دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد آمارریاضی

آزمون‌های استوار M در مدل‌های خطی با خطاهای زبرجمعی منفی

نگارنده: سمانه درویشی بهلولی

استاد راهنما

دکتر نگار اقبال

استاد مشاور

دکتر حسین باغیشنی

بهمن ۱۳۹۹

تقدیم با بوسه بر دستان پدرم که راه را به من نشان داد و تقدیم به مادر عزیزتر از جانم که چگونه رفتن را به من آموخت.

سپاس‌گزاری...

به مصداق لم یشکرالمخلوق، لم یشکر الخالق سپاس بی‌کران نثار پدر و مادرم، که ناتوان شدند تا من توانا شوم، مویشان سپید شد تا من روسفید شوم و عاشقانه سوختند تا گرمابخش وجودم و روشنگر را هم باشند. و هم‌چنین از استاد گرامی سرکار خانم دکتر اقبال و و استاد محترم جناب آقای دکتر باغیشنی به خاطر زحمت‌ها و راهنمایی‌هایی که در طول این پایان‌نامه برای من کشیده‌اند تقدیر و تشکر می‌کنم.

سمانه درویشی بهلولی
بهمن ۱۳۹۹

تعهد نامه

اینجانب **سمانه درویشی بهلولی** دانشجوی کارشناسی ارشد رشته **آماررياضی** دانشکده **علوم ریاضی** دانشگاه صنعتی شاهرود، نویسنده پایان‌نامه با عنوان **آزمون‌های استوار M در مدل‌های خطی با خطاهای زبرجمعی 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.

سمانه درویشی بهلولی
بهمن ۱۳۹۹

مالکیت نتایج و حق نشر

چکیده:

در حالت کلی استواری یک روش آماری را می‌توان به عنوان حساس نبودن آن به انحراف کم از پذیره‌های پایه‌ای تعریف کرد. به عبارت دیگر هرگاه ویژگی‌های یک روش آماری تحت تاثیر انحراف کم از پذیره‌های اساسی قرار نگیرد، آن روش آماری را استوار گوییم. با توجه به اینکه برآوردگرهای کلاسیک ضرایب رگرسیونی در مدل‌های خطی به شدت تحت تاثیر داده‌های پرت قرار می‌گیرند، برآوردگرهای استوار مانند M ـ برآوردگرها پیشنهاد می‌شوند. همچنین سطح و توان آزمون‌های کلاسیک با حضور داده‌های پرت به شدت کاهش می‌یابند، برای حل این مشکل، آزمون‌های استوار معرفی می‌شوند. برای انجام آزمون استوار پارامترهای رگرسیونی، توزیع مجانبی آماره آزمون مشخص می‌شود و بر اساس پایایی برآوردگرهای مدل و توان آزمون، روش‌ها مورد ارزیابی قرار می‌گیرند. در این پایان نامه علاوه بر بررسی برخی ویژگی‌های اساسی متغیرهای تصادفی وابسته زبرجمعی منفی، آزمون‌های استوار M در مدل‌های خطی با خطای وابسته زبرجمعی منفی را تحت برخی شرایط مورد مطالعه قرار می‌دهیم و با استفاده از روش‌های شبیه‌سازی مونت کارلو به مقایسه روش‌های کلاسیک و استوار بر اساس پایایی برآوردگرها و توان آزمون می‌پردازیم.

کلمات کلیدی: آزمون استوار، وابسته زبرجمعی‌منفی، مدل خطی رگرسیون، شبیه‌سازی‌های مونت‌کارلو

فهرست مطالب

فهرست تصاویر

فهرست جداول

پیشگفتار

در متون آماری، استوار بودن یک روش، به معنی معتبر و مناسب بودن آن است. به عبارت دیگر، هرگاه ویژگی‌های یک روش آماری، تحت تاثیر انحراف کم از پذیره‌های اساسی قرار نگیرد، آن را روش آماری استوار گوییم. متغیرهای تصادفی زبرجمعی منفی یک رده بسیار وسیع از متغیرهای 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. در این پایان‌نامه برخی از روش‌های آماری استوار و آزمون استوار در مدل خطی با جملات خطای وابسته منفی و برخی ویژگی‌های اساسی متغیرهای تصادفی وابسته زبرجمعی‌منفی را رمورد مطالعه قرار می‌دهیم. هم‌چنین در انتها با مطالعه شبیه‌سازی به ارزیابی نتایج نظری می‌پردازیم. محتوای این پایان‌نامه به صورت زیر تدوین شده است:

- در فصل اول به بیان برخی تعاریف و مفاهیم اولیه که در اثبات نتایج اصلی به آن‌ها نیاز داریم، می پردازیم.
- در فصل دوم مفهوم استواری و برخی روش‌های آماری استوار مانند برآورد پارامترهای مدل رگرسیونی استوار معرفی می‌کنیم.
- در فصل سوم مدل خطی آزمون استوار با خطاهای وابسته زبرجمعی منفی را بررسی می‌کنیم.
- در فصل چهارم با استفاده از ابزارهای شبیه‌سازی نتایج نظری استواری فصل سوم را ارزیابی می‌کنیم.
- پیوست پایان‌نامه شامل کدهای نرم‌افزار R برای تولید نتایج شبیه‌سازی می‌باشد.

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 گات (۱۹۹۴)، پایان نامه‌های کارشناسی ارشد حیدری نژاد (۱۳۹۷)، حبیبی کمنی (۱۳۹۵) و رحمانی (۱۳۹۵) استفاده کرده‌ایم.

۱.۱ 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 هرچه کران به مقدار واقعی نزدیک‌تر باشد، جواب نهایی قابل اعتمادتر است. در زیر برخی از نابرابری‌های مهم آماری را از گات (۱۹۹۴) معرفی می‌کنیم.

۱ـ نابرابری مارکف[1]: فرض کنید X متغیر تصادفی که برای عدد حقیقی مثبت r که $E|X|^r < \infty$ آن‌گاه اگر $x > 0$

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

[1] Markov's Inequalitiy

۲ـ نابرابری هافدینگ[1]: فرض کنید $X_1, X_2, ..., X_n$ متغیرهای تصادفی مستقل باشند طوری که برای $k = 1,2,...,n$ و $P(a_k \leq X_k \leq b_k) = 1$. هم چنین فرض کنید برای $n \geq 1$ ، $S_n = \sum_{k=1}^{n} X_k$ مجموع‌های جزیی باشد. آن‌گاه برای $x > 0$ و a_n, b_n که اعداد حقیقی ثابت هستند و $a_n \leq b_n$

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

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

۳ـ نابرابری C_r[2] : فرض کنید $\{X_n; n \geq 1\}$ دنباله ای از متغیرهای تصادفی مستقل باشند. اگر عدد حقیقی مثبت $r > 0$ وجود داشته باشد که به ازای $i = 1,2,...,n$ ، $E|X_i|^r < \infty$ باشد، آن‌گاه

$$E|\sum_{i=1}^{n} X_i|^r \leq C_r \sum_{i=1}^{n} E|X_i|^r$$

که C_r مقادیر زیر را اختیار می‌کند

$$C_r = \begin{cases} 1 & \text{اگر } r \leq 1 \\ 2^{r-1} & \text{اگر } r \geq 1 \end{cases}$$

۴ـ نابرابری هولدر[3]: فرض کنید برای هر عدد حقیقی $q > 1$ و $p > 1$ که

$$p^{-1} + q^{-1} = 1$$

اگر $E|X|^p < \infty$ و $E|Y|^q < \infty$ آن‌گاه

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

۵ـ نابرابری کوشی‌ـشوارتز[4]: در نابرابری هولدر اگر $p = q = 2$ قرار دهیم آن‌گاه نابرابری کوشی شوارتز به صورت زیر به دست می آید:

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

۶ـ نابرابری مینکوفسکی[5]: فرض کنید X و Y متغیرهای تصادفی باشد و برای هر عدد حقیقی $p \geq 1$ ، $E|X|^p < \infty$ و $E|Y|^p < \infty$ باشد. آن‌گاه

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

[1] Hoeffding's Inequality
[2] C_r *Inequality*
[3] Holder's Inequality
[4] Cauchy-Schwarz Inequality
[5] Mincowski Inequality

۲.۱ انواع وابستگی

در این بخش به معرفی چند نوع از وابستگی‌ها که رده‌ی بزرگی از دنباله متغیرهای تصادفی را در بر می‌گیرند می‌پردازیم.

۱.۲.۱ وابسته ربعی منفی

تعریف ۱.۲.۱. (لهمن ۱۹۶۶) متغیرهای تصادفی X و Y دارای وابستگی ربعی منفی[1] (NQD) هستند اگر برای هر x و y حقیقی داشته باشیم:

$$P(X \leq x, Y \leq y) \leq P(X \leq x)\, P(Y \leq y)$$

یا به طور معادل

$$F(x,y) \leq F_1(x)\, F_2(y).$$

که F تابع توزیع توام و F_1, F_2 توابع توزیع حاشیه‌ای هستند، اگر علامت نابرابری روابط بالا را برعکس کنیم به وابستگی ربعی مثبت[2] (PQD) تبدیل می‌شود. شرایط تعریف شده برای NQD نتیجه می‌دهد که $Cov(X,Y) \leq 0$ است.

۲.۲.۱ وابستگی پیوندی منفی

نتایج مختلفی در آمار و احتمال تحت فرض برخی از متغیرهای تصادفی دارای وابستگی پیوندی منفی[3] (NA) هستند توسط آلام و سکنا (۱۹۸۱) معرفی شده‌اند و جاج و پروشان (۱۹۸۳) آنها را به دقت مورد مطالعه قرار داده‌اند.

تعریف ۲.۲.۱. یک دنباله متناهی از متغیرهای تصادفی $\{X_i, : 1 \leq i \leq n\}$ ، وابسته پیوندی منفی (NA) نامیده می‌شود اگر برای هر جفت زیرمجموعه جدا از هم A_1 و A_2 از $\{1,2,...,n\}$ بتوان نوشت

$$Cov\Big(f_1(X_i, i \in A_1)\,, f_2(X_j, j \in A_2)\Big) \leq 0$$

هرگاه دو تابع f_1 و f_2 نسبت به هر مولفه صعودی باشند و کواریانس تعریف شده در رابطه بالا وجود داشته باشد.

نکته ۱.۲.۱. (جاج و پروشان ۱۹۸۳). اگر توابع f_1 و f_2 هر دو نزولی باشند نابرابری رابطه $Cov\Big(f_1(X_i, i \in A_1), f_2(X_j, j \in A_2)\Big) \leq 0$ نیز برقرار است. یکی از مزیت‌های وابستگی NA

[1] Negative Quadrant Dependence
[2] Positive Quadrant Dependence
[3] Negatively Associated

این است که توابع f_1 و f_2 در رابطه بالا می‌توانند صورتهای مختلفی داشته باشند و این نتیجه به بسیاری از نابرابری‌های چند متغیره مفید منجر می‌شود.

بدون از دست دادن کلیت می‌توان فرض کرد $A_1 \bigcup A_2 = \{1,2,...,n\}$. هم‌چنین یک دنباله نامتناهی پیوندی منفی است اگر هر زیر دنباله متناهی آن پیوندی منفی باشد.

۳.۲.۱ چند ویژگی متغیرهای تصادفی پیوندی منفی

چند ویژگی برای متغیرهای تصادفی NA توسط جاج و پروشان (۱۹۸۳) بیان شده‌اند که در زیر به آنها اشاره می‌کنیم:

۱ـ هرجفت از متغیرهای تصادفی NQD، NA می‌باشد.

۲ـ فرض کنید $A_1, A_2, ..., A_m$ زیرمجموعه‌های مجزا از $\{1,2,..,n\}$ و هم‌چنین توابع $f_1, f_2, ..., f_m$ نسبت به هر مولفه صعودی باشند، در این حالت NA بودن $X_1, X_2, ..., X_n$ دلالت بر این دارد که

$$E\prod_{i=1}^{m} f_i(X_j, j \in A_i) \leq \prod_{i=1}^{m} Ef_i(X_j, j \in A_i).$$

۳ـ یک نتیجه بدیهی از ویژگی ۲ این است که برای زیرمجموعه‌های جدا از هم A_1 و A_2 از $\{1,2,...,n\}$ و هم‌چنین مقادیر حقیقی $x_1, x_2, ..., x_n$ می‌توان نوشت

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

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

۴ـ یک زیرمجموعه دوتایی یا بیشتر از متغیرهای تصادفی NA، NA است.

۵ـ یک مجموعه از متغیرهای تصادفی مستقل، NA است.

۶ـ توابع صعودی تعریف شده روی زیرمجموعه‌های جدا از هم یک مجموعه از متغیرهای تصادفی NA، NA هستند.

۴.۲.۱ مثال‌هایی از متغیرهای تصادفی پیوندی منفی

جاج و پروشان (۱۹۸۳) نشان دادند برای دو مثال زیر ویژگی NA می‌تواند برقرار باشد:

۱) فرض کنید $\mathbf{Z} = (Z_1, Z_2, ..., Z_k)$ بردار تصادفی دارای توزیع چند جمله‌ای است که تنها دارای یک مشاهده است. بنابراین فقط یک Z_i برابر با ۱ می‌باشد و بقیه اعضا صفر هستند. به عبارتی $\sum_{i=1}^{k} Z_i = 1$ است. پس با افزایش یک Z_i بقیه یا ثابت می‌مانند یا کاهش پیدا می‌کنند.

پس کوواریانس بین آنها منفی است. بنابراین از تعریف پیوندی‌منفی نتیجه می‌شود که ویژگی NA برای **Z** بدیهی است. با توجه به ویژگی (۶) به NA بودن **Z** پی می‌بریم.

۲) جعبه‌ای شامل N توپ با رنگ‌های متمایز را در نظر بگیرید. نمونه‌ای شامل n توپ (بدون جای‌گذاری) از این مجموعه انتخاب کرده و متغیرهای تصادفی X_i، $i = ۱,۲,...N$ را تابع نشان‌گر وجود یک توپ از رنگ i ام در نمونه تعریف کنید. به عبارت دیگر $X_i = ۱$ است اگر توپی با رنگ i ام در نمونه موجود باشد و در غیر این صورت $X_i = ۰$ است. با توجه به این که از هر رنگ فقط یک توپ موجود است پس بردار تصادفی $(X_۱, X_۲, ..., X_n)$ برای هر $n < N$، زیرمجموعه $(X_۱, X_۲, ..., X_N)$ بوده که دارای توزیع جایگشتی است. بنابراین با توجه به ویژگی‌های (۵) و (۶) پس NA می‌باشد. به طور کلی فرض کنید N_i توپ از رنگ i ام $۱ \le i \le k$ در جعبه موجود باشد. پس $\sum_{i=۱}^{k} Z_i = N$ است. اگر Y_i تعداد توپ‌هایی از رنگ i ام در نمونه n تایی باشد آن‌گاه Y_i را می‌توان به عنوان مجموع N_i متغیر تصادفی نشان‌گری که در قسمت قبل تعریف کردیم، در نظر گرفت. با توجه به ویژگی (۶) Y_i ها هم NA هستند.

۵.۲.۱ وابستگی زبرجمعی منفی

وابسته زبرجمعی منفی[۱] (NSD) نوعی وابستگی است که توسط هو (۲۰۰۰) مطرح شد که در پایه آن از توابع زبرجمعی[۲] استفاده شده است.

تعریف ۳.۲.۱. (کمپرمن، ۱۹۷۷) تابع $\phi : R^n \to R$ زبرجمعی نامیده می شود اگر برای هر $\mathbf{x}، \mathbf{y} \in R^n$

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

در رابطه بالا $\vee$ نشان دهنده‌ی ماکسیمم مولفه‌ای[۳] و $\wedge$ نشان دهنده‌ی مینیمم مولفه‌ای[۴] است، که به صورت زیر تعریف می‌شود.

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

و

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

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

[1] Negatively Superadditive Dependent
[2] Superadditive Function
[3] Maximum Componentwise
[4] Minimum Componentwise

لم ۱.۲.۱. فرض کنید تابع ϕ دارای مشتق جزیی مرتبه دوم پیوسته باشد، آن‌گاه ϕ زبرجمعی است هرگاه برای $۱ \leq i \neq j \leq n$ داریم:

$$\frac{\partial^۲ \phi(\mathbf{x})}{\partial x_i \partial x_j} \geq ۰ \qquad (۱.۱)$$

تعریف ۴.۲.۱. بردار تصادفی $\mathbf{X} = (X_۱, X_۲, ..., X_n)$ را NSD گویند، هر گاه برای هر تابع زبرجمعی ϕ

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

که Y_i ها متغیرهای تصادفی مستقل هستند، هم چنین به ازای هر $i = ۱, ۲, ..., n$ ، X_i و Y_i هم‌توزیع هستند.

تعریف ۵.۲.۱. $\{X_n, n \geq ۱\}$ را دنباله‌ای از متغیرهای تصادفی NSD گویند هرگاه برای هر $n \geq ۱$ ، $(X_۱, X_۲, ..., X_n)$، NSD باشد.

لم ۲.۲.۱. (هو، ۲۰۰۰) فرض کنید بردار $(X_۱, X_۲, ..., X_n)$، NSD باشد. آن‌گاه

۱ــ $(-X_۱, -X_۲, ..., -X_n)$ نیز NSD هستند.

۲ــ برای توابع صعودی دلخواه $f_۱, f_۲, ..., f_n$ ، $(f_۱(X_۱), f_۲(X_۲), ..., f_n(X_n))$ نیز NSD هستند.

۳ــ برای هر $x_۱, x_۲, ..., x_n \in R$، می‌توان نتیجه گرفت

$$P(X_۱ \leq x_۱, X_۲ \leq x_۲, ..., X_n \leq x_n) \leq P(X_۱ \leq x_۱)P(X_۲ \leq x_۲)...P(X_n \leq x_n)$$

برهان. با توجه به فرض لم، داریم

۱ــ از آنجا که بردار $(X_۱, X_۲, ..., X_n)$، NSD است و با توجه به لم ۱.۲.۱ کافی است نشان دهیم که برای هر تابع زبرجمعی ϕ و به ازای هر $i \neq j$، تابع $\phi(-X_۱, -X_۲, ..., -X_n)$ نیز زبرجمعی است. در نتیجه NSD بودن آن روشن است. زیرا

$$\begin{aligned}\frac{\partial^۲ \phi(-x_۱, -x_۲, ..., -x_n)}{\partial x_i \partial x_j} &= \frac{\partial^۲ \phi(-x_۱, -x_۲, ..., -x_n)}{\partial(-x_i)\partial(-x_j)} \times \frac{\partial(-x_i)\partial(-x_j)}{\partial x_i \partial x_j} \\ &= \frac{\partial^۲ \phi(x_۱, x_۲, ..., x_n)}{\partial x_i \partial x_j} \geq ۰\end{aligned}$$

۲ــ چون f تابعی صعودی است، با توجه به لم ۱.۲.۱ کافی است نشان دهیم برای هر تابع زبرجمعی ϕ و هر $i \neq j$، تابع $\phi(f_۱(x_۱), f_۲(x_۲), ..., f_n(x_n))$ نیز زبرجمعی است. در نتیجه برهان کامل می‌شود، زیرا

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

۳ـ با توجه به تعریف وابستگی زبرجمعی منفی، همچنین بر اساس قسمت دوم این لم با قرار دادن تابع صعودی نشانگر به جای تابع صعودی f، می‌توان نشان داد

$$\begin{aligned} P(X_۱ \leq x_۱, X_۲ \leq x_۲, ..., X_n \leq x_n) &= E(I\{X_۱ \leq x_۱, X_۲ \leq x_۲, ..., X_n \leq x_n\}) \\ &= E\Big(\prod_{i=۱}^{n} I\{X_i \leq x_i\}\Big) \\ &\leq E\Big(\prod_{i=۱}^{n} I\{Y_i \leq x_i\}\Big) \\ &= \prod_{i=۱}^{n} E(I\{Y_i \leq x_i\}) \\ &= \prod_{i=۱}^{n} P(Y_i \leq x_i) \\ &= \prod_{i=۱}^{n} P(X_i \leq x_i). \end{aligned}$$

□

ملاحظه ۱.۲.۱. قسمت ۲ لم ۲.۲.۱ اگر توابع $f_۱, f_۲, ..., f_n$ ناصعودی باشند نیز برقرار است.

ملاحظه ۲.۲.۱. (کریستوفیدز و واگلاتو، ۲۰۰۴) اگر $X_۱, X_۲, ..., X_n$ پیوندی منفی باشند، آن‌گاه وابسته زبرجمعی‌منفی هستند ولی عکس آن همیشه برقرار نیست.

در زیر برخی ویژگی‌های متغیرهای تصادفی وابسته زبرجمعی‌منفی را از هو (۲۰۰۰)، بدون برهان بیان می‌کنیم.

ملاحظه ۳.۲.۱. وابستگی پیوندی منفی یک زوج متغیر تصادفی با وابستگی زبرجمعی‌منفی آن معادل است.

ملاحظه ۴.۲.۱. برای یک زوج متغیر تصادفی NQD با NA معادل است، لذا روابطی که میان NA و NSD برقرار هستند، میان NSD و NQD نیز برقرار خواهند بود.

۶.۲.۱ مثال‌ها و کاربردهایی از وابسته زبرجمعی منفی

خانواده‌ای از توزیع‌های چندمتغیره با ویژگی NSD

●خانواده توزیع‌های بیضی‌گون: برای این منظور، در ابتدا به تعریفی از خانواده توزیع‌های بیضی‌گون می‌پردازیم.

تعریف۶.۲.۱. (بولک و سامپسون، ۱۹۸۸) خانواده $F(x,\sum)$ خانواده‌ای از توزیع‌های بیضی‌گون[1] پایه‌گذاری شده بر اساس تابع g است، هرگاه تابع چگالی آن به صورت زیر تعریف شود:

$$f(x,\sum)=\left|\sum\right|^{-\frac{1}{2}}g(x^T\sum x) \qquad :x\in\Re^m$$

که در آن $\int_0^\infty g(t)t^{\frac{m}{2}-1}dt<\infty$.

بولک و سامپسون (۱۹۸۸) نشان دادند که $F(x,\sum)$ دارای وابستگی از نوع NSD خواهد بود اگر برای هر $i\neq j$، $\delta_{ij}\leq 0$ باشد به طوری که δ_{ij} ها مولفه‌های ماتریس $\sum$ هستند. در این میان، توزیع نرمال چندمتغیره نیز به عنوان حالت خاصی از توزیع‌های بیضی‌گون محسوب می‌شود. بنابراین توزیع نرمال چندمتغیره NSD خواهد بود اگر عناصر خارج از قطر اصلی ماتریس کواریانس آن نامثبت باشند.

●خانواده توابع فراوانی پولیای نوع۲[2] :به جهت آشنایی با این دست از توزیع‌ها تعریفی از آن را ارائه می‌کنیم.

تعریف ۷.۲.۱. (افرون، ۱۹۶۵) تابع چگالی احتمال روی خط حقیقی $r(x)$ را تابع فراوانی پولیای نوع ۲ (PF_2) نامند، هرگاه به ازای $x_1\geq x_2$ و $z_1\geq z_2$ داشته باشیم

$$\begin{bmatrix} r(x_1-z_1) & r(x_1-z_2)\\ r(x_2-z_1) & r(x_2-z_2)\end{bmatrix}\geq 0$$

این خانواده از توزیع‌ها دارای ویژگی‌هایی است که در زیر به طور مختصر به آن‌ها اشاره می‌نماییم.

آ. تعداد زیادی از توزیع‌های رایج در نظریه احتمال، PF_2 هستند که به عنوان حالت خاص می‌توان به توزیع‌های پواسن، دوجمله‌ای، گاما دوجمله‌ای منفی و همچنین توزیع نرمال با تابع چگالی

$$r(x)=(2\pi\sigma^2)^{-\frac{1}{2}}exp\Big\{\frac{1}{2}\Big[\frac{x-m}{\sigma}\Big]^2\Big\}$$

با $\sigma>0$، اشاره کرد.

[1] Elliptically Contoured D0istribution
[2] Polya FrequencyFunction of order 2

ب. توابع چگالی $PF_۲$ به طور لگاریتمی مقعرند[1]. بنابراین تقریبا همه جا دارای مشتق اول و دوم هستند.

پ. توابع چگالی $PF_۲$ کران‌دارند و همچنین بر روی یک فاصله، غیرصفر و در خارج از آن، صفر هستند.

ت. فرض کنید $r(x)$ و $q(x)$، $PF_۲$ باشند آن‌گاه پیچش آن‌ها به شکل

$$P(x) = \int_{-\infty}^{\infty} r(t)q(x-t)dt$$

نیز $PF_۲$ است.

بنابراین توزیع‌هایی مانند پواسن، دوجمله‌ای، گاما و دوجمله‌ای منفی دارای ویژگی $PF_۲$ هستند.

بر اساس آن‌چه هو (۲۰۰۰) ارائه کرده است، اگر $X_۱, X_۲, ..., X_n$ متغیرهای تصادفی مستقل با تابع چگالی احتمال $PF_۲$ باشند، $S_n = \sum_{i=۱}^{n} X_i$ به طور تقریبا حتمی NSD است.

۳.۱ ترتیب‌های غیرتصادفی

ترتیب‌های غیرتصادفی را با o و O را نشان می‌دهند که نمادهایی برای توصیف مرتبه‌ی مجانبی کمیت‌های غیرتصادفی می‌باشند. ون و 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 ۱.۳.۱. دو دنباله از اعداد ثابت $\{a_n; n \geq ۱\}$ و $\{b_n; n \geq ۱\}$ را در نظر بگیرید. گوییم $b_n = O(a_n)$، هرگاه برای هر $\varepsilon > ۰$، $K(\varepsilon) > ۰$ و عدد صحیح $N(\varepsilon)$ وجود داشته باشند به طوری‌که برای هر $n \geq N$، آن‌گاه

$$|b_n| < K(\varepsilon)|a_n|$$

تعریف ۲.۳.۱. دو دنباله از اعداد ثابت $\{a_n; n \geq ۱\}$ و $\{b_n; n \geq ۱\}$ را در نظر بگیرید. گوییم $b_n = o(a_n)$، هرگاه

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

با استفاده از تعاریفی که در بالا به آن اشاره شد، برای هر ثابت حقیقی مثل C، ترتیب‌های $O(a_n)$ و $o(a_n)$، که به ترتیب معادل با $Ca_nO(۱)$، $Ca_no(۱)$ هم‌چنین اگر $X_n = O(n^C)$، آن‌گاه $X_n = O(n^{C+۱})$ قرار می‌گیرد اما عکس رابطه برقرار نیست.

[1] Log-Concave

فصل ۲

روش های آماری استوار

در این فصل به شرح روش‌های آماری استوار از جمله برآورد پارامترهای مکانی استوار، رگرسیون استوار و تاریخچه‌ی آن و مقایسه آن با کمترین توان‌های دوم درمورد نقاط پرت می‌پردازیم. مطالب این فصل از هوگ (۱۷۹۹)، پایان نامه کارشناسی ارشد اقبال (۱۳۸۴) و حیدری نژاد (۱۳۹۷) می‌پردازیم.

۱.۲ استواری

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 تأثیر 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 داشته باشد و از جامعه‌ی مورد نظر نیامده باشد، داده پرت باشد، بسیار متفاوت از میانگین واقعی 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 در نمونه باشد، مقدار میانه تغییر نمی‌کند.

این سوال مطرح می‌شود که نرمال نبودن مشاهدات بر دقت و کارایی بازه اطمینان چقدر تأثیر 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 تأثیر می‌گذارد؟

جوابی که می‌توان به این سوال داد این است که تاثیر نرمال نبودن، در حالی 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 ـ 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 بر روی برآوردگر نهایی دارد. یکی از اقسام رگرسیون، رگرسیون استوار می‌باشد (در مورد تجزیه و تحلیل داده‌های پرت) که توسط جورج باکس (۱۹۵۳) برای اولین بار معرفی گردید. مقالات و کتاب‌هایی در این زمینه به چاپ رسیده است که می‌توان از آنها استفاده کرد. از جمله جان توکی (۱۹۶۲) 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 نقطه 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 علاقه به تاثیر منحنی داشته باشد می‌تواند به مقاله همپل (۱۹۷۴) مراجعه کند. پس از آنکه هوبر (۱۹۷۳) M ـ برآوردگرها را مطالعه کرد، بسیاری از آمارشناسان علاقه‌مند به مطالعه این موضوع هستند. مراجعه به کارهای اخیر M ـ برآوردگرها را می‌توان در چن و ژائو (۱۹۹۵) یافت. پایداری مجانبی M ـ برآوردگرهای مدل‌های خطی با خطاهای وابسته عملا مهم است اما از نظر علم نظری چالش‌برانگیز است. هوبر (۱۹۷۳) اظهار داشت که محدودیت درمورد فرض استقلال مهم است و بیان کرد که فرض استقلال را در نظریه M ـ برآوردگرهای کلاسیک کاهش خواهیم داد تا خطاهای وابسته مناسب‌تر شود. به طور کلی بدست آوردن خصوصیات استواری (سازگاری) قوی برای خطاهای وابسته کار آسانی نیست.

مدل خطی رگرسیون $\mathbf{Y}_i = \mathbf{X}_i^\top \boldsymbol{\beta} + e_i \quad i = ۱, ۲, ..., n$ را در نظر بگیرید. وقتی مشاهدات $\mathbf{Y}_i$ ها دارای توزیع نرمال هستند برآورد خوبی از $\boldsymbol{\beta}$ حاصل می‌شود. اما اگر مشاهدات دارای توزیع نرمال نباشند نمی‌تواند برآورد خوبی را به‌دست آورد. زیرا در انتهای منحنی دارای مشاهداتی هستیم که منحنی دم‌پهن‌تر یا به عبارتی ضخیم تر از نرمال می‌باشد که روش کمترین توان های دوم کارا نیست که این شاخه‌های ضخیم توزیع را معمولا نقاط پرت ایجاد می‌کنند. این نقاط پرت بر برآوردگرهای کمترین توان های 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 را 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[1] را پیشنهاد می‌دهند که برای کاهش اثر مشاهداتی به کار می‌رود. اگر روش کمترین توان‌های دوم تاثیرگذاری بالایی خواهند داشت یعنی اینکه روش استواری می‌خواهد

[1] Robust Regression

باقیمانده‌های مرتبط با دور افتاده‌های بزرگ را کنار بگذارد که به این وسیله تشخیص نقاط پرت خیلی ساده‌تر می‌شود. رگرسیون استوار علاوه بر حساس نبودن نسبت به داده‌های پرت زمانی که مشاهدات دارای توزیع نرمال هستند دارای کارایی ۹۰ تا ۹۵ درصد 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 ـ برآوردگر[1]، L ـ برآوردگر[2]. هدف اصلی ما در این پایان‌نامه 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 به‌دست آوردن تابع چگالی $f(x-\theta)$ فرض کنید که $x_1, x_2, \cdots, x_n$ یک نمونه‌ی تصادفی از این توزیع را نشان می‌دهد یکی از روش‌های مشهور برای برآورد پارامتر مکان استفاده از روش درست نمایی ماکسیمم است. اگر تابع درست نمایی را با $L(\theta)$ نشان دهیم.

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

که $\rho(x) = -lnf(x)$ که با مشتق گرفتن نتیجه به صورت زیر حاصل می شود

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

اگر $\rho' = \psi(x)$ آنگاه

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

$L(\theta)$ را ماکسیمم می کند. به θ ای که در رابطه بالا صدق می کند برآوردگر درست نمایی ماکسیمم گوییم و آن را با $\hat{\theta}$ نمایش می دهیم به مثال های زیر توجه کنید:

۱ـ نرمال؛ $\rho(x) = \frac{x^2}{2} + c$ در این صورت $\psi(x) = x$ و

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

۲ـ نمایی دوگانه

[1] M- estimator
[2] L- estimator

$$\rho(x) = |x| + c$$

با توجه به اینکه

$$\psi(x) = \begin{cases} -۱ & x < ۰ \\ ۱ & x > ۰ \end{cases}$$

است در این صورت $\hat{\theta}$ میانه نمونه می باشد.

۳ـ کوشی

$$\rho(x) = ln(۱ + x^۲) + c$$

و با توجه به اینکه

$$\psi(x) = \frac{۲x}{۱ + x^۲}$$

بنابراین $\hat{\theta}$ با روش تکرار حل می شود.

توجه داشته باشید که تابع ψ در مثال‌های مربوط نمایی دوگانه و کوشی کراندار است هم چنین در مثال مربوط به توزیع کوشی با افزایش x به طور مجانبی به صفر نزدیک می شود که این جواب ها خیلی تحت تاثیر نقاط پرت قرار نمی گیرند 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 دم‌پهن مانند کوشی در نظر گرفته شود خوب نیست. بنابراین در برآورد استوار 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 کارایی (تقریبا نزدیک به ۵ درصد) کاسته می شود.

تعریف ۱.۲.۲. اگر $U_۱$ و $U_۲$ دو برآوردگر نااریب برای θ باشند برآوردگری کاراتر است که واریانس کمتری داشته باشد. کارایی برآوردگر $U_۱$ نسبت به $U_۲$ به صورت زیر تعریف می‌کنیم

$$e(U_۱, U_۲) = \frac{Var(U_۲)}{Var(U_۱)}$$

اگر $e < ۱$ باشد، $U_۲$ کاراتر از $U_۱$ است و اگر $e > ۱$ باشد، $U_۱$ کاراتر است.

برای یک دلیل نظری خاص (مینیمم کردن ماکسیمم واریانس مجانبی از برآوردگرهای مربوط به یک کلاس از توزیع‌ها)، هوبر (۱۹۶۴) پیشنهاد داد که برای برآوردگرهای استوار، برآوردکننده ماکسیمم درست‌نمایی 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

$$\rho(x) = \begin{cases} \frac{۱}{۲}x^۲ & |x| \le k \\ k|x| - \frac{k^۲}{۲} & k < |x| \end{cases} \qquad (۱.۲)$$

هم‌چنین تابع ψ مربوط به آن به صورت زیر می‌باشد

$$\psi(x) = \begin{cases} -k & x < -k \\ x & -k \le x \le k \\ k & k < x \end{cases} \qquad (۲.۲)$$

که با روش تکراری حل می‌شود. یک برآوردگر از این نوع (لزوما این تابع ψ نیست)، برآوردگر برای این حالت را با $\hat{\theta}$ نمایش می‌دهیم که این را M ـ برآوردگر می‌نامیم، یکی‌دیگر از ویژگی‌های M ـ برآوردگرها به صورت زیر است.

فرض کنید $\hat{\theta}$ یک برآوردگر برای نمونه خاص $x_1, x_2, ..., x_n$ باشد، سپس این موارد توسط بعضی از آنها جایگزین شده باشد به عنوان مثال انحراف از $\hat{\theta}$ سه برابر شده باشد. این برآوردگر جدید $\hat{\theta}$ با استفاده از نمونه تعدیل‌یافته لزوما یکسان نیست. برای به‌دست آوردن مقیاس ثابت از این برآوردگر ما می‌توانیم معادله زیر را حل کنیم

$$\sum_{i=1}^{n} \psi\Big(\frac{x_i - \theta}{d}\Big) = 0 \qquad (۳.۲)$$

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

$$d = \frac{median|x_i - median(x_i)|}{(0/6745)} \qquad (۴.۲)$$

صورت کسر d، میانه انحراف از میانه 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 استاندارد 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 هموارساز k باعث بالا بردن کارایی $\hat{\theta}$ می‌شود. در حالت نرمال، ما بیشترین محدودیت‌ها را برای برآورد نابرابری خواهیم داشت

$$\Big|\frac{x_i - \theta}{d}\Big| \le k \qquad (۵.۲)$$

زیرا

$$\psi\Big(\frac{x_i - \theta}{d}\Big) = \frac{x_i - \theta}{d} \qquad (۶.۲)$$

در واقع اگر همه‌ی محدودیت‌ها 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 توزیع دم سنگین کارایی بهتر می‌شود، اما در بسیاری از موارد که توزیع‌ها غیرنرمال 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.

۳.۲ رگرسیون استوار

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

$$\mathbf{Y} = \mathbf{X}^\top \boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (۱.۲)$$

$\mathbf{Y}$ برداری تصادفی $n \times ۱$، $\mathbf{X}$ ماتریس طرح است با بعد $n \times p$ طوری که $\mathbf{X}'\mathbf{X}$ یک ماتریس رتبه کامل است، $\boldsymbol{\beta}$ یک بردار $n \times ۱$ از ضرایب مجهول است و $\boldsymbol{\varepsilon}$ بردار تصادفی $n \times ۱$ است که عناصر آن نمونه ای از یک توزیع متقارن حول صفر(مثلا نرمال) هستند. مشابه رهیافت پارامتر مکان ما می‌خواهیم مجموعه زیر را مینیمم کنیم

$$\sum_{i=۱}^{n} \rho\Big(\frac{Y_i - X_i^\top \beta}{\mathbf{d}}\Big) \qquad (۲.۲)$$

که در آن Y_i، i امین عنصر از $\mathbf{Y}$ و X_i، i مین سطر از ماتریس $\mathbf{X}$ است و $\mathbf{d}$ برآورد پارامتر مقیاس توزیع مربوط به $\boldsymbol{\varepsilon}$ 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 ماکسیمم جواب از رابطه زیر می‌باشد

$$\sum_{i=۱}^{n} x_i \psi\Big(\frac{Y_i - X_i^\top \beta}{d}\Big) = ۰ \quad i = ۰, ۱, ..., p \qquad (۳.۲)$$

فرض کنید بخواهیم ابتدا برآورد $\boldsymbol{\beta}$ را به دست 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 (مینیمم مجموع قدر مطلق خطاها) بتواند ضرایب رگرسیونی را به خوبی پیدا کند در این صورت کار ما را آسان‌تر می‌کند به عبارتی شبیه به میانه در یک نمونه واحد

است. برای برآورد استوار d حاصل از پارامتر مقیاس، میانه مقادیر قدر مطلق باقیمانده های غیر صفر را تقسیم بر ۰/۶۷۴۵ مانند پارامتر مکان می‌باشد. اگر چه آماردانان تلاش زیادی در زمینه L ـ برآوردگرها کردند اما با این حال ناراضی بودند بنابراین برای برآورد پارامتر β و مقیاس به صورت همزمان سراغ الگوریتم معرفی شده توسط داتر (۱۹۷۷) می‌رویم که این الگوریتم فقط برای تابع ψ در هوبر (۲۰۰۰) توضیح داده شده است. همان طور که مشاهده می کنیم این روش شبیه کمترین توان های دوم می باشد. البته باید به این نکته هم توجه کنیم که این روش فقط یکی از روش هایی است که در داتر (۱۹۷۷) ذکر شده است. ابتدا با برخی از برآوردهای اولیه β_0 و d_0 شروع می‌کنیم، و این دو را برآوردهای معمولی برای پارامتر β و مقیاس در نظر می‌گیریم که به صورت زیر عمل می کنیم

۱. باقیمانده‌ها را به صورت زیر محاسبه می کنیم

$$z_i = Y_i - X_i^\top \beta_0 \quad i = 1, 2, ..., n$$

۲. یک برآوردگر از پارامتر مقیاس به صورت زیر به دست می آوریم

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

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

۳. باقیمانده های وینزوریده[1] شده به صورت زیر تعیین می شوند

$$\triangle_i = \psi(\frac{z_i}{d_1}) d_1$$

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

۴. برآورد کمترین توان‌های دوم[2] ضرایب رگرسیونی به شرطی که باقیمانده های وینزوریده شده مشاهده شده باشند به صورت زیر می باشد

$$\hat{\xi}_0 = (X'X)^{-1} X'\Delta_0 \qquad (۴.۲)$$

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

۵. یک برآورد جدید از β محاسبه کنید یعنی

$$\beta = \beta_0 + q\hat{\xi}_0$$

[1] Winsorize
[2] Least Squares Estimates

داتر (۱۹۷۷) دریافت یک انتخاب مناسب از فاکتور q به شکل زیر می باشد

$$q = min[\frac{۱}{\Phi(k) - \Phi(-k), ۱/۹}]$$

که Φ تابع توزیع نرمال استاندارد می باشد.

با برآوردهای جدید $\beta_۱$ و $d_۱$ به عنوان مقادیر اولیه برای مراحل یک تا پنج تکرار می‌شوند. این روند تکراری را $i = ۱, ۲, ..., p$ ادامه می‌دهیم در این صورت

$$|q\hat{\xi}_m^i| < \varepsilon d_{m+۱} \quad |d_{m+۱} - d_m| < \varepsilon d_{m+۱}$$

که در آن $\varepsilon > ۰$ یک سطح تحمل مناسب، $\hat{\xi}_m^i$، i ـ امین مولفه از $\hat{\xi}$ و x_{ii} i امین عنصر قطری از $(X'X)^{-۱}$ است در این زمان تکرار را متوقف کنید و β و پارامتر مقیاس را با به بکار بردن $\beta_{m+۱}$ و $d_{m+۱}$ برآورد کنید. خواننده باید دقیقا مشخص کند که این روش برای کمترین توان های دوم است یا خیر.

البته مرحله چهارم کمترین توان های دوم برای باقیمانده های وینزوریده شده است و مرحله پنج کمترین توان دوم را با β شامل $q = ۱$ و $\Delta_۰$ هایی برابر حاصل از باقیمانده های واقعی (غیر وینزوریده شده) برای هر برآورد اولیه $\beta_۰$ ارایه می دهد. هم چنین جالب است که برای محاسبه ی $(X'X)$ نیاز به محاسبه‌ی یک بار روند تکراری است که این خود باعث صرفه جویی در محاسبات می شود سپس یک برآورد استوار خوب از $\beta_۰$ پارامتر مقیاس با استفاده از الگوریتم معرفی شده توسط داتر با روش های کمترین مقدار مطلق یا دیگر روش ها (برخی روش های ناپارامتری که ممکن است بسیار مناسب تر باشد) به دست آوریم. ما می توانیم مشکل نقاط پرت را با به کار بردن برآوردهای استوار $d_۰$ و $\beta_۰$ یک تابع صعودی ψ بر طرف کنیم مانند موجی یا وزنی در این موارد کمترین توان های دوم موزون می تواند جایگزین مناسبی برای استفاده باشد.

۱.۳.۲ M ـ برآوردگر

معادله (۲.۲) را نسبت به پارامترهای β_j، $j = ۱, ۲, .., k$ مینیمم کنیم با مشتق‌گیری از معادله (۲.۲) نسبت به تک تک پارامترها تعداد $p = k + ۱$ معادله به صورت زیر حاصل می‌شود

$$\sum_{i=۱}^{n} x_{ij} \psi \frac{(Y_i - X_i^\top \beta)}{d} = ۰ \qquad (۵.۲)$$

که $\mathbf{X'}_i = (X_{i۱}, X_{i۲}, ..., X_{ik})$ می‌باشد. 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

$$W_{i\beta} = \frac{\psi\left\{\frac{(Y_i - X_i^\top \beta)}{d}\right\}}{\frac{(Y_i - X_i^\top \beta)}{d}} \qquad (۶.۲)$$

در صورتی که دقیقا $\mathbf{Y}_i - \mathbf{X}_i^\top \beta$ اتفاق بیافتد معادله (۵.۲) را می‌توان به صورت زیر بازنویسی کرد

$$\sum_{i=1}^{n} x_{ij} W_{i\beta}(Y_i - \mathbf{X}_i^\top \beta) = 0 \quad j = 0, 1, 2, ..., n \tag{۷.۲}$$

آن‌گاه

$$\sum_{i=1}^{n} x_{ij} W_{i\beta} \mathbf{X'}_i^\top \beta = \sum_{i=1}^{n} x_{ij} W_{j\beta} Y_i \quad j = 0, 1, 2, ..., n \tag{۸.۲}$$

که می‌توان به صورت ماتریسی زیربازنویسی کرد

$$\mathbf{X'} W_\beta \mathbf{X} \beta = \mathbf{X'} W_\beta \mathbf{Y} \tag{۹.۲}$$

که در روابط بالا $\mathbf{W}_{\boldsymbol{\beta}=(W_{1\beta}, W_{2\beta}, ..., W_{n\beta})}$ این معادلات به وضوح به صورت معادلات نرمال کمترین توان‌های دوم تعمیم یافته هستند. (به طور دقیق‌تر، معادلات کمترین توان‌های دوم وزنی هستند زیرا W_β یک ماتریس قطری از وزن‌ها می‌باشد). مشکل حل کردن آن این است که W_β به $\boldsymbol{\beta}$ بستگی دارد. با استفاده از یک روش ساده قدیمی این مشکل را حل می‌نماییم به این صورت که یک برآورد اولیه $\hat{\beta}_0$ از $\boldsymbol{\beta}$ را در نظر می‌گیریم (با استفاده از کمترین توان‌های دوم) مقدار W_0 را محاسبه می‌کنیم، یعنی W_β برای همان $\hat{\beta}_0$ و معادله (۹.۲) را برای رسیدن به جواب $\hat{\beta}_1$ حل می‌کنیم. با استفاده از $\hat{\beta}_1$ به W_1 می‌رسیم. سپس با استفاده از W_1 به $\hat{\beta}_1$ می‌رسیم و به همین ترتیب ادامه می‌دهیم. معمولا همگرایی به سرعت رخ می‌دهد.(این روند می‌تواند بعد از یک تعداد منتخبی از گام‌ها متوقف شود). تنها یک برنامه برای کمترین توان‌های دوم احتیاج است، مثلا در $MINITAB$ می‌تواند مکرر به صورت زیر نوشته شود

$$\hat{\beta}_{q+1} = (\mathbf{x'} W_q \mathbf{X})^{-1} X' W_q \mathbf{Y} \tag{۱۰.۲}$$

و این روند ممکن است زمانی که همه‌ی برآوردها کمتر از یک مقدار از پیش تعیین شده مثلا ۰.۱ درصد یا ۱.۰ درصد، تغییر کنند یا بعد از یک تعداد گام تعیین شده، متوقف می‌شود. از معادله (۶.۲) مشاهده می‌شود که وزن‌ها به مقدار $\frac{(\dot{\rho}(x))}{x}$ وابسته‌اند که در آن $x = \frac{(Y_i - \mathbf{X}_i^\top \beta)}{d}$ می‌باشد. بنابراین می‌توانیم برای همه‌ی انتخاب‌های $\rho(x)$ که در جدول زیر آمده به تابع $W(x) = \frac{(\dot{\rho}(x))}{x} x$ نگاه کنیم. به وضوح انتخاب یک $\rho(x)$ مشخص مبنای انتخاب یک تابع وزنی است که کاربر برای اختصاص دادن به باقیمانده‌های مقیاس شده در انواع مقادیر مختلف نیاز دارد یادآوری می‌کنیم که وزن‌ها در (۶.۲) بر حسب $x = \frac{(Y_i - \mathbf{X}_i^\top \beta)}{d}$ تعریف شده‌اند و نیاز به برآورد مقیاس d دارد. یک انتخاب استوار که تا حدی محبوبیت پیدا کرده به صورت زیر است

$$d = \frac{median|e_i - median(e_i)|}{۰/۶۷۴۵} \qquad (۱۱.۲)$$

اگر n به قدر کافی بزرگ و خطاها به صورت نرمال توزیع شده باشند برآوردگری تقریبا نااریب که از $Y_i = \sigma$ به‌دست خواهد آورد. چندین عامل مقیاسی جایگزین نیز پیشنهاد و بررسی شده‌اند، برای مثال هیل و هولاند (۱۹۷۷) بررسی کردند انتخاب واقعی d چه اثری روی باقیمانده‌ها خواهد گذاشت، به این معنا که برآوردهای پارامتر رگرسیونی استوار بر اساس انتخاب d ثابت نیستند.

هلند و والش (۱۹۷۷) از مقایسه برآوردهای استوار حاصل از ۸ تابع ψ به‌دست آمده که از جمله آن‌ها بسته‌های آماری استوار که در مرکز کامپیوتری دفتر ملی تحقیقات اقتصادی توسعه یافته است. یکی از ویژگی‌های مطلوب تابع ψ این است که تابع موج اندروز یا تابع وزن دوتایی توکی برای شناسایی نقاط دورافتاده زمانی که ما از روش هوبر استفاده می‌کنیم بهتر می‌شوند و در بسیاری از موارد به نقاط پرت یا نقطه داده‌های بد وزن صفر داده می‌شود. البته هر زمان که روش‌های کمترین توان‌های دوم(تعمیم آن‌ها) به‌کار می‌روند به ظاهر امکان استفاده از روش‌های استواری نیز وجود دارد. که شامل $ANOVA$، رگرسیون، سری‌زمانی، اسپلاین، تجزیه و تحلیل چندمتغیره و تشخیص، برای اطلاعات بیشتر درمورد توابع ρ و توابع وزن به جدول صفحه‌ی بعد 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 ثابت‌های تنظیم کننده و
۱	حداقل مربعات (توزیع مانده‌های نرمال)	$\frac{۱}{۲}x^۲$	$\infty \le x \le \infty$	x	۱	
۲	Huber	$\begin{cases} \frac{۱}{۲}x^۲ \\ a\lvert u\rvert - \frac{۱}{۲}a^۲ \end{cases}$	$-a \le x \le +a$ $u < -a, x > a$	x $a * sign(x)$	۱ $\frac{a}{\lvert x\rvert}$	$a = ۲$
۳	Ramsay	$\frac{\left\{۱-e^{a\lvert x\rvert}(۱+a\lvert x\rvert)\right\}}{a^۲}$	$\infty \le x \le \infty$	$ue^{a\lvert x\rvert}$	$e^{a\lvert x\rvert}$	$a = ۰/۳$
۴	Andrew	$\begin{cases} a\left\{۱-\cos(\frac{x}{a})\right\} \\ ۲a \end{cases}$	$-a \le x\pi \le +a\pi$ $x < -a\pi, x > a\pi$	$\sin(\frac{x}{a})$ ۰	$\sin(\frac{x}{a})\frac{x}{a}$ ۰	$a = ۱/۳۳۹$
۵	Tukey	$\begin{cases} \frac{۱}{۲}x^۲ - \frac{x^۲}{۴a^۲} \\ \frac{۱}{۴}a^۲ \end{cases}$	$-a \le x \le +a$ $x < -a, x > a$	$x(۱-\frac{x^۲}{a^۲})$ ۰	$۱-\frac{x^۲}{a^۲}$ ۰	$a = ۵ \le a \le ۶$
۶	$Hampel$ $a, b, c > ۰$	$\begin{cases} \frac{۱}{۲}x^۲ \\ a\lvert x\rvert - \frac{۱}{۲}a^۲ \\ a\frac{(c\lvert x\rvert - \frac{۱}{۲}x^۲)}{c-b} - \frac{۷a^۲}{۶} \\ a(b+c-a) \end{cases}$	$-a \le x \le +a$ $-b \le x \le -a, a \le x \le b$ $-c \le x \le -b, b \le x \le c$ $x < -c, ux > c$	x $a \times sign(x)$ $csign(x) - (x)a/c$ ۰	۱ $\frac{a}{\lvert x\rvert}$ $(c/\lvert x\rvert - ۱)(a/c - b)$ ۰	$a = ۱/۷, b = ۳/۴, c = ۸/۵$

۲.۳.۲ L ـ برآوردگر

یک نمونه تصادفی از یک توزیع پیوسته با پارامتر مکان θ در نظر بگیرید آماره‌های ترتیبی نمونه، $X_{(۱)}, X_{(۲)}, ..., X_{(n)}$ یک L ـ برآوردگر ترکیب خطی از آماره‌های ترتیبی می‌باشد. که موارد زیر نمونه‌هایی از L ـ برآوردگرها می‌باشند

۱ـ میانه نمونه

$$\bar{X}_\alpha = \frac{\sum_{i=[n\alpha]+۱}^{n-[n\alpha]} X_{(i)}}{n - ۲[n\alpha]} \quad i = [n\alpha]+۱, ..., n-[n\alpha] \tag{۱۲.۲}$$

۲ـ برآوردگر گستویرث[1] که یک میانگین وزنی با صدکهای $۳۳\frac{۱}{۳}$، ۵۰، $۶۶\frac{۲}{۳}$ با وزن‌های ۰٫۳، ۰٫۴ و ۰٫۳ می‌باشد.

۳ـ توکی سه میانگین، که میانگین موزون اولین، دومین و سومین، چارک‌های مربوط به آن‌ها دارای وزن $\frac{۱}{۴}$، $\frac{۱}{۲}$ و $\frac{۱}{۴}$ می‌باشند.

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

$$\rho(x) = |x| \tag{۱۳.۲}$$

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

$$\psi(x) = \begin{cases} -(۱-p)x & x < ۰ \\ p(x) & x \geq ۰ \end{cases} \tag{۱۴.۲}$$

صدک (۱۰۰_p) ام در یک تک نمونه و برآوردی از سطح تراز صدک (۱۰۰_p) ام است، یا اینکه سطح وضعیت رگرسیون که برای اطلاعات بیشتر به کوکنر و باست (۱۹۷۸) مراجعه کنید.

[1] Gastwirth's Estimator

فصل ۳

آزمون‌های استوارمدل خطی با خطاهای وابسته زبرجمعی منفی

۱.۳ مقدمه

در این فصل، به آزمون فرضیه‌های پارامترهای رگرسیون در مدل‌های خطی با خطاهای وابسته زبرجمعی منفی NSD می‌پردازیم. توزیع مجانبی آماره آزمون بدست آمده و برآوردهایی از پارامترهای زیادی در توزیع بدون علامت ایجاد می‌شود. اطلاعات این فصل از یو و همکاران (۲۰۱۷) و هوبر (۱۹۸۱) گرفته شده است. مدل خطی رگرسیون زیر را در نظر بگیرید

$$Y_i = X_i'\beta + e_i \quad i = ۱,۲,...,n \tag{۱.۳}$$

که در (۱.۳) X_i بردار $p \times ۱$ معلوم می‌باشد، $\beta \in \mathbb{R}^p$ پارامتر مجهول رگرسیونی، $\mathbf{e_i}$ جمله خطاهای تصادفی هستند. $\hat{\beta}_n$ یک M ـ برآورد برای β به صورت زیر تعریف می‌شود

$$min \sum_{i=۱}^{n} \rho(e)_i = \sum_{i=۱}^{n} \rho(Y_t - X_i'\beta) \tag{۲.۳}$$

در رابطه (۲.۳) ρ تابعی محدب و غیریکنواخت می‌باشد، برای نمونه می‌توان از ρ هوبر استفاده کرد.

$$\rho(x) = \Big(x^{۲} I(|x| \le k)\Big)/۲ + \Big(k|x| - k^{۲}/۲\Big) I|x| > k, \quad k > ۰$$

رفتار مجانبی $\hat{\beta}_n$ برای انتخاب‌های مختلف تابع ρ توسط بای و همکاران (۱۹۹۲)، الش (۱۹۸۶)، یوهای (۱۹۷۴) و یوهای و مارونا (۱۹۷۹) بررسی شده است. غالب نتایج متکی به فرض استقلال باقیمانده‌هاست. برای مثال هوبر (۱۹۸۱) نشان داد شرط ضروری برای درستی نتایج، استقلال باقیمانده‌هاست. برلینت و همکاران (۲۰۰۰) با فرض باقیمانده‌های وابسته شرایط سازگاری M ـ برآوردها را در مدل‌های خطی با خطاهای آمیخته قوی را ارائه کردند. کوی و همکاران (۲۰۰۴) توزیع مجانبی M ـ برآوردها را در مدل‌های خطی با خطاهای همبسته فضایی مشخص کردند. وو (۲۰۰۷) در مورد نمایش‌های بهادر قوی و ضعیف برای M ـ برآوردها در مدل‌های خطی با جملات خطای وابسته تحقیق کرد. وو (۲۰۱۱) شرایط سازگاری قوی M ـ برآوردها را در مدل‌های خطی با خطاهای پیوندی منفی بررسی کرد.

هو(۲۰۰۰) با ارائه مثالی نشان داد که NSD به معنای پیوند منفی نیست، هم‌چنین برخی از خواص پایه‌ای و دو قضیه اساسی برای NSD ارائه داد و نشان داد که تعدادی از توزیع‌های چند متغیره شناخته شده دارای خاصیت NSD هستند. NSD توسط آلام و ساکسنا (۱۹۹۸) مطرح و این که NSD اشاره به NSD دارد، نلسن (۲۰۰۶) خاصیت وابستگی را می‌توان برای کاربرد نظریه مفصل اعمال کرد، جاورکی و همکاران (۲۰۱۳) با دید اقتصادی به متغیرهای تصادفی NSD نگاه کردند و بیشتر تمرکزشون روی کاربرد اقتصادی قرار دادن، کریسدوندز و واگلاتو (۲۰۰۴) متغیرهای وابسته‌ای که NSD باشند دارای خاصیت NSD نیز هستند مورد بررسی قرار دادند، اقبال و همکاران (۲۰۱۰) قانون قوی اعدادبزرگ برای فرم‌های درجه دوم در متغیرهای تصادفی NSD تحت این فرض که این متغیرها نامنفی هستند، و یک دنباله $\{X_n, n \ge ۱\}$

با $EX_n^r \le \infty$ برای هر $n \ge ۱$ و $r \ge ۱$ مطرح کردند، هم‌چنین اقبال و همکاران (۲۰۱۱) نامساوی کلموگروف را برای فرم‌های درجه دوم در دنباله نامنفی از متغیرهای تصادفی کراندار یکنواخت NSD ارائه 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.

هدف اصلی این بخش بررسی مساله آزمون استوار M برای پارامترهای رگرسیون در مدل $Y_i = X_i'\beta + e_i$ با خطاهای تصادفی NSD است. فرضیه‌های زیر را در نظر بگیرید:

$$H_۰ : \mathbf{H}^\top(\boldsymbol{\beta} - \mathbf{b}) = \mathbf{0} \quad H_۱ : \mathbf{H}^\top(\boldsymbol{\beta} - \mathbf{b}) \ne \mathbf{0} \qquad (۳.۳)$$

که $\mathbf{H}$ یک ماتریس معلوم $p \times q$ با رتبه q درجه، $(۰ < q \le p)$، $\mathbf{b}$ یک بردار p تایی معلوم است. فرضیه مقابل را می‌توان با فرضیه زیر جایگزین کرد:

$$H_{۲,n} : \mathbf{H}^\top(\boldsymbol{\beta} - \mathbf{b}) = \mathbf{H}^\top \omega_{\mathbf{n}} \qquad (۴.۳)$$

که ω_n بردار p معلوم است. به طوریکه

$$\|\mathbf{S_n^{\frac{1}{2}}}\omega_\mathbf{n}\| = O(1) \qquad (۵.۳)$$

که $\mathbf{S_n} = \sum_{\mathbf{t=1}}^{\mathbf{n}} \mathbf{x_t x_t}^\top$ و $\|.\|$ یک نرم اقلیدسی است. روابط زیر را در نظر بگیرید:

$$\min_{\mathbf{H}^\top(\boldsymbol{\beta}-\mathbf{b})=\mathbf{0}} \sum_{\mathbf{t=1}}^{\mathbf{n}} \boldsymbol{\rho}(\mathbf{y_t} - \mathbf{x_t}^\top \boldsymbol{\beta}) = \sum_{\mathbf{t=1}}^{\mathbf{n}} \boldsymbol{\rho}(\mathbf{y_t} - \mathbf{x_t}^\top \tilde{\beta})$$

$$\min_{\beta \in \mathbf{R^p}} \sum_{\mathbf{t=1}}^{\mathbf{n}} \boldsymbol{\rho}(\mathbf{y_t} - \mathbf{x_t}^\top \boldsymbol{\beta}) = \sum_{\mathbf{t=1}}^{\mathbf{n}} \boldsymbol{\rho}(\mathbf{y_t} - \mathbf{x_t}^\top \hat{\beta})$$

$$\mathbf{M_n} = \sum_{\mathbf{t=1}}^{\mathbf{n}} \boldsymbol{\rho}(\mathbf{y_t} - \mathbf{x_t}^\top \tilde{\beta}) - \sum_{\mathbf{t=1}}^{\mathbf{n}} \boldsymbol{\rho}(\mathbf{y_t} - \mathbf{x_t}^\top \hat{\beta})$$

. که $\hat{\beta}$ و $\tilde{\beta}$ M ـبرآوردگرهای پارامترهای مدل $Y_i = X_i'\beta + e_i$ در مدل محدود و بدون محدودیت است. برای آزمون فرضیه (۳.۳) ما معیار M را در نظر می‌گیریم که M_n را معیار اندازه‌گیری سطح انحراف از فرضیه صفر می‌داند. چندین نتیجه گیری کلاسیک زمانی که خطاها به نظر می‌رسد مستقل هستند، در مقاله‌های (۸)، (۵۴) و (۵۵) ارائه شده‌اند، این مورد را به خطاهای تصادفی NSD تعمیم می‌دهد. در این بخش C یک مقدار ثابت مثبت می‌باشد. $|\tau| = \max_{۱\le t\le p}\{|\tau_۱|, |\tau_۲|, ..., |\tau_p|\}$ به طوریکه τ یک بردار p تایی است. و هم‌چنین داریم $x^+ = xI(x \ge ۰)$ و $x^- = -xI(x \le ۰)$. یک دنباله تصادفی X_n گفته می‌شود روی نرم L_q، و $q > ۰$ است اگر $E|X_n|^q < \infty$.

ρ را یک تابع محدب و غیریکنوا فرض می‌کنیم، مشتقات چپ و راست تابع ψ را با ψ_- و ψ_+ نمایش می‌دهیم، $\psi_-(u) \le \psi(u) \le \psi_+(u)$.

چندین فرضیه به شرح زیر در نظر می‌گیریم:

- $A۱$: تابع $G(u) = E\psi(e_t + u)$ به طوریکه $G(۰) = E\psi(e_t) = ۰$ و تابع λ دارای مشتق مثبت $u = ۰$ است.
- $A۲$: $۰ < E\psi^۲(e_۱) = \sigma^۲ < \infty$ ، و $\lim_{u\to ۰} E|\psi(e_۱ + u) - \psi(e_۱)|^۲ = ۰$.
- $A۳$: مقدار ثابت مثبت Δ وجود دارد طوری که برای هر $h \in (۰, \Delta)$ و تابع

$$\psi(u + h) - \psi(u)$$

نسبت به u یکنوا است.

- $A۴$:

$$\sum_{t=۲}^{\infty} |Cov(\psi(e_۱), \psi(e_t))| < \infty$$

- $A۵$: از آن‌جایی که $S_n = \sum_{t=۱}^{n} x_t x_t^\top$ فرض می‌کنیم $S_n > ۰$، برای n به قدر کافی بزرگ

$$d_n = \max_{۱\le t\le n} x_t^\top S_n^{-۱} x_t = O(n^{-۱}).$$

ملاحظه ۱.۱.۳. شرایط A۱ تا A۴ استاندارد و اغلب در نظریه مجانبی M ـبرآوردها در مدل‌های رگرسیون خطی کاربرد دارند. برای مثال به ژائو (۲۰۰۲) مراجعه کنید.

ملاحظه ۲.۱.۳. شرط A۵ منطقی است زیرا معادل با کران $\max_{۱\le t\le n}|x_t x_t^\top|$، است، در این‌جا یک حالت شرط $d_n = O(n^{-\delta})$ برای $۰ < \delta \le ۱$، که بوسط اوانگ و همکاران (۲۰۱۵) مورد استفاده قرار گرفت. مثال‌هایی که در زیر آورده شده شرایط A۵ برای آن‌ها برقرار است و توابع $\psi(x)$ که در مثال‌های زیر آورده شده شرایط A۱ و A۵ برقرار است.

مثال ۱.۱.۳. برآوردگر کمترین توان‌های دوم[1] به صورت زیر می باشد

$$\rho(x) = x^۲$$

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

$$\psi(x) = ۲x.$$

مثال ۲.۱.۳. برآورد کمترین انحراف مطلق[2] به صورت زیر می‌باشد

$$\rho(x) = |x|$$

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

$$\psi(x) = sign(x) = \begin{cases} ۱ & x > ۰ \\ ۰ & x = ۰ \\ -۱ & x < ۰ \end{cases}$$

لم ۱.۱.۳. فرض کنید $\{X_n, n \ge ۰\}$ دنباله ای از متغیرهای تصادفی NSD باشد که $EX_n = ۰$ و برای $p \ge ۲$ و $E|X_n|^p < \infty$ آنگاه ثابت مثبت C_p وابسته به p وجود دارد به طوری که برای هر $n \ge ۱$

برای اثبات این قضیه می‌توانید به (۴۰) مراجعه کنید.

لم ۲.۱.۳. وجود دارد یک D زیرمجموعه محدب از $\mathbb{R}^n$ و $\{f_n\}$ تابع محدب روی D برای هر $x \in D$،

$$f_n(x) \xrightarrow{\mathbb{P}} f(x)$$

اگر f یک تابع واقعی روی D، و سپس f محدب باشد. و برای همه زیرمجموعه‌های فشرده $D_۰ \subset D$،

[1] Least Squares Estimator
[2] Absolute Deviation Estimator

$$\sup_{x\in D_\circ}|f_n(x)-f(x)|\xrightarrow{\mathbb{P}}\circ \qquad (۶.۳)$$

علاوه بر این اگر f یک تابع متمایز روی D، $g(x)$ و $g_n(x)$ به ترتیب گرادیان(شیب) و گرادیان فرعی از f را نشان می‌دهند. سپس رابطه (۶.۳) دلالت بر این دارد که برای همه $D_\circ$ داریم

$$\sup_{x\in D_\circ}|g_n(x)-g(x)|\xrightarrow{\mathbb{P}}\circ$$

برای اثبات این قضیه می‌توانید به (۷) مراجعه کنید.

لم ۳.۱.۳. فرض کنید $\{X_n, n\geq ۱\}$ یک دنباله تصادفی از توزیع NSD با واریانس محدود و آرایه‌ای از مقدار حقیقی $\{a_{nj}, ۱\leq j\leq n\}$ متقاعد است. $\sum_{j=۱}^{n}a_{nj}^{۲}=O(۱)$، $\max_{۱\leq j\leq n}|a_{nj}|\to\circ$. برای هر $r_j, j=۱,۲,...,n$

$$\Big|Eexp\Big(i\sum_{j=۱}^{n}r_jZ_{nj}\Big)-\prod_{j=۱}^{n}Eexp(ir_jZ_{nj})\Big|\leq\frac{۱}{۲}\sum_{j\neq\ell,\ell=۱}^{n}|r_jr_\ell Cov(Z_{nj},Z_n\ell)|$$

وجود دارد $Z_{nj}=\sum_{\ell\in\gamma_j}a_{n\ell}X_\ell$، $\{\gamma_j\}$ زیرمجموعه‌های ناپیوسته از $\{۱,۲,...,n\}$

برهان: برای دوتایی متغیرهای تصادفی NSD، X,Y با استفاده از ویژگی ۳ در لم ۲.۲.۱ داریم

$$H(x,y)=P(X\leq x,Y\leq y)-P(X\leq x)P(Y\leq y)\leq\circ. \qquad (۷.۳)$$

اگر $F(x,y)$ تابع توزیع توام از (X,Y) و $F_X(x),F_Y(y)$ تابع توزیع حاشیه‌ای از X,Y داریم:

$$\begin{aligned}Cov(X,Y)=E(XY)-E(X)E(Y) &= \int\int[F(x,y)-F_X(x)F_Y(y)]dxdy\\ &=\int\int H(x,y)dxdy\end{aligned} \qquad (۸.۳)$$

از رابطه فوق بدست می‌آوریم:

$$Cov(f(X),g(Y))=\int\int f'(X)g'(Y)H(x,y)dxdy$$

وجود دارد f,g توابع برآورد شده روی R به طوریکه $f',g'<\infty$ از ترکیب رابطه (۷.۳) و (۸.۳) بدست می‌اوریم:

$$|Cov(f(X),g(Y))|\leq\int\int|f'(X)||g'(Y)||H(x,y)|dxdy\leq\|f'\|_\infty\|g'\|_\infty|Cov(X,Y)|. \qquad (۹.۳)$$

به طوریکه $f(X)=exp(irX), g(Y)=exp(iuY)$ به راحتی دیده می‌شود

$$\|f'(X)\|_\infty \le r<\infty, \quad \|g'(Y)\|_\infty \le u<\infty$$

در نتیجه برای هر مقادیر r,u داریم

$$|Eexp(irX+iuY)-Eexp(irX)Eexp(iuY)| \le |ruCov(X,Y)| \qquad (۱۰.۳)$$

این لم برای $n=۱$ بدیهی است و $n=۲$ برای رابطه (۱۰.۳) صدق می‌کند. فرض کنید نتیجه برای همه $n\le M\,(n\ge ۳)$. برای $n=M+۱$ وجود دارد $\epsilon^۲=۱$، $\delta^۲=۱$، $k\in ۱,۲,...,M$ به طوریکه

$$\begin{cases} \epsilon r_j \ge ۰, & ۱\le j\le k \\ \delta r_j \ge ۰, & ۱+k\le j\le M+۱ \end{cases}$$

اگر $X=\sum_{j=۱}^{k}\epsilon r_j Z_{nj}$, $Y=\sum_{j=k+۱}^{M+۱}\delta r_j Z_{nj}$ آن‌گاه

$$\sum_{j=۱}^{n} r_j Z_{nj}=\epsilon X+\delta Y.$$

توجه داشته باشید X,Y توابع ناصعودی هستند. با توجه به فرضیات داریم:

$$\begin{aligned}
&\Big|Eexp\Big(i\sum_{j=۱}^{n} r_j Z_{nj}\Big)-\prod_{j=۱}^{n} Eexp(ir_j Z_{nj})\Big| \\
&\le |E(T_۱T_۲)-E(T_۱)E(T_۲)|+\Big|E(T_۱)E(T_۲)-E(T_۱)\prod_{j=۱+k}^{M+۱}E(R)\Big| \\
&+\Big|E(T_۱)\prod_{j=۱+k}^{M+۱}E(R)-\prod_{j=۱}^{k}E(R)\Big| \\
&\le |\epsilon\delta||Cov(X,Y)|+\Big|E(T_۲)-\prod_{j=۱+k}^{M+۱}E(R)\Big|+\Big|E(T_۱)-\prod_{j=۱+k}^{M+۱}E(R)\Big| \\
&\le \Big|Cov\Big(\sum_{j=۱}^{k}\epsilon r_j Z_{nj},\sum_{j=k+۱}^{M+۱}\delta r_\ell Z_{n\ell}\Big)\Big|+\frac{۱}{۲}\sum_{j\ne\ell,j,\ell=k+۱}^{M+۱}|r_j r_\ell Cov(Z_{nj},Z_{n\ell})| \\
&+\frac{۱}{۲}\sum_{j\ne\ell,j,\ell=۱}^{M+۱}|r_j r_\ell Cov(Z_{nj},Z_{n\ell})| \\
&\le \frac{۱}{۲}\sum_{j\ne\ell,j,\ell=۱}^{n}|r_j r_\ell Cov(Z_{nj},Z_{n\ell})|
\end{aligned}$$

به طوریکه $T_۱=exp(i\epsilon X)$, $T_۲=exp(i\delta Y)$, $R=exp(ir_j Z_{nj})$. در نتیجه اثبات لم کامل می‌شود.

لم ۴.۱.۳. اگر $X_{nj} \xrightarrow{L} X_j$، $X_j \xrightarrow{L} X$ برای هر j و اندازه یکنواخت ϱ و برای هر $\varepsilon > ۰$،

$$\lim_{j\to\infty}\lim_{n\to\infty} sup\{\varrho(X_{nj}, Y_n) \geq \varepsilon\} = ۰$$

سپس

$$Y_n \xrightarrow{L} X.$$

برای اثبات این قضیه می‌توانید به (۱۰) مراجعه کنید.

لم ۵.۱.۳. فرض کنید $\{X_n, n \geq ۱\}, \{a_{nj} ۱ \leq j \leq n\}$ همان شرایط لم از ۳.۱.۳ را داشته باشند فرض کنید $\{X_n, n \geq ۱\}$ از خانواده انتگرال از $L_۲ - norm$ سپس

$$\sigma_n^{-۱} \sum_{j=۱}^{n} a_{nj} X_j \xrightarrow{D} N(۰, ۱)$$

که $\sigma_n^۲ = var(\sum_{j=۱}^{n} a_{nj} X_j)$.

برهان: بدون از دست دادن کلیت، فرض می‌کنیم $a_{nj} = ۰$ برای همه $j > n$. با استفاده از رابطه (۸.۳)، $Cov(X, Y) \leq ۰$ سپس برای $۱ \leq m \leq n - ۱$ داریم:

$$\begin{aligned}
&\sum_{l,j=۱,|l-j|\geq m}^{n} |a_{nl} a_{nj} cov(X_l, X_j)| = ۲ \sum_{l,j=۱,l-j\geq m}^{n} |a_{nl} a_{nj} cov(X_l, X_j)| \\
&\leq \sum_{l,j=۱,l-j\geq m}^{n} (a_{nl}^۲ + a_n^۲ j)|cov(X_j, X_l)| \\
&\leq \sum_{l,j=۱,|l-j|\geq m}^{n} ۲ a_{nj}^۲ |cov(X_j, X_l)| \\
&= \sum_{l,j=۱}^{n} \sum_{|l-j|\geq m}^{n} a_{nj}^۲ |cov(X_j, X_l)| \\
&= \sum_{j=۱}^{n} a_{nj}^۲ \sum_{|l-j|\geq m}^{n} |cov(X_j, X_l)| \\
&\leq \sum_{j=۱}^{n} a_{nj}^۲ (\sup_j \sum_{|l-j|\geq m, l=۱}^{n} |cov(X_j, X_l)|) \\
&= \sum_{j=۱}^{n} a_{nj}^۲ (\sup_j |\sum_{|l-j|\geq m}^{n} cov(X_j, X_l)|) \\
&= \sup_j |\sum_{l=۱,|l-j|\geq m}^{n} cov(X_l, X_j)| (\sum_{j=۱}^{n} a_{nj}^۲)
\end{aligned}$$

اگر در شرط $A۴$ و $\psi(x)=x$ در نظر بگیریم برای همه $\ell \geq ۱$ و j های به قدر کافی بزرگ داریم

$$\sum_{j:|l-j|\geq m}^{\infty} \Big|Cov(X_\ell, X_j)\Big| \to ۰.$$

بنابراین، برای ε کوچک ثابت، وجود دارد عدد صحیح مثبت $m=m_\varepsilon$ چنان که داریم

$$\sum_{j,\ell=۱,|\ell-j|\geq m} \Big|a_{n\ell}a_{nj}Cov(X_\ell, X_j)\Big| \leq \varepsilon \qquad (۱۱.۳)$$

$N_۰=[۱/\varepsilon]$، در نظر می‌گیریم که $[x]$ قسمت صحیح x است، $Y_{nj}=\sum_{k=mj+۱}^{m(j+۱)} a_{nk}X_k,\ j=$ $۰,۱,۲,...,n,$

$$\gamma_j=\Big\{j: ۲N_۰\ell \leq j \leq ۲N_۰\ell+N_۰, |Cov(Y_{nj},Y_{nj+۱})| \leq \frac{۲}{N_۰}\sum_{j=۲N_۰\ell}^{۲N_۰\ell+N_۰} Var(Y_{n\ell})\Big\}$$

تعریف می‌کنیم

$$s_۰=۰, s_{j+۱}=min\{s: s>s_j, s\in\gamma_j\}$$

با بکار بردن روابط زیر داریم

$$Z_{nj}=\sum_{\ell=s_j+۱}^{s_{j+۱}} Y_{n\ell}, j=۰,۱,۲,...,n$$

$$\Lambda_j=\{m(s_j+۱)+۱,...,m(s_{j+۱}+۱)\}$$

توجه کنید که

$$Z_nj=\sum_{\ell\in\Lambda_j} a_{n\ell}X_\ell, j=۰,۱,۲,...,n$$

به راحتی می‌توان دید $\#\Lambda_j \leq ۳N_۰m,$ وجود دارد که در آن $\#$ برابر تعداد عناصر یک مجموعه است. اثبات خواهیم کرد که $\{Z_{nj}, ۱\leq j\leq n\}$، شرایط لیندبرگ را دارد. از آنجا که $B_n^۲=\sum_{j=۱}^n EZ_{nj}^۲$

است با استفاده از لم ۱.۱.۳ پس داریم

$$
\begin{aligned}
B_n^{۲} &= \sum_{j=۱}^{n} E\Big(\sum_{\ell\in\Lambda_j} a_{n\ell}X_\ell\Big)^{۲} \\
&\le \sum_{j=۱}^{n} a_{n\ell}^{۲} E\Big(\sum_{\ell\in\Lambda_j} X_\ell\Big)^{۲} \\
&\le \sum_{j=۱}^{n}\sum_{\ell\in\Lambda_j} a_{n\ell}^{۲} E(X_\ell)^{۲} \\
&\le \sum_{j=۱}^{n} E(a_{n\ell X_j})^{۲} = \sigma_n^{۲} < \infty
\end{aligned}
$$

طبق لم ۳.۱.۳ واریانس $E(X_n)^{۲} < \infty$ و $\sum_{j=۱}^{n} = O(۱).$ پس طبق لم ۳.۱.۳ $E(X_j)^{۲}$ واریانس متناهی است و وقتی واریانس متناهی باشد E هم متناهی است. گوییم $\{Z_{nj}^{۲}\}$ یکنواخت انتگرال پذیر است .

از آنجایی که $\{X_j^{۲}, j \ge ۱\}$ به طور یکنواخت انتگرال پذیر است. از این رو برای هر ε مثبت، داریم $\{Z_{nj}, ۱ \le j \le n\}$ بنابراین

$$
\begin{aligned}
\frac{۱}{B_n^{۲}}\sum_{j=۱}^{n} EZ_{nj}^{۲} I\{|z_{nj} \ge \varepsilon B_n\} &\le \sum_{j=۱}^{n}\Big(\sum_{\ell\in\Lambda_j} a_{n\ell}^{۲}\Big)\max_{\ell\in\Lambda_j} EX_\ell^{۲} I\Big\{\sum_{\ell\in\Lambda_j}|X_\ell| \ge \varepsilon/\max_{\ell\in\Lambda_j}|a_{n\ell}|\Big\} \\
&\le \Big(\sum_{j=۱}^{n} a_{n\ell}^{۲}\Big)\max_{۱\le j\le n} EX_\ell^{۲} I\Big\{\sum_{\ell\in\Lambda_j} X_\ell^{۲} \ge \varepsilon^{۲}/\Big(\max_{\ell\in\Lambda\ell}|a_{nl}|\Big)^{۲}\Big\} \\
&\le \Big(\sum_{j=۱}^{n} a_{n\ell}^{۲}\Big)\max_{۱\le j\le n} E\sum_{\ell\in\Lambda_j} X_\ell^{۲} I\Big\{\sum_{\ell\in\Lambda_j} X_\ell^{۲} \ge \varepsilon^{۲}/\Big(\max_{\ell\in\Lambda\ell}|a_{nl}|\Big)^{۲}\Big\}
\end{aligned}
$$

حال $\{Z_{nj}^{*}, j = ۱,۲,...,n\}$ دنباله مستقل تصادفی که دارای توزیع حاشیه‌ای یکسانی با Z_{nj} برای هر j. و $F(Z_{n\ell ۱}, Z_{n\ell ۲}, ..., Z_{nn})$ و $G(Z_{n\ell}^{*}۱, Z_{n\ell ۲}^{*}, ..., Z_{nn}^{*})$ توابع ویژه به ترتیب برای $\sum_{j=۱}^{n} Z_{nj}$ و $\sum_{j=۱}^{n} Z_{nj}^{*}$. با انتخاب $r = max(r_\ell, r_j)$، با استفاده از لم ۳.۱.۳ و رابطه (۱۱.۳) داریم

$$
\begin{aligned}
|F-G| &= \Big|Eexp\Big(i\sum_{j=۱}^{n} r_j Z_{nj}\Big) - Eexp\Big(i\sum_{j=۱}^{n} r_j Z^*_{nj}\Big)\Big| \\
&= \Big|Eexp\Big(i\sum_{j=۱}^{n} r_j Z_{nj}\Big) - \prod_{j=۱}^{n} Eexp(ir_j Z_{nj})\Big) \\
&\le \frac{۱}{۲}\sum_{j\neq \ell, j,\ell=۱}^{n} |r_j r_\ell Cov(Z_{nj}, Z_{n\ell})| \\
&\le r^۲\Big(\sum_{۱\le \ell<j\le n, |\ell-j=۱}^{n} |Cov(Z_{nj}, Z_{n\ell})| \sum_{۱\le \ell<j\le n, |\ell-j>۱}^{n} |+Cov(Z_{nj}, Z_{n\ell})|\Big) \\
&\le r^۲\Big(\sum_{j=۱}^{n} |Cov(Y_{n\ell s}, Y_{ns_{\ell+۱}}) + \sum_{۱\le \ell<j\le n, |\ell-j>m}^{n} |a_{n\ell}a_{nj}|Cov(X_{nj}, X_{n\ell})|\Big) \\
&\le r^۲\Big(\frac{C}{N_\circ}\sum_{j=۱}^{n} Var(Y_{nj}) + \varepsilon\Big) \le \varepsilon(r^۲ + C\sigma_n^۲)
\end{aligned}
$$

با جمع‌بندی قضیه‌ها، Z^*_{nj} حالت نرمال مجانبی را کسب می‌کند. با استفاده از لم ۴.۱.۳، توزیع مشخصه به طور یکسان از $\{X_j\}$ نتیجه می‌گیریم

$$B_n^{-۱}\sum_{j=۱}^{n} Z_{nj} = \sigma_n^{-۱}\sum_{j=۱}^{n} a_{nj}X_j \xrightarrow{D} N(\circ, ۱)$$

بنابراین اثبات لم کامل می‌شود.

لم ۶.۱.۳. در مدل $Y_i = X_i'\beta + e_i \quad i = ۱,۲,...,n$ فرض می‌کنیم دنباله $\{e_i, ۱ \le i \le n\}$ متغیر تصادفی NSD به طور یکسان شرایط $A۱ - A۴$ ایجاد کردند. برای هر δ مثبت، n بزرگ ایجاد کردند، سپس

$$\sup_{|\varsigma|\le\delta}\Big|\sum_{t=۱}^{n}\rho(e_t - x_n t^\top\zeta) - \rho(e_t) + \psi(e_t)x_n t^\top\zeta^\top - \frac{۱}{۲}\lambda\zeta^\top\Big| \underset{\rightarrow}{p} \circ,$$

$$\sup_{|\varsigma|\le\delta}\Big|\sum_{t=۱}^{n}\psi(e_t - x_n t^\top\zeta) - \psi(e_t)x_n t^\top + \lambda\zeta\Big| \underset{\rightarrow}{p} \circ.$$

و ζ یک بردار p تایی است.

برهان:

$$
\begin{aligned}
f_n(\zeta) &= \sum_{t=۱}^{n}\rho(e_t - x_n t^\top\zeta) - \rho(e_t) + \psi(e_t)x_n t^\top\zeta \\
&= \sum_{t=۱}^{n}\int_\circ^{-} x_n t^\top\zeta\{\psi(e_t + u) - \psi(e_t)\}du
\end{aligned}
$$

برای ζ ثابت، از شرط $A۵$ پیروی می‌کند چنانچه

$$\max_{۱\le t\le n}|x_n t^\top \zeta| \to O(n^{-۱/۲}) \qquad (۱۲.۳)$$

از شرط $A۱$ و رابطه‌ی فوق دنباله‌ای از مقادیر مثبت $\varepsilon_n \to ۰$ و $\theta_{nt} \in (-۱,۱)$ وجود دارد، چنانچه، برای n های به قدر کافی بزرگ داریم

$$\begin{aligned} Ef_n(\zeta) &= \sum_{t=۱}^{n}\int_{۰}^{-x_n t^\top \zeta} E(\psi(e_t+u)-\psi(e_t))du \\ &= \sum_{t=۱}^{n}\int_{۰}^{-x_n t^\top \zeta} \{\lambda u + o(|u|)\}du \\ &= \frac{۱}{۲}\lambda\sum_{t=۱}^{n}(x_n t^\top \zeta)^۲(۱+\varepsilon_n\theta_{nt}) \to \frac{۱}{۲}\lambda\zeta^\top\zeta. \end{aligned}$$

از دیدگاه همگرایی $\psi(e_t+u)-\psi(e_t)$، مجموعه‌ای از $f_n(\zeta)$ همچنین همگرایی در رابطه با e_t از ویژگی ۲ در لم ۲.۲.۱ با جدا کردن مجموعه از $f_n(\zeta)$ در دو قسمت منفی و مثبت، با استفاده از ویژگی ۱ در لم ۲.۲.۱ که NSD هستند. پس با استفاده از نامساوی شوارتز و رابطه (۱۲.۳) بدست می‌آوریم

$$\begin{aligned} var[f_n(\zeta)] &= E\Big\{\sum_{t=۱}^{n}\Big[\int_{۰}^{-x_n t^\top \zeta}(\psi(e_t+u)-\psi(e_t))du]^+ \\ &\quad -\sum_{t=۱}^{n}\Big[\int_{۰}^{-x_n t^\top \zeta}(\psi(e_t)+u)-\psi(e_t)]^- du\Big\}^۲ \\ &\le E\Big\{\sum_{t=۱}^{n}\Big[\int_{۰}^{-x_n t^\top \zeta}(\psi(e_t+u)-\psi(e_t))du]^+\Big\}^۲ \\ &\quad + E\Big\{\sum_{t=۱}^{n}\Big[\int_{۰}^{-x_n t^\top \zeta}(\psi(e_t+u)-\psi(e_t))du]^-\Big\}^۲ \\ &\le \sum_{t=۱}^{n}E[\int_{۰}^{-x_n t^\top \zeta}(\psi(e_t+u)-\psi(e_t))du]^۲ \\ &\le \sum_{t=۱}^{n}|x_{nt}^\top|\Big|\int_{۰}^{-x_n t^\top \zeta}E[\psi(e_t+u)-\psi(e_t)]^۲ du\Big| \\ &= o(۱)\sum_{t=۱}^{n}(x_{nt}^\top)^۲ \to ۰. \end{aligned}$$

بنابراین برای n های به قدر کافی بزرگ داریم

$$f_n(\zeta) \xrightarrow{p} \frac{۱}{۲}\lambda\zeta^\top\zeta. \qquad (۱۳.۳)$$

با استفاده از رابطه ۱۳.۳ و لم ۲.۱.۳ اثبات شد.

لم ۷.۱.۳. تحت شرایط از لم ۶.۱.۳ و با جایگزین کردن رابطه $H_{۲,n} : H^\top(\beta - b) = H^\top \omega_n$ و $\|S_n^{-۱/۲}\omega_n\| = O(۱)$ برای هر δ ثابت مثبت و n های به قدر کافی بزرگ.

$$\sup_{|\xi_۱|\le\delta} \Big| \sum_{t=۱}^{n} \{\rho(y_{nt} - x_{nt}^\top \eta) - \rho(e_t) + \psi(e_t) x_{nt}^\top \xi_۱\} - \frac{۱}{۲}\lambda\|\xi_۱\|^۲ \Big| \xrightarrow{p} ۰ \tag{۱۴.۳}$$

$$\sup_{|\xi_۱|\le\delta} \Big| \sum_{t=۱}^{n} [\psi(y_{nt} - x_{nt}^\top \eta) - \psi(e_t)] x_{nt}^\top + \lambda\xi_۱ \Big| \xrightarrow{p} ۰ \tag{۱۵.۳}$$

$$\sup_{|\xi_۲|\le\delta} \Big| \sum_{t=۱}^{n} \{\rho(y_{nt} - x_{nt}^\top K_n \zeta) - \rho(e_t)\} + \sum_{t=۱}^{n} \psi(e_t) x_{nt}^\top (K_n\zeta - \beta(n)) - \frac{۱}{۲}\lambda\|\xi_۲\|^۲ + \|\omega(n)\|^۲ \Big| \xrightarrow{p} ۰ \tag{۱۶.۳}$$

$$\sup_{|\xi_۲|\le\delta} \Big| \sum_{t=۱}^{n} [\psi(y_{nt} - x_{nt}^\top K_n \zeta) - \psi(e_t)] x_{nt} + \lambda(K_n\zeta - \beta(n)) \Big| \xrightarrow{p} ۰ \tag{۱۷.۳}$$

از آنجایی که $\xi_۱ = \eta - \beta(n), \xi_۲ = \zeta - \tau(n)$

اثبات رابطه (۱.۳):

$$\begin{aligned}
&\sup_{|\xi_۱|\le\delta} \Big| \sum_{t=۱}^{n} \{\rho(y_{nt} - X_{nt}^\top(\xi_۱ + \beta(n))) - \rho(e_t) + \psi(e_t) X_{nt}^\top \xi_۱\} - \frac{۱}{۲}\lambda\|\xi_۱\|^۲ \Big| \\
&= \sup_{|\xi_۱|\le\delta} \Big| \sum_{t=۱}^{n} \{\rho(y_{nt} - X_{nt}^\top \beta(n) - X_{nt}^\top \xi_۱) - \rho(e_t) + \psi(e_t) X_{nt}^\top \xi_۱\} - \frac{۱}{۲}\lambda\|\xi_۱\|^۲ \Big| \\
&= \sup_{|\xi_۱|\le\delta} \Big| \sum_{t=۱}^{n} \{(e_t - X_{nt}^\top \xi_۱) - \rho(e_t) + \psi(e_t) X_{nt}^\top \xi_۱\} - \frac{۱}{۲}\lambda\|\xi_۱\|^۲ \Big| \xrightarrow{p} ۰
\end{aligned}$$

اثبات رابطه (۲.۳):

$$\begin{aligned}
&\sup_{|\xi_۱|\le\delta} \Big| \sum_{t=۱}^{n} [\psi(y_{nt} - x_{nt}^\top(\xi_۱ + \beta(n)) - \psi(e_t)] X_{nt}^\top + \lambda\xi_۱ \Big| \\
&= \sup_{|\xi_۱|\le\delta} \Big| \sum_{t=۱}^{n} [\psi(y_{nt} - X_{nt}^\top \beta(n) - X_{nt}^\top \xi_۱) - \psi(e_t)] X_{nt}^\top + \lambda\xi_۱ \\
&= \sup_{|\xi_۱|\le\delta} \Big| \sum_{t=۱}^{n} [\psi(e_t - X_{nt}^\top \xi_۱) - \psi(e_t)] X_{nt}^\top + \lambda\xi_۱ \Big| \xrightarrow{p} ۰
\end{aligned}$$

اثبات رابطه‌ی (۱۴.۳) و (۱۵.۳) مشابه اثبات رابطه‌ی (۱۶.۳) و (۱۷.۳) هستند.

●اثباتی از قضیه‌ها

با در نظر گرفتن مدل مجدد

$$\mathbf{y_{nt}} = \mathbf{x_{nt}^{\top}}\beta(\mathbf{n}) + \mathbf{e_t} \quad \mathbf{t = 1, 2, ..., n} \tag{۱۸.۳}$$

از آنجایی که $\mathbf{x_{nt}} = \mathbf{S_n^-1/2x_t}$، $\boldsymbol{\beta(n)} = \mathbf{S_n^1}/2(\boldsymbol{\beta} - \mathbf{b})$، $y_{nt} = y_t - \mathbf{x_t^{\top}b}$. به راحتی می‌توان آن را بررسی کرد

$$\sum_{t=1}^{n} \left\| \mathbf{x_{nt}x_{nt}^{\top}} \right\| = \mathbf{p} \tag{۱۹.۳}$$

فرض کنید $q < p$ باشد. ماتریس $K_p \times (p-q)$ با $(p-q)$ درجه آزادی چنانچه که $\mathbf{H^{\top}K = 0}$ و $\mathbf{K^{\top}\omega_n = 0}$ سپس برای برخی از $\gamma \in R^{p-q}$، که $H_\circ$ و $H_{۲,n}$ را می‌توان به صورت زیر نوشت

$$H_\circ : \boldsymbol{\beta} - \mathbf{b} = \mathbf{K}\gamma \quad \mathbf{H_{2,n}} : \boldsymbol{\beta} - \mathbf{b} = \mathbf{K}\gamma + \omega_\mathbf{n} \tag{۲۰.۳}$$

دلالت بر اینکه $\mathbf{H_n} = \mathbf{S_n^{-1/2}H(H^{\top}S_n^{-1}H)^{-1/2}}, \mathbf{K_n} = \mathbf{S_n^{1/2}K(K^{\top}S_nK)^{-1/2}}$ سپس

$$\mathbf{H_n^{\top}H_n = I_q}, \quad \mathbf{K_n^{\top}K_n = I_{p-q}}, \quad \mathbf{H_n^{\top}K_n = 0}, \quad \mathbf{H_nH_n^{\top} + K_nK_n^{\top} = I_p} \tag{۲۱.۳}$$

با توجه به اینکه $\gamma_\circ(n) = (\mathbf{K^{\top}S_nK})^{1/2}\gamma$. تحت فرضیه صفر مدل (۱۸.۳) می‌توان به صورت زیر بازنویسی کرد

$$y_{nt} = \mathbf{x_{nt}^{\top}K_n\gamma_0(n)} + \mathbf{e_t} \quad \mathbf{t = 1, 2, ..., n}$$

مجموعه $\omega(n) = \mathbf{H_n^{\top}S_n^{1/2}}\omega_\mathbf{n}, \gamma_(\mathbf{n}) = \boldsymbol{\gamma_0(n)} + \mathbf{K_n^{\top}S_n^{1/2}}\omega_\mathbf{n}$، تحت جایگزین‌های مکانی (۲۰.۳).

$$\boldsymbol{\beta}(n) = \mathbf{K_n}\gamma(\mathbf{n}) + \mathbf{H_n}\omega(\mathbf{n}) \tag{۲۲.۳}$$

تعریف می‌کنیم $\hat{\boldsymbol{\beta}}(n) = \mathbf{S}_n^{۱/۲}(\hat{\beta} - \mathbf{b})$ و $\hat{\gamma}(n)$ حاصل می‌شود

$$\min_{\varsigma \in R^{p-q}} \sum_{t=۱}^{n} \rho(y_{nt} - \mathbf{x_{nt}^{\top}K_n}\boldsymbol{\zeta}) = \sum_{\mathbf{t=1}}^{\mathbf{n}} \rho(\mathbf{y_{nt} - x_{nt}^{\top}K_n}\hat{\gamma}(\mathbf{n}))$$

$\hat{\boldsymbol{\beta}}(\boldsymbol{n}), \hat{\boldsymbol{\gamma}}(\boldsymbol{n})$ به ترتیب M ـ برآوردهایی از $\boldsymbol{\beta}(n)$ و $\boldsymbol{\gamma}(n)$ هستند. بنابراین

$$\tilde{\beta} = \mathbf{b} + \mathbf{S}_n^{-۱/۲}\mathbf{K}_n\hat{\gamma}(n)$$

قضیه ۱.۱.۳. در مدل $Y_{it} = X_i'\beta + e_i, i = ۱,۲,...,n$ فرض کنید $\{e_t, ۱ \le t \le n\}$ که یک دنباله از توزیع NSD متغیر تصادفی یک خانواده کامل است. از $L^۲ - norm$ و شرایط $A۱ - A۵$ را شامل می‌شوند. سپس $۲\lambda\sigma^{-۲}M_n$ مجانب مرکزی نیست توزیع $\chi^۲$ با p درجه آزادی است به عبارت دیگر پارامتر $v(n)$ مرکزی نیست داریم

$$۲\lambda\sigma^{-۲}M_n \xrightarrow{D} \chi^۲_{p,v(n)}$$

وجود دارد $v(n) = \lambda^۲\sigma^{-۲}\|w(n)\|$ ، $w(n) = H_n^\top S_n^{۱/۲} w(n)$ ، $H_n = S_n^{-۱/۲}H(H^\top S_n^{-۱/۲}H)^{-۱/۲}$.
به طور خاص زمانی که $\|S_n^{۱/۲}w(n)\| \to ۰$ به این معنی است که پارامترهای واقعی از صفر منحرف شده‌اند. پس $۲\lambda\sigma^{-۲}M_n$ توزیع مربع $\chi^۲$ با p درجه آزادی

$$۲\lambda\sigma^{-۲}M_n \xrightarrow{D} \chi^۲_p$$

برای یک سطح معنی‌دار مشخص، ما می‌توانیم ناحیه رد را به صورت زیر تعیین کنیم:

$$W = (۰, \chi^۲_p(۱-\alpha/۲)) \cup (\chi^۲_p(\alpha/۲), +\infty) \qquad (۲۳.۳)$$

به طوریکه $\chi^۲_p(۱-\alpha/۲)$ ، $\chi^۲_p(\alpha/۲)$ به ترتیب چندک $\alpha/۲$ از $\chi^۲$ مرکزی توزیعی با p درجه آزادی.

اثبات: طبق رابطه $|\hat{\beta}n| = O_p(۱)$ و رابطه زیر در لم ۷.۱.۳ داریم

$$\sup_{|\xi_۲|\le\delta}\Big|\sum_{t=۱}^{n}\{\rho(y_{nt}-x_{nt}^\top K_n\zeta)-\rho(e_t)\}+\sum_{t=۱}^{n}\psi(e_t)x_{nt}^\top(K_n\zeta-\beta(n))-\frac{۱}{۲}\lambda\|\xi_۲\|^۲+\|\omega(n)\|^۲\Big| \xrightarrow{p} ۰$$

$$\begin{aligned}&\sum_{t=۱}^{n}[\rho(y_{nt}-x_{nt}^T\hat{\beta}(n)-\rho(e_t)]+\sum_{t=۱}^{n}\psi(e_t)x_{nt}(\hat{\beta}-\beta(n))-\frac{\lambda}{۲}\|\hat{\beta}-\beta(n)\|^۲ \xrightarrow{p} ۰\\ &,\sum_{t=۱}^{n}[\rho(y_{nt}-x_{nt}^\top\hat{\beta}(n)-\rho(e_t)]+\frac{۱}{۲\lambda}\Big\|\sum_{t=۱}^{n}\psi(e_t)x_{nt}\Big\|^۲ \xrightarrow{p} ۰.\end{aligned} \qquad (۲۴.۳)$$

با استفاده از رابطه بالا داریم

$$\begin{aligned}&\sum_{t=۱}^{n}[\rho(y_{nt}-x_{nt}^\top K_n\hat{\gamma}(n)-\rho(e_t)]-\sum_{t=۱}^{n}\psi(e_t)x_{nt}^\top\Big(K_n\hat{\gamma}(n)-K_n\gamma(n)-H_nw(n)\Big)\\ &+\frac{۱}{۲\lambda}\|\sum_{t=۱}^{n}\psi(e_t)K_n^\top x_{nt}\|^۲-\frac{\lambda}{۲}\|w(n)\|^۲ \xrightarrow{p} ۰\\ &=\sum_{t=۱}^{n}[(y_{nt}-x_{nt}^\top K_n\hat{\gamma}(n)-\rho(et)]-\sum_{t=۱}^{n}\psi(et)x_{nt}^\top H_nw(n)\\ &+\frac{۱}{۲\lambda}\|\sum_{t=۱}^{n}\psi(et)K_n^\top X_{nt}\|^۲-\sum_{t=۱}^{n}\psi(et)x_{nt}^\top H_n\omega(n)-\frac{\lambda}{۲}\|\omega(n)\|^۲ \xrightarrow{p} ۰\end{aligned} \qquad (۲۵.۳)$$

با استفاده از رابطه‌های زیر
$H_n^\top H_n = I_q, K_n^\top K_n = I_{p-q}, H_n^\top K_n = 0, H_n H_n^\top + K_n K_n^\top = I_p$ و روابط (۲۵.۳) و (۲۴.۳) داریم

$$\begin{aligned} 2\lambda\sigma_n^2 M_n &= \|\sum_{t=1}^n H_n^\top x_{nt}\psi(et)\|^2 + \lambda^2\|\omega(n)\|^2 + 2\lambda\omega^\top(n)\sum_{t=1}^n H_n^\top x_{nt}\psi(et) + o_(p)(1) \\ &= \|\sum_{t=1}^n H_n^\top x_{nt}\psi(et) + \lambda\omega(n)\|^2 + o_p(1) \end{aligned}$$

(۲۶.۳)

از آنجایی که $E\psi(et) = 0$ $E\psi^2(et) = \sigma^2 < \infty$ $max_{1\le t\le n}\| x_{nt}\zeta\| = O(n^{-1/2})$
همچنین با استفاده از رابطه‌ی $A4$ و کراندار بودن σ_n^2 داریم

$$\begin{aligned} \|Var\Big(\sum_{t=1}^n x_{nt}\psi(et)\Big)\| &= \|\sum_{t=1}^n (x_{nt})^2 E\psi^2(et)\| + 2\|\sum_{t=1}^{n-1}\sum_{j=t+1}^n x_{nt}x_{nj}^\top E(\psi(et)\psi(e_j))\| \\ &= \rho\sigma^2 + 2\|S_n^{-1}\|x_t x_j^\top\|\sum_{t=1}^{n-1}\sum_{j=t+1}^n |E(\psi(et)\psi(e_j))| \\ &= \rho\sigma^2 + 2C\|x_t x_j^\top\| \\ &= \rho\sigma^2 + 2\rho C = O(1) \end{aligned}$$

از دید $H_n^\top H_n = I_q$ و لم ۵.۱.۳ داریم:

$$\sigma^{-1}\sum_{t=1}^n H_n^\top x_{nt}\psi(et) \xrightarrow{D} N(0,1) \qquad (۲۷.۳)$$

بنابراین قضیه فوق از رابطه‌های (۲۶.۳) و (۲۷.۳) پیروی می‌کند.

لم ۸.۱.۳. تحت شرایطی از قضیه ۱.۱.۳ ، زمانی که $n \to \infty$ داریم

$$\hat\beta(n) - \beta(n) = \lambda^{-1}\sum_{t=1}^n x_{nt}\psi(e_t) + o_p(1) \qquad (۲۸.۳)$$

$$\hat\gamma(n) - \gamma(n) = \lambda^{-1}\sum_{t=1}^n K_n^\top x_{nt}\psi(e_t) + o_p(1) \qquad (۲۹.۳)$$

برهان: از برآورد از رابطه $\sum_{t=1}^n \rho(y_t - x_t^\top\beta)$ تعریف کنیم

$$\Big\|S_n^{-1/2}\sum_{t=1}^n \psi(y_t - x_t^\top\hat\beta)x_t\Big\| = o_p(1). \qquad (۳۰.۳)$$

به طوریکه $\hat{\beta}(n) = S_n^{۱/۲}\hat{\beta}_n$ رابطه (۳۰.۳) به صورت زیر بازنویسی می‌کنیم

$$\Big\|\sum_{t=۱}^{n}\psi(e_t - x_{nt}^{\top}\hat{\beta}(n))x_{nt}\Big\| = o_p(۱) \qquad (۳۱.۳)$$

با یک استدلال معمولی، باید ثابت کنیم

$$\hat{\beta}(n) = O_p(۱) \qquad (۳۲.۳)$$

u یک واحد شمارای متراکم در حدود $\mathbb{R}^p$ است. می‌نویسیم:

$$D(\zeta, L) = \sum_{t=۱}^{n}\psi(e_t - Lx_{nt}^{\top}\zeta)x_{nt}^{\top}\zeta$$

به طوریکه $L \geq ۰,\ \zeta \in \mathbb{R}^p$. $D(.,L)$ بدیهی است که، ζ داده شده صعودی روی L زمانی که ψ صعودی باشد. برای هر $\varepsilon > ۰$،

$$L_۰ = \sqrt{p}\sigma/(\lambda\sqrt{\varepsilon}).$$

پس با استفاده از رابطه (۳۱.۳) وجود دارد $n_۱$ که $n \geq n_۱$،

$$Pr\Big\{\Big|D\Big(\frac{\hat{\beta}(n)}{\|\hat{\beta}(n)}, \|\hat{\beta}(n)\Big)\Big| \geq L_۰\lambda\Big\} < \varepsilon.$$

توجه کنید که $D(\zeta,.)$ یک تابع صعودی روی ζ برای L سپس داریم:

$$Pr\{\|\hat{\beta}(n)\|L_۰\} < Pr\{\sup_{\zeta\in u} D(\zeta, L_۰) \leq -L_۰\lambda\} + \varepsilon.$$

با توجه در لم ۷.۱.۳ و $\max_{۱\leq t\leq n}|x_{nt}^{\top}\zeta| = O(n^{-۱/۲})$ داریم

$$\sup_{\zeta} \in u\Big|\sum_{t=۱}^{n}[\psi((e_t) - L_۰x_{nt}^{\top}\zeta) - \psi(e_t)]x_{nt}^{\top}\zeta + L_۰\lambda\Big| \to ۰ \qquad (۳۳.۳)$$

از طرف دیگر با استفاده از نامساوی شوارتز داریم:

$$\sup_{\zeta\in u}\Big|\sum_{t=۱}^{n}\psi(e_t)x_{nt}^{\top}\zeta\Big| \leq \Big\|\sum_{t=۱}^{n}\psi(e_t)x_{nt}\Big\| \qquad (۳۴.۳)$$

از ترکیب رابطه (۳۳.۳) و (۳۴.۳) وجود دارد $n_۲$ $(n_۱ \leq n_۲ \leq n)$ چنان‌چه

$$Pr\Big\{\sup_{\zeta\in u} D(\zeta, L_۰) < -L_۰\lambda + \Big\|\sum_{t=۱}^{n}\psi(e_t)x_{nt}\Big\|\Big\} > ۱ - \varepsilon \qquad (۳۵.۳)$$

با استفاده از نامساوی چبیشف و نابرابری C_r و لم ۷.۱.۳ داریم:

$$\begin{aligned} Pr\Big\{\Big\|\sum_{t=۱}^{n}\psi(e_t)x_{nt}\Big\| \geq L_۰\lambda\Big\} &\leq \Big(\frac{۱}{L_۰\lambda}\Big)^{۲}E\Big\|\sum_{t=۱}^{n}\psi(e_t)x_{nt}\Big\|^{۲} \\ &\leq \Big(\frac{۱}{L_۰\lambda}\Big)^{۲}E\sum_{t=۱}^{n}\|\psi(e_t)x_{nt}\|^{۲} \\ &\leq \Big(\frac{۱}{L_۰\lambda}\Big)^{۲}E\sum_{t=۱}^{n}\sigma^{۲}\|x_{nt}\|^{۲} \leq \varepsilon \end{aligned} \qquad (۳۶.۳)$$

از رابطه (۳۵.۳) و (۳۶.۳) می‌توان نوشت

$$Pr\{\sup_{\zeta\in u} D(\zeta, L_\circ) < -L_\circ\lambda\} > 1 - 2\varepsilon$$

همچنین، زمانی که $n \geq n_2$ بدست می‌آوریم

$$Pr\{\|\hat{\beta}(n)\| \geq L_\circ\} < \varepsilon - 1 - (1 - 2\varepsilon) = 3\varepsilon \qquad (۳۷.۳)$$

بنابراین نتیجه رابطه (۳۲.۳) از رابطه (۳۷.۳) پیروی می‌کند و استبدادی از ε شکل می‌گیرد. با استفاده از لم ۷.۱.۳ و $\max_{1\leq t\leq n}|x_{nt}^\top\zeta| = O(n^{-1/2})$ از این قرار است که

$$I(\|\hat{\beta}(n)\| \geq L_\circ)\Big|\sum_{t=1}^{n}[\psi(e_t - x - nt^\top\hat{\beta}(n)) - \psi(e_t)]x_{nt} + \lambda\hat{\beta}(n)\Big| \xrightarrow{p} \circ$$

دلالت بر این دارد

$$\hat{\beta}(n) = \lambda^{-1}\sum_{t=1}^{n} x_{nt}\psi(e_t) + o_p(1)$$

در نتیجه، رابطه (۲۸.۳) ثابت شده است. همان‌طور که تعریف شد، رابطه $\beta(n) = K_n\gamma(n) + H_n\omega(n)$ و $\hat{\beta}(n)$ می‌توان به طور مشابه به صورت $\hat{\beta}(n) = K_n\hat{\gamma}(n) + H_n\omega(n)$ نوشته شود. پس به ترتیب با تعویض $\beta(n)$، $\hat{\beta}(n)$ و $\gamma(n)$، $\hat{\gamma}(n)$، رابطه (۲۹.۳) ثابت شده است با $K_n^\top K_n = I_{p-q}$.

قضیه ۲.۱.۳. از آنجایی که

$$\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^{n}\psi^2(y_t - x_t^\top\hat{\beta}_(n))$$

$$\hat{\lambda}_n = (2nh)^{-1}\sum_{t=1}^{n}\psi(y_t - x_t^\top\hat{\beta}_n + h) - \psi(y_t - x_t^\top\hat{\beta}_n - h)$$

وجود دارد $h = h_n > \circ$ و h_n ترتیبی است برای پیروی کردن

$$h_n/d_n^{1/2} \to \infty \quad h_n \to \circ \quad \lim_{n\to\infty} nh_n^2 > \circ$$

با توجه به قضیه یک داریم

$$\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$$

$$\hat{\lambda}_n^2 \to \lambda$$

تحت این فرض $\|S_n^{1/2}\omega_n\| \to \circ$ به ترتیب با تعویض λ, σ^2 برآورد ثابت آن‌ها $\hat{\sigma}_n$ و $\hat{\lambda}_n$ سپس داریم:

$$2\hat{\lambda}_n\hat{\sigma}_n^{-2}M_n \xrightarrow{D} \chi_p^2$$

اثبات: مدل زیر را در نظر بگیرید $y_{nt} = x_{nt}^{\top}\beta(n) + e_t \quad t = ۱, ۲, ..., n$، بدون از دست دادن کلیت، فرض کنید پارامتر $\beta(n)$ صحیح، و برابر صفر باشد. برای هر $\delta > ۰$ می‌نویسیم:

$$V_n = E|\psi(e_۱ + \delta d_n^{۱/۲}) - \psi(e_۱ - \delta d_n^{۱/۲})|$$

با استفاده از همگرایی ψ، نامساوی شوارتز و شرط $A۲$ برای n های به اندازه کافی بزرگ داریم:

$$\begin{aligned} & EI(\|\hat{\beta}(n)\| \le \delta)\Big|\hat{\sigma}_n^{۲} - n^{-۱}\sum_{t=۱}^{n}\psi^{۲}(e_t)\Big| \\ & \le V_n + ۲E|\psi(e_۱)[\psi(e_۱ + \delta d_n^{۱/۲}) - \psi(e_۱ - \delta d_n^{۱/۲})]| \\ & \le V_n + ۲\delta V_n^{۱/۲} \to ۰ \end{aligned}$$

.

با استفاده از لم ۸.۱.۳ داریم:

$$\hat{\delta}_n^{۲} = n^{-۱}\sum_{t=۱}^{n}\psi(e_t + h) - \psi(e_t - h) \xrightarrow{p} E\psi^{۲}(e_۱) = \sigma^{۲}.$$

در نتیجه رابطه $\hat{\delta_n^{۲}} \xrightarrow{p} \delta^{۲}$ ثابت شده است. چن و همکاران به منظور اثبات رابطه $\hat{\lambda^{۲}} \xrightarrow{p} \lambda$ می‌توان اثبات کرد که

$$(۲nh)^{-۱}\sum_{t=۱}^{n}\psi(e_t + h) - \psi(e_t - h) \xrightarrow{p} \lambda.$$

در واقع همگرایی از $\psi(e_t + h) - \psi(e_t - h)$ و فرض رابطه

$$h_n/d_n^{۱/۲} \to \infty,\ h_n \to ۰,\ \lim_{n\to\infty} nh_n^{۲} > ۰$$

با استفاده از لم ۶.۱.۳ و لم ۱.۱.۳ داریم:

$$\begin{aligned} Var\Big\{(۲nh)^{-۱}\sum_{t=۱}^{n}[\psi(e_t + h) - \psi(e_t - h)]\Big\} & \le (۴n^{۲}h^{۲})^{-۱}E\Big[\sum_{t=۱}^{n}\psi(e_t + h) - \psi(e_t - h))\Big]^{۲} \\ & \le (۴nh^{۲})^{-۱}E[\psi(e_t + h) - \psi(e_t - h))]^{۲} \to ۰. \end{aligned}$$

از طرف دیگر ازآنجایی که $\lim_{n\to\infty}[G(h) - G(-h)]/(۲h) = \lambda$،

$$(۲nh)^{-۱}\sum_{t=۱}^{n}[\psi(e_t + h) - \psi(e_t - h)] = [G(h) - G(-h)]/(۲h) + o_p(۱) \to \lambda.$$

اثبات قضیه کامل شد.

فصل ۴

ارزیابی برآوردها و آزمون‌های M با مطالعه شبیه‌سازی

در این فصل، برای ارزیابی روش برآوردیابی استوار و توان آزمون‌های M مطرح‌شده در این پایان‌نامه، از مطالعه شبیه‌سازی مونت کارلویی و نتایج گزارش‌شده در مقاله یو و همکاران (۲۰۱۷) استفاده کرده‌ایم. برای اجرای شبیه‌سازی از نرم‌افزار R استفاده شده و کدهای مربوط برای این هدف در پیوست آمده است.

۱.۴ مثال شبیه‌سازی

در این مطالعه، تحت فرضیه صفر، برآوردگرهای ضرایب رگرسیونی و پارامترهای مازاد توسط روش‌های M مطرح‌شده شامل روش‌های LS، LAD و هوبر محاسبه شدند. همچنین، تحت فرضیه جانشین موضعی، توابع توان آزمون‌های M پیشنهادی با محاسبه ناحیه رد داده‌شده در قضیه ۱.۱.۳ به دست آمدند.

برای تولید یک دنباله از متغیرهای NSD، از رابطه زیر استفاده شد:

$$X_t = a_n Y_t + b_n Z_t \qquad t = ۱, ۲, ..., n \tag{۱.۴}$$

که در آن a_n و b_n دنباله‌هایی از اعداد حقیقی مثبت هستند و Y_t و Z_t متغیرهای تصادفی وابسته

منفی (متناظر با $\rho_{\circ} < \circ$) هستند که از توزیع توام

$$(Y, Z) \sim \mathrm{N}(\mu_{۱}, \mu_{۲}, \sigma_{۱}^{۲}, \sigma_{۲}^{۲}, \rho_{\circ})$$

پیروی می‌کنند. برای اطمینان از درستی فرآیند تولید مشاهدات از یک دنباله NSD، ابتدا نشان خواهیم داد $(e_{۱}, e_{۲}, \ldots, e_{n})$ که به کمک رابطه (۱.۴) ساخته می‌شوند، یک دنباله متناهی NSD است.

با توجه به توزیع (Y_t, Z_t)، می‌توان نشان داد

$$Cov(X_{tt}, X_j) < \circ \quad ۱ \leq t < j \leq n.$$

از طرفی، هو (۲۰۰۰) نشان داده است اگر تابع ϕ دارای مشتق‌های جزیی مرتبه دوم پیوسته باشد، آن‌گاه زبرجمعی بودن ϕ معادل است با

$$\frac{\partial^{۲}\phi}{\partial x_t \partial x_j} \geq \circ \quad ۱ \leq t \neq j \leq n.$$

پس تابع $\phi(x_{۱}, x_{۲}, \ldots, x_n) = \exp(\sum_{t=۱}^{n} t)^{۲}$ یک تابع زبرجمعی است. اکنون با توجه به این‌که مولفه‌های $\{X_t^*, ۱ \leq t \leq n\}$ توزیع‌های کناری مشابهی با مولفه‌های $\{X_t, ۱ \leq t \leq n\}$ دارند، با استفاده از نامساوی ینسن، می‌توان نوشت

$$\begin{aligned}\frac{E\phi(X_{۱}^*, X_{۲}^*, \ldots, X_n^*)}{E\phi(X_{۱}, X_{۲}, \ldots, X_n)} &= E\exp\Big\{\Big(\sum_{t=۱}^{n} X_t^*\Big)^{۲} - \Big(\sum_{t=۱}^{n} X_t\Big)^{۲}\Big\} \\ &\geq \exp E\Big\{\Big(\sum_{t=۱}^{n} X_t^*\Big)^{۲} - \Big(\sum_{t=۱}^{n} X_t\Big)^{۲}\Big\} \geq ۱.\end{aligned}$$

بنابراین $(X_{۱}, \ldots, X_n)$ یک دنباله NSD است.

برای برآورد پارامترها در مثال شبیه‌سازی، تابع هوبر به صورت

$$\rho(x) = (x^{۲} I(|x| \leq k))/۲ + (k|x| - k^{۲}/۲) I(|x| > k)$$

در نظر گرفته شد، که در آن $k = ۱/۳۴۵\sigma_{\circ}$. برای تولید متغیرهای توضیحی رگرسیونی x_t دو سناریو مختلف در نظر گرفته شدند:

(I) با تولید u_t، برای $t = ۱, \ldots, n$، از توزیع یکنواخت استاندارد، متغیر توضیحی به صورت $x_t = ۵u_t$ تعریف شد.

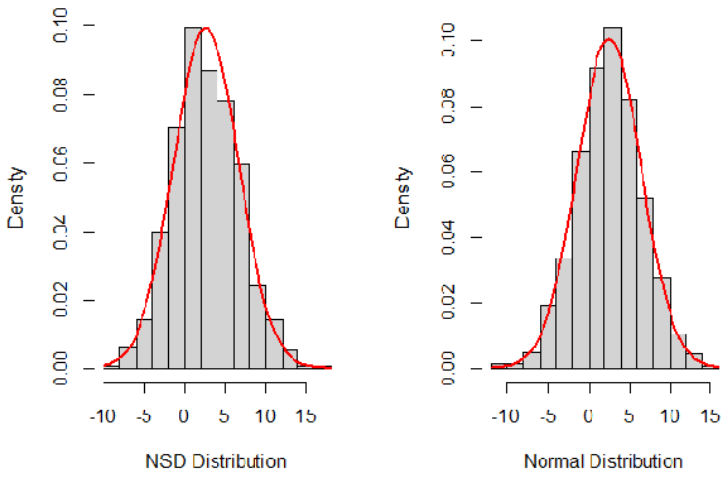
(II) با همان تحقق‌های یکنواخت استاندارد، متغیر توضیحی به صورت $\sin(۲t) + ۱/۵u_t$ در نظر گرفته شد.

سپس مشاهدات از مدل خطی با جمله خطای NSD به صورت زیر تولید شدند:

$$y_t = \beta_{\circ} + \beta_{۱} x_t + e_t, \quad t = ۱, ۲, \ldots, n$$

که در آن خطاهای NSD $\{e_t = Y_t + Z_t,\ t = ۱,۲,\ldots,n\}$ از یک آمیخته چندمتغیره از توزیع نرمال $(Y,Z) \sim \mathrm{N}(\mu_۱,\mu_۲,\sigma_۱^۲,\sigma_۲^۲,\rho_۰)$ با $\rho_۰ < ۰$ پیروی می‌کنند. به‌طور دقیق توزیع نرمال دومتغیره به صورت $(Y,Z) \sim \mathrm{N}(۰,۰,۱,۱۶,-۰/۵)$ در نظر گرفته شد. مقادیر واقعی ضرایب رگرسیونی نیز برابر $(\beta_۰,\beta_۱)^\top = (۱,۲)^\top$ در نظر گرفته شدند. این مقادیر همان وضعیت فرضیه صفر را تعیین می‌کنند؛ یعنی $H_۰ : (\beta_۰,\beta_۱)^\top = (۱,۲)^\top$. افزون بر این برای بررسی اندازه نمونه، سه حجم نمونه $n = ۱۰۰,۵۰۰,۱۰۰۰$ انتخاب شدند. برای کاهش تاثیر تولید تصادفی مشاهدات در برآورد پارامترها، شبیه‌سازی مدل خطی ۱۰۰۰ بار تکرار شد و میانگین برآوردهای پارامترها محاسبه شدند.

شکل ۱.۴ منحنی تابع چگالی توزیع‌های برازش‌شده NSD و نرمال را روی هیستوگرام باقی‌مانده‌های حاصل از مدل رگرسیونی برازش‌شده با روش برآورد M، برای حجم نمونه ۱۰۰۰، نشان می‌دهد. از روی نتیجه این شکل می‌توان نزدیکی دو توزیع برازش‌شده روی باقی‌مانده‌ها را نتیجه گرفت. البته، به‌طور نسبی، می‌توان گفت توزیع NSD برازش‌شده رفتاری نزدیک به یک توزیع بریده‌شده دارد.



شکل ۱.۴: نمودارهای هیستوگرام و منحنی‌های چگالی برازش‌شده باقی‌مانده‌های مدل رگرسیونی با خطای NSD و مستقل برای حجم نمونه $n = ۱۰۰۰$ به ازای یکی از روش‌های M ـ برآورد

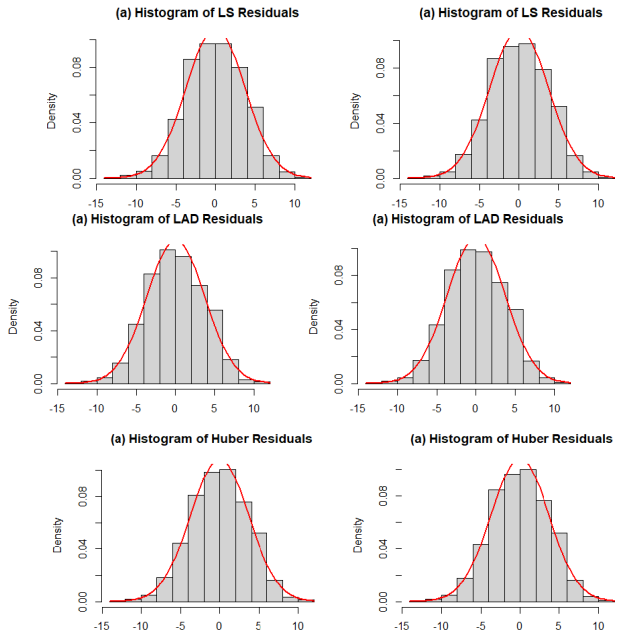
جدول ۱.۴ حاوی نتایج ارزیابی برآوردگرهای ضرایب رگرسیونی و پارامترهای مازاد تحت فرضیه صفر است. نتایج جدول نشان می‌دهند که روش‌های M ـ برآورد معتبر هستند؛ زیرا برآوردهای متناظر پارامترهای رگرسیونی به مقادیر واقعی $\beta_۰ = ۱$ و $\beta_۱ = ۲$ نزدیک هستند. از طرفی برآوردگرهای پارامترهای مازاد در روش‌های مختلف مناسب و موثر هستند. مثلا برای تابع محدب روش LS، مقادیر واقعی پارامترهای مازاد $\sigma^۲ = ۱۳$ و $\lambda = ۱$ هستند. برای دو روش دیگر LAD و هوبر نیز وضعیت مشابهی برقرار است. البته مقادیر واقعی پارامترهای مازاد در این روش‌ها متفاوت هستند ولی از روی علامت یکسان برآوردها و معنی‌دار بودن آن‌ها، می‌توان به‌طور ضمنی تشخیص داد که عملکرد روش‌ها قابل قبول و موثر هستند. همچنین می‌توان

گفت که برآوردها حتی با حجم نمونه نه چندان بزرگ مثل $n = ۱۰۰$ دقت و رفتار خوبی دارد، اما از روی نتایج جدول واضح است که با افزایش اندازه نمونه، دقت برآوردها افزایش می‌یابند.

جدول ۱.۴: ارزیابی ضرایب رگرسیون و پارامترهای زیاد

پارامتر	n	ls.I	ls.II	lad.I	lad.II	هوبر.I	هوبر.II
$\hat{\beta}_۰$	۱۰۰	۱/۱۶۳	۱/۵۰۴	۰/۹۹۰	۱/۴۹۳	۱/۰۸۹	۱/۵۶۸
	۵۰۰	۱/۱۷۷	۱/۰۱۳	۱/۴۸۶	۱/۱۶۰	۱/۲۱۷	۱/۰۲۵
	۱۰۰۰	۱/۴۳۳	۱/۱۹۳	۱/۲۴۷	۰/۸۹۲	۱/۳۶۶	۱/۱۶۹
$\hat{\beta}_۱$	۱۰۰	۱/۹۲۷	۱/۸۱۴	۱/۸۹۲	۱/۶۹۰	۱/۹۷۷	۱/۷۵۳
	۵۰۰	۱/۹۲۷	۲/۱۰۳	۱/۸۴۸	۱/۸۷۹	۱/۹۲۶	۲/۰۶۸
	۱۰۰۰	۱/۸۶۵	۱/۸۷۴	۱/۸۹۱	۲/۰۰۲	۱/۸۷۶	۱/۹۱۰
$\hat{\sigma}^۲_n$	۱۰۰	۱۳/۹۵۰	۱۰/۹۵۳	۱/۰۰۰	۱/۰۰۰	۱۰/۴۵۱	۸/۴۹۹
	۵۰۰	۱۳/۲۵۸	۱۲/۱۱۱	۱/۰۰۰	۱/۰۰۰	۹/۶۸۷	۹/۲۰۷
	۱۰۰۰	۱۲/۵۶۵	۱۲/۷۱۴	۱/۰۰۰	۱/۰۰۰	۹/۳۵۴	۹/۳۹۹
$\hat{\lambda}_n$	۱۰۰	۱/۰۰۰	۱/۰۰۰	۰/۲۰۱	۰/۲۳۳	۰/۴۹۹	۰/۳۷۱
	۵۰۰	۱/۰۰۰	۱/۰۰۰	۰/۲۰۰	۰/۲۰۴	۰/۶۷۸	۰/۵۸۵
	۱۰۰۰	۱/۰۰۰	۱/۰۰۰	۰/۲۴۶	۰/۲۴۵	۰/۷۹۰	۰/۷۵۷

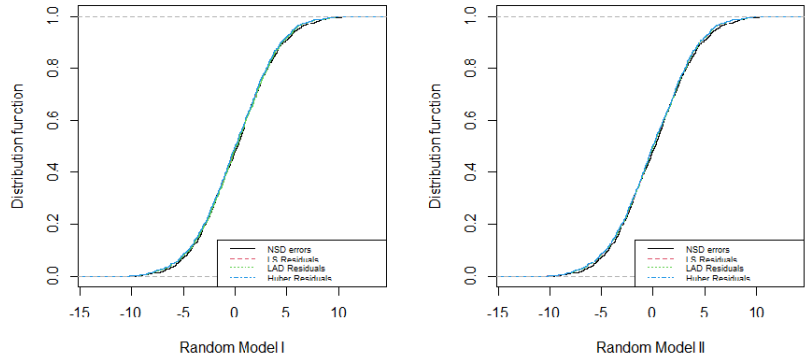
همان‌طور که انتظار داشتیم، بر اساس شکل ۲.۴، منحنی‌های توابع چگالی برازش‌شده باقی‌مانده‌ها به خطاهای NSD فرض‌شده نزدیک هستند و همه آن‌ها باز هم ویژگی یک توزیع بریده‌شده را نشان می‌دهند.



شکل ۲.۴: نمودارهای هیستوگرام و منحنی چگالی برازش‌شده باقی‌مانده‌های M ـ برآورد با متغیرهای توضیحی محتلف در دو سناریو و روش‌های مختلف M ـ برآورد به ازای حجم نمونه $n = ۱۰۰۰$

شکل ۳.۴ نموداری است که NSD بودن باقی‌مانده‌ها رو ارزیابی می‌کند. برای این کار توزیع تجربی باقی‌مانده‌ها به‌عنوان برآوردی از توزیع واقعی آن‌ها با توابع توزیع برازش‌شده NSD، برای دو سناریو تولید متغیر توضیحی، مقایسه شده‌اند. نتیجه از NSD بودن باقی‌مانده‌ها به خوبی پشتیبانی می‌کند.

به‌عنوان آخرین بخش مطالعه شبیه‌سازی، سطح معنی‌داری تجربی آزمون‌های M پیشنهادی



شکل ۳.۴: مقایسه توابع توزیع NSD فرض‌شده برای باقی‌مانده‌های رگرسیونی با توزیع تجربی برآوردشده باقی‌مانده‌ها برای حجم نمونه $n = ۱۰۰۰$ و دو سناریو مختلف تولید متغیر توضیحی

ارزیابی شدند. بنا به قضیه ۱.۱.۳ و قضیه ۲.۱.۳، تحت فرضیه جانشین موضعی، $۲\hat{\lambda}_n\hat{\sigma}_n^{-}۲M_n$ دارای توزیع مجانبی کی‌دو مرکزی با ۲ درجه آزادی است. بنابراین فرضیه جانشین رد می‌شود اگر مقدار شبیه‌سازی‌شده $۲\hat{\lambda}_n\hat{\sigma}_n^{-}۲M_n$ در W قرار گیرد، به‌طوری که

$$W = (۰, \chi^{۲}_{p}(۱ - \alpha/۲)) \cup (\chi^{۲}_{p}(\alpha/۲), +\infty).$$

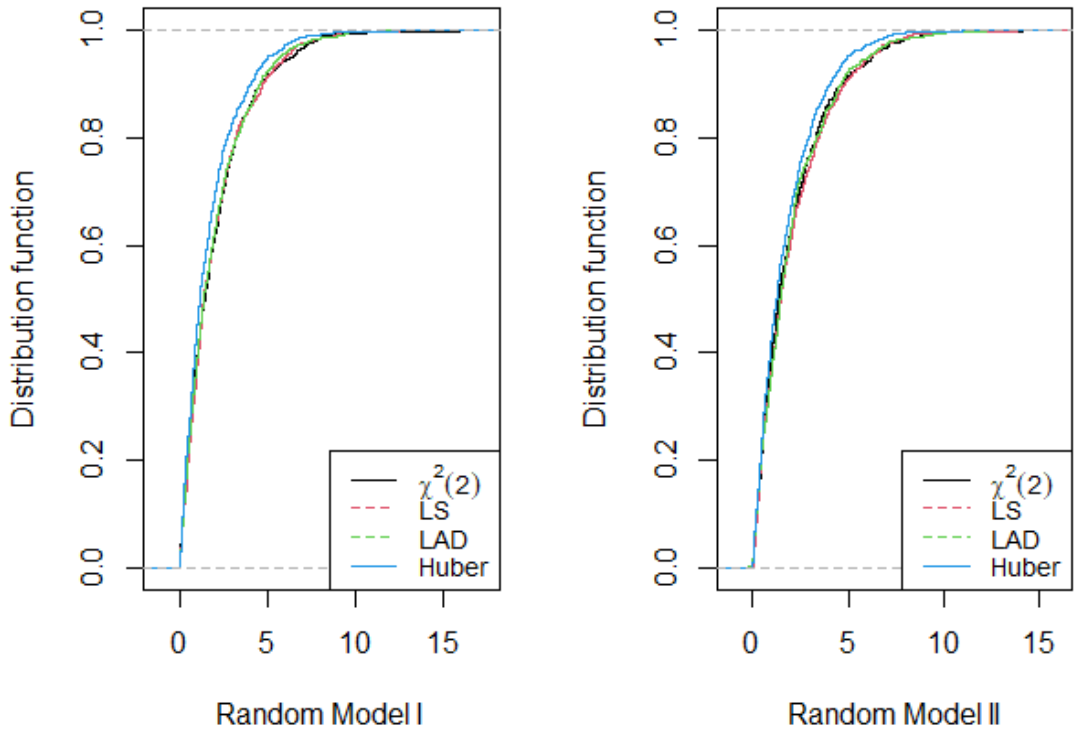
جدول ۲.۴ نتایج اندازه‌های تجربی آزمون‌ها را برای سطوح معنی‌داری $\alpha = ۰/۰۵$ و $\alpha = ۰/۰۱$ به ازای انتخاب‌های مختلف روش‌های M ـ برآورد، سناریوهای تولید متغیر توضیحی و اندازه‌های نمونه مختلف نشان می‌دهد.

جدول ۲.۴: اندازه‌های تجربی آزمون‌های M مختلف در مدل رگرسیون خطی با خطاهای NSD ـ مقادیری که با $*$ مشخص شده‌اند، بیانگر اندازه‌های معنی‌داری اسمی هستند.

سطح معنی‌داری	LS.I	LS.II	LAD.I	LAD.II	هوبر.I	هوبر.II
۰/۰۵*	۰۵۷.۰	۰۴۷.۰	۰۴۲.۰	۰۳۸.۰	۰۴۱.۰	۰۳۵.۰
۰/۱۰*	۰۱۶.۰	۰۱۳.۰	۰۱۵.۰	۰۰۷.۰	۰۰۹.۰	۰۰۷.۰
۰/۰۵*	۰۵۱.۰	۰۵۴.۰	۰۴۳.۰	۰۴۸.۰	۰۴۹.۰	۰۴۵.۰
۰/۱۰*	۰۰۹.۰	۰۰۹.۰	۰۱۳.۰	۰۱۲.۰	۰۰۸.۰	۰۱۰.۰
۰/۰۵*	۰۴۹.۰	۰۴۱.۰	۰۵۵.۰	۰۵۰.۰	۰۴۲.۰	۰۳۷.۰
۰/۱۰*	۰۱۰.۰	۰۰۵.۰	۰۱۸.۰	۰۱۴.۰	۰۰۹.۰	۰۱۱.۰

نتایج جدول بیانگر آن است که اندازه‌های معنی‌داری تجربی به اندازه‌های اسمی خود نزدیک هستند. بنابراین آزمون‌های M پیشنهادی معتبر و قابل دفاع هستند.

شکل ۴.۴ نیز، با مقایسه تابع توزیع تجربی $۲\hat{\lambda}_n\hat{\sigma}_n^{۲}M_n$ برای روش‌های M ـ برآورد مختلف و تابع توزیع کی‌دو با دو درجه آزادی، نشان می‌دهد که $۲\hat{\lambda}_n\hat{\sigma}_n^{۲}M_n$ می‌تواند توزیع کی‌دو را به خوبی تقریب بزند. این نتیجه، تایید دیگری بر معتبر بودن آزمون‌های M پیشنهادی تحت فرضیه‌های جانشین موضعی است.



شکل ۴.۴: مقایسه توابع توزیع برازش‌شده $2\hat{\lambda}_n\hat{\sigma}_n^2 M_n$ برای روش‌های M ـ برآورد مختلف و توزیع کی‌دو مرکزی با دو درجه آزادی به ازای اندازه نمونه $n = ۱۰۰۰$ و سناریوهای تولید متغیر توضیحی

نتیجه‌گیری مطالعه شبیه‌سازی

نتایج حاصل از شبیه‌سازی نشان می‌دهد که آزمون‌های M برای مدل خطی با خطاهای NSD به انتخاب روش‌های M ـ برآورد و همین‌طور متغیرهای توضیحی حساس نیستند و بنابراین یک تنومندی را نشان می‌دهند. این نتیجه موثر بودن آزمون M را در چارچوب وابستگی نوع NSD در مدل‌های خطی را تشریح می‌کند.

مراجع

[۱] اقبال، ن. (۱۳۸۴)، پایان‌نامه ارشد: '' **روش‌های استواری** '' دانشکده علوم ریاضی، دانشگاه فردوسی مشهد.

[۲] حبیبی کمنی، س. (۱۳۹۵)، پایان‌نامه ارشد: '' **همگرایی کامل برای مجموع موزون متغیرهای تصادفی وابسته زبرجمعی منفی و کاربرد آن در مدل رگرسیون خطی با متغیرهخای تبیینی آلوده به خطا** '' دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود.

[۳] حیدری نژاد، ح. (۱۳۹۷)، پایان‌نامه ارشد: '' **سازگاری قوی M ـ برآوردها در مدل خطی با خطاهای وابسته زبرجمعی‌منفی** '' دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود.

[۴] رحمانی، س. (۱۳۹۵)، پایان‌نامه ارشد: '' **همگرایی کامل برآوردگر مدل رگرسیون ناپارامتری با جملات خطای وابسته زبرجمعی منفی** '' دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود.

[۵] 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.

[6] 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.

[7] Andersen P. K. and Gill R. D. (1982). "Cox's regression model for counting processes: a large sample study" **The annals of statistics** 1100-1120.

[8] Bai Z. D. Rao C. R. and Zhao L. C. (1993). "MANOVA type tests under a convex discrepancy function for the standard multivariate linear model" **Journal of statistical planning and inference** 36(1), 77-90.

[9] 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.

[10] Billingsley P. (1968). "Convergence of probability measures John Wiley" **New York**

[11] Block, H. W., and Sampson, A. R. (1988). Conditionally orderd distributions, **J Multi Anal**, 27,91-104.

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

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

[14] 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.

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

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

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

[18] Efron, D. (1965). Increasing Properteies of polya frequency function, Ann Math Statist, **36,** 272-279.

[19] Eghbal N. Amini M. Bozorgnia A. (2010). "Some maximal inequalities for quadratic forms of negative superadditive dependence random variables" **Statistics and probability letters** 80(7-8), 587-591.

[20] Eghbal N. Amini M. and Bozorgnia A. (2011). "On the Kolmogorov inequalities for quadratic forms of dependent uniformly bounded random variables" **Statistics and probability letters** 81(8), 1112-1120.

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

[22] Hampel, F.R. (1974). "The Influence Curve and Its Robust Estimation", **Journal of the American Statistical Association,** 69, pp 383-393.

[23] Hill, R.W. and Holland, P.W. (1977). "two Robust to least squares regression ," **Journal of the American Statistical Association,** 72, pp 828-833.

[24] Hogg, R.E. (1799). "Statistical Robustness: One view of its use in applications today," **American Statistical.** 33:3, pp 108-115.

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

[26] Hu, T.Z. (2000). "Negatively supperadditive dependenc of random variables with applications." **Chinese J. Appl. Probab. Statist.** 16, pp 133-144.

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

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

[29] Huber, P.J.(1972)."Robust Statistic: A Review," Annals of Mathematical Statisstics, 43, pp 1041-1067.

[30] Huber, P.J.(1973)."Robust Regression: Asymptotics, Conjectures, and Monte Carlo," Anals of Statistics, 1, pp 799-821.

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

[32] 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.

[33] 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.

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

[35] 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.

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

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

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

[39] 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.

[40] Shen A. Zhang Y. and Volodin, A. (2015). ”Applications of the Rosenthal-type inequality for negatively superadditive dependent random variables” **Metrika** 78(3), 295-311.

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

[42] 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.

[43] 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.

[44] Welsh A. H. (1986), ”Bahadur representations robust scale estimators based on regression residuals” **The Annals of Statistics**, 1246-1251.

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

[46] 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.

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

[48] 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.

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

[50] Yohai V. J., and Maronna R. A. (1979), "Asymptotic behavior of M-estimators for the linear model" **The Annals of Statistics** 258-268.

[51] Yu, Y., Hu, H., Liu, L., and Huang, S. (2017). M-test in linear models with negatively superadditive dependent errors. **Journal of inequalities and applications,** 2017(1), .1-21

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

[53] Zhao, L.C. (2000). "Strong consistency of M-estimates in linear models." **China Ser.** A 45, pp 1420-1427.

[54] Zhao L. C. and Chen X. R. (1991). "Asymptotic behavior of M-test statistics in linear models" **J. Comput. Inf. Syst** 16, 234-248.

[55] Zhao L. (2008). "Bai's Contribution to M-estimation and Relevant tests in linear models" **In Advances In Statistics** (pp. 19-26).

پیوست آ

کدهای شبیه‌سازی جداول و نمودارهای فصل چهار

آ.۱ کدهای مربوط به جدول دو و نمودار شکل چهار

```
################################ Model I
rm(list=ls())

library(mvtnorm)
library(MASS)

set.seed(8888)
Beta_0 = 1 ; Beta_1 = 2
n =1000

ro = -.5
cov.0 = ro * sqrt(1*16)
```

```
R <- matrix(c(1, cov.0, cov.0, 16),
nrow = 2, ncol = 2)

mu <- c(X = 0, Y = 0)
X2_ls= X2_lad= X2_huber=c()

# Make loop to find the Statistics
rep=1000

for(rr in 1:rep){
YZ <- rmvnorm(n, mean = mu, sigma = R)

u = runif(n)

###################### First model (I)
x = 5 * u
###################### Second model (II)
#x = c(); for (t in 1:n) { x[t] = sin(2*t) + 1.5 * u[t]}
######################### Generate Linear model with NSD errors
e = YZ[,1]+YZ[,2]
y = Beta_0 + Beta_1 * x + e

#plot(x,y)



#fit <- fitdistr(y, densfun="normal")
#fit
#hist(y, pch=20, breaks=13, prob=TRUE, main="" , xlab = "Normal Distribution")
#curve(dnorm(x, fit$estimate[1], fit$estimate[2]), col="red", lwd=2, add=T)
#par(mfrow=c(1,1))

#################### LS
fit.ls = lm(y~x)
var.ls = var(predict(fit.ls)-y)
```

```
res1<-predict(fit.ls)-y
#fit <- fitdistr(res1, densfun="normal")
#fit
#hist(res1, pch=20, breaks=13, prob=TRUE, main="(a) Histogram of LS Residuals",

 xlab = "Model I")
#curve(dnorm(x, fit$estimate[1], fit$estimate[2]), col="red", lwd=2, add=T)

################## LAD

library(L1pack)
fit.lad = lad(y~x,  method = "EM")
var.lad = var(predict(fit.lad)-y)

res2<-predict(fit.lad)-y
#fit <- fitdistr(res2, densfun="normal")
#fit
#hist(res2, pch=20, breaks=13, prob=TRUE, main="(a) Histogram of LAD Residuals",

 xlab = "Model I")
#curve(dnorm(x, fit$estimate[1], fit$estimate[2]), col="red", lwd=2, add=T)

#################### Huber

fit.huber <- rlm(y~x, scale.est = "Huber", method = "M")
var.huber = var(predict(fit.huber)-y)

res3<-predict(fit.huber)-y
#fit <- fitdistr(res3, densfun="normal")
#fit
#hist(res3, pch=20, breaks=13, prob=TRUE, main="(a) Histogram of Huber Residuals",

 xlab = "Model I")
#curve(dnorm(x, fit$estimate[1], fit$estimate[2]), col="red", lwd=2, add=T)
```

```
######################### another way of lad
#library("quantreg")
#fit.lad2 <- rq(y ~ x)
################################ ecdf plots Figure 3

#par(mfrow=c(1,1))
#plot(ecdf(e), xlab="Random Model I", ylab = "Distribution function", main = "")
#lines(ecdf(res1), col = 2)
#lines(ecdf(res2), col = 3)
#lines(ecdf(res3), col = 4)
#legend("bottomright", c("NSD errors", "LS Residuals","LAD Residuals",

 "Huber Residuals"), col=c(1,2,3,4),
#       lty= 1:4, cex = 0.55)

######################### Mn Functions

Mn.ls <- sum((y-Beta_0-Beta_1*x)^2)/2-sum((y-fit.ls$coefficients[1]-fit.
ls$coefficients[2]*x)^2)/2

Mn.lad <- sum(abs(y-Beta_0-Beta_1*x))-sum(abs(y-fit.lad$coefficients[1]-fit.
lad$coefficients[2]*x))

w0=y-Beta_0-Beta_1*x
w1=y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x
#.................. Find the Best k in Huber
MAR = median(abs(predict(fit.huber)-y))   #median absolute residual
k = 1.345 * (MAR/ 0.6745)
k
#.................................

Mn.huber <- sum(0.5*((w0^2)*ifelse(abs(w0)<=k,1,0))+
(k*abs(w0)-0.5*(k^2))*ifelse(abs(w0)>k,1,0))-
```

```
sum(0.5*((w1^2)*ifelse(abs(w1)<=k,1,0))+
(k*abs(w1)-0.5*(k^2))*ifelse(abs(w1)>k,1,0))

Mn.ls; Mn.lad; Mn.huber

############################ Sigma

sigma.ls = n^(-1)*sum((y-fit.ls$coefficients[1]-fit.ls$coefficients[2]*x)^2)

sigma.lad = n^(-1)*sum((sign(y-fit.lad$coefficients[1]-fit.lad$coefficients[2]*x))^2)

sigma.huber = n^(-1)*sum(((-k*ifelse((y-fit.huber$coefficients[1]-fit.
huber$coefficients[2]*x)< -k,1,0))+
(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x)*
ifelse(abs(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x)<=k,1,0)
+k*ifelse((y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x)>k,1,0))^2)

############################ Lambda
h=1
lamb.ls = (2*n*h)^(-1)*sum((y-fit.ls$coefficients[1]
-fit.ls$coefficients[2]*x+h)-(y-fit.ls$coefficients[1]-fit.ls$coefficients[2]*x-h))

lamb.lad = (2*n*h)^(-1)*sum(sign(y-fit.
lad$coefficients[1]-fit.lad$coefficients[2]*x+h)
-sign(y-fit.ls$coefficients[1]-fit.ls$coefficients[2]*x-h))

lamb.huber = (2*n*h)^(-1)*sum((-k*ifelse((y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x+h)< -k,1,0))+
(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x+h)
*ifelse(abs(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x+h)<=k,1,0)
+k*ifelse((y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x+h)>k,1,0)-(-k*ifelse((y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x-h)< -k,1,0))+
(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x-h)
```

```
*ifelse(abs(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x-h)<=k,1,0)
+k*ifelse((y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x-h)>k,1,0))

################################################
#                          Model I
################################################ Figure 4

X2_ls[rr] = 2*lamb.ls*(1/sigma.ls)*Mn.ls

X2_lad[rr] = 2*lamb.lad*(1/sigma.lad)*Mn.lad

X2_huber[rr] = 2*lamb.huber*(1/sigma.huber)*Mn.huber

}
par(mfrow=c(1,2))
plot(ecdf(rchisq(rep,2)), xlab="Random Model I", ylab = "Distribution function", main = "")
lines(ecdf(X2_ls), col = 2)
lines(ecdf(X2_lad), col = 3)
lines(ecdf(X2_huber), col = 4)
legend("bottomright", c(expression(chi^2*(2)), "LS","LAD", "Huber"), col=c(1,2,3,4),
lty= c(1,2,2,1), cex = 0.85)

############################################## density plot

plot(density(X2_huber), col=4, main= "densities plot Model I")
lines(density(rchisq(rep, 2)), col=1)
lines(density(X2_lad), col=3)
lines(density(X2_ls), col=2)
legend("topright", c(expression(chi^2*(2)),"LS","LAD","Huber"), col=1:4, lwd = 1)

################################### Table 2
alpha=0.01
X2.L <- qchisq(alpha/2, df = 2, lower.tail = T)
X2.R <- qchisq(alpha/2, df = 2, lower.tail = F)
```

```
pvalue.ls = length(which(X2_ls<=X2.L))/rep+length(which(X2_ls>=X2.R))/rep
pvalue.lad = length(which(X2_lad<=X2.L))/rep+length(which(X2_lad>=X2.R))/rep
pvalue.huber = length(which(X2_huber<=X2.L))/rep+length(which(X2_huber>=X2.R))/rep

p100.05.II = c(pvalue.ls , pvalue.lad, pvalue.huber)
p100.01.II = c(pvalue.ls , pvalue.lad, pvalue.huber)

p500.05.II = c(pvalue.ls , pvalue.lad, pvalue.huber)
p500.01.II = c(pvalue.ls , pvalue.lad, pvalue.huber)

p1000.05.II = c(pvalue.ls , pvalue.lad, pvalue.huber)
p1000.01.II = c(pvalue.ls , pvalue.lad, pvalue.huber)
####################
p100.05.I = c(pvalue.ls , pvalue.lad, pvalue.huber)
p100.01.I = c(pvalue.ls , pvalue.lad, pvalue.huber)

p500.05.I = c(pvalue.ls , pvalue.lad, pvalue.huber)
p500.01.I = c(pvalue.ls , pvalue.lad, pvalue.huber)

p1000.05.I = c(pvalue.ls , pvalue.lad, pvalue.huber)
p1000.01.I = c(pvalue.ls , pvalue.lad, pvalue.huber)

LS.I=c(p100.05.I[1],p100.01.I[1],p500.05.I[1],p500.01.I[1],p1000.05.I[1]
,p1000.01.I[1])
LS.II=c(p100.05.II[1],p100.01.II[1],p500.05.II[1],p500.01.II[1]
,p1000.05.II[1],p1000.01.II[1])

LAD.I=c(p100.05.I[2],p100.01.I[2],p500.05.I[2],p500.01.I[2]
,p1000.05.I[2],p1000.01.I[2])
LAD.II=c(p100.05.II[2],p100.01.II[2],p500.05.II[2],p500.01.II[2]
,p1000.05.II[2],p1000.01.II[2])

Huber.I=c(p100.05.I[3],p100.01.I[3],p500.05.I[3],p500.01.I[3]
```

```
,p1000.05.I[3],p1000.01.I[3])
Huber.II=c(p100.05.II[3],p100.01.II[3],p500.05.II[3],p500.01.II[3]
,p1000.05.II[3],p1000.01.II[3])


table_2 = cbind(rep(c(0.05,0.1), times=3),LS.I,LS.II,LAD.I,LAD.II,Huber.I,Huber.II)
save(table_2, file = "table2.RData")

library(xtable)

table_2=data.frame(table_2)
xtable(table_2, digits = 3)

################################################ Table 1
##..................... n= 100

out100_1 <- rbind(c(fit.ls$coefficients[1],
fit.lad$coefficients[1],fit.huber$coefficients[1]),
c(fit.ls$coefficients[2],fit.lad$coefficients[2],fit.huber$coefficients[2]),
c(sigma.ls, sigma.lad, sigma.huber),
c(lamb.ls, lamb.lad, lamb.huber))
colnames(out100_1)=c("LS","LAD","Huber")
rownames(out100_1)=c("B0","B1","Sigma2","Lambda")
#save(out100_1, file = "out100_1.Rdata")

##..................... n=500

out500_1 <- rbind(c(fit.ls$coefficients[1],
fit.lad$coefficients[1],fit.huber$coefficients[1]),
c(fit.ls$coefficients[2],fit.lad$coefficients[2],fit.huber$coefficients[2]),
c(sigma.ls, sigma.lad, sigma.huber),
c(lamb.ls, lamb.lad, lamb.huber))
colnames(out500_1)=c("LS","LAD","Huber")
rownames(out500_1)=c("B0","B1","Sigma2","Lambda")
```

```
save(out500_1, file = "out500_1.Rdata")

##..................... n=1000

out1000_1 <- rbind(c(fit.ls$coefficients[1],
fit.lad$coefficients[1],fit.huber$coefficients[1]),
c(fit.ls$coefficients[2],fit.lad$coefficients[2],
fit.huber$coefficients[2]),
c(sigma.ls, sigma.lad, sigma.huber),
c(lamb.ls, lamb.lad, lamb.huber))
colnames(out1000_1)=c("LS","LAD","Huber")
rownames(out1000_1)=c("B0","B1","Sigma2","Lambda")
save(out1000_1, file = "out1000_1.Rdata")

############################### table of results


ls.I=c(out100_1[1,1],out500_1[1,1],out1000_1[1,1],
out100_1[2,1],out500_1[2,1],out1000_1[2,1],
out100_1[3,1],out500_1[3,1],out1000_1[3,1],
out100_1[4,1],out500_1[4,1],out1000_1[4,1])
ls.II=c(out100_2[1,1],out500_2[1,1],out1000_2[1,1],
out100_2[2,1],out500_2[2,1],out1000_2[2,1],
out100_2[3,1],out500_2[3,1],out1000_2[3,1],
out100_2[4,1],out500_2[4,1],out1000_2[4,1])

lad.I=c(out100_1[1,2],out500_1[1,2],out1000_1[1,2],
out100_1[2,2],out500_1[2,2],out1000_1[2,2],
out100_1[3,2],out500_1[3,2],out1000_1[3,2],
out100_1[4,2],out500_1[4,2],out1000_1[4,2])
lad.II=c(out100_2[1,2],out500_2[1,2],out1000_2[1,2],
out100_2[2,2],out500_2[2,2],out1000_2[2,2],
out100_2[3,2],out500_2[3,2],out1000_2[3,2],
out100_2[4,2],out500_2[4,2],out1000_2[4,2])
```

```
huber.I=c(out100_1[1,3],out500_1[1,3],out1000_1[1,3],
out100_1[2,3],out500_1[2,3],out1000_1[2,3],
out100_1[3,3],out500_1[3,3],out1000_1[3,3],
out100_1[4,3],out500_1[4,3],out1000_1[4,3])
huber.II=c(out100_2[1,3],out500_2[1,3],out1000_2[1,3],
out100_2[2,3],out500_2[2,3],out1000_2[2,3],
out100_2[3,3],out500_2[3,3],out1000_2[3,3],
out100_2[4,3],out500_2[4,3],out1000_2[4,3])

table=cbind(ls.I,ls.II,lad.I,lad.II,huber.I,huber.II)
colnames(table)=c("ls.I","ls.II","lad.I","lad.II","huber.I","huber.II")

nn = rep(c(100,500,1000),times=4)
model=rep(c("I","II"), times=3)
parameters=c("B0","B1","Sigma2","Lambda")

table=data.frame(table)
#save(table, file = "table.RData")
xtable(cbind(parameters,nn,table))
```

آ.۲ کدهای مربوط به جدول یک نمودارهای یک الی سه

```
############################## Model II

rm(list=ls())

library(mvtnorm)
library(MASS)

set.seed(8888)
Beta_0 = 1 ; Beta_1 = 2
n =1000
```

```
ro = -.5
cov.0 = ro * sqrt(1*16)
R <- matrix(c(1, -2, -2, 16),
nrow = 2, ncol = 2)

mu <- c(X = 0, Y = 0)
YZ <- rmvnorm(n, mean = mu, sigma = R)

u = runif(n)

###################### First model (I)
#x = 5 * u
###################### Second model (II)
x = c(); for (t in 1:n) { x[t] = sin(2*t) + 1.5 * u[t]}
######################### Generate Linear model with NSD errors
e = YZ[,1]+YZ[,2]
y = Beta_0 + Beta_1 * x + e

#plot(x,y)



fit <- fitdistr(y, densfun="normal")
fit
hist(y, pch=20, breaks=13, prob=TRUE, main="" , xlab = "Normal Distribution")
curve(dnorm(x, fit$estimate[1], fit$estimate[2]), col="red", lwd=2, add=T)
par(mfrow=c(1,2))

#################### LS
fit.ls = lm(y~x)
var.ls = var(predict(fit.ls)-y)

res1<-predict(fit.ls)-y
fit <- fitdistr(res1, densfun="normal")
fit
```

```
hist(res1, pch=20, breaks=13, prob=TRUE, main="(a) Histogram of LS Residuals",
 xlab = "Model II")
curve(dnorm(x, fit$estimate[1], fit$estimate[2]), col="red", lwd=2, add=T)

################## LAD
library(L1pack)
fit.lad = lad(y~x,  method = "EM")
var.lad = var(predict(fit.lad)-y)

res2<-predict(fit.lad)-y
fit <- fitdistr(res2, densfun="normal")
fit
hist(res2, pch=20, breaks=13, prob=TRUE, main="(a) Histogram of LAD Residuals",
 xlab = "Model II")
curve(dnorm(x, fit$estimate[1], fit$estimate[2]), col="red", lwd=2, add=T)

###################### Huber
library(MASS)

fit.huber <- rlm(y~x, scale.est = "Huber", method = "M")
var.huber = var(predict(fit.huber)-y)

res3<-predict(fit.huber)-y
fit <- fitdistr(res3, densfun="normal")
fit
hist(res3, pch=20, breaks=13, prob=TRUE, main="(a) Histogram of Huber Residuals",
 xlab = "Model II")
curve(dnorm(x, fit$estimate[1], fit$estimate[2]), col="red", lwd=2, add=T)

######################### another way of lad
#library("quantreg")
#fit.lad2 <- rq(y ~ x)
################################### ecdf plots
```

```
par(mfrow=c(1,2))
plot(ecdf(e), xlab="Random Model II", ylab = "Distribution function", main = "")
lines(ecdf(res1), col = 2)
lines(ecdf(res2), col = 3)
lines(ecdf(res3), col = 4)
legend("bottomright", c("NSD errors", "LS Residuals","LAD Residuals",
"Huber Residuals"), col=c(1,2,3,4),
lty= 1:4, cex = 0.55)

######################### Mn Functions

Mn.ls <- sum((y-Beta_0-Beta_1*x)^2)/2-sum((y-fit.ls$coefficients[1]
-fit.ls$coefficients[2]*x)^2)/2

Mn.lad <- sum(abs(y-Beta_0-Beta_1*x))-sum(abs(y-fit.lad$coefficients[1]
-fit.lad$coefficients[2]*x))

w0=y-Beta_0-Beta_1*x
w1=y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x

#.................. Find the Best k in Huber

MAR = median(abs(predict(fit.huber)-y))   #median absolute residual
k = 1.345 * (MAR/ 0.6745)
k
#.................................

Mn.huber <- sum(0.5*((w0^2)*ifelse(abs(w0)<=k,1,0))+
(k*abs(w0)-0.5*(k^2))*ifelse(abs(w0)>k,1,0))-
sum(0.5*((w1^2)*ifelse(abs(w1)<=k,1,0))+
(k*abs(w1)-0.5*(k^2))*ifelse(abs(w1)>k,1,0))

Mn.ls; Mn.lad; Mn.huber
```

```
############################ Sigma

sigma.ls = n^(-1)*sum((y-fit.ls$coefficients[1]-fit.ls$coefficients[2]*x)^2)

sigma.lad = n^(-1)*sum((sign(y-fit.lad$coefficients[1]-fit.lad$coefficients[2]*x))^2)

sigma.huber = n^(-1)*sum(((-k*ifelse((y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x)< -k,1,0))+
(y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x)
*ifelse(abs(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x)<=k,1,0)
+k*ifelse((y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x)>k,1,0))^2)

############################ Lambda
h=1
lamb.ls = (2*n*h)^(-1)*sum((y-fit.ls$coefficients[1]-fit.
ls$coefficients[2]*x+h)-(y-fit.ls$coefficients[1]-fit.ls$coefficients[2]*x-h))

lamb.lad = (2*n*h)^(-1)*sum(sign(y-fit.lad$coefficients[1]-fit.
lad$coefficients[2]*x+h)-sign(y-fit.ls$coefficients[1]-fit.ls$coefficients[2]*x-h))

lamb.huber = (2*n*h)^(-1)*sum((-k*ifelse((y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x+h)< -k,1,0))+
(y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x+h)*ifelse(abs(y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x+h)<=k,1,0)
+k*ifelse((y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x+h)>k,1,0)
-(-k*ifelse((y-fit.huber$coefficients[1]
-fit.huber$coefficients[2]*x-h)< -k,1,0))+
(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x-h)
*ifelse(abs(y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x-h)<=k,1,0)
+k*ifelse((y-fit.huber$coefficients[1]-fit.huber$coefficients[2]*x-h)>k,1,0))

################################################
```

```
#                          Model II
################################################## Table 2
X2_ls = 2*lamb.ls*(1/sigma.ls)*Mn.ls

X2_lad = 2*lamb.lad*(1/sigma.lad)*Mn.lad

X2_huber = 2*lamb.huber*(1/sigma.huber)*Mn.huber

alpha = 0.05

pchisq(X2_ls, df=2, lower.tail = F)
pchisq(X2_lad, df=2, lower.tail = F)
pchisq(X2_huber, df=2, lower.tail = F)



################################################ Table 1
##..................... n=100

out100_2 <- rbind(c(fit.ls$coefficients[1],fit.
lad$coefficients[1],fit.huber$coefficients[1]),
c(fit.ls$coefficients[2],fit.lad$coefficients[2],fit.huber$coefficients[2]),
c(sigma.ls, sigma.lad, sigma.huber),
c(lamb.ls, lamb.lad, lamb.huber))
colnames(out100_2)=c("LS","LAD","Huber")
rownames(out100_2)=c("B0","B1","Sigma2","Lambda")
save(out100_2, file = "out100_2.Rdata")

##..................... n=500

out500_2 <- rbind(c(fit.ls$coefficients[1],
fit.lad$coefficients[1],fit.huber$coefficients[1]),
c(fit.ls$coefficients[2],fit.lad$coefficients[2],fit.huber$coefficients[2]),
c(sigma.ls, sigma.lad, sigma.huber),
c(lamb.ls, lamb.lad, lamb.huber))
```

```
colnames(out500_2)=c("LS","LAD","Huber")
rownames(out500_2)=c("B0","B1","Sigma2","Lambda")
save(out500_2, file = "out500_2.Rdata")

##..................... n=1000

out1000_2 <- rbind(c(fit.ls$coefficients[1],fit.lad$coefficients[1],
fit.huber$coefficients[1]),
c(fit.ls$coefficients[2],fit.lad$coefficients[2],fit.huber$coefficients[2]),
c(sigma.ls, sigma.lad, sigma.huber),
c(lamb.ls, lamb.lad, lamb.huber))
colnames(out1000_2)=c("LS","LAD","Huber")
rownames(out1000_2)=c("B0","B1","Sigma2","Lambda")
save(out1000_2, file = "out1000_2.Rdata")
```

Aabstract

In general, the robustness of a statistical method can be defined as its insensitivity to mild departures from the basic assumptions. In other words, if the characteristics of a statistical method are not affected by a small deviation from the basic assumptions, the method is described as robust. Because outliers strongly influence classical estimates of regression coefficients in linear models, robust estimators such as M-estimators are suggested. The level and power of classical tests are also significantly reduced in the presence of suspicious data. For solving the problem, robust tests are introduced. To perform a robust test about regression parameters, the asymptotic distribution of test statistic is determined. The methods are evaluated based on the reliability of the model estimators and power of tests. In addition to examining some basic features of negatively super-additive dependent random variables in this thesis, we study robust M-tests in linear models with negatively super-additive dependent errors under certain conditions. Using Monte Carlo simulation studies, we compare classic and robust methods based on estimators' reliability and power of tests as well.

Keywords: Robust tests, negatively super-additive dependent random variables, linear model, Monte Carlo simulation.

Faculty Of Mathematical Sciences

MSc Thesis in Mathematical Statistics

Robust M-tests in linear models with negatively superadditive dependent errors

By: Samane Darvishi Bohloli

Supervisor:

Negar Eghbal

Advisor:

Hossein Baghishani

February 2021