Asymptotic Distribution(Peng, Wang & Wang, 2008, JSPI)

Theorem 2 Case I. Suppose F is discontinuous. Then

$$\mathbb{P}(n^{\nu}(\hat{\beta}_n - \beta_0) \to 0) = 1, \quad \nu \geq 0.$$

Hence $\hat{\beta}_n$ is *super-efficient*.

Asymptotic Distribution(Peng, Wang & Wang, 2008, JSPI)

Theorem 2 Case I. Suppose F is discontinuous. Then

$$\mathbb{P}(n^{\nu}(\hat{\beta}_n - \beta_0) \to 0) = 1, \quad \nu \geq 0.$$

Hence $\hat{\beta}_n$ is *super-efficient*.

Cae II. Suppose F is continuous.

Asymptotic Distribution(Peng, Wang & Wang, 2008, JSPI)

Theorem 2 Case I. Suppose F is discontinuous. Then

$$\mathbb{P}(n^{\nu}(\hat{\beta}_n - \beta_0) \to 0) = 1, \quad \nu \geq 0.$$

Hence $\hat{\beta}_n$ is *super-efficient*.

Cae II. Suppose F is continuous.
Let $C_n^2 = \sum c_i^2$ where $c_i = \sum \mathbf{1}[x_j > x_i] - \mathbf{1}[x_j < x_i]$.

Asymptotic Distribution(Peng, Wang & Wang, 2008, JSPI)

Theorem 2 Case I. Suppose F is discontinuous. Then

$$\mathbb{P}(n^{\nu}(\hat{\beta}_n - \beta_0) \to 0) = 1, \quad \nu \geq 0.$$

Hence $\hat{\beta}_n$ is *super-efficient*.

Cae II. Suppose F is continuous.
Let $C_n^2 = \sum c_i^2$ where $c_i = \sum \mathbf{1}[x_j > x_i] - \mathbf{1}[x_j < x_i]$.
If $k_n \to \infty$, $\max_{i,j} |x_i - x_j|/k_n \to 0$, $\liminf_n C_n/n^{3/2} > 0$, and

$$C_n^{-1} \sum_{i<j} [1 - 2F_2(t|x_i - x_j|/k_n)] \to m(t), \quad F_2(t) = \mathbb{E}F(\varepsilon + t),$$

Asymptotic Distribution(Peng, Wang & Wang, 2008, JSPI)

Theorem 2 Case I. Suppose F is discontinuous. Then

$$\mathbb{P}(n^{\nu}(\hat{\beta}_n - \beta_0) \to 0) = 1, \quad \nu \geq 0.$$

Hence $\hat{\beta}_n$ is *super-efficient*.

Cae II. Suppose F is continuous.
Let $C_n^2 = \sum c_i^2$ where $c_i = \sum \mathbf{1}[x_j > x_i] - \mathbf{1}[x_j < x_i]$.
If $k_n \to \infty$, $\max_{i,j} |x_i - x_j|/k_n \to 0$, $\liminf_n C_n/n^{3/2} > 0$, and

$$C_n^{-1} \sum_{i<j} [1 - 2F_2(t|x_i - x_j|/k_n)] \to m(t), \quad F_2(t) = \mathbb{E}F(\varepsilon + t),$$

then

$$\lim_{n\to\infty} \mathbb{P}(k_n(\hat{\beta}_n - \beta_0) \leq t) = \Phi(-\sqrt{3}m(t)), \quad t \in \mathbb{R},$$

where Φ is the cdf of the standard normal.

Asymptotic Distribution (Cont'd)

- The asymptotic distribution is normal iff $m(t)$ in linear in t.

Asymptotic Distribution (Cont'd)

- The asymptotic distribution is normal iff $m(t)$ in linear in t.
- Sufficient Conditions.

  Corollary Suppose F is absolutely continuous with pdf f such that $B(F) = \int f^2(t)\,dt < \infty$. If $\liminf C_n/n^{3/2} > 0$, then

  $$(D_n/C_n)(\hat{\beta}_n - \beta_0) \Rightarrow \mathcal{N}(0, 1/3B^2(F)),$$

  where $D_n = \sum_{i=1}^n d_i$ with $d_i = \sum_j |x_i - x_j|$.

Asymptotic Distribution (Cont'd)

- The asymptotic distribution is normal iff $m(t)$ in linear in t.
- Sufficient Conditions.

  Corollary Suppose F is absolutely continuous with pdf f such that $B(F) = \int f^2(t)\,dt < \infty$. If $\liminf C_n/n^{3/2} > 0$, then

  $$(D_n/C_n)(\hat{\beta}_n - \beta_0) \Rightarrow \mathcal{N}(0, 1/3B^2(F)),$$

  where $D_n = \sum_{i=1}^{n} d_i$ with $d_i = \sum_j |x_i - x_j|$.

  This corresponds to Theorem 6.2 of Sen (1968, JASA).

Simulation On Super-Efficiency

Table: Proportions of $\hat{\beta}_n = \beta_0$ with $N = 500$ replications & different sample sizes.

x	Err	5	20	50	80	100	150	250	400
Bin	± 1	0.564	0.998	1.00	1.00	1.00	1.00	1.00	1.00
	Poi	0.222	0.932	1.00	1.00	1.00	1.00	1.00	1.00
	Bin	0.034	0.270	0.47	0.58	0.66	0.72	0.86	0.92

Multiple Theil-Sen Estimator (Dang, Peng, Wang & Zhang, 2008, To appear in JMVA)

- Multiple Linear Regression

$$Y_i = \alpha + \beta^\top X_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where X_i's are *IID random covariates*, ϵ_i's IID $\sim F$, and $\theta = (\alpha, \beta^\top)^\top \in \mathbb{R}^{p+1}$ are parameters.

Multiple Theil-Sen Estimator (Dang, Peng, Wang & Zhang, 2008, To appear in JMVA)

- Multiple Linear Regression

$$Y_i = \alpha + \beta^\top X_i + \epsilon_i, \quad i = 1, \cdots, n,$$

  where X_i's are *IID random covariates*, ϵ_i's IID $\sim F$, and $\theta = (\alpha, \beta^\top)^\top \in \mathbb{R}^{p+1}$ are parameters.
- For sub-sample $\xi_{(K)} = \{(X_k, Y_k) : k \in K\}$ with $K = \{i_1, \cdots, i_m\}$ a subset of $\{1...n\}$ with $p + 1 \le m \le n$, the LSE:

$$\tilde{\theta}_{(K)} = (X_{(K)}^\top X_{(K)})^{-1} X_{(K)}^\top Y_{(K)}.$$

Multiple Theil-Sen Estimator (Dang, Peng, Wang & Zhang, 2008, To appear in JMVA)

- Multiple Linear Regression

$$Y_i = \alpha + \beta^\top X_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where X_i's are *IID random covariates*, ϵ_i's IID $\sim F$, and $\theta = (\alpha, \beta^\top)^\top \in \mathbb{R}^{p+1}$ are parameters.

- For sub-sample $\xi_{(K)} = \{(X_k, Y_k) : k \in K\}$ with $K = \{i_1, \cdots, i_m\}$ a subset of $\{1...n\}$ with $p + 1 \leq m \leq n$, the LSE:

$$\tilde{\theta}_{(K)} = (X_{(K)}^\top X_{(K)})^{-1} X_{(K)}^\top Y_{(K)}.$$

- The proposed Multiple Theil-Sen Estimator (MTSE):

$$\hat{\theta}_n = \text{Multivariate Median} \left\{ \tilde{\theta}_{(K)} : \forall K \right\}$$

Difference-based MTSE

- Pairwise Differences

$$Y_i - Y_j = \beta^\top (X_i - X_j) + \epsilon_i - \epsilon_j, \quad i, j = 1, \dots, n$$

Difference-based MTSE

- Pairwise Differences

$$Y_i - Y_j = \beta^\top (X_i - X_j) + \epsilon_i - \epsilon_j, \quad i, j = 1, ..., n$$

- For a sub-sample

$$\xi_{(K)} = \{(X_k - X_l, Y_k - Y_l) : k, l \in K\}$$

The least squares etimator is $\beta^*_{(K)}$.

Difference-based MTSE

- Pairwise Differences

$$Y_i - Y_j = \beta^\top (X_i - X_j) + \epsilon_i - \epsilon_j, \quad i, j = 1, ..., n$$

- For a sub-sample

$$\xi_{(K)} = \{(X_k - X_l, Y_k - Y_l) : k, l \in K\}$$

The least squares etimator is $\beta^*_{(K)}$.

- The proposed Difference-based MTSE of the slope β:

$$\beta^*_n = \text{Multivariate Median}\left\{\beta^*_{(K)} : \forall K\right\}.$$

Difference-based MTSE

- Pairwise Differences

$$Y_i - Y_j = \beta^\top (X_i - X_j) + \epsilon_i - \epsilon_j, \quad i, j = 1, \ldots, n$$

- For a sub-sample

$$\xi_{(K)} = \{(X_k - X_l, Y_k - Y_l) : k, l \in K\}$$

  The least squares etimator is $\beta^*_{(K)}$.
- The proposed Difference-based MTSE of the slope β:

$$\beta^*_n = \text{Multivariate Median}\left\{\beta^*_{(K)} : \forall K\right\}.$$

- Similarly, define non-overlapped difference-based MTSE $\tilde{\beta}_n$, so β^*_n is overlapped diff-based.

Depth Functions and Depth Medians

- Depth functions provide center-outward ordering of a point in high dim space w.r.t. a distribution.

Depth Functions and Depth Medians

- Depth functions provide center-outward ordering of a point in high dim space w.r.t. a distribution.
- High depth corresponds to centrality, low depth to outlyingness.

Depth Functions and Depth Medians

- Depth functions provide center-outward ordering of a point in high dim space w.r.t. a distribution.
- High depth corresponds to centrality, low depth to outlyingness.
- Depth median consists of the point(s) with the deepest depth.

Depth Functions and Depth Medians

- Depth functions provide center-outward ordering of a point in high dim space w.r.t. a distribution.
- High depth corresponds to centrality, low depth to outlyingness.
- Depth median consists of the point(s) with the deepest depth.
- Popular depth functions:
  - Tukey depth (halfspace depth)(Tukey, '75, Proc. ICM)
  - Simplicial depth (Liu, '90, Ann. Statist.)
  - Spatial depth (Zhang, et al., '00, JMVA)
  - Projection depth (Zuo & Serfling, '00, Ann. Statist.)
  - Tangent depth (Mizera, '02, Ann. Statist.)

Spatial Depth and Spatial Median

- Spatial Depth:

$$D(x, F) = 1 - \|\mathbb{E}_F S(x - X)\|, \quad x \in R^d$$

where $S(x) = x/\|x\|$ is the spatial sign function.

Spatial Depth and Spatial Median

- Spatial Depth:

$$D(x, F) = 1 - \|\mathbb{E}_F S(x - X)\|, \quad x \in R^d$$

  where $S(x) = x/\|x\|$ is the spatial sign function.
- Sample Version:

$$D(x, F_n) = 1 - \left\| \frac{1}{n} \sum_{i=1}^{n} S(x - X_i) \right\|$$

Spatial Depth and Spatial Median

- Spatial Depth:

$$D(x, F) = 1 - \|\mathbb{E}_F S(x - X)\|, \quad x \in R^d$$

where $S(x) = x/\|x\|$ is the spatial sign function.

- Sample Version:

$$D(x, F_n) = 1 - \left\| \frac{1}{n} \sum_{i=1}^{n} S(x - X_i) \right\|$$

- Spatial Median:

$$\text{Spatial Median}\,\{x; X_1, ..., X_n\} = \arg\max_{x \in \mathbb{R}^d} D(x, F_n)$$

$$= \arg\min_{x \in \mathbb{R}^d} \left\| \frac{1}{n} \sum_{i=1}^{n} S(x - X_i) \right\|$$

Uniqueness and Existence of Spatial Median

Let Z be a r.v. on $\mathbb{R}^d$ with distribution Q. Z has a unique spatial median if one of the following holds.

Uniqueness and Existence of Spatial Median

Let Z be a r.v. on $\mathbb{R}^d$ with distribution Q. Z has a unique spatial median if one of the following holds.

- Q is not concentrated on a line (Milasevic and Ducharme, 1987, Ann. Statist.)

Uniqueness and Existence of Spatial Median

Let Z be a r.v. on $\mathbb{R}^d$ with distribution Q. Z has a unique spatial median if one of the following holds.

- Q is not concentrated on a line (Milasevic and Ducharme, 1987, Ann. Statist.)
- There are two one-dimensional marginal distributions each of which is not point mass for $d \geq 2$.

Uniqueness and Existence of Spatial Median

Let Z be a r.v. on $\mathbb{R}^d$ with distribution Q. Z has a unique spatial median if one of the following holds.

- Q is not concentrated on a line (Milasevic and Ducharme, 1987, Ann. Statist.)
- There are two one-dimensional marginal distributions each of which is not point mass for $d \geq 2$.
- There are at least two absolute continuous one-dimensional marginal distributions.

Uniqueness and Existence of Spatial Median

Let Z be a r.v. on $\mathbb{R}^d$ with distribution Q. Z has a unique spatial median if one of the following holds.

- Q is not concentrated on a line (Milasevic and Ducharme, 1987, Ann. Statist.)
- There are two one-dimensional marginal distributions each of which is not point mass for $d \geq 2$.
- There are at least two absolute continuous one-dimensional marginal distributions.
- Q is *angularly symmetric* about its median.

Uniqueness and Existence of Spatial Median

Let Z be a r.v. on $\mathbb{R}^d$ with distribution Q. Z has a unique spatial median if one of the following holds.

- Q is not concentrated on a line (Milasevic and Ducharme, 1987, Ann. Statist.)
- There are two one-dimensional marginal distributions each of which is not point mass for $d \geq 2$.
- There are at least two absolute continuous one-dimensional marginal distributions.
- Q is *angularly symmetric* about its median.
- Q is *centrally symmetric* about its median.

Strong Consistency

Denote $\theta_0 = (\alpha_0, \beta_0^\top)^\top$ the true parameter value and $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ the estimator and $k_0 = (1...m)$.

Theorem 3 Suppose the distribution of $h(\xi_0) = (X_{k_0}^\top X_{k_0})^{-1} X_{k_0}^\top Y_{k_0}$ is not concentrated on a line and the map $\vartheta \mapsto \mathbb{E}\|\vartheta - h(\xi_0)\|$ is maximized at true θ_0. Then the MTSE $\hat{\theta}_n$ is strongly consistent, i.e. $\hat{\theta}_n \to \theta_0$ a.s.

Strong Consistency

Denote $\theta_0 = (\alpha_0, \beta_0^\top)^\top$ the true parameter value and $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ the estimator and $k_0 = (1 \ldots m)$.

Theorem 3 Suppose the distribution of $h(\xi_0) = (X_{k_0}^\top X_{k_0})^{-1} X_{k_0}^\top Y_{k_0}$ is not concentrated on a line and the map $\vartheta \mapsto \mathbb{E}\|\vartheta - h(\xi_0)\|$ is maximized at true θ_0. Then the MTSE $\hat{\theta}_n$ is strongly consistent, i.e. $\hat{\theta}_n \to \theta_0$ a.s.

Theorem 4 Suppose ϵ is not concentrated on a point mass and its distribution is symmetric about zero. Then the overlapped diff-based MTSE β_n^* is strongly consistent: $\beta_n^* \to \beta_0$ a.s.

Strong Consistency

Denote $\theta_0 = (\alpha_0, \beta_0^\top)^\top$ the true parameter value and $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ the estimator and $k_0 = (1...m)$.

Theorem 3 Suppose the distribution of $h(\xi_0) = (X_{k_0}^\top X_{k_0})^{-1} X_{k_0}^\top Y_{k_0}$ is not concentrated on a line and the map $\vartheta \mapsto \mathbb{E}\|\vartheta - h(\xi_0)\|$ is maximized at true θ_0. Then the MTSE $\hat{\theta}_n$ is strongly consistent, i.e. $\hat{\theta}_n \to \theta_0$ a.s.

Theorem 4 Suppose ϵ is not concentrated on a point mass and its distribution is symmetric about zero. Then the overlapped diff-based MTSE β_n^* is strongly consistent: $\beta_n^* \to \beta_0$ a.s.

Theorem 5 Suppose the distribution of the error ϵ is not concentrated on a point mass. Then the non-overlapped diff-based MTSE $\tilde{\beta}_n$ are strongly consistent: $\tilde{\beta}_n \to \beta_0$ a.s.

Symmetry

Obviously $\epsilon_1 - \epsilon_2 \overset{d}{=} \epsilon_2 - \epsilon_1$.

Symmetry

Obviously $\epsilon_1 - \epsilon_2 \stackrel{d}{=} \epsilon_2 - \epsilon_1$. But

$$(\epsilon_1 - \epsilon_2, \epsilon_1 - \epsilon_3, \epsilon_2 - \epsilon_3) \stackrel{d}{\neq} (\epsilon_2 - \epsilon_1, \epsilon_3 - \epsilon_1, \epsilon_3 - \epsilon_2).$$

Symmetry

Obviously $\epsilon_1 - \epsilon_2 \stackrel{d}{=} \epsilon_2 - \epsilon_1$. But

$$(\epsilon_1 - \epsilon_2, \epsilon_1 - \epsilon_3, \epsilon_2 - \epsilon_3) \stackrel{d}{\neq} (\epsilon_2 - \epsilon_1, \epsilon_3 - \epsilon_1, \epsilon_3 - \epsilon_2).$$

Certainly

$$(\epsilon_1 - \epsilon_2, \epsilon_3 - \epsilon_4, \epsilon_5 - \epsilon_6) \stackrel{d}{=} (\epsilon_2 - \epsilon_1, \epsilon_4 - \epsilon_3, \epsilon_6 - \epsilon_5).$$

Symmetry

Obviously $\epsilon_1 - \epsilon_2 \stackrel{d}{=} \epsilon_2 - \epsilon_1$. But

$$(\epsilon_1 - \epsilon_2, \epsilon_1 - \epsilon_3, \epsilon_2 - \epsilon_3) \stackrel{d}{\neq} (\epsilon_2 - \epsilon_1, \epsilon_3 - \epsilon_1, \epsilon_3 - \epsilon_2).$$

Certainly

$$(\epsilon_1 - \epsilon_2, \epsilon_3 - \epsilon_4, \epsilon_5 - \epsilon_6) \stackrel{d}{=} (\epsilon_2 - \epsilon_1, \epsilon_4 - \epsilon_3, \epsilon_6 - \epsilon_5).$$

Theorem 6 A random variable is symmetric about its median iff the random vectors whose components are the differences of three i.i.d. copies of the random variable are symmetric about zero.

Super-Efficiency

Let $h_b(\xi_0) = I_p h(\xi_0)$ with $I_p = \text{diag}(0, 1, \ldots, 1)$ a diagonal matrix.

Super-Efficiency

Let $h_b(\xi_0) = I_p h(\xi_0)$ with $I_p = \text{diag}(0, 1, \ldots, 1)$ a diagonal matrix.
Theorem 7 Suppose the error has a distribution symmetric about zero. Assume $h_b(\xi_0)$ is not concentrated on a line. If the error distribution is discontinuous, then

$$\mathbb{P}(\hat{\beta}_n = \beta_0) \to 1.$$

Super-Efficiency

Let $h_b(\xi_0) = I_p h(\xi_0)$ with $I_p = \text{diag}(0, 1, \ldots, 1)$ a diagonal matrix.
Theorem 7 Suppose the error has a distribution symmetric about zero. Assume $h_b(\xi_0)$ is not concentrated on a line. If the error distribution is discontinuous, then

$$\mathbb{P}(\hat{\beta}_n = \beta_0) \to 1.$$

Consequently, we have super-efficiency:

$$n^\nu(\hat{\beta}_n - \beta_0) \to 0, \quad \nu \geq 0.$$

Assumptotic Normality

Denote $\mu(\vartheta) = \mathbb{E}(\|\vartheta - h(\xi_0)\| - \|h(\xi_0)\|)$ and

$$D_1(\vartheta) = \mathbb{E}\left\{\frac{1}{\|\vartheta - h(\xi_0)\|}\left(I_m - \frac{(\vartheta - h(\xi_0))^{\otimes 2}}{\|\vartheta - h(\xi_0)\|^2}\right)\right\}.$$

Assumptotic Normality

Denote $\mu(\vartheta) = \mathbb{E}(\|\vartheta - h(\xi_0)\| - \|h(\xi_0)\|)$ and

$$D_1(\vartheta) = \mathbb{E}\left\{\frac{1}{\|\vartheta - h(\xi_0)\|}\left(I_m - \frac{(\vartheta - h(\xi_0))^{\otimes 2}}{\|\vartheta - h(\xi_0)\|^2}\right)\right\}.$$

Theorem 8 Suppose

- ϵ is symmetric about zero;

Assumptotic Normality

Denote $\mu(\vartheta) = \mathbb{E}(\|\vartheta - h(\xi_0)\| - \|h(\xi_0)\|)$ and

$$D_1(\vartheta) = \mathbb{E}\left\{\frac{1}{\|\vartheta - h(\xi_0)\|}\left(I_m - \frac{(\vartheta - h(\xi_0))^{\otimes 2}}{\|\vartheta - h(\xi_0)\|^2}\right)\right\}.$$

Theorem 8 Suppose

- ϵ is symmetric about zero;
- the map $\vartheta \mapsto \mathbb{E}\|\vartheta - h(\xi_0)\|$ is maximized at true θ_0;

Assumptotic Normality

Denote $\mu(\vartheta) = \mathbb{E}(\|\vartheta - h(\xi_0)\| - \|h(\xi_0)\|)$ and

$$D_1(\vartheta) = \mathbb{E}\left\{ \frac{1}{\|\vartheta - h(\xi_0)\|} \left(I_m - \frac{(\vartheta - h(\xi_0))^{\otimes 2}}{\|\vartheta - h(\xi_0)\|^2} \right) \right\}.$$

Theorem 8 Suppose

- ϵ is symmetric about zero;
- the map $\vartheta \mapsto \mathbb{E}\|\vartheta - h(\xi_0)\|$ is maximized at true θ_0;
- the distributions of ϵ and $(X_{k_0}^\top X_{k_0})^{-1} X_{k_0}^\top$ are absolutely continuous w.r.t. the Lebesgue measure;

Assumptotic Normality

Denote $\mu(\vartheta) = \mathbb{E}(\|\vartheta - h(\xi_0)\| - \|h(\xi_0)\|)$ and

$$D_1(\vartheta) = \mathbb{E}\left\{\frac{1}{\|\vartheta - h(\xi_0)\|}\left(I_m - \frac{(\vartheta - h(\xi_0))^{\otimes 2}}{\|\vartheta - h(\xi_0)\|^2}\right)\right\}.$$

Theorem 8 Suppose

- ϵ is symmetric about zero;
- the map $\vartheta \mapsto \mathbb{E}\|\vartheta - h(\xi_0)\|$ is maximized at true θ_0;
- the distributions of ϵ and $(X_{k_0}^\top X_{k_0})^{-1} X_{k_0}^\top$ are absolutely continuous w.r.t. the Lebesgue measure;
- $\Delta\mu(\vartheta)$ is continuously differentiable with derivative $\Delta^2\mu(\vartheta) = D_1(\vartheta)$ in a neighborhood of θ_0.

Assumptotic Normality

Denote $\mu(\vartheta) = \mathbb{E}(\|\vartheta - h(\xi_0)\| - \|h(\xi_0)\|)$ and

$$D_1(\vartheta) = \mathbb{E}\left\{\frac{1}{\|\vartheta - h(\xi_0)\|}\left(I_m - \frac{(\vartheta - h(\xi_0))^{\otimes 2}}{\|\vartheta - h(\xi_0)\|^2}\right)\right\}.$$

Theorem 8 Suppose

- ϵ is symmetric about zero;
- the map $\vartheta \mapsto \mathbb{E}\|\vartheta - h(\xi_0)\|$ is maximized at true θ_0;
- the distributions of ϵ and $(X_{k_0}^\top X_{k_0})^{-1} X_{k_0}^\top$ are absolutely continuous w.r.t. the Lebesgue measure;
- $\Delta\mu(\vartheta)$ is continuously differentiable with derivative $\Delta^2\mu(\vartheta) = D_1(\vartheta)$ in a neighborhood of θ_0.

Then we have($D = D_1(\theta_0)$ invertible):

$$\hat{\theta}_n = \theta_0 + D^{-1}\bar{S}_n + R_n, \tag{1}$$

where $\bar{S}_n = \sum_k S(\theta_0 - h(\xi_{(k)}))/\binom{n}{m}$, $R_n = o_p(n^{-1/2})$.

Assumptotic Normality: Theorem 6 (Cont'd)

Hence

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, \Sigma)$$

where $\Sigma = D_1^{-1}(\theta_0)\mathbb{E}\tilde{h}(\xi_1)^{\otimes 2} D_1^{-1}(\theta_0)$ with

$$\tilde{h}(\xi_1) = \mathbb{E}(S(\theta_0 - h(\xi_1, ..., \xi_m))|\xi_1).$$

Assumptotic Normality

Theorem 9 Suppose ϵ has distribution symmetric about zero. Assume

$$\mathbb{E}\|h(\xi_0) - \theta_0\|^{(3+\nu)/2} < \infty$$

for some $0 \leq \nu \leq 1$. Then the MTSE $\hat{\theta}_n$ satisfies (1) with the remainder

$$R_n = O_p(n^{-(3+\nu)/4}(\log n)^{1/2}(\log\log n)^{(1+\nu)/4}).$$

Assumptotic Normality

Theorem 9 Suppose ϵ has distribution symmetric about zero. Assume

$$\mathbb{E}\|h(\xi_0) - \theta_0\|^{(3+\nu)/2} < \infty$$

for some $0 \leq \nu \leq 1$. Then the MTSE $\hat{\theta}_n$ satisfies (1) with the remainder

$$R_n = O_p(n^{-(3+\nu)/4}(\log n)^{1/2}(\log\log n)^{(1+\nu)/4}).$$

Theorem 10 Under similar assumptions, β_n^* satisfies (1) with the remainder satisfying the above equality.

Assumptotic Normality

Theorem 9 Suppose ϵ has distribution symmetric about zero. Assume

$$\mathbb{E}\|h(\xi_0) - \theta_0\|^{(3+\nu)/2} < \infty$$

for some $0 \leq \nu \leq 1$. Then the MTSE $\hat{\theta}_n$ satisfies (1) with the remainder

$$R_n = O_p(n^{-(3+\nu)/4}(\log n)^{1/2}(\log \log n)^{(1+\nu)/4}).$$

Theorem 10 Under similar assumptions, β_n^* satisfies (1) with the remainder satisfying the above equality.

Our results slightly improve Bose's (1998, Ann. Statist.) and Zhou and Serfling's (2007, preprint).

Robustness

- Breakdown point

$$BP = 1 - (1/2)^{1/m}$$

Robustness

- Breakdown point

$$BP = 1 - (1/2)^{1/m}$$

- Balance between robustness, efficiency and computation intensity
  robustness: decreasing
  efficiency: increasing
  intensity (complexity): increasing then decreasing

Robustness

- Influence function $IF((y, \mathbf{x}); \hat{\beta}_n)$ is

$$D^{-1}\mathbb{E}S(\beta_0 - (X_{\mathbf{x}}^\top X_{\mathbf{x}})^{-1}X_{\mathbf{x}}^\top Y_y)$$

where D is the previous D_1 or D_1^*, $X_{\mathbf{x}} = [\mathbf{1}_m, X(\mathbf{x})]$ with column $\mathbf{1}_m \in \mathbb{R}^m$ of all entries 1 and $X(\mathbf{x}) = [\mathbf{x}, X_1, \cdots, X_{m-1}]^\top$ and $Y_y = (y, Y_1, \cdots, Y_{m-1})^\top$.

Robustness

- Influence function $IF((y, \mathbf{x}); \hat{\beta}_n)$ is

$$D^{-1}\mathbb{E}S(\beta_0 - (X_{\mathbf{x}}^\top X_{\mathbf{x}})^{-1}X_{\mathbf{x}}^\top Y_y)$$

  where D is the previous D_1 or D_1^*, $X_{\mathbf{x}} = [\mathbf{1}_m, X(\mathbf{x})]$ with column $\mathbf{1}_m \in \mathbb{R}^m$ of all entries 1 and $X(\mathbf{x}) = [\mathbf{x}, X_1, \cdots, X_{m-1}]^\top$ and $Y_y = (y, Y_1, \cdots, Y_{m-1})^\top$.

- This shows that the estimator is only influenced by the direction and is irrelevant to the magnitudes of y and $\mathbf{x}$. Consequently our MTSE is robust against both $\mathbf{x}$ and y outlying.

Simulation: Robustness

Samples are generated from

$$Y_i = 1 + 5X_{1i} + 10X_{2i} + \epsilon_i,$$

where $X_{1i} \sim \mathcal{N}(0,1)$, $X_{2i} \sim U(0,1)$, $\epsilon_i \sim \mathcal{N}(0,0.5)$.

Simulation: Robustness

Samples are generated from

$$Y_i = 1 + 5X_{1i} + 10X_{2i} + \epsilon_i,$$

where $X_{1i} \sim \mathcal{N}(0,1)$, $X_{2i} \sim U(0,1)$, $\epsilon_i \sim \mathcal{N}(0,0.5)$.

	true parameter $\theta_0 = (5,10)$		
	Theil-Sen	Diff Theil-Sen	LSE
n=20	(4.31,10.43)	(4.38,10.93)	(4.38,10.59)
n=40	(4.97,9.88)	(4.98,9.66)	(5.01,9.87)
$n_1 = 16, n_2 = 4$	(5.01,9.95)	(5.06,9.71)	(4.18,7.76)
$n_1 = 15, n_2 = 5$	(5.30,9.46)	(5.25,9.33)	(5.65,2.27)
$n_1 = 14, n_2 = 6$	(4.37,9.68)	(4.22,9.41)	(-2.65,7.72)
$n_1 = 13, n_2 = 7$	(4.14,9.17)	(4.88,9.59)	(-2.37,3.34)
$n_1 = 12, n_2 = 8$	(3.98,9.12)	(0.72,5.65)	(-3.37,5.18)

(a) Upper: estimators with sample size n without outliers.
(b) Lower: estimators with sample size $n = n_1 + n_2$: n_1=# of "good" observations, n_2=# of outliers.

Simulation: Efficiency Comparison

		Normal		T_3		Cauchy	
		TSE	LSE	TSE	LSE	TSE	LSE
n=10	EMSE	3.716	2.643	7.058	7.628	45.97	2613
	RE	0.711	1.000	1.081	1.000	56.84	1.000
n=20	EMSE	1.339	1.075	2.111	2.627	5.667	816.2
	RE	0.803	1.000	1.245	1.000	144.0	1.000
n=30	EMSE	0.739	0.596	1.161	1.569	3.032	2207
	RE	0.806	1.000	1.352	1.000	728.0	1.000

EMSE=Empirical Mean Squared Error.
RE=Ratio of EMSE of LSE to EMSE of MTSE

Theil-Sen Estimator in Multiple Regression With non-random Covariate

- Multiple Regression

$$Y_i = \alpha + \beta^\top x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where x_i's are non-random.

Theil-Sen Estimator in Multiple Regression With non-random Covariate

- Multiple Regression

$$Y_i = \alpha + \beta^\top x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where x_i's are non-random.

- Asymptotic Behavior: Consistency and Asymptotic Normality.

Theil-Sen Estimator in Multiple Regression With non-random Covariate

- Multiple Regression

$$Y_i = \alpha + \beta^\top x_i + \epsilon_i, \quad i = 1, \cdots, n,$$

where x_i's are non-random.

- Asymptotic Behavior: Consistency and Asymptotic Normality.
- Asymptotic Behavior of TSE as $m = m_n \to \infty$.

Theil-Sen Estimator in Multivariate Multiple Regression

- Multivariate Multiple Regression

$$\mathbf{Y} = B\mathbf{X} + \mathscr{E}$$

where $\mathbf{Y}, \mathbf{X}$ are observation matrices and $\mathscr{E}$ is randon error matrix, and B is matrix parameter of interest.

Theil-Sen Estimator in Multivariate Multiple Regression

- Multivariate Multiple Regression

$$\mathbf{Y} = B\mathbf{X} + \mathscr{E}$$

  where $\mathbf{Y}, \mathbf{X}$ are observation matrices and $\mathscr{E}$ is randon error matrix, and B is matrix parameter of interest.
- Construction of TSE and Asymptotic Behavior.

Theil-Sen Estimator in Multivariate Multiple Regression

- Multivariate Multiple Regression

$$\mathbf{Y} = B\mathbf{X} + \mathscr{E}$$

  where $\mathbf{Y}, \mathbf{X}$ are observation matrices and $\mathscr{E}$ is randon error matrix, and B is matrix parameter of interest.
- Construction of TSE and Asymptotic Behavior.
  Two considerations:

Theil-Sen Estimator in Multivariate Multiple Regression

- Multivariate Multiple Regression

$$\mathbf{Y} = B\mathbf{X} + \mathscr{E}$$

  where $\mathbf{Y}, \mathbf{X}$ are observation matrices and $\mathscr{E}$ is randon error matrix, and B is matrix parameter of interest.
- Construction of TSE and Asymptotic Behavior.
  Two considerations:
  (1) $\mathbf{X}$ is random
  (2) $\mathbf{X}$ is non-random.

TSE in Semiparametric Mixed Models

Semiparametric mixed model

$$y_j = x_j^\top \beta + z_j^\top u_j + \rho(t_j) + \varepsilon_j, \quad j = 1, \cdots, n,$$

where β is parameter, u_j is random vector with $\mathbb{E}u_j = 0$, ρ is unknown nonparametric function. $\{\varepsilon_j\}$ are IID errors independent of $\{(x_j, u_j, t_j)\}$.

TSE in Semiparametric Mixed Models

Semiparametric mixed model

$$y_j = x_j^\top \beta + z_j^\top u_j + \rho(t_j) + \varepsilon_j, \quad j = 1, \cdots, n,$$

where β is parameter, u_j is random vector with $\mathbb{E}u_j = 0$, ρ is unknown nonparametric function. $\{\varepsilon_j\}$ are IID errors independent of $\{(x_j, u_j, t_j)\}$.

The model includes

- Linear Mixed Model: $\rho(t) \equiv 0$,

TSE in Semiparametric Mixed Models

Semiparametric mixed model

$$y_j = x_j^\top \beta + z_j^\top u_j + \rho(t_j) + \varepsilon_j, \quad j = 1, \cdots, n,$$

where β is parameter, u_j is random vector with $\mathbb{E}u_j = 0$, ρ is unknown nonparametric function. $\{\varepsilon_j\}$ are IID errors independent of $\{(x_j, u_j, t_j)\}$.

The model includes

- Linear Mixed Model: $\rho(t) \equiv 0$, including
  - (1a) Weighted multiple linear model: $u \equiv 0$
  - (1b) BLUE of common mean
  - (1c) Kriging estimator

TSE in Semiparametric Mixed Models

Semiparametric mixed model

$$y_j = x_j^\top \beta + z_j^\top u_j + \rho(t_j) + \varepsilon_j, \quad j = 1, \cdots, n,$$

where β is parameter, u_j is random vector with $\mathbb{E}u_j = 0$, ρ is unknown nonparametric function. $\{\varepsilon_j\}$ are IID errors independent of $\{(x_j, u_j, t_j)\}$.
The model includes

- Linear Mixed Model: $\rho(t) \equiv 0$, including
  - (1a) Weighted multiple linear model: $u \equiv 0$
  - (1b) BLUE of common mean
  - (1c) Kriging estimator
- Nonparametric: $\beta \equiv 0, u \equiv 0$

TSE in Semiparametric Mixed Models

Semiparametric mixed model

$$y_j = x_j^\top \beta + z_j^\top u_j + \rho(t_j) + \varepsilon_j, \quad j = 1, \cdots, n,$$

where β is parameter, u_j is random vector with $\mathbb{E}u_j = 0$, ρ is unknown nonparametric function. $\{\varepsilon_j\}$ are IID errors independent of $\{(x_j, u_j, t_j)\}$.

The model includes

- Linear Mixed Model: $\rho(t) \equiv 0$, including
  - (1a) Weighted multiple linear model: $u \equiv 0$
  - (1b) BLUE of common mean
  - (1c) Kriging estimator
- Nonparametric: $\beta \equiv 0, u \equiv 0$
- Partially Linear Additive: $u \equiv 0$

Sub-Sampling

- The computation of TSE will be enormous.

Sub-Sampling

- The computation of TSE will be enormous.
- To deal with this, we consider sub-sampling.

Sub-Sampling

- The computation of TSE will be enormous.
- To deal with this, we consider sub-sampling.
- We compute the spatial median based a *simple random subsample* $\mathcal{N}$ of all $\binom{n}{m}$ LSE's, $\mathcal{M}$.

Sub-Sampling

- The computation of TSE will be enormous.
- To deal with this, we consider sub-sampling.
- We compute the spatial median based a *simple random subsample* $\mathcal{N}$ of all $\binom{n}{m}$ LSE's, $\mathcal{M}$.
- We propose TSE $\hat{\theta}_{n,N}$ by

$$\hat{\theta}_{n,N} = \mathrm{SpatialMedian}\left\{\hat{\theta}_{(m)} : (m) \in \mathcal{N}\right\}.$$

Sub-Sampling

- The computation of TSE will be enormous.
- To deal with this, we consider sub-sampling.
- We compute the spatial median based a *simple random subsample* $\mathcal{N}$ of all $\binom{n}{m}$ LSE's, $\mathcal{M}$.
- We propose TSE $\hat{\theta}_{n,N}$ by

$$\hat{\theta}_{n,N} = \text{SpatialMedian}\left\{\hat{\theta}_{(m)} : (m) \in \mathcal{N}\right\}.$$

- Asymptotic behavior and properties of $\hat{\theta}_{n,N}$.

Projects

- Efficiency Comparison of TSE's with LSE, ETC.

Projects

- Efficiency Comparison of TSE's with LSE, ETC.
- Robustness Comparison of TSE's with Other Robust Estimators.

Projects

- Efficiency Comparison of TSE's with LSE, ETC.
- Robustness Comparison of TSE's with Other Robust Estimators.
- Applications: astronomy, remote sensoring, psychology, etc.

Projects

- Efficiency Comparison of TSE's with LSE, ETC.
- Robustness Comparison of TSE's with Other Robust Estimators.
- Applications: astronomy, remote sensoring, psychology, etc.
- TSE's Based on Other Depth-defined Medians.

Spatial depth-based outlier detector

A disadvantage of Spatial Depth

(a) Triangle data

(b) Ring data

Positive Definition Kernel

A positive definite kernel, $\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, implicitly defines an embedding map

$$\phi : \mathbf{x} \in \mathbb{R}^d \longmapsto \phi(\mathbf{x}) \in \mathbb{F}$$

via the inner product in the feature space $\mathbb{F}$, i.e.

$$\kappa(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle \ .$$

Examples of kernels:

Gaussian kernel: $\kappa(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}\right)$

Polynomial kernel: $\kappa(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x}^t\mathbf{y})^p$

Rational quadratic kernel: $\kappa(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|/(\theta + \|\mathbf{x} - \mathbf{y}\|)$

Kernelized Spatial Depth

Rewrite sample spatial depth as

$$D(\mathbf{x}, F_n) = 1 - \frac{1}{n}\left(\sum_{\mathbf{y},\mathbf{z}\in\mathcal{X}} \frac{\mathbf{x}^T\mathbf{x} + \mathbf{y}^T\mathbf{z} - \mathbf{x}^T\mathbf{y} - \mathbf{x}^T\mathbf{z}}{\delta(\mathbf{x},\mathbf{y})\delta(\mathbf{x},\mathbf{z})}\right)^{1/2}.$$

where $\delta(\mathbf{x},\mathbf{y}) = \sqrt{\mathbf{x}^T\mathbf{x} + \mathbf{y}^T\mathbf{y} - 2\mathbf{x}^T\mathbf{y}}$. Replacing the inner products with kernel κ, we have *the kernelized spatial depth*:

$$D_\kappa(\mathbf{x}, F_n) = 1 - \frac{1}{n}\left(\sum_{\mathbf{y},\mathbf{z}\in\mathcal{X}} \frac{\kappa(\mathbf{x},\mathbf{x}) + \kappa(\mathbf{y},\mathbf{z}) - \kappa(\mathbf{x},\mathbf{y}) - \kappa(\mathbf{x},\mathbf{z})}{\delta_\kappa(\mathbf{x},\mathbf{y})\delta_\kappa(\mathbf{x},\mathbf{z})}\right)^{1/2} \quad (2)$$

where $\delta_\kappa(\mathbf{x},\mathbf{y}) = \sqrt{\kappa(\mathbf{x},\mathbf{x}) + \kappa(\mathbf{y},\mathbf{y}) - 2\kappa(\mathbf{x},\mathbf{y})}$.

(a) Triangle data

(b) Ring data

Figure: Contour plots of kernelized spatial depth functions.

Synthetic Data

Figure: Decision boundaries of outlier detectors.

Synthetic Data

Figure: Decision boundaries of outlier detectors.

Chen, Dang, Peng and Bart (2007). **Outlier Detection with the**

Local Spatial Depth

Zuo and Serfling (2000, Ann. Statist.) gave defining properties of statistical depth function.

- Affine invariance.

Local Spatial Depth

Zuo and Serfling (2000, Ann. Statist.) gave defining properties of statistical depth function.

- Affine invariance.
- Maximality at center.

Local Spatial Depth

Zuo and Serfling (2000, Ann. Statist.) gave defining properties of statistical depth function.

- Affine invariance.
- Maximality at center.
- Monotonicity relative to the deepest point.

Local Spatial Depth

Zuo and Serfling (2000, Ann. Statist.) gave defining properties of statistical depth function.

- Affine invariance.
- Maximality at center.
- Monotonicity relative to the deepest point.
- Vanishing at infinity.

Local Spatial Depth

- Existence and Uniqueness of LSD.

Local Spatial Depth

- Existence and Uniqueness of LSD.
- Nestedness of the LSD contours.

Local Spatial Depth

- Existence and Uniqueness of LSD.
- Nestedness of the LSD contours.
- Relationship between LSD and kernel density estimates.

Applications of Local Spatial Depth

- LSD-based TSE's.

Applications of Local Spatial Depth

- LSD-based TSE's.
- LSD-based Spatial Rank Statistics: Skewness, Kurtosis, ETC.

Applications of Local Spatial Depth

- LSD-based TSE's.
- LSD-based Spatial Rank Statistics: Skewness, Kurtosis, ETC.
- LSD-based Spatial Quantile.

Applications of Local Spatial Depth

- LSD-based TSE's.
- LSD-based Spatial Rank Statistics: Skewness, Kurtosis, ETC.
- LSD-based Spatial Quantile.
- LSD-based Robust Estimators of Scatter Matrices.

Applications of Local Spatial Depth

- LSD-based TSE's.
- LSD-based Spatial Rank Statistics: Skewness, Kurtosis, ETC.
- LSD-based Spatial Quantile.
- LSD-based Robust Estimators of Scatter Matrices.
- LSD-based Clustering/Classification/Outlier Detection.

THANKS