روش کمتر ن قدر م لق خ ا در مدل‌ها خ

حمزه ترابی

دانشکده ر ا دانشگاه زد

چکیده روش کمتر ن قدرم لق خ ا رای رورد پارامترهای مدل‌های خ ی ه عنوان ک روش ؛ رومند[1] گاهی ه جای روش کمتر ن توان دوم خ ا ه کار رده می‌شود ه‌و ژه اگر در ن داده‌ها داده پرت[2] وجود داشته اشد و ا شر نرمال ودن خ اها رقرار نباشد
رورد‌گرهای کمتر ن قدر م لق خ ا تحت شرا ع فی روی خ اها ه ور مجانبی کاراتر از رورد‌گرهای کمتر ن توان دوم خ ا هستند
در ا ن مقاله پس از ان چند تعر ف و نمادگذاری ه تار خچه‌ای اجمالی از ا ن دو روش رورد البته جهت‌دار ه سمت روش کمتر ن قدر م لق خ ا و سپس ه ررسی چند روش ی شنهادی رای افتن رورد‌گرهای روش کمتر ن قدر م لق خ ا می‌پرداز م در پا ان پس از ان توز مجانبی ا ن رورد‌گرها چگونگی استنبا یرامون پارامترهای مدل را ررسی خواه م کرد

مقدمه

مدل خ ی

$$y_i = \sum_{j=1}^{k} x_{ij}\beta_j + \epsilon_i, \quad i = \quad , ..., n, \qquad ( )$$

ا هم‌ارز ن

$$y_i = \underline{x}'_i \underline{\beta} + \epsilon_i \qquad ( )$$

و ا فرم ماتر سی ن عنی

$$\underline{y} = X\underline{\beta} + \underline{\epsilon} \qquad ( )$$

را در ; ر گ ر د که در ن

$\underline{y} = [y_1, ..., y_n]', \quad X = [x_{ij}]_{n\times k}, \quad \underline{\beta} = [\beta_1, ..., \beta_k]', \quad X' = [\underline{x}_1, ..., \underline{x}_n],$

1) Robust 2) Outlier 3) Design Matrix 4) Minkowsky Metric

$\underline{x}'_i = [x_{i1}, ..., x_{ik}], \quad \underline{\epsilon} = [\epsilon_1, ..., \epsilon_n]', \quad rank(X) = r \leq k \leq n,$

$\underline{\epsilon}$ ردار خ اها $\underline{\beta}$ ردار پارامترها X ماتر س رح[3] و $\underline{y}$ ردار مشاهدات است

رای رورد $\underline{\beta}$ ک مع ار م لو   متر م نکوفسکی[4] ا نرم L_p[5] است که ه صورت ز ر تعر ف می‌شود

$$||\underline{\epsilon}||_p = [\sum_{i=1}^{n} |\epsilon_i|^p]^{1/p}, \quad p \geq$$

روشن است که رای ک مقدار p هتر ن رورد $\underline{\beta}$ مقداری است که $||\underline{\epsilon}||_p$ را م ن مم می‌کند سه حالت ز ر عموما در رورد ه کار رده می‌شوند

الف $p =$ در ا ن حالت رورد $\underline{\beta}$ را ا $\tilde{\underline{\beta}}$ نما ش می‌ده م و ه ن رورد کمتر ن قدرم لق خ ا[6] از ا ن پس LAE وگاهی رورد L_1 می‌گو م

$p =$ در ا ن حالت رورد $\underline{\beta}$ را ا $\hat{\underline{\beta}}$ نما ش می‌ده م و ه ن رورد کمتر ن توان دوم خ ا[7] از ا ن پس LSE وگاهی رورد L_2 می‌گو م

پ $p = \infty$ در ا ن حالت رورد $\underline{\beta}$ را ا $\underline{\beta}_\infty$ نما ش می‌ده م و ه ن رورد م ن ماکس[8] ا کمتر ن ماکس مم قدرم لق خ ا[9] از ا ن پس $LMAE$ وگاهی رورد L_∞ می‌گو م

گاهی ه رورد‌های کمتر ن قدرم لق خ ا رورد‌های کمتر ن قدرم لق مقدار[10] LAV کمتر ن قدرم لق انحراف[11] LAD م ن مم قدرم لق خ ا[12] MAE ن ز گفته می‌شود

تار خچه LAE و LSE

پ دا ش رگرس ون و درحالت کلی‌تر مدل‌های خ ی از مس له رازش منحنی وده است   کی ازگام‌های اول ه رای رازش ک منحنی رای ک مجموعه داده توس گال له[13] در سال رداشته شده است   او نرم L_1 را رای مشخص کردن موقع ت ستارگان ه کار می ند

وسکو چ[14]   ک روش هندسی رازش خ $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$ را ا م ن مم کردن مجمو $\sum |y_i - \beta_1 - \beta_2 x_i|$ ا محدود ت $\sum (y_i - \beta_1 - \beta_2 x_i) =$ در ن ر می‌گ رد لاپلاس[15] مس له پ شنهادی وسکو چ را حل می‌کند اما ک روش جبری را جا گز ن روش هندسی او می‌کند در حدود سال عد از مس له پ شنهادی وسکو چ روش LSE توس لژاندر   ه عنوان ک روش جبری ارا ه می‌شود و سپس ه عنوان ک روش ماری توس سی اف گاوس در سال   مورد ررسی و تعد ل قرار می‌گ رد

5) L_p-norm 6) Least Absolute Error 7) Least Square Error 8) MINIMAX 9) Least of Maximum Absolute Error 10) Least Absolute Value 11) Least Absolute Deviation 12) Minimum Absolute Error 13) Galileo 14) Boscovich 15) Laplace

پس از ک وقفه نسبتا ولانی ادورث[16] رای روش LSE استفاده از توز نرمال رای خ اها و همچن ن ررسی نقا پرت را پ شنهاد می‌کند او از روش وسکو چ شر صفر ودن مجمو خ ا را حذف می‌کند و ن را ه عنوان اول ن روش قا ل کار رد رای رورد $(\beta_1, ..., \beta_k)$ در رگرس ون چندگانه[17] م رح می‌کند او همچن ن رای مدل رگرس ون خ ی ک الگور تم رای افتن ش خ روش LAE زمانی که عرض از مبدا خ صفر اشد ارا ه می‌دهد اما روش او ه دل ل دشواری محاسبه مورد استقبال قرار نمی‌گ رد

از ن ه عد علاقه ه روردهای LAE کم می‌شود ادورث دو اره مس له خود را م رح می‌کند رودز[18] ک روش تکراری[19] رای مس له او پ شنهاد می‌کند فار رادر[20] خش را ب ن د س نگل‌تون[21] ه پ شنهاد رودز م البی می‌افزا د

وقتی چارنس[22] و همکاران ارتبا روردهای LAE و رنامه‌ر زی خ ی[23] LP را نشان دادند او ا تغ رکرد ا ه‌کارگ ری LP افتن روردهای LAE ساده شد ا ظهور کامپ وترهای د ج تال رنامه‌های کامپ وتری ن ز توانمندتر شدند اررودال و رو رتز[24] ه ا ن من ور الگور تمی ه نام الگور تم ارا ه دادند

در است و کو نکر[25] ن ر ه توز حدی رورگرهای LAE را توسعه می‌دهند ا ن ن ر ه س اری از حدس ات قبلی را تا د کرد مثلا اگر در توز خ ا م انه نمونه‌ای در مقا سه ا م انگ ن نمونه‌ای رورد هتری رای پارامتر مکانی اشد رورگرهای LAE در مقا سه ا رورگرهای LSE دارای ناح ه‌های ا م نان اکدا کوچکتری خواهند ود

عد از سال ن ز کارهای متعددی روی LAE انجام می‌شود اما اساسا پ شنهاد الگور تم‌های محاسباتی کاراتر رای ا ن روش وده است

استنبا در اره پارامترها ا رورگرهای L_1 توس کو نکر و است م رح وگسترش داده می‌شود

در سال اول ن کنفرانس ن‌المللی تحت عنوان تحل ل داده‌های ماری ر پا ه نرم L_1 و روش‌های وا سته در ن وشاتل[26] سو س رگزار می‌شود

ک پ شرفت معنی‌دار زمانی رخ داد که توز مجانبی رورگرهای L_1 تحت شرا ع فتری اثبات شد

راو[27] و ای راو و ن[28] چند روش رای رورد پارامترهای مزاحم[29]

16) Edgeworth 17) Multiple Regression 18) Rhodes 19) Iterative Procedure
20) Farebrother 21) Singleton 22) Charnes 23) Linear Programing
24) Barrodale and Roberts 25) Bassett and Koenker 26) Neuchâtel 27) Rao
28) Bai, Rao and Yin 29) Nuisance Parameters

مدل و وار انس م انه نمونه‌ای را پ شنهاد کردند
رخی د گر از ماردانان همچون د ل‌من و فافن رگر[30] رای رورد پارامتر مزاحم از شب ه‌سازی استفاده کردند

چه وقت ا د از LAE استفاده کن م

عدا ثا ت می‌کن م که هرگاه خ اها iid و دارای توز لاپلاس دو نما ی اشند روش LAE هتر ن روش رورد است در ا ن حالت ا ن رودگرها MLE هستند و همچن ن ه ور مجانبی ناار و کارا هستند و کارا ی نسبی مجانبی رودگرهای LAE نسبت ه LSE را ر است

اگر خ اها دارای توز لاپلاس نباشند اما توز نها سنگ ن دم[31] اشد مانند توز کوشی لجست ک ا ... ا توز متقارنی که م انه در مقا سه ا م انگ ن رورد هتری اشد از هم رودگرهای LAE هتر از LSE هستند د ل‌من و فافن رگر را ب ن د
ک راه د گر تشخ ص ر مبنای ر کش دگی نمونه‌ای[32] است که ه صورت

$$k = \frac{\mu_4}{\sigma^4} = \frac{E[\epsilon_i^4]}{Var^2[\epsilon_i]}$$

تعر ف می‌شود هارتر[33] پ شنهاد داده است که اگر $k > /$ از L_1 اگر $/ \le k \le /$ از L_2 و اگر $k < /$ از L_∞ استفاده شود وی ا ن نت جه را ا مقا سه MSE رودگرها ه‌دست ورده است نا را ن رای توز های خ ای سنگ ن دم ه دل ل زرگ ودن k ا د از L_1 استفاده کرد

شولت نوردهالت[34] ان کرده است که رای توز نرمال رودگرهای L_2 خ ای کمتری نسبت ه L_1 دارند در حالی که رای توز های کوشی و لاپلاس عکس ن درست است اما چون توز خ ا معمولا مجهول است می‌توان از هر دو روش استفاده کرد و خ اها را مقا سه نمود

رورد پارامترها

در ا ن خش چند روش پ شنهادی را رای رورد مدل رگرس ون و در حالت کلی‌تر پارامترهای مدل خ ی ا ارا ه می‌ده م

30) Dielman and Pfaffenberger 31) Heavy Tail 32) Sample Kurtosis Ceofficient
33) Harter 34) Schulte-Nordholt

۱ مدل رگرسـ ون خـ ی

فرض کنـ د $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$ ، $i = \ ,...,n$ تعر ف می‌کنـ م

$$S = \sum S_i = \sum |\epsilon_i|.$$

لاپلاس ا تدا β_2 را ه صورت ز ر رورد می‌کند فرض کنـ د خـ رگرسـ ون از نقـ ه دلخواه (x_p, y_p) عبور کند نا را ن $y_p = \beta_1 + \beta_2 x_p + \epsilon_p$ همچـ ن فرض کنـ د $x_{ip} = x_i - x_p$ و $y_{ip} = y_i - y_p$ در نتـ جه

$$\begin{aligned} S_i &= |y_i - \beta_1 - \beta_2 x_i| = |y_i - \beta_1 - \beta_2 x_i - y_p + \beta_1 + \beta_2 x_p| \\ &= |y_{ip} - \beta_2 x_{ip}|. \end{aligned}$$

چون (x_{ip}, y_{ip}) معلوم است نا را ن S_i تنها تا عی از β_2 است S_i تا عی محد و در نتـ جه پـ وسته و دارای مـ نـ مم کتای $\beta_2 = \frac{y_{ip}}{x_{ip}}$ است پس $S = \sum S_i$ نـ ز محد است نمودار S ه صورت پاره‌خـ های ه هم پـ وسته‌ای است که در نقا ی ه ول $\frac{y_{ip}}{x_{ip}}$ $i = \ ,...,n$ شـ نـ‌ها تغـ ر می‌کند

اکنون دو حالت پـ ش می د

حالت اول نقـ ه‌ای ه فرم $\frac{y_{ip}}{x_{ip}}$ مـ نـ مم کتای S است که در ا ن صورت ا ن نقـ ه $\tilde{\beta}_2$ خواهد ود دقت می‌کنـ م که در ا ن حالت خـ از نقـ ه (x_i, y_i) می‌گذرد [۳۵]

حالت دوم حداقل دو نقـ ه ه فرم $\frac{y_{ip}}{x_{ip}}$ دارای عرض کسان هستند که در ا ن حالت $\tilde{\beta}_2$ ه صورت کتا ه‌دست نمی د و می‌تواند هر نقـ ه ن ا ن نقا اشد

ا نک در حالت کلی ه ررسی روردهای β_1 و β_2 مدل $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$ $i = \ ,...,n$ می‌پرداز م

مسـ له رورد مسـ له مـ نـ مم کردن $S = \sum |\epsilon_i| = \sum |y_i - \beta_1 - \beta_2 x_i|$ نسبت ه β_1 و β_2 است S_i شامل دو نـ م‌صفحه در فـ ای $(S_i; \beta_1, \beta_2)$ است که مـ نـ مم مقدار خود را زمانی می‌گـ رد که

$$y_i - \beta_1 - \beta_2 x_i = \ . \qquad (\ )$$

دقت می‌کنـ م که S_i ک تا دو متغـ ره محد و نا را ن پـ وسته است S عنی جمـ عمودی نـ م‌صفحه‌ها شامل قـ عه نـ م‌صفحات ه هم پـ وسته‌ای است که تشکـ ل ک چندوجهی[۳۶] می‌دهد کارست[۳۷] ا تـ لور[۳۸] را بـ نـ د n لبه چندوجهی S در امتداد خـ و ی است که در صدق می‌کنند پس مـ نـ مم S هم روی همـ ن خـ و خواهد ود که می‌تواند ک نقـ ه اشد ا خـ افقی و ا ک رو ه سـ ح [۳۹] الای صفحه (β_1, β_2)

ا ن مو و عدا ه عنوان ک از و ژگـ ها روردها LAE مـ رح مـ شود

36) Polyhedron 37) Karst 38) Taylor 39) Surface

حالت اول اگر م ن مم ک نق ه اشد   نگاه ا ن نق ه که فرض کن د اشتراک دو خ $y_j = \beta_1 - \beta_2 x_j$ و $y_k = \beta_1 - \beta_2 x_k$ اشد $\tilde{\beta}_1$ و $\tilde{\beta}_2$ را ه   ور کتا مشخص می‌کنند   در ا ن حالت خ LAE از نقا (x_j, y_j) و (x_k, y_k) می‌گذرد چون $e_j = e_k =$ نارولا[40] و همکاران   ا ن دو مشاهده را مشاهدات تعر ف شده[41] نام ده‌اند

حالت دوم اگر م ن مم ک خ   اشد ا ن فاصله سته مر و   ه مشاهده (x_j, y_j) است که توس   دو ن م‌صفحات مر و   ه دو مشاهده (x_g, y_g) و (x_h, y_h) محصور شده‌اند

حالت سوم اگر م ن مم ک رو ه اشد نگاه ا ن رو ه ک چند   لعی است که ه وس له حداقل سه خ   ه دست مده است

برورد پارامترها در مدل‌ها   خ ی

 ار د گر مدل   را در ن ر گ ر د $(\tilde{\beta}_1, ..., \tilde{\beta}_k)$ مقدار $S = \sum |\epsilon_i|$ را م ن مم می‌کنند   ا ه کارگ ری روشی مشا ه حالت خ ی می‌توان مس له را ه صورت ز ر تع ر کرد

S_i شامل دو ن م ا رصفحه[42] k عدی در ف   ای $(S_i; \beta_1, ..., \beta_k)$ است که م ن مم مقدار خود را زمانی می‌گ رد که $S_i =$ در نت جه S شامل ق عه ن م ا ر صفحات محد   ه هم پ وسته‌ای است که تشک ل ک ا ر چندوجهی[43] می‌دهد   نا ر ا ن م ن مم S می‌تواند ک نق ه   ا ک رو ه تخت[44] در ف   ای $(\beta_1, ..., \beta_k)$ اشد   ک نق ه از اشتراک k ا رصفحه ه شکل $S_i =$ ه دست می د   در ا ن حالت ا ن k نق ه تعر ف شده هستند و   را   را ه صورت کتا مشخص می‌کنند   ک رو ه ه صورت ک چند   لعی که ه وس له حداقل $k +$ ا رصفحه ه شکل $S_i =$ ه دست می د   کتا ن ست   رای حث   شتر ت لور   را   ن د

۳   براوردها   ماکس مم درستنما ی

فرض کن د ϵ_iها iid و دارای توز   لاپلاس اشند   عنی

$$f(\epsilon_i) = -exp\{-|\epsilon_i|\}, \quad \epsilon_i \in \Re$$

در ا ن صورت   رای مدل خ   ی   خواه م داشت

$$f(y_i) = -exp\{-|y_i - \sum_{j=1}^{k} x_{ij}\beta_j|\},$$

40) Narula 41) Defining Observations 42) Half-Hyperplane 43) Polytope
44) Flat Surface

و در نت جه تا درستنمای ه صورت ز ر در می د

$$L(\beta) = \prod_{i=۱}^{n} f(y_i) = \quad {}^{-n} exp\{-\sum_{i} |y_i - \sum_{j=۱}^{k} x_{ij}\beta_j|\}$$

و در نت جه

$$l(\beta) = ln\ L(\beta) = -n\ ln \quad - \sum_{i} |y_i - \sum_{j=۱}^{k} x_{ij}\beta_j|\}$$

روشن است که MLE پارامتر $\underline{\beta}$ ا $\underline{\tilde{\beta}}$ عنی رورد روش LAE کی است است و کو نکر ان کرده‌اند که تحت شرا ی عف ه شر توز لاپلاس رای خ اها روردگرهای $L_۱$ سازگار و مجانبا نرمال هستند

برنامه‌ر ز خ ی

مدل خ ی $\underline{y} = X\underline{\beta} + \underline{\epsilon}$ را می‌توان ه‌صورت $\underline{y} = X\underline{\beta} + \underline{\epsilon}^{+} - \underline{\epsilon}^{-}$ نوشت که در ن $\underline{\epsilon}^{+} = [\epsilon_۱^{+}, ..., \epsilon_n^{+}]'$ $\underline{\epsilon}^{-} = [\epsilon_۱^{-}, ..., \epsilon_n^{-}]'$ $\epsilon_i^{+} = max(\epsilon_i, \ )$ و $\epsilon_i^{-} = max(-\epsilon_i, \ )$
روشن است که $|\epsilon_i| = \epsilon_i^{+} + \epsilon_i^{-}$ و نا را ن

$$\sum |\epsilon_i| = \sum_{i} \epsilon_i^{+} + \sum_{i} \epsilon_i^{-} = \ '\underline{\epsilon}^{+} + \ '\underline{\epsilon}^{-}.$$

که در ن $[\ ,...,\ ] = \underline{\ }'$ روش LAE را می‌توان ه صورت مس له اول ه ز ر درورد

$$min\ \underline{\ }'\underline{\epsilon}^{+} + \underline{\ }'\underline{\epsilon}^{-} \qquad (\ )$$

ه شر ن که داشته اش م

$X\underline{\beta} + \underline{\epsilon}^{+} - \underline{\epsilon}^{-} = \underline{y}\ .\ \ \epsilon_i^{+}, \epsilon_j^{-} \geq$

در فرم استاندارد مس له اول ه را می‌توان ه صورت ز ر نوشت

$$min\ [\underline{\ }'\ \underline{\ }'\ \underline{\ }'] \begin{bmatrix} \underline{\beta} \\ \underline{\epsilon}^{+} \\ \underline{\epsilon}^{-} \end{bmatrix}$$

ه شر ن که داشته اش م

$$[XI - I] \begin{bmatrix} \underline{\beta} \\ \underline{\epsilon}^{+} \\ \underline{\epsilon}^{-} \end{bmatrix} = \underline{y}.$$

$\epsilon_i^{+}, \epsilon_j^{-} \geq \ .$

در ا نجا n محدود ت معادله و $n + k$ متغ ر ساختاری β_iها و ϵ_i^{+}ها و ϵ_i^{-}ها وجود دارد که می‌توان ا نرم‌افزارهای موجود مس له را ه سادگی حل کرد

روش کاهش مستق م

روش کاهش مستق م[45] که توسـ وسولوسکی[46] و جوس‌وانگر و اسپوز تو[47] پ شنهاد شده است ک روش مستق م رای م نـ مم کردن است ه خا ر دار م که S ک ا رچندوجهی محد ود که لبه‌ها ش هوسـ له اشتراک ق عه ن م ا رصفحه‌ها تشک ل می‌شد در ا ن روش از ک نق ه اول ه غاز می‌کن م و ه نزد کتر ن لبه می‌رس م در ول ا ن لبه پا نتر ن نق ه اشتراک را می ا م جا ی که ا ن لبه ه لبه د گر وصل می‌شود ا ن روش را تا رسـ دن ه پا ن‌تر ن نق ه ادامه می‌ده م انگ زه ا ن روش ر ا ن اساس است که هنگامی که ماتر س رح X دارای رتبه r اشد ک ا رصفحه L_1 ه نه از حداقل r نق ه مشاهدات عبور می‌کند که اگر دق قا از r نق ه گذرد روردها کتا هستند شتر نقا غاز ن روش‌های کاهش ه وسـ له روردهای LSE مشخص می‌شوند چون ا ن روردها اکثرا ه جوا ه نه نزد کتر هستند عنی پس از افتن روردهای LSE $k-$ داده مر و ه کمتر ن خ ا را در ن ر می‌گ ر م فرض کن د ا ن داده‌ها داده‌های l_1 l_{k-1} اشند پس تقر با دار م

$$y_{l_i} = \sum_j x_{l_ij}\beta_j + \epsilon_{l_i}, \quad i = \quad ,..., k -$$

از ا ن دستگاه خ ی $k-$ معادله همه پارامترها را ر اساس ک پارامتر و ژه مثلا β_m ه دست می ور م فرض کن د

$\beta_j = b_j\beta_m + c_j, j = \quad , \quad ,..., k, \quad j \neq m.$

در نت جه مسـ له م نـ مم کردن ه صورت ز ر در می د

$\min_{\beta_1,...,\beta_k} \sum_i |y_i - \sum_j x_{ij}\beta_j|$
$= \min_{\beta_m} \sum_{i\neq l_1,...,l_{k-1}} |y_i - \sum_j x_{ij}\beta_j|$
$= \min_{\beta_m} \sum_{i\neq l_1,...,l_{k-1}} |y_i - x_{i1}(b_1\beta_m + c_1) - ... - x_{im}\beta_m - ... - x_{ik}(b_k\beta_m + c_k)|$
$= \min_{\beta_m} \sum_i |y_i - \sum_{j\neq m} x_{ij}c_j - \beta_m \sum_{j\neq m} x_{ij}b_j - x_{im}\beta_m|$
$= \min_{\beta_m} \sum_i |x_{im} + \sum_{j\neq m} x_{ij}b_j| \times |\frac{y_i - \sum_{j\neq m} x_{ij}c_j}{x_{im} + \sum_{j\neq m} x_{ij}b_j} - \beta_m|$

ا نک فرض کن د $d_i = \frac{y_i - \sum_{j\neq m} x_{ij}c_j}{x_{im} + \sum_{j\neq m} x_{ij}b_j}$ $w_i = |x_{im} + \sum_{j\neq m} x_{ij}b_j|$ و $e = \beta_m$ در ا ن صورت دار م

$\min_{\beta_1,...,\beta_k} \sum |\epsilon_i| = min_e \sum w_i|d_i - e|$

دقت می‌کن م که w_iها و d_iها مقاد ر معلومی هستند مسـ له م نـ مم کردن مجمو اخ ر ک مسـ له شناخته شده‌ای است که لاپلاس ن را در قرن ه جدهم حل کرده است رای حل مسـ له فرض کن د متغ ر تصادفی Y مقاد ر d_1 d_n را ا احتمال‌های متنا ر p_1 p_n اخت ار می‌کند

45) Direct Descent Method 46) Wesolowsky 47) Josvanger and Sposito

که در آن $p_i = \frac{w_i}{\sum w_i}$ در این صورت داریم
$E|Y-e| = \sum p_i|d_i - e| = \frac{1}{\sum w_i}\sum wi|d_i - e|.$
پس مسئله مینیمم کردن $\sum wi|d_i - e|$ هم ارز مینیمم کردن $E|Y-e|$ است. از طرفی می‌دانیم که e باید میانه متغیر تصادفی Y باشد.
در حالت رگرسیون خطی داریم

$$min_{\beta_1,\beta_2}|y_i-\beta_1-\beta_2x_i| = min_{\beta_1}\sum_i |1-\frac{x_i}{x_j}|\times|\frac{y_i-\frac{y_jx_i}{x_j}}{1-\frac{x_i}{x_j}}-\beta_1|$$

$$= min_{\beta_2}\textstyle\sum_i|x_i-x_j|\times|\frac{y_i-y_j}{x_i-x_j}-\beta_2|.$$

در تساوی اول $w_i = |1-\frac{x_i}{x_j}|$ و $d_i = \frac{y_i-\frac{y_jx_i}{x_j}}{1-\frac{x_i}{x_j}}$ و $e=\beta_1$ و در تساوی دوم $w_i=|x_i-x_j|$ و $d_i = \frac{y_i-y_j}{x_i-x_j}$ و $e=\beta_2$ است.

الگوریتمی برای مینیمم کردن نرم L_1 برای مدل‌های غیرخطی با محدودیت‌های غیرخطی

سلیمان و کریستنسن[48] الگوریتمی برای مینیمم کردن نرم L_1 برای مدل‌های غیرخطی با محدودیت‌های غیرخطی ارائه کرده‌اند. به صورت خلاصه این الگوریتم به صورت زیر است. در آغاز مدل خطی $\underline{y}=X\underline{\beta}+\underline{\epsilon}$ را با محدودیت خطی $B\underline{\beta}=\underline{c}$ در نظر بگیرید که در آن B یک ماتریس $l\times k$ و $\underline{c}$ یک بردار l بعدی است. تعریف می‌کنیم $H=[X'\ B']'$، $\underline{z}=[\underline{y}'\ \underline{c}']'$ و $\underline{\epsilon}^*=[\underline{\epsilon}'\ \underline{\ }']'$ برای مدل خطی $\underline{z}=H\underline{\beta}+\underline{\epsilon}^*$ برآورد $\underline{\beta}$ به روش LSE و سپس انحراف معیار باقیمانده‌ها را می‌یابیم. از داده‌های مدل اصلی داده‌های پرت یعنی داده‌هایی که قدر مطلق باقیمانده آن‌ها از انحراف معیار باقیمانده‌ها بیشتر است را کنار می‌گذاریم. حداکثر به تعداد $n+l-k$ و سپس برای بقیه داده‌ها $\underline{\beta}$ را به روش LSE برآورد می‌کنیم و باقیمانده‌ها را دوباره محاسبه می‌کنیم. در پایان $k-l$ داده با کمترین باقیمانده و همچنین l محدودیت را در یک معادله به فرم $X^*\underline{\beta}=z^*$ قرار می‌دهیم. $\tilde{\underline{\beta}}$ جواب معادله فوق خواهد بود.

به عنوان مثال فرض کنید می‌خواهیم پارامترهای مدل رگرسیون خطی $y_i=\beta_1+\beta_2x_i$ را با توجه به نقاط $\{(\ , \ ),(\ , \ ),(\ , \ ),(\ , \ ),(\ , \ )\}$ و با محدودیت $\ \beta_2 + \beta_1 = \ $ برآورد کنیم. روشن است که $n=\ $ ، $k=\ $ و $l=\ $ و همچنین

48) Soliman and Christensen

$$H = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}, \quad \underline{z} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_۱ \\ \beta_۲ \end{bmatrix}$$

رورد LSE ه صورت ز ر ه‌دست می د

$$\hat{\underline{\beta}} = (H'H)^{-۱}H'\underline{z} = \begin{bmatrix} / \\ / \end{bmatrix}.$$

از روی اق مانده‌ها مقدار انحراف معار رار / $s =$ می‌شود از رفی ردار قدر م لق اق مانده‌ها داده‌ها ه صورت ز ر ه دست می د

$$|\underline{e}| = \begin{bmatrix} / \\ / \\ / \\ / \\ / \end{bmatrix},$$

پس داده پنجم عنی (,) داده پرت و اد حذف شود ا قه داده‌ها رورد $\underline{\beta}$ رار $\hat{\hat{\underline{\beta}}} = [\ /\ \ \ /\ \ ]$ می‌شود و اان رورد ردار اق مانده‌ها ه صورت ز ر ه دست می د

$$|\hat{\underline{e}}'| = [\ /\ \ \ /\ \ \ /\ \ \ /\ \ \ /\ \ ].$$

حال چون رتبه ماتر س H رار است ا توجه ه حث‌های پ ش ن رورد‌گر LAE حداقل دو نق ه را رازش می‌دهد کی نق ه محدود ت که دون خ است و کی نق ه‌ای که در ن داده‌ها دارای کمتر ن اق مانده است نا را ن دار م

$$H^*\underline{\beta} = z^*,\ H^* = \begin{bmatrix} \\ \end{bmatrix}, \underline{z}^* = \begin{bmatrix} \\ \end{bmatrix}$$

و در نت جه

$$\tilde{\underline{\beta}} = (h^*)^{-۱}\underline{z}^* = \begin{bmatrix} ۲ \\ \frac{}{۳} \end{bmatrix}.$$

ا نک مدل غ رخ ی $y_i = f_i(x_i, \underline{\beta}) + \epsilon_i$ $i = \ ,...,n$ ا محدود ت $g_i(x_i, \underline{\beta}) =$ $i = \ ,...,l$ را در ن ر گ ر د رای مدل فوق ا تدا y_i و $g_i(x_i, \underline{\beta})$ را ا استفاده از دو جمله اول سه تا لور تقر می‌زن م و سپس از روش خ ی قبل استفاده می‌کن م رای شرح شتر ه سلمان و کر ستن‌سن مراجعه کن د

استنبا   در باره پارامترها   مدل

بررسی جنبه‌های استنبا ی در مدل‌های خ ی نقش به‌سزا ی در تجز ه وتحل ل داده‌های ماری دارد   در بخش بعد به بررسی توز   مجانبی برآوردگرهای L_1 و چگونگی استنبا   در باره پارامترهای مدل می‌پرداز م

۱   توز   مجانبی برآوردگرها L_1

فرض کن د در مدل خ ی   توز ϵ_iها به صورت $F(.)$ باشد   در ا ن صورت توز y_iها به صورت $F(y_i - \sum_{j=1}^{k} x_{ij}\beta_j)$ می‌شود   همچن ن فرض کن د که م انه توز F صفر باشد و داشته باش م

الف F پ وسته و دارای تا   چگالی مثبت و پ وسته در م انه باشد

$\lim n^{-1}(X'X) = Q$ ک ماتر س مثبت مع ن[49] باشد [50]

است و کو نکر   ثا ت کرده‌اند که تحت شرا   فوق $\tilde{\beta}$ ک برآوردگر سازگار $\underline{\beta}$ است و همچن ن دار م

$$\sqrt{n}(\tilde{\underline{\beta}} - \underline{\beta}) \rightarrow N_k[\ ,\lambda^2(n^{-1}X'X)^{-1}] \quad (\ )$$

که در ن $\frac{\lambda^2}{n}$ وار انس مجانبی م انه نمونه‌ای مر و   ه متغ ر تصادفی با توز F است   دقت کن د که ن ازی ن ست که توز F حول م انه متقارن باشد شر $F(\ ) = \frac{1}{2}$ کافی است

بنا بر ق ه (iv) شاو[51]   وار انس م انه نمونه‌ای با تا   چگالی f با م انه توز m به ور مجانبی برا ر

$$Var[m] = \frac{\quad}{n[f(m)]^2};$$

است   در نت جه $\lambda = \frac{1}{2f(m)}$

شباهت $\sigma^2(X'X)^{-1}$ در LSE و $\lambda^2(X'X)^{-1}$ در LAE شا ان توجه است   بنا برا ن ملاکی ن کو برای استفاده از LAE یا LSE مقا سه σ^2 و λ^2 است   می‌توان نت جه گرفت که وقتی توز   خ ا به گونه‌ای باشد که م انه در مقا سه با م انگ ن برآورد بهتری باشد وار انس کمتری داشته باشد   برآوردگرهای L_1 ی‌های با م نان اک دا کوچکتری خواهند داشت   ا ن شر   برای گستره وس عی از توز ها که دارای ماکس مم م لق در م انه و با سنگ ن دم باشند برقرار است

49) Positive Definite

اقتصاددانان به ماتر س X که در ا ن شر   صدق م کند ماتر س خوش رح $well - design$ م گو ند

51) Shao

برورد پارامتر مزاحم λ

استنبا در اره $\underline{\beta}$ ا استفاده از ه دل ل وجود پارامتر مزاحم λ[۵۲] ممکن ست د د م که ه ور مجانبی دار م $\lambda = \frac{۱}{۲f(m)}$ فرض کن د اق مانده‌های مرت شده ه صورت $e_{(۱)}$ $e_{(n)}$ اشند در ا ن صورت دار م

$$\hat{f}(m) \simeq \frac{F_n(e_{(t)}) - F_n(e_{(s)})}{e_{(t)} - e_{(s)}} = \frac{\frac{t}{n} - \frac{s}{n}}{e_{(t)} - e_{(s)}} = \frac{(t-s)/n}{e_{(t)} - e_{(s)}},$$

که البته ا د t و s متقارن حول اند س م انه نمونه‌ای اق مانده‌ها اشند و اختلاف چندانی هم از ا ن اند س نداشته اشند نا ر ا ن مقدار $t - s$ ا د کوچک و زوج اشد در حالت و ژه اگر $t = [n/\ ] + l$ و $s = [n/\ ] - l$ l ا د عدد بعی کوچکی اشد ه عنوان مثال در ک نمونه تا ی اند س م انه را ر می‌شود و در نت جه اگر داشته اش م $t - s =$ نگاه خواه م داشت $t =$ و $s =$ پس از افتن t و s مناس رورد λ ه صورت ز ر ه‌دست می د

$$\hat{\lambda} = [\ \hat{f}(m)]^{-۱}.$$

اسپوز تو ه صورت تجر ی مقاد ر مناس $t - s$ را رای چند توز رای مقاد ر مختلف n ارا ه داده است

نقصان اساسی ا ن رز رورد λ عدم کتا ی ن است چون مقاد ر متفاوت $t - s$ ه روردهای گوناگون می‌انجامد نا ر ا ن ک زمون س ح α واقعا دارای س ح α و ک ازه ا م نان ا ر ا م نان $\alpha -$ واقعا دارای ر ا م نان $\alpha -$ ست

روش‌های س اری رای رورد λ وجود دارد از جمله روش وت‌استراپ[۵۳] که توس مک‌ک ن و شرادر[۵۴] استاجن‌هوس[۵۵] د ل‌من و فافن رگر و . . . پ شنهاد شده است

۳ زمون فرض در مدل‌ها خ ی

فرض خ ی $B\underline{\beta} = \underline{c}$ که در ن B ک ماتر س $p \times k$ ا رتبه $p \leq k$ و ا ک ف ای ستونی که ز ر ف ای ستونی X است را در ر گ ر د در LSE ا فرض نرمال ودن ماره‌های زمون کوچک نمونه‌ای دارای توز نرمال t ا F هستند اما در مبحث LAE رای نمونه‌های کوچک توز های دق ق معلوم ستند در روش LSE سه روش رای زمون فرض در اخت ار است ا ن روش‌ها را می‌توان رای روش LAE تعم م داد ا ن سه روش عبارتند از روش والد[۵۶] W روش نسبت درستنما ی[۵۷] LR و روش را لاگرانژ[۵۸] LM در خش‌های تی

ا ن پارامتر ه تا پراکندگ $sparcity\ function$ ز معروف است

53) Bootstrap 54) Mckean and Schrader 55) Stagenhaus 56) Wald
57) Likelihood Ratio 58) Lagrange Multiplier

ه ررسی ا ن روش‌هامی‌پرداز م

۱ ۳ زمون والد

توج ه و انگ زه زمون والد ساده است اما ممکن است ه دست وردن ماره‌ها گاهی ز اد ساده نباشد ا توجه ه فرض $B\underline{\beta}=\underline{c}$ دار م

$$\sqrt{n}(B\underline{\tilde{\beta}}-B\underline{\beta}) \rightarrow_d N_k(\ ,\lambda^{۲}B(n^{-۱}X'X)^{-۱}B'), \qquad (\ )$$

و همچن ن

$$W=n(B\underline{\tilde{\beta}}-B\underline{\beta})'[\lambda^{۲}B(n^{-۱}X'X)^{-۱}B']^{-۱}(B\underline{\tilde{\beta}}-B\underline{\beta}) \rightarrow_d \chi^{۲}_p.$$

نا را ن اگر فرض $B\underline{\beta}=c$ درست اشد توز حدی

$$\hat{W}=\hat{\lambda}^{-۲}(B\underline{\tilde{\beta}}-\underline{c})'[B(n^{-۱}X'X)^{-۱}B']^{-۱}(B\underline{\tilde{\beta}}-\underline{c})$$

$\chi^{۲}_p$ می‌شود و می‌توان ا مقا سه $\hat{W}$ ا $\chi^{۲}_{p;\alpha}$ زمون را در س ح α انجام داد

توج ه زمون والد ا ن است که وقتی فرض درست اشد $B\underline{\tilde{\beta}}$ ه مقدار $B\underline{\beta}=c$ نزد ک می‌شود و نا را ن مقدار $B\underline{\tilde{\beta}}-c$ ه صفر نزد ک می‌شود و در نت جه زرگی ماره $\hat{W}$ نادرستی فرض را نت جه می‌دهد

اگر B ساختار و ژه‌ای داشته اشد زمون فرض را می‌توان ه ور مستق م از فرمول زمون کرد در دنباله چند حالت و ژه ررسی شده است

$B_۱=(b_۱,...,b_k)$ و $\underline{c}$ ردار ثا تی اشد در ا ن صورت ماره زمون ه صورت ز ر است

$$\frac{B_۱\underline{\tilde{\beta}}-c}{\hat{\lambda}\sqrt{B_۱(X'X)^{-۱}B'_۱}} \rightarrow_d N(\ , \ ).$$

$B_۲=(\ ,..., \ , \ , \ ,..., \ )$ و $c=$ در ا ن حالت از ماره

$$\frac{\underline{\tilde{\beta}}_i}{\hat{\lambda}\sqrt{(X'X)^{-۱}}} \rightarrow_d N(\ , \ ),$$

استفاده می‌کن م

۳ زمون نس ت درستنما

فرض کن د مقدار سوپر مم تا درستنما ی را تحت فرض ا $L(\underline{\tilde{\beta}}_\circ)$ و دون فرض ا $L(\underline{\tilde{\beta}})$ نما ش ده م در ا ن صورت ماره زمون نسبت درستنما ی ه صورت ز ر تعر ف می‌شود

$$\Lambda = \frac{L(\underline{\tilde{\beta}}_\circ)}{L(\underline{\tilde{\beta}})}.$$

اگر فرض درست اشد Λ تقر با را ر و اگر نادرست اشد تقر با صفر می‌شود نا را ن کوچکی مقدار ن حاکی از نادرستی فرض است و در نت جه گواه رای رد کردن ن ز اد است ا فرض ا ن که توز خ اها لاپلاس اشد چگالی توام $\underline{\epsilon}$ ه صورت

$$f(\underline{\epsilon};\underline{\beta}) = (-)^n exp\{-\sum |y_i - x_i'\underline{\beta}|\},$$

است که در ن x_i' سه ر i ام ماتر س X است پس

$$-۲ln\ \Lambda = ۲\sum |y_i - X_i\underline{\tilde{\beta}}_\circ| - ۲\sum |y_i - X_i\underline{\tilde{\beta}}| = ۲[SAD(\underline{\tilde{\beta}}_\circ) - SAD(\underline{\tilde{\beta}})] \rightarrow_d \chi^۲_p,$$

که SAD نشان دهنده مجمو قدرم لق انحرافات[59] است در واق روش LRT کاهش SAD را اندازه‌گ ری می‌کند از د دگاه هندسی ا ن روش فاصله نرم $L_۱$ ن $\underline{y}$ و دو ز ر ف ای تحت فرض و دون فرض را مقاس ه می‌کند مک ک ن و شرادر را ن د ا داشتن دو شر اول ه الف و کو نکر ثا ت کرده است که اگر فرض درست اشد دار م

$$\hat{\lambda}^{-۱}[SAD(\underline{\tilde{\beta}}_\circ) - SAD(\underline{\tilde{\beta}})] \rightarrow_d \chi^۲_p.$$

روش LRT زمانی مناس است که $\underline{\tilde{\beta}}_\circ$ ه سادگی دست افتنی نباشد

۳ ۳ زمون را لاگرانژ

زمون را لاگرانژ LMT ر مبنای مس له ه نه‌سازی ز ر است

$min \underline{\ }'\underline{\epsilon}^+ + \underline{\ }'\underline{\epsilon}^-$

ه شر ن که داشته اش م

$B\underline{\beta} = \underline{c}$ $X\underline{\beta} + \underline{\epsilon}^+ - \underline{\epsilon}^- = \underline{y}$ $\epsilon_i^+, \epsilon_j^- \geq$ ا ن مس له را می‌توان ه صورت ز ر ن ز فرمول ندی کرد

$$min[\underline{\ }'\ \underline{\ }'\ \underline{\ }'] \begin{bmatrix} \underline{\beta} \\ \underline{\epsilon}^+ \\ \underline{\epsilon}^- \end{bmatrix}$$

59) Some of Absolute Error

به شرط آنکه

$$\begin{bmatrix} X & I & I \\ B & - & - \end{bmatrix} \begin{bmatrix} \underline{\beta} \\ \underline{\epsilon}^{+} \\ \underline{\epsilon}^{-} \end{bmatrix} = \begin{bmatrix} \underline{y} \\ \underline{c} \end{bmatrix}.$$

کونکر ثابت کرده است که تحت فرض داریم

$$L = (X'\underline{\delta})'Q^{-1}(X'\underline{\delta}) \to \chi^2_p,$$

که در آن Q در شرط صدق می‌کند و $\underline{\delta} = (\delta_1, ..., \delta_n)'$ و $\delta_i = sign(y_i - X_i\tilde{\underline{\beta}})$
$i = \ , ..., n$

روش LMT به ویژه هنگامی مناسب و ساده است که مدل خطی را بتوان به صورت $\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{\epsilon}$ نوشت و بخواهیم فرض $\underline{\beta}_2 = \_$ را آزمون کنیم. در این حالت داریم

$$B = [\_\ I], \ \ \underline{c} = \_, \ \ X = [X_1, X_2], \quad Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}.$$

بنابراین

$$\begin{bmatrix} \\ X_2^{-}\underline{\delta} \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} \\ X_2^{-}\underline{\delta} \end{bmatrix} = (X_2'\underline{\delta})Q_{22}(X_2'\underline{\delta}) \to_d \chi^2_p.$$

ضریب تعیین

یک معیار برای بررسی صحت و درستی مدل در روش LSE ضریب تعیین[60] است که به صورت

$$R^2 = \frac{SST - SSE}{SST},$$

تعریف می‌شود که در آن SST مجموع مربعات کل[61] و SSE مجموع مربعات انحرافات[62] است.

در واقع R^2 درصد یا میزان تغییراتی را نشان می‌دهد که توسط مدل توضیح داده می‌شود. در نتیجه برای یک مجموعه داده بهترین مدل، مدلی است که با کمترین تعداد پارامتر دارای بیشترین R^2 باشد.

مک‌کین و سی‌ورس[63] معیاری همانند R^2 در مبحث LAE به صورت زیر پیشنهاد داده‌اند

$$R_1 = \frac{SAD(med) - SAD(\tilde{\underline{\beta}})}{SAD(med)},$$

60) Ceofficient of Detemination 61) the Total of Sum of Squares 62) the Sum of Square Deviations 63) Mckean and Sievers

که در ن $SAD(med) = \sum |y_i - y_{med}|$

$R_۱$ دارای و ژگی‌های ز ر است

ا ن مع ار می‌تواند ه عنوان مقا اسی رای صحت رازش مدل ه کار رده شود

ا ن مع ار دون عد است و تحت تبد لات مکانی و مقا اسی[64] در y_iها و x_iها ناوردا[65] است

$\leq R_۱ \leq$ که صفر و ک ه ترت عنی مناس نبودن و مناس ودن مدل

هنگامی که ک مدل ساده را ه ک مدل پ چ ده‌تر ا تعداد پارانتر شتر تبد ل می‌کن م $R_۱$ صعود می‌کند

رای زمون فرض $\mathcal{H}_۰ : \underline{\beta} = \_$ ا ن مع ار نمی‌تواند استفاده شود رعکس $R^۲$ در مبحث LSE

در را ر داده‌های پرت ن رومند ن ست درست مثل $R^۲$

مراج

[1] Bai, Z. D., Rao, C. Radhakrishna and Yin, Y. Q., (1990) Least Absolute Deviations Analysis of Variance. *Sankhyā* : The Indian Journal of Statistics, Vol 52, series A, Part 2, 166-177.

[2] Barrodale, I. and Roberts, F. D. K. (1974) Algoritm 478: Solution of an Obverdetermined System of Equations in the L_1 Norm. Communications of the ACM, Vol 17, No 6, 319-320.

[3] Bassett, Gilbert and Koenker, Roger (1978) Asymptotic Theory of Least Absolute Error Regression. Journal of the American Statistical Association, Vol 73, 618-622.

[4] Boscovich, R. J.,(1757) De Literaria Expeditione per Pontificiam Ditionem, et Synopsis Amplioris Operis, ac Habentur Plura Eius Ex Exemplaria Etiam Sensorum Impressa. Bononiensi Scientiarum et Artium Instituto Atque Academia Commentarii, Vol 4, 353-396.

[5] Charnes, A., Cooper, W. W. and Ferguson, R. O. (1955) Optimal Estimation of Executive Compensation by Linear Programing. Management Science, Vol 1, 138-150.

[6] Dielman, Terry and Pfaffenberger, Roger (1992) A Further Comparison of Tests of Hypothsis in LAV Regression. Computational Statistics & Data Analysis, Vol 14, 375-384.

[7] Edgeworth, F. Y. (1887) On Observations Relating to Several Quantities. Philosophical Magazine(5th series, Vol 24, 222.

[8] Edgeworth, F. Y. (1923) On the use of Medians for Reducing of Observations. Philosophical Magazine, Vol 66, 1074-1088.

64) Location and Scale Transformations 65) Invariant

[9] Farebrother, R. W. (1987) The Historical Development of the L_1 and L_∞ Estimation Procedures: 1793-1930. In Y. Dodge (ed.), Statistical Data Analysis Based on the L_1 Norm and Related Methods, Amesterdam: North Holland, 37-63.

[10] Josvanger, Lee Ann and Sposito, Vincent A., (1983) L_a Norm Estimates for the Simple Regression Problem. Communaictions in Statistics, Simulation and Computation, Vol 12(2), 215-221.

[11] Harter, H. L. (1977) Nonuniqueness of Least Absolute Values Regression. Communaictions in Statistics Theory and Methods, Vol A6, 829-838.

[12] Koenker, Roger (1987) A Comparison of Asymptotic Testing Methods for L_1 Regression. In Y. Dodge(ed.) Statistical Data Analysis Based on L_1 Norm and Related Methods, Amesterdam: North Holland, 287-295.

[13] Koenker, Roger and Bassett, Gilbert (1982) Tests of Hypothesis and L_1 Estimation. Econometrica, Vol 50, No 6, 1577-1583.

[14] Karst, Otto J. (1958) Linear Curve Fitting Using Least Deviations. Journal of the American Statistical Associations, Vol 53, 118-132.

[15] Laplace, Pierre Simon(1793) Sur Queques Points du Systeme du Monde. Memories de L'Academie Royale des Sciences de Paris, 1-87. Reprinted (1895) in Ouvers Completes de Laplace, Vol 11, Paris: Gauthier-Villars, 447-558.

[16] Legendre, Adrian Marie(1805) Nouvelles Methodes pour la Determination des Orbites des Cometes. Appendice sur la Methodes des Moindres Quarres, Paris: Courcier, 72-80.

[17] Mckean, Joseph W. and Schrader, Roland M. (1987) Least Absolute Errors Analysis of Varince. In Y. Dodge(ed.), Statistical Dta Analysis Based on L_1 Norm and Related Methods, Amesterdam: North Holland, 297-305.

[18] Mckean, Joseph W. and Sievers, gerland L. (1987) Coefficients of Determination for Least Absolute Deviation Analysis. Statistics & Probability Letters, Vol 5, 49-54.

[19] Narula, Subash C., Sposito, Vincent A. and Wellington, John F., (1993) Intervals Which Leave the Minimum Sum of Absolute Error Regression Unchanged. Applied Statistics, Vol 26 , 106-111.

[20] Peou, Poon Chee (1993) The Least Absolute Errors in Regression. (B.Sc. thesis), National University of Singapore.

[21] Rao, C. Radhakrishna (1988) Methodology Based on the L_1- $Norm$ in Statistical Inferences. $Sankhy\bar{a}$: The Indian Journal of Statistics, Vol 50, series A, Part 3, 289-313.

[22] Rhodes E. C., (1930), Reducing Observations by the Method of minimum deviation, Philosophical magazine 9, 974-992.

[23] Shao, Jun (1999) Mathematical Statistics, Springer-Verlag, New York.

[24] Schulte-Nordholt, E., (1992) LAV Regression, in Dutch. Kwantitatieve Methoden, Vol 39, 39-52.

[25] Singleton, R. R., (1940) A Method for Minimizing the Sum of Absolute Values of Deviatins. Annals of Mathematical Statistics, Vol 9, 301-310.

[26] Soliman S.A., and Christensen G.S.,(1991) A New Algoritm for Non Linear L_1 Norm Minimization with Non-Linear Equality Constraints, Computational Statistics & Data Analysis, 11, 97-109.

[27] Sposito, Vincent A. (1990) Some Properties of $L-p$ Estimators. In K.D. Lawrence and J. L. Arthur(eds.), Robust Regression: Analysis and Applications, NY: Marcel Dekker, 23-58.

[28] Stagenhaus, Gabrila (1987) Bootstrap and Inference Procedures for L_1 Regression. In Y. Dodge (ed.), Statistical Data Analysis Based on The L_1 Norm and Related Methods, Amesterdam: North Holland, 323-332.

[29] Taylor, Lester D. (1974) Estimation by Minimizing the Sum of Absolute Errors. In P. Zarembka (ed.), Fronriers in Econometrics, NY: Academic Press, 169-180.

[30] Wesolowsky, G. O. (1981) A New Descent Algoritm for the Least Absolute Value Regression Problem. Communaictions in Statistics, Simulation and Computation, Vol B10(5), 479-491.