بسم الله الرحمن الرحیم

دانشکده علوم ریاضی

رشته گروه آمار، گرایش آمار ریاضی

پایان‌نامه کارشناسی ارشد

سازگاری قوی M ـ برآوردها در مدل‌های خطی با خطاهای وابسته زیرجمعی منفی

نگارنده: حسین حیدری نژاد

استاد راهنما

دکتر نگار اقبال

استاد مشاور

دکتر حسین باغیشنی

بهمن ۱۳۹۷

پایان‌نامه خود را تقدیم می‌کنم به پدر و مادر و خانواده‌ام که در تمام سختی‌ها و مشکلات در کنارم بوده‌اند.

سپاس‌گزاری...

خدا را بی‌نهایت سپاس‌گزارم به خاطر تمام نعمت‌هایی را که به من داده و از زحمت‌های خانواده‌ام که همیشه در کنار من بوده‌اند سپاس‌گزارم هم‌چنین از استاد گرامی سرکار خانم دکتر نگار اقبال و استاد محترم جناب آقای دکتر حسین باغیشنی به خاطر زحمت‌ها و راهنمایی‌هایی که در طول این پایان نامه برای من کشیده‌اند تقدیر و تشکر می‌کنم.

حسین حیدری نژاد
بهمن ۱۳۹۷

تعهد نامه

اینجانب **حسین حیدری نژاد** دانشجوی کارشناسی ارشد رشته **آمارریاضی** دانشکده **علوم ریاضی** دانشگاه صنعتی شاهرود، نویسنده پایان‌نامه با عنوان **سازگاری قوی M ـ برآوردها در مدل‌های خطی با خطاهای وابسته زیرجمعی منفی** ، تحت راهنمایی **نگار اقبال** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های دیگر پژوهش‌گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان‌نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارایه نشده است.
- حقوق معنوی این اثر، به دانشگاه صنعتی شاهرود تعلق دارد، و مقالات مستخرج با نام ‘‘ دانشگاه صنعتی شاهرود ’’ یا “ Shahrood University of Technology “ به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به‌دست آوردن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند، در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

حسین حیدری نژاد
بهمن ۱۳۹۷

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان‌نامه بدون ذکر منبع مجاز نمی‌باشد.

چکیده

در این پایان‌نامه شرایط سازگاری قوی M ـ برآوردهای پارامترهای مدل خطی با خطاهای تصادفی وابسته زبرجمعی منفی (NSD) مورد بررسی قرار گرفته است. نتایج موجود بهبود شرایط گشتاوری و تعمیمی از روش‌های ارائه شده در حالت استقلال به خطاهای NSD می‌باشند.

کلمات کلیدی: سازگاری قوی، وابسته زبرجمعی منفی، رگرسیون استوار.

پیشگفتار

همگرایی کامل نقش مهمی را در قضیه حد مرکزی و به‌دست آوردن مرتبه همگرایی یک دنباله از متغیرهای تصادفی ایفا می‌کند. متغیرهای تصادفی زبرجمعی منفی یک رده بسیار وسیع از دنباله‌های وابسته هستند که به عنوان مواردی خاص از این رده، می‌توان به دنباله متغیرهای مستقل و متغیرهای پیوندی منفی اشاره نمود. مفهوم متغیرهای تصادفی وابسته زبرجمعی منفی مبتنی بر رده توابع زبرجمعی است. از آن‌جا که وابستگی زبرجمعی منفی تعمیمی از وابستگی پیوندی منفی است، در نامساوی‌های گشتاوری و هم‌چنین نامساوی‌های احتمالی متعددی مفید واقع می‌شود. محققین این حوزه از همگرایی کامل متغیرهای تصادفی وابسته زبرجمعی منفی در بسیاری از زمینه‌های استنباط آماری، از جمله مدل‌های رگرسیونی، استفاده کرده‌اند. در اکثر موارد شرط مستقل بودن برای داده‌هایی که در اختیار داریم، برقرار نیست. بنابراین، در این پایان‌نامه علاوه بر بررسی برخی ویژگی‌های اساسی متغیرهای تصادفی وابسته زبرجمعی منفی، سازگاری قوی M ـ برآوردها در مدل‌های خطی با جملات خطای وابسته زبرجمعی منفی را تحت برخی شرایط مورد مطالعه قرار می‌دهیم. هم‌چنین در انتها با مطالعه شبیه‌سازی به ارزیابی نتایج نظری می‌پردازیم. محتوای مطالب این پایان‌نامه به‌صورت زیر تدوین شده است:

- در فصل اول به بیان برخی تعاریف و مفاهیم اولیه که در اثبات نتایج اصلی به آن‌ها نیاز داریم، می‌پردازیم.
- در فصل دوم مفهوم استواری و برخی روش‌های برآورد پارامترهای مدل رگرسیونی استوار معرفی می‌کنیم.
- در فصل سوم شرایط همگرایی کامل سازگاری قوی M ـ برآوردها با خطاهای NSD را بررسی می‌کنیم.
- در فصل چهارم با استفاده از ابزارهای شبیه‌سازی نتایج نظری همگرایی کامل فصل سوم را ارزیابی می‌کنیم.
- پیوست پایان‌نامه شامل کدهای نرم افزار R برای باز تولید نتایج شبیه‌سازی می‌باشد.

فهرست مطالب

فهرست تصاویر

فهرست جداول

فصل ۱

مفاهیم اولیه

در این فصل برخی تعاریف و قضیه‌هایی که در فصل‌های بعدی به آن احتیاج داریم بیان می‌کنیم، برای گردآوری عمده مطالب این فصل از هو (۲۰۰۰)، گات (۱۹۹۴)، مازی و کاتز (۲۰۰۱)، پایان نامه ارشد پاشاکایی، مقاله اسدی، فلاح زاده، نوقایی و کتاب رگرسیون خطی عبدالرضابازرگان لاری (۱۳۹۱) استفاده کرده‌ایم.

۱.۱ انواع همگرایی

همگرایی کامل نقش مهمی در احتمال، قضایای حدی، آمار ریاضی به ویژه در تعیین نرخ همگرایی دارد، این همگرایی اولین بار توسط هسو و رابینز (۱۹۴۷) بیان شد، به صورت زیر تعریف شد.

تعریف ۱.۱.۱. دنباله‌ای از متغیرهای تصادفی $\{X_n, n \geqslant 1\}$ را همگرای کامل[1] به متغیر تصادفی X گویند، هرگاه به ازای هر $\varepsilon > 0$ داشته باشیم

$$\sum_{n=1}^{+\infty} P(|X_n - X| > \varepsilon) < \infty$$

هم‌چنین آن را با نماد $X_n \xrightarrow{c.c.} X$ نشان می‌دهیم.

[1] Compelet Convergence

تعریف ۲.۱.۱. دنباله‌ای از متغیرهای تصادفی $\{X_n, n \geq 1\}$ به متغیر تصادفی X را همگرای قریب به یقین[۲] گویند، هرگاه

$$P(\{\omega : X_n(\omega) \longrightarrow X(\omega); n \longrightarrow \infty\}) = 1$$

یا به‌طور معادل

$$P(\{\omega : X_n(\omega) \nrightarrow X(\omega); n \longrightarrow \infty\}) = 0$$

و آن را با نماد $X_n \xrightarrow{a.s.} X$ نشان می‌دهیم.

قضیه ۱.۱.۱. فرض کنید $\{X_n, n \geqslant 1\}$ دنباله‌ای از متغیرهای تصادفی و به طور کامل به متغیر تصادفی X همگرا باشد، این دنباله به طور قریب به یقین نیز به X همگرا می‌باشد.

قضیه ۲.۱.۱. فرض کنید $\{X_n, n \geqslant 1\}$ دنباله‌ای از متغیرهای تصادفی و به طور قریب به یقین به ثابتی مثل C همگرا باشد، این دنباله به طور کامل نیز به ثابت C همگرا می‌باشد.

لم ۱.۱.۱ (فاتو). فرض کنید $\{X_n, n \geqslant 1\}$ دنباله‌ای از متغیرهای تصادفی نامنفی باشند به شرط وجود امید ریاضی‌ها، آن‌گاه

$$E(\liminf_{n\longrightarrow\infty} X_n) \leq \liminf_{n\longrightarrow\infty} EX_n.$$

تعریف ۳.۱.۱. فرض کنید $\{A_n, n \geq 1\}$ یک دنباله از پیشامدهای دلخواه باشد آن‌گاه

$$P(A_n \quad i.o.) = P(\limsup_{n\longrightarrow\infty} A_n)$$

$i.o.$[۳] یعنی تعداد زیادی از A_n اتفاق بی‌افتد (تقریبا بی‌نهایت بار) که در این‌صورت لم زیر را داریم

لم ۲.۱.۱. لم برل کانتلی[۴] به دو حالت زیر می‌باشد

۱ــ فرض کنید $\{A_n, n \geqslant 1\}$ یک دنباله از پیشامدهای دلخواه باشد، اگر

$$\sum_{n=1}^{+\infty} P(A_n) < \infty$$

آن‌گاه

$$P(A_n; \quad i.o.) = 0.$$

۲ــ فرض کنید $\{A_n, n \geqslant 1\}$ یک دنباله از پیشامدهای مستقل باشد

$$\sum_{n=1}^{+\infty} P(A_n) = +\infty$$

اگر و تنها اگر

$$P(A_n \quad i.o.) = 1.$$

[۲] Almost Surely

[۳] Infinity Often

[۴] Borel-Cantelli Lemmas

۲.۱ نابرابری‌ها

نابرابری‌ها نقش کلیدی در نظریه احتمال دارند. زیرا اگر نتوانیم مقدار دقیق احتمال‌ها و گشتاورها را محاسبه کنیم، با استفاده از نابرابری‌ها می‌توان کران‌هایی برای مقادیر مورد نظر پیدا کرد. یک نابرابری مناسب کلیدی جهت رفع مشکلات محسوب می‌شود. به عبارت دیگر هر چه کران به مقدار واقعی نزدیک‌تر باشد جواب نهایی قابل اعتمادتر است. در زیر برخی از نابرابری‌ها از کتاب گات (۱۹۹۴) را معرفی می‌کنیم.

۱ـ نابرابری مارکف[۵] : فرض کنید X متغیر تصادفی آن‌گاه برای هر عدد حقیقی مثبت r و $x > 0$

$$P(|X| > x) \leqslant \frac{E|X|^r}{x^r}.$$

۲ـ نابرابری هافدینگ[۶] : فرض کنید $X_1, X_2, \dots, X_n$ متغیرهای تصادفی مستقل باشند طوری‌که برای $k = 1, 2, \dots, n$، $P(a_k \leq X_k \leq b_k) = 1$. همچنین فرض کنید $S_n = \sum_{k=1}^{n} X_k$، $n \geq 1$، مجموعه‌های جزئی باشد. آن‌گاه برای $x > 0$

$$P(S_n - ES_n > x) \leq \exp\left\{\frac{-2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right\}.$$

۴ـ نابرابری گشتاوری[۷] : فرض کنید X و Y متغیرهای تصادفی مستقل باشند طوری‌که $E|X|^r < \infty$ و $E|Y|^r < \infty$، و r یک عدد حقیقی مثبت باشد، آن‌گاه

$$E|X + Y|^r \leqslant 2^r(E|X|^r + E|Y|^r).$$

۵ـ نابرابری هولدر[۸] : فرض کنید برای هر عدد حقیقی $q > 1$ و $p > 1$، $p^{-1} + q^{-1} = 1$ اگر $E|X|^p < \infty$ و $E|Y|^q < \infty$ آن‌گاه

$$E|XY| \leq (E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}}.$$

۶ـ نابرابری کوشی شوارتز[۹] : در نابرابری هولدر اگر $p = q = 2$ قرار دهیم آن‌گاه نابرابری کوشی شوارتز به‌صورت زیر به‌دست می‌آید:

$$E|XY| \leq (E|X|^2)^{\frac{1}{2}}(E|Y|^2)^{\frac{1}{2}}.$$

[۵] Markov's Inequality
[۶] Hoffiding's Inequality
[۷] R-mean's Inequality
[۸] Holder's Inequality
[۹] Cauchy-Shwats Inequality

۷ـ نابرابری مینکوفسکی[۱۰] : فرض کنید X و Y متغیرهای تصادفی باشد و برای هر عدد حقیقی $p > ۱$، $E|X|^p < \infty$ و $E|Y|^p < \infty$ باشد. آن‌گاه

$$(E|X+Y|^p)^{\frac{1}{p}} \leq (E|X|^p)^{\frac{1}{p}} + (E|Y|^p)^{\frac{1}{p}}.$$

۳.۱ انواع وابستگی

دراین بخش به معرفی انواع وابستگی‌ها که رده‌ی بزرگی از دنباله متغیرهای تصادفی را دربر می‌گیرند می‌پردازیم.

۱.۳.۱ وابستگی زبرجمعی منفی

متغیرهای تصادفی وابسته زبرجمعی منفی[۱۱] (NSD) نوعی از وابستگی‌ها است که توسط هو (۲۰۰۰) مطرح شد که در پایه آن از توابع زبرجمعی استفاده شده است

تعریف ۱.۳.۱. (کمپرمن، ۱۹۷۷) تابع $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ زبرجمعی نامیده می‌شود اگر برای هر $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\phi(\mathbf{x} \vee \mathbf{y}) + \phi(\mathbf{x} \wedge \mathbf{y}) \geqslant \phi(\mathbf{x}) + \phi(\mathbf{y}).$$

در رابطه بالا $\vee$ نشان دهنده‌ی ماکسیمم مولفه‌ای[۱۲] و $\wedge$ نشان دهنده‌ی مینیمم مولفه‌ای[۱۳] است، که به صورت زیر تعریف می‌شود.

$$\mathbf{x} \vee \mathbf{y} = (x_۱ \vee y_۱, x_۲ \vee y_۲, \ldots, x_n \vee y_n)$$

و

$$\mathbf{x} \wedge \mathbf{y} = (x_۱ \wedge y_۱, x_۲ \wedge y_۲, \ldots, x_n \wedge y_n).$$

کمپرمن (۱۹۷۷) روشی را بیان کرد که می‌توان با استفاده از آن زبرجمعی بودن یا نبودن یک تابع را مشخص کرد.

لم ۱.۳.۱. فرض کنید تابع ϕ دارای مشتق جزئی مرتبه دوم پیوسته باشد، آن‌گاه ϕ زبرجمعی است هرگاه

$$\frac{\partial^۲ \phi(\mathbf{x})}{\partial x_i \partial x_j} \geq ۰ \qquad ۱ \leqslant i \neq j \leq n. \tag{۱.۱}$$

[۱۰] Mincowski Inequality

[۱۱] Negatively Superadditive Dependent

[۱۲] Maximum Componentwise

[۱۳] Minimum Componentwise

تعریف ۲.۳.۱. بردار تصادفی $\mathbf{X}=(X_۱,X_۲\ldots,X_n)$ را NSD گویند، هرگاه برای هر تابع زبرجمعی ϕ

$$E\phi(X_۱,X_۲,\ldots,X_n)\leq E\phi(Y_۱,Y_۲,\ldots,Y_n).$$

که Y_iها متغیرهای تصادفی مستقل هستند، همچنین به ازای هر $i=۱,۲,\ldots,n$، X_i و Y_i هم‌توزیع هستند.

تعریف ۳.۳.۱. $\{X_n,n\geqslant ۱\}$ را دنباله‌ای از متغیرهای تصادفی NSD گویند هرگاه برای هر $n\geq ۱$، $(X_۱,X_۲,\ldots,X_n)$، NSD باشد.

لم ۲.۳.۱. فرض کنید $(X_۱,X_۲,\ldots,X_n)$، NSD و توابع $f_۱,f_۲,\ldots,f_n$ نانزولی باشند آن‌گاه $f_۱(X_۱),f_۲(X_۲),\ldots,f_n(X_n)$، NSD می‌باشند.

برهان. بنابر معادله (۱.۱) برای هر تابع زبرجمعی ϕ، داریم $\phi(f_۱(x_۱),f_۲(x_۲),\ldots,f_n(x_n))$ نیز زبرجمعی است، چون

$$\begin{aligned}&\frac{\partial^۲\phi(f_۱(x_۱),f_۲(x_۲),\ldots,f_n(x_n))}{\partial x_i\partial x_j}\\&=\frac{\partial^۲\phi(f_۱(x_۱),f_۲(x_۲),\ldots,f_n(x_n))}{\partial f_i(x_i)\partial f_j(x_j)}\cdot\frac{\partial f_i(x_i)}{\partial x_i}\cdot\frac{\partial f_j(x_j)}{\partial x_j}\geq ۰.\end{aligned}$$

□

ملاحظه ۱.۳.۱. روابط بالا برای $f_۱,f_۲,\ldots,f_n$ ناصعودی باشند نیز برقرار است.

۲.۳.۱ معرفی برخی دیگر از وابستگی‌ها

تعریف ۴.۳.۱. بردار تصادفی $\mathbf{X}=(X_۱,X_۲,\ldots,X_n)$ را پیوند منفی[۱۴] (NA) گویند اگر برای هر جفت زیرمجموعه‌های جدا از هم مانند A و B از $\{۱,۲,\ldots,n\}$ داشته باشیم

$$Cov\left(f_۱\left(\mathbf{X}_i,i\in A\right),f_۲\left(\mathbf{X}_j,j\in B\right)\right)\leq ۰. \qquad (۲.۱)$$

که $f_۱$ و $f_۲$ توابع صعودی هستند که کواریانس آن‌ها موجود می‌باشد. یکی از مزیت‌های متغیرهای تصادفی NA این است که $f_۱$ و $f_۲$ در معادله (۲.۱) ممکن است شکل‌های مختلفی را داشته باشند از این‌رو به نامساوی‌های چند متغیره مفید زیادی منجر می‌شوند.

تعریف ۵.۳.۱. متغیرهای تصادفی X و Y را وابسته منفی[۱۵] (ND) گویند هرگاه برای هر x و y عضو $\mathbb{R}$

$$P\left(X\leq x,Y\leq y\right)\leq P\left(X\leq x\right)P\left(Y\leq y\right). \qquad (۳.۱)$$

[۱۴] Negatively Associate

[۱۵] Negatively Dependent

تعریف ۶.۳.۱. برای هر $x, y \in \mathbb{R}$ مجموعه A دو به دو وابسته منفی[۱۶] (PND) گویند اگر برای هر جفت متغیر تصادفی در رابطه (۳.۱) برقرار است. با توجه به آن‌چه در رابطه (۳.۱) اشاره شده آن‌گاه

$$P\left(X > x, Y > y\right) \leq P\left(X > x\right) P\left(Y > y\right). \qquad (۴.۱)$$

برای هر جفت $x, y \in \mathbb{R}$، (۴.۱) به معنی (۳.۱) است از این‌رو (۴.۱) و (۳.۱) معادل هستند. ابراهیمی و قوش (۱۹۸۱) نشان دادند (۴.۱) و (۳.۱) برای مجموعه‌ای سه یا بیشتر از سه متغیر تصادفی معادل نیستند.
مثال‌هایی که در زیر آورده شده از مقاله هو (۲۰۰۰) و وو و جی یانگ (۲۰۱۱) می‌باشد.

مثال ۱.۳.۱. متغیرهای تصادفی X_1، X_2 و X_3 را در نظر بگیرید که مقادیر $(1,1,0)$، $(1,0,1)$، $(0,1,1)$ و $(0,0,0)$ هر کدام احتمال $\frac{1}{4}$ دارند، برای هر جفت متغیرهای تصادفی X_1، X_2، X_3 و دوبه دو مستقل و از این‌رو (۳.۱) و (۴.۱) برقرار است، بنابراین برای هر x_1, x_2, x_3

$$P\left(X_1 > x_1, X_2 > x_2, X_3 > x_3\right) \leq P\left(X_1 > x_1\right) P\left(X_2 > x_2\right) P\left(X_3 > x_3\right)$$

$$P\left(X_1 \leq 0, X_2 \leq 0, X_3 \leq 0\right) = P\left(X_1 = 0, X_2 = 0, X_3 = 0\right) = P(0,0,0) = \frac{1}{4}$$

$$P(X_1 \leq 0) = P(X_1 = 0) = P(0,1,1) + P(0,0,0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X_2 \leq 0) = P(X_2 = 0) = P(1,0,1) + P(0,0,0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X_3 \leq 0) = P(X_3 = 0) = P(1,1,0) + P(0,0,0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

آن‌گاه

$$P\left(X_1 \leq 0, X_2 \leq 0, X_3 \leq 0\right) = \frac{1}{4} \nleq \frac{1}{8} = P\left(X_1 \leq 0\right) P\left(X_2 \leq 0\right) P\left(X_3 \leq 0\right).$$

مثال ۲.۳.۱. فرض کنید $\mathbf{X} = (X_1, X_2, X_3, X_4)$ یک بردار تصادفی از متغیرهای تصادفی گسسته با تابع احتمال زیر باشد فرض کنید

		$(X_1،X_2)$				
		(۰،۰)	(۰،۱)	(۱،۰)	(۱،۱)	تابع احتمال حاشیه‌ای
(X_3, X_4)	(۰،۰)	۰/۰۵۷۷	۰/۰۶۲۳	۰/۰۶۲۳	۰/۰۵۷۷	۰/۲۴
	(۰،۱)	۰/۰۶۲۳	۰/۰۶۷۷	۰/۰۶۷۷	۰/۰۶۲۳	۰/۲۶
	(۱،۰)	۰/۰۶۲۳	۰/۰۶۷۷	۰/۰۶۷۷	۰/۰۶۲۳	۰/۲۶
	(۱،۱)	۰/۰۵۷۷	۰/۰۶۲۳	۰/۰۶۲۳	۰/۰۵۷۷	۰/۲۴
	تابع احتمال حاشیه‌ای	۰/۲۴	۰/۲۶	۰/۲۶	۰/۲۴	

[۱۶] Pairwise Negatively Dependent

$$(X_۱, X_۲, X_۳, X_۴) \overset{\mathrm{d}}{=\!=} (X_۲, X_۱, X_۳, X_۴) \overset{\mathrm{d}}{=\!=} (X_۲, X_۱, X_۴, X_۳) \overset{\mathrm{d}}{=\!=} (X_۳, X_۴, X_۱, X_۲)$$

بدون این‌که از کلیت مساله کم شود، تابعی زبرجمعی مانند ϕ را در نظر می‌گیریم طوری که به ازای هر $\mathbf{x} = (x_۱, x_۲, x_۳, x_۴)$

$$\phi\,(x_۱, x_۲, x_۳, x_۴) = \phi\,(x_۲, x_۱, x_۳, x_۴) = \phi\,(x_۲, x_۱, x_۴, x_۳) = \phi\,(x_۱, x_۲, x_۴, x_۳)$$

در واقع تابع زبرجمعی X نسبت به جایگشت متغیرهای تصادفی ثابت هستند با استناد به تعریف تابع زبرجمعی که در آن

$$\phi\,(\mathbf{x}) + \phi\,(\mathbf{y}) \le \phi\,(\mathbf{x} \wedge \mathbf{y}) + \phi\,(\mathbf{x} \vee \mathbf{y})$$

در نتیجه

$$۲\phi_{۰۰۰۱} \le \phi_{۰۰۰۰} + \phi_{۰۰۱۱}, \quad ۲\phi_{۰۱۱۱} \le \phi_{۰۰۱۱} + \phi_{۱۱۱۱}, \quad ۲\phi_{۰۱۰۱} \le \phi_{۰۰۰۱} + \phi_{۰۱۱۱}$$

که در آن $\phi_{ijkl} = \phi(i, j, k, l)$ به عنوان نمونه دو تابع $\phi_{۰۰۰۱}$ و $\phi_{۰۰۱۰}$ را در نظر بگیرید، برای حالت می‌نیمم مولفه‌ای در اولین مولفه به دلیل این‌که هر دو صفر هستند، بنابراین می‌نیمم صفر خواهد بود به همین ترتیب برای مولفه‌های بعدی هم‌چنین حالت ماکسیمم مولفه‌ای را به‌دست می‌آوریم بنابراین با استدلالی مشابه و با استفاده از روابط بالا داریم

$$۲۶\phi_{۰۱۰۱} \le ۶\phi_{۰۰۰۰} + ۱۲\phi_{۰۰۱۱} + ۶\phi_{۱۱۱۱} + \phi_{۰۰۰۱} + \phi_{۰۱۱۱}$$

حال با توجه به اعداد جدول نابرابری فوق درست می‌باشد. در این صورت X، NSD می‌باشد از طرفی با توجه به وابستگی پیوند منفی

$$P\left(X_i = ۱; i = ۱, ۲, ۳, ۴\right) < P\,(X_۱ = X_۲ = ۱)\,P\,(X_۳ = X_۴ = ۱)$$

با توجه به این‌که

$$cov(f(\mathbf{X}_A), g(\mathbf{X}_B) \le ۰, \quad f(\mathbf{X}_A) = I\{X_۱ = X_۲ = ۱\}, g(\mathbf{X}_B) = I\{X_۳ = X_۴ = ۱\}$$

بنابراین

$$\begin{aligned} E(f(\mathbf{X}_A)g(\mathbf{X}_B)) - E(f\mathbf{X}_A)Eg(\mathbf{X}_B) &= P(X_۱ = X_۲ = X_۳ = X_۴ = ۱) \\ &- P(X_۱ = X_۲ = ۱)P(X_۳ = X_۴ = ۱) \qquad (۵.۱) \\ &= ۰/۰۵۷۷ - ۰/۰۵۷۶ \nleq ۰. \end{aligned}$$

این در حالی است با جایگذاری از اعداد جدول ملاحظه می‌کنیم که نامساوی فوق برقرار نیست که نتیجه می‌گیریم X، NA نخواهد بود که از این مثال به این نتیجه می‌رسیم که زبرجمعی منفی تعمیمی از وابستگی پیوندی منفی می‌باشد که این نتیجه مهم می‌تواند در برخی از نابرابری‌های احتمال و نابرابری گشتاوری نقش بسزایی را ایفا کند.

لم ۳.۳.۱. اگر $(X_1, X_2, \ldots, X_n)$، NSD باشد برای هر عدد حقیقی $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ آن‌گاه

$$P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \leq P(X_1 \leq x_1)P(X_2 \leq x_2)\ldots P(X_n \leq x_n)$$
$$= \prod_{i=1}^{n} P(X_i \leq x_i).$$

برهان. برای هر $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ با توجه به نزولی بودن تابع $I\{X_i \leq x_i\}$ نسبت به x_i و زبرجمعی بودن تابع $\phi(\mathbf{x}) = \prod_{i=1}^{n} I(X_i \leq x_i)$ و NSD بودن X_iها داریم

$$\begin{aligned}
P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) &= E(I\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\}) \\
&= E\left(\prod_{i=1}^{n} I\{X_i \leq x_i\}\right) \\
&\leq E\left(\prod_{i=1}^{n} I\{Y_i \leq x_i\}\right) \qquad (۶.۱) \\
&= \prod_{i=1}^{n} E(I\{Y_i \leq x_i\}) \\
&= \prod_{i=1}^{n} P(X_i \leq x_i)
\end{aligned}$$

که در رابطه فوق Y_iها همان شرایط (NSD) را دارند. □

لم ۴.۳.۱. اگر برای $(X_1, X_2, \ldots, X_n)$، NSD باشد برای هر عدد حقیقی $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ آن‌گاه

$$P(X_1 \geq x_1, X_2 \geq x_2, \ldots, X_n \geq x_n) \leq P(X_1 \geq x_1)P(X_2 \geq x_2)\ldots P(X_n \geq x_n)$$
$$= \prod_{i=1}^{n} P(X_i \geq x_i).$$

برهان. با استدلال مشابه برهان لم قبل داریم:

$$\begin{aligned}
P(X_1 \geq x_1, X_2 \geq x_2, \ldots, X_n \geq x_n) &= E(I\{X_1 \geq x_1, X_2 \geq x_2, \ldots, X_n \geq x_n\} \\
&= E\left(\prod_{i=1}^{n} I\{X_i \geq x_i\}\right) \\
&\leq E\left(\prod_{i=1}^{n} I\{Y_i \geq x_i\}\right) \qquad (۷.۱) \\
&= \prod_{i=1}^{n} E(I\{Y_i \geq x_i\}) \\
&= \prod_{i=1}^{n} P(X_i \geq x_i).
\end{aligned}$$

که در رابطه فوق Y_iها همان شرایط (NSD) را دارند. □

۴.۱ رگرسیون

رگرسیون[۱۷] در لغت به معنی برگشت است. نخستین بار گالتن این واژه را به کار برد و با آن ارتباط بین میانگین طول قامت پدران و پسران را مورد مطالعه قرار داد. گالتن از روی مشاهداتش نتیجه گرفت که طول قامت پسران متعلق به والدین خیلی بلند قد(یا کوتاه)، روی همرفته بلندتر (یا کوتاهتر) از طول قامت متوسط بوده است، ولی البته نه به بلندی (یا کوتاهی) قد والدینشان این نتیجه گیری در سال ۱۸۵۵ به عنوان برگشت قد ارثی حد متوسط منتشر شد.

۱.۴.۱ رگرسیون خطی

مدل ارتباطی (ضابطه ریاضی) حاکم بر متغیر پاسخ و مجموعه متغیرهای تبیینی، توسط افراد متخصص در زمینه مورد مطالعه، بر مبنای دانش یا قضاوت‌های عینی و ذهنی آن‌ها تعیین می‌شود. این مدل که می‌تواند پارامترهای نامعلوم زیادی را در بر داشته باشد، یک مدل پارامتری نامیده می‌شود. اگر در معادله رگرسیونی تنها یک متغیر تبیینی وجود داشته باشد، آن را مدل رگرسیونی خطی ساده می‌نامند. در یک مدل رگرسیونی، هدف بررسی اثر متغیرهای تبیینی بر روی متغیر پاسخ است. معمولا متغیرهای تبیینی را با X (که می‌توان بعدی بزرگتر از یک داشته باشد) و متغیر پاسخ را با Y نمایش می‌دهند. هم‌چنین این ضابطه می‌تواند خطی یا غیرخطی باشد. باید توجه داشته باشیم که منظور از رابطه خطی یا غیرخطی، در مورد رابطه بین Y و X_i ها نیست بلکه در واقع پارامترهای رگرسیونی به طور خطی یا غیرخطی وارد معادله رگرسیون می‌شوند. به عنوان مثال هریک از روابط زیر خطی‌اند هر چند که ممکن است رابطه بین X و Y ممکن است خطی نباشد:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

$$Y = \beta_0 + \beta_1 e^X + \varepsilon$$

$$Y = \beta_0 + \beta_1 \log X + \varepsilon$$

هم‌چنین ε جمله خطا می‌باشد که متغیری تصادفی و غیر قابل مشاهده است. مثال‌هایی از یک مدل رگرسیونی عبارتند از

- بررسی اثر تعداد قطعات لازم در ساخت یک دستگاه (متغیر تبیینی) بر مدت زمان لازم بر تولید آن (متغیر پاسخ)
- بررسی تاثیر سرانه خالص ملی (متغیر تبیینی) بر تولید خالص ملی (متغیر پاسخ)
- بررسی تاثیر سن (متغیر تبیینی) بر تعداد ضربان قلب (متغیر پاسخ)

[۱۷] Regression

در مدل رگرسیونی خطی، متغیر پاسخ Y یک متغیر تصادفی پیوسته است، در حالی که X تصادفی نبوده و توسط تحلیل‌گر کنترل و با خطای قابل اغماضی قابل اندازه‌گیری می‌شود. بنابراین به ازای هر مقدار ممکن X برای Y یک توزیع احتمال وجود دارد. به‌طور کلی، منظور از مدل رگرسیونی این است که $E\left(Y|x\right)$ را بر حسب تابعی از X، یا Y را بر حسب تابعی از X، به همراه یک جمله خطا ε ، که امید ریاضی آن صفر فرض می‌شود، بنابراین

$$Y=\beta_0+\beta_1X+\varepsilon, \qquad E\left(Y|x\right)=\beta_0+\beta_1X$$

که هر دو معادل هستند.
هر مدل آماری مبتنی بر پذیره‌هایی ساخته می‌شود. برای مدل رگرسیونی خطی نیز چهار پذیره در نظر گرفته می‌شود که به‌صورت زیر می‌باشند:

- خطی بودن ضابطه تابع رگرسیونی $E\left(Y|x\right)$؛
- ثابت بودن واریانس جمله خطا؛
- ناهمبسته بودن مولفه‌های خطا؛
- (بر حسب نیاز) نرمال بودن خطاها.

۲.۴.۱ رگرسیون چندگانه

در حالت کلی ممکن است متغیر پاسخ به k متغیر رگرسیونی $x_1, x_2, \ldots, x_k$ مربوط باشد، بنابراین مدل رگرسیونی به‌صورت زیر است

$$Y=\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_kx_k+\varepsilon, \quad i=1,2,\ldots,n \tag{8.1}$$

را می‌توان ارائه کرد که در آن εها متغیرهای تصادفی کنترل‌ناپذیر می‌باشند و متغیرهای خطا نامیده می‌شود و فرض می‌شود دارای توزیع نرمال با میانگین صفر و واریانس σ^2 و ناهمبسته هستند. پارامترهای σ^2 و $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ مجهول هستند و باید با استفاده از داده‌ها برآورد شوند. فرض می‌شود داده‌ها عبارتند از $Y_1, Y_2, \ldots, Y_k$ که در آن Y_iها پاسخ متناظر با k سطح از متغیرهای مستقل $x_{i1}, x_{i2}, \ldots, x_{ik}$ است، یعنی بنابر معادله‌ی (۸.۱) برای مقادیر $i=1,2,\ldots,n$ می‌توان نوشت

$$Y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_kx_{ik}+\varepsilon_i. \tag{9.1}$$

که هدف ما به دست آوردن برآوردهایی برای $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ به ترتیب $b_0, b_1, \ldots, b_k$ است در نتیجه به دست آوردن معادله‌ی

$$\hat{Y}=b_0+b_1x_1+b_2x_2+\ldots+b_kx_k \tag{10.1}$$

برای برآورد رگرسیون، که در آن $\hat{Y}$ نشان دهنده‌ی مقدار برآورد در ازای $x_1, x_2, \ldots, x_k$ است. آن‌گاه معادله (۱.۱۰) به عنوان معادله‌ی پیش‌بینی کننده مورد استفاده قرار می‌گیرد. که روش کمترین توان‌های دوم روشی بسیار مفید برای برآورد پارامترهای رگرسیون خطی و خطی چندگانه می‌باشد.

برآوردهای $b_0, b_1, \ldots, b_k$ را به گونه‌ای انتخاب می‌کنیم که مجموع توان‌های دوم انحراف‌ها را کمینه کند یعنی آن‌ها را به گونه‌ای به دست می‌آوریم که در معادله زیر هنگامی که به ترتیب جایگزین $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ می‌شوند، کمترین مقدار ممکن را تولید کنند

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_k x_{ik})^2 \tag{۱.۱۱}$$

با استفاده از حساب دیفرانسیل و انتگرال، برآوردهای $b_0, b_1, \ldots, b_k$ با مشتق گرفتن از معادله‌ی (۱.۱۱) نسبت $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ و مساوی صفر قرار دادن آن‌ها، به دست می‌آیند. یعنی

$$\begin{aligned}
\frac{\partial S}{\partial \beta_0} &= -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_k x_{ik}) = 0 \\
\frac{\partial S}{\partial \beta_1} &= -2 \sum_{i=1}^{n} x_{i1} (Y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_k x_{ik}) = 0 \\
&\vdots \\
\frac{\partial S}{\partial \beta_k} &= -2 \sum_{i=1}^{n} x_{ik} (Y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_k x_{ik}) = 0
\end{aligned}$$

با جایگزین کردن $b_0, b_1, \ldots, b_k$ به ترتیب به جای $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ در معادله‌های بالا و مساوی صفر قرار دادن آن‌ها پس از مرتب کردن داریم

$$\begin{aligned}
\sum_{i=1}^{n} Y_i &= n b_0 + b_1 \sum_{i=1}^{n} x_{i1} + b_2 \sum_{i=1}^{n} x_{i2} + \ldots + b_k \sum_{i=1}^{n} x_{ik} \\
\sum_{i=1}^{n} x_{i1} Y_i &= b_0 \sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} x_{i1}^2 + b_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \ldots + b_k \sum_{i=1}^{n} x_{i1} x_{ik} \\
&\vdots \\
\sum_{i=1}^{n} x_{ik} Y_i &= b_0 \sum_{i=1}^{n} x_{ik} + b_1 \sum_{i=1}^{n} x_{i1} x_{ik} + b_2 \sum_{i=1}^{n} x_{i2} x_{ik} + \ldots + b_k \sum_{i=1}^{n} x_{ik}^2
\end{aligned}$$

ملاحظه می‌شود که برای حل این معادله‌های نرمال بهتر است که از روش ماتریسی استفاده شود که به‌صورت زیر تعریف می‌کنیم

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

هم‌چنین داریم

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$$

$\mathbf{Y}$ یک بردار $n \times 1$، $\mathbf{X}$ یک ماتریس $n \times (k+1)$، $\boldsymbol{\beta}$ یک بردار $(k+1) \times 1$ و $\boldsymbol{\varepsilon}$ یک بردار $n \times 1$ هستند. به این ترتیب ملاحظه می‌شود که مدل رگرسیون خطی چندگانه (۹.۱) را می‌توان به صورت زیر نوشت

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{۱۲.۱}$$

هم‌چنین معادله‌ی نرمال را می‌توان به صورت زیر نوشت

$$\left(\mathbf{X}'\mathbf{X}\right)\mathbf{b} = \mathbf{X}'\mathbf{Y} \tag{۱۳.۱}$$

زیرا

$$\begin{aligned} S = \sum_{i=1}^{n} \varepsilon_i^2 &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\ &= \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right) \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

چون $\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y}$ یک ماتریس 1×1 است و در نتیجه با ترانهاده‌ی خود برابر است، پس $\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{X}\boldsymbol{\beta}$ و خواهیم داشت

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \tag{۱۴.۱}$$

با مشتق گرفتن از (۱۴.۱) نسبت به بردار $\boldsymbol{\beta}$ و با جایگزین کردن $\mathbf{b}$ به جای $\boldsymbol{\beta}$ و مساوی صفر قرار دادن آن، معادله‌های نرمال (۱۳.۱) به دست می‌آیند. ماتریس‌های $\mathbf{X}'\mathbf{X}$ و $\mathbf{X}'\mathbf{Y}$ عبارتند از

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{i1}x_{ik} & \sum_{i=1}^{n} x_{i2}x_{ik} & \cdots & \sum_{i=1}^{n} x_{ik}^2 \end{pmatrix}, \quad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} x_{i1}Y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik}Y_i \end{pmatrix}$$

با فرض معکوس پذیر بودن ماتریس $\mathbf{X}'\mathbf{X}$ داریم

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}.$$

۵.۱ نقاط پرت

تعریف ۱.۵.۱. داده‌های پرت یا به عبارتی نقاط پرت به نقاطی گفته می‌شود که فاصله‌ی زیادی از سایر نقاط داشته باشند، که این گونه داده‌ها به دلایل مختلفی از جمله اشتباه در جمع‌آوری داده‌ها، ابزار اندازه‌گیری نامناسب، وجود افراد کم‌تجربه برای نمونه‌گیری و ... به وجود می‌آیند و در رگرسیون تعاریف مختلفی را دارند که در زیر به مواردی از آن‌ها اشاره می‌کنیم:

- نقاط دور افتاده‌ی رگرسیونی: نقاطی هستند که از برازش خطی $n - ۱$ نقطه‌ی دیگر فاصله‌ی زیادی را داشته باشند.
- نقاط دور افتاده‌ی باقیمانده‌ای: نقاطی هستند که قدرمطلق باقیمانده‌ای استیودنت شده‌ای بزرگی را داشته باشند یا به عبارتی استاندارد شده‌ی بزرگی را داشته باشند.
- نقاط دور افتاده از فضای (x, y): نقاطی هستند که مختصات x و y آن‌ها از دیگر نقاط فاصله‌ی زیادی را داشته باشند.

با توجه به این‌که داده‌های پرت در تمام مراحل مربوط به آنالیز و تفسیر اطلاعات چه از لحاظ ساختاری و چه از لحاظ مفهومی تاثیرگذار هستند و بعضی موارد مکان نتیجه‌گیری منطقی از اطلاعات جمع‌آوری شده وجود ندارد و دچار خطاهای علمی آماری از لحاظ پایایی و روایی می‌شویم، بنابراین شناسایی و تشخیص داده‌های پرت و اقدام در جهت اصلاح آن‌ها امری ضروری در تمامی تحقیقات می‌باشد، که آماردانان روش‌های مختلفی را برای شناسایی این نقاط پرت ارائه دادند. از جمله این روش‌ها می‌توان به موارد زیر اشاره کرد:

- استفاده از باقیمانده‌های حذف شده استیودنت شده: باقیمانده‌های حذف شده استیودنت شده به باقیمانده‌هایی گفته می‌شود که در مدل رگرسیونی که در یک مجموعه داده‌ی iام پرت به نظر می‌رسد با حذف مورد iام باقیمانده‌ها محاسبه می‌شود، بنابراین برای پیش بینی مقدار پیش بینی مقدار مشاهده به صورت زیر استفاده می‌کنیم

$$\hat{Y}_{(-i)}\left(x_{i۱}, x_{i۲}, \ldots, x_{in}\right) = \hat{\beta}_{۰(-i)} + \hat{\beta}_{(-i)} x_{i۱} + \ldots + \hat{\beta}_{(-i)} x_{ik}$$

اگر تفاوت این مقدار برازش شده را از مقدار واقعی آن به‌دست آوریم و بر خطای استاندارد آن تقسیم کنیم باقیمانده‌های حذف شده‌ی استیودنت شده به‌دست می‌آید که به صورت زیر می‌باشد

$$T = \frac{Y_i - \hat{Y}_{(-i)}\left(x_{i۱}, x_{i۲}, \ldots, x_{in}\right)}{\frac{\hat{\sigma}_{(-i)}}{\sqrt{۱-h_{ii}}}}$$

h_{ii}، همان عامل iام در ماتریس $\mathbf{H} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-۱}\mathbf{X}'$ که تمام مشاهدات $\mathbf{X}$ را مشاهده می‌کنیم. T دارای توزیع t- استیودنت با $(n - ۱) - k - ۱ = n - k - ۲$ درجه آزادی می‌باشد، به همین دلیل به باقیمانده‌های حذف شده استیودنت شده یا t- باقیمانده

نیز می‌گویند که معمولا اگر T بزرگتر از ۲ یا ۳ باشد نشان دهنده‌ی آن است که داده‌ی پرت وجود دارد و مورد نیاز است که علت را بررسی کنیم.

- استفاده از نمودار جعبه‌ای: برای تشخیص داده‌ی پرت با استفاده از نمودار جعبه‌ای برای هر مجموعه داده‌ای که مشاهدات خارج از دامنه نمودار قرار گیرند جزء داده‌های پرت به شمار می‌آیند.

- نمودار پراکنش: در نمودار پراکنش برای شناسایی نقاط پرت داده‌هایی را که خارج از محدوده‌ی سایر داده‌ها قرار بگیرد را جزء داده‌ی پرت در نظر می‌گیریم

- فاصله کوک[۱۸]: برای تشخیص داده‌ی پرت با استفاده از فاصله کوک از اندازه مربع فاصله بین $\hat{\beta}$ برآورد حداقل مربعات روی تمام نقاط و $\hat{\beta}_{(-i)}$ برآورد به‌دست آمده براساس حذف iامین نقطه را پیشنهاده کرده است که این اندازه‌ی فاصله‌ای به صورت زیر می‌باشد

$$D_i\left(M,c\right)=\frac{\left(\hat{\beta}_{(-i)}-\hat{\beta}\right)'M\left(\hat{\beta}_{(-i)}-\hat{\beta}\right)}{c},\quad i=۱,۲,\ldots,n \tag{۱۵.۱}$$

که در (۱۵.۱)، M و c به ترتیب برابر $M=\mathbf{X'X}$ و $c=pMSE$ تعریف می‌شوند که معمولا نقاطی را که $D_i>۱$ باشد به عنوان نقطه‌ی پرت شناسایی می‌شوند.

در پایان این قسمت باید اشاره کنیم که روش‌های زیادی برای شناسایی نقاط پرت وجود دارد که ما در بالا فقط به چند مورد اشاره کردیم. از جمله روش‌های دیگری را که می‌توان نام برد استفاده از نمودار هیستوگرام، استفاده از مجموع مربعات موزون فاصله iامین نقطه از مرکز داده‌ها و ... هستند.

۶.۱ ترتیب‌های غیرتصادفی

ترتیب‌های غیرتصادفی را با o و O را نشان می‌دهند که نمادهایی برای توصیف مرتبه‌ی مجانبی کمیت‌های غیرتصادفی می‌باشند. ون در وارت (۱۹۹۸) تعاریفی از ترتیب‌های غیرتصادفی با توجه به ترتیب همگرایی در احتمال به همراه نمادهای مناسب ارائه داد که به شرح زیر می‌باشد.

تعریف ۱.۶.۱. دو دنباله از اعداد ثابت $\{a_n;n\geq ۱\}$ و $\{b_n;n\geq ۱\}$ را در نظر بگیرید. گوییم $b_n=O\left(a_n\right)$، هرگاه برای هر $\varepsilon>۰$، $K\left(\varepsilon\right)>۰$ و عدد صحیح $N\left(\varepsilon\right)$ وجود داشته باشند به طوری‌که برای هر $n\geq N\left(\varepsilon\right)$، آن‌گاه

$$\left|b_n\right|<K\left(\varepsilon\right)\left|a_n\right|.$$

[۱۸] Cook Interval

تعریف ۲.۶.۱. دو دنباله از اعداد ثابت $\{a_n; n \geq ۱\}$ و $\{b_n; n \geq ۱\}$ را در نظر بگیرید. گوییم $b_n = o\left(a_n\right)$، هرگاه

$$\lim_{n \longrightarrow \infty} \left| \frac{b_n}{a_n} \right| = ۰.$$

با استفاده از تعاریفی که در بالا به آن اشاره شد، برای هر ثابت حقیقی مثل C، ترتیب‌های $O\left(a_n\right)$، و $o\left(a_n\right)$، که به ترتیب معادل با $Ca_nO\left(۱\right)$ ، $Ca_no\left(۱\right)$ هم‌چنین اگر $X_n = O\left(n^C\right)$، آن‌گاه $X_n = O\left(n^{C+۱}\right)$ قرار می‌گیرد اما عکس رابطه برقرار نیست. برخی از ویژگی‌های ترتیب‌های غیرتصادفی در جدول زیر آورده شده است که از پایان نامه کارشناسی ارشد اسحاقی (۱۳۹۲) استفاده شده که در این جدول a و b اعداد حقیقی و $k = \max\left\{a, b\right\}$ فرض می‌کنیم.

جدول ۱.۱: تعدادی از ترتیب‌های غیرتصادفی

	عملگرد ضرب	عملگرد جمع	عملگر ترکیب
۱	$o\left(n^a\right) o\left(n^b\right) = o\left(n^{a+b}\right)$	$o\left(n^a\right) + o\left(n^b\right) = o\left(n^k\right)$	$o\left(O\left(n^a\right)\right) = o\left(n^a\right)$
۲	$O\left(n^a\right) O\left(n^b\right) = O\left(n^{a+b}\right)$	$O\left(n^a\right) + O\left(n^b\right) = O\left(n^k\right)$	$O\left(o\left(n^a\right)\right) = O\left(n^a\right)$
۳	$O\left(n^a\right) o\left(n^b\right) = o\left(n^{a+b}\right)$	$O\left(n^a\right) + o\left(n^b\right) = O\left(n^k\right)$	$O\left(O\left(n^a\right)\right) = O\left(n^a\right)$

فصل ۲
روش‌های آماری استوار

در این فصل به شرح استواری، روش‌های آماری استوار از جمله برآورد پارامترهای مکانی، مقیاسی، رگرسیون استوار و تاریخچه‌ی آن و مقایسه‌ی آن با کمترین توان‌های دوم درمورد نقاط پرت عمده مطالب این فصل از هوگ (۱۷۹۹)، پایان نامه کارشناسی ارشد اقبال (۱۳۸۴)، کتاب مقدمه‌ای بر تحلیل رگرسیون خطی داگلاس مونتگمری، الیزابت پک ترجمه سید ابراهیم رضوی پاریزی (۱۳۸۶) و کتاب تجزیه و تحلیل رگرسیون کاربردی دراپر (۱۹۹۸) می‌پردازیم.

۱.۲ استواری

روش‌های استوار[۱] در متون آماری، استوار بودن یک روش، به معنی معتبر بودن و مناسب بودن آن است به عبارت دیگر هرگاه ویژگی یک روش آماری تحت تاثیر انحراف کم از پذیره‌های اساسی قرار نگیرند، آن روش آماری را استوار گوییم، برای مثال $\bar{x}$، میانگین نمونه، برآوردگری برای میانگین جامعه است، ولی استوار نیست. یعنی اگر حداقل یکی از اعضای نمونه مقدار خیلی بزرگی را داشته باشد و از جامعه‌ی مورد نظر نیامده باشد (داده‌ی پرت باشد)، $\bar{x}$ بسیار متفاوت از میانگین جامعه است در مقابل میانه‌ی نمونه برآوردگری استوار برای پارامتر مکان (در صورت تقارن توزیع جامعه، برای میانگین جامعه) است. زیرا حتی اگر نقاط پرت به تعداد $\left[\frac{n-۱}{۲}\right]$ نیز باشند، مقدار میانه تغییر نمی‌کند. استواری در مبحث بازه‌ی اطمینان و آزمون

[۱] Robust Method

فرضیه‌ها نیز مطرح می‌شود، برای مثال $\bar{X} \pm t_{n-1}(\frac{\alpha}{2})\frac{S}{\sqrt{n}}$ یک بازه اطمینان $(1-\alpha)$ درصد برای میانگین جامعه‌ی نرمال، استوار است.
این سوال مطرح می‌شود که نرمال نبودن مشاهدات بر دقت و کارایی بازه‌ی اطمینان چقدر تاثیر می‌گذارد؟ یعنی انحراف از پذیره‌ی نرمال بودن جامعه به میزان کم، به چه اندازه بر ضریب اطمینان واقعی بازه و به چه میزان بر متوسط طول بازه تاثیر می‌گذارد؟
جوابی را که می‌توان به این سوال داد این است که تاثیر نرمال نبودن جامعه در صورتی که جامعه متقارن باقی بماند بر ضریب اطمینان واقعی بسیار اندک است ولی اگر جامعه متقارن نباشد، ضریب اطمینان به شدت کاهش می‌یابد. بنابراین می‌توان گفت برآورد فاصله‌ای t ــ استیودنت مادامی که جامعه متقارن باشد، استوار و اگر توزیع از تقارن فاصله بگیرد، نااستوار است. بر این اساس روش‌های استواری برای اصلاح روش کمترین توان‌های دوم به گونه‌ای هستند، که نقاط دور افتاده تاثیر کمتری را بر روی برآوردگر نهایی دارد. از آنجا که یکی از اقسام رگرسیون، رگرسیون استوار می‌باشد (در مورد تجزیه و تحلیل داده‌های پرت)، توسط جورج باکس (۱۹۵۳) برای اولین بار معرفی گردید که مقالات و کتاب‌هایی در این زمینه به چاپ رسیده است که می‌توان از آنها استفاده کرد از جمله این کتاب‌ها جان توکی (۱۹۶۲) است که در مورد مکان و موقعیت داده‌ها توضیح می‌دهد، شاید بیشترین سهم را هوبر داشته باشد که مقاله‌های اساسی او در سال‌های ۱۹۶۴، ۱۹۷۲ و ۱۹۷۳ نقطه‌ی عطف در این زمینه هستند. هم‌چنین فرانک همپل از تاثیر منحنی در برآورد استواری که یک مفهوم مرکزی است صحبت می‌کند، از این رو اگر کسی علاقه‌ای به تاثیر منحنی داشته باشد می‌تواند به مقاله همپل (۱۹۷۴) مراجعه کند.
مدل خطی رگرسیون $\mathbf{Y}_i = \mathbf{X}_i'\beta + e_i$ را در نظر بگیرید. وقتی مشاهدات Y_iها دارای توزیع نرمال هستند برآورد خوبی از β را می‌دهد، اما اگر مشاهدات دارای توزیع نرمال نباشند، نمی‌تواند برآورد خوبی را ارائه دهد. زیرا در انتهای منحنی دارای مشاهداتی هستیم که منحنی نرمال دم پهن‌تر یا به عبارتی ضخیم‌تر از نرمال می‌باشد که روش کمترین توان‌های دوم کارا نیست که این شاخه‌های ضخیم توزیع را معمولا نقاط پرت ایجاد می‌کنند. این نقاط پرت بر برآوردگرهای کمترین توآن‌های دوم تاثیر می‌گذارند. داده‌های پرت در برازش کمترین توان‌های دوم را در حد زیادی به سمت خود می‌کشد که در نتیجه تعیین و تشخیص این دور افتاده‌ها مشکل می‌شود زیرا باقیمانده‌ی مربوط به آن‌ها به طور ساختگی و مصنوعی کوچک می‌باشند. بنابراین باید دنبال راه‌حلی باشیم که این مشکل را برطرف کند که کاربران آماری رگرسیون استوار[۲]را پیشنهاد می‌دهند، برای کاهش اثر مشاهداتی به کار می‌رود که اگر روش کمترین توان‌های دوم تاثیر گذاری بالایی خواهند داشت یعنی اینکه روش استواری می‌خواهد باقیمانده‌های مرتبط با دور افتاده‌های بزرگ را کنار بگذارد که به این وسیله تشخیص نقاط پرت خیلی ساده‌تر می‌شود. رگرسیون استوار علاوه برحساس نبودن نسبت به داده‌های پرت زمانی که مشاهدات داری توزیع نرمال هستند رگرسیون استوار دارای کارایی ۹۰ تا ۹۵ درصد

[۲] Robust Regression

نسبت به کمترین توان‌های دوم می‌باشد.
امروزه آماردانان بیشتر از روش‌های استواری و تعمیم آنها استفاده می‌کنند، که ما در این بخش به معرفی چند مورد از آنها می‌پردازیم از جمله M_برآوردگر[۳]، L_ بروآردگرها[۴]و R_ برآوردگرها[۵]. هدف اصلی ما در این پایان نامه بررسی برخی از روش‌های استواری اشاره شده است، روش‌هایی که تاکید می‌کنند این است که به وسیله برخی از این سازگاری‌ها می‌توان هر زمان کمترین توان‌های دوم یا تعمیم آن‌ها استفاده شود مورد استفاده قرار گیرد.

۲.۲ برآورد استوار پارامتر مکان

در یک توزیع با تابع چگالی احتمال $f(x)$، و پارامتر مکان θ نامعلوم $-\infty < \theta < +\infty$ برای بدست آوردن تابع چگالی $f(x-\theta)$، فرض کنید که $X_1, X_2, \ldots, X_n$ یک نمونه‌ی تصادفی از این توزیع را نشان می‌دهد، یکی از روش‌های مشهور برای برآورد پارامتر مکان استفاده از درست‌نمایی ماکسیمم است. اگر تابع درست نمایی را با $L(\theta)$ نشان دهیم

$$lnL(\theta) = \sum_{i=1}^{n} lnf(x_i - \theta) = -\sum_{i=1}^{n} \rho(x_i - \theta) \tag{۱.۲}$$

که $\rho(x) = -\ln f(x)$، که با مشتق گرفتن نتیجه به صورت زیر حاصل می‌شود

$$\frac{\partial lnL(\theta)}{\partial \theta} = -\sum_{i=1}^{n} \frac{f'(x_i - \theta)}{f(x_i - \theta)} = \sum_{i=1}^{n} \psi(x_i - \theta) \tag{۲.۲}$$

اگر $\rho'(x) = \psi(x)$ آن‌گاه

$$\sum_{i=1}^{n} \psi(x_i - \theta) = 0 \tag{۳.۲}$$

$L(\theta)$ را ماکسیمم می‌کند. به θای که در رابطه بالا صدق کند برآوردگر درست‌نمایی ماکسیمم گوییم و آن را با $\hat{\theta}$ نمایش می‌دهیم، به مثال‌های زیر توجه کنید:

۱_ نرمال

$$\rho(x) = \frac{x^2}{2} + c \tag{۴.۲}$$

در این‌صورت $\psi(x) = x$ و

$$\sum_{i=1}^{n} (x_i - \theta) = 0 \Longrightarrow \hat{\theta} = \bar{x} \tag{۵.۲}$$

---
[۳] M estimator
[۴] L estimator
[۵] R estimator

۲ـ نمایی دوگانه[۶]

$$\rho(x) = |x| + c \qquad (۶.۲)$$

با توجه به این‌که $\psi(x) = \begin{cases} -۱ & x < ۰ \\ ۱ & x > ۰ \end{cases}$ است، در این‌صورت $\hat{\theta}$ میانه نمونه می‌باشد.

۳ـ کوشی[۷]

$$\rho(x) = \ln(۱ + x^۲) + c \qquad (۷.۲)$$

و با توجه به اینکه $\psi(x) = \frac{۲x}{۱+x^۲}$ بنابراین $\hat{\theta}$ با روش تکرار حل می‌شود.

توجه داشته باشید که تابع ψ در مثال‌های مربوط نمایی دوگانه و کوشی کراندار است هم‌چنین در مثال مربوط به توزیع کوشی با افزایش x به طور مجانبی به صفر نزدیک می‌شود که این جواب‌ها خیلی تحت تاثیر نقاط پرت قرار نمی‌گیرند ولی از سوی دیگر برآوردگر کمترین توان دوم $\bar{x}$ در مثال نرمال به شدت تحت تاثیر نقاط پرت قرار دارد، برآوردگر کمترین توان دوم $\bar{x}$ به اندازه حالتی که توزیع دم پهن مانند کوشی در نظر گرفته شود، خوب نیست. بنابراین در برآورد استواری ما به دنبال برآوردگری هستیم که تقریبا کارا باشد یعنی کارایی(۹۰ تا ۹۵ درصدی)، برای حالتی که توزیع نرمال است، تحت فرضیه نرمال گاهی اوقات مقدار کمی از کاریی (تقریبا نزدیک به ۵ درصد) کاسته می‌شود.

برای یک دلیل نظری خاص (در واقع مینیمم کردن ماکسیمم واریانس مجانبی از برآوردگرهای مربوط به یک کلاس از توزیع‌ها)، هوبر (۱۹۶۴) پیشنهاد داد که برای برآوردگر استوار، برآورد کننده ماکسیمم درست نمایی از پارامتر مکان مربوط به چگالی مانند یک نرمال در مرکز و نمایی دوگانه در دم‌ها استفاده می‌شود. به طور خاص تابع ρ تابع هوبر است اگر

$$\rho(x) = \begin{cases} \frac{۱}{۲}x^۲ & |x| \leqslant k \\ k|x| - \frac{k^۲}{۲} & k < |x| \end{cases} \qquad (۸.۲)$$

هم‌چنین تابع ψ مربوط به آن به صورت زیر می‌باشد

$$\psi(x) = \begin{cases} -k & x < -k \\ x & -k \leq x \leq k \\ k & k < x \end{cases} \qquad (۹.۲)$$

که با روش تکراری حل می‌شود. یک برآوردگر از این نوع (لزوما از این تابع ψ خاص نیست)، برآوردگر برای این حالت را با $\hat{\theta}$ نمایش می‌دهیم که این را M ـ برآوردگر می‌نامیم، یکی دیگر از

[۶] Double Exponential

[۷] Cauchy

ویژگی‌های M– برآوردگرها به صورت زیر است.
فرض کنید $\hat{\theta}$ یک برآوردگر برای نمونه خاص $x_۱, x_۲, \ldots, x_n$ باشد، سپس این موارد توسط بعضی از آنها جایگزین شده باشد به عنوان مثال انحراف از $\hat{\theta}$ سه برابر شده باشد. این برآورگر جدید $\hat{\theta}$ با استفاده از نمونه تعدیل یافته لزوما یکسان نیست. برای به‌دست آوردن مقیاس ثابت از این برآوردگر ما می‌توانیم معادله زیر را حل کنیم

$$\sum_{i=۱}^{n} \psi\left(\frac{x_i - \theta}{d}\right) = ۰ \tag{۱۰.۲}$$

که d یک برآورد استوار برابر پارامتر مقیاس است. به طور خاص d یک آماره خوب استفاده شده در این راه حل است

$$d = \frac{median|x_i - median(x_i)|}{(۰/۶۷۴۵)}. \tag{۱۱.۲}$$

صورت کسر d، میانه انحراف قدر مطلق تقسیم بر ۰/۶۷۴۵ است. اگر n بزرگ باشد تقریبا $d \approx \sigma$ در واقع نمونه از یک توزیع نرمال حاصل می‌شود. معمولا انحراف معیار استاندار نمونه s به عنوان یک مقدار d استفاده نمی‌شود زیرا خیلی تحت تاثیر نقاط دور افتاده قرار می‌گیرد و در نتیجه استوار نیست.
این طرح ویژه از انتخاب d مقدار مناسب را پیشنهاد می‌دهد. اگر توزیع پایه‌ای نرمال باشد، ثابت هموار ساز k باعث بالا بردن کارایی $\hat{\theta}$ می‌شود. در حالت نرمال، ما بیشترین سوال‌ها را برای برآورد نابرابری خواهیم داشت

$$\left|\frac{x_i - \theta}{d}\right| \leq k \tag{۱۲.۲}$$

زیرا که

$$\psi\left(\frac{x_i - \theta}{d}\right) = \frac{x_i - \theta}{d}. \tag{۱۳.۲}$$

در واقع اگر همه‌ی سوال‌ها از این نابرابری باشند سپس $\hat{\theta} = \bar{X}$ به عنوان برآوردگری از توزیع نرمال است.

اگر $d \approx \sigma$، آن‌گاه k معمولا نزدیک به ۱/۵ است. وقتی $k = ۱/۵$ است از روش هوبر استفاده می‌شود. اگر σ معلوم (d معلوم) باشد، تحت فرض نرمال دارای کارایی بیش از ۰/۹۵ درصدی می‌باشد (هوبر ۱۹۶۴). اندروز و همکاران (۱۹۷۲) به این نتیجه رسیدند که در شرایط دم سنگین کارایی بهتر می‌شود، اما در بسیاری از موارد که توزیع‌ها غیر نرمال هستند و σ نامعلوم، پارامتر σ باید برآورد شود که این باعث می‌شود مقدار بسیار کمی از کارایی کاهش یابد زمانی که $n \geq ۱۰$ است.
توابع ψ دیگر که عموما مورد استفاده قرار می‌گیرد در زیر آورده شده

۱. (a,b,c) همپل[۸]

$$\psi(x)=(sign x)\begin{cases}|x| & 0\le |x|<a\\ a & a\le |x|<b\\ a\frac{c-|x|}{c-b} & b\le |x|<c\\ 0 & c\le |x|\end{cases} \qquad (۱۴.۲)$$

مقادیر خوب و ثابتی را که می‌توان برای ثابتها a,b,c انتخاب کرد به صورت زیر می‌باشد

$$a=1/7 \qquad b=3/4 \qquad c=8/5. \qquad (۱۵.۲)$$

۲. k– موج اندروز[۹] که به‌صورت زیر می‌باشد

$$\psi(x)=\begin{cases}sin(\frac{x}{k}) & |x|\le k\pi\\ 0 & |x|>k\pi\end{cases} \qquad (۱۶.۲)$$

با مقادیر $k=1/5$، یا برابر ۲ است، در واقع اگر پارامتر مقیاس معلوم، و $k=1/339$، شرایط برای ما ایده‌ال است به عبارت دیگر مقدار کمی از کارایی کم می‌شود چیزی نزدیک به ۵ درصد.

۳. k– دو وزنی توکی[۱۰]

$$\psi(x)=\begin{cases}x\left(1-(\frac{x}{k})^2\right)^2 & |x|\le k\\ 0 & |x|>k\end{cases} \qquad (۱۷.۲)$$

با مقادیر $k=5$ یا ۶ است، در واقع اگر پارامتر مقیاس معلوم باشد، و $k=4/685$، شرایط برای ما ایده‌ال است به عبارت دیگر مقدار کمی از کارایی کم می‌شود چیزیی نزدیک به ۵ درصد. لازم به ذکر است که موج اندروز و دو وزنی توکی بسیار شبیه به هم هستند و جایگزین مناسبی برای یکدیگر می‌باشند زیرا تابع ρ با این سه تابع ψ پیوند داده شده و این سه تابع محدب نیستند، اگرچه ممکن است مشکلات همگرایی خاصی در روش تکرار وجود داشته باشد با توجه به اینکه این روش موفقیت آمیز بوده اما باید تحت یک سری شرایط مورد استفاده قرار گیرد.

به عنوان مثال فرض کنید با روش تکرار ما جوابی را برای $\hat{\theta}$ پیدا می‌کنیم یعنی $\hat{\theta}$، M– برآوردگر باشد بطوریکه $\sum_{i=1}^{n}\psi\frac{(x_i-\hat{\theta})}{d}=0$ برای هر تابع ψ فرد کراندار معادله آخر تقریبا آسان می‌شود

[۸] Hampel

[۹] (k) Wave of Andrews

[۱۰] (k) Biweight of Tukey

با جایگزین کردن اعضای سمت چپ توسط سری تیلور، بسط مقدار درستی درباره پارامتر θ می‌باشد. از این تقریب زمانی که $d \approx \sigma$

$$\sqrt{n}(\hat{\theta}-\theta) \approx \frac{\sigma[\sum_{i=1}^{n}\psi(\frac{x_i-\theta}{\sigma})]/\sqrt{n}}{[\sum_{i=1}^{n}\psi'(\frac{x_i-\theta}{\sigma})]/n} \tag{۱۸.۲}$$

و این عبارت دارای توزیع نرمال با میانگین صفر و واریانس زیر می‌باشد

$$\frac{\sigma^2 E\left[\psi^2(\frac{x-\theta}{\sigma})\right]}{\left[E(\psi'(\frac{x-\theta}{\sigma}))\right]^2} \tag{۱۹.۲}$$

البته از آن‌جا که چگالی آن را نمی‌دانیم، ما باید تقریبی از مقدار مورد انتظار و σ که به صورت مجانبی است را به‌دست آوریم.

یک تقریب بر ای واریانس $\sqrt{n}(\hat{\theta}-\theta)$ به‌صورت زیر است

$$s_1^2 = \frac{d^2[\frac{1}{n}\sum_{i}^{n}\psi^2(\frac{x_i-\hat{\theta}}{d})]}{[(\frac{1}{n})\sum_{i=1}^{n}(\psi'(\frac{x_i-\hat{\theta}}{d}))]^2} \tag{۲۰.۲}$$

برای n های بزرگ $\sqrt{n}(\hat{\theta}-\theta)/s_1$ که تقریبا دارای توزیع نرمال استاندارد می‌باشد.

هوبر (۱۹۷۰) بررسی کرد توزیع تقریبی بهتری برای $\sqrt{n-1}(\hat{\theta}-\theta)/s_1$ می‌توان با استفاده از عضوی از خانواده t با درجه آزادی $n-1$ پیدا کرد یا حتی دارای درجه آزادی کمتر از $n-1$ باشد. با یکی از این توزیع‌های تقریبی ما می‌توانیم در مورد θ استنباط آماری انجام دهیم.

گروس (۱۹۷۶) فواصل اطمینان را براساس این ایده هم‌چنین نتایج حاصل از طرح‌های دیگر را بررسی کرد و اعتبار استواری (ضریب اطمینان ۹۵ درصدی) و کارایی (ایجاد فواصل نسبتا کوتاه) را پیدا کرد.

۳.۲ رگرسیون استوار

در حالتی که رگرسیون کلاسیک تحت تاثیر نقاط دور افتاده قرار می‌گیرند از این‌رو رگرسیون استوار با استفاده از روش‌های استوار تاثیر این نقاط دور افتاده را کم می‌کند مدل رگرسیون زیر را در نظر بگیرید

$$\mathbf{Y} = \mathbf{X}_i^\top \boldsymbol{\beta} + \varepsilon, \qquad i = 1, 2, \ldots, n \tag{۲۱.۲}$$

$\mathbf{Y}$ برداری تصادفی $n \times 1$، $\mathbf{X}$ ماتریس طرح است با بعد $n \times p$ طوری که $\mathbf{X'X}$ یک ماتریس رتبه کامل است، $\boldsymbol{\beta}$ یک بردار $p \times 1$ از ضرایب مجهول است و ε بردار تصادفی $n \times 1$ است که عناصر آن نمونه‌ای از یک توزیع متقارن حول صفر (مثلا نرمال) هستند. برای مشابه رهیافت پارامتر مکان، ما می‌خواهیم مجموعه زیر را مینیمم کنیم

$$\sum_{i=1}^{n} \rho\left(\frac{\mathbf{Y}_i - \mathbf{X}_i^\top \boldsymbol{\beta}}{d}\right) \tag{۲۲.۲}$$

که در آن Y_i، i امین عنصر از $\mathbf{Y}$ و X_i، i مین سطر از ماتریس $\mathbf{X}$ است و d برآورد پارامتر مقیاس توزیع مربوط به ε است. معادله مشتقات جزیی مرتبه اول مربوط به عناصر $\boldsymbol{\beta}$ را، که با β_j نمایش می‌دهیم، برابر صفر قرار می‌دهیم. همانطور که مشاهده می‌کنیم این کار هم ارز با یافتن ماکسیم جواب از رابطه زیر می‌باشد

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{\mathbf{Y}_i-\mathbf{X}_i^{\top}\boldsymbol{\beta}}{d}\right)=0, \qquad j=0,1,\ldots,p \tag{۲۳.۲}$$

فرض کنید بخواهیم ابتدا برآورد β_0 حاصل از $\boldsymbol{\beta}$ را به‌دست بیاوریم برای این کار به دو چیز نیاز داریم

- برآورد استوار d حاصل از پارامتر مقیاس
- نقطه شروع برای بدست آوردن $\hat{\boldsymbol{\beta}}$

اگر برآوردگر L_1 (مینیمم کردن مجموع قدر مطلق خطاها) بتواند ضرایب رگرسیونی را به خوبی پیدا کند که این کار مارا راحت می‌کند چیزیی شبیه به میانه در یک نمونه واحد است. علاوه بر این برای برآورد پارمتر مقیاس، میانه مقادیر قدرمطلق باقیمانده‌های غیرصفر تقسیم بر ۶۷۴۵/ مانند مورد پارامتر مکان می‌باشد. اگر چه آمار دانان تلاش زیادی در زمینه L_1 برآوردگرها کردند اما با این حال ناراضی بودند، بنابراین برای برآورد پارامتر $\boldsymbol{\beta}$ و مقیاس به صورت همزمان سراغ الگوریتم معرفی شده توسط داتر(۱۹۷۷) می‌رویم، که این الگریتم فقط برای تابع ψ در هوبر توضیح داده شده است. همان‌طور که مشاهده می‌کنیم این روش شبیه کمترین توان‌های دوم می‌باشد. البته باید به این نکته هم توجه کنیم که این روش فقط یکی از روش‌هایی است که در مقاله داتر ذکر شده است اجازه دهید با برخی از برآوردهای اولیه شروع کنیم، β_0 و d_0 ممکن است به خوبی برآوردهای معمولی برای پارامتر $\boldsymbol{\beta}$ و مقیاس باشند ما به صورت زیر عمل می‌کنیم

- باقیمانده‌ها را به‌صورت زیر محاسبه می‌کنیم

$$z_i=\mathbf{Y}_i-\mathbf{X}_i^{\top}\boldsymbol{\beta}_0, \qquad i=1,2,\ldots,n \tag{۲۴.۲}$$

- یک برآوردگر از پارامتر مقیاس به‌صورت زیر به‌دست می‌آوریم

$$d_1^2=\frac{1}{(n-1)\,E\left(\psi^2\right)}\sum_{i}^{n}\left(\psi\left(\frac{z_i}{d_0}\right)\right)^2 d_0^2 \tag{۲۵.۲}$$

که در آن $E(\psi^2)$، امید ریاضی مقادیر $\psi^2(w)$ هوبراست و ω یک متغیر تصادفی است که از توزیع نرمال استاندارد پیروی می‌کند.

- باقیمانده‌های وینزوریده شده[11] به‌صورت زیر تعیین می‌شوند

$$\Delta_i=\psi\left(\frac{z_i}{d_1}\right)d_1 \tag{۲۶.۲}$$

[11] Winsorize the Residuals

توجه کنید که باقیمانده‌ها وینزوریده شده $d_1\,\psi\left(\frac{z_i}{d_1}\right)$ معادل z_i هستند، اگر $|\frac{z_i}{d_1}| \leq k$ اما برابر است با $kd_1\,(-kd_1)$ اگر $kd_1 < z_i\,(z_i < -kd_1)$، برای کسب اطلاعات بیشتر درمورد وینزوریده کردن به دیکسون و توکی (۱۹۶۸) مراجعه شود.

- برآورد کمترین توان‌های دوم[12] ضرایب رگرسیونی به شرطی که باقیمانده‌های وینزوریده شده مشاهده شده باشند به‌صورت زیر می‌باشد

$$\hat{\xi_0} = (X'X)^{-1}\,X'\Delta_0 \tag{۲۷.۲}$$

که (۲۷.۲) Δ_0 یک بردار ستونی $p \times 1$ حاصل از $\Delta_1, \Delta_2, \ldots, \Delta_n$ است.

- یک برآورد جدید از $\boldsymbol{\beta}$ محاسبه کنید یعنی

$$\boldsymbol{\beta} = \boldsymbol{\beta}_0 + q\hat{\xi_0} \tag{۲۸.۲}$$

داتر دریافت یک انتخاب مناسب از فاکتور q به شکل زیر می‌باشد

$$q = min\left[\frac{1}{\Phi(k) - \Phi(-k)}, 1/9\right] \tag{۲۹.۲}$$

که Φ تابع توزیع نرمال استاندارد می‌باشد.

با برآوردهای جدید، $\boldsymbol{\beta}_1$ و d_1 به عنوان مقادیر اولیه برای مراحل یک تا پنج تکرار می‌شوند. این روند تکراری را $i = 1, 2, \ldots, p$ ادامه می‌دهیم در این‌صورت داریم

$$|q\hat{\xi}^i_m| < \varepsilon d_{m+1}, \qquad |d_{m+1} - dm| < \varepsilon d_{m+1} \tag{۳۰.۲}$$

که در آن $\varepsilon > 0$ یک سطح تحمل مناسب $\hat{\xi}^i_m$، i- امین مولفه از $\hat{\xi}_m$ و x_{ii}، i امین عنصر قطری از $(\mathbf{X'X})^{-1}$ است در این زمان تکرار را متوقف کنید و $\boldsymbol{\beta}$ و پارامتر مقیاس را با به کار بردن $\boldsymbol{\beta}_{m+1}$ و d_{m+1} برآورد کنید. خواننده باید دقیقا مشخص کند که این روش برای کمترین توان‌های دوم است یا خیر.

البته مرحله چهار کمترین توان‌های دوم برای باقیمانده‌های وینزوریده شده است و مرحله پنج برآورد کمترین دوم را با $\boldsymbol{\beta}$ شامل $q = 1$ و Δ_0هایی برابر حاصل از باقیمانده‌های واقعی (غیر وینزوریده شده) برای هر برآورد اولیه $\boldsymbol{\beta}_0$ ارائه می‌دهد. همچنین جالب است که برای محاسبه‌ی

$(\mathbf{X'X})^{-1}$ نیاز به محاسبه‌ی یک بار روند تکراری است که این خود باعث صرف جویی در محاسبات می‌شود، سپس یک برآورد استوار خوب از $\boldsymbol{\beta}_0$ و پارامتر مقیاس با استفاده از الگوریتم معرفی شده توسط داتر با روش‌های کمترین مقدار مطلق یا دیگر روش‌ها (برخی روش‌های ناپارامتری که ممکن است بسیار مناسب‌تر باشد) بدست می‌آوریم. ما می‌توانیم مشکل نقاط پرت را با به کار بردن برآوردهای استوار $\tilde{d}_0$ و $\tilde{\boldsymbol{\beta}}_0$ و یک تابع صعودی ψ برطرف کنیم مانند موجی یا وزنی در این موارد کمترین توان‌های دوم موزون، می‌تواند جایگزین مناسبی برای استفاده باشد.

[12] Least squares Estimates

۱.۳.۲ M ـ برآوردگر

معادله (۲۲.۲) را نسبت به پارامترهای β_j، $j = ۱,۲,\ldots,k$ مینیمم کنیم با مشتق گیری از معادله (۲۲.۲) نسبت به تک به تک پارامترها منجر به $p = k + ۱$ معادله به شکل زیر می‌شود

$$\sum_{i=۱}^{n} x_{ij}\psi \frac{(\mathbf{Y}_i - \mathbf{X}_i^\top \boldsymbol{\beta})}{d} \tag{۳۱.۲}$$

$\mathbf{X}_i' = ۱, X_{i۱}, X_{i۲}, \ldots, X_{ik}$ باشد. این معادله‌ها در کل جواب واضحی ندارند و جواب عددی مکرر نیاز است. همانند بیتون و توکی (۱۹۷۴) وزن‌ها را به صورت زیر تعریف می‌کنیم

$$W_{i\beta} = \frac{\psi\left\{\frac{(\mathbf{Y}_i - \mathbf{X}_i^\top \boldsymbol{\beta})}{d}\right\}}{\frac{(\mathbf{Y}_i - \mathbf{X}_i^\top \boldsymbol{\beta})}{d}} \tag{۳۲.۲}$$

و در صورتی که دقیقا $\mathbf{Y}_i - \mathbf{X}_i^\top \beta$ اتفاق بیافتد آن را با یک تعریف می‌کنیم از اینرو معادله (۳۱.۲) را می‌توان به‌صورت زیر بازنویسی کرد

$$\sum_{i=۱}^{n} x_{ij} W_{i\boldsymbol{\beta}} \left(\mathbf{Y}_i - \mathbf{X}_i^\top \boldsymbol{\beta}\right) \qquad j = ۰,۱,۲,\ldots,n \tag{۳۳.۲}$$

آن‌گاه

$$\sum_{i=۱}^{n} x_{ij} W_{i\boldsymbol{\beta}} \mathbf{X}_i' \boldsymbol{\beta} = \sum_{i=۱}^{n} x_{ij} W_{j\boldsymbol{\beta}} \mathbf{Y}_i, \qquad j = ۰,۱,۲,\ldots,n \tag{۳۴.۲}$$

که می‌توان به‌صورت ماتریسی زیر بازنویسی کرد

$$\mathbf{X}' W_{\boldsymbol{\beta}} \mathbf{X} \boldsymbol{\beta} = \mathbf{X}' W_{\boldsymbol{\beta}} Y \tag{۳۵.۲}$$

که در روابط بالا $W_{\boldsymbol{\beta}} = W_{۱\boldsymbol{\beta}}, W_{۲\boldsymbol{\beta}}, \ldots, W_{n\boldsymbol{\beta}}$ این معادلات به وضوح به صورت معادلات نرمال کمترین توان‌های دوم تعمیم یافته هستند. (به طور دقیق‌تر آن‌ها معادلات نوع کمترین توان‌های دوم وزنی هستند زیرا $W_{\boldsymbol{\beta}}$ یک ماتریس قطری از وزن‌ها می‌باشد). مشکل حل کردن آن این است که $W_{\boldsymbol{\beta}}$ به $\boldsymbol{\beta}$ بستگی دارد. با استفاده از یک روش ساده قدیمی این مشکل را حل می‌نماییم به طریقی یک برآورد اولیه $\hat{\boldsymbol{\beta}}_۰$ از $\boldsymbol{\beta}$ را می‌گیریم (شاید با استفاده از کمترین توان‌های دوم). مقدار $W_۰$ را محاسبه می‌کنیم، یعنی $W_{\boldsymbol{\beta}}$ برای همان $\hat{\boldsymbol{\beta}}_۰$ و معادله (۳۵.۲) را برای رسیدن به جواب $\hat{\boldsymbol{\beta}}_۱$ حل می‌کنیم. با استفاده از $\hat{\boldsymbol{\beta}}_۱$ به $W_۱$ می‌رسیم. سپس با استفاده از $W_۱$ به $\hat{\boldsymbol{\beta}}_۱$ می‌رسیم و به همین ترتیب ادامه می‌دهیم. معمولا همگرایی به سرعت رخ می‌دهد. (روند می‌تواند بعد یک تعداد منتخبی از گام‌ها متوقف شود). تنها یک برنامه کمترین توان‌های دوم احتیاج است، مثلا در MINITAB می‌تواند مکرر به صورت زیر نوشته شود

$$\hat{\boldsymbol{\beta}}_{q+۱} = \left(\mathbf{X}' W_q \mathbf{X}\right)^{-۱} \mathbf{X}' W_q \mathbf{Y} \tag{۳۶.۲}$$

و روند ممکن است زمانی که همه‌ی برآوردها کمتر از یک مقدار از پیش تعیین شده مثلا ۱. درصد یا ۰۱. درصد، تغییر کنند یا بعد از یک تعداد گام تعیین شده، متوقف می‌شود. از معادله (۳۲.۲) مشاهده می‌شود که وزن‌ها به مقدار $\frac{\frac{(\partial\rho(x))}{\partial x}}{x}$ وابسته‌اند که در آن $x = \frac{(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta})}{d}$ می‌باشد. بنابراین می‌توانیم برای همه‌ی انتخاب‌های $\rho(x)$ در جدول زیر که آورده شده به تابع $W(x) = \frac{\frac{(\partial\rho(x))}{\partial x}}{x}$ نگاه کنیم. به وضوح انتخاب یک $\rho(x)$ بخصوص اساس یک انتخاب از تابع وزنی است که کاربر برای اختصاص دادن به باقیمانده‌های مقیاس شده در انواع مقادیر مختلف نیاز دارد شخصی می‌تواند حتی به راحتی تابع وزن خود را برای $\rho(x)$ تعریف کند.
یادآوری می‌کنیم که وزن‌ها در (۳۲.۲) برحسب $x = \frac{(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta})}{d}$ تعریف شده‌اند و نیاز به برآورد مقیاس d دارد. یک انتخاب استوار تا حدی محبوبیت پیدا کرده به صورت زیر است

$$d = \frac{median|e_i - median(e_i)|}{۰/۶۷۴۵} \qquad (۳۷.۲)$$

اگر n به قدر کافی بزرگ و خطاها به صورت نرمال توزیع شده باشند برآوردگری تقریبا ناریب را از $SdY_i = \sigma$ به‌دست خواهد داد. (این پیشنهاد از همپل با یک اشتباه چاپی در مخرج در اندروز و همکاران ۱۹۷۲ داده شده است. هم‌چنین گاهی در جاهای دیگر با مخرجی که در حال حاضر معکوس شده همانند ۱/۴۸۲۶ نشان داده شده است). چندین عامل مقیاسی جایگزین نیز پیشنهاد و بررسی شده‌اند، برای مثال هیل و هولاند(۱۹۷۷) را ببینید انتخاب واقعی d اثری روی باقیمانده‌ها خواهد گذاشت، به این معنا که برآوردهای پارامتر رگرسیونی استوار بر انتخاب d ثابت نیستند.
هلند و والش (۱۹۷۷) مقایسه برآوردهای استوار حاصل از ۸ تابع ψ است از جمله این ۸ تابع برآورد بسته های آماری استوار که در مرکز کامپیوتری دفتر ملی تحقیقات اقتصادی Ins توسعه یافته است. برای مثال، از گزارش خودشان یک ثابت هموار ساز، $k = ۱/۵$ برای برآورد موج مناسب است و در صورتی که $d = \sigma$ و معلوم شده باشد دارای توزیع نرمال است، دارای کارایی بیش از ۹۵درصد است. یکی از ویژگی‌های مطلوب تابع ψ مانند تابع موج اندروز یا تابع وزن دوتایی توکی برای شناسایی نقاط دور افتاده زمانی که ما از روش هوبر استفاده می‌کنیم بهتر می‌شوند و در بسیاری از موارد وزن صفر به نقاط پرت یا نقطه داده‌های بد داده می‌شود. البته هر زمان که روش‌های کمترین توان‌های دوم(تعمیم آن‌ها) مورد استفاده قرار می گیرند، به ظاهر امکان استفاده از روش‌های استواری وجود دارد.که شامل ANOVA، رگرسیون، سری زمانی، اسپلاین، تجزیه و تحلیل چند متغیره و تشخیص، برای اطلاعات بیشتر درمورد توابع ρ و توابع وزن به جدول زیر رجوع کنید

جدول ۱.۲: بعضی از توابعی که برای توزیع وزن مانده‌ها پیشنهاد شده‌اند

	معیار	$\rho(x)$	دامنه $\rho(x)$	$\psi(x) = \frac{\partial\rho(x)}{\partial x}$	$w(x) = \frac{\psi(x)}{x}$	مقدار ثابت‌های تنظیم کننده و توصیحات
۱	حداقل مربعات (توزیع مانده‌های نرمال)	$\frac{1}{2}x^2$	$\infty \leq x \leq \infty$	x	۱	
۲	Huber	$\begin{cases}\frac{1}{2}x^2\\ a\lvert u\rvert - \frac{1}{2}a^2\end{cases}$	$-a \leq x \leq +a$ $u < -a and x > a$	x $a * sign(x)$	۱ $\frac{a}{\lvert x\rvert}$	$a = 2$
۳	Ramsay	$\frac{\left\{1-e^{a\lvert x\rvert}(1+a\lvert x\rvert)\right\}}{a^2}$	$\infty \leq x \leq \infty$	$ue^{a\lvert x\rvert}$	$e^{a\lvert x\rvert}$	$a = 0/3$
۴	Andrew	$\begin{cases}a\left\{1-\cos(\frac{x}{a})\right\}\\ 2a\end{cases}$	$-a \leq x\pi \leq +a\pi$ $x < -a\pi, x > a\pi$	$\sin(\frac{x}{a})$ 0	$\sin(\frac{x}{a})\frac{x}{a}$ 0	$a = 1/339$
۵	Tukey	$\begin{cases}\frac{1}{2}x^2 - \frac{x^2}{4a^2}\\ \frac{1}{4}a^2\end{cases}$	$-a \leq x \leq +a$ $x < -a, x > a$	$x(1-\frac{x^2}{a^2})$ 0	$1-\frac{x^2}{a^2}$ 0	$a = 5 \leq a \leq 6$
۶	$Hampel$ $a, b, c > 0$	$\begin{cases}\frac{1}{2}x^2\\ a\lvert x\rvert - \frac{1}{2}a^2\\ a\frac{(c\lvert x\rvert - \frac{1}{2}x^2)}{c-b} - \frac{7a^2}{6}\\ a(b+c-a)\end{cases}$	$-a \leq x \leq +a$ $-b \leq x \leq -a, a \leq x \leq b$ $-c \leq x \leq -b, b \leq x \leq c$ $x < -c, ux > c$	x $a \times sign(x)$ $cSign(x) - (x)a/c$ 0	۱ $\frac{a}{\lvert x\rvert}$ $(c/\lvert x\rvert - 1)(a/c-b)$ 0	$a = 1/7, b = 3/4, c = 8/5$

۲.۳.۲ L- برآوردگر

یک نمونه تصادفی از یک توزیع پیوسته با پارامتر مکان θ در نظر بگیرید آماره‌های ترتیبی نمونه، $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ یک L- برآوردگر ترکیب خطی از آماره‌های ترتیبی می‌باشد. که موارد زیر نمونه‌هایی از L- برآوردگرها می‌باشند

۱ میانه نمونه

$$\bar{X}_\alpha = \frac{\sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)}}{n-2\,[n\alpha]}, \qquad i=[n\alpha]+1,\ldots,n-[n\alpha] \tag{۳۸.۲}$$

۳- برآوردگر گست‌ویرث[۱۳] که یک میانگین وزنی با صدکهای $33\frac{1}{3}$، ۵۰ و $66\frac{2}{3}$ با وزن‌های ۰/۳، ۰/۴ و ۰/۳ می‌باشد.

۳- توکی سه میانگین، که میانگین موزون اولین، دومین و سومین، چارک‌های مربوط به آن‌ها دارای وزن $\frac{1}{4}$، $\frac{1}{2}$ و $\frac{1}{4}$ می‌باشند.

این و سایر L- برآورگرها توسط اندروز وهمکاران (۱۹۷۲)، توضیح داده شده است که به طور کلی L- برآوردگرها دارای وضعیت رگرسیون روشنی درمورد M- برآورگرها نیست زیرا با استفاده از ρ به صورت زیر می‌باشد

$$\rho(x) = |x| \tag{۳۹.۲}$$

که میانه را به عنوان یک برآوردگر (سطح میانه و سطح وضعیت رگرسیون) در نظر می‌گیرد. و این می‌تواند به آسانی تعدیل بیابد برای بدست آوردن سایر صدک‌ها تابع ρ، به صورت زیر می‌باشد

$$\psi(x) = \begin{cases} -(1-p)\,x & x<0 \\ p(x) & x\geq 0 \end{cases} \tag{۴۰.۲}$$

صدک $(100p)th$ در یک تک نمونه و برآوردی از سطح تراز صدک $(100p)th$ است، یا اینکه سطح وضعیت رگرسیون که برای اطلاعات بیشتر به کوکنر و باست (۱۹۷۸) مراجعه کنید.
روشن است تعمیم (تعاریف) از برآوردها مانند آن‌ها از توکی و گست‌ویرث در مسائل رگرسیونی به کار برده می‌شود، علاوه بر این در بسیاری از موارد مانند اطلاعات آموزشی شامل پیش‌بینی از عملکرد دبیرستان، رتبه دبیرستان و نمرات SAT,ACT مورد استفاده قرار می‌گیرد، که علاوه بر این احتمال دارد برخی علاقه‌مند به برخی از صدک‌های دیگر باشند بنابراین برآوردهای صدکی می‌توانند به تنهایی هم‌چنین با سایر برای پیش‌بینی یک سطح یا سطح متوسط به کار برده شود.

[۱۳] Gastwirth' s Estimator

۳.۳.۲ R ـ برآوردگر

برآوردگر رتبه‌ای علاوه بر Mـ برآوردگر که به طور مفصل درباره‌ی آن بحث شد روش دیگری برای رگرسیون استوار وجود دارد روش Rـ برآوردگر براساس رتبه‌ها می‌باشد. برای روشن‌تر شدن روش کلی، جانشین کردن یک عامل رتبه‌ای در تابع حداقل مربعات $\sum_{i=1}^{n}\left(\mathbf{Y}_i-\mathbf{X}_i^{\top}\boldsymbol{\beta}\right)^2$ اگر Rـ برآوردگر رتبه $\left(\mathbf{Y}_i-\mathbf{X}_i^{\top}\boldsymbol{\beta}\right)$، R-برآوردگر می‌گوییم. بنابراین رتبه‌ی $\mathbf{R_i}$ یک تابع از β می‌باشد. بنابراین داریم

$$\sum_{i=1}^{n}\left(\mathbf{Y}_i-\mathbf{X}_i^{\top}\boldsymbol{\beta}\right)\mathbf{R}_i \tag{۴۱.۲}$$

که می‌توان با جایگزین کردن رتبه‌ها $1,2,\ldots,n$ با نمرات $a(1)\le a(2)\le\ldots\le a(n)$ تعمیم دهیم(۴۱.۲) که در اینصورت تابع هدف به صورت زیر می‌باشد

$$\min_{\boldsymbol{\beta}}\sum_{i=1}^{n}\left(\mathbf{Y}_i-\mathbf{X}_i^{\top}\boldsymbol{\beta}\right)a\mathbf{R_i} \tag{۴۲.۲}$$

البته دو مثال از نمره‌ها به صورت زیر می‌باشد

مثال ۱.۳.۲. اگر تابع امتیاز را با رتبه‌ها مساوی قرار بگیرد نمرات (رتبه‌ها) ویلککسون که به صورت زیر می‌باشد

$$a(i)=i \tag{۴۳.۲}$$

مثال ۲.۳.۲. میانه نمرات که به صورت زیر می‌باشد

$$a(i)=\begin{cases}-1 & i<\frac{(n+1)}{2}\\ 1 & i>\frac{(n+2)}{2}\end{cases} \tag{۴۴.۲}$$

جیکل (۱۹۷۲) ثابت کرد که این حداقل رساندن‌ها هم ارز با حل کردن p است

$$\sum_{i=1}^{n}x_{ij}a(\mathrm{R_i})=0,\qquad j=1,2,\ldots,p \tag{۴۵.۲}$$

معادله (۴۵.۲) به صورت تقریبی حل می‌شود زیرا به دلیل ناپیوستگی در $\mathrm{a(.)}$ و $\mathrm{R_i}$ علاوه بر این به‌خوبی معلوم است که نمرات خوب (داشتن خواص مجانبی مشخص) داده می‌شود توسط

$$a(i)=\phi\left[\frac{i}{(n+1)}\right] \tag{۴۶.۲}$$

و همچنین داریم

$$\phi(t)=\frac{\left[-f'F^{-1}(t)\right]}{f\left[F^{-1}(t)\right]} \tag{۴۷.۲}$$

مثال‌های مربوط به (۴۶.۲) و (۴۷.۲)

مثال ۳.۳.۲. f نرمال باشد آن‌گاه داریم

$$\phi(t) = \Phi^{-1}(t), \qquad 0 < t < 1 \tag{۴۸.۲}$$

در این‌صورت نمرات نرمال می‌باشند.

مثال ۴.۳.۲. اگر f نمایی دو گانه باشد آن‌گاه

$$\phi(t) = \begin{cases} -1 & 0 < t < \frac{1}{2} \\ 1 & \frac{1}{2} < t < 1 \end{cases} \tag{۴۹.۲}$$

در این صورت میانه نمرات را می‌دهد.

مثال ۵.۳.۲. f لجستیک باشد آن‌گاه

$$\phi(t) = 2(t) - 1, \qquad 0 < t < 1 \tag{۵۰.۲}$$

در این‌صورت نمرات ویلککسون را می‌دهد.

جورکاکووا (۱۹۷۷) ثابت کرد با نمرات مشخص $\mathrm{a}(.)$ و تابع ψ، R–برآوردگر و M–برآوردگر به طور مجانبی معادل خواهند بود از جمله شراط دیگری که به آن نیاز داریم

$$\phi(t) = c_1 \psi \left[\mathrm{F}^{-1}(\mathrm{t})\right] + c_2 \tag{۵۱.۲}$$

که در این‌صورت لازم است که c_1 و c_2 در (۵۱.۲)ثابت باشند که معادل باشند.

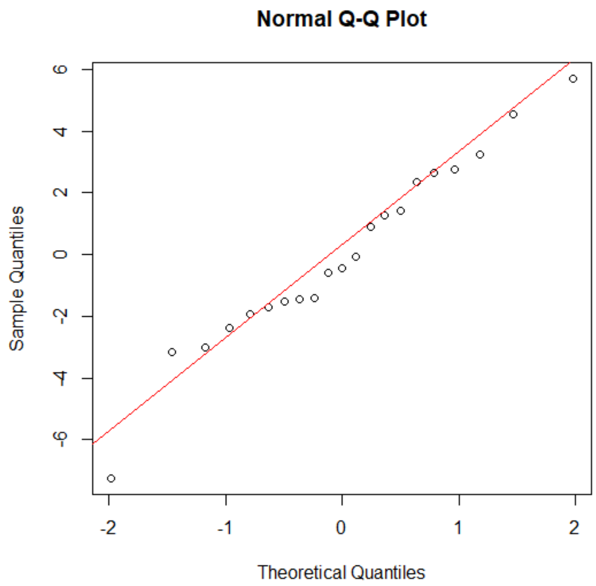
ملاحظه ۱.۳.۲. با توجه به (۵۱.۲) منطقی‌تر این است که از M ـ برآوردگر استفاده کنیم زیرا هم محاسبات کمتری را احتیاج دارد و هم زمان کمتری صرف می‌شود. کاربرد برآورد L ـ برآوردگر در زمینه رگرسیون به سادگی M ـ برآوردگر و R ـ برآوردگر نمی‌باشد. دنی و لارسن رگرسیون قطعه قطعه را تشریح می کند که در آن برای یک متغیر رگرسیون داده‌ها را به گروه‌هایی تقسیم می‌کنند و به کار گیری مرکز نقل گروه‌ها خط مستقیم را برازش می‌دهند. تعمیمم قدم به قدم این تکنیک برای رگرسیون چندگانه به کار می‌رود. موسا ـ هامودا و لئون (۱۹۷۴, $۱۹۷۷a$) روش‌هایی را برای رگرسیون خطی ساده با مشاهدات تکراری y به ازای x پیشنهاد می کنند که مستلزم آراستن و کنار گذاشتن مقادیر حذف شده y می‌باشد.
هرگاه تحقیقات در مورد R ـ برآورد و L ـ برآورد ادامه یابد باور داریم که در حال حاضر M ـ برآوردگرها برای مسائل رگرسیونی قابل قبول‌تر می‌باشند. آن‌ها به سادگی از یک برنامه کامپیوتری حداقل مربعات موزون بدست می‌آیند و در آخرین تکرار وزن‌ها نقاط تاثیر گذار را مشخص می‌کنند و به‌دست آوردن R و L برآورگر از نظر محاسباتی از M ـ برآوردگر مشکل‌تر هستند.

مثالی که در زیر آورده شده داده‌های آن مربوط به کاهش توده می‌باشد برای ارتباط بین مدل رگرسیون استوار و رگرسیون کلاسیک که چه موقع از رگرسیون استوار و چه موقع از رگرسیون کلاسیک استفاده کنیم.

مثال ۶.۳.۲. داده‌های کاهش توده که از یک کارخانه تبدیل اکسید آمونیاک به اسید نیتریک توسط دانیل و وود ۱۹۸۰ تشریح شده است که دارای ۲۱ مشاهده می‌باشند و سه متغیر مستقل (جریان هوا، حرارت آب خنک کننده ورودی و غلظت اسید)، رگرسیون خطی همراه با نمودار برای این مجموعه داده ارائه می‌دهیم، حالت اول کمترین توان‌های دوم برای ۲۱ مشاهده، حالت دوم ۲۱ مشاهده دارای چهار داده‌ی پرت می‌باشند که در این حالت چهار داده‌ی پرت را کنار می‌گذاریم و برای ۱۷ مشاهده باقیمانده رگرسیون خطی ساده را انجام می‌هیم و حالت آخر برای ۲۱ مشاهده رگرسیون استوار استفاده شده، انجام می‌دهیم

- مدل رگرسیون خطی برای ۲۱ مشاهده تقریبا به صورت زیر می‌باشد و $MSE = ۹/۳۵$ می‌باشد.

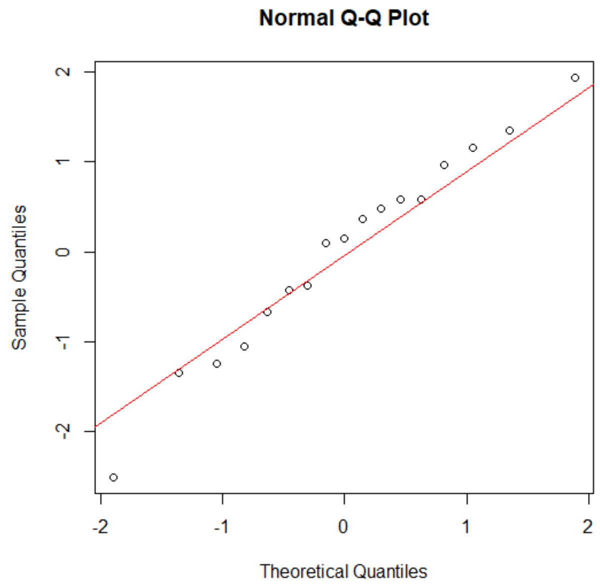
$$\hat{Y} = -۴۰ + ۰/۷ X_۱ + ۱/۳ X_۲ - ۰/۱۵ X_۳$$



شکل ۱.۲: کمترین توان‌های دوم با ۲۱ مشاهده

- مدل رگرسیونی خطی برای ۱۷ مشاهده که ۴ داده‌ی پرت به صورت زیر می‌باشد و اینکه $MSE = ۳۴/۳۷$ است.

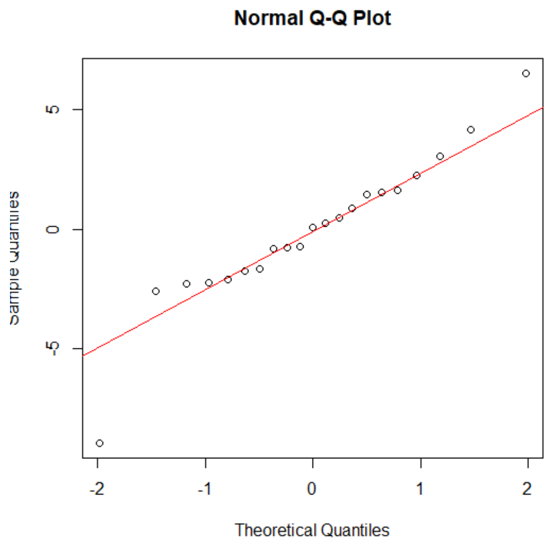
$$\hat{Y} = -۳۷/۷ + ۰/۸ X_۱ + ۰/۶ X_۲ - /۰۷ X_۳$$



شکل ۲.۲: کمترین توان‌های دوم با ۱۷ مشاهده

- مدل رگرسیون استوار برای۲۱ مشاهده به صورت زیر است و $MSE = ۸/۵$ می‌باشد.

$$\hat{Y} = -۴۱ + ۰/۸X_۱ + ۰/۹X_۲ - ۰/۱۲X_۳$$



شکل ۳.۲: رگرسیون استوار با ۲۱ مشاهده

همان‌طور که مشاهده می‌کنیم فرق سه مدل در این است که در حالت استوار که از روش M– برآوردگر با استفاده از نرم‌افزار R و پکیج MASS استفاده شده با خطای کمتری مواجه هستیم

نسبت به دو حالت دیگر شاید در ظاهر با حذف داده‌های پرت شکل بهتر باشد اما خطای زیادی را داریم چون داده‌هایی را که از مدل حذف می‌کنیم احتمالا داده‌های مهمی هستند که نباید آن‌ها را حذف کنیم پس بهتر است از روشی استفاده کنیم که نتیجه‌ی بهتری را در بر داشته باشد که رگرسیون استوار این مشکلات را حل می‌کند.

فصل ۳

سازگاری قوی M ـ برآوردها با خطاهای وابسته زبرجمعی منفی

با توجه به وجود داده‌ی پرت در مدل نیاز داریم که از روش بهتری برای شناسایی آن‌ها استفاده کنیم که از روش M ـ برآورد استفاده می‌کنیم ولی بیشتر کارهایی که انجام شده برای متغیرهای تصادفی مستقل نتایج موجود است برای متغیرهای تصادفی وابسته خصوصا متغیرهای تصادفی وابسته NSD.

۱.۳ سازگاری قوی M ـ برآودهای پارامترهای مدل رگرسیونی

در این بخش که از مقاله وانگ و هو (۲۰۱۵) استفاده کردیم درباره سازگاری قوی M ـ برآوردها[۱] پارامترهای رگرسیونی در مدل خطی با خطاهای تصادفی زبرجمعی منفی وابسته بحث می‌کنیم که نتیجه‌ی آن بهبود شرایط گشتاوری و تعمیم نتایج موجود برای خطاهای تصادفی مستقل است.
مدل خطی رگرسیون زیر را در نظر بگیرید

[۱] strong consistency of M-estimates

$$\mathbf{Y}_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \mathbf{e}_i, \qquad i = ۱, ۲, \ldots, n \tag{۱.۳}$$

که در (۱.۳) $\mathbf{x}_i$ بردار $p \times ۱$ معلوم می‌باشد، $\boldsymbol{\beta} \in \mathbb{R}^p$ پارامتر مجهول رگرسیونی، $\mathbf{e}_i$ بردار خطاهای تصادفی هستند. $\hat{\boldsymbol{\beta}}_n$ یک M_ برآورد برای $\boldsymbol{\beta}$ به‌صورت زیر تعریف می‌شود

$$\min \sum_{i=۱}^{n} \rho(e_i) = \min_{\beta} \sum_{i=۱}^{n} \rho \left(\mathbf{Y}_i - \mathbf{X}_i^\top \boldsymbol{\beta} \right) \tag{۲.۳}$$

در رابطه (۲.۳)، ρ تابعی محدب و غیریکنوا می‌باشد، برای نمونه می‌توان به موارد زیر اشاره کرد.

- هوبر

$$\rho(x) = \left(x^۲ I \left\{ |x| \leqslant c \right\} \right) /۲ + \left(c|x| - c^۲/۲ \right) I \left\{ |x| > c \right\}, \quad c > ۰,$$

- چارک‌های رگرسیونی[۲]

$$\rho(x) = \rho_\alpha(x) = \alpha x I \left\{ x > ۰ \right\} - (۱ - \alpha) x I \left\{ x \leqslant ۰ \right\}, \qquad ۰ < \alpha < ۱.$$

مقدار قابل توجهی از کار مربوط به ویژگی مجانبی M_ برآوردها را هوبر (۱۹۷۳) و (۱۹۸۱) بررسی کرد، سازگاری ضعیف M_ برآوردها ژائو، رائو و چن (۱۹۹۳) را مورد مطالعه قرار دادند، هی و شائو (۱۹۹۶) نمایشی از بهادر قوی برای یک کلاس عمومی M_ برآوردها را مورد مورد مطالعه قرار دادند، کول و سارگالیس (۲۰۰۰) یک مرتبه بالاتر از بسط مجانبی از یک کلاس M_ برآوردها از پارامتر رگرسیونی زمانی که خطاها یک میانگین متحرک با طول عمر طولانی را تشکیل می‌دهند ارائه کردند، ژائو (۲۰۰۲) سازگاری قوی M_ برآوردهای پایدار از ضرایب رگرسیون تحت برخی شرایط ضعیف را مورد مطالعه قرار داد، توزیع‌های مجانبی M_ برآوردها با خطاهای مرتبط با حالت فضایی را کویی و هی (۲۰۰۴) استخراج کردند، وو (۲۰۰۷) خاصیت مجانبی M_ برآوردها از پارامترهای رگرسیونی با خطاهایی که وابسته هستند، ضعف و نمایش بهادر قوی از M_ برآوردها و یک قضیه‌ی حد مرکزی به‌دست آورد، گاو، لی و لین (۲۰۰۹) حرکت مجانبی M_ برآوردها از پارامتر رگرسیونی در مدل خطی هنگامی که هریک از متغیرهای تصادفی در مدل ثابت، α_آمیخته شده را بررسی کردند، میا، وو و لیو (۲۰۰۹) خواص مجانبی بازگشتی M_ برآوردها در ضرایب رگرسیونی استخراج کردند، وو (۲۰۱۱) سازگاری قوی M_ برآوردها برای پارامترهای رگرسیونی با خطاهای تصادفی وابسته منفی مطالعه کرد، فان (۲۰۱۲) انحراف متوسط M_ برآوردها با پارامتر رگرسیونی زمانی که خطا به‌صورت یک دنباله مانا ϕ آمیخته باشد بررسی کرد، برای اطلاعات بیشتر می‌توان به

[۲] Regression Quantiles

مراجع زیر مراجعه کنید M ـ برآوردها با مراجعه به یوهای مارونا (۱۹۷۹)، چن و لی (۱۹۸۴)، چن و وو (۱۹۸۸)، رائو و ژائو ،(۱۹۹۲a،b) بای رائو وو (۱۹۹۲)، چن ژائو (۱۹۹۶،۱۹۹۵)، میا و وو (۱۹۹۶)، ژائو و شی (۱۹۹۷)، برلاینت، لیس و وجدا (۲۰۰۰)، ژائو (۲۰۰۰)، یانگ (۲۰۰۲)، وو (۲۰۰۶،۲۰۰۵)، ویدرز و نادراجا (۲۰۰۱۴).

در ادامه به مطالعاتی که برای همبستگی‌ها انجام شده می‌پردازیم.

هو (۲۰۰۰) با ارائه مثالی نشان داد که NSD به معنای پیوند منفی نیست، هم‌چنین برخی از خواص پایه‌ای و سه قضیه اساسی برای NSD ارائه داد و نشان داد که تعدادی از توزیع‌های چند متغیره شناخته شده دارای خاصیت NSD هستند. NA توسط آلام و ساکسنا (۱۹۹۸) مطرح و این‌که NA اشاره به NSD دارد، نلسن (۲۰۰۶) خاصیت وابستگی را می‌توان برای کاربرد نظریه مفصل اعمال کرد، جاورکی،دارنته و هاردله (۲۰۱۳) با دید اقتصادی به متغیرهای تصادفی NSD نگاه کردند و بیشتر تمرکزشون روی کاربرد اقتصادی قرار دادن، چریستوفیتس و واگلاتو (۲۰۰۴) متغیرهای وابسته‌ای که NSD باشند دارای خاصیت NA نیز هستند مورد بررسی قرار دادند، اقبال، امینی و بزرگنیا (۲۰۱۰) قانون قوی اعداد بزرگ برای فرم‌های درجه دوم در متغیرهای تصادفی NSD تحت این فرض که این متغیرها نامنفی هستند، و یک دنباله $\{X_n, n \geq ۱\}$ با $EX_n^r \leq \infty$ برای هر $n \geq ۱$ و $r \geq ۱$ مطرح کردند، همچنین آن‌ها (۲۰۱۱) نامساوی کلموگروف را برای فرم‌های درجه دوم در دنباله نامنفی از متغیرهای تصادفی کراندار یکنواخت NSD ارئه دادند. شن و همکاران (۲۰۱۳) قضیه ثابت همگرایی قریب به یقین و هم‌چنین پایداری قوی برای مقادیر وزنی از متغیرهای تصادفی NSD را مورد مطالعه قرار دادند.

در این پایان نامه ρ را یک تابع محدب و غیریکنوا فرض می‌کنیم که در $\mathbb{R}$ بامشتقات چپ و راست ψ_- و ψ_+ برای هر تابع ψ، $u \in \mathbb{R}$ و $\psi_-(u) \leq \psi(u) \leq \psi_+(u)$ و

$$S_n = \sum_{i=۱}^{n} \mathbf{x}_i \mathbf{x}_i^\top, \qquad d_n = \max_{۱\leq i\leq n} \mathbf{x}_i^\top \mathbf{S}_n^{-۱} \mathbf{x}_i \tag{۳.۳}$$

که برای هر عدد طبیعی $n_۰$، $S_{n_۰} > ۰$ است.

در رابطه با سازگاری قوی M ـ برآورد از $\hat{\beta}_n$ با حداقل رساندن رابطه (۲.۳) نتیجه به صورت زیر می‌باشد

قضیه ۱.۱.۳. فرض کنید که $\{\mathbf{e}_i, i \geq ۱\}$ یک دنباله از متغیرهای تصادفی مستقل و هم‌توزیع و ثابت‌های مثبت $l_۱$ و $l_۲$ وجود داشته باشد، طوری‌که

$$E\left(\rho\left(\mathbf{e}_i + u\right) - \rho\left(\mathbf{e}_i\right)\right) \geq l_۱ u^۲, \quad |\, u \,| < l_۲, i \geq ۱ \tag{۴.۳}$$

و

$$\begin{aligned} E\mid \psi_+\left(\mathbf{e}_i \pm \Delta\right)\mid^m \leq h_m < \infty, \qquad i \geq ۱, m \geq ۱, \\ d_n = O\left(n^{-\delta}\right) \end{aligned} \tag{۵.۳}$$

برای برخی از ثابت‌های مثبت Δ، $h_m\ (m=1,2,\dots)$ و δ آن‌گاه وقتی $n \longrightarrow \infty$، $\boldsymbol{\beta} \xrightarrow{a.s.} \hat{\boldsymbol{\beta}}_n$.

از این‌رو برای سازگاری قوی $\hat{\boldsymbol{\beta}}$ محدودیت اعمال شده توسط شرایط گشتاوری در قضیه (۱.۱.۳) خیلی قوی است، یانگ (۲۰۰۲) نشان داد که می‌توان با جایگزین کردن شرط زیر به جای (۵.۳) نتیجه را بهبود بخشید.

$$E \mid \psi_+ \left(\mathbf{e}_i \pm \Delta\right) \mid^{\frac{2}{\delta}} < \infty \qquad 0 < \delta \leq 1$$

برای تعمیم نتایج بالا در حالت خطاهای NSD نیاز به معرفی برخی شرایط برای توابع ρ و ψ می‌باشد:

- $A1$: مقادیر ثابت مثبت C_0 و Δ وجود دارد طوری‌که برای هر ψ، $h \in (0,\Delta)$ و $u \in \mathbb{R}$ آن‌گاه
$$\psi\left(u+h\right) - \psi\left(u\right) \leq C_0.$$
- $A2$: $E\left(\psi(e_1)\right) = 0$ و ثابت‌های مثبت C_1 و Δ وجود دارد طوری‌که اگر $\mid u \mid < \Delta$ آن‌گاه
$$\mid E\psi\left(e_1 + u\right) \mid \geq C_1 \mid u \mid .$$
- $A3$: برای $0 < \delta \leq 1$
$$d_n = O\left(n^{-\delta}\right).$$
- $A4$: برای $0 < \delta < 1$
$$E \mid \psi(e_1) \mid^{1/\delta} < \infty.$$
- $A5$: $\psi\left(u+h\right) - \psi(u)$ نسبت به u یکنوا باشد یا $\mid \psi(u) \mid \leq C_2$ برای بعضی $C_2 > 0$، $u \in \mathbb{R}$ و $h \in (0,\Delta)$

ملاحظه ۱.۱.۳. شرایط $A1$ تا $A4$ استاندارد است و اغلب آن‌ها بررسی می‌کنند M ـ برآوردها از مدل‌های خطی زمانی که خطاها مستقل هستند برای مثال ژائو (۲۰۰۲) شرط $A1$ محدودیت و مشکلی در خانواده M ـ برآوردها ایجاد نمی‌کند، و برقرار است که $A1$ شامل حداقل کمترین انحراف مطلق و کمترین برآورد توان‌های دوم و همه‌ی برآورد بین آن‌ها استوارتر از دیگر برآوردها می‌باشد برای مثال ژائو و رائو (۱۹۹۲b) و ژائو (۲۰۰۲) را مطالعه کنید.

ملاحظه ۲.۱.۳. شرط $A5$ برای توزیع کردن به خطاهای وابسته و سازگاری قوی M ـ برآوردها مورد نیاز است. مثال‌هایی که در زیر آورده شده شرایط $A5$ برای آن‌ها برقرار است و توابع $\psi(x)$ که در مثال‌های زیر آورده شده شرایط$A1$ و $A5$ برقرار است.

مثال ۱.۱.۳. برآوردگر هوبر

$$\rho(x)=\frac{\left(x^{۲}I\left\{\mid x\mid\leq c\right\}\right)}{۲}+\left(c\mid x\mid-\frac{c^{۲}}{۲}\right)I\left\{\mid x\mid>c\right\}$$

مثال ۲.۱.۳. برآوردگر کمترین توان‌های دوم[۳] به‌صورت زیر می‌باشد

$$\rho(x)=x^{۲}.$$

که تابع $\psi(x)$ آن بصورت زیر می‌باشد:

$$\psi(x)=۲x.$$

مثال ۳.۱.۳. برآورد کمترین انحراف مطلق[۴] به‌صورت زیر می‌باشد

$$\rho(x)=\mid x\mid.$$

که تابع $\psi(x)$ آن به‌صورت زیر می‌باشد:

$$\psi(x)=signx.$$

مثال ۴.۱.۳. رگرسیون چندکی به‌صورت زیر می‌باشد

$$\rho(x)=\alpha xI\left\{x>\circ\right\}-\left(۱-\alpha\right)xI\left\{x\leq\circ\right\}$$

قبل از این‌که $\psi(x)$ را بنویسیم لازم است به این نکات توجه کنیم

$$\alpha I\left\{x>\circ\right\}=\begin{cases}\alpha & x>\circ\\ \circ & o.w.\end{cases}\tag{۶.۳}$$

$$\left(۱-\alpha\right)I\left\{x\leq\circ\right\}=\begin{cases}۱-\alpha & x\leq\circ\quad \circ<\alpha<۱\\ \circ & o.w.\end{cases}\tag{۷.۳}$$

با توجه به (۶.۳) و (۷.۳) تابع $\psi(x)$ به‌صورت زیر می‌باشد

$$\psi(x)=\alpha-I\left\{x\leq\circ\right\}$$

قضیه زیر توسط ژائو (۲۰۰۲) تعمیمی از قضیه یانگ (۲۰۰۲) در شرایط گشتاوری است به صورت زیر بیان شد

[۳] Least Squares Estimator

[۴] absolute deviation estimator

قضیه ۲.۱.۳. فرض کنید $\{\mathrm{e}_i, i \geq ۱\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع و شرایط $A۱$ تا $A۴$ برقرار باشد طوری‌که حداقل یک $q > ۱$، $E \mid \psi(e_۱) \mid^q < \infty$ و $\delta = ۱$ آن‌گاه وقتی $n \longrightarrow \infty$، $\hat{\boldsymbol{\beta}}_n \xrightarrow{a.s.} \boldsymbol{\beta}$.

وانگ و همکاران (۲۰۱۴) یک نتیجه همگرایی کامل برای ارایه‌های هر سطر از متغیرهای تصادفی NSD مطرح کردند که هدف اصلی در این بخش سازگاری قوی M ـ برآوردها $\hat{\beta}_n$ برای β در مدل(۱.۳) با خطای NSD است که ما فرض استقلال و محدودیت شرایط گشتاوری کم می‌کنیم بنابراین قضیه زیر داریم

قضیه ۳.۱.۳. در مدل (۱.۳) فرض کنید $\{\mathrm{e}_i, i \geq ۱\}$ دنباله‌ای از متغیرهای تصادفی هم‌توزیع و NSD و شرایط $A۱$ تا $A۵$ برقرار باشد، آن‌گاه وقتی $n \longrightarrow \infty$، $\hat{\boldsymbol{\beta}}_n \xrightarrow{a.s.} \boldsymbol{\beta}$.

قضیه (۲.۱.۳) و شرط $E \mid \psi(e_۱) \mid^q \leq \infty$ برای برخی $q > ۱$ همچنین مورد نیاز است $\delta = ۱$ و قضیه (۳.۱.۳) بدون نیاز $E \mid \psi(e_۱) \mid^q \leq \infty$، $q > ۱$ و $\delta = ۱$ از این‌رو قضیه(۳.۱.۳) و تعمیم دادن قضیه (۲.۱.۳) برای خطاهای تصادفی مستقل به صورت خطاهای تصادفی NSD تحت شرایط گشتاوری ضعیف‌تر اعمال می‌کنیم در این بخش این نمادها $C, C_۰, C_۱, C_۲, C_۳, \ldots$ ثابت‌های مثبت هستند که ممکن است در جاهای مختلف متفاوت باشند، $I(A)$ تابع نشانگر مجموعه و $[x]$ قسمت صحیح x را نشان می‌دهد. در زیر ابتدا سه لم را آوردیم توضیح می‌دهیم و لم‌های دیگری که مورد نیاز است در فصل یک توضیح داده شده برای اثبات قضیه‌ها این فصل

لم ۱.۱.۳. فرض کنید که $\{X_n, n \geq ۱\}$ یک دنباله از متغیرهای تصادفی NSD باشد طوری‌که برای هر $n \geq ۱$، $EX_n = ۰$ و $b > ۰$، $\mid X_n \mid \leq b \quad a.s.$. فرض کنید $B_n^۲ = \sum_{i=۱}^n EX_i^۲$ آن‌گاه برای هر $\varepsilon > ۰$

$$P\left(\left|\sum_{i=۱}^{n} X_i\right| > \varepsilon\right) \leq ۲\exp\left\{-\frac{\varepsilon^۲}{۲\left(۲B_n^۲ + b\varepsilon\right)}\right\}. \qquad (۸.۳)$$

لم زیر برگرفته از شن و همکاران (۲۰۱۳) می‌باشد.برای اطلاعات بیشتر به آن مراجعه کنید.

لم ۲.۱.۳. فرض کنید $\{X_n, n \geq ۱\}$ یک دنباله از متغیرهای تصادفی NSD باشد، آن‌گاه

۱ـ اگر برای هر n، $EX_n = ۰$ و برخی p، $\sum_{i=۱}^n E \mid X_n \mid^p < \infty$ که $۱ < p \leq ۲$ آن‌گاه $\sum_{i=۱}^n X_i$ همگرای قریب به یقین می‌باشد.

۲ـ اگر X_nها هم‌توزیع و $E \mid X_۱ \mid < \infty$ آن‌گاه وقتی $n \longrightarrow \infty$

$$\frac{۱}{n}\sum_{i}^{n} X_i \xrightarrow{a.s.} EX_۱$$

لم ۳.۱.۳. فرض کنید $\{X_n, n \geq ۱\}$ یک دنباله از متغیرهای تصادفیNSD و $\{a_{nk}, n \geq ۱, k \geq ۱\}$ یک آرایه از اعداد حقیقی باشد همچنین

۱ـ فرض کنید X_nها هم‌توزیع و $EX_n = ۰$ هم‌چنین $max_{۱\le k\le n} \mid a_{nk} \mid= O(n^{-۱})$ وقتی $n \longrightarrow \infty$ آن‌گاه

$$\sum_{i=۱}^{n} a_{nk}X_k \xrightarrow{a.s.} ۰. \qquad (۹.۳)$$

۲ـ فرض کنید که یک ثابت $C > ۰$ وجود دارد به طوری‌که برای هر $t \ge ۰$ و متغیر تصادفی X، $P(\mid X_n \mid > t) \le CP(\mid X \mid > t)$
فرض کنید برای بعضی $۰ < q \le ۲$ و $\mu > ۰$، $E \mid X \mid^{max\left\{q,\frac{۱}{\mu}\right\}} < \infty$ و به‌علاوه فرض کنید برای هر $n \ge ۱$ و $k \ge ۱$، $\mid a_{nk} \mid\le Dk^{-\mu}$ که D یک عدد حقیقی مثبت است و این‌که برای بعضی از $\nu > ۰$، و هر $n \ge ۱$، $\sum_{k=۱}^{\infty} \mid a_{nk} \mid^q \le Dn^{-\nu}$ مقدار q که وجود آن در بالا فرض شود بزرگتر از یک است، برای مثال $q \in (۱, ۲]$ بعلاوه فرض کنید $EX_n = ۰$ آن‌گاه برای هر n

$$T_n = \sum_{k=۱}^{\infty} a_{nk}X_k \xrightarrow{a.s.} ۰, \quad n \longrightarrow \infty. \qquad (۱۰.۳)$$

برهان. بدون این‌که از کلیت مساله چیزیی کم شود فرض کنید برای هر $k \ge ۱$، $n \ge ۱$، $a_{nk} \ge ۰$ به‌علاوه فرض کنید $\varepsilon > ۰$ داده شده باشد M را طوری انتخاب کنیم که $E \mid X_۱ \mid I(\mid X_۱ \mid > M) < \varepsilon$ و تعریف می‌کنیم

$$X'_k = -MI(X_k < -M) + X_kI(\mid X_k \mid\le M) + MI(X_k > M)$$

هم‌چنین $X''_k = X_k - X'_k$ و تعریف می‌کنیم

$$Y'_k = X'_K - EX'_k, \qquad Y''_K = X''_k - EX''_k$$

با توجه به این‌که $EY'_k = ۰$ و Y'_k به طور یکنواخت دارای کران $۲M$ به صورت زیر می‌باشد

$$|Y'_k| = |X'_k - E(X'_k)| \le |X'_k| + |E(X'_k)| \le |X'_k| + E|X'_k| \le M + M$$

هم‌چنین a_{nk} دارای کران مشخصی می‌باشد تمام این کران‌ها را می‌بریم داخل یک C ثابت و با استفاده از نابرابری‌ها و $e^x \le ۱ + x + \frac{۱}{۲}x^۲e^{|x|}$ برای هر $x \in \mathbb{R}$ نتیجه به‌صورت زیر حاصل می‌شود

$$\begin{aligned} E\exp\left\{ta_{nk}Y'_k\right\} &\le ۱ + E\left\{\frac{t^۲}{۲}a_{nk}^۲Y_k'^۲\exp\left(ta_{nk} \mid Y'_k \mid\right)\right\} \\ &\le ۱ + \frac{Ct^۲}{n^۲}\exp\left(\frac{Ct}{n}\right) \\ &\le \exp\left\{\frac{Ct^۲}{n^۲}\exp\left(\frac{Ct}{n}\right)\right\} \end{aligned}$$

که رابطه آخر از $1 + x \leq e^x$ به‌دست می‌آید برای هر $t > 0$ با استفاده از لم‌های ۲.۳.۱ و ۳.۳.۱ فصل یک که در مورد خاصیت NSD می‌باشد آن‌گاه برای هر $u > 0$

$$E\left\{\sum_{i=1}^{n} uX_i\right\} \leq \prod_{i=1}^{n} E\exp\{uX_i\} \qquad (11.3)$$

با قرار دادن $t = \frac{2logn}{\varepsilon}$ و با استفاده از (۱۱.۳) و نامساوی مارکف نتیجه به‌صورت زیر را حاصل می‌شود

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left(\sum_{k=1}^{n} a_{nk}Y'_k > \varepsilon\right) &\leq \sum_{n=1}^{\infty} \frac{E\exp\left\{\sum_{k=1}^{n} a_{nk}tY'_k\right\}}{e^{t\varepsilon}} \\
&\leq \sum_{n=1}^{\infty} \frac{\prod_{k=1}^{n} E\exp\left\{\sum_{k=1}^{n} a_{nk}tY'_k\right\}}{e^{t\varepsilon}} \\
&\leq \sum_{n=1}^{\infty} \exp\left\{-t\varepsilon + n\left(\frac{Ct^2}{n^2}\exp\left(\frac{Ct}{n}\right)\right)\right\} \\
&= \sum_{n=1}^{\infty} \exp\left\{-2\log n + \frac{\frac{4C\log_n^2}{\varepsilon^2}}{n}\exp\left\{\frac{2C\log n}{n\varepsilon}\right\}\right\} \\
&= \sum_{n=1}^{\infty} \exp\{-2\log n + C\} \\
&\leq C\sum_{n=1}^{\infty} \exp\{-2logn\} \\
&= C\sum_{n=1}^{\infty} \exp\left(\log n^{-2}\right) \\
&= C\sum_{n=1}^{\infty} n^{-2} < \infty
\end{aligned}$$

با استفاده از لم برل کانتلی نتیجه می‌شود که که نابرابری به سمت صفر میل می‌کند. □

برهان. فرض کنیم $N = [\frac{2}{\nu} + 1]$ همچنین تعریف می‌کنیم

$$\begin{aligned}
X_k^{(1)} &= X_k I\left(X_k > \frac{\varepsilon_k k^{\mu}}{ND}\right) \\
X_{nk}^{(2)} &= X_k I\left(a_{nk}X_k \leq n^{\frac{-\nu}{2q}}\right) + a_{nk}^{(-1)} n^{\frac{-\nu}{2q}} I\left(a_{nk}X_k > n^{\frac{-\nu}{2q}}\right) \\
X_{nk}^{(3)} &= X_k - X_k^{(1)} - X_{nk}^{(2)} \\
T_n^{(1)} &= \sum_{k=1}^{\infty} a_{nk}X_k^{(1)} \\
T_n^{(i)} &= \sum_{k=1}^{\infty} a_{nk}X_{nk}^{(i)} \qquad i = 2,3
\end{aligned}$$

ملاحظه می‌کنید که حاصل مجموع T_n به صورت زیر حاصل می‌شود

$$T_n = T_n^{(1)} + T_n^{(2)} + T_n^{(3)}$$

$$\begin{aligned} T_n = T_n^{(1)} + T_n^{(2)} + T_n^{(3)} &= \sum_{k=1}^{\infty} a_{nk} X_k^{(1)} + \sum_{k=1}^{\infty} a_{nk} X_{nk}^{(2)} + \sum_{k=1}^{\infty} a_{nk} X_{nk}^{(3)} \\ &= \sum_{k=1}^{\infty} a_{nk} X_k^{(1)} + \sum_{k=1}^{\infty} a_{nk} X_{nk}^{(2)} + \sum_{k=1}^{\infty} a_{nk} \left(X_k - X_{nk}^{(1)} - X_{nk}^{(2)} \right) \end{aligned}$$

که بعد از ساده کردن نتیجه به صورت زیر حاصل می‌شود

$$T_n = \sum_{k=1}^{\infty} a_{nk} X_k$$

می‌توان نشان داد که $\sum_{k=1}^{\infty} P\left(| X |> \frac{k^{\mu}\varepsilon}{ND}\right) < \infty$ برای هر $\varepsilon > 0$ و $E \mid X \mid^{\frac{1}{\mu}} < \infty$ بنابراین ε وابسته به k وجود دارد که با وقتی $\varepsilon_k \downarrow 0$، مجموعه‌ی احتمال $\sum_{k=1}^{\infty} P\left(| X_k |> \frac{k^{\mu}\varepsilon_k}{ND}\right)$ متناهی می‌شود که با استفاده از لم برل کانتلی آن‌گاه

$$P\left(| X_k |> \frac{k^{\mu}\varepsilon_k}{ND}, \quad i.o.\right) = 0$$

با توجه به این‌که رابطه‌ی بالا همیشه رخ نمی‌دهد مگر در تعداد خیلی متناهی رخ دهد که رابطه‌ی زیر با بالا متناظر است پس در تعداد انگشت شماری رخ می‌دهد یعنی برای k خیلی محدودی رخ می‌دهد، بنابراین برای هر $t > 0$

$$\sum_{k=1}^{\infty} | X_k |^t \, I\left(| X_k |> \frac{k^{\mu}\varepsilon_k}{ND}\right) < \infty \quad a.s. \tag{۱۲.۳}$$

سپس با استفاده از نامساوی کوشی شوارتز و به ازای $q \in (0,2)$ متناهی است بنابراین ضریب C را طوری انتخاب می‌کنیم که رابطه‌ی زیر برقرار باشد

$$\sum_{k=1}^{\infty} a_{nk}^2 \leq C \sum_{k=1}^{\infty} | a_{nk} |^q$$

آن‌گاه به ازای هر t مثبتی رابطه‌ی (۱۲.۳)متناهی است، آن‌گاه نامساوی آخر به سمت صفر میل می‌کند.

$$\begin{aligned} 0 \leq T_n^{(1)} = \sum_{k=1}^{\infty} a_{nk} X_k^{(1)} &\leq \left(\sum_{k=1}^{\infty} a_{nk}^2\right)^{\frac{1}{2}} \left(\sum_{k=1}^{\infty} X_k^2 I\left(X_k > \frac{k^{\mu}\varepsilon_k}{ND}\right)\right)^{\frac{1}{2}} \\ &\leq C\left(\sum_{k=1}^{\infty} | a_{nk} |^q\right)^{\frac{1}{2}} \left(\sum_{k=1}^{\infty} X_k^2 I\left(X_k > \frac{k^{\mu}\varepsilon_k}{ND}\right)\right)^{\frac{1}{2}} \\ &\leq C n^{\frac{-\nu}{2}} \left(\sum_{k=1}^{\infty} X_k^2 I\left(X_k > \frac{k^{\mu}\varepsilon_k}{ND}\right)\right)^{\frac{1}{2}} \longrightarrow 0 \quad a.s. \end{aligned} \tag{۱۳.۳}$$

با در نظر گرفتن Y_{nk} برابر $n^{\frac{\nu}{2q}}a_{nk}X_{nk}^{2}$ داریم

$$Y_{nk}=n^{\frac{\nu}{2q}}a_{nk}X_kI\left\{a_{nk}\leq n^{-\frac{\nu}{2q}}\right\}+I\left\{a_{nk}X_k>n^{-\frac{\nu}{2q}}\right\}$$

با توجه به این‌که Y_{nk} یک تابع صعودی از X_k است، لم ۲.۳.۱ و برای هر ثابت $n\geq 1$، $\{Y_{nk},k\geq 1\}$ یک دنباله از متغیرهای تصادفی NSD است، چون X_k، NSD و اگر $f_1,f_2,\ldots,f_n$ نانزولی باشد، آن‌گاه $f_1(X_1),f_2(X_2),\ldots,f_n(X_n)$، NSD است بنابراین Y_{nk}، NSD است. با استفاده از نابرابری $e^y\leq 1+2\mid y\mid^q$ برای هر $y\leq 1$ و $0<q\leq 1$ نتیجه می‌شود

$$E\exp\left(Y_{nk}\right)\leq 1+2E\mid Y_{nk}\mid^q\leq\exp\left(2E\mid Y_{nk}\mid^q\right).$$

توجه داشته باشید $Y_{nk}\leq 1$ و $EY_{nk}\leq 1$ وقتی $1<q\leq 2$ که براساس این نامساوی $e^y\leq 1+y+\mid y\mid^q$ برای $y\leq 1$ و $1<q\leq 2$ نتیجه می‌شود که

$$\begin{aligned}
E\exp\left(Y_{nk}\right)&\leq 1+EY_{nk}+E\mid Y_{nk}\mid^q\\
&\leq 1+1+E\mid Y_{nk}\mid^q\\
&\leq 1+\exp\left\{E|Y_{nk}|^q\right\}\\
&\leq 2\exp\left(E\mid Y_{nk}\mid^q\right)\\
&\leq 2\exp\left(2E\mid Y_{nk}\mid^q\right).
\end{aligned}$$

بنابراین برای $0<q\leq 2$ داریم

$$E\exp\left(Y_{nk}\right)\leq 2\exp\left(2E\mid Y_{nk}\mid^q\right).\tag{۱۴.۳}$$

با توجه به اینکه $X_{nk}^{(2)}\leq X_k$، همچنین $E\mid X_{nk}^{(2)}\mid^q\leq E\mid X_k\mid^q\leq CE\mid X\mid^q$ بنابراین با استفاده از شرایط لم ۱.۱.۱ که دنباله‌ای نامنفی از متغیرهای تصادفی می‌باشد، exp با هر توانی نامنفی می‌باشد سپس از لم ۱.۱.۱ و (۱۱.۳) که درمورد خاصیت NSD می‌باشد، همچنین (۱۴.۳) داریم آن‌گاه نتیجه به‌صورت زیر حاصل می‌شود

$$\begin{aligned}
E\exp\left(n^{\frac{\nu}{2q}}T_n^{(2)}\right)&=E\exp\left(n^{\frac{\nu}{2q}}\sum_{k=1}^{\infty}a_{nk}X_{nk}^{(2)}\right)\\
&=E\liminf\exp\left(\sum_{k=1}^{m}Y_{nk}\right)\\
&\leq\liminf_{m\longrightarrow\infty}\prod_{k=1}^{m}E\exp\left(Y_{nk}\right)\\
&\leq\liminf_{m\longrightarrow\infty}2\exp\left(2Cn^{\frac{\nu}{2}}\sum_{k=1}^{m}\mid a_{nk}\mid^q E\mid X\mid^q\right)\\
&\leq 2\exp\left(Cn^{\frac{\nu}{2}}.n^{-\nu}\right)\leq C.
\end{aligned}$$

با توجه به این‌که $\sum_{n=1}^{\infty} P\left(T_n^{(2)} \geq \varepsilon\right) \leq C\sum_{n=1}^{\infty} \exp\left(-n^{\frac{\nu}{2q}}\varepsilon\right)$ چون سری متناهی است زیرا $\alpha > 0$ $\sum_{n=1}^{\infty} \frac{1}{\exp^{n^{\alpha}}}$ یک سری همگراست پس در نتیجه $C\sum_{n=1}^{\infty} \frac{1}{\exp^{n^{\alpha}}}$ یک سری همگرا و متناهی است آن‌گاه به‌ازای هر $\varepsilon > 0$ رابطه‌ی زیر برقرار است

$$\sum_{n=1}^{\infty} P\left(T_n^{(2)} \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} \exp\left(-n^{\frac{\nu}{2q}}\varepsilon\right) E\exp\left(n^{\frac{\nu}{2q}}T_n^{(2)}\right) \leq \infty$$

با استفاده از لم برل کانتلی نتیجه می‌شود برای هر $\varepsilon > 0$

$$P\left(\limsup_{n\longrightarrow\infty} T_n^{(2)} > \varepsilon\right) \leq P\left(T_n^{(2)} \geq \varepsilon, i.o.\right) = 0$$

بنابراین

$$\limsup_{n\longrightarrow\infty} T_n^{(2)} \leq 0 \quad a.s. \tag{۱۵.۳}$$

یک عدد صحیح ثابت k را انتخاب می‌کنیم به طوری که $\varepsilon_k < \frac{\varepsilon}{2}$ برای هر $k > K$ از این شرط $\sum_{k=1}^{\infty} \mid a_{nk} \mid^q \leq Dn^{-\nu}$، n_0 وجود دارد به‌ازای $n \geq n_0$ آن‌گاه

$$\sum_{k=1}^{K} a_{nk}X_{nk}^{(3)} \leq \left(Dn^{-\nu}\right)^{\frac{1}{q}} \sum_{k=1}^{K} \frac{\varepsilon_k}{ND}k^{\mu} < \frac{\varepsilon}{2}$$

برای هر $n > n_0$ فرض کنید

$$D_n = \left\{k|k > K, a_{nk}X_k \geq n^{\frac{-\nu}{2q}}\right\}$$

باید به این نکته توجه کنیم که برای ثابت n، $\sum_{k=1}^{\infty} a_{nk}^q E \mid X_k \mid^q < \infty$ با $0 < q \leq 2$ و $EX_k = 0$ برای $1 < q \leq 2$ با استفاده از لم ۲.۱.۳ و قضیه (۲.۱۲.۲) از استوت (۱۹۷۴)، به‌دست می‌آید که T_nها همگی متناهی و همگرای قریب به یقین هستند در نتیجه از آنجا که $\mid T_n \mid < \infty \quad a.s.$ هستند، بنابراین D_n یک مجموعه متناهی با احتمال یک هستند، تعدادی از عناصر d_n در D_n مشخص می‌کند بنابراین

$$\sum_{k=K+1}^{\infty} a_{nk}X_{nk}^{(3)} \leq \sum_{k\in D_n} Dk^{-\mu}\frac{k^{\mu}\varepsilon_k}{ND} \leq d_n\frac{\varepsilon}{2N}$$

بنابراین $\sum_{k=K+1}^{\infty} a_{nk}X_{nk}^{(3)} \geq \frac{\varepsilon}{2}$ اشاره دارد که $d_n \geq N$ ازین‌رو با توجه به این رابطه $P\left(X+Y > \varepsilon\right) \leq P\left(X > \frac{\varepsilon}{2}\right) + P\left(Y > \frac{\varepsilon}{2}\right)$ و d_n متناهی هستند، و معادل این $P\left(Y > \frac{\varepsilon}{2}\right)$، d_n نامتناهی هستند پس ما قسمت اول نامساوی را در نظر می‌گیریم که نتیجه به صورت زیر حاصل می‌شود

$$P\left(T_n^{(3)} \geq \varepsilon\right) \leq P\left(\sum_{k}^{K} a_{nk}X_{nk}^{(3)} \geq \frac{\varepsilon}{2}\right)$$

سپس تعدا متناهی N وجود دارد باتوجه به خاصیت NSD و احتمال توام کمتر مساوی حاصلضرب همچنین با توجه به رابطه‌ی $P(X_k > \varepsilon) \leq CP(X > \varepsilon)$ و نامساوی مارکف و متناهی بودن عبارت $\sum_{n=1}^{\infty} a_{nk}^q \leq Dn^{-\nu}$ آن‌گاه نتایج به صورت زیر حاصل می‌شود

$$\begin{aligned}
P\left(T_n^{(3)} \geq \varepsilon\right) &\leq P\left(\sum_{k}^{K} a_{nk} X_{nk}^{(3)} \geq \frac{\varepsilon}{2}\right) \\
&\leq P\left\{k > K, X_k > a_{nk}^{-1} n^{\frac{-\nu}{2q}}\right\} \\
&\leq \sum_{1 \leq k_1 < k_2 < \ldots < k_N \leq n} P\left(X_{k_1} > a_{nk_1}^{-1} n^{\frac{-\nu}{2q}}, \ldots, X_{k_N} > a_{nk_N}^{-1} n^{\frac{-\nu}{2q}}\right) \\
&\leq \sum_{1 \leq k_1 < k_2 < \ldots < k_N \leq n} \prod_{i=1}^{N} P\left(\mid X_{k_i} \mid > a_{nk_i}^{-1} n^{\frac{-\nu}{2q}}\right) \\
&\leq \left(C \sum_{k=1}^{n} P\left(\mid X \mid > a_{nk}^{-1} n^{\frac{-\nu}{2q}}\right)\right)^N \\
&\leq C\left(\sum_{k=1}^{n} a_{nk}^q n^{\frac{\nu}{2}} E \mid X \mid^q\right)^N
\end{aligned}$$

از متناهی بودن رابطه‌ی $\sum_{n=1}^{\infty} a_{nk}^q \leq Dn^{-\nu}$ نتیجه می‌شود

$$\sum_{n=1}^{\infty} a_{nk}^q n^{\frac{\nu}{2}} \leq Dn^{-\nu} n^{\frac{\nu}{2}} = Dn^{\frac{-\nu}{2}} \leq Cn^{\frac{-N\nu}{2}}.$$

با انتخاب N، $\frac{N\nu}{2}$ بنابراین برای هر $\varepsilon > 0$، آن‌گاه

$$\sum_{n=1}^{\infty} P\left(T_n^{(3)} \geq \varepsilon\right) < \infty$$

با توجه به لم برل کانتلی آن‌گاه نتیجه به صورت زیر حاصل می‌شود

$$\sum_{n=1}^{\infty} P\left(T_n^{(3)} \geq \varepsilon\right) = 0$$

□

اثبات نتیجه‌ی اصلی

براساس لم‌ها و قضیه‌هایی که تا اینجا داشتیم سراغ اثبات نتیجه‌ی اصلی می‌رویم

برهان. قضیه (۳.۱.۳) فرض کنید $\mathbf{x}_{ni} = S_n^{-\frac{1}{2}} \mathbf{x}_i$ و $\boldsymbol{\beta}_n = S_n^{\frac{1}{2}} \boldsymbol{\beta}$ و اینکه $\hat{\boldsymbol{\beta}}_n^* = S_n^{\frac{1}{2}} \hat{\boldsymbol{\beta}}_n$ آن‌گاه

$$\sum_{i=1}^{n} \mathbf{x}_{ni} \mathbf{x}_{ni}^{\top} = I_p, \quad \sum_{i=1}^{n} \|\mathbf{x}_{ni}\|^2 = p, \quad d_n = \max_{1 \leq i \leq n} \|\mathbf{x}_{ni}\|^2 \tag{۱۶.۳}$$

که I_p یک ماتریس همانی $p\times p$ و $\|.\|$ یک اندازه اقلیدوسی روی $\mathbb{R}^p$ می‌باشد، که در این صورت می‌توانیم مدل (۱.۳) را به فرم زیر بازنویسی کنیم

$$Y_i=\mathbf{x}_{ni}^{\top}\boldsymbol{\beta}_n+\mathbf{e}_i, \quad i=1,2,\ldots,n \tag{۱۷.۳}$$

آن‌گاه مجموع باقیمانده‌ها به صورت زیر می‌باشد

$$\sum_{i=1}^{n}\rho\left(Y_i-\mathbf{x}_{ni}^{\top}\hat{\boldsymbol{\beta}}_n^{*}\right)=\min_{\beta}\sum_{i=1}^{n}\rho\left(Y_i-\mathbf{x}_{ni}^{\top}\hat{\boldsymbol{\beta}}\right) \tag{۱۸.۳}$$

فرض کنید در مدل (۱.۳) $\boldsymbol{\beta}=0$ و در مدل (۱۷.۳)، $\boldsymbol{\beta}_n=0$ و اینکه u گوی واحد، $\{\boldsymbol{\beta}:\boldsymbol{\beta}\in\mathbb{R}^p,\|\boldsymbol{\beta}\|=1\}$، که این پیروی می‌کند از شرط $A3$ که $d_n\leq C_4n^{-\delta}$ فرض کنید برای هر ثابت $\varepsilon>0$ داده شده باشد، بدون از دست دادن کلیت مساله می‌توان فرض کرد $\Delta>2C_3\varepsilon$، تعریف می‌کنیم قبل از اینکه تعریف اصلی را بنویسیم احتیاج است به نکات زیر توجه کنیم

$$\begin{aligned}\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma}\left\{\psi\left(e_i+t\right)-\psi\left(e_i\right)\right\}\mathrm{d}t&=\sum_{i=1}^{n}\left\{\rho\left(e_i+\omega_{ni}^{\top}\gamma\right)-\rho\left(e_i\right)\right\}-\sum_{i=1}^{n}\psi\left(e_i\right)\omega_{ni}^{\top}\gamma\\&=\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma}\left\{\psi\left(e_i+t\right)-\psi\left(e_i\right)\right\}\mathrm{d}t+\sum_{i=1}^{n}\psi\left(e_i\right)\omega_{ni}^{\top}\gamma\end{aligned}$$

آن‌گاه داریم

$$\begin{aligned}D_n\left(\boldsymbol{\beta}\right)&=\sum_{i=1}^{n}\left\{\rho\left(e_i-\mathbf{x}_{ni}^{\top}\boldsymbol{\beta}\right)-\rho\left(e_i\right)\right\}, \quad \boldsymbol{\beta}\in\mathbb{R}^p\\D_n\left(\varepsilon n^{\frac{\delta}{2}}\gamma\right)&=\sum_{i=1}^{n}\left\{\rho\left(e_i-\varepsilon n^{\frac{\delta}{2}}\mathbf{x}_{ni}^{\top}\gamma\right)-\rho\left(e_i\right)\right\}\\&=\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma}\left\{\psi\left(e_i+t\right)-\psi\left(e_i\right)\right\}\mathrm{d}t+\sum_{i=1}^{n}\omega_{ni}^{\top}\gamma\psi\left(e_i\right)\\&=I_{1n}\left(\gamma\right)+I_{2n}\left(\gamma\right)\end{aligned}$$

جایی که $\omega_{ni}=-\varepsilon n^{\frac{\delta}{2}}\mathbf{x}_{ni}$ و $\gamma\in U$ همچنین با توجه به افراز $D_n\left(\varepsilon n^{\frac{\delta}{2}}\gamma\right)$ به مجموع دو تابع $I_{1n}\left(\gamma\right)$ و $I_{2n}\left(\gamma\right)$ و اینکه $\sup -I_{2n}\left(\gamma\right)\leq\sup\mid -I_{2n}\left(\gamma\right)\mid$ آن‌گاه

$$\begin{aligned}\inf_{\gamma\in U}D_n\left(\varepsilon n^{\frac{\delta}{2}}\gamma\right)&\geq\inf_{\gamma\in U}I_{1n}\left(\gamma\right)+\inf_{\gamma\in U}I_{2n}\left(\gamma\right)\\&=\inf_{\gamma\in U}I_{1n}\left(\gamma\right)-\sup_{\gamma\in U}-I_{2n}\left(\gamma\right)\\&\geq\inf_{\gamma\in U}I_{1n}\left(\gamma\right)-\sup\mid I_{(2n)}\left(\gamma\right)\mid\end{aligned} \tag{۱۹.۳}$$

U را به N بخش تقسیم می‌کنیم، $U_1,U_2,\ldots,U_N$ بنابراین قطر هر بخش کم می‌شود، سپس n^{-2} و $N\leq\left(2n^2+1\right)^p$ قرار می‌گیرد، همچنین فرض کنید T_j کوچکترین مجموعه محدب بسته U_j باشند که برای ثابت T_j سه مورد ذکر شده در یک همسایگی بسته تابع مقادیر sup و

inf خود را اختیار می‌کند، و قسمت اول نامساوی که درباره‌ی inf است چون بزرگتر مساوی صفر قرار می‌گیرد بنابراین دارای inf می‌باشد و برای نامساوی بعدی هم inf تبدیل شده sup که بصورت زیر می‌باشد

- برای هر $\gamma \in T_j$ اگر $\omega_{ni}^{\top}\gamma \geq 0$ وجود دارد $\gamma_{ij} \in T_j$ طوری‌که
$$\omega_{ni}^{\top}\gamma_{ij} = \inf\left\{\omega_{ni}^{\top}\gamma : \gamma \in T_j\right\}$$
- برای هر $\gamma \in T_j$ اگر $\omega_{ni}^{\top}\gamma \leq 0$ وجود دارد $\gamma_{ij} \in T_j$ طوری‌که
$$\omega_{ni}^{\top}\gamma_{ij} = \sup\left\{\omega_{ni}^{\top}\gamma : \gamma \in T_j\right\}$$
- اگر $\omega_{ni}^{\top}\gamma_1 > 0$ برای $\gamma_1 \in T_j$ و $\omega_{ni}^{\top}\gamma_2 < 0$ و برای $\gamma_2 \in T_j$ بنابراین وجود دارد $\gamma_{ij} \in T_j$ طوری‌که

$$\surd\gamma_1 \in T_j \quad \exists \quad \omega_{ni}^{\top}\gamma_1 > 0 \Longrightarrow \inf\left\{\omega_{ni}^{\top}\gamma_1, \gamma_1 \in T_j\right\} = \omega_{ni}^{\top}\gamma_{ij}$$
$$\surd\gamma_2 \in T_j \quad \exists \quad \omega_{ni}^{\top}\gamma_2 < 0 \Longrightarrow \sup\left\{\omega_{ni}^{\top}\gamma_2, \gamma_2 \in T_j\right\} = \omega_{ni}^{\top}\gamma_{ij}$$

که با تبدیل کردن sup به inf برابر $-\omega_{ni}^{\top}\gamma_{ij}$ هم‌چنین از حاصل $\sup + \inf$ نتیجه به صورت زیر حاصل می‌شود
$$\omega_{ni}^{\top}\gamma_{ij} = 0$$

بنابراین تعریف می‌کنیم
$$G(t) = E\psi(e_i + t), \quad \Psi_i(t) = \psi(e_i + t) - \psi(e_i) - G(t), \quad t \in \mathbb{R}$$

با در نظر گرفتن یکنوایی ψ و اینکه $1 \leq j \leq N$ و T_j کوچکترین مجموعه‌ی محدب بسته‌ی U می‌باشد و اجتماع T_j که $U_1, U_2, \ldots, U_N$ می‌باشد، پوشش بزرگتری برای U می‌باشد که در این‌صورت $U \subseteq \bigcup_j T_j$ یا $U \subseteq T_j$ هم‌چنین از طرفین inf بگیریم نتایج به صورت زیر حاصل می‌شود

$$\begin{aligned}
\inf_{\gamma\in U} I_{1n}(\gamma) &\geq \inf_{1\leq j\leq N}\inf_{\gamma\in T_j} I_{1n}(\gamma)\\
&\geq \inf_{1\leq j\leq N}\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}}\left\{\psi(e_i+t)-\psi(e_i)\right\}\mathrm{d}t\\
&= \inf_{1\leq j\leq N}\sum_{i=1}^{n}\int_0^{\omega'_{ni}\gamma_{ij}} G(t)\,\mathrm{d}t\left\{\frac{\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t + \sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}}\Psi_i(t)\,\mathrm{d}t}{\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}} G(t)}\right\}\\
&= \inf_{1\leq j\leq N}\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t\left\{1+\frac{\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}}\Psi_i(t)\,\mathrm{d}t}{\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t}\right\}\\
&\geq \inf_{1\leq j\leq N}\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t\left\{1-\frac{\left|\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}}\Psi_i(t)\,\mathrm{d}t\right|}{\sum_{i=1}^{n}\int_0^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t}\right\}
\end{aligned} \tag{۲۰.۳}$$

که در آخرین نامساوی(۲۰.۳) از ژائو (۲۰۰۰،p.۱۴۲۵) است، فرض کنید $\gamma \in U_j$ و $\gamma_{ij} \in T_j$ باشد، با استفاده از(۱۶.۳) و تعاریفی که از U_j و T_j شده است برای n به اندازه کافی بزرگ داریم

$$\begin{aligned}
\sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma\right)^{2}-\sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma_{ij}\right)^{2} &\leq \left|\sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma_{ij}\right)^{2}-\sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma\right)^{2}\right| \\
&= \left|\sum_{i=1}^{n}\gamma_{ij}^{\top}\mathbf{x}_{ni}\mathbf{x}_{ni}^{\top}\gamma_{ij}-\sum_{i=1}^{n}\gamma^{\top}\mathbf{x}_{ni}\mathbf{x}_{ni}^{\top}\gamma\right| \\
&= \left|\sum_{i=1}^{n}\left(\gamma_{ij}-\gamma\right)^{\top}\mathbf{x}_{ni}\mathbf{x}_{ni}^{\top}\left(\gamma_{ij}+\gamma\right)\right| \\
&\leq \sum_{i=1}^{n}\|\gamma_{ij}-\gamma\|\|\mathbf{x}_{ni}\|^{2}\left(\|\gamma_{ij}-\gamma\|+2\|\gamma\|\right) \\
&\leq \sum_{i=1}^{n}n^{-2}\left(n^{-2}+2\right)\|\mathbf{x}_{ni}\|^{2} \\
&\leq 3pn^{-2} \\
&< \frac{1}{2}.
\end{aligned} \tag{۲۱.۳}$$

سپس برای $1 < j \leq N$ و nهای به اندازه‌ی کافی بزرگ بدست می‌آوریم

$$\begin{aligned}
\sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma\right)^{2}-\sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma_{ij}\right)^{2} &\leq \frac{1}{2} \\
\sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma_{ij}\right)^{2} &\geq \sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma\right)^{2}-\frac{1}{2} \\
&= \gamma^{\top}\left(\sum_{i=1}^{n}\mathbf{x}_{ni}\mathbf{x}_{ni}^{\top}\right)\gamma-\frac{1}{2} \\
&= \|\gamma\|-\frac{1}{2} \\
&= 1-\frac{1}{2} \\
&= \frac{1}{2}
\end{aligned} \tag{۲۲.۳}$$

با استفاده از (۱۳.۳) و انتخاب ε با توجه به اینکه $d_n = \|\mathbf{x}_{ni}\|^{2}$ و $\mathbf{x}_{ni} = S_n^{\frac{-1}{2}}\mathbf{x}_i$ آن‌گاه

$$d_n = \mathbf{x}_{ni}^{\top}S_n^{\frac{-1}{2}}\mathbf{x}_i = \mathbf{x}_i^{\top}\left(S_n^{\frac{-1}{2}}\right)^{\top}S_n^{\frac{-1}{2}}\mathbf{x}_i$$

چون که $d_n = o\left(n^{-\delta}\right)$ و $\mathbf{x}_{ni}^{\top}$ از مرتبه $o\left(n^{\frac{-\delta}{2}}\right)$ می‌باشد بنابراین $\mathbf{x}_{ni}^{\top} = o\left(n^{\frac{-\delta}{2}}\right) \leq C_3$ که رابطه به‌صورت زیر قرار می‌گیرد

$$\mathbf{x}_i^{\top}S_n^{-1}\mathbf{x}_i \leq d_n = o\left(n^{-\delta}\right)$$

مقدمات بالا برای برقراری نامساوی زیر لازم است

$$|\omega_{ni}^{\top}\gamma_{ij}| = |\varepsilon n^{\frac{\delta}{2}}\mathbf{x}_{ni}^{\top}\gamma_{ij}| \leq C_3\varepsilon\|\gamma_{ij}\| \leq C_3\varepsilon\left(1+n^{-2}\right) < 2C_3\varepsilon < \Delta \tag{۲۳.۳}$$

برای n های به اندازه کافی بزرگ و برای $i = ۱,۲,\ldots,n$ همچنین $j = ۱,۲,\ldots,N$ با شرط $A۲$ و (۲۲.۳) نتیجه به صورت زیر می‌باشد

$$\begin{aligned}\inf_{۱\leq j\leq N}\sum_{i=۱}^{n}\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t &\geq \inf_{۱\leq j\leq N} C_{۱}\sum_{i=۱}^{n}\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}} t\mathrm{d}t\\ &= \inf_{۱\leq j\leq N}\frac{\mathrm{C}_{۱}\varepsilon^{۲}n^{\delta}}{۲}\sum_{i=۱}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma_{ij}\right)^{۲}\\ &\geq \frac{\mathrm{C}_{۱}\varepsilon^{۲}n^{\delta}}{۴}\end{aligned}$$

مجموعه $Y_{ni}^{(j)} = \int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}}\Psi_{i}(t)$ برای $۱ \leq j \leq N$ آن‌گاه تعریف می‌کنیم

$$\begin{aligned}|Y_{ni}^{(j)}| &= \left|\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}}\left(\psi(e_i+t)-\psi(e_i)-G(t)\right)\mathrm{d}t\right|\\ &= \left|\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}}\left(\psi(e_i+t)-\psi(e_i)-E\psi(e_i+t)\right)\mathrm{d}t\right|\end{aligned} \qquad (۲۴.۳)$$

همچنین با توجه به (۲۴.۳) و رابطه $\int(X-E(X))\,\mathrm{d}t$ که اگر $X < C_{۰}$ آن‌گاه $X - E(X) < ۲C_{۰}$ می‌باشد و با توجه به کران‌هایی که دارند نامساوی‌هایی که در ادامه به بحث در درباره آن‌هایی می‌پردازیم برقرار است

- اگر $\psi(e_i+t)-\psi(e_i)$ در e_i یکنوا باشد، و همچنین $Y_{ni}^{(j)}$ در e_i برای $۱ \leq j \leq N$ نیز یکنوا باشد. بنابراین با شرط $A۱$ و (۲۳.۳) یکی نتیجه به صورت زیر می‌باشد

$$|\mathrm{Y}_{ni}^{(j)}| \leq ۲\mathrm{C}_{۰}|\omega_{ni}^{\top}\gamma_{ij}| < ۲C_{۰}۲C_{۳}\varepsilon = ۴\mathrm{C}_{۰}\mathrm{C}_{۳}\varepsilon = \mathrm{C}_{۴}$$

برای $۱ \leq j \leq N$ با توجه به قضیه‌ی ۹.۲ گات و شرط $A۱$ چون عبارت نامنفی و انتگرال‌پذیر است نامساوی‌ها را به صورت زیر تعریف می‌کنیم

$$\begin{aligned}\sum_{i=۱}^{n} var\left(Y_{ni}^{(j)}\right) &\leq \sum_{i=۱}^{n} E\left\{\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}}\left(\psi(e_i+t)-\psi(e_i)\right)\mathrm{d}t\right\}^{۲}\\ &= \sum_{i=۱}^{n} E\left[\left\{\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}}\left(\psi(e_i+t)-\psi(e_i)\right)\mathrm{d}t\right\}\left\{\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}}\left(\psi(e_i+t)-\psi(e_i)\right)\mathrm{d}t\right\}\right]\\ &\leq \mathrm{C}_{۰}\sum_{i=۱}^{n}|\omega_{ni}^{\top}\gamma_{ij}|\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}} E\psi(e_i+t)-E\psi(e_i)\,\mathrm{d}t\\ &= \mathrm{C}_{۰}\sum_{i=۱}^{n}|\omega_{ni}^{\top}\gamma_{ij}|\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t\\ &\leq \frac{\mathrm{C}_{۴}}{۲}\sum_{i=۱}^{n}\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t\end{aligned}$$

و پیشامد A_n را بصورت زیر تعریف می‌کنیم

$$A_n = \left\{\sup_{۱\leq j\leq N}\frac{\left|\sum_{i=۱}^{n}\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}}\Psi_i(t)\,\mathrm{d}t\right|}{\sum_{i=۱}^{n}\int_{۰}^{\omega_{ni}^{\top}\gamma_{ij}} G(t)\,\mathrm{d}t} \geq \frac{۱}{۲}\right\}$$

از $N \le \left(2n^{2}+1\right)^{p}$ و لم ۱.۱.۳ نتیجه می‌شود

$$
\begin{aligned}
P\left(A_n\right) &\le \sum_{j=1}^{N} P\left(\left|\sum_{i=1}^{n} Y_{ni}^{(j)}\right| \ge \frac{1}{2}\sum_{i=1}^{n}\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}} G\left(t\right)\mathrm{dt}\right)\\
&\le 2\sum_{j=1}^{N}\exp\left\{-\frac{\frac{1}{4}\left(\sum_{i=1}^{n}\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}} G\left(t\right)\mathrm{dt}\right)^{2}}{2\left(2\times\frac{\mathrm{C}_4}{2}\sum_{i=1}^{n}\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}} G\left(t\right)\mathrm{dt}+\mathrm{C}_4\times\frac{1}{2}\sum_{i=1}^{n}\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}} G\left(t\right)\mathrm{dt}\right)}\right\}\\
&= 2\sum_{j=1}^{N}\exp\left\{-\frac{1}{12\mathrm{C}_4}\sum_{i=1}^{n}\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}} G\left(t\right)\mathrm{dt}\right\}\\
&\le 2\left(2n^{2}+1\right)^{p}\exp\left(-\mathrm{C}_5 n^{\delta}\right)
\end{aligned}
\tag{25.3}
$$

- حالت دیگر این است زمانی که $\left|\psi\left(x\right)\right| \le \mathrm{C}_2$ و اینکه

$$
\begin{aligned}
Y_{ni}^{(j)} &= \int_{0}^{\omega_{ni}^{\top}\gamma_{ij}}\left(\psi\left(e_i+t\right)-E\psi\left(e_i+t\right)\right)\mathrm{dt}-\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}}\psi\left(e_i\right)\mathrm{dt}\\
&= Y_{ni1}^{(j)}-Y_{ni2}^{(j)}
\end{aligned}
$$

توجه داشته باشید $Y_{ni1}^{(j)}$ و $Y_{ni2}^{(j)}$ در e_i یکنوا برای $1 \le j \le N$ آن‌گاه مشاهده می‌کنیم

$$
\begin{aligned}
|Y_{nil}^{(j)}| \le 2\mathrm{C}_2|\omega_{ni}^{\top}\gamma_{ij}| &\le 4\mathrm{C}_2\mathrm{C}_3\varepsilon = \mathrm{C}_4^{\top}, \quad l=1,2\\
\sum_{i=1}^{n} var\left(Y_{nil}^{(j)}\right) &\le \sum_{i=1}^{n} 4\mathrm{C}_2^{2}\left(\omega_{ni}^{\top}\gamma_{ij}\right)^{2}\\
&\le 4\mathrm{C}_2^{2}\varepsilon^{2}n^{\delta}\sum_{i=1}^{n}\left(\mathbf{x}_{ni}^{\top}\gamma_{ij}\right)^{2}\\
&\le 4\mathrm{C}_2^{2}\varepsilon^{2}n^{\delta}\left(n^{-2}+1\right)^{2}\\
&\le \mathrm{C}\sum_{i=1}^{n}\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}} G\left(t\right)\mathrm{dt}.
\end{aligned}
$$

به طور مشابه از اثبات (۲۵.۳) می‌توانیم نشان دهیم

$$
\begin{aligned}
P\left(A_n\right) \le &\sum_{j=1}^{N} P\left(\left|\sum_{i=1}^{n} Y_{ni1}^{(j)}\right| \ge \frac{1}{4}\sum_{i=1}^{n}\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}} G\left(t\right)\mathrm{dt}\right)\\
&+\sum_{j=1}^{N} P\left(\left|\sum_{i=1}^{n} Y_{ni2}^{(j)}\right| \ge \frac{1}{4}\sum_{i=1}^{n}\int_{0}^{\omega_{ni}^{\top}\gamma_{ij}} G\left(t\right)\mathrm{dt}\right)\\
&\le 4\left(2n^{2}+1\right)^{p}\exp\left(-\mathrm{C}_5' n^{\delta}\right).
\end{aligned}
$$

بدین ترتیب $\sum_{i=1}^{n} P\left(A_n\right) < \infty$. با استفاده از لم بورل کانتلی نتیجه می‌شود که

$P(A_n, i.o.) = 0$، ازاین‌رو با احتمال یک برای n های به اندازه کافی بزرگ داریم

$$\sup_{1\le j\le N} \frac{\left|\sum_{i=1}^{n} \int_0^{\omega_{ni}^\top \gamma_{ij}} \Psi_i(t)\,dt\right|}{\sum_{i=1}^{n} \int_0^{\omega_{ni}^\top \gamma_{ij}} G(t)\,dt} < \frac{1}{2}$$

با توجه به رابطه‌ی (۲۰.۳) با احتمال یک برای n های به اندازه‌ی کافی بزرگ آن‌گاه

$$\inf_{\gamma\in U} I_{1n}(\gamma) \ge \frac{C_1 \varepsilon^2 n^\delta}{8} \qquad (۲۶.۳)$$

اکنون ما از تعاریف زیر استفاده خواهیم کرد

$$\mathbf{x}_{ni} = \begin{pmatrix} x_{ni1} \\ x_{ni2} \\ \vdots \\ x_{nip} \end{pmatrix} \qquad \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{pmatrix} \qquad a_{ni} = \begin{cases} \frac{x_{nik}}{n^{\frac{\delta}{2}}} & 1 \le i \le n \quad k = 1, 2, \dots, p \\ 0 & i \ge n \end{cases}$$

ترکیب (۱۶.۳) و (۵.۳) و با ثابت $k = 1, 2, \dots, p$، $|x_{nik}| \le \|\mathrm{x}_{ni}\| \le d_n^{\frac{1}{2}} \le \sqrt{C_3}\, n^{\frac{-\delta}{2}}$ و اینکه $\sum_{i=1}^{n} \mathrm{x}_{nik}^2 = 1$ بنابراین می‌بینیم که

$$|a_{ni}| \le \sqrt{C_3}\, n^{-\delta}, \quad n \ge 1 \qquad (۲۷.۳)$$

برای $\delta = 1$ توسط قسمت اول از ۳.۱.۳ و (۲۷.۳) e_iها NSD هستند و اینکه تابع ψ یکنوا می‌باشد پس $\psi(e_i)$ نیز NSD می‌باشد بنابراین شرایط (۹.۳) برقرار است که نتیجه به‌صورت زیر حاصل می‌شود

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathrm{x}_{nik} \psi(e_i) = \sum_{i=1}^{n} a_{ni} \psi(e_i) \longrightarrow 0 \quad a.s.$$

و برای $\delta \le \frac{1}{2}$ نتیجه به صورت زیر حاصل می‌شود

$$\sum_{i=1}^{n} a_{ni}^2 = \frac{1}{n^\delta} \sum_{i=1}^{n} \mathrm{x}_{nik}^2 = n^{-\delta} \qquad (۲۸.۳)$$

برای $\frac{1}{2} < \delta < 1$ و اینکه از نامساوی کوشی شوارتز استفاده می‌کنیم نتیجه به صورت زیرحاصل می‌شود

$$\sum_{i=1}^{n} |a_{ni}|^{\frac{1}{\delta}} = \frac{1}{n^{\frac{1}{2}}} \sum_{i=1}^{n} |x_{nik}|^{\frac{1}{\delta}} \le n^{\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^{n} x_{nik}^2 \right)^{\frac{1}{2\delta}} = n^{-\frac{1}{2\delta} + \frac{1}{2}} \qquad (۲۹.۳)$$

از قسمت دوم لم ۳.۱.۳ شرط C_4 e_iها NSD هستند و اینکه تابع ψ یکنوا می‌باشد پس $\psi(e_i)$ نیز NSD می‌باشد بنابراین شرایط لم (۳.۱.۳) برقرار است، و فرمول (۲۷.۳) تا (۲۹.۳) نتیجه به صورت زیر می‌شود

$$n^{-\frac{\delta}{2}} \sum_{i=1}^{n} x_{nik} \psi(e_i) = \sum_{i=1}^{n} a_{ni} \psi(e_i) \longrightarrow 0 \quad a.s.$$

بنابراین برای $0 < \delta \leq 1$ که در ادامه نامساوی از نامساوی کوشی شوارتز استفاده می‌کنیم آن‌گاه

$$\begin{aligned}
n^{-\delta} \sup_{\gamma\in U} \left| I_{2n}(\gamma) \right| &= n^{-\delta} \sup_{\gamma\in U} \left| \sum_{i=1}^{n} \omega_{ni}^{\top} \gamma \psi(e_i) \right| \\
&= n^{-\delta} \sup_{\gamma\in U} \left| \sum_{i=1}^{n} \varepsilon n^{\frac{\delta}{2}} \mathbf{x}'_{ni} \gamma \psi(e_i) \right| \\
&= \varepsilon n^{-\frac{\delta}{2}} \sup_{\gamma\in U} \left| \sum_{i=1}^{n} \sum_{k=1}^{p} x_{nik} \gamma_k \psi(e_i) \right| \\
&\leq \varepsilon n^{-\frac{\delta}{2}} \sup_{\gamma\in U} \sqrt{\sum_{k=1}^{p} \left(\sum_{i=1}^{n} x_{nik} \psi(e_i) \right)^2} \cdot \sqrt{\sum_{k=1}^{p} \gamma_k^2} \\
&\leq \varepsilon \sqrt{\sum_{k=1}^{p} \left(n^{-\frac{\delta}{2}} \sum_{i=1}^{n} x_{nik} \psi(e_i) \right)^2} \longrightarrow 0 \quad a.s.
\end{aligned}$$

که نشان می‌دهد با احتمال یک برای nهای به اندازه‌ی کافی بزرگ نتیجه به صورت زیر می‌باشد

$$\sup_{\gamma\in U} \left| I_{2n}(\gamma) \right| \leq \frac{C_3 \varepsilon^2 n^{\delta}}{16} \tag{۳۰.۳}$$

با ترکیب (۱۹.۳)، (۲۶.۳) و (۳۰.۳) با احتمال یک برای n های به اندازه‌ی کافی بزرگ نتیجه به صورت زیر است

$$\inf_{\gamma\in U} D_n\left(\varepsilon n^{\frac{\delta}{2}} \gamma \right) \geq \frac{C_3 \varepsilon^2 n^{\delta}}{16}$$

از محدب بودن $D_n(0)$ دانستن $D_n(0) = 0$ همچنین با استفاده از تعریف (۱۸.۳) برای $\hat{\beta}_n^*$ نتیجه‌ی زیر حاصل می‌شود

$$\left\{ \inf_{\gamma\in U} D_n\left(\varepsilon n^{\frac{\delta}{2}} \gamma \right) > 0 \right\} \subset \left\{ \|\hat{\beta}_n^*\| \leq \varepsilon n^{\frac{\delta}{2}} \right\}.$$

در نتیجه برای هر $\varepsilon > 0$ داده شده با احتمال یک برای nهای به اندازه‌ی کافی بزرگ

$$\|\hat{\beta}_n^*\| \leq \varepsilon n^{\frac{\delta}{2}}$$

که نتیجه به صورت زیر حاصل می‌شود

$$n^{-\frac{\delta}{2}} \|\hat{\beta}_n^*\| \longrightarrow 0 \quad a.s. \tag{۳۱.۳}$$

توجه داشته باشید $\xi(A)$ کوچکترین مقدار ویژه ماتریس متناهی A است ما از این نتیجه استفاده می‌کنیم که اگر A و B دو ماتریس متناهی از مرتبه p باشند آن‌گاه

$$\operatorname{tr}(AB) \geq \mu(A)\, \xi(B)$$

جایی که $\mu(A)$ بزرگترین مقدار ویژه از A که با مراجعه به چن و ژائو (۱۹۹۵) ببنید.
اکنون با در نظر گرفتن $M = n_\circ$ و $n > n_\circ$ طوری‌که $S_n \geq S_M > \circ$ آن‌گاه

$$\begin{aligned}
\xi\left(S_M\right)\left(\xi\left(S_n\right)\right)^{-۱} &= \xi\left(S_M\right)\mu\left(S_n^{-۱}\right) \\
&\leq \operatorname{tr}\left(S_n^{-۱}S_M\right) \\
&= \sum_{i=۱}^{M}\mathbf{x}_i'\mathbf{S}_n^{-۱}\mathbf{x}_i \\
&\leq \mathbf{M}d_n \\
&\leq \mathbf{M}\mathrm{C}_۳ n^{-\delta}
\end{aligned}$$

در نتیجه یک ثابت مثبت $C_۶$ وجود دارد طوری‌که $\left(\xi\left(\mathbf{S}_n^{\frac{۱}{۲}}\right)\right)^{-۱} \leq \mathrm{C}_۶ n^{-\frac{\delta}{۲}}$ بنابراین نتیجه زیر بدست می‌آید

$$\|\hat{\beta}_n\| \leq \left(\xi\left(\mathbf{S}_n^{\frac{۱}{۲}}\right)\right)^{-۱}\|\mathbf{S}_n^{\frac{۱}{۲}}\hat{\beta}_n\| \leq \mathrm{C}_۶ n^{-\frac{\delta}{۲}}\|\mathbf{S}_n^{\frac{۱}{۲}}\hat{\beta}_n\| = \mathrm{C}_۶ n^{-\frac{\delta}{۲}}\|\hat{\beta}_n^*\| \longrightarrow \circ \quad a.s$$

□

فصل ۴

مطالعات شبیه‌سازی

شبیه‌سازی یک ابزار قدرتمند برای طراحی و تحلیل سیستم‌ها یا فرایندهای پیچیده مهندسی است که به استفاده کنندگان این امکان را می‌دهد که آزمایش‌هایی را انجام دهند که در عمل غیرممکن یا دارای هزینه‌ی بسیار زیادی می‌باشند.
به عنوان مثال قبل از پرتاب یک ماهواره عملیات پرتاب بارها شبیه‌سازی می‌شود. مثال‌های زیادی را مطرح کرده‌اند بلند شدن یک هواپیما، شلیک یک موشک و... . فرآیند شبیه‌سازی در علم آمار از قرن بیستم مطابق با شروع انتشار رادیو و تلویزیون آغاز گردید:
همانطور که استفاده از این وسایل در زندگی هر کسی معمول و متداول شده است، استفاده از شبیه‌سازی در بسیاری از شاخه‌های آماری گسترش یافته است.
از کاربردهای شبیه سازی در آمار می‌توان به تولید داده‌های تصادفی از توزیع‌های گسسته و پیوسته، تولید اعداد شبه تصادفی، محاسبه انتگرال‌ها و ... اشاره کرد هم‌چنین برای پاسخ به سوالاتی که نظریه آمار برای آن‌ها جواب ندارد می‌توان از شبیه‌سازی استفاده کرد و جواب تجربی به‌دست آورد در این پایان‌نامه برای ارزیابی سازگاری برآوردگرهای معرفی شده با استفاده از یک مدل شبیه‌سازی، مدل رگرسیون خطی زیر را در نظر بگیرید.

$$Y_i = -۱ + ۰/۶\mathbf{X} + e_i, \qquad i = ۱, ۲, \dots, n$$

مطالعات شبیه سازی با نرم افزار R انجام گرفته و از پکیج MASS و mvtnorm استفاده شده است که کدهای مربوط به شبیه سازی در پیوست آمده است. مشاهدات شبیه‌سازی شده براساس مدل رگرسیون خطی و e_iهای NSD هستند. برای این‌که توزیع خطاها NSD باشد از

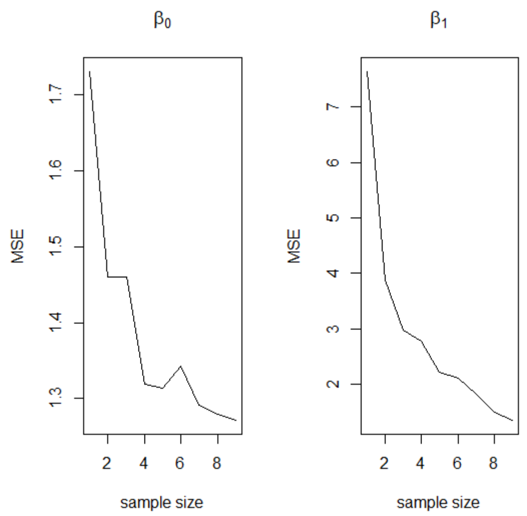
دو توزیع نرمال چندمتغیره و یا تی چندمتغیره استفاده کرده‌ایم. براساس مقاله هو (۲۰۰۰) باید درایه‌های غیر قطر اصلی ماتریس کواریانس توزیع نرمال و ماتریس مقیاس تی منفی باشد در توزیع نرمال و تی چند متغیره ماتریس کواریانس مقیاس در مراحل شبیه سازی به‌صورت زیر می‌باشد

$$\sigma_{ii} = 1 \qquad \sigma_{ij} = -\frac{1}{n-1}.$$

بعد از این‌که داده‌های NSD را تولید کردیم مرحله‌ی بعدی سراغ مدل رگرسیون استوار می‌رویم. همان‌طور که اشاره شد در فصل دو و سه برای مدل رگرسیونی استوار احتیاج به تابع $\rho(x)$ داریم. ما در این شبیه‌سازی تابع هوبر را در نظر گرفته‌ایم. از مدل رگرسیونی استوار در R براساس rlm استفاده می‌کنیم و حجم نمونه‌ها را $n = 20, 40, 60, 80, 100, 150, 200, 500, 1000$ در نظر می‌گیریم. نتیجه شبیه سازی به‌صورت زیر می‌باشد

۱.۴ توزیع نرمال چندمتغیره

با توجه به منحنی برازش شده واضح است که برای حجم نمونه‌های کم MSE[1] زیاد است ولی با افزایش حجم نمونه‌ها MSE کاهش می‌یابد، در نتیجه منحنی برازش شده به مقدار واقعی نزدیک‌تر می‌شود و برآورد قابل قبول‌تری را نسبت به حجم نمونه‌های کم می‌دهد.

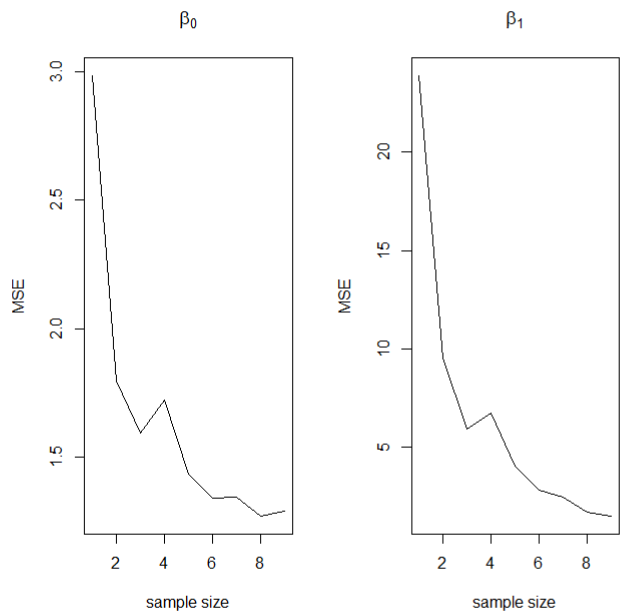


شکل ۱.۴: سازگاری ضرایب با استفاده از نرمال چندمتغیره

[1] Mean Square Error

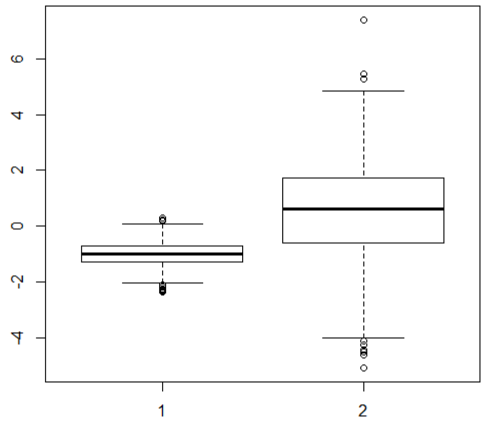
۲.۴ توزیع تی چندمتغیره

همان طور که مشاهده می‌کنیم توزیع تی چندمتغیره دارای MSE نسبتا زیادی می‌باشند ولی باز با افزایش حجم نمونه MSE کاهش می‌یابد و تمام شرایطی که در بالا به آن اشاره کردیم برای توزیع تی نیز برقرار است که سازگاری $\hat{\beta}$ را نشان می‌هد و چون ما دنبال نتیجه ایده‌آل‌تری می‌باشیم پس بهتر است از توزیع نرمال چند متغیره استفاده کنیم زیرا سازگاری بهتری را به ما می‌دهد و منحنی هموارتر است یعنی از یک مرحله‌ای به بعد سازگاری بهتر و مهم‌تر از همه، برازش منحنی بهتر می‌شود، که نمودار آن به‌صورت زیر می‌باشد

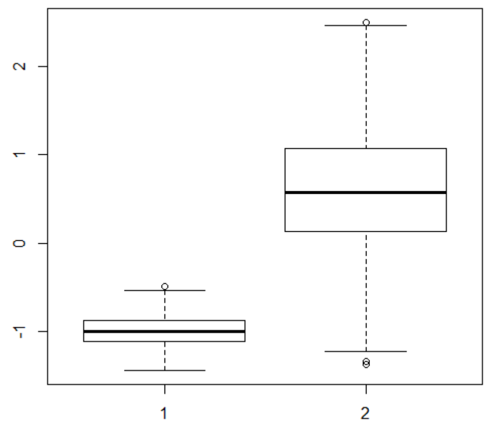


شکل ۲.۴: سازگاری ضرایب با استفاده از چندمتغیره توزیع تی

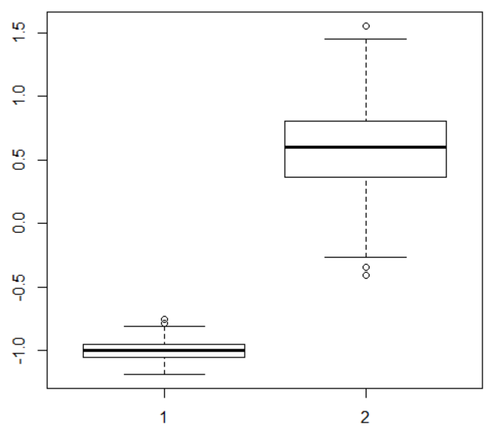
نمودار جعبه‌ای که در زیر آمده برای نمونه‌های $n = ۴۰, ۲۰۰, ۱۰۰۰$ همان‌طور که مشاهده می‌کنید در مدل رگرسیون استوار با افزایش حجم نمونه به مقدار واقعی ضرایب نزدیک می‌شود که داده‌هایی که خارج از فواصل اطمینان هستند نشان دهنده‌ی داده‌ی پرت می‌باشند که یعنی در مدل هیچ تاثیری ندارند



شکل ۳.۴: نمودار باکس‌پلات مربوط به برآورد ضرایب برای حجم نمونه ۴۰



شکل ۴.۴: نمودار باکس‌پلات مربوط به برآورد ضرایب برای حجم نمونه ۲۰۰



شکل ۵.۴: نمودار باکس‌پلات مربوط به برآورد ضرایب برای حجم نمونه ۱۰۰۰

جدول ۱.۴: میانگین مربوط به برآورد ضرایب برای حجم نمونه‌های مختلف

تعداد	$\beta_۰$	$\beta_۱$
$n=$ ۲۰	۰/۹۸۶۰−	۰/۶۱۹۵
$n=$ ۴۰	۰/۹۹۱۹−	۰/۵۲۵۷
$n=$ ۶۰	۱/۰۲۰۱−	۰/۵۹۱۹
$n=$ ۸۰	۰/۹۸۵۰−	۰/۶۰۶۵
$n=$ ۱۰۰	۰/۹۹۱۳−	۰/۵۹۶۶
$n=$ ۱۵۰	۱/۰۰۳۴−	۰/۵۹۰۹
$n=$ ۲۰۰	۰/۹۸۹۷−	۰/۵۷۸۵
$n=$ ۵۰۰	۰/۹۹۶۳−	۰/۶۰۰۸
$n=$ ۱۰۰۰	۰/۹۹۶۹−	۰/۵۹۰۸
$n=$ ۱۰۰۰۰	۰/۹۹۶۰۵−	۰/۵۹۲۵

میانگین برآورد ضرایب در مدل رگرسیون استوار برای حجم نمونه‌های مختلف که در جدول (۱.۴) آورده شده برآورد میانگین ضرایب برای حجم نمونه‌های کم تقریبا برابر ضرایب رگرسیون مدل خطی کمترین توان‌های دوم است ولی هرچه حجم نمونه را افزایش می‌دهیم این برآوردها تقریبا به طور کامل به ضرایب مدل رگرسیونی خطی کمترین توان‌های دوم نزدیک می‌شود، همانطور که مشاهده با افزایش حجم نمونه و تکرار زیاد حلقه تقریبا به مقدار واقعی نزدیک شدن که این همون مفهوم سازگاری را می‌رساند.

۱.۲.۴ بحث و نتیجه‌گیری

- در این پایان نامه یک مدل رگرسیون استوار را فرض کردیم هم‌چنین جمله‌ی خطاها را متغیرهای تصادفی وابسته با ساختار وابستگی NSD می‌باشد که هدف ما در این مدل رگرسیونی استوار نشان دادن سازگاری قوی $\hat{\beta}_n$ بوده که این نتیجه تضمینی برای همگرایی کامل در مدل رگرسیون استوار می‌باشد، که با استفاده از مطالعات شبیه‌سازی این سازگاری را نشان دادیم.

- روش استوار در متون آماری، به معنی معتبر بودن و مناسب بودن آن است در حالت کلی می‌توان حساس نبودن به انحراف کم از پذیره‌ها کم را استواری تعریف کرد. رگرسیون استوار علاوه بر حساس نبودن نسبت به داده‌های پرت زمانی که مشاهدات دارای توزیع نرمال هستند دارای کارایی ۹۰ تا ۹۵ درصدی نسبت به کمترین توان‌های دوم می‌باشد.

- رده وابستگی NSD یک رده‌ی بسیار بزرگ از وابستگی‌هاست که متغیرهای تصادفی NA و مستقل را شامل می‌شود بنابراین نتیجه ارایه شده برای این دو رده از متغیرها برقرار می‌باشد که می‌توان جمله خطا در مدل رگرسیونی با این دو نوع متغیر تصادفی از نظر همگرایی کامل نتیجه گرفت که بررسی همگرایی کامل برآوردگر رگرسیون استوار برای سایر رده‌های وابستگی موضوعی است که در آینده می‌توان از آن استفاده کرد.

مراجع

[۱] اسحاقی ا، (۱۳۹۲)، پایان‌نامه ارشد: ''مدل‌های نیمه پارامتری تحلیل بقا برای داده‌های بازگشتی با روش هسته''، دانشکده ریاضی، دانشگاه صنعتی شاهرود.

[۲] اقبال ن، (۱۳۸۴)، پایان نامه ارشد:'' روش‌های استواری '' دانشکده ریاضی، دانشگاه فردوسی مشهد.

[۳] لاری ع، (۱۳۹۱)، **''رگرسیون خطی کاربردی''** چاپ سوم، انتشارات دانشگاه شیراز.

[۴] مونتگمری د،پک ا، ترجمه سید ابراهیم رضوی پاریزی، (۱۳۸۶)، **''مقدمه‌ای بر تحلیل رگرسیون خطی''** چاپ سوم، انتشارات دانشگاه شهید باهنر کرمان.

[۵] ولیپور پاشاکایی م، (۱۳۹۲)، پایان نامه ارشد:''مدل بندی داده‌های نرخ و نسبت با رگرسیون بتا ''، دانشکده ریاضی، دانشگاه صنعتی شاهرود.

[۶] فریبا ا، حسین ف، الهام خ. نشریه دانشجویی آمار (ندا)ـ سال سیزدهمـ شماره اول، ''روش‌های شناسایی و تحلیل داده‌های پرت در تحلیل رگرسیون''، دانشگاه بیرجند، دانشکده ریاضی، گروه آمار.

[7] Alam, K. and Saxena, K.M.L. (1981). "Positive dependence in multivariate distributions.Positive dependence in multivariate distributions" **Comm Statist Theor Meth.** A10, pp 1183–1196.

[8] Bai, Z.D., Rao, C.R. and Wu, Y. (1992). "M-estimation of multivariate linear regression parameters under a convex discrepancy function" **Statist. Sinica** 2, pp 237–254.

[9] Beaton, A. E. and Tukey, J. W. (1974)."The fitting of power series, meaning polynomials, illustrated on band spectroscopic data,"**Technometrics,**16, pp 147-185.

[10] Berlinet, A., Liese, F. and Vajda, I. (2000). "Necessary and sufficient conditions for consistency of M estimates in regression models with general errors. " **J. Statist. Plann. Infer.** 89, pp 243–267.

[11] Box, G. E.P. (1953). ”Non-Normality and Tests on Variances,” **Biometrika**, 40, pp 318-335.

[12] Chen, X.R. and Wu, Y.H. (1988). ”Strong consistency of M-estimates in linear models.” **J. Multi. Anal.** 27, pp 116–130.

[13] Chen, X.R. and Zhao, L.C. (1995). ”Strong consistency of M-estimates of multiple regression coefficients.” **J. Syst. Sci. Complex.** 8, pp 82–87.

[14] Cheng, C.S. and Li, K.C. (1984). ”The strong consistency of M-estimators in linear models.”**J. Multi. Anal.** 15, pp 91–98.

[15] Christofides, T.C. and Vaggelatou, E. (2004). ”A connection between supermodular ordering and positive/ negative association.”**J. Multi. Anal.** 88, pp 138–151.

[16] Cui, H.J., He, X.M. and Ng, K.W. (2004). ”M-estimation for linear models with spatially-correlated errors.” **Statist. Probab. Lett.** 66, pp 383–393.

[17] Ebrahimi, N. and Ghosh, M. (1981). ”Multivariate negative dependence.” **Commun. Statist. Theor. Meth.** A 10, pp 307–337.

[18] Eghbal, N. Amini, M. and Bozorgnia, A. (2010). ”Some maximal inequalities for quadratic forms of negative superadditive dependence random variables.” **Statist. Probab. Lett.** 80, pp 587–591.

[19] Eghbal, N., Amini, M. and Bozorgnia, A. (2011). ”On the Kolmogorov inequalities for quadratic forms of dependent uniformly bounded random variables.” **Statist. Probab. Lett.** 81, pp 1112–1120.

[20] Daniel, C. and Wood, F. S. (1980). **Fitting Equations to Data,** 2nd ed., Wiley, Newyork

[21] Derper, N.R. and Smith, H. (1998). ”**Aplied regression analysis**”, Johan wiley and Sons

[22] Dixon, W. J. and Tukey, J. W. (1968). ”Approximate Behavior of the Distribution of Winsorized t (Trimming/Winsorization 2),” **Technometrics,** 10, pp 83-98.

[23] Dutter, R. (1977). ”Numerical Solution of Robust Regression Problems: Computational Aspects, a Comparison,” **Journal of Statistical Computation and Simulation,** 5, pp 207-238.

[24] Gao J.T. and Shi, P.D. (1997). ”M-type smoothing spline in nonparametric and semiparametric regression models.” **Statist. Sinica** 7, pp 1155–1169.

[25] Gao, J.T., Li, D.G. and Lin, Z.Y. (2009). ”Robust estimation in parametric time series models under long-and short-range-dependent structures.” **Aust. N. Z. J. Stat.** 51, pp 161–181.

[26] Gut, A. (1994). **Probability: A Graduate Course** , Spring T Stat, Sweden.

[27] Hampel, F. R. (1974). ”The Influence Curve and Its Role in Robust Estimation”, **Journal ofthe American Statistical Association**, 69, pp 383-393.

[28] He, X.M. and Shao, Q.M. (1996). ”A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs.” **Ann. Statist.** 18, pp 2608–2630.

[29] Hill, R. W. and Holland, P. W. (1977). ”Two robust alternatives to least squares regression,”**journal of the American Statisical Association,** 72, pp 828-833.

[30] Hogg, R. E. (1799). ”statistical robustness: One view of its use in applications today,”**The American Statistician.** 33:3, pp 108-115.

[31] Holland, P. W. and Welsch R. E. (1977). ”Robust Regression Using Iteratively Reweighted Least-Squares,”**Communications in Statistics,** A6, pp 813-828.

[32] Hsu, P. and Robbins, H. (1947).”Complete convergence and the law of large numbers,” **Proc Natl Acad Sci USA**, 33, 25–31.

[33] Hu, T.Z. (2000). ”Negatively supperadditive dependence of random variables with applications.” **Chinese J. Appl. Probab. Statist.** 16, pp 133–144.

[34] Huber, P.J. (1981). **Robust Statistics,** New York: John Wiley.

[35] Huber, P. J. (1964), ”Robust Estimation of a Location Parameter.” **Annals of Mathematical Statistics**, 35, pp 73-101.

[36] Huber, P. J. (1972). ”Robust Statistics: A Review,” **Annals of Mathematical Statisticz**, 43, pp 1041-1067.

[37] Huber, P. J. (1973). ”Robust Regression: Asymptotics, Conjectures, and Monte Carlo,” **Annals of Statistics**, 1, pp 799-821.

[38] Jaeckel, L.A. (1972). ”Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals,” **Annals of Mathematical Statistics,** 43, pp 1449-1458.

[39] Jaworski, P., Durante, F. and Hardle, W.K. (2013). ”Copulae in Mathematical and Quantitative Finance.” **Heidelberg: Springer-Verlag.**

[40] Jureckova, J. (1977). ”Asymptotic Relations of M-estimates and R-estimates in Linear Regression Models,” **Annals of Statistics**, 5,pp 464-472.

[41] Kemperman, J.H.B. (1977). ”On the FKG-inequalities for measures on a partially ordered space.” **Proc. Akad. Wetenschappen Ser. A.** 80, pp 313–331.

[42] Koenker, R. and Bassett, G. (1978). ”Regression Quantities,” **Econometrics,** 46, pp 33-50.

[43] Koul, H.L. and Surgailis, D. (2000). ”Second-order behavior of M-estimators in linear regression with long-memory errors.” **J. Statist. Plann. Infer.** 91, pp 399–412.

[44] Mari, D. D. and Kotz, S. (2001). ”Correlation And Dependence”, **Imp Coll Press London.**,

[45] Miao, B.Q. and Wu, Y. (1996). ”Limiting behavior of recursive M-estimators in multivariate linear regression models.” **J. Multi. Anal.** 59, pp 60–80.

[46] Miao, B.Q., Wu, Y. and Liu, D. (2009). ”Limiting behavior of recursive M-estimators in multivariate linear regression models and their asymptotic efficiencies.” **Acta Math. Sci.** 29B, pp 1–11.

[47] Moussa-Hamouda, E. and Leon, F. C. (1977a).”The robustness of efficitency of abjusted trimmed estimators in linear regression.”**Technometrics.**19, pp 19-34.

[48] Moussa-Hamouda, E. and Leon, F. C. (1974). ”The 0-blue estimators for complete and censored samples in liner regression.”**Technometrics.**16, pp 441-446.

[49] Nelsen, R.B. (2006). **An Introduction to Copulas.** New York: Springer Science + Business Media.

[50] Rao, C.R. and Zhao, L. C. (1992a). ”On the consistency of M-estimate in a linear model obtained through an estimating equation.” **Statist. Probab. Lett.** 14, pp 79–84.

[51] Rao, C.R. and Zhao, L. C. (1992b). ”Linear representation of M-estimates in linear models. Canad.” **J. Statist.** 20, pp 359–368.

[52] Shen Y., Wang X.J., Yang W.Z. and Hu S.H. (2013). ”Almost sure convergence theorem and strong stability for weighted sums of NSD random variables.” **Acta. Math. Sinica English Ser.** 29, pp 743–756.

[53] Stout, W.F. (1974). ”Almost Sure Convergence.” **New York: Academic Press.**

[54] Tukey, J. W. (1962). ”The Future of Data Analysis,” **Annals of Mathematical Statistics,** 33, pp 1-67.

[55] Wang, X.and Hu, S. (2015). ”On the strong consistency of M-Estimates in linear models for negatively superadditive dependent erros,” **Australian And New Zealand Journal of Statistics,** 57(2), pp 259-274.

[56] Wang, X.J., Deng, X., Zheng L.L. and Hu S.H. (2014). ”Complete convergence for arrays of rowwise negatively supperadditive-dependent random variables and its applications.” **Statistics** 48, pp 834–850.

[57] Withers, C.S. and Nadarajah, S. (2014). ”Asymptotic properties of M-estimators in linear and nonlinear multivariate regression models.” **Metrika** 77, pp 647–673.

[58] Wu, Q.Y. (2006). ”Strong consistency of M estimator in linear model for negatively associated samples.” **J. Syst. Sci. Complex.** 19, pp 592–600.

[59] Wu, Q.Y. and Jiang Y.Y. (2011). ”The strong consistency of M- estimator in a linear model for negatively dependent random samples,” **Communications in statistics theory and methods.**40, pp 467–491.

[60] Wu, Q.Y. (2007). ”M-estimation of linear models with dependent errors.” **Ann. Statist.** 35, pp 495–521.

[61] Wu, Q.Y. (2011). The strong consistency of M estimator in a linear model for negatively dependent random samples. Commun. Statist. Theor. Meth. 40, 467–491.

[62] Yang, S.C. (2002). ”Strong consistency of M-estimator in linear model.” **Acta Math. Sinica** 45, pp 21–28.(in Chinese).

[63] Yohai, V.J. and Maronna, R.A. (1979). ”Asymptotic behavior of M-estimators for the linear model.” **Ann. Statist.** 7, pp 258–268.

[64] Zhao, L.C. (2000). ”Some contributions to M-estimation in linear models.” **J. Statist. Plann. Infer.** 88, pp 189–203.

[65] Zhao, L.C. (2002). ”Strong consistency of M-estimates in linear models.” **China Ser.** A 45, pp 1420–1427.

پیوست آ

کدهای شبیه سازی فصل دو و چهار

آ.۱ کدهای مربوط به مدل رگرسیون کمترین توان‌های دوم و رگرسیون استوار

```
rm(list=ls())
library(MASS)
library(HoRM)
data <- read.csv("G:/data1.csv")
Y <- c(data$Y)
X0 <- c(rep(1,21))
X1 <- c(data$X1)
X2 <- c(data$X2)
X3 <- c(data$X3)
X <- matrix(c(Y,X1 ,X2 ,X3 ), nrow = 21, ncol = 4)
X
Xls <- matrix(c(X0,X1 ,X2 ,X3 ), nrow = 21, ncol = 4)
LS <- lm(Y ~. , data=data)
```

```
summary(LS)
names(LS)
beta1 <- c(LS$coefficients)
beta01 <- as.matrix(beta1)
Y01 <- as.matrix(Y)
dim(t(Y01))
dim(Y01)
dim(t(beta01))
dim(t(Xls))
SSE_b0 <- t(Y01)%*%Y01 - t(beta01)%*%t(Xls)%*%Y01
MSE_b0 <- (SSE_b0)/(21-3)
MSE_b0
se.ls <- LS$residuals
plote<-qqnorm(se.ls)
qqline(se.ls, col="red",lWd=3)
names(LS)
data_o <- X[-c(1,3,4,21),]
data_o
da <- as.data.frame(data_o)
LS1 <- lm(V1 ~. , data=da)
summary(LS1)
se.LS1<-LS1$residuals
Yhat <- (-37.6)+(0.8*X1)+(0.58*X2)+((-0.07)*X3)
resi <- Y-Yhat
resi
beta2 <- c(LS1$coefficients)
beta02 <- as.matrix(beta2)
SSE_b2 <- t(Y01)%*%Y01 - t(beta02)%*%t(Xls)%*%Y01
MSE_b02 <- (SSE_b2)/(21-3)
MSE_b02
plote<-qqnorm(se.LS1)
qqline(se.LS1, col="red",lWd=4)
#########################
mod.bisq1 <- rlm(Y ~., data=data, method="M")
```

```
names(mod.bisq1)
summary(mod.bisq1)
se.M <- mod.bisq1$residuals
plote<-qqnorm(se.M)
qqline(se.M, col="red",lWd=4)
beta5 <- c(mod.bisq1$coefficients)
beta05 <- as.matrix(beta5)
SSE_b5 <- t(Y01)%*%Y01 - t(beta05)%*%t(Xls)%*%Y01
MSE_b05 <- (SSE_b5)/(21-3)
MSE_b05
```

آ.۱.۱ کدهای مربوط به توزیع نرمال چندمتغیره

```
library(MASS)
library(mvtnorm)
remove(list=ls())
set.seed(2314)
gen.data<-function(beta0, beta1, n){
### generating NSD errors from a multivariate normal dist.
   a=-1/(n-1)
   M=a*matrix(rep(1,n*n),ncol=n)
   diag(M)=rep(1,n)
   eigen(M)$values
   epsilon<-as.vector(rmvnorm(1, mean=rep(0, nrow(M)), sigma=100*M))
### generating covariate x
   x=rnorm(n)
### generating data from a linear model with NSD errors
   y = beta0 + beta1*x+ epsilon
##
   dt = data.frame(cbind(x,y))
return(dt)
}
dt = gen.data(beta0=-1, beta1=.6, n=100)
### M-estimation of the model
```

```
#fit = rlm(y~x, psi= psi.huber, data=dt, method='M', scale.est="Huber")
#summary(fit)
#fit$coef
MSE <- function(B, n){
 Est = matrix(NA, nrow=B, ncol=2)
 bias = mse = rep(NA, 2)
 beta0 = -1; beta1 = 0.6
 for (i in 1:B){
dt <- gen.data(beta0=beta0, beta1=beta1, n=n)
fit = rlm(y~x, psi= psi.huber, data=dt, method='M', scale.est="Huber")
Est[i,] = fit$coef
 }
 bias = apply(Est, 2, mean)-c(beta0, beta1)
 mse = apply((Est-c(beta0, beta1))^2, 2, mean)
return(list(Est=Est, bias=bias, mse=mse))
}
n = c(20, 40, 60, 80, 100, 150, 200, 500, 1000)
res.n20 = MSE(n=n[1], B=500)
res.n40 = MSE(n=n[2], B=500)
res.n60 = MSE(n=n[3], B=500)
res.n80 = MSE(n=n[4], B=500)
res.n100 = MSE(n=n[5], B=500)
res.n150 = MSE(n=n[6], B=500)
res.n200 = MSE(n=n[7], B=500)
res.n500 = MSE(n=n[8], B=500)
res.n1000 = MSE(n=n[9], B=500)
par(mfrow=c(1,2))
plot(c(res.n20$mse[1], res.n40$mse[1], res.n60$mse[1],
res.n80$mse[1], res.n100$mse[1], res.n150$mse[1], res.n200$mse[1],
res.n500$mse[1], res.n1000$mse[1]), type="l", ylab="MSE",
 xlab="sample size", main=expression(beta[0]))
plot(c(res.n20$mse[2], res.n40$mse[2], res.n60$mse[2],
res.n80$mse[2], res.n100$mse[2], res.n150$mse[2],
res.n200$mse[2], res.n500$mse[2], res.n1000$mse[2]),
```

```
type="l", ylab="MSE", xlab="sample size", main=expression(beta[1]))
```

آ.۱.۲ کدهای مربوط به توزیع تی چندمتغیره

```
library(MASS)
library(mvtnorm)
remove(list=ls())
set.seed(2314)
gen.data<-function(beta0, beta1, n){
### generating NSD errors from a multivariate normal dist.
   a=-1/(n-1)
   M=a*matrix(rep(1,n*n),ncol=n)
   diag(M)=rep(1,n)
   eigen(M)$values
   epsilon<-as.vector(rmvt(1, sigma=100*M, df=3))
### generating covariate x
   x=rnorm(n)
### generating data from a linear model with NSD errors
   y = beta0 + beta1*x+ epsilon
##
   dt = data.frame(cbind(x,y))
return(dt)
}
dt = gen.data(beta0=-1, beta1=.6, n=100)
### M-estimation of the model
#fit = rlm(y~x, psi= psi.huber, data=dt, method='M', scale.est="Huber")
#summary(fit)
#fit$coef
MSE <- function(B, n){
 Est = matrix(NA, nrow=B, ncol=2)
 bias = mse = rep(NA, 2)
 beta0 = -1; beta1 = 0.6
 for (i in 1:B){
dt <- gen.data(beta0=beta0, beta1=beta1, n=n)
```

```
fit = rlm(y~x, psi= psi.huber, data=dt, method='M', scale.est="Huber")
Est[i,] = fit$coef
 }
 bias = apply(Est, 2, mean)-c(beta0, beta1)
 mse = apply((Est-c(beta0, beta1))^2, 2, mean)
return(list(Est=Est, bias=bias, mse=mse))
}
n = c(20, 40, 60, 80, 100, 150, 200, 500, 1000)
res.n20 = MSE(n=n[1], B=500)
res.n40 = MSE(n=n[2], B=500)
res.n60 = MSE(n=n[3], B=500)
res.n80 = MSE(n=n[4], B=500)
res.n100 = MSE(n=n[5], B=500)
res.n150 = MSE(n=n[6], B=500)
res.n200 = MSE(n=n[7], B=500)
res.n500 = MSE(n=n[8], B=500)
res.n1000 = MSE(n=n[9], B=500)
par(mfrow=c(1,2))
plot(c(res.n20$mse[1], res.n40$mse[1], res.n60$mse[1],
res.n80$mse[1], res.n100$mse[1], res.n150$mse[1], res.n200$mse[1],
res.n500$mse[1], res.n1000$mse[1]), type="l", ylab="MSE", xlab="sample size",
 main=expression(beta[0]))
plot(c(res.n20$mse[2], res.n40$mse[2], res.n60$mse[2],
res.n80$mse[2], res.n100$mse[2], res.n150$mse[2],
res.n200$mse[2], res.n500$mse[2], res.n1000$mse[2]),
type="l", ylab="MSE", xlab="sample size", main=expression(beta[1]))
```

آ.۱.۳ کدهای مربوط به میانگین ضرایب

```
res.n20 = MSE(n=n[1], B=500)
mean(res.n20$Est[,1])
mean(res.n20$Est[,2])
res.n40 = MSE(n=n[2], B=500)
mean(res.n40$Est[,1])
```

```
mean(res.n40$Est[,2])
res.n60 = MSE(n=n[3], B=500)
mean(res.n60$Est[,1])
mean(res.n60$Est[,2])
res.n80 = MSE(n=n[4], B=500)
mean(res.n80$Est[,1])
mean(res.n80$Est[,2])
res.n100 = MSE(n=n[5], B=500)
mean(res.n100$Est[,1])
mean(res.n100$Est[,2])
res.n150 = MSE(n=n[6], B=500)
mean(res.n150$Est[,1])
mean(res.n150$Est[,2])
res.n200 = MSE(n=n[7], B=500)
mean(res.n200$Est[,1])
mean(res.n200$Est[,2])
res.n500 = MSE(n=n[8], B=500)
mean(res.n500$Est[,1])
mean(res.n500$Est[,2])
res.n1000 = MSE(n=n[9], B=500)
mean(res.n1000$Est[,1])
mean(res.n1000$Est[,2])
res.n10000 = MSE(n=n[9], B=1000)
mean(res.n10000$Est[,1])
mean(res.n10000$Est[,2])
```

آ.۱.۴ کدهای مربوط به نمودارهای جعبه‌ای

```
res.n40 = MSE(n=n[2], B=50)
boxplot(res.n40$Est)
res.n200 = MSE(n=n[7], B=500)
boxplot(res.n200$Est)
res.n1000 = MSE(n=n[9], B=500)
boxplot(res.n1000$Est)
```

Aabstract

In this thesis we discuss the strong consistency of M-estimates of the regression parameters in a linear model with negatively superadditive dependent (NSD) random errors. The result improves the moment condition and generalises the case of independent random errors to that of NSD random errors.
key words: Strong consistency, Negatively superadditive dependent, Robust regression.

Shahrood University of Technology

Faculty Of Mathematical Sciences

MSc Thesis in: Mathematical Statistic

On the strong consistency of M-estimates in linear models for negatively superadditive dependent errors

By: Hossein Heidari Nezhad

Supervisor

Negar Eghbal

Advisor

Hossein Baghishani

January 2019