

حاشا لله
الرحمن الرحيم



دانشکده علوم ریاضی

رساله دکتری آمار

برآوردگرهای بهبود یافته در برخی مدل‌های رگرسیونی خطی جریمه‌شده

نگارنده: مینا نوروزی‌راد

استاد راهنما

دکتر محمد آرشی

اسفند ۱۳۹۶

تقدیم خالصانہ و خاضعانہ بہ:

پدرم بہ پاس سال ہا حمایت ہا بی بی درغش
مادرم بہ پاس سال ہا صبوری ہا بی انتہائش
خواہرم بہ پاس سال ہا مہربانی ہا فراوانش

و

استادم

بہ پاس وہ سال صبوری ہا و چشم پوشی ہائش

سپاسگزاری

منت خدای را عزوجل که طاعتش موجب قربت است و به شکر اندرش فرید نعمت.

به مصداق لم یشکر المخلوق، لم یشکر الخالق، سپاس بی کران 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 راهم باشند. اگر پدرم و حمایت‌های بی دریغش نبود، این سطرها هیچ‌گاه نگاشته نمی‌شد.

سال ۸۴، همان سال‌هایی که دانشجوی کارشناسی ریاضی محض بودم، استاد آمار جوانی به دانشکده ریاضی پیوست. تقدیر به‌گونه‌ای رقم خورد که در مهرماه سال ۸۵ در اولین کلاس او در دانشکده ریاضی حضور داشته باشم. آن روزها هیچ‌کس نمی‌دانست و حتی تصور نمی‌کرد دانشجوی بیش‌فعال آن کلاس، امروز از رساله دکتری‌اش در رشته آمار دفاع کند. بی‌نهایت سپاس گزارم از استاد خوبم، جناب آقای دکتر محمدرضا ربیعی که بی‌شک دروازه ورود به مسیر آمار را او نشانم داد و اولین گام‌ها را در پیمودن این مسیر سخت به یاری او برداشته‌ام.

در ادامه راه حضور در کلاس تک ستارگان آسمان علم آمار ایران، دکتر حسن صادقی، دکتر ابوالقاسم بزرگ‌نیا و دکتر حسنعلی آذرنوش، گام برداشتن در این مسیر سخت را برایم آسان‌تر نمود. به‌ویژه از دکتر حسن صادقی که سپیدی را بر تخته سیاه زندگیم نگاشت و زندگی، بودن، امید، انگیزه و به‌ویژه انسان بودن را به من آموخت، بی‌نهایت سپاس‌گزارم.

به جرات می‌توانم بگویم خدا از روی رحمت بی‌کرانش، استادی نصیبم ساخته است که نزدیک به ده سال است صبورانه، با ارائه رهنمودها، انتقادهای و پیشنهادهایش مرا حمایت و تشویق می‌کند و راه را نشانم می‌دهد. بوسه می‌زنم بر دستانش که با همه وجودش سال‌هاست که مرا تحمل می‌کند و می‌آموزد تا بیاموزم. سپاسگزاری و قدردانی خالصانه و خاضعانه‌ام نثار استاد عزیزم، دکتر محمد آرشی که با حمایت‌های بی‌دریغش تاکنون هیچ‌گاه تنه‌ایم نگذاشته‌اند و با صبر و تلاش فراوان ایشان، توانستم پایان‌نامه کارشناسی‌ارشد و این رساله دکتری را گردآوری نمایم.

از اساتید محترم، دکتر احمد نزاکتی، دکتر داود شاهسونی، دکتر حسین باغیشنی که در طول دوران تحصیلی‌ام در دوره دکتری صادقانه و عاشقانه تلاش نمودند تا برایم نقالی دانسته‌ها را به نقادی اندیشه‌ها تبدیل کنند، قدردانی می‌کنم. از استاد و دوست خوبم، دکتر نگار اقبال به خاطر همه حرف‌ها و راهنمایی‌های با ارزششان از صمیم قلب تشکر می‌کنم. از مدیر گروه دلسوز و پرتلاش گروه آمار دانشگاه صنعتی شاهرود، جناب آقای دکتر ربیعی به دلیل خدمات، راهنمایی‌ها و زحماتشان سپاسگزارم.

بر خود واجب می‌دانم از زحمات پروفسور ابوالخیر محمد احسانز صالح (دانشگاه کارلتون کانادا)، پروفسور سید اجاز احمد (دانشگاه براک کانادا)، دکتر محمد سخوات حسین (دانشگاه وینیپگ کانادا) و دکتر بهادر یوزباشی (دانشگاه اینئو ترکیه) که در استخراج مقالاتی از این رساله تلاش زیادی کرده‌اند، صمیمانه تشکر کنم.

از اساتید داور محترم، جناب آقای دکتر محسن محمدزاده (دانشگاه تربیت مدرس تهران) و دکتر فرزاد اسکندری (دانشگاه علامه طباطبایی تهران)، دکتر محمدرضا ربیعی (دانشگاه صنعتی شاهرود) که با حضور دلگرمشان تصحیح و داوری این رساله را برعهده گرفته‌اند، سپاسگزارم.

و دوباره، بارها و بارها از همدلی و همراهی و همگامی **مهربان خانواده‌ام** که همواره با قلبی آکنده از عشق و معرفت، محیطی سرشار از سلامت، امنیت، آرامش و آسایش برایم فراهم آورده‌اند و مرا در رسیدن به اهدافم یاری رساندند بی‌نهایت سپاس گزارم.

مینا نوروزی‌راد

اسفند ۱۳۹۶

تعهد نامه

اینجانب **مینا نوروزی راد** دانشجوی دکتری رشته **آمار** دانشکده **علوم ریاضی** دانشگاه شاهرود، نویسنده پایان نامه با عنوان **برآوردگرهای بهبودیافته در برخی مدل های رگرسیونی خطی جریمه شده**، تحت راهنمایی **محمد آرشی** متعهد می شوم:

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش های دیگر پژوهش گران، به مرجع مورد استفاده استناد شده است.
- مطالب این پایان نامه، تا کنون توسط خود، یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا 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 به دست آوردن نتایج اصلی پایان نامه تاثیرگذار بوده اند، در مقالات مستخرج از پایان نامه رعایت می گردد.
- در تمام مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافت های آنها) استفاده شده است، ضوابط و اصول اخلاقی رعایت شده است.
- در تمام مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته (یا استفاده شده است)، اصل رازداری و اصول اخلاق انسانی رعایت شده است.

مینا نوروزی راد

اسفند ۱۳۹۶

مالکیت نتایج و حق نشر

- تمام حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می باشد. این مطلب باید به نحو مقتضی، در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در این پایان نامه بدون ذکر منبع مجاز نمی باشد.

چکیده

در این رساله، اطلاعات پیشین غیرقطعی (غیرنمونه‌ای) در برآورد پارامترها در مدل‌های رگرسیونی خطی جریمه‌شده ریج، لاسو و استوار جریمه‌شده (M ریج) به مدل تزریق شده که منجر به ظهور برآوردگرهای آزمون اولیه، انقباضی نوع استاین و جزء مثبت انقباضی نوع استاین گردیده است. ویژگی‌های این برآوردگرهای بهبودیافته با در نظر گرفتن چندین محدودیت در مدل‌های رگرسیونی M ریج و لاسو بررسی شده است. همچنین، از اطلاعات فرعی به‌دست آمده از رگرسیون جریمه‌شده برای ساختن برآوردگرهای انقباضی در حضور داده‌های پرت و نقاط اهرمی بهره گرفته شده است. در تمامی مدل‌های رگرسیونی بیان شده، نتایج به‌دست آمده، به کمک مطالعه‌های شبیه‌سازی و تحلیل داده‌های واقعی مورد بررسی قرار گرفته که برآوردگرهای انقباضی ارائه شده در تمامی موارد بهتر از برآوردگر معمول در مدل مورد بررسی رفتار می‌کند.

کلمات کلیدی: برآوردگر استوار، برآوردگر آزمون اولیه، برآوردگر انقباضی، رگرسیون خطی، ریج، لاسو.

لیست مقالات مستخرج از پایان نامه

1. نوروزی‌راد، م. و آرشی، م. (۱۳۹۶). مطالعه رفتار حدی برآوردگرهای انقباضی در مدل رگرسیونی توانیده با نرم مستطیلی، *مجله علوم آماری*، ۱۱(۱)، ۱۴۹-۱۷۴.
2. **Norouzirad, M.**, Arashi, M. and Ahmed, S. E. (2017). Improved robust ridge M-estimation, *Journal of Statistical Computations and Simulation*, **87**(18), 3469-3490.
3. **Norouzirad, M.** and Arashi, M. (2016). Preliminary test and Stein-type shrinkage ridge estimators in robust regression. *Statistical Papers*, DOI: 10.1007/s00362-0899-3.
4. **Norouzirad, M.**, Hossain, S. and Arashi, M. (2018). Shrinkage and penalized estimators in weighted least absolute deviations regression models, *Journal of Statistical Computation and Simulation*, **88**(8), 1557-1575.
5. **Norouzirad, M.** and Arashi, M. (2017) Preliminary test and Stein-type shrinkage LASSO-based estimators, *SORT-Statistics and Operations Research Transactions*, Accepted.
6. **Norouzirad, M.**, Arashi, M. and Ahmed, S. E. (2017) Asymptotic Properties of Shrinkage LAD estimators, 11th Seminar on Probability and Stochastic Processes, Qazvin, Iran.
7. Hossain, S., **Norouzirad, M.**, and Arashi, M. (2017). Estimation Strategies for Weighted Least Absolute Deviations Regression Models, 45th Annual Meeting of the Statistical Society of Canada, Winnipeg, Canada.
8. Arashi, M. and **Norouzirad, M.** (2016). Steinian Shrinkage Estimation in High Dimensional Regression, 13th Iranian Statistical Inference, Shahid Bahonar University of Kerman, Kerman, Iran.
9. **Norouzirad, M.** and Arashi, M. (2016). From Lad-LASSO to PR-LAD Estimator, 1nd Seminar on Nonparametric Statistics and its Application, Allame Tabatabaeyi University, Tehran, Iran.
10. Arashi, M and **Norouzirad, M.** (2015). Improved ridge M-estimators, 46th Annual Iraninan Mathematics Conference, Yazd University, Yazd, Iran.
11. **Norouzirad, M.** , Arashi, M. and Saleh, A. K. Md. E. (2015). Some Asymptotic Results on Improved LASSO Estimators, 10th Seminar on Probability and Stochastic Process, Yazd University, Yazd, Iran.
12. **Norouzirad, M.** and Arashi, M. and Saleh, A. K. Md. Ehsanes (2015). Shrinkage estimation through convex optimization, 8th International Conference of the Iranian Society of Opertions research, Ferdowsi University, Mashhad, Iran.

پیش‌گفتار

استاین (۱۹۵۶) نشان داد در توزیع نرمال، برای بعد فضای پارامتر حداقل ۳، میانگین نمونه، برآوردگر غیرمجاز برای پارامتر میانگین توزیع نرمال است. تاکنون مطالعات زیادی در راستای بهبود برآوردگر میانگین در این حالت انجام شده اما هنوز برآوردگر مجاز به دست نیامده است. معمولاً در مدل رگرسیونی خطی فرض می‌شود توزیع خطاها نرمال است که در این حالت، یکی از متداول‌ترین روش‌های برآورد ضرایب رگرسیونی، کمترین توان‌های دوم (OLS) است. اما در اغلب مسائل کاربردی، برقراری پذیره‌های رگرسیونی چالش‌برانگیز است.

در مدل رگرسیونی خطی، وقتی خطاها نرمال باشد، اگر بین ستون‌های ماتریس متغیرهای توضیحی، رابطه خطی وجود داشته باشد، که در اصطلاح رگرسیون، آن را هم‌خطی چندگانه می‌نامند، برآوردگر OLS از کارایی قابل قبولی برخوردار نیست. روش‌های متعددی برای مقابله با مسئله هم‌خطی پیشنهاد شده است که معروف‌ترین روش، ریج (**هورل و کنارد، ۱۹۷۰**) است. این برآوردگر، با اعمال محدودیت بر روی فضای پارامتر و حل مسئله رگرسیون جریمه‌شده تحت این محدودیت، به دست می‌آید.

اگر در مدل رگرسیون خطی، توزیع خطاها نرمال نباشد، احتمال حضور داده‌های پرت و نقاط اهرمی در داده‌ها افزایش می‌یابد که از کارایی برآوردگر OLS می‌کاهد. این موضوع، انگیزه ابداع برآوردگرهای استوار در مدل رگرسیونی بوده است که در این راستا، می‌توان به برآوردگرهای LAD و نیز برآوردگرهای M اشاره نمود.

در حالتی که تعداد متغیرها در یک مدل رگرسیونی زیاد است، حضور همه متغیرها در مدل، تفسیرپذیری آن را دشوار می‌کند. در واقع، حضور متغیرهایی که با متغیر پاسخ در مدل ارتباطی معنی‌داری ندارند، منجر به برآوردگری با واریانس زیاد می‌شود. یکی از روش‌های مدرن انتخاب یک زیرمجموعه از متغیرهای توضیحی بااهمیت و تاثیرگذار در مدل، روش پیشنهادی **تیبشیرانی (۱۹۹۶)** است. برآوردگر به دست آمده با این روش، برآوردگر لاسو نام دارد.

گاهی اطلاعات پیشین یا غیرنمونه‌ای نیز وجود دارد که در روش‌های کلاسیک آماری معمولاً نادیده گرفته می‌شود. استفاده از آن‌ها در رویه برآورد، منجر به به وجود آمدن روش‌های انقباضی برای برآورد پارامترها شده است. در برآوردگرهای انقباضی، با ترکیب اطلاعات پیشین با برآوردگرهای متداول (برآوردگر OLS و) می‌توان مسئله برآورد را بهبود بخشید.

مسئله‌ای را در نظر بگیرید که در داده‌ها هم‌خطی و داده پرت به‌طور هم‌زمان وجود داشته باشد. در این حالت، معمولاً از روش‌های رنج استوار استفاده می‌شود. در این رساله، با استفاده از روش‌های انقباضی، این برآوردگرها را با در نظر گرفتن اطلاعات پیشین متفاوت بهبود بخشیده‌ایم. این ایده را ابتدا در مقاله **آرشی و نوروزی‌راد (۲۰۱۵)** مطرح کردیم و سپس در مقاله‌های **نوروزی‌راد و همکاران (۲۰۱۷b)** و **نوروزی‌راد و آرشی (۲۰۱۷b)** با جزئیات بیشتری مورد بررسی قرار دادیم که فصل ۲ رساله به این موضوع اختصاص یافته است.

ایده اصلی در برآوردگر لاسو این است که ضرایب رگرسیونی که معنادار نیستند یا اثرات ضعیف (قابل اغماض) دارند، به سمت بردار صفر منقبض شود. هنگامی که علاوه بر اطلاعات به‌دست آمده از نمونه، از سایر اطلاعات در برآورد پارامترها استفاده گردد، معمولاً نتایج پرباری به‌همراه دارد. در این حالت موضوع دوباره انقباضی را مطرح نمودیم که برآوردگر لاسو را به درون یک زیرفضا از فضای پارامتر که همان محدودیت است، منقبض می‌کنیم. این ایده ابتدا در مقاله **نوروزی‌راد و همکاران (۲۰۱۵a)** با نگاه بهینه‌سازی در یافتن پاسخ مسئله بهینه‌سازی مطرح گردید. در مقاله **نوروزی‌راد و همکاران (۲۰۱۵b)** روش دیگری برای یافتن برآوردگرهای انقباضی لاسو پیشنهاد شد که این ایده در مقاله **نوروزی‌راد و آرشی (۱۳۹۶)** و نیز مقاله **نوروزی‌راد و آرشی (۲۰۱۷a)** برای انواع مختلفی از برآوردگرهای انقباضی کامل گردید که در فصل ۳ این رساله به آن پرداخته‌ایم.

همان‌طور که قبلاً اشاره شد برای ساخت برآوردگرهای انقباضی به اطلاع پیشینی در خصوص پارامتر مدل که عمدتاً برگرفته از نظر متخصص می‌باشد، نیاز داریم. در حالی که اگر این اطلاعات پیشین از قبل وجود نداشته باشد، می‌توان از روش‌های معمول انتخاب متغیر استفاده کرد تا متغیرهای معنی‌دار را شناسایی کنیم. روش‌های جریمه‌شده، در انتخاب متغیرها خودکار هستند و به تحلیل‌گر اجازه نمی‌دهد که فکر کند و خودش مدل مناسب را انتخاب می‌کند. از این رو، در رگرسیون کمترین قدرمطلق انحراف‌ها (LAD) با استفاده از برآوردگر جریمه‌شده LAD-LASSO متغیرهای مهم را شناسایی نموده و از محدودیت صفر قرار دادن متغیرهایی که معنی‌دار نیستند، در مدل برای ساختن برآوردگرهای انقباضی استفاده کردیم که مقاله‌های **نوروزی‌راد و آرشی (۲۰۱۶)** و **نوروزی‌راد و همکاران (۲۰۱۷a)** حاصل کار است. همچنین در حالتی که در مشاهده‌ها نقاط اهرمی وجود داشته باشد، از برآوردگر جریمه‌شده WLAD-LASSO در مدل رگرسیونی WLAD استفاده کرده‌ایم که منجر به استخراج مقاله‌های **حسین و همکاران (۲۰۱۷)** و **نوروزی‌راد و همکاران (۲۰۱۸)** شده است که فصل ۴ رساله، این موضوع را در بر می‌گیرد.

ایده دیگری که در این رساله مورد بررسی قرار داده‌ایم، انتخاب برآوردگر رنج به عنوان یک برآوردگر اولیه و برآوردگر لاسو به عنوان یک محدودیت برای ساختن برآوردگرهای انقباضی در مدل‌های رگرسیونی بعدالاً است که در مقاله **آرشی و نوروزی‌راد (۲۰۱۶)** بررسی کرده‌ایم.

این مجموعه، پنج فصل و دو پیوست دارد:

فصل اول، به مروری بر روش‌های برآورد پارامترهای مدل رگرسیون خطی چندگانه در موارد بیان شده اختصاص دارد.

در **فصل دوم**، با در نظر گرفتن دو اطلاع پیشین مختلف و ویژگی‌های آن‌ها، برآوردگرهای بهبودیافته در مدل رگرسیون خطی در حالی که توزیع خطاها غیرنرمال بوده و داده‌ها هم‌خطی داشته باشند، مورد بررسی قرار می‌گیرد.

فصل سوم، برآوردگرهای بهبودیافته لاسو را معرفی می‌کند و با انجام مطالعه شبیه‌سازی و تحلیل یک مثال واقعی، ویژگی‌های آن‌ها را بررسی می‌نماید.

در **فصل چهارم**، برآوردگرهای بهبودیافته را در حالتی که (الف) نقاط پرت و (ب) نقاط اهرمی در داده‌ها وجود داشته، با در نظر گرفتن اطلاعات فرعی به جای اطلاعات پیشین، معرفی کرده و ویژگی‌های آن‌ها را از زوایای مختلف مورد بررسی قرار می‌دهیم.

فصل پنجم، پیشنهادات برای پژوهش‌های آینده را در بر می‌گیرد.
پیوست آ، تعاریف، لم‌ها و قضایای اساسی را که در این رساله از آنها استفاده شده است، یادآوری می‌نماید.

پیوست ب، شامل گزیده‌ای از برنامه‌های نوشته شده با نرم‌افزار R است.

فهرست مطالب

ق	فهرست تصاویر
ر	فهرست جداول
ت	فهرست نمادها
ذ	فهرست کلمات اختصاری
۱	۱ برآوردگرهای انقباضی و جریمه‌شده در مدل رگرسیونی خطی چندگانه
۱	۱.۱ مقدمه
۳	۲.۱ روش‌های برآورد پارامترها
۳	۱.۲.۱ پذیره‌های مدل
۷	۲.۲.۱ برآوردگرهای انقباضی
۱۵	۳.۱ تاریخچه برآوردگرهای انقباضی
۱۷	۱.۳.۱ برآوردگرهای جریمه‌شده
۲۴	۲.۳.۱ ترکیب روش‌های انقباضی و جریمه‌شده
۲۵	۴.۱ مدل رگرسیون استوار
۲۹	۲ برآوردگرهای بهبودیافته بر پایه M ریج
۲۹	۱.۲ مقدمه
۳۱	۲.۲ نوآوری مسئله
۳۱	۳.۲ مدل کامل
۳۳	۱.۳.۲ ویژگی برآوردگرها
۳۸	۲.۳.۲ کارایی برآوردگرها: تجزیه و تحلیل مخاطره جانبی
۵۲	۳.۳.۲ انتخاب پارامتر ریج
۵۲	۴.۳.۲ نمایش گرافیکی
۵۸	۵.۳.۲ مطالعه شبیه‌سازی

۶۴	تحلیل مثال‌های کاربردی	۶.۳.۲
۷۱	مدل افراز شده	۴.۲
۷۶	برآوردگرهای بهبودیافته M ریج	۱.۴.۲
۷۷	انتخاب پارامتر ریج	۲.۴.۲
۷۷	مطالعه شبیه‌سازی	۳.۴.۲
۸۱	تحلیل مجدد داده‌های چربی بدن	۴.۴.۲
۸۳	نتیجه‌گیری	۵.۲
۸۵		برآوردگرهای بهبودیافته بر پایه برآوردگر لاسو	۳
۸۵	مقدمه	۱.۳
۸۷	برآوردگرهای انقباضی بر پایه لاسو	۱.۱.۳
۸۸	ویژگی برآوردگرها	۲.۳
۹۶	مطالعه شبیه‌سازی	۳.۳
۹۹	تحلیل داده‌های آمار و ارقام ایالات آمریکا	۴.۳
۱۰۲	نتیجه‌گیری	۵.۳
۱۰۳		برآوردگرهای بهبودیافته در رگرسیون LAD	۴
۱۰۳	مقدمه	۱.۴
۱۰۷	برآوردگرهای انقباضی LAD	۲.۴
۱۱۰	ویژگی برآوردگرها	۱.۲.۴
۱۱۴	مطالعه شبیه‌سازی	۲.۲.۴
۱۱۶	تحلیل داده‌های نرخ رشد	۳.۲.۴
۱۲۰	برآوردگرهای انقباضی LAD موزون	۳.۴
۱۲۲	ویژگی برآوردگرها	۱.۳.۴
۱۲۶	مطالعه شبیه‌سازی	۲.۳.۴
۱۲۹	تحلیل مجدد داده‌های چربی بدن	۳.۳.۴
۱۳۱	نتیجه‌گیری	۴.۴
۱۳۳		پیشنهادات برای پژوهش‌های آینده	۵
۱۳۳	مقدمه	۱.۵
۱۳۴	برآوردگر انقباضی لاسو از دیدگاه بهینه‌سازی	۲.۵
۱۳۶	برآوردگر انقباضی LAD از دیدگاه روش‌های عددی	۳.۵
۱۳۸	برآوردگرهای انقباضی در بعد بالا	۴.۵
۱۴۰	ایده‌های دیگر	۵.۵

۱۴۳	تعاریف، لم‌ها و قضایای اساسی	آ
۱۴۳ جبر ماتریس‌ها	۱.آ
۱۴۵ توابع ریاضی و توزیع‌ها	۲.آ
۱۴۷ همگرایی‌ها	۳.آ
۱۵۰ برخی توابع خاص	۴.آ
۱۵۳	گزیده‌ای از برنامه‌های رایانه‌ای	ب
۱۷۱	مراجع	
۱۹۰	واژه‌نامه فارسی به انگلیسی	

فهرست تصاویر

۲۲	۱.۱	نمودار برآورهای لاسو و ریج
		۱.۲	رفتار توان دوم اریبی مجانبی برآوردگرها برای $p = 5$ ، $\lambda = 0.1$ ، $\eta^2 = 1$ و
۵۳		مقادیر مختلف
		۲.۲	رفتار توان دوم اریبی مجانبی برآوردگرها برای $p = 5$ ، $\alpha = 0.05$ ، $\eta^2 = 1$ و
۵۴		مقادیر مختلف λ
		۳.۲	رفتار مخاطره‌ی مجانبی برآوردگرها برای $p = 5$ ، $\lambda = 0.1$ ، $\eta^2 = 1$ و مقادیر
۵۵		مختلف α
		۴.۲	رفتار مخاطره‌ی مجانبی برآوردگرها برای $p = 5$ ، $\alpha = 0.05$ ، $\eta^2 = 1$ و
۵۵		مقادیر مختلف λ
۶۶	۵.۲	نمودار ماتریسی- نقطه‌ای از متغیرهای مستقل مجموعه داده V02
۶۶	۶.۲	نمودار شناسایی داده‌های پرت در مجموعه داده V02
۶۹	۷.۲	نمودار ماتریسی نقطه‌ای برای متغیرها در مجموعه داده چربی بدن
۷۰	۸.۲	نمودارهای شناسایی داده پرت در مجموعه داده چربی بدن
۱۱۵	۱.۴	توزیع‌های لاپلاس و نرمال
۱۱۸	۲.۴	نمودارها برای شناسایی داده‌های پرت در مجموعه داده نرخ رشد
۱۱۹	۳.۴	چگالی باقیمانده‌های برآوردگرهای LAD در مجموعه داده بارو
۱۳۰	۴.۴	فاصله کوک و نقاط اهرمی برای داده‌های چربی بدن

فهرست جداول

۵۷	کارایی‌های نسبی برآوردگرها برای $p = 5$ ، $\eta^2 = 1$ ، $\alpha = 0.5$ و مقادیر مختلف λ .	۱.۲
۵۸	کارایی‌های نسبی برآوردگرها برای $p = 20$ ، $\eta^2 = 1$ ، $\alpha = 0.5$ و مقادیر مختلف λ .	۲.۲
۶۱	کارایی‌های نسبی برآوردگرهای پیشنهادی به‌ازای $n = 50$ ، $p = 10$ و $\alpha = 0.1$.	۳.۲
۶۲	کارایی‌های نسبی برآوردگرهای پیشنهادی به‌ازای $n = 50$ ، $p = 20$ و $\alpha = 0.1$.	۴.۲
۶۳	کارایی‌های نسبی برآوردگرهای پیشنهادی به‌ازای $n = 50$ ، $p = 5$ و $\alpha = 0.1$.	۵.۲
۶۵	آماره‌های توصیفی برای مجموعه داده V02.	۶.۲
۶۵	مقادیر VIF برای مجموعه داده V02.	۷.۲
۶۷	برآوردگرهای OLS، M، M و جزء مثبت انقباضی M ريج برای مجموعه داده V02.	۸.۲
۶۸	آماره‌های توصیفی از متغیرها در مجموعه داده چربی بدن.	۹.۲
۶۸	مقادیر VIF برای مجموعه داده چربی بدن.	۱۰.۲
۷۰	برآوردگرهای OLS، M، M و جزء مثبت انقباضی M ريج برای داده‌های چربی بدن.	۱۱.۲
۷۰	کارایی نسبی برآوردگرهای پیشنهادی در حالت $n = 50$ ، $\sigma^2 = 1$ ، $\alpha = 0.5$.	۱۲.۲
۷۸	کارایی نسبی برآوردگرهای پیشنهادی در حالت $n = 100$ ، $\sigma^2 = 1$ ، $\alpha = 0.5$.	۱۳.۲
۷۹	کارایی نسبی برآوردگرهای پیشنهادی در حالت $n = 50$ ، $\sigma^2 = 1$ ، $\alpha = 0.5$.	۱۴.۲
۸۰	کارایی نسبی برآوردگرهای پیشنهادی در حالت $n = 100$ ، $\sigma^2 = 1$ ، $\alpha = 0.5$.	۱۵.۲
۸۱	برآوردگرهای M و M ريج برای مجموعه داده‌های چربی بدن.	۱۶.۲
۸۲	مقادیر RPE برآوردگرهای پیشنهادی مجموعه داده چربی بدن.	۱۷.۲
۹۷	کارایی‌های نسبی برآوردگرها برای Δ^2 های ثابت، $\rho = 0$ ، $q = p - 6$ و مقادیر مختلف p .	۱.۳

۲.۳	کارایی‌های نسبی برآوردگرها به ازای Δ^2 های ثابت، $\rho = 0.2$ ، $q = p - 6$ و مقادیر مختلف p	۹۸
۳.۳	کارایی‌های نسبی برآوردگرها به ازای Δ^2 های ثابت، $\rho = 0.9$ ، $q = p - 6$ و مقادیر مختلف p	۹۸
۴.۳	آماره‌های توصیفی از متغیرهای داده‌های آمار و ارقام ایالات آمریکا	۱۰۰
۵.۳	جدول خطاهای پیشگویی نسبی برآوردگرها در داده آمار و ارقام ایالت‌های آمریکا	۱۰۱
۱.۴	مقادیر RMSE شبیه‌سازی شده برای برآوردگرهای PRLAD و SLAD، PTLAD، RLAD برای $n = 50$ و $\alpha = 0.05$ ، $(p_1, p_2) = (3, 5)$	۱۱۶
۲.۴	مقادیر RMSE شبیه‌سازی شده برآوردگرها نسبت به برآوردگر uLAD به‌ازای $p_1 = 3$ و $\Delta^2 = 0$	۱۱۷
۳.۴	آماره‌های توصیفی برای مجموعه داده نرخ رشد	۱۱۸
۴.۴	برآورد ضرایب و کارایی آن‌ها برای مجموعه داده بارو	۱۲۰
۵.۴	مقادیر RMSE برآوردگرها نسبت به برآوردگر UWLAD به‌ازای $p_1 = 3$ و $\Delta^2 = 0$	۱۲۸
۶.۴	مقادیر RMSE برآوردگرهای پیشنهادی نسبت به برآوردگر UWLAD به‌ازای $\Delta^2 \geq 0$ ، $p_1 = 3$ و $p_2 = 15$	۱۲۸
۷.۴	مقادیر RMSE برآوردگرهای پیشنهادی نسبت به برآوردگر UWLAD به‌ازای $\Delta^2 = 0$ و $p_1 = 6$	۱۲۹
۸.۴	برآوردها (خطای استاندارد) برای ضرایب و RMSE بر اساس شبیه‌سازی بوت‌استرپ نسبت به برآوردگر UWLAD	۱۳۱

فهرست نمادها

$\text{tr}(\mathbf{A})$	اثر ماتریس A
\mathbb{R}	اعداد حقیقی
$I(f)$	اطلاع فیشر تابع f
$E[\cdot]$	امید ریاضی
E_R	امید ریاضی متغیر پواسون R با پارامتر $\Delta^2/2$
$\hat{\beta}^M$	برآوردگر M
$\hat{\beta}^{\text{PT}}$	برآوردگر آزمون اولیه
$\hat{\beta}^{\text{ptLAD}}$	برآوردگر آزمون اولیه LAD
$\hat{\beta}^{\text{ptWLAD}}$	برآوردگر آزمون اولیه LAD موزون
$\hat{\beta}^{\text{ptM}}$	برآوردگر آزمون اولیه M
$\hat{\beta}^{\text{ptM}}(\lambda)$	برآوردگر آزمون اولیه M - ریج
$\hat{\beta}^{\text{ptLso}}$	برآوردگر آزمون اولیه لاسو
$\hat{\beta}^{\text{SR}}$	برآوردگر انقباضی محدودشده
$\hat{\beta}^{\text{S}}$	برآوردگر انقباضی نوع - استاین
$\hat{\beta}^{\text{sLAD}}$	برآوردگر انقباضی نوع استاین LAD
$\hat{\beta}^{\text{sWLAD}}$	برآوردگر انقباضی نوع استاین LAD موزون
$\hat{\beta}^{\text{sM}}$	برآوردگر انقباضی نوع استاین M
$\hat{\beta}^{\text{sM}}(\lambda)$	برآوردگر انقباضی نوع استاین M - ریج
$\hat{\beta}^{\text{sLso}}$	برآوردگر انقباضی نوع استاین لاسو
$\hat{\beta}^{\text{IPT}}$	برآوردگر بهبودیافته آزمون اولیه
$\hat{\beta}^{\text{iptM}}$	برآوردگر بهبودیافته آزمون اولیه M
$\hat{\beta}^{\text{iptM}}(\lambda)$	برآوردگر بهبودیافته آزمون اولیه M - ریج
$\hat{\beta}^{\text{iptLso}}$	برآوردگر بهبودیافته آزمون اولیه لاسو
$\hat{\beta}^{\text{S+}}$	برآوردگر جزء مثبت انقباضی نوع استاین
$\hat{\beta}^{\text{sLAD+}}$	برآوردگر جزء مثبت انقباضی نوع استاین LAD
$\hat{\beta}^{\text{sWLAD+}}$	برآوردگر جزء مثبت انقباضی نوع استاین LAD موزون

$\hat{\beta}^{sM+}$	برآوردگر جزء مثبت انقباضی نوع استاین M
$\hat{\beta}^{sM+}(\lambda)$	برآوردگر جزء مثبت انقباضی نوع استاین M-ریج
$\hat{\beta}^{sLso+}$	برآوردگر جزء مثبت انقباضی نوع استاین لاسو
$\hat{\beta}^{Rdg}$	برآوردگر ریج
$\hat{\beta}^{PT}(\lambda)$	برآوردگر ریج آزمون اولیه
$\hat{\beta}^U(\lambda)$	برآوردگر ریج جزء مثبت نوع استاین
$\hat{\beta}^R(\lambda)$	برآوردگر ریج محدود شده
$\hat{\beta}^U(\lambda)$	برآوردگر ریج محدود نشده
$\hat{\beta}^{Ols}$	برآوردگر کمترین توان های دوم
$\hat{\beta}^{Lso}$	برآوردگر لاسو
$\hat{\beta}^R$	برآوردگر محدود شده
$\hat{\beta}^{rLAD}$	برآوردگر محدود شده LAD
$\hat{\beta}^{rWLAD}$	برآوردگر محدود شده LAD موزون
$\hat{\beta}^{rM}$	برآوردگر محدود شده M
$\hat{\beta}^{rM}(\lambda)$	برآوردگر محدود شده M-ریج
$\hat{\beta}^{rLso}$	برآوردگر محدود شده لاسو
$\hat{\beta}^U$	برآوردگر محدود نشده
$\hat{\beta}^{uLAD}$	برآوردگر محدود نشده LAD
$\hat{\beta}^{rWLAD}$	برآوردگر محدود نشده LAD موزون
$\hat{\beta}^{uM}$	برآوردگر محدود نشده M
$\hat{\beta}^{uM}(\lambda)$	برآوردگر محدود نشده M-ریج
$\hat{\beta}^{uLso}$	برآوردگر محدود نشده لاسو
β	بردار پارامتر
y_n	بردار پاسخ
ε_n	بردار خطا
max	بیشینه
$\Phi(\cdot; \mu, \Sigma)$	تابع توزیع تجمعی $\mathcal{N}_k(\mu, \Sigma)$
$\phi(\cdot; \mu, \Sigma)$	تابع چگالی $\mathcal{N}_k(\mu, \Sigma)$
$I(A)$	تابع نشان گر مجموعه A
A^T	ترانواده ماتریس A
F_{k_1, k_2}	توزیع فیشر مرکزی با k_1 و k_2 درجه آزادی
$F_{k_1, k_2}(\Delta^2)$	توزیع فیشر مرکزی با k_1 و k_2 درجه آزادی و پارامتر غیرمرکزی Δ^2
χ_k^2	توزیع کی دو مرکزی با k درجه آزادی
$\chi_k^2(\Delta^2)$	توزیع کی دو غیرمرکزی با k درجه آزادی و پارامتر غیرمرکزی Δ^2

$\mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	توزیع نرمال k متغیره با بردار میانگین $\boldsymbol{\mu}$ و ماتریس واریانس-کوواریانس $\boldsymbol{\Sigma}$
α	سطح معنی داری
$\kappa(\mathbf{A})$	عدد شرطی ماتریس \mathbf{A}
a^+	قسمت مثبت a
min.	کمینه
$\mathbf{A}(m \times n)$	ماتریس با m سطر و n ستون
\mathbf{X}_n	ماتریس طرح
\mathbf{I}_n	ماتریس همانی از مرتبه $n \times n$
\mathbf{A}^{-1}	معکوس ماتریس \mathbf{A}
argmin $f(\cdot)$	مولفه‌ای از دامنه‌ی تابع f است که منجر به کمینه شدن تابع f می‌شود.
$\ \cdot\ $	نرم اقلیدسی
Var	واریانس
\sim	هم‌توزیع
$\xrightarrow{\mathcal{P}}$	همگرایی در احتمال
$\xrightarrow{\mathcal{D}}$	همگرایی در توزیع

فهرست کلمات اختصاری

aLASSO (Adaptive Least Absolute Shrinkage and Selection Operator)

ADB (Asymptotic Distributional Bias)

ADF (Asymptotic Distributional Function)

ADQB (Asymptotic Distributional Quadratic Bias)

ADR (Asymptotic Distributional Risk)

BLUE (Best Linear Unbiased Estimator)

BP (Breakdown Point)

DES (Data Embedded Sample)

DFFITs (Difference in Fits)

IPTE (Improved Preliminary Test Estimator)

IPTRME (Improved Preliminary Test Ridge M Estimator)

IPTLE (Improved Preliminary Test LASSO Estimator)

LAD (Least Absolute Deviations)

LAR (Least Angle Regression)

LASSO (Least Absolute Shrinkage and Selection Operator)

LAV (Least Absolute Values)

LMS (Least Median of Squares)

LTS (Least Trimmed Squares)

MAD (Median Absolute Deviations)

MLE (Maximum Likelihood Estimator)

MSAE (Minimum Sum of Absolute Errors)

NSI (Non-Sample Information)

OLSE (Ordinary Least Squares Estimator)

PE (Prediction Error)

PLS (Penalized Least Squares)

PTE (Preliminary Test Estimator)

PTLE (Preliminary Test LASSO Estimator)
PTME (Preliminary Test M Estimator)
PTRME (Preliminary Test Ridge M Estimator)
PRSME (Positive-rule Stein-type M Estimator)
PRRME(Positive-rule Stein-type Ridge M estimator)
PRSLE (Positive-rule Stein-type LASSO estimator)
PRSE(Positive-rule Stein-type estimator)
RE (Restricted Estimator)
REff (Relative Efficiency)
RLSE (Restricted Least Squares Estimator)
RLE (Restricted LASSO Estimator)
RME (Restricted M Estimator)
RPE (Relative Prediction Error)
RRME (Restricted Ridge M Estimator)
SE (Shrinkage estimator)
SEL (Square Error Loss)
SME (Stein-type M Estimator)
SRE (Shrinkage Restricted Estimator)
SRME (Stein-type Ridge M Estimator)
SSE (Stein-type shrinkage estimator)
SUR (Seemingly Unrelated Regression)
UE (Unrestricted Estimator)
ULE (Unrestricted LASSO Estimator)
UME (Unrestricted M Estimator)
URME (Unrestricted Ridge M Estimator)
UPI (Uncertain Prior Information)
VIF (Variance Inflation Factor)
WLAD (Weighted Least Absolute Deviations)

فصل ۱

برآوردگرهای انقباضی و جریمه‌شده در مدل رگرسیونی خطی چندگانه

در این فصل، مفاهیم استفاده شده در این رساله به اختصار بیان شده است. مطالب این فصل از منابع متعددی برداشت شده است که از آن جمله می‌توان به **مونت‌گمری و همکاران (۲۰۱۱)**، **صالح (۲۰۰۶)**، **هورل و کنارد (۱۹۷۰)**، **تیشیرانی (۱۹۹۶)** و **هرماتزیگر و مک‌گین (۱۹۹۸)** اشاره نمود.

۱.۱ مقدمه

تحلیل رگرسیون، ابزار آماری مهمی برای بررسی و مدل‌سازی ارتباط بین متغیرها در مجموعه‌ای از داده‌ها است. روش‌های رگرسیونی، به‌صورت گسترده‌ای با هدف پیش‌گویی به‌کار برده می‌شوند و بدین جهت، تقریباً در هر زمینه‌ای می‌توان ردپایی از آن‌ها را مشاهده نمود. پیرو این مطالب می‌توان اشاره کرد که اولین بار رگرسیون به‌عنوان یک مفهوم زیست‌شناختی در مقاله **گالتون (۱۸۷۷)** مطرح و بعدها توسط **پیرسون (۱۸۹۷، ۱۹۰۱)** برای مفاهیم آماری توسعه داده شد (**فینی، ۱۹۹۶؛ سنتون، ۲۰۰۱**).

ساختار رگرسیون بدین صورت است که متغیر وابسته یا پاسخ Y با یک یا چند متغیر توضیحی (تبیینی یا پیشگو) $X = (X_1, \dots, X_p)$ مرتبط می‌باشد. رگرسیون خطی را می‌توان امیدریاضی شرطی Y به شرط X تعریف نمود که با نماد $E[Y|X]$ نشان داده می‌شود. معمولاً داده‌ها به‌صورت مجموعه $\{(x_1, y_1), \dots, (x_n, y_n)\}$ که $x_i = (x_{i1}, \dots, x_{ip})^T \in \mathbb{R}^p$ ، متغیر i ام

توضیحی و y_i مولفه i ام پاسخ است، در دسترس هستند. در بیشتر مسائل کاربردی، معمولاً متغیرهای توضیحی با دقت بالایی اندازه‌گیری می‌شوند و این در حالی است که متغیر پاسخ، شامل مولفه‌های سیگنال و نوفه یا خطا است که به‌سختی از هم تفکیک می‌شوند. از آنجایی که امید ریاضی شرطی، به‌صورت خطی متغیر پاسخ را تصویر می‌کند، می‌توان تجزیه منحصر به‌فرد

$$Y = E[Y|X = x] + \varepsilon$$

را در نظر گرفت، که در آن x بردار یافته‌های X و ε نوفه یا خطا، یک متغیر تصادفی است. بدیهی است که $E[\varepsilon] = 0$. علاوه بر این، $Y - E[Y|x]$ بر بردار x عمود است. برای جزئیات بیشتر در مورد رگرسیون از دیدگاه امید ریاضی شرطی، **انگریست و پیشک (۲۰۰۸)** را ببینید. رگرسیون خطی یکی از گسترده‌ترین تکنیک‌های مدل‌سازی است. در این حالت، ماهیت خط رگرسیون، به‌صورت خطی می‌باشد و برای مشاهده i ام می‌توان رابطه

$$E[y_i|x_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}$$

را نوشت، که در آن $\boldsymbol{\beta}$ ضرایب خط است. بنابراین در حالت ماتریسی، مدل خطی

$$\mathbf{y} = \beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

نتیجه می‌شود که در آن

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

عدد نامعلوم β_0 عرض از مبدا، $\mathbf{1}_n = (1, \dots, 1)^\top$ یک بردار n بعدی، \mathbf{X}_n ، ماتریس مشاهدات در اندازه $n \times p$ ، $\boldsymbol{\beta}$ ، بردار پارامترها و $\boldsymbol{\varepsilon}$ ، بردار خطاست. متغیرهای تصادفی ε_i و هم‌توزیع با تابع توزیع F بوده که روی خط اعداد حقیقی \mathbb{R} دارای اطلاع فیشر $I(f)$ به صورت

$$I(f) = \int_{-\infty}^{+\infty} \left\{ -\frac{f'(x)}{f(x)} \right\}^2 f(x) dx < \infty, f(x) = \frac{dF(x)}{dx}, f'(x) = \frac{d^2 F(x)}{dx^2}$$

است. در اینجا هیچ فرضی درباره صورت تابعی F وجود ندارد. فرض کنید $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ و $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top] = \sigma^2 \mathbf{I}_n$ که σ مجهول، متناهی و مثبت بوده و \mathbf{I}_n ماتریس همانی از مرتبه $n \times n$ است.

فرض کنید در مدل رگرسیونی، متغیرهای توضیحی استاندارد شده باشند. بدون از دست دادن کلیت مسئله، در این رساله پارامتر عرض از مبدا صفر است و می‌توان مدل رگرسیون خطی را به صورت

$$\mathbf{y}_n = \mathbf{X}_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}_n \quad (1.1)$$

نوشت که در آن اندیس n ، وابستگی به تعداد مشاهده‌ها را نشان می‌دهد.

۲.۱ روش‌های برآورد پارامترها

مدل‌های آماری به منظور دستیابی به بینشی درباره کمیت‌های مجهول (پارامترها) برآورد می‌شوند. ذکر این نکته ضروری به نظر می‌رسد که اگر مجموعه‌ای از داده‌ها در اختیار باشد که با رگرسیون خطی تحلیل شوند، باید در نظر داشت که یک الگوی نظری هرگز نمی‌تواند یک فرآیند واقعی را کاملاً شرح دهد زیرا دنیای واقعی خیلی پیچیده است. از این رو برای اینکه مدل، مناسب باشد، داده‌ها باید نشان دهند که به‌طور تقریبی از مدل پیروی می‌کنند. بنابراین می‌توان هدف اصلی مدل را به این صورت دانست که تابع رگرسیونی، «نزدیک» به مقدار Y باشد. سوالی که در این جا مطرح می‌شود آن است که مفهوم واژه «نزدیکی» و روش‌های مختلف اندازه‌گیری آن چیست؟ در این جا منظور از پارامتر، بردار ضرایب رگرسیونی β است. معمولاً بردار پارامتر رگرسیونی β به گونه‌ای برآورد می‌شود که تابع زیان (اختلاف بین مقدار واقعی y_n و مقدار پیشگویی شده \hat{y}_n) کمینه گردد. به‌طور معمول از تابع زیان توان دوم خطا (SEL) که بر پایه نرم اقلیدسی (نرم L_2) است، استفاده شده و به‌صورت

$$L(\mathbf{y}_n, \hat{\mathbf{y}}_n) = \|\mathbf{y}_n - \hat{\mathbf{y}}_n\|_2^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

تعریف می‌شود، که در آن برای بردار دلخواه $\mathbf{a} = (a_1, \dots, a_p)^T$ ، $\|\mathbf{a}\|_p = \left(\sum_{j=1}^p |a_j|^p\right)^{1/p}$ ، با استفاده از معیار SEL، بردار ضرایب رگرسیونی β به گونه‌ای برآورد می‌شود که بردار پیش‌گویی شده $\hat{\mathbf{y}}_n$ به بردار مشاهده‌های \mathbf{y}_n نزدیک باشد. به عبارت دیگر مجموع توان‌های دوم خطا یعنی $\sum_{i=1}^n \varepsilon_i^2$ ($\varepsilon^T \varepsilon$) کمینه گردد. در این بخش، برآوردگرهای انقباضی و جریمه‌شده را که اساس کار ما در این رساله می‌باشند، معرفی می‌کنیم. قبل از آن به معرفی یک سری مفاهیم مرتبط در این راستا می‌پردازیم.

۱.۲.۱ پذیره‌های مدل

به منظور یافتن برآوردگر OLS لازم است که یک سری پذیره‌های زیربنایی در مدل (۱.۱) برقرار باشند که عبارتند از (۱) $E[\varepsilon_n] = 0$ ، (۲) $E[\varepsilon_n \varepsilon_n^T] = \sigma^2 \mathbf{I}_n$ که در آن $\sigma^2 > 0$. (۳) ماتریس طرح \mathbf{X}_n غیرتصادفی و ثابت است که در نتیجه \mathbf{X}_n از ε_n مستقل است. برآوردگر OLS بردار ضرایب رگرسیونی β به صورت

$$\begin{aligned} \hat{\beta}^{\text{OLS}} &= \operatorname{argmin}_{\beta} \{S(\beta)\} = \operatorname{argmin}_{\beta} \left\{ (\mathbf{y}_n - \mathbf{X}_n \beta)^T (\mathbf{y}_n - \mathbf{X}_n \beta) \right\} \\ &= (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T \mathbf{y}_n \\ &= \mathbf{C}_n^{-1} \mathbf{X}_n^{-1} \mathbf{y}_n \end{aligned} \quad (2.1)$$

است، که در آن $\mathbf{C}_n = \mathbf{X}_n^T \mathbf{X}_n$.

به سادگی نتیجه می‌شود $\hat{\beta}^{Ols} \sim N_p(\beta, \sigma^2 C^{-1})$ که با استفاده از قضیه گاوس-مارکوف، بهترین برآوردگر ناریب خطی (BLUE) است (کاریا و کوراتا، ۲۰۰۴). برآوردگر ناریب واریانس σ^2 بر اساس این برآوردگر به صورت

$$S^2 = \frac{1}{m} (\mathbf{y}_n - \mathbf{X}_n \hat{\beta}^{Ols})^\top (\mathbf{y}_n - \mathbf{X}_n \hat{\beta}^{Ols}), \quad m = n - p.$$

است، که آماره $(ms^2)/\sigma^2$ ، مستقل از $\hat{\beta}^{Ols}$ می‌باشد و توزیع آن، χ_m^2 است. روش OLS، متداول‌ترین روش برای برآورد ضرایب رگرسیونی است که در سال ۱۸۰۹ توسط گاوس ارائه شد. در سال ۱۸۲۳، گاوس در مقاله‌ای نظریه کمترین توان‌های دوم را با بیان کردن قضیه گاوس-مارکوف کامل نمود (مارتین، ۱۹۹۸).

اگر داده‌ها از یک مدل رگرسیونی خطی با پذیره‌های مطرح‌شده پیروی کنند، آن‌گاه روش کمترین توان‌های دوم برآوردهای خوبی را ارائه می‌دهد اما هنگامی که حداقل یکی از این پذیره‌ها برای خطاهای تصادفی برقرار نباشد، مثلاً، مدل رگرسیونی با مشکل هم‌خطی یا خودهمبستگی یا ناهمپراشی مواجه شود، برآوردهای OLS بهینه نیستند. خطاها ممکن است از توزیعی با دم‌های پهن‌تری نسبت به توزیع نرمال تولید شده باشند. در این حالت، احتمال تولید داده‌های پرت در توزیع‌های دم پهن نسبت به توزیع نرمال بیشتر است. داده‌های پرت، داده‌هایی است که متفاوت (به معنی خیلی بزرگ یا خیلی کوچک) از سایر داده‌ها است.

طبق تعریف روسیو و لروی (۱۹۸۷) در مدل‌های رگرسیونی، سه نوع داده پرت وجود دارد: (۱) داده پرت در راستای \mathbf{y}_n ، اغلب باقیمانده‌های (مقادیر پیش‌گویی شده منهای مقادیر مشاهده شده) متناظر با این داده‌ها، مقادیر نسبتاً بزرگی دارند. این نوع از داده‌ها به راحتی بر اساس نمودار باقیمانده‌ها مشخص می‌شود. (۲) داده پرت در فضای \mathbf{X}_n که آن‌ها را نقاط اهرمی می‌نامند. (۳) داده پرت در راستای \mathbf{y}_n و فضای \mathbf{X}_n . (روسیو و لروی، ۱۹۸۷).

پس از برازش مدل رگرسیونی، بررسی اعتبار پذیره‌های رگرسیونی اهمیت بسزایی دارد. از این رو، اجرای روش‌های تشخیصی در مدل ضروری است. این روش‌ها، ترکیبی از روش‌های عددی و نموداری بوده که کلاس مهمی از این روش‌ها بر اساس اصل شناسایی یک مشاهده در یک زمان است. فرض کنید $\hat{\beta}(-i)$ برآورد β را بر اساس نمونه بدون مشاهده i ام نشان دهد. در این صورت اختلاف بین $\hat{\beta}(-i)$ و $\hat{\beta}$ معیاری را ارائه می‌دهد که تاثیر مشاهده i ام را بر ضرایب رگرسیونی بیان می‌کند. یکی از ابزارهای اصلی روش‌های تشخیص کلاسیک، مولفه‌های قطری ماتریس تصویر برآوردگر OLS به صورت

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$$

است، بر اساس روش کمترین توان‌های دوم، بردار مقادیر برازش شده $\hat{\mathbf{y}}_n = \mathbf{H}\mathbf{y}_n$ بوده و بردار مانده‌ها، $e_n = \hat{e}_n = (\mathbf{I}_n - \mathbf{H})\mathbf{y}_n$ است. همچنین $\text{Var}(\hat{\mathbf{y}}_n) = \sigma^2 \mathbf{H}$ و $\text{Var}(\hat{e}_n) = \sigma^2 (\mathbf{I}_n - \mathbf{H})$. همان‌طور که ملاحظه می‌شود ماتریس \mathbf{H} نقش کلیدی و بسیار مهمی در نتایج روش کمترین توان‌های دوم بازی می‌کند.

ماتریس H که اولین بار توسط **هواگلین و ولش** (۱۹۷۸) معرفی شد، معمولاً به ماتریس هت شناخته می‌شود زیرا این ماتریس علامت \hat{y} را روی بردار y می‌گذارد، اما نام‌های دیگری نیز برای این ماتریس به کار می‌رود که می‌توان به ماتریس تصویر (تصویر بردار y_n را روی فضای تولید شده متغیرهای X_n ایجاد می‌کند) یا ماتریس پیشگویی (ماتریس Hy_n بردار مقادیر پیشگویی شده را ارائه می‌دهد) اشاره کرد. ماتریس H به عنوان ماتریس اهرمی نیز شناخته می‌شود زیرا i امین مقدار برازش شده، \hat{y}_i ، می‌تواند به صورت $\hat{y}_i = \sum_{j=1}^n h_{ij}y_j$ نوشته شود. بنابراین، h_{ij} اهرم (وزن) j امین مشاهده y_j در تعیین i امین مقدار برازش شده \hat{y}_i است. به‌ویژه، h_{ii} اهرم i امین مشاهده y_i در تعیین مقدار \hat{y}_i است. بنابراین، مشاهده y_j با مقادیر بزرگ h_{ij} می‌تواند تاثیر نامناسب را روی نتایج کمترین توان‌های دوم نشان دهد.

تاکنون ماتریس H از زوایای مختلف بررسی شده است. بیشترین موضوع مورد مطالعه، یافتن معیاری برای شناسایی نقاط اهرمی در رگرسیون خطی است. بیشتر این معیارها بر پایه مولفه‌های قطری h_{ii} ماتریس H می‌باشد. از آنجایی که ماتریس H ، ماتریس تصویر است، $\sum_{j \neq i} h_{ij}^2 = h_{ii}(1 - h_{ii})$ به ازای $i = 1, \dots, n$ برقرار بوده و در نتیجه $0 \leq \sum_{i=j=1}^n h_{ij}^2 = h_{ii} \leq 1$. اگر $h_{ii} = 1$ ، آن‌گاه $h_{ij} = \delta_{ij}$ که در آن δ_{ij} تابع دلتای کرونکر است، بنابراین $\hat{y}_i = y_i$. لذا ابرصفحه رگرسیون صرف‌نظر از مشاهده‌های دیگر، از نقطه (x_i^T, y_i) می‌گذرد. چنین نقطه‌ای می‌تواند مشاهده تاثیرگذار (نقطه اهرمی) باشد، اما حتی در حالتی که $h_{ii} \approx 1$ ، مقادیر h_{ij} ، $i \neq j$ نزدیک ۱ بوده و ابرصفحه رگرسیون از نزدیکی نقطه (x_i^T, y_i) می‌گذرد. اگر مولفه قطری h_{ii} مساوی صفر باشد، تساوی بالا نتیجه می‌دهد که $h_i^T = 0$ ، به عبارت دیگر، سطر متناظر ماتریس H برابر ۰ است و نشان می‌دهد نقطه y_i تاثیری در برازش ندارد.

تاثیر مولفه‌های قطری ماتریس H روی رفتار برآوردگر OLS، به خوبی معیار رفتار برآوردگر دلخواه $\hat{\beta}^*$ را در دم‌های توزیع خطای F در مدل رگرسیونی (۱.۱) برای پارامتر β نشان می‌دهد. این معیار به صورت زیر تعریف می‌شود:

$$B(a, \hat{\beta}^*) = \frac{-\ln P_{\beta}(\max_i |x_i^T(\hat{\beta}^* - \beta)| > a)}{-\ln(1 - F(a))}$$

مقادیر بزرگ $B(a, \hat{\beta}^*)$ برای $a \gg 0$ از اهمیت بسزایی برخوردارند زیرا مقدار $B(a, \hat{\beta}^*)$ نشان می‌دهد هنگامی که $a \rightarrow \infty$ ، عبارت $P_{\beta}(\max_i |x_i^T(\hat{\beta}^* - \beta)| > a)$ چه قدر سریع‌تر در مقایسه با دم‌های توزیع پایه‌ای F به صفر همگرا می‌شود.

فرض کنید $h^* = \max_{1 \leq i \leq n} h_{ii}$ بیشینه مولفه قطری ماتریس تصویر باشد. در این صورت روابط زیر برای برآوردگر OLS برقرار است:

$$(۱) \text{ اگر } F \text{ توزیع نرمال } (\circ, \sigma^2), \sigma > 0 \text{ باشد، آن‌گاه } \frac{1}{h^*} = \lim_{a \rightarrow \infty} B(a, \hat{\beta}^{\text{Ols}})$$

$$(۲) \text{ اگر } F \text{ توزیع لجستیک یا لاپلاس باشد، آن‌گاه}$$

$$\frac{1}{h^*} \leq \liminf_{a \rightarrow \infty} B(a, \hat{\beta}^{\text{Ols}}) \leq \limsup_{a \rightarrow \infty} B(a, \hat{\beta}^{\text{Ols}}) \leq \frac{1}{h^*}$$

(۳) اگر F توزیعی دم‌پهن باشد، وقتی $a \rightarrow \infty$ ، $\frac{-\ln(1-F(a))}{m \ln a} \rightarrow 1$ و آن‌گاه

$$\lim_{a \rightarrow \infty} B(a, \hat{\beta}^{\text{Ols}}) = 1.$$

این روابط نشان می‌دهند اگر h^* به‌طور منطقی کوچک نباشد، آن‌گاه حتی زمانی که توزیع نرمال باشد، رفتار برآوردگر OLS مطلوب نیست.

از آن جایی که $\text{tr}(\mathbf{H}) = \sum_{i=1}^n h_{ii} = p$ ، نتیجه می‌شود $h^* \geq p/n$. در حالت کلی، میانگین h_{ii} ها p/n است و مشاهدات y_i با $h_{ii} > 2p/n$ می‌تواند به‌عنوان نقطه اهرمی در نظر گرفته شود. برای تعیین تاثیر i امین مشاهده، از معیارهای DFFITS و فاصله کوک استفاده می‌شود. فاصله کوک تغییرات در ضرایب رگرسیون بعد از حذف i امین مشاهده را اندازه‌گیری می‌کند و به‌صورت

$$D_i = \frac{(\hat{\beta}^{\text{Ols}} - \hat{\beta}^{\text{Ols}}(i))^{\top} \mathbf{X}^{\top} \mathbf{X} (\hat{\beta}^{\text{Ols}} - \hat{\beta}^{\text{Ols}}(i))}{ps^2}$$

تعریف می‌شود که در آن $\hat{\beta}^{\text{Ols}}(i)$ برآوردگر OLS پارامتر β بدون i امین مشاهده است. مقدار بزرگتر D_i نشان می‌دهد که i امین مشاهده، تاثیر قابل ملاحظه‌ای بر تعیین $\hat{\beta}^{\text{Ols}}$ دارد. نظرهای متفاوتی برای تعیین مقدار آستانه‌ای برای تعیین داده‌های پرت وجود دارد. معمولاً از روش **بولن و جکمن (۱۹۹۰)** استفاده می‌شود و بدین صورت است که اگر $D_i > 4/n$ ، آن‌گاه i امین مشاهده به‌عنوان نقطه پرت در نظر گرفته می‌شود.

آماره DFFITS به‌صورت

$$\text{DFFITS}_i = \frac{(y_i - \hat{y}_i) (h_{ii})^{\frac{1}{2}}}{s(i) \sqrt{1 - h_{ii}}}$$

تعریف می‌شود که در آن $s(i)$ برآورد σ محاسبه‌شده بدون i امین مشاهده است. مشاهده‌هایی که مقدار DFFITS بیشتر از $2(p/n)^{1/2}$ دارند، باید حذف شوند. این معیار، تاثیر پیش‌گویی حذف مشاهده i ام را نشان می‌دهد.

درباره هم‌خطی می‌توان اشاره نمود که دو نوع هم‌خطی در بین متغیرهای توضیحی وجود دارد:

(۱) دقیق: با نماد ریاضی، هم‌خطی دقیق به‌صورت $\sum_{j=1}^p a_j X_j = 0$ است.

(۲) شدید اما نادقیق: وقتی همبستگی بین متغیرهای توضیحی، شدید و نادقیق باشد، آن‌گاه هم‌خطی شدید را می‌توان به‌صورت زیر نشان داد

$$\sum_{j=1}^p a_j X_j \approx 0,$$

که در آن $\mathbf{a} = (a_1, \dots, a_p)^{\top}$ یک بردار ثابت دلخواه است ($a_j \neq 0$).

برای اینکه هم‌خطی تشخیص داده شود، چندین روش وجود دارد:

(۱) ماتریس همبستگی: مقادیر بزرگ همبستگی پیرسون برای ستون‌های ماتریس X_n می‌تواند نشانه‌ای از هم‌خطی باشد. گاهی تمامی مقادیر همبستگی کوچک هستند، اما هنوز هم‌خطی وجود دارد. مقادیر و بردارهای ویژه ماتریس همبستگی می‌تواند برای شناسایی هم‌خطی به‌عنوان یک روش ساده به‌کار گرفته می‌شود. وقتی مقادیر و بردارهای ویژه نزدیک به صفر باشند، آن‌گاه هم‌خطی می‌تواند وجود داشته باشد.

(۲) عامل تورم واریانس (VIF): از معیارهای معروف و کاربردی برای شناسایی هم‌خطی است. اگر هر یک از مقادیر VIF بیشتر از 10^0 باشد، هم‌خطی در داده‌ها وجود دارد و اگر هر یک از مقدار آن‌ها بیشتر از 30^0 باشد، هم‌خطی از نوع شدید است. (گجراتی، ۲۰۰۴). VIF می‌تواند به‌صورت

$$VIF_j = (1 - R_j^2)^{-1}, \quad j = 1, 2, \dots, p.$$

فرمول‌بندی شود، که در آن R_j^2 ضریب تعیین در رگرسیون زامین متغیر مستقل بر روی متغیرهای دیگر است.

(۳) عدد شرطی: عدد شرطی معمولاً با κ نشان داده می‌شود و به‌صورت نسبت بزرگترین به کوچکترین مقدار ویژه از تجزیه مقدار ویژه (SVD) X_n است. به‌طور کلی، اگر $\kappa < 10^0$ ، هم‌خطی وجود ندارد. اگر $10^0 < \kappa < 10^00$ ، هم‌خطی متوسطی وجود دارد و اگر بیشتر از 10^00 باشد، بین متغیرهای مستقل، هم‌خطی شدید وجود دارد.

۲.۲.۱ برآوردگرهای انقباضی

استفاده از قوانین آماری استاندارد که معمولاً فرض می‌کند هیچ اطلاعی درباره نتیجه وجود ندارد ممکن است چندان مفید نباشد. از این رو، برخی از آماردان‌ها علاوه بر اطلاعات نمونه‌ای، از سایر اطلاعات نیز در برآورد پارامترها استفاده می‌کنند. منظور از سایر اطلاعات، اطلاعات غیرنمونه‌ای (NSI) یا اطلاعات پیشین غیرقطعی (UPI) درباره پارامترهای موردعلاقه است. در برآورد، شناسایی NSI اهمیت بسزایی دارد به خصوص زمانی که اطلاعات نمونه‌ای نسبتاً محدود باشند. استفاده از UPI در خصوص بعضی از پارامترها و گاهی در مورد تمام پارامترهای مورد بررسی، اغلب منجر به بهبود یعنی دستیابی به مخاطره توان دوم کمتر در فرآیند برآورد می‌شود (صالح و سن، ۱۹۷۸، ۱۹۸۵).

استفاده از اطلاعات قابل اطمینان غیرنمونه‌ای نتایج مفیدی دارد اما در برخی از مطالعات، این اطمینان وجود ندارد که آیا این اطلاعات صحیح هستند یا خیر. به‌عنوان مثال، فرض کنید با اندازه‌گیری تومور خاصی در نوعی از موش‌ها در زمان‌های مختلف پس از تزریق مواد سرطان‌زا، داده‌هایی به دست آمده است. زیست‌شناسان علاقه‌مند هستند پارامتر نرخ رشد تومور (λ) را در حالی برآورد کنند که فرض شده است $\lambda = \lambda_0$. چنین λ_0 می‌تواند در شرایط آزمایشگاهی مختلف (رفتار سلول‌ها می‌تواند یا نمی‌تواند متفاوت باشد و ...) به دست آید.

از آن جا که شرایط آزمایشگاهی متفاوت است، احتمال دارد زیست‌شناسان فرض کنند که λ_0 مقدار درستی برای پارامتر رشد در مطالعات تجربی است که ممکن است در این باره مطمئن نباشند.

از آن جایی که برآوردگر OLS، بی‌ارتباط با اطلاعات پیشین غیرقطعی است، در حالت کلی آن را برآوردگر محدود نشده نامیده و با نماد $\hat{\beta}^U$ نشان می‌دهند.

استفاده از اطلاعات غیرنمونه‌ای در مدل‌های آماری، معمولاً به صورت یک محدودیت مطرح می‌شود که منجر به مدل‌های محدود شده گشته و برآوردگری که بر اساس این مدل‌ها به دست می‌آید، برآوردگر محدود شده (RE) نامیده می‌شود. اعتبار و کارایی این برآوردگر بر اساس قید اعمال شده بر روی فضای پارامتر مورد سنجش و بررسی قرار می‌گیرد. بنابراین نتایج به دست آمده بر اساس مدل‌های محدود شده و محدود نشده تنها با سنجیدن درستی و اعتبار محدودیت‌های مطرح شده، می‌تواند معقول و کاربردی باشد. باید توجه نمود که انتخاب برآوردگرهای محدود شده در شرایطی که NSI درست نباشد، توجیه‌پذیر نیست.

محدودیت می‌تواند به یکی از دلایل زیر به وجود آید: (۱) حقیقت ناشی از ملاحظات نظری یا تجربی (آزمایشگاهی)، (۲) فرضیه‌ای که باید آزمون شود، (۳) نظر محقق یا متخصص، (۴) شرط تحمیل شده غیرمعمول برای کاهش یا حذف فزونگی در مدل مورد بررسی. جزئیات بیشتر را می‌توان در **دان (۱۹۸۲)**، **سن‌گوپتا و جامالامادا (۲۰۰۳)** و **آرشی و ناداراجا (۲۰۱۷)** مشاهده نمود.

برآوردگر انقباضی، برآوردگری است که از ترکیب برآوردگرهای متداول (MLE)، برآوردگر OLS، برآوردگر بیز و ... با NSI یا UPI به منظور دستیابی به برآوردگرهای بهبودیافته به دست می‌آیند. بهبود می‌تواند به صورت کاهش یا افزایش برخی توابع هدف مانند میانگین توان دوم خطا (MSE)، توان دوم مخاطره، اریبی، تابع درست‌نمایی و ... باشد.

برای درک بهتر از این نوع برآوردگرها، برآورد پارامتر δ را در نظر بگیرید. اغلب در کاربرد، محقق با توجه به اطلاعاتی که از رفتار سیستم دارد یا با توجه به تجربه‌های گذشته می‌تواند حدسی آگاهانه یا برآوردی اولیه به صورت δ_0 را در خصوص مقدار δ ، در دسترس قرار دهد. از آن جایی که این نظر مهم است و نباید تجربه‌ها و دانش محقق را نادیده گرفت، می‌توان یک برآوردگر اولیه برای پارامتر، مانند $\hat{\delta}$ در نظر گرفت و سپس آن را به مقدار برآورد اولیه δ_0 نزدیک (منقبض) نمود.

برای ساختن برآوردگر انقباضی، مقدار ثابتی تحت عنوان ثابت انقباضی در تفاضل $\hat{\delta} - \delta_0$ ضرب کرده و به δ_0 اضافه می‌شود. به عبارت دیگر، برآوردگر جدید انقباضی به صورت زیر به دست می‌آید:

$$\hat{\delta}^{\text{Shrinkage}} = \hat{\theta}^S = \delta_0 + c(\hat{\delta} - \delta_0) = c\hat{\delta} + (1 - c)\delta_0; \quad 0 \leq c \leq 1.$$

برای روشن شدن این موضوع، فرض کنید می‌خواهید مسئله برآورد پارامتر بردار میانگین توزیع نرمال p متغیره را فرمول بندی کنید. برای انجام آن، بردار تصادفی $X = (X_1, \dots, X_p)^T$

را در نظر بگیرید که توزیع آن، نرمال p متغیره است که میانگین آن $\theta = (\theta_1, \dots, \theta_p)^T$ و ماتریس واریانس - کوواریانس آن I_p باشد. پارامتر مورد علاقه، θ است که باید برآورد شود. **استاین (۱۹۵۶)** نشان داد در توزیع نرمال، هنگامی که بعد فضای پارامتر حداقل ۳ باشد، MLE، برآوردگر غیرمجاز برای پارامتر θ است که انگیزه‌ای را برای بهبود برآوردگرها در مسائل برآوردیابی در بین محققین ایجاد نمود. دقت کنید غیر مجاز بودن برای حالت‌های $p = 1, 2$ برقرار نیست. لازم به ذکر است که در حالت‌های $p = 1, 2$ ، MLE تحت تابع زیان درجه دوم برای θ مجاز است.

جیمز و استاین (۱۹۶۱) برآوردگری معرفی کردند که به ازای $p \geq 3$ بر MLE برتری دارد که تحول بزرگی در شاخه‌های مختلف آمار ایجاد نمود. به افتخار پروفسور چارلز استاین، برآوردگری را که به‌طور یکنواخت بر برآوردگرهای استاندارد، مانند برآوردگر OLS و MLE برتری دارد، برآوردگر نوع استاین می‌نامند (**کسلا، ۱۹۸۵**).

فرض کنید $\hat{\theta}^U = (\hat{\theta}_1^U, \dots, \hat{\theta}_p^U)^T$ برآوردگر ML پارامتر θ باشد. بنابراین، صورت ساده‌ای از برآوردگر جیمز-استاین پارامتر θ به صورت

$$\hat{\theta}^S = \left\{ 1 - \frac{c}{(\hat{\theta}^U)^T (\hat{\theta}^U)} \right\} \hat{\theta}^U \quad (3.1)$$

است، که در آن c ، ثابت انقباضی است و معمولاً $0 \leq c \leq 2(p-2)$ است. لازم به یادآوری است که $\hat{\theta}^S$ ، برآوردگر $\hat{\theta}^U$ را به سمت صفر منقبض می‌کند و چنین برآوردگری در حالت کلی یک برآوردگر انقباضی (SE) نامیده می‌شود. با این وجود، دلیلی وجود ندارد که $\hat{\theta}^U$ باید حتماً به سمت صفر منقبض شود. در حالت کلی، ممکن است برآوردگر $\hat{\theta}^U$ به سمت بردار ثابت و دلخواه θ_0 منقبض شود. در این حالت، برآوردگر جیمز-استاین به صورت

$$\hat{\theta}^S = \left\{ 1 - \frac{c}{(\hat{\theta}^U - \theta_0)^T (\hat{\theta}^U - \theta_0)} \right\} (\hat{\theta}^U - \theta_0) + \theta_0, \quad p \geq 3. \quad (4.1)$$

به دست می‌آید.

اگرچه برآوردگر ML به سمت یک بردار ثابت و مشخص منقبض شده است، می‌تواند به سمت یک برآوردگر نیز انقباض پیدا کند. به عنوان مثال، **رائو (۱۹۴۸)** یک ماتریس 4×28 از داده‌هایی را در نظر گرفت که وزن‌های انبوه ذخایر ۲۸ درخت را در چهار منطقه شمال، جنوب، شرق و غرب نشان می‌دهد. اگر وزن نمونه تصادفی ذخایر درختان با X_{north} ، X_{south} ، X_{east} و X_{west} نشان داده شود، آن‌گاه نمونه شامل ۲۸ مشاهده از بردار $x = (x_{north}, x_{south}, x_{east}, x_{west})^T$ خواهد بود. فرض کنید که توزیع X ، نرمال با میانگین $\theta = (\theta_{north}, \theta_{south}, \theta_{east}, \theta_{west})^T$ باشد. هدف، برآورد بردار میانگین ذخایر است. یکی از فرضیه‌های ممکن این است که میانگین ذخایر برای ناحیه‌های مختلف، یکسان باشد. به عبارت دیگر، $H_0: \theta_{north} = \theta_{south} = \theta_{east} = \theta_{west} = \theta_0$.

در حالت کلی، ممکن است $\theta_1 = \dots = \theta_p = \theta_0$ برقرار باشد که θ_0 مجهول است. در این مورد، ابتدا برآوردگر محدودشده به صورت $\hat{\theta}^R = (\hat{\theta}^R, \dots, \hat{\theta}^R)^T = \hat{\theta}^R \mathbf{1}_p$ تعریف می‌شود که در آن $\mathbf{1}_p = (1, \dots, 1)^T$ و $\hat{\theta}^R = \frac{1}{p} \sum_{i=1}^p \bar{x}_i$ میانگین ستونی ماتریس X_n است. در مطالعات گذشته نشان داده شده است وقتی اختلاف بین مقدار حدس زده شده و واقعی پارامتر کوچک باشد، برآوردگر محدود شده بهتر از برآوردگر محدود نشده است و هرچه قدر مطلق این اختلاف بزرگتر شود، $\hat{\theta}^R$ برآوردگر مطلوبی نخواهد بود زیرا به طور قابل ملاحظه‌ای اریب، ناکارا و ناسازگار است در حالی که اختلاف، اثری بر روی کارایی برآوردگر $\hat{\theta}^U$ ندارد. می‌توان برآوردگر معمولی $\hat{\theta}^U$ پارامتر θ را به سمت $\hat{\theta}^R$ منقبض نمود. در این حالت برآوردگر انقباضی نوع جیمز-استاین به صورت زیر است:

$$\hat{\theta}^S = \hat{\theta}^R + \left\{ 1 - \frac{c}{(\hat{\theta}^U - \hat{\theta}^R)^T (\hat{\theta}^U - \hat{\theta}^R)} \right\} (\hat{\theta}^U - \hat{\theta}^R)$$

افرون و موریس (۱۹۷۵) و رابینز (۱۹۸۳) بیان کردند که این برآوردگر، برآوردگر بیز تجربی است و مبنای آماری دارد. **براندوین و استرادرن (۱۹۹۰) و استیگلر (۱۹۹۰)** تفسیری برای برآوردگر جیمز-استاین ارائه کردند. برای مطالعه بیشتر در مورد برآوردگرهای نوع استاین، به **جاج و باک (۱۹۷۸)؛ اندرسون (۱۹۸۴)؛ کوپاس (۱۹۸۳)؛ برگر (۱۹۸۵)؛ احمد و صالح (۱۹۹۰)، (۱۹۹۳) و روخین (۱۹۹۵)** مراجعه کنید. همچنین نظریه جانبی برآوردگرها توسط **سن (۱۹۸۶)؛ علی و صالح (۱۹۹۱)؛ احمد (۱۹۹۲، ۲۰۰۲)** مورد بررسی قرار گرفته است. نظریه جانبی ناپارامتری برآوردگرها توسط **صالح و سن (۱۹۷۸، ۱۹۸۵) و احمد و صالح (۱۹۹۹)** بررسی شده است.

برآوردگرهای جیمز-استاین روی کل فضای پارامتر مخاطره کمتری از برآوردگرهای ML دارند. با این وجود، آن‌ها غیرمجاز هستند. زیرا برآوردگر دیگری وجود دارد که بر برآوردگرهای جیمز-استاین تسلط دارد. برآوردگر انقباضی در (۴.۱) با جزء مثبت آن بهبود می‌یابد که به صورت زیر تعریف می‌شود:

$$\hat{\theta}^{S+} = \left\{ 1 - \frac{c}{(\hat{\theta}^U)^T (\hat{\theta}^U)} \right\}^+ \hat{\theta}^U \quad p \geq 3 \quad (5.1)$$

که در آن $a^+ = \max\{0, a\}$. این برآوردگر از تغییر علامت $\hat{\theta}^U$ جلوگیری می‌کند. مقدار عامل انقباضی، $c / ((\hat{\theta}^U)^T (\hat{\theta}^U))$ ، $1 - c / ((\hat{\theta}^U)^T (\hat{\theta}^U))$ به مقدار c وابسته است. اگر $c / ((\hat{\theta}^U)^T (\hat{\theta}^U)) \geq 1$ ، عامل انقباضی منفی می‌شود. بنابراین، علامت‌های تعدادی از ضرایب ممکن است برعکس شود. پدیده بیش انقباضی کارایی مخاطره برآوردگر جیمز-استاین را تحت تاثیر قرار نمی‌دهد اما باعث می‌شود تفسیرپذیری مدل پیچیده‌تر شود. از نگاه کاربردی، تغییر در علامت ممکن است باعث سردرگمی تحلیل‌گر شود، به ویژه هنگامی که این برآوردگر با برآوردگر محدود نشده مقایسه شود. برآوردگر جدیدی که هرگز علامت متغیرها را تغییر نمی‌دهد، برآوردگر جزء مثبت انقباضی نوع استاین (PRSE) خواهد بود.

در ادامه برآوردگر انقباضی محدودشده (SRE) به صورت زیر معرفی می‌شود

$$\hat{\theta}^{SR} = \pi \hat{\theta}^U + (1 - \pi) \hat{\theta}^R$$

که در آن π ضریبی است که درجه بی‌اعتمادی به اطلاع پیشین را منعکس می‌کند. مقدار آن ممکن است کاملاً توسط کاربر، بر اساس درجه عدم اعتقاد او به NSI تعیین شود. مقدار نزدیک به یک، باعث می‌شود که $\hat{\theta}^{SR}$ فقط به داده‌های نمونه به تنهایی بستگی داشته باشد. به ازای $\pi = I(\mathcal{L}_n > \mathcal{L}_{n,\alpha})$ ، که در آن $I(A)$ تابع نشان‌گر مجموعه A ، \mathcal{L}_n آماره آزمون مناسب برای فرضیه صفر $H_0: \theta_1 = \dots = \theta_k = \theta_0$ و $\mathcal{L}_{n,\alpha}$ چندک بالایی $100\alpha\%$ توزیع آماره آزمون \mathcal{L}_n در حالت فرضیه صفر است، برآوردگر آزمون اولیه (PTE) برای بردار پارامتر θ نتیجه می‌شود. آن را با نماد $\hat{\theta}^{PT}$ نشان می‌دهند و به صورت زیر است:

$$\hat{\theta}^{PT} = \hat{\theta}^R I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha}) + \hat{\theta}^U I(\mathcal{L}_n > \mathcal{L}_{n,\alpha}). \quad (6.1)$$

برآوردگرهای آزمون اولیه توسط محقق‌های بسیاری به کار برده شده‌اند. شاهد این ادعا، ارجاعات متعدد به مقاله **بنکرافت و هان (۱۹۷۷)** است. با این وجود، برآوردگرهای آزمون اولیه بر برآوردگرهایی که فقط بر اساس UPI به دست آمده‌اند در ناحیه نسبتاً کوچکی از فضای پارامتر که با NSI ایجاد شده است، برتری دارد. کارایی PTE به صحت UPI بستگی دارد. اگر این اطلاعات درست یا تقریباً درست باشد، آن‌گاه PTE بهتر از برآوردگری است که فقط بر اساس اطلاعات نمونه‌ای به دست آمده و در غیر این صورت، عکس این نتیجه برقرار است.

یادآوری می‌شود که برآوردگرهای انقباضی در حالت کلی بر MLE به ازای $p \geq 3$ برتری دارد در حالی که به ازای $p = 1, 2$ ، برآوردهای ML، برآوردهای مجاز هستند. بنابراین، استفاده از برآورد انقباضی ممکن است در محدودیت بعد گرفتار شود (در حالت‌های یک و دوبعدی کاربرد نیستند). در این حالت‌ها رهیافت آزمون اولیه به کار برده می‌شود.

برای مطالعه بیشتر در نظریه ناپارامتری برآوردگرهای آزمون اولیه در فضای متناهی، به **احمد و صالح (۱۹۸۹، ۱۹۹۰، ۱۹۹۳)** و **جاج و باک (۱۹۷۸)** مراجعه کنید. نظریه جانبی این برآوردگرها توسط **سن (۱۹۸۶)**؛ **احمد (۱۹۹۲)**؛ **کالپرگر و احمد (۱۹۹۲)** مطالعه شده است. درباره برآوردگرهای آزمون اولیه در مبحث ناپارامتری می‌توان مقاله‌های **صالح و سن (۱۹۷۸)**، **(۱۹۸۵)**؛ **احمد و صالح (۱۹۹۹)** را مشاهده نمود.

استفاده از برآوردگر آزمون اولیه به اندازه سطح معنی‌داری آزمون بستگی دارد. بنابراین، تکنیک انقباضی برای برآوردگر آزمون اولیه معرفی شد تا بر این مشکل غلبه کند. برآوردگر انقباضی آزمون اولیه (SPTE) پارامتر θ که با نماد $\hat{\theta}^{SPT}$ نشان داده می‌شود، از جایگذاری $\hat{\theta}^R$ با $\hat{\theta}^{SR}$ در برآوردگر آزمون اولیه به صورت

$$\hat{\theta}^{SPT} = \hat{\theta}^U - (1 - \pi) (\hat{\theta}^U - \hat{\theta}^R) I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha}) \quad (7.1)$$

به دست می‌آید که به ازای $\pi = 0$ ، $\hat{\theta}^{SPT} = \hat{\theta}^{PT}$ و برای $\pi = 1$ ، $\hat{\theta}^{SPT} = \hat{\theta}^U$. ویژگی‌های این برآوردگر در مسائل مختلف در مقاله‌های **احمد (۱۹۹۴، ۱۹۹۵)** بررسی شده است.

با این وجود، برآوردگر آزمون اولیه، ممکن است از جانشینی $\hat{\beta}^U$ با $\hat{\beta}^S$ بهبود یابد و برآوردگر آزمون اولیه بهبودیافته (IPTE) نامیده شود که اولین بار توسط **سالو و همکاران (۱۹۷۲)** معرفی شده است. این برآوردگر که از رابطه

$$\hat{\theta}^{IPT} = \hat{\theta}^R I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha}) + \hat{\theta}^S I(\mathcal{L}_n > \mathcal{L}_{n,\alpha}).$$

به دست می‌آید، بر PTE برتری دارد اما در حالت کلی، محدودیت $p \geq 3$ وجود دارد در حالی که هیچ محدودیتی برای استفاده از PTE نیست. با این وجود، برآوردگر آزمون اولیه بهبودیافته، برآوردگر $\hat{\theta}^U$ را به طور یکنواخت بهبود نمی‌بخشد و بنابراین همان انتقادی که در مورد برآوردگر آزمون اولیه مطرح شد، در این جا نیز دوباره بحث می‌شود. **احمد و اولاه (۱۹۹۹)** و دیگران ویژگی‌های برآوردگر آزمون اولیه بهبودیافته را بررسی کردند. رفتار مجانبی این برآوردگر در مسائل ناپارامتری توسط **احمد و صالح (۱۹۹۹)** مطالعه شده است.

اکنون برآورد ضرایب رگرسیونی را در مدل (۱.۱) وقتی که ضرایب ممکن است به یک زیرفضا محدود شوند، در نظر بگیرید. در اینجا، پارامتر β در حالتی برآورد می‌شود که این بردار در زیرفضایی که به صورت

$$H\beta = h \quad (۸.۱)$$

تعریف می‌شود، قرار بگیرد که در آن H ماتریس معلوم $q \times p$ با رتبه کامل سطری q و h بردار $q \times 1$ از ثابت‌هاست.

اطلاع پیشین مطرح شده در رابطه (۸.۱)، ممکن است صریحا در رویه برآورد با اصلاح کردن فضای پارامتر به کار گرفته شود. در این حالت فضای پارامتر جدید (محدود شده) یک زیرفضا از فضای اصلی و اولیه است (بعد آن کاهش یافته است). در حالت کلی، کاهش بعد منجر به برآوردهای کاراتری از پارامتر می‌شود. با این وجود، وقتی که محدودیت درست نباشد، نتیجه عکس برقرار است.

با در نظر گرفتن محدودیت (۸.۱)، مشابه **صالح (۲۰۰۶)** برآوردگر کمترین توان‌های دوم محدودشده (RLSE) یا برای سادگی RE پارامتر β برابر است با

$$\hat{\beta}^R = \hat{\beta}^U - C_n^{-1} H^T (HC_n^{-1} H^T)^{-1} (H\hat{\beta}^U - h). \quad (۹.۱)$$

اگر فرض شود که محدودیت (۸.۱) صحیح است، آن گاه $\hat{\beta}^R$ برآوردگر ناریب β است و بر $\hat{\beta}^U$ برتری دارد. البته این حالت ممکن است همیشه برقرار نباشد (**راویشانکر و دی، ۲۰۰۲**). منطقی است که برآوردگر $\hat{\beta}^U$ به سمت برآوردگر $\hat{\beta}^R$ منقبض شود (**تامپسون، ۱۹۶۸**). بنابراین برآوردگر محدودشده انقباضی پارامتر β برابر است با

$$\hat{\beta}^{SR} = \hat{\beta}^U - (1 - \pi) (\hat{\beta}^U - \hat{\beta}^R). \quad (۱۰.۱)$$

برآوردگر $\hat{\beta}^{SR}$ ترکیب محدب از $\hat{\beta}^U$ و $\hat{\beta}^R$ با مقدار ثابت π است. همواره،

$$\hat{\beta}^{SR} = \begin{cases} \hat{\beta}^R & \pi = 0 \\ \hat{\beta}^U & \pi = 1 \end{cases}$$

در واقعیت، اطلاع پیشین $H\beta = h$ غیرقطعی است. تردید درباره این اطلاع را می توان با رهیافت فیشر برطرف نمود. بدین صورت که در ابتدا فرضیه $H_0: H\beta = h$ در مقابل فرضیه ثابت $H_A: H\beta \neq h$ آزمون شود. سپس بر اساس نتیجه این آزمون، اگر فرضیه H_0 پذیرفته شود، $\hat{\beta}^R$ و اگر فرضیه H_0 رد شود، $\hat{\beta}^U$ انتخاب گردد. به عبارت دیگر، برآوردگری می توان ساخت که وابسته به نتیجه آزمون باشد.

برآوردگر آزمون اولیه را با نماد $\hat{\beta}^{PT}$ نشان می دهند و به صورت زیر تعریف می شود:

$$\begin{aligned} \hat{\beta}^{PT} &= \begin{cases} \hat{\beta}^R; & \mathcal{L}_n \leq \mathcal{L}_{n,\alpha} \\ \hat{\beta}^U; & \mathcal{L}_n > \mathcal{L}_{n,\alpha} \end{cases} \\ &= \hat{\beta}^R I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha}) + \hat{\beta}^U I(\mathcal{L}_n > \mathcal{L}_{n,\alpha}) \\ &= \hat{\beta}^U + (\hat{\beta}^R - \hat{\beta}^U) I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha}) \end{aligned} \quad (11.1)$$

برای یافتن $\mathcal{L}_{n,\alpha}$ ، قضیه های زیر مطرح می شوند.

قضیه ۱.۲.۱. (تیل، ۱۹۷۱) فرض کنید مولفه های ε_n به طور مستقل و هم توزیع با میانگین صفر و واریانس متناهی σ^2 توزیع شده اند. اگر عناصر X_n به طور یکنواخت کراندار باشند و $\lim_{n \rightarrow \infty} n^{-1} X_n^{-1} X_n = C$ که C ماتریس متناهی و وارون پذیر است. آن گاه

$$\sqrt{n} (\hat{\beta}_n^U - \beta) \xrightarrow{D} \mathcal{N}_p(0, \sigma^2 C^{-1}).$$

قضیه ۲.۲.۱. فرض کنید که مولفه های ε_n مستقل و هم توزیع با توزیعی با میانگین صفر و واریانس متناهی σ^2 باشند. اگر $\lim_{n \rightarrow \infty} n^{-1} X_n^T X_n = C$ که C ماتریس وارون پذیر و متناهی است، آن گاه تحت فرضیه $H_0: H\beta_n = h$ ، آماره آزمون نسبت درستنمایی به صورت

$$\mathcal{L}_n = (H\hat{\beta}_n^U - h)^T [H(X_n^T X_n)^{-1} H^T]^{-1} (H\hat{\beta}_n^U - h) / s^2$$

است. (الف) تحت فرض نرمال بودن خطا، آماره \mathcal{L}_n دارای توزیع فیشر $F_{q,m}(\Delta^2)$ است که در آن،

$$\Delta^2 = \frac{(H\beta - h)^T (H(X^T X)^{-1} H^T)^{-1} (H\beta - h)}{\sigma^2}.$$

(ب) به طور مجانبی وقتی $n \rightarrow \infty$ ، آماره \mathcal{L}_n دارای توزیع $\chi_q^2(\Delta^2)$ است.

بنابراین تحت فرض های قضیه ۲.۲.۱، $\mathcal{L}_{n,\alpha}$ چندک بالایی $100\alpha\%$ توزیع درحالت (الف)

و در حالت (ب) χ_q^2 است.

با دقت در رابطه (۱۱.۱)، مشخص می‌شود که برآوردگر آزمون اولیه از جایگذاری π با $I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha})$ در $\hat{\beta}^{SR}$ به دست می‌آید تا یک وزن تصادفی به وجود آید. اما واضح است که اگر $\alpha = 1$ ، آن‌گاه فرضیه H_0 همواره رد می‌شود، $\hat{\beta}^{PT} = \hat{\beta}^U$ و اگر $\alpha = 0$ ، آن‌گاه فرضیه H_0 همواره پذیرفته می‌شود، $\hat{\beta}^{PT} = \hat{\beta}^R$. از آنجایی که $0 < \alpha < 1$ ، ترکیبی از برآوردگرهای $\hat{\beta}^R$ و $\hat{\beta}^U$ است.

بنکرافت (۱۹۴۴، ۱۹۶۴) اولین کسی بود که ایده آزمون اولیه را برای برآورد واریانس در جدول تحلیل واریانس به کار برد. وی ایده‌اش را از روش پیشنهادی **سندگر (۱۹۳۸)** و یا **برکسون (۱۹۴۲)** الهام گرفته بود. سندگر با فرض تساوی واریانس‌های مجهول، برابری میانگین‌ها را مورد آزمون قرار داده بود. بدین صورت که اگر آزمون برابری واریانس‌ها رد نشود، آن‌گاه در آزمون t استودنت، از واریانس جمع‌شده و در غیر این صورت از دسته‌بندی بهرنز-فیشر استفاده می‌گردد. تعمیم‌های مختلفی از این روش را می‌توان در مقاله‌های **هان و بنکرافت (۱۹۶۸)** و **بنکرافت و هان (۱۹۸۳)** مشاهده کرد. اخیراً **کلین و همکاران (۲۰۱۷)** در مدل‌های اقتصادسنجی، کاربردی از این برآوردگر را ارائه نموده‌اند. **کربلایی (۱۳۹۶)** برآوردگر آزمون اولیه را با در نظر گرفتن آزمون‌های آماری مختلف بررسی کرده است.

ذکر این نکته ضروری است که در نهایت در هر آزمون، $\hat{\beta}^{PT}$ به یکی از دو حالت فرین $\hat{\beta}^U$ یا $\hat{\beta}^R$ تبدیل می‌شود و بنابراین خاصیت گسسته‌ای دارد. خاصیت گسسته‌سازی تابع نشان‌گر باعث می‌شود که نتوان رفتار برآوردگرهای $\hat{\beta}^U$ و $\hat{\beta}^R$ را همزمان در برآوردگر آزمون اولیه بررسی کرد. از طرفی رفتار برآوردگر آزمون اولیه شدیداً به سطح معنی‌داری آزمون وابسته است. بنابراین ایده استفاده از برآوردگری که بین دو برآوردگر محدودشده و محدودنشده تغییر کند، مطرح می‌شود. در این راستا، برآوردگر انقباضی نوع استاین بردار پارامترهای رگرسیونی β به صورت زیر تعریف می‌شود:

$$\begin{aligned}\hat{\beta}^S &= \hat{\beta}^R + (1 - c\mathcal{L}_n^{-1})(\hat{\beta}^U - \hat{\beta}^R) \\ &= \hat{\beta}^U - c\mathcal{L}_n^{-1}(\hat{\beta}^U - \hat{\beta}^R).\end{aligned}\quad (12.1)$$

که در آن، ضریب $(1 - c\mathcal{L}_n^{-1})$ را ثابت انقباضی گویند. دقت کنید که $\hat{\beta}^S$ ترکیب محدبی از دو برآوردگر $\hat{\beta}^U$ و $\hat{\beta}^R$ نیست بلکه بر اساس برآوردگر کلاسیک جیمز-استاین ساخته می‌شود (**جیمز و استاین، ۱۹۶۱**).

با توجه به ساختار برآوردگرهای $\hat{\beta}^{PT}$ و $\hat{\beta}^S$ به ازای مقادیر بزرگ آماره آزمون \mathcal{L}_n ، هر دوی آن‌ها، برآوردگر $\hat{\beta}^R$ را نتیجه می‌دهند. در حالی که برای مقادیر کوچک \mathcal{L}_n ، رفتار $\hat{\beta}^{PT}$ متفاوت از $\hat{\beta}^S$ است.

همان‌طور که قبلاً اشاره شد، برآوردگر آزمون اولیه می‌تواند به‌طور یکنواخت بهبود داده شود. در این راستا، برآوردگر آزمون اولیه بهبودیافته پارامتر β از رابطه زیر به دست می‌آید:

$$\hat{\beta}^{IPT} = \hat{\beta}^{PT} - c\mathcal{L}_n^{-1}I(\mathcal{L}_n > \mathcal{L}_{n,\alpha})(\hat{\beta}^U - \hat{\beta}^R).\quad (13.1)$$

از آن جایی که به ازای مقادیر $c > \mathcal{L}_n$ ، ثابت انقباضی $(1 - c\mathcal{L}_n^{-1})$ منفی می‌شود و منجر به حذف اثر انقباض می‌گردد، تعمیمی از برآوردگر انقباضی نوع استاین معرفی می‌شود که فقط از قسمت مثبت آن استفاده می‌کند و PRSE نامیده و به صورت زیر تعریف می‌شود.

$$\begin{aligned}\hat{\beta}^{S+} &= \hat{\beta}^R + (1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n > c)(\hat{\beta}^R - \hat{\beta}^U) \\ &= \hat{\beta}^S - (1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \leq c)(\hat{\beta}^R - \hat{\beta}^U)\end{aligned}\quad (14.1)$$

این برآوردگر جدید، ترکیبی محدب از برآوردگرهای محدودنشده $\hat{\beta}^U$ و محدودشده $\hat{\beta}^R$ است و از این جهت اهمیت دارد که مسئله بیش انقباضی را در برآوردگر انقباضی نوع استاین کنترل می‌کند و به طور یکنواخت بر برآوردگر $\hat{\beta}^S$ برتری دارد.

برای سادگی، در ادامه از اصطلاح برآوردگرهای بهبودیافته به جای برآوردگرهای آزمون اولیه، آزمون اولیه بهبودیافته، انقباضی نوع استاین و جزء مثبت انقباضی نوع استاین استفاده می‌شود، زیرا این برآوردگرها مخاطره برآوردگر محدودنشده را کاهش داده و کارایی این برآوردگر را بهبود می‌دهند.

۳.۱ تاریخچه برآوردگرهای انقباضی

مطالعات زیادی در بهبود برآوردگرها انجام شده است. ایده اصلی در این مطالعات، استفاده مناسب از اطلاعات نمونه‌ای و غیرنمونه‌ای در توسعه هرچه بهتر برآوردگرهای انقباضی نوع استاین است. **جاج و باک (۱۹۷۸)** برآوردگرهای آزمون اولیه و نوع استاین و کاربردی از آن را در اقتصادسنجی بررسی کرده است. **کلارک و همکاران (۱۹۸۷)** برآوردگر آزمون اولیه خطای واریانس را در مدل رگرسیون خطی مطالعه کرده است. **بندا (۱۹۹۶)**؛ **صالح و هان (۱۹۹۰)** برآوردگرهای آزمون اولیه و انقباضی را در مدل رگرسیون خطی بررسی کرده‌اند.

برآوردگرهای انقباضی در حالتی که توزیع خطاها نرمال نباشد، توسط نویسندگان بسیاری بررسی شده‌اند. **سینگ (۱۹۹۱)** توزیع خطاها را t چندمتغیره در نظر گرفت. **خان و صالح (۱۹۹۷)** برآورد انقباضی پارامتر عرض از مبدا را در این خانواده‌ها معرفی کرده‌اند. **جایلز (۱۹۹۱)**، **(۱۹۹۲)** خانواده بزرگتر توزیع‌های چندمتغیره کروی متقارن را برای خطا در نظر گرفتند و برآوردگر آزمون اولیه را با در نظر گرفتن پذیره‌های مختلف بررسی کردند. **طباطبایی (۱۹۹۵)** برآوردگرهای انقباضی را در این خانواده به طور کامل معرفی و ویژگی‌های مختلف آن‌ها را مورد مطالعه قرار داده است.

گروبر (۱۹۹۸) به مروری بر روش انقباضی جیمز - استاین و ریج در مدل‌های رگرسیونی می‌پردازد. **شالاب (۱۹۹۸)** برآوردگرها را در مدل رگرسیونی با خطای اندازه‌گیری با به‌کارگیری تکنیک انقباضی بهبود می‌بخشد. **شالاب (۲۰۰۱)** بر اساس معیار پیتمن، مقایسه‌ای بین برآوردگر OLS و نوع - استاین در مدل‌های رگرسیونی هنگامی که توزیع خطاها نرمال نباشد، انجام داده است.

آرشی و طباطبایی (۲۰۰۹) برآوردگر بهبودیافته واریانس را با در نظر گرفتن یک زیرفضای محدودشده محاسبه کرده است. **رحیم و احمد (۲۰۱۱)** با به کارگیری روش مونت کارلو در یک مطالعه شبیه‌سازی، کارایی برآوردگرهای انقباضی جزء مثبت نوع استاین و آزمون اولیه را در مدل رگرسیون خطی چندگانه بررسی نموده‌اند. **آرشی و همکاران (۲۰۱۴)** برآوردگرهای انقباضی لیو را در حالتی که توزیع خطاها بیضی‌گون باشند، معرفی کرده‌اند. **تولومیس (۲۰۱۵)** برآوردگر ناپارامتری نوع استاین ماتریس واریانس-کوواریانس را در مدل‌های رگرسیونی بعد بالا مورد مطالعه قرار داده‌اند. **آرشی و روزبه (۲۰۱۵)** برآوردگر انقباضی را در دستگاه معادلات رگرسیونی معرفی کرده‌اند.

احمد (۲۰۰۱) برآورد انقباضی داده‌های سانسور شده را در مدل رگرسیونی چندگانه ارائه کرده است. **احمد و همکاران (۲۰۰۷)**، **رحیم (۲۰۱۲)**، **آرشی و همکاران (۲۰۱۲)**، **رحیم و همکاران (۲۰۱۲)**، **روزبه و آرشی (۲۰۱۳، ۲۰۱۶)** برآوردگرهای بهبودیافته را در رگرسیون خطی جزئی بررسی نموده‌اند. **احمد و نیکل (۲۰۱۲)** کاربردی از برآوردگرهای انقباضی را در مدل رگرسیونی غیرخطی نشان داده‌اند. **خان و احمد (۲۰۰۳)** مسئله برآورد بردار ضرایب رگرسیون کلاسیک را با در نظر گرفتن اطلاع پیشینی که بیان می‌کرد پارامترها متعلق به یک زیرفضا هستند، در نظر گرفتند و از لحاظ تئوری و عددی برآوردگرهای بهبودیافته را مقایسه کردند. **آرشی و همکاران (۲۰۱۳، ۲۰۱۵)** در خانواده توزیع‌های بیضی‌گون، برآوردگرهای انقباضی را مورد بررسی قرار دادند. **توره‌زاده و آرشی (۲۰۱۴)** برآوردگرهای انقباضی موجک‌ها را ارائه نموده‌اند. **روزبه (۲۰۱۵)** این برآوردگرها را در مدل‌های ناپارامتری بررسی کرده است. **عربی‌بلاغی و همکاران (۲۰۱۵)** برآوردگرهای بهبودیافته را رگرسیونی مطالعه نموده است.

به جرأت می‌توان گفت **صالح (۲۰۰۶)** کامل‌ترین مرجعی است که برآوردگرهای بهبودیافته را در برخی مدل‌های رگرسیونی با در نظر گرفتن محدودیت‌های تصادفی و غیر تصادفی بررسی کرده است. **صالح و همکاران (۲۰۱۴)** این برآوردگرها را در حالتی که خطاها دارای توزیع t چندمتغیره باشد، در مدل‌های مختلف رگرسیونی تحلیل نموده‌اند. **احمد (۲۰۱۴)** نیز از این تکنیک‌ها در کنار روش برآورد جریمه‌شده برای برآورد بردار پارامتر در رگرسیون چندگانه با خطاهای دارای توزیع نرمال و پواسون بهره گرفته است. با توجه به ماهیت این برآوردگرها، در شاخه‌های مختلف آمار و علوم دیگر کاربرد گسترده‌ای دارند و پایان‌نامه‌ها و مقاله‌های بسیاری منتشر شده است.

طباطبایی (۱۹۹۵) با در نظر گرفتن توزیع t چندمتغیره و هم‌چنین توزیع‌های متقارن کروی برای بردار خطا برآوردگرهای بهبودیافته را بررسی کرده است. **آرشی (۱۳۸۷)** این برآوردگرها را در مدل‌های رگرسیونی با بردار خطای دارای خانواده توزیع کلی‌تر بیضی‌گون ارائه نموده است. در این راستا، **آرشی و طباطبایی (۲۰۰۹، ۲۰۱۰)**؛ **آرشی و همکاران (۲۰۱۱)** و **آرشی (۲۰۱۲)** را ببینید. **روزبه (۱۳۹۰)** در مدل‌های خطی جزئی، **عربی‌بلاغی (۱۳۹۳)** در رگردها، برآوردگرهای بهبودیافته را معرفی کرده‌اند. **فلاح (۱۳۹۶)** برآوردگرهای انقباضی را در برخی از مدل‌های خطی چندگانه مورد بررسی قرار داده است.

۱.۳.۱ برآوردگرهای جریمه‌شده

گاهی ممکن است یکی از مولفه‌های بردار پارامتر β بزرگ باشد. بزرگی این مقدار، افزایش تابع زیان را به دنبال خواهد داشت و از آنجایی که هدف کمینه کردن این تابع است، و احتمال اینکه مولفه متناظر برآوردگر به اندازه مقدار پارامتر، بزرگ باشد، تقریباً صفر است؛ کمینه کردن تابع زیان منجر به بیش‌برازشی می‌شود. یکی از روش‌هایی که می‌توان در برطرف کردن این مشکل به کار گرفت این است که به اصطلاح پارامتر جریمه شود. ایده رگرسیون جریمه‌شده از اینجا سرچشمه می‌گیرد.

برای درک بهتر رهیافت جریمه‌شده، بهتر است از دیدگاه بهینه‌سازی به آن نگرسته شود. در حالت کلی، مسئله بهینه‌سازی زیر را در نظر بگیرید:

$$\min_{x \in X} f(x) \quad \text{s.t.} \quad g(x) \geq g_0$$

برای حل مسائل بهینه‌سازی مقید، روش‌های مختلفی وجود دارد که یکی از آنها استفاده از تابع جریمه است. در این روش، اجازه تخطی از قیدهای مسئله به پاسخ‌های پیشنهادی داده می‌شود اما هر پاسخ، بر اساس میزان تخطی‌اش باید جریمه‌ای بپردازد. این جریمه در قالب بدتر کردن کیفیت پاسخ با دستکاری در مقدار هدف، پیاده‌سازی می‌شود. مثلاً در مسائل کمینه‌سازی، تابع جریمه باعث افزایش مقدار تابع هدف و بدتر شدن پاسخ می‌شود. میزان تخطی از یک قید را با معیار تابع تخطی نشان می‌دهند. اگر شرط به صورت $g(x) \geq g_0$ باشد، در این صورت، تابع تخطی مطلق به صورت زیر تعریف می‌شود:

$$V \{g(x) \geq g_0\} = \begin{cases} 0 & g(x) \geq g_0 \\ g_0 - g(x) & g(x) < g_0 \end{cases} \quad (15.1)$$

اما معمولاً از تابع تخطی نسبی به صورت زیر در مسائل کاربردی استفاده می‌شود:

$$V \{g(x) \geq g_0\} = \begin{cases} 0 & g(x) \geq g_0 \\ 1 - \frac{g(x)}{g_0} & g(x) < g_0 \end{cases}$$

لذا، تابع هدف زیر تعریف می‌شود

$$f(x) + \lambda_n P(V)$$

که در آن $P(V)$ و $P(0) = 0$ صعودی است.

در رابطه (۱۵.۱)، بدون از دست دادن کلیت مسئله، فرض کنید $g_0 = 0$ ، در این صورت تابع تخطی به صورت زیر تعریف می‌شود:

$$V \{g(x) \geq 0\} = \begin{cases} 0 & g(x) \geq 0 \\ -g(x) & g(x) < 0 \end{cases}$$

برای تابع ثابت (و البته صعودی!) $P(V) = V$ ، تابع هدف به صورت زیر خواهد بود:

$$f(x) - \lambda_n g(x).$$

عبارت $\lambda_n g(x)$ تابع جریمه نامیده می‌شود. از آن جایی که $\lambda_n > 0$ ، اگر $g(x)$ مثبت باشد، $-\lambda_n g(x)$ به کاهش تابع هدف کمک می‌کند و اگر $g(x) < 0$ ، وضعیت تابع هدف بدتر می‌شود. ضریب تنظیم λ ، ضریب جریمه است که متفاوت از ضریب لاگرانژ می‌باشد. ضریب لاگرانژ، متغیر است ولی ضریب جریمه، حالت دینامیک ضریب لاگرانژ را از آن گرفته است. مقدار ضریب جریمه، باید با ظرافت تعیین شود. دقیقاً مانند نظام حقوقی و جزایی انسان‌ها، در دنیای بهینه‌سازی مقید نیز باید مقدار جریمه متناسب با میزان تخطی و جرم انتخاب شود. اگر جریمه کمتر از مقدار مناسب باشد، باعث تکرار جرم می‌شود. در مقابل، اگر جریمه برای هر نوع جرمی بسیار بالا در نظر گرفته شود، افراد در مقابل جریمه، بی تفاوت می‌شوند و تمایل به ارتکاب جرایم کوچکتر، بیشتر خواهد شد.

یافتن ضریب جریمه، از اهمیت بسزایی برخوردار است و حتی گاهی اوقات از اصل مسئله دشوارتر است. روش‌های کلاسیک، مشابه ضریب لاگرانژ با ضریب جریمه رفتار می‌کند اما در هر مسئله‌ای قابل اجرا نیست. بنابراین، روش‌های متفاوتی ممکن است برای دستیابی به مقدار مناسب λ_n پیشنهاد شود که از آن جمله می‌توان به روش اعتبارسنجی متقابل که مرسوم می‌باشد، اشاره کرد.

مسئله بهینه‌سازی در آمار، روش منظم‌سازی نامیده می‌شود. در حالت کلی در رگرسیون خطی جریمه‌شده، هدف کمینه کردن تابع زیر است

$$\begin{aligned} \phi(\beta) &= S(\beta) + \lambda_n P(\beta) \\ &= (\mathbf{y}_n - \mathbf{X}_n \beta)^\top (\mathbf{y}_n - \mathbf{X}_n \beta) + \lambda_n P(\beta). \end{aligned} \quad (16.1)$$

با دقت در ساختار تابع (16.1)، استنباط می‌شود که به تابع توان دوم خطا در رگرسیون معمولی، عبارت جریمه $P(\beta)$ اضافه شده است. در واقع با جریمه کردن پارامتر، سعی شده است مقدار خطا کمینه شود. این عبارت جریمه، تابعی از پارامترهاست و بنا به روش رگرسیون جریمه شده، صورت‌های متفاوتی دارد.

در این حالت، $\lambda_n > 0$ نیز پارامتر تنظیم‌کننده از قبل تعیین شده است. λ_n در اینجا، موازنه‌ای بین خطای پیش‌گویی و پیچیدگی مدل انجام می‌دهد و بنابراین، پاسخ این مسئله، به انتخاب λ به شدت وابسته است. تعیین مقدار مناسب برای پارامتر تنظیم‌کننده در آمار نیز، یک مسئله چالش‌برانگیز است. برای جزئیات بیشتر درباره ارتباط بین مسائل بهینه‌سازی ریاضی با آمار و به‌ویژه روش‌های رگرسیونی، وانگ و همکاران (۲۰۰۴) و یائو (۲۰۰۴) را ببینید. جریمه‌های متفاوتی می‌تواند در نظر گرفته شود که در هر حالت، برآوردگر حاصل از کمینه‌سازی (16.1) نام خاصی به خود می‌گیرد. جریمه‌های $\sum_{j=1}^p |\beta_j|$ و $\sum_{j=1}^p \beta_j^2$ از معروفترین توابع جریمه هستند که به ترتیب بر اساس نرم‌های مستطیلی (L_1) و اقلیدسی (L_2) هستند.

برآوردگر ریج

با توجه به اینکه در مدل رگرسیون خطی (۱.۱)، $\text{Var}(\hat{\beta}^U) = \sigma^2 C_n^{-1}$ ، هم خط بودن متغیرهای کمکی می‌تواند واریانس برآوردهای ضرایب رگرسیونی را افزایش دهد که باعث می‌شود از مقادیر درست و واقعی پارامتر β فاصله گرفته و برآوردها را به کمترین تغییرات در مدل، خیلی حساس کند. بنابراین، در این حالت برآوردهای ضرایب ناپایدار هستند و بعضی از ضرایب ممکن است از نظر آماری معنی‌دار نباشند و علامت اشتباه داشته باشند. برای درک بهتر، مثال ۱.۳.۱ را ملاحظه کنید.

مثال ۱.۳.۱. فرض کنید مدل به صورت $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ بوده که در آن متغیرهای X_1 و X_2 به شدت همبسته‌اند. برای ایجاد همبستگی شدید، ۲۰ عدد از توزیع نرمال استاندارد به عنوان متغیر X_1 و ۲۰ عدد از توزیع نرمال با میانگین X_1 و انحراف معیار ۱٪ تولید شده که همبستگی بین این دو متغیر، ۹۹٪ به دست می‌آید. هم‌چنین مقدار واقعی بردار پارامترهای β به صورت $(3, 1, 1)$ باشد. به روش OLS، برآورد ضرایب به صورت $(-20/39, 22/60, 3/09)$ به دست می‌آید. همان‌طور که ملاحظه می‌شود برآوردها با مقدار واقعی پارامتر اختلاف زیادی دارند که به دلیل همبسته بودن متغیرهای X_1 و X_2 است. مقدار تابع $S(\beta)$ به ازای مقدار واقعی β ، برابر $12/84$ و به ازای برآوردگر OLS برابر $10/72$ است که اختلاف زیادی مشاهده نمی‌شود. همین‌طور که مشاهده می‌شود $\beta_1 + \beta_2 = 2$ و مجموع مولفه‌های دوم و سوم برآوردگر OLS برابر $2/22$ است که بسیار به هم نزدیک هستند. اما $12 + 12 = 24$ و $\beta_1^2 + \beta_2^2 = 2$ عدد بسیار کوچکتری از $926/6298 = (-20/39)^2 + (22/60)^2$ است. بنابراین از روش‌های موثر در این مسئله این است که تابع جریمه $\lambda \sum_{j=1}^p \beta_j^2$ در رابطه (۱۶.۱) در نظر گرفته شود. بردار $(1, 1)$ به عنوان پاسخ، جریمه کمتری را منجر می‌شود و نسبت به $(-20/39, 22/60)$ از علاقه‌مندی بیشتری برخوردار است.

برآوردگر ریج، مشهورترین عضو کلاس برآوردگرهای جریمه‌شده است. **هورل و کنارد (۱۹۷۰)** برآوردگر ریج را معرفی کردند که دروازه ورود به دنیای برآوردگرهای جریمه‌شده بر اساس روش منظم‌سازی **تیکونوف (۱۹۶۳)** بود. و رایج‌ترین روشی است که برای از بین بردن مشکل هم‌خطی برآوردگر OLS به کار می‌رود (**گروبر، ۱۹۹۸**). این رگرسیون، مسئله هم‌خطی را با جریمه کردن پارامتر β به صورت زیر حل می‌کند:

$$\begin{aligned} \hat{\beta}^{\text{Rdg}} &= \operatorname{argmin}_{\beta} \left\{ (\mathbf{y}_n - \mathbf{X}_n \beta)^\top (\mathbf{y}_n - \mathbf{X}_n \beta) + \lambda \sum_{j=1}^p \beta_j^2 \right\} \\ &= \operatorname{argmin}_{\beta} \left\{ (\mathbf{y}_n - \mathbf{X}_n \beta)^\top (\mathbf{y}_n - \mathbf{X}_n \beta) + \lambda \beta^\top \beta \right\} \\ &= (\mathbf{X}_n^\top \mathbf{X}_n + \lambda \mathbf{I}_p)^{-1} \mathbf{X}_n^\top \mathbf{y}_n \\ &= (\mathbf{C}_n + \lambda \mathbf{I}_p)^{-1} \mathbf{X}_n^\top \mathbf{y}_n \end{aligned} \quad (17.1)$$

که در آن $\mathbf{C}_n = \mathbf{X}_n^\top \mathbf{X}_n$ و λ ، پارامتر تنظیم‌کننده یا پارامتر ریج است.

در حالتی که $C_n = I_p$ ، ماتریس طرح، متعامد خواهد بود. در این حالت، برآورد ریج به صورت زیر به دست می‌آید:

$$\hat{\beta}^{\text{Rdg}} = \frac{1}{1 + \lambda} \hat{\beta}^{\text{Ols}}$$

این رابطه، ویژگی اساسی رگرسیون جریمه‌شده یعنی ماهیت انقباضی بودن آن را به خوبی نشان می‌دهد. در واقع، در روش‌های جریمه‌شده مشابه روش‌های انقباضی، اطلاع پیشین غیرقطعی وجود دارد. این اطلاع در روش‌های جریمه‌شده بر اساس این ادعا است که مقدار ضرایب پیرامون صفر قابل‌اعتمادتر از ضرایبی هستند که از صفر فاصله زیادی دارند. در رگرسیون جریمه‌شده ریج، پارامتر λ را پارامتر انقباضی نیز می‌نامند زیرا مقدار انقباض به سمت صفر را کنترل می‌کند.

در برآورد OLS، اگر ماتریس X_n رتبه کامل (پرتبه) نباشد، آن‌گاه $X_n^T X_n$ وارون‌پذیر نخواهد بود و بنابراین، پاسخ یکتایی برای برآوردگر OLS وجود ندارد. اما برای هر ماتریس دلخواه X_n ، عبارت $X_n^T X_n + \lambda I_p$ همواره وارون‌پذیر است و در نتیجه، پاسخ یکتا برای برآوردگر ریج وجود دارد. همچنین برای هر مقدار دلخواه پارامتر β ، همیشه λ ای وجود دارد که MSE آن کمتر از MSE برآوردگر OLS باشد. به عبارت دیگر، اگر مدل برازش شده کاملاً درست باشد و همچنین از توزیع مشخصی دقیقاً پیروی کند، همواره با منقبض کردن ضرایب به سمت صفر، می‌توان برآوردگر بهتری یافت.

به کار بردن جریمه ریج، منجر به منقبض شدن برآوردها به سمت صفر می‌شود. همچنین، برآوردگر ریج، یک برآوردگر آریب است اما واریانس برآورد را در مقایسه با برآوردگر OLS کاهش می‌دهد که در نهایت کاهش MSE را به دنبال خواهد داشت، می‌توان بیان کرد برآوردگر ریج در مقایسه با برآوردگر OLS، برآوردگر بهتری است. لازم به ذکر است که به کار بردن جریمه ریج باعث نمی‌شود هیچ برآوردی صفر شود، بلکه یک برآورد هموار را ارائه می‌کند.

رهیافت ریج در مثال ۱.۳.۱ به ازای مقدار دلخواه $\lambda = 1$ ، بردار β را به صورت $(3/09, 1/12, 1/08)$ برآورد می‌کند که به مقادیر واقعی پارامتر نزدیک می‌باشد.

برآوردگرهای ریج با توزیع خطای نرمال توسط نویسندگان متعددی بررسی شده است. به گیبونز (۱۹۸۱)؛ سارکار (۱۹۹۲)؛ گروبر (۱۹۹۸)؛ کیریا (۲۰۰۳)؛ حسن‌زاده‌بشتیان و همکاران (۲۰۱۱b)؛ مونیز و همکاران (۲۰۱۲)؛ نجاریان و همکاران (۲۰۱۳) می‌توان اشاره کرد. رگرسیون ریج به طور گسترده‌ای در مدل‌های مختلف آماری به کار گرفته شده که می‌توان هورل و همکاران (۱۹۷۵)، اینوآ (۲۰۰۱)، طباطبایی و همکاران (۲۰۰۴)، منسون و همکاران (۲۰۱۰)، مونیز و همکاران (۲۰۱۲) و تاباکان و آکدنیز (۲۰۱۰) را برای جزئیات بیشتر مشاهده نمود.

همان‌طور که اشاره شد، انتخاب پارامتر جریمه (در اینجا، پارامتر ریج) از اهمیت بسزایی برخوردار است. در این راستا روش‌های مختلفی پیشنهاد شده که می‌توان به کیریا (۲۰۰۳)؛ مونیز و کیریا (۲۰۰۹)؛ صالح و کیریا (۲۰۱۱)؛ کیریا و همکاران (۲۰۱۵) مراجعه کرد.

برآوردگر لاسو

اگر چه رگرسیون ريج، بهبود پيش گویی را نتيجه می دهد ولی با توجه به حضور همه متغيرها در آن، تفسير پذیری به سادگی امکان پذیر نیست. از طرف دیگر، حضور متغيرهایی که با متغير پاسخ در یک مدل ارتباطی ندارند، در فرآیند برآورد، منجر به برآوردگری با واریانس زیاد می شود. روشی ساده و بدیهی برای حل این مسئله، انتخاب یک زیرمجموعه از متغيرهای توضیحی با اهمیت و تاثیرگذار در مدل می باشد. انتخاب این متغيرها می تواند بر اساس معنی داری آن ها در مدل یا برخی از معیارهای اطلاع مانند AIC (آکائیک، ۱۹۷۳)، BIC (شوارتز، ۱۹۷۸)، C_p - مالو (مالوز، ۱۹۷۳) و ... انجام شود. اما مشکل بزرگی که در هنگام استفاده از این معیارها مطرح می شود تعداد کل زیرمجموعه هایی از متغيرها است که باید در نظر گرفته شوند، که به طور نمایی با افزایش p رشد می کند و از نظر محاسباتی برای تعداد متغيرهای بیشتر از ۴۰ یا ۵۰، تقریباً نشدنی است. مشکل دیگر این روش ها، ویژگی گسسته بودن آن ها است. به عبارت دیگر، تغییر بسیار کوچکی در داده ها، می تواند برآوردهای کاملاً متفاوتی را نتيجه دهد. در مجموع، روش های انتخاب بهترین زیرمجموعه، اغلب ناپایدار و به شدت تغییرپذیر هستند. روش دیگر، برای رویارویی با این مسئله، استفاده از رگرسیون جریمه شده است. اگر چه به کاربردن جریمه ريج، در رگرسیون جریمه شده، به دستیابی به برآوردگرهایی با واریانس کمتر کمک بسیاری می کند اما وقتی قدرمطلق مقدار ضرایب رگرسیونی زیاد باشد، برآوردگر ريج اریبی زیادی به سمت صفر ایجاد می کند و از طرف دیگر، با وجودی که متغيرهای بی اهمیت به سمت صفر منقبض شده اند، هم چنان در مدل حضور دارند که باعث می شود تقسیرپذیری مدل به راحتی انجام نشود. در این راستا، تیشیرانی (۱۹۹۶) روشی جدید برای انتخاب متغيرها پیشنهاد نمود که منجر به مدل های دقیق، پایا و صرفه جو (تفسیرپذیرتر) می شود و نام آن را لاسو نامید.

لاسو به حلقه طنابی گفته می شود که یک گاوچران به سمت هدف مشخصی پرتاب می کند و پس از اصابت کردن به آن، طناب کشیده و حلقه کوچکتر می شود. از دیدگاه آماری لاسو، مخفف عملگر انتخاب کننده، منقبض کننده و کمینه کننده از جنس قدرمطلق است. برآوردگر لاسو، نوع جریمه شده برآوردگر OLS است که بر اساس ویژگی تنکی نرم L_1 ، در سال های اخیر مورد توجه قرار گرفته است. برآوردگر لاسو تیشیرانی (۱۹۹۶) از حل مسئله بهینه سازی زیر به دست می آید:

$$\hat{\beta}_n^{\text{Lso}} = \min_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p X_{ij}\beta_j)^2 \right\} \quad \text{s.t.} \quad \sum_{i=1}^p |\beta_j| \leq t, \quad (18.1)$$

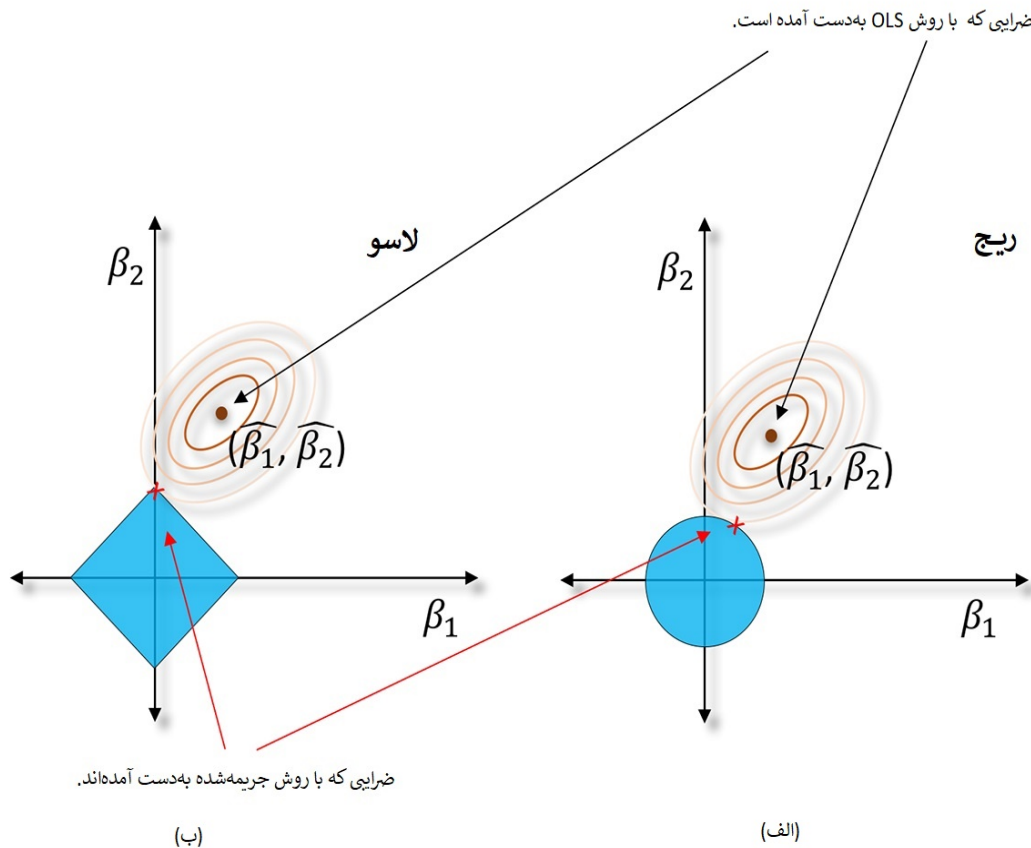
که در آن t یک مقدار ثابت است. اگر $t = 0$ ، آن گاه متغیری در مدل وجود نخواهد داشت در حالی که اگر $t = \infty$ ، آن گاه مدل، کامل خواهد بود. فرض کنید $\hat{\beta}^* = (\hat{\beta}_1^*, \dots, \hat{\beta}_p^*)^T$ یک برآوردگر اولیه برای β است که معمولاً این برآوردگر را $\hat{\beta}^{\text{Ols}}$ در نظر می گیرند. اگر $t > \sum_{i=1}^p |\hat{\beta}_i^*|$ ، روش لاسو منجر به برآوردگر OLS می شود.

اگر $0 < t < \sum_{j=1}^p |\hat{\beta}_j^{OLS}|$ باشد، آن‌گاه مسئله بهینه‌سازی (۱۸.۱) معادل با

$$\hat{\beta}^{Lso} = \operatorname{argmin}_{\beta} \left\{ (\mathbf{y}_n - \mathbf{X}_n \beta)^T (\mathbf{y}_n - \mathbf{X}_n \beta) + \lambda \sum_{j=1}^p |\beta_j| \right\}. \quad (19.1)$$

است، که در آن λ پارامتر تنظیم‌کننده است و سطح تنکی (تعداد پارامترهای صفر) را در برآوردگر لاسو کنترل می‌کند. وقتی $\lambda \rightarrow 0$ ، $\hat{\beta}^{Lso} \rightarrow \hat{\beta}^{OLS}$ و وقتی $\lambda \rightarrow \infty$ ، $\hat{\beta}^{Lso} \rightarrow \mathbf{0}$. به عبارت دیگر، هرچه جریمه بزرگتری به کار برده شود، تعداد بیشتری از ضرایب به سمت صفر منقبض می‌شوند. در حقیقت λ و t رفتار یا رابطه‌ای معکوس با یکدیگر دارند.

به رگرسیون لاسو، به خاطر شکل تابع جریمه آن، رگرسیون جریمه‌شده L_1 نیز می‌گویند. انتخاب این تابع جریمه باعث می‌شود که برآوردگر لاسو، ضرایب را به سمت صفر منقبض کرده و متغیرهای اضافی را از مدل حذف نماید. در واقع، برآوردگر لاسو، همزمان هم انتخاب متغیر انجام می‌دهد و هم ضرایب را منقبض می‌کند، شکل ۱.۱ این ویژگی را به خوبی نشان می‌دهد.



شکل ۱.۱: نمودار برآورد دوبعدی در رگرسیون ریج (الف) و رگرسیون لاسو (ب). ناحیه محدودیت لاسو (تابع جریمه) $|\beta_1| + |\beta_2| = c$ و در رگرسیون ریج، $\beta_1^2 + \beta_2^2 = c$ است. بیضی‌ها، منحنی‌های تراز خطای برآورد کمترین توان‌های دوم هستند.

تابع جریمه قدرمطلق باعث می‌شود که جواب صریحی برای برآوردگر لاسو وجود نداشته باشد. **تیبشیرانی (۱۹۹۶)** این برآوردگر را از طریق برنامه‌ریزی درجه دوم به‌دست آورده بود. **افرون و همکاران (۲۰۰۴)** نوع دیگری از رگرسیون گام به گام را به نام رگرسیون کمترین زاویه (LAR) معرفی نمودند که برآوردهای لاسویی را نتیجه می‌دهد که هزینه محاسباتی آن با برآوردهای کمترین توان‌های دوم برابر است.

انتخاب پارامتر جریمه بخش مهمی از برازش مدل در رگرسیون جریمه‌شده است. روش‌های بسیاری در این مورد وجود دارد اما مهمترین روش‌های کاربردی بر اساس این است که برآوردگر جریمه‌شده چه قدر خوب می‌تواند مقدار واقعی متغیر پاسخ را پیش‌گویی کند. منصفانه نیست از داده‌ها دوبار استفاده شود، یک بار برای برازش مدل و دوباره از همان داده‌ها برای برآورد دقت پیش‌گویی. این کار، منجر به بیش‌برازشی می‌شود.

ایده‌ای که به‌ذهن می‌رسد، تقسیم‌بندی داده‌ها به دو مجموعه است. از یک مجموعه (مجموعه آموزش) برای برازش برآوردگر جریمه‌شده (مثلاً β^*) استفاده شده و از مجموعه دیگر (مجموعه آزمون) برای محاسبه اینکه کمیت $X\beta^*$ چه قدر خوب توانسته است مشاهدات مجموعه دیگر را پیش‌گویی کند. اما به‌ندرت تعداد داده‌ها آن قدر زیاد است که بتوان آن‌ها را به دو قسمت تقسیم کرد به‌طوری که هر یک از مجموعه‌ها به‌تنهایی برای انتخاب پارامتر جریمه کافی باشند. از این رو، روش اعتبارسنجی متقابل پیشنهاد می‌شود.

در اعتبارسنجی متقابل، داده‌ها به‌طور تصادفی به K زیرمجموعه تقسیم می‌شود. یک زیرمجموعه کنار گذاشته می‌شود که نام مجموعه آزمون را به خود می‌گیرد، $K - 1$ مجموعه باقیمانده، مجموعه آموزشی نامیده و برای برازش مدل استفاده می‌گردد. سپس مدل برازش شده به‌منظور پیش‌گویی متغیرهای پاسخ در مجموعه داده آزمون به‌کار برده می‌شود. سرانجام، خطاهای پیش‌گویی با میانگین توان دوم انحراف مقادیر مشاهده شده از مقادیر پیش‌گویی در مجموعه داده آزمون به‌دست می‌آیند. معمولاً $K = 5, 10, n$ انتخاب می‌شود.

فن و لی (۲۰۰۱) نشان دادند که یک تابع جریمه خوب باید برآوردگری با سه ویژگی مطلوب ناریبی، تُنکی و پیوستگی را نتیجه دهد. جریمه L_1 برآوردگرهایی را نتیجه می‌دهد که تُنک، نسبت به داده‌ها پیوسته و اریب هستند زیرا اندازه انقباضی که به ضرایب رگرسیونی کوچک و بزرگ اختصاص می‌دهد، یکسان است (لازم به ذکر است این مورد در خصوص جریمه L_2 صادق نیست). هم‌چنین، لاسو برای یک مدل رگرسیونی خطی با p متغیر توضیحی و n مشاهده، حداکثر n متغیر انتخاب می‌کند و بنابراین اگر متغیرهای بیشتری (بیشتر از n) در مدل معنی‌دار باشند، توسط لاسو انتخاب نمی‌شوند. لاسو، نمی‌تواند متغیرهایی که اثر معنی‌داری کمی دارند را تشخیص دهد و آن را از مدل حذف می‌کند. برای داده‌های هم‌خط مناسب نیست زیرا در بین گروه متغیرهایی که وابستگی شدیدی دارند، تنها یکی از آن‌ها را انتخاب می‌کند. اگر همبستگی بالایی بین متغیرهای توضیحی وجود داشته باشد، کارایی پیش‌گویی لاسو به خوبی برآوردگر ریج نیست. بزرگترین عیب لاسو این است که خاصیت پیش‌گویی را ندارد.

فرض کنید $A = \{j : \beta_j \neq 0\}$ به طوری که $|A| = p_0 < p$. در این حالت، مدل درست تنها به زیرمجموعه‌ای از متغیرها (p_0 تا متغیر) وابسته است. هم‌چنین برآوردگری که از روش δ به دست می‌آید را به صورت $\hat{\beta}(\delta)$ در نظر بگیرید. بر اساس فن و لی (۲۰۰۱) گفته می‌شود δ خاصیت پیش‌گویی دارد اگر $\hat{\beta}(\delta)$ به‌طور مجانبی دارای خواص زیر باشد:

(۱) زیر مدل درست را مشخص کند به عبارتی، مجموعه $\{j : \hat{\beta}_j \neq 0\}$ را طوری مشخص کند که $\{j : \hat{\beta}_j \neq 0\} = A$.

(۲) دارای نرخ برآورد بهینه باشد به عبارتی،

$$\sqrt{n}(\hat{\beta}_A(\delta) - \beta_A) \xrightarrow{D} \mathcal{N}(0, \Sigma^*)$$

که در آن ماتریس کوواریانس زیرمدل درست است.

ژو (۲۰۰۶) نشان داد که برآوردگر لاسو، خاصیت پیش‌گویی ندارد و لاسوی تطبیقی (aLASSO) را با تصحیح جریمه لاسو به صورت زیر معرفی کرد:

$$\hat{\beta}^{\text{aLso}} = \operatorname{argmin}_{\beta} \left\{ (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) + \lambda \sum_{j=1}^p \hat{w}_j |\beta_j| \right\}, \quad (20.1)$$

که در آن $\hat{w} = (\hat{w}_1, \dots, \hat{w}_p)^\top$ بردار وزن بوده و هم‌چنین $\hat{w} = 1/|\hat{\beta}^*|^\gamma$ برای $\gamma > 0$ مقدار ثابتی است. $\hat{\beta}^*$ هر برآوردگر \sqrt{n} سازگار برای پارامتر β مانند OLS یا لاسو است و منظور از $1/|\hat{\beta}^*|^\gamma$ برداری با p مولفه است که مولفه i ام آن $1/|\hat{\beta}_i^*|^\gamma$ بوده و $\hat{\beta}_i^*$ مولفه i ام $\hat{\beta}^*$ است. یکی از مزیت‌های مهم برآوردگر لاسوی تطبیقی این است که برخلاف برآوردگر لاسو دارای خاصیت پیش‌گویی است. (ژو، ۲۰۰۶).

۲.۳.۱ ترکیب روش‌های انقباضی و جریمه‌شده

می‌توان رابطه (۱۷.۱) را به صورت

$$\begin{aligned} \hat{\beta}^{\text{Rdg}} = \hat{\beta}^{\text{U}}(\lambda) &= (\mathbf{X}_n^\top \mathbf{X}_n + \lambda \mathbf{I}_p)^{-1} \mathbf{X}_n^\top \mathbf{y}_n \\ &= [\mathbf{I}_p + \lambda (\mathbf{X}_n^\top \mathbf{X}_n)^{-1}]^{-1} \hat{\beta}^{\text{Ols}} = \mathbf{R}_n(\lambda) \hat{\beta}^{\text{Ols}}. \end{aligned} \quad (21.1)$$

بازنویسی کرد که در آن $\mathbf{R}_n(\lambda) = [\mathbf{I}_p + \lambda \mathbf{C}_n^{-1}]^{-1}$. ضریب از پیش تعیین شده $\mathbf{R}_n(\lambda)$ عامل ریج نامیده می‌شود که برآوردگر OLS را به برآوردگر ریج تبدیل می‌کند.

برآوردگرهای ریج با صورت ارائه شده در (۲۱.۱) بسیار مورد مطالعه قرار گرفته است، می‌توان به گیبونز (۱۹۸۱)؛ گروبر (۱۹۹۸)؛ اینوا (۲۰۰۱)؛ کیریا و صالح (۲۰۰۳، ۲۰۰۴)؛ مالتھوس (۱۹۹۹)؛ صالح و کیریا (۱۹۹۳)؛ سارکار (۱۹۹۲)؛ نجاریان و همکاران (۲۰۱۳) و روزبه (۲۰۱۵) اشاره کرد.

وقتی $\lambda \rightarrow 0$ ، $\hat{\beta}^{Ols} \rightarrow \hat{\beta}^{Rdg}$ و هنگامی که $\lambda \rightarrow \infty$ ، $\hat{\beta}^{Rdg} \rightarrow 0$. بر طبق (۲۱.۱)، برآوردهای بهبودیافته در رگرسیون ريج به صورت زیر تعریف می‌شوند:

$$\begin{aligned}\hat{\beta}^R(\lambda) &= \mathbf{R}_n(\lambda)\hat{\beta}^R \\ \hat{\beta}^{PT}(\lambda) &= \mathbf{R}_n(\lambda)\hat{\beta}^{PT} \\ \hat{\beta}^S(\lambda) &= \mathbf{R}_n(\lambda)\hat{\beta}^S \\ \hat{\beta}^{S+}(\lambda) &= \mathbf{R}_n(\lambda)\hat{\beta}^{S+}\end{aligned}$$

صالح و کیریا (۱۹۹۳)؛ آکدنیز (۲۰۰۲)؛ مونیز و کیریا (۲۰۰۹)؛ روزبه و همکاران (۲۰۱۱)؛ حسن‌زاده‌بشتیان و همکاران (۲۰۱۱a)؛ آلهیتی و کیریا (۲۰۱۴) و آرشى و همکاران (۲۰۱۴) این برآوردها را در مدل‌های مختلف ريج با در نظر گرفتن محدودیت $H\beta = h$ به دست آورده‌اند.

تکنیک ترکیب روش‌های انقباضی با جریمه‌شده را می‌توان علاوه بر برآوردها ريج بر سایر برآوردهای جریمه‌شده از قبیل لاسو (فصل ۳) به کار برد.

۴.۱ مدل رگرسیون استوار

معمولاً وقتی رگرسیون OLS با استفاده از n مشاهده، برای p متغیر تحت مدل خطی (۱.۱) در نظر گرفته می‌شود، فرض می‌شود توزیع خطاها نرمال است در حالی که در واقعیت ممکن است این چنین نباشد. **فیشر (۱۹۵۶)** برای اولین بار به نتایج نامناسب استفاده از توزیع نرمال به عنوان توزیع خطای تصادفی در بعضی از مدل‌های خطی آماری اشاره کرده و در این راستا رگرسیون با خطاهای دارای توزیع t را ارائه کرد (صالح و همکاران، ۲۰۱۴).

نظریه‌های زیادی در زمینه اقتصاد برای توضیح رفتار عوامل اقتصادی و روابط بین متغیرها بیان می‌شود که ممکن است به صورت نظری جذاب و از نظر منطقی صحیح باشند، ولی نمی‌توانند به صورت کاربردی ارائه شوند مگر اینکه داده‌های واقعی آن‌ها را پشتیبانی کنند. معمولاً از مدل‌های رگرسیونی برای تحلیل روابط بین متغیرهای اقتصادی و بررسی مسائل مربوط به آن بسیار استفاده می‌شود. امروزه در مطالعه رفتارها و مدل‌های پیچیده اقتصادی به دلیل وجود داده‌های پرت و فرین، به این نتیجه رسیده‌اند که استفاده از توزیع نرمال به عنوان مدل خطای تصادفی، باعث کاهش استواری و ناتوانی مدل در تفسیر و تایید مسائل مطرح شده می‌شود. لذا به سمت به کار بردن روش‌های رگرسیونی استوار حرکت کرده‌اند. بر اساس نظر **هوبر (۱۹۸۱)**، استواری به معنای عدم حساسیت به تغییرات کوچک از پذیره‌هایی است که می‌تواند بر روی مدل داده‌ها اعمال شود.

اصطلاح «استوار» دقیق نیست، اما به طور کلی یک شیوه رگرسیونی «استوار» به شیوه‌ای اشاره دارد که علاوه بر اینکه در زمانی که توزیع خطاها نرمال نیست، خوب انجام شود بلکه

نسبت به انحراف‌های کوچک از فرض نرمال بودن نیز حساس نبوده و تغییر چندانی نداشته باشد.

روش OLS برای به‌دست آوردن پارامترها برای هر مشاهده وزن یکسان تعیین می‌کند در حالی که روش‌های استوار قادر هستند به مشاهدات وزن‌های نابرابر اختصاص دهند. در این روش‌ها، مشاهداتی که مانده‌های بزرگی تولید می‌کنند، وزن‌های کمتری دریافت می‌کنند. روش‌های رگرسیون استوار نسبت به روش OLS محاسبات بیشتری و هم‌چنین به در نظر گرفتن برخی فرض‌ها در مورد اختصاص وزن کم به مشاهدات با مانده‌های بزرگ نیاز دارند.

معیارهای کارایی و نقطه فروریزش (BP) دو معیار مهم و کلاسیک برای مقایسه روش‌های مختلف استوار می‌باشد (دونوهو و هوبر، ۱۹۸۳). معیار کارایی به‌منظور اندازه‌گیری کارایی نسبی برآوردهای استوار نسبت به برآورد OLS، هنگامی که توزیع خطا دقیقاً نرمال باشد و داده پرت در بین داده‌ها وجود نداشته باشد، به کار می‌شود.

نمونه دلخواه $z = (z_1, \dots, z_n)$ را در نظر بگیرید که در آن $z_i = (x_i^T, y_i)$. اگر $T(z)$ یک برآورد پارامتر β و z' نمونه دست‌کاری‌شده‌ای باشد که m داده از مجموعه اولیه داده‌های z با داده‌های آلوده (داده‌های پرت) جایگزین شده باشد، در این صورت، نقطه فروریزش نمونه متناهی بر اساس **دونوهو و هوبر (۱۹۸۳)** به‌صورت زیر تعریف می‌شود:

$$BP(z, T) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{z'} \|T(z') - T(z)\| = \infty \right\},$$

به‌عبارت دیگر، کوچکترین سهم m تا از داده‌های پرت در برآورد است که اگر مقادیر $[nm]$ به سمت بی‌نهایت میل کند، آن‌گاه برآوردگر بی‌نهایت شود. برای مثال، میانگین متناهی x_1, \dots, x_n می‌تواند بر اساس یک تابع از فقط یک مشاهده در نمونه به‌صورت زیر نوشته شود:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \left[\sum_{i=1}^{n-1} x_i + x_n \right] = \frac{n-1}{n} \bar{x}_{n-1} + \frac{1}{n} x_n$$

و بنابراین اگر x_n به‌اندازه کافی بزرگ باشد، آن‌گاه \bar{x}_n می‌تواند صرف‌نظر از مقدار سایر داده‌ها، بزرگ شود و به بی‌نهایت میل کند.

میان‌نمونه، به‌عنوان برآوردگر یک میان‌جامعه، می‌تواند تا ۵٪ مقادیر بد را در بین داده‌ها تحمل کند. در حالت کلی، نقطه فروریزش، نمی‌تواند بیشتر از ۵٪ باشد. **روسیو و یوهای (۱۹۸۴)** نشان دادند که برآوردهای OLS مقدار نقطه فروریزش برابر $1/n$ دارند و هنگامی که حجم نمونه بزرگتر شود، به سمت صفر میل می‌کند. بنابراین، یک داده نامتعارف در مشاهده‌ها می‌تواند تأثیر زیادی روی برآورد OLS داشته باشد.

روش‌های استوار زیادی برای دستیابی به بالاترین کارایی یا بیشترین نقطه فروریزش یا هر دو پیشنهاد شده است که می‌توان به برآوردهای زیر اشاره نمود:

L – برآوردها (**روسیو، ۱۹۸۴، ۱۹۸۵**) اغلب از نظر آماری، استوار هستند و مقدار نقطه فروریزش بالایی دارند. کارایی این برآوردها، به‌اندازه سایر برآوردها نیست اما گاهی اوقات، کارایی نسبتاً بالایی دارند.

M- برآوردگرها (هوبر، ۱۹۸۱) به مشاهده‌های پرت در راستای y استوار هستند اما به نقاط اهرمی در x حساس هستند. نقطه فروریزش یک M- برآوردگر، $1/n$ است.

R- برآوردگرها (جیکل، ۱۹۷۲) که مجموع امتیازهای باقیمانده‌های مرتب‌شده را کمینه می‌کند، نسبتاً کارایی بالایی دارد اما نقطه فروریزش آن، مشابه برآوردگر OLS مقدار کمی است. برآوردگرهای کمترین میانه توان دوم (LMS)، (سیجل، ۱۹۸۲)، کمترین توان‌های دوم پیراسته (LTS) (روسیو، ۱۹۸۵) و S- برآوردگرها (روسیو و یوهای، ۱۹۸۴) که به ترتیب میانه توان دوم، مجموع پیراسته توان دوم و واریانس باقیمانده‌ها را کمینه می‌کنند، نقطه فروریزش بالایی اما کارایی کمی دارند.

برآوردهای S تعمیم‌یافته (کروس و همکاران، ۱۹۹۴) مقدار نقطه فروریزشی مشابه با S- برآوردگرها دارد ولی کارایی آن کمی بیشتر شده است.

MM- برآوردها (یوهای، ۱۹۸۷) می‌تواند همزمان به نقطه فروریزش و کارایی بالایی دست یابد.

برآوردگرهای M- تعمیم‌یافته مالوز (مالوز، ۱۹۷۵) و شوایی (هنسچین و همکاران، ۱۹۷۵) وزن نقاط اهرمی در x ها را کاهش می‌دهند اما نمی‌توانند نقاط اهرمی «خوب» یا «بد» را نشان دهند، بنابراین منجر به کاهش کارایی می‌شوند. علاوه بر این، هر دو برآوردگر هنگامی که تعداد متغیرهای توضیحی زیاد باشد، مقدار نقطه فروریزش بالایی دارند.

برآوردهای M- تعمیم‌یافته شوایی یک مرحله‌ای (کواکلی و هرمازبرگر، ۱۹۹۳) مشکل برآوردهای M- تعمیم‌یافته شوایی را حل می‌کند و در یک مرحله محاسبه می‌شود. آن‌ها مقدار نقطه فروریزش و کارایی بالایی دارند. اخیراً گروینی و یوهای (۲۰۰۲) کلاس جدیدی را از برآوردهای استوار با کارایی و مقدار نقطه فروریزش بالایی که برآوردگر کمترین توان‌های دوم موزون کارا و استوار نامیده شده است معرفی کردند.

جزئیات بیشتر درباره برآوردگرهای استوار را می‌توان در ربیعی (۱۳۸۰) و هرمازبرگر و مک‌گین (۱۹۹۸)؛ ساسانتی و همکاران (۲۰۱۴)؛ یو و یائو (۲۰۱۷) مشاهده کرد.

فصل ۲

برآوردگرهای بهبودیافته بر پایه M ریج

در این فصل، روش‌های انقباضی برای دستیابی به برآوردگرهای بهبودیافته زمانی که در داده‌ها هم‌خطی و داده پرت به‌طور همزمان وجود دارد، در مدل‌های کامل و افرازشده رگرسیونی، برآوردگرها معرفی، ویژگی‌های آن‌ها بررسی و در یک مطالعه شبیه‌سازی و مثال واقعی تحلیل شده‌اند. از این فصل، مقاله‌های **نوروزی‌راد و همکاران (۲۰۱۷b)**، **نوروزی‌راد و آرشی (۲۰۱۷b)** و **آرشی و نوروزی‌راد (۲۰۱۵)** نتیجه شده است.

۱.۲ مقدمه

خانواده برآوردگرهای M، اولین بار برای مدل‌های مکانی توسط **هوبر (۱۹۶۴، ۱۹۶۸)** پایه‌گذاری شدند. سپس **رلز (۱۹۶۸)** و **هوبر (۱۹۷۳)** به مدل‌های رگرسیونی تعمیم دادند. در برآورد OLS، ضرایب رگرسیونی β طوری برآورد می‌شوند که $\sum_{i=1}^n \varepsilon_i^2$ کمینه شود. برآوردها به روش کمترین قدرمطلق انحرافها (LAD) طوری انتخاب می‌شوند که $\sum_{i=1}^n |\varepsilon_i|$ کمینه گردد. مزیت اصلی برآوردهای LAD نسبت به برآوردهای OLS این است که نسبت به داده‌های پرت در متغیر پاسخ حساس نیستند. با وجود این، هنگامی که داده پرت وجود ندارد، برآوردهای OLS ممکن است دقیق‌تر باشند. می‌توان مزایای این دو روش را ترکیب کرده و تابع $\rho(\varepsilon)$ را طوری تعریف کرد که در نقاطی که مقدار ε تقریباً صفر است، برابر با ε^2 و نقاطی که ε خیلی از صفر فاصله دارد، برابر $|\varepsilon|$ (یا حداقل مشابه $|\varepsilon|$) باشد. در برآورد M، ضرایب رگرسیونی به‌گونه‌ای برآورد می‌شوند که $\sum_{i=1}^n \rho(\varepsilon_i)$ کمینه گردد. برآوردهای OLS و LAD، حالت خاصی از برآورد M به ترتیب با $\rho(\varepsilon) = \varepsilon^2$ و $\rho(\varepsilon) = |\varepsilon|$ هستند.

بر اساس **مارونا و همکاران (۲۰۰۶)** تابع حقیقی مقدار $\rho(\cdot)$ یک ρ -تابع نامیده می‌شود که در شرایط زیر صدق کند: (۱) $\rho(\circ) = \circ$ ، (۲) $\rho(-u) = \rho(u)$ ، (۳) به‌ازای $\circ \leq u \leq v$ ، $\rho(u) \leq \rho(v)$ ، (۴) یک تابع پیوسته است. (۵) اگر $\rho(u) \leq \sup_u \rho(u)$ و $\circ \leq u \leq v$ ، آن‌گاه $\rho(u) < \rho(v)$ اگر ρ کران‌دار باشد آن‌گاه $\sup_u \rho(u) = 1$.
برآوردگر $\hat{\beta}$ مقدار β ای است که عبارت زیر به‌ازای آن کمینه می‌شود:

$$\sum_{i=1}^n \rho(\varepsilon_i) = \sum_{i=1}^n \rho(y_i - \mathbf{x}_i^\top \beta)$$

به عبارت دیگر،

$$\psi(y_i - \mathbf{x}_i^\top \hat{\beta}) = \frac{\partial}{\partial \beta} [\rho(y_i - \mathbf{x}_i^\top \beta)] \Big|_{\beta=\hat{\beta}} = \circ.$$

از آن‌جایی که مولفه‌های بردار مشتقات جزئی به‌صورت $-\mathbf{x}_i \psi(y_i - \mathbf{x}_i \beta)$ نشان داده می‌شوند، می‌توان $\hat{\beta}$ را از حل معادله زیر به‌دست آورد:

$$\mathbf{M}_n(\mathbf{b}) = (M_1(\mathbf{b}), \dots, M_p(\mathbf{b}))^\top = \sum_{i=1}^n \mathbf{x}_i \psi(y_i - \mathbf{x}_i \mathbf{b}) = \circ, \quad \mathbf{b} \in \mathbb{R}^p$$

تابع ψ در شرایط زیر صدق می‌کند: (۱) تابع $\psi(\cdot)$ غیرنزولی، چوله-متقارن و پیوسته باشد. (۲) یک عدد ثابت مثبت متناهی d وجود دارد که به‌ازای $|u| \geq d$ ، $\psi(u) = \psi(d) \operatorname{sgn}(u)$ ، (۳) $\psi(\cdot)$ روی بازه $[-d, d]$ غیرثابت است.
برای اینکه برآوردگر M ، هم‌پایای مقیاس شود، باید آن را به کمک یک آماره مقیاس، استودنت‌سازی کرد. در واقع، به‌جای کمینه کردن $\sum_{i=1}^n \rho(y_i - \mathbf{x}_i^\top \beta)$ عبارت

$$\sum_{i=1}^n \rho\left(\frac{\varepsilon_i}{s_n}\right) = \sum_{i=1}^n \rho\left(\frac{y_i - \mathbf{x}_i^\top \beta}{s_n}\right)$$

را کمینه کرد که در آن $s_n = s_n(\mathbf{y})$. به‌عنوان مثال، s_n می‌تواند آماره‌ای به‌صورت

$$s_M = \frac{\text{MAD}}{\circ.۶۷۴۵} = \frac{\text{median}|\varepsilon - \text{median}(\varepsilon)|}{\circ.۶۷۴۵}$$

باشد، که در آن MAD، میانه قدرمطلق انحراف‌ها از میانه‌شان است. به‌طور مشابه می‌توان p معادله‌های غیرخطی زیر را حل کرد:

$$\mathbf{M}_n(\mathbf{b}) = (M_1(\mathbf{b}), \dots, M_p(\mathbf{b}))^\top = \sum_{i=1}^n \mathbf{x}_i \psi\left(\frac{y_i - \mathbf{x}_i^\top \mathbf{b}}{s_M}\right) = \circ. \quad (1.2)$$

بررسی دقیق‌تری از برآوردگرهای M را می‌توان در **هوبر (۱۹۸۱)** و **همپیل و همکاران (۱۹۸۶)** ملاحظه نمود. برخی از محققین اشاره کرده‌اند که برآوردگر M در صورتی که داده‌های پرت به‌خصوص در راستای متغیر پاسخ وجود داشته باشد، جایگزین بهتری برای برآوردگر OLS است (**برایکس و داگ، ۱۹۹۳**) و برآوردگر رنج معمولی در صورت وجود هم‌خطی در بین متغیرها بهتر

رفتار می کند (هورل و کنارد، ۱۹۷۰؛ هورل و همکاران، ۱۹۷۵). با این وجود، این روش‌ها در صورتی که در داده‌ها همزمان هم خطی و داده پرت وجود داشته باشد، منجر به نتایج نادرستی می‌شوند (جادهاو و کشید، ۲۰۱۱).

وجود هم خطی در متغیرها و غیرنرمال بودن توزیع خطاها باعث ظهور داده پرت می‌شود.

آسکین و مونت گمری (۱۹۸۰)؛ مونت گمری و آسکین (۱۹۸۱)؛ لارنس و مارش (۱۹۸۴)؛ فافنبرگر و دایلمن (۱۹۸۵)؛ سیلواپول (۱۹۹۱)؛ جادهاو و کشید (۲۰۱۴) برآوردگرهای ریج استوار را پیشنهاد دادند که نسبت به داده‌های پرت حساسیت کمتری نشان می‌دهند.

۲.۲ نوآوری مسئله

فرض کنید در مدل رگرسیونی، UPI یا NSI نیز درباره پارامتر β ، وجود داشته باشد و علاقه‌مندیم به وسیله آن‌ها برآورد را بهبود ببخشیم. این اطلاعات (ادعای یک متخصص) در مدل رگرسیونی (۱.۱) می‌تواند به صورت محدودیت

$$\mathcal{H}_o : H\beta = h \quad (2.2)$$

روی β بیان شود که در آن H ماتریس معلوم $q \times p$ با رتبه کامل سطری q و h بردار $q \times 1$ از ثابت‌هاست.

به عنوان یک حالت خاص از محدودیت (۲.۲)، فرض کنید $H = I_p$ و h بردار p تایی از صفرها باشد، در این صورت، محدودیت

$$\mathcal{H}_o : \beta = o. \quad (3.2)$$

به دست می‌آید.

مدل رگرسیونی (۱.۱) را مدل کامل می‌نامند. در ادامه، برآوردگرهای بهبودیافته در حالتی که در داده‌ها هم خطی وجود دارد، توزیع خطاها غیرنرمال است و اطلاعات غیرنمونه‌ای نیز در دسترس باشند، معرفی شده‌اند و کارایی آنها مورد بررسی قرار گرفته است.

۳.۲ مدل کامل

از آن جا که برآوردگر M ، به UPI بستگی ندارد، آن را برآوردگر محدود نشده M (UME) از پارامترهای β نامیده، جواب معادله زیر است.

$$M_n(\mathbf{b}) = (M_1(\mathbf{b}), \dots, M_p(\mathbf{b}))^\top = \sum_{i=1}^n \mathbf{x}_i \psi \left(\frac{y_i - \mathbf{x}_i^\top \mathbf{b}}{s_M} \right) = o, \quad (4.2)$$

تحت محدودیت (۳.۲)، برآوردگر محدود شده M (RME) نامیده و با نماد $\hat{\beta}^{rM}$ نشان داده می‌شود. بدیهی است که $\hat{\beta}^{rM} = o$.

به منظور تعریف برآوردگرهای انقباضی M، شرایط نظم زیر را در نظر بگیرید
(۱) اگر $C_n = n^{-1} X_n^T X_n$ ، آن گاه $\lim_{n \rightarrow \infty} C_n = C$ که C ماتریس معین مثبت است.

$$\max_{1 \leq i \leq n} \{x_i^T C_n^{-1} x_i\} = o(n) \quad (2)$$

صالح و شیرایشی (۱۹۸۹) نشان دادند وقتی $n \rightarrow \infty$ ،

$$\sqrt{n}(\hat{\beta}^{uM} - \beta) \xrightarrow{D} \mathcal{N}_p(o, \eta^2 C^{-1}), \quad (5.2)$$

که در آن $\eta^2 = \gamma^{-2} \sigma_\psi^2$ و

$$\gamma = \int_{-\infty}^{+\infty} \psi(x) \left\{ -\frac{f'(x)}{f(x)} \right\} dF(x)$$

$$\sigma_\psi^2 = \int_{-\infty}^{+\infty} \psi^2(x) dF(x).$$

به منظور پیشنهاد برآوردگر آزمون اولیه M (PTME)، برای فرضیه H_0 به آماره آزمون

$$\mathcal{L}_n^M = \frac{M_n^T(o) C_n^{-1} M_n(o)}{A_n^2}, \quad (6.2)$$

نیاز است که در آن $A_n^2 = \sum_{i=1}^n \psi^2(y_i - x_i \hat{\beta}^{rM})/n$. آماره آزمون \mathcal{L}_n^M به طور مجانبی دارای توزیع χ^2 با p درجه آزادی است (صالح و شیرایشی، ۱۹۸۹).

برآوردگر آزمون اولیه M با در نظر گرفتن یکی از $\hat{\beta}^{uM}$ یا $\hat{\beta}^{rM}$ بر حسب پذیرش یا رد H_0 براساس \mathcal{L}_n^M تعریف می شود. آن را با نماد $\hat{\beta}^{ptM}$ نشان می دهند و به صورت

$$\hat{\beta}^{ptM} = \hat{\beta}^{uM} - \hat{\beta}^{uM} I(\mathcal{L}_n^M \leq \chi_{p,\alpha}^2) = I(\mathcal{L}_n^M > \chi_{p,\alpha}^2) \hat{\beta}^{uM}. \quad (7.2)$$

تعریف می شود. برآوردگر انقباضی نوع استاین M (SME) پارامتر β به صورت

$$\hat{\beta}^{sM} = \hat{\beta}^{uM} - c[\mathcal{L}_n^M]^{-1} \hat{\beta}^{uM} = (1 - c[\mathcal{L}_n^M]^{-1}) \hat{\beta}^{uM} \quad (8.2)$$

تعریف می شود که c ثابت انقباضی است و معمولاً برابر با $p - 2$ در نظر گرفته می شود. برآوردگر انقباضی نوع استاین M در رابطه (۸.۲) با جزء مثبت آن بهبود می یابد. برآوردگر جزء مثبت انقباضی نوع استاین M (PRSME) به صورت

$$\hat{\beta}^{sM+} = \hat{\beta}^{sM} - \hat{\beta}^{sM} (1 - c[\mathcal{L}_n^M]^{-1}) I(\mathcal{L}_n^M \leq c)$$

$$= (1 - c[\mathcal{L}_n^M]^{-1}) I(\mathcal{L}_n^M > c) \hat{\beta}^{sM}. \quad (9.2)$$

تعریف می شود.

برآوردگرهای بهبودیافته M در رابطه های (۷.۲) - (۹.۲) زمانی به کار می روند که در داده ها هم خطی وجود ندارد. در حالی که اگر هم خطی وجود داشته باشد، باید این برآوردگرها را تصحیح کرد. استراتژی تصحیح پیشنهادی این است که برآوردگر محدود نشده/ شده M را می توان با برآوردگر ریج M جایگذاری نمود.

در ابتدا ماتریس ريج $R_n(\lambda)$ به صورت $(I_p + \lambda(n^{-1}C_n)^{-1})$ تعريف می‌شود که

$$\lim_{n \rightarrow \infty} R_n(\lambda) = \lim_{n \rightarrow \infty} \left(I_p + \lambda(n^{-1}C_n)^{-1} \right)^{-1} = \left(I_p + \lambda C^{-1} \right)^{-1} = R(\lambda) \quad (10.2)$$

برآوردگر محدود نشده M ريج (URME) پارامتر β به صورت

$$\hat{\beta}^{uM}(\lambda) = R_n(\lambda) \hat{\beta}^{uM}. \quad (11.2)$$

تعريف می‌شود. برآوردگر M ريج محدود شده (RRME) برابر با $\hat{\beta}^{rM}(\lambda) = o$ است. با توجه به استراتژی تصحيح فوق، برآوردگر آزمون اوليه M ريج (PTRME)، انقباضی نوع استاین M ريج (SRME) و جزء مثبت انقباضی نوع استاین M ريج (PRSRME) به ترتیب به صورت‌های

$$\hat{\beta}^{ptM}(\lambda) = R_n(\lambda) \hat{\beta}^{ptM} = I(\mathcal{L}_n^M > \chi_{p,\alpha}^2) \hat{\beta}^{uM}(\lambda), \quad (12.2)$$

$$\hat{\beta}^{sM}(\lambda) = R_n(\lambda) \hat{\beta}^{sM} = (1 - c[\mathcal{L}_n^M]^{-1}) \hat{\beta}^{uM}(\lambda), \quad (13.2)$$

$$\hat{\beta}^{sM+}(\lambda) = R_n(\lambda) \hat{\beta}^{sM+} = (1 - c[\mathcal{L}_n^M]^{-1}) I(\mathcal{L}_n^M > c) \hat{\beta}^{uM}(\lambda). \quad (14.2)$$

تعريف می‌شوند.

۱.۳.۲ ویژگی برآوردگرها

فرضیه مقابل \mathcal{H}_0 می‌تواند به صورت $H\beta = h + \delta$ (فرضیه مقابل ثابت) در نظر گرفته شود که در آن $\delta = (\delta_1, \dots, \delta_q)^T$ یک بردار ثابت است. تحت برقراری این فرضیه، اگر $\hat{\beta}^*$ یک برآوردگر دلخواه از بین برآوردگرهای پیشنهادی باشد، توزیع مجانبی $\sqrt{n}(\hat{\beta}^* - \beta)$ هنگامی که $n \rightarrow \infty$ ، با توزیع $\sqrt{n}(\hat{\beta}^{Ols} - \beta)$ برابر خواهد بود (صالح، ۲۰۰۶). بنابراین، در این حالت هنگامی که حجم نمونه زیاد است مقایسه برآوردگرها بی‌معنی خواهد بود. این ایده به ذهن می‌رسد که در حالت مجانبی، نباید یک فرضیه مقابل ثابت در نظر گرفت.

تحت فرضیه مقابل ثابت، توزیع‌های مجانبی برآوردگرهای بهبودیافته معادل با برآوردگر محدود نشده ريج M است (صالح، ۲۰۰۶). به‌عنوان یک نتیجه، با در نظر گرفتن فرضیه مقابل ثابت، نمی‌توان تفاوتی بین ویژگی‌های آریبی و مخاطره سه برآوردگر قائل شد. بنابراین، مجموعه فرضیه‌های مقابل موضعی به صورت زیر در نظر گرفته می‌شوند:

$$\mathcal{K}_{(n)} : \beta = n^{-\frac{1}{2}} \delta, \quad \delta = (\delta_1, \dots, \delta_p)^T \neq o.$$

برای برآوردگر دلخواه $\hat{\beta}^*$ از β ، تابع توزیع تجمعی به صورت زیر تعريف می‌شود:

$$G^*(x) = \lim_{n \rightarrow \infty} P \left(\sqrt{n}(\hat{\beta}^* - \beta) \leq x | \mathcal{K}_{(n)} \right), \quad x \in \mathbb{R}^p, \quad (15.2)$$

که در آن فرض می‌شود G^* غیرتباهیده است. آریبی و توان دوم آریبی مجانبی را به صورت زیر محاسبه می‌کنیم

$$ADB(\hat{\beta}_n^*) = \lim_{n \rightarrow \infty} E \left[\sqrt{n}(\hat{\beta}_n^* - \beta) \right] = \int x dG^*(x), \quad (16.2)$$

$$ADQB(\hat{\beta}_n^*) = [ADB(\hat{\beta}_n^*)]^T ADB(\hat{\beta}_n^*). \quad (17.2)$$

بنابراین، با تابع زیان توان دوم خطای $(\hat{\beta}^* - \beta)^\top (\hat{\beta}^* - \beta)$ ، مخاطره جانبی $\hat{\beta}^*$ به صورت

$$\text{ADR}(\hat{\beta}_n^*) = \lim_{n \rightarrow \infty} E \left[n(\hat{\beta}_n^* - \beta)^\top (\hat{\beta}_n^* - \beta) \right] = \text{tr} \{ \Gamma^* \}, \quad (18.2)$$

تعریف می‌شود که در آن ماتریس Γ^* ماتریس واریانس-کوواریانس به صورت

$$\Gamma^* = \int \mathbf{x}_n \mathbf{x}_n^\top dG(\mathbf{x}_n) = \lim_{n \rightarrow \infty} E \left[n(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)^\top \right]$$

است.

برای سادگی در محاسبات، شکلی از برآوردگرها که همه برآوردگرهای پیشنهادی را در بر می‌گیرد به صورت

$$\hat{\beta}^{*M} = \hat{\beta}^{uM} - g(\mathcal{L}_n^M) \hat{\beta}^{uM}. \quad (19.2)$$

تعریف می‌شود.

(الف) اگر $g(\mathcal{L}_n^M) = 0$ ، آن‌گاه $\hat{\beta}^{*M} = \hat{\beta}^{uM}$ ، UME است.

(ب) اگر $g(\mathcal{L}_n^M) = 1$ ، آن‌گاه $\hat{\beta}^{*M} = \mathbf{o} = \hat{\beta}^{rM}$ ، RME است.

(ج) اگر $g(\mathcal{L}_n) = I(\mathcal{L}_n \leq \chi_{p,\alpha}^2)$ ، آن‌گاه $\hat{\beta}^{*M} = \hat{\beta}^{ptM}$ ، PTME است.

(د) اگر $g(\mathcal{L}_n) = c[\mathcal{L}_n^M]^{-1}$ ، آن‌گاه $\hat{\beta}^{*M} = \hat{\beta}^{sM}$ ، SME است.

(ه) اگر $g(\mathcal{L}_n) = c[\mathcal{L}_n^M]^{-1} + (1 - c[\mathcal{L}_n^M]^{-1})I(\mathcal{L}_n^M \leq c)$ ، آن‌گاه $\hat{\beta}^{*M} = \hat{\beta}^{sM+}$ ، PRSME است.

قضیه ۱.۳.۲. تحت برقراری شرایط نظم و کلاس فرضیه‌های مقابل موضعی $\mathcal{K}_{(n)}$ ، آریبی جانبی برآوردگرهای بهبودیافته M به صورت

$$\begin{aligned} \text{ADB}(\hat{\beta}^{uM}) &= \mathbf{o}, \\ \text{ADB}(\hat{\beta}^{ptM}) &= -\delta H_{p+\nu}(\chi_{p,\alpha}^2; \Delta^\nu), \quad \Delta^\nu = \eta^{-\nu} \delta^\top C \delta \\ \text{ADB}(\hat{\beta}^{sM}) &= -c \delta E \left[\chi_{p+\nu}^{-\nu}(\Delta^\nu) \right], \\ \text{ADB}(\hat{\beta}^{sM+}) &= -\delta \left\{ c E \left[\chi_{p+\nu}^{-\nu}(\Delta^\nu) \right] + E \left[(1 - c \chi_{p+\nu}^{-\nu}(\Delta^\nu)) I(\chi_{p+\nu}^2(\Delta^\nu) \leq c) \right] \right\}, \end{aligned}$$

است که در آن

$$\begin{aligned} E \left[\chi_{p+\nu}^{-\nu}(\Delta^\nu) \right] &= E_R \left[(p + \nu R)^{-1} \right] \\ E \left[\chi_{p+\nu}^{-\nu}(\Delta^\nu) I(\chi_{p+\nu}^2(\Delta^\nu) \leq c) \right] &= E_R \left[(p + \nu R)^{-1} H_{p+\nu R}(c; \circ) \right] \\ E \left[\chi_{p+\nu}^{-\nu}(\Delta^\nu) I(\chi_{p+\nu}^2(\Delta^\nu) \leq c) \right] &= E_R \left[(p + \nu R)^{-1} (p + \nu R - \nu)^{-1} H_{p-\nu+\nu R}(c; \circ) \right] \end{aligned}$$

منظور از E_R ، امید ریاضی متغیر تصادفی پواسون R با میانگین $\Delta^2/2$ است.

برهان. طبق تعریف ADB، رابطه (۱۶.۲)، تحت $\mathcal{K}_{(n)}$ می توان نوشت

$$\begin{aligned} \text{ADB}(\hat{\beta}^{*M}) &= \lim_{n \rightarrow \infty} E \left[\sqrt{n} \left(\hat{\beta}^{*M} - \beta \right) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\sqrt{n} \left(\hat{\beta}^{\text{uM}} - g(\mathcal{L}_n^M) \hat{\beta}^{\text{uM}} - \beta \right) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\sqrt{n} \left(\hat{\beta}^{\text{uM}} - \beta \right) \right] - \lim_{n \rightarrow \infty} E \left[\sqrt{n} g(\mathcal{L}_n^M) \hat{\beta}^{\text{uM}} \right] \\ &= \text{ADB}(\hat{\beta}^{\text{uM}}) - \lim_{n \rightarrow \infty} E \left[\sqrt{n} g(\mathcal{L}_n^M) \hat{\beta}^{\text{uM}} \right] \\ &= o - \delta E \left[g(\chi_{p+r}^{\vee}(\Delta^{\vee})) \right]. \end{aligned} \quad (۲۰.۲)$$

خط آخر از رابطه (۵.۲) و قضیه ۵.۲.آ نتیجه شده است. با جانشین کردن هر یک از توابع g متناظر با هر برآوردگر در رابطه (۲۰.۲) اثبات کامل می شود. \square

نتیجه ۱.۳.۲. تحت برقراری شرایط نظم و کلاس فرضیه های مقابل موضعی $\mathcal{K}_{(n)}$ ، توان دوم اریبی مجانبی برآوردگرهای بهبودیافته M به راحتی از رابطه (۱۷.۲) به صورت

$$\begin{aligned} \text{ADQB}(\hat{\beta}^{\text{uM}}) &= o, \\ \text{ADQB}(\hat{\beta}^{\text{ptM}}) &= \Delta^{\vee} \{ H_{p+r}(\chi_{p,\alpha}^{\vee}; \Delta^{\vee}) \}^{\vee}, \\ \text{ADQB}(\hat{\beta}^{\text{sM}}) &= c^{\vee} \Delta^{\vee} \left\{ E \left[\chi_{p+r}^{-\vee}(\Delta^{\vee}) \right] \right\}^{\vee}, \\ \text{ADQB}(\hat{\beta}^{\text{sM}+}) &= \Delta^{\vee} \left\{ c E \left[\chi_{p+r}^{-\vee}(\Delta^{\vee}) \right] - E \left[(1 - c \chi_{p+r}^{-\vee}(\Delta^{\vee})) I(\chi_{p+r}^{\vee}(\Delta^{\vee}) \leq c) \right] \right\}^{\vee}, \end{aligned}$$

به دست می آید.

قضیه ۲.۳.۲. تحت برقراری شرایط نظم و کلاس فرضیه های موضعی $\mathcal{K}_{(n)}$ ، مخاطره مجانبی برآوردگرهای بهبودیافته M به صورت

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{uM}}) &= \eta^{\vee} \text{tr}(\mathbf{C}^{-1}), \\ \text{ADR}(\hat{\beta}^{\text{ptM}}) &= \eta^{\vee} \text{tr}(\mathbf{C}^{-1}) (1 - H_{p+r}(\chi_{p,\alpha}^{\vee}; \Delta^{\vee})) + \delta^{\top} \delta Z(\alpha; \Delta^{\vee}), \\ \text{ADR}(\hat{\beta}^{\text{sM}}) &= \eta^{\vee} \text{tr}(\mathbf{C}^{-1}) - c \eta^{\vee} \text{tr}(\mathbf{C}^{-1}) X(\Delta^{\vee}) + c(c + \mathcal{F}) \delta^{\top} \delta E \left[\chi_{p+r}^{-\mathcal{F}}(\Delta^{\vee}) \right] \\ \text{ADR}(\hat{\beta}^{\text{sM}+}) &= \text{ADR}(\hat{\beta}^{\text{sM}}) + \delta^{\top} \delta Q(\Delta^{\vee}) \\ &\quad - \eta^{\vee} \text{tr}(\mathbf{C}^{-1}) E \left[(1 - c \chi_{p+r}^{-\vee}(\Delta^{\vee}))^{\vee} I(\chi_{p+r}^{\vee}(\Delta^{\vee}) \leq c) \right], \end{aligned}$$

است که در آن

$$\begin{aligned} Z(\alpha; \Delta^{\vee}) &= \vee H_{p+r}(\chi_{p,\alpha}^{\vee}; \Delta^{\vee}) - H_{p+r}(\chi_{p,\alpha}^{\vee}(\alpha); \Delta^{\vee}), \\ X(\Delta^{\vee}) &= \vee E \left[\chi_{p+r}^{-\vee}(\Delta^{\vee}) \right] - c E \left[\chi_{p+r}^{-\mathcal{F}}(\Delta^{\vee}) \right], \\ Q(\Delta^{\vee}) &= \vee E \left[(1 - c \chi_{p+r}^{-\vee}(\Delta^{\vee})) I(\chi_{p+r}^{\vee}(\Delta^{\vee}) \leq c) \right] \\ &\quad - E \left[(1 - c \chi_{p+r}^{-\mathcal{F}}(\Delta^{\vee}))^{\vee} I(\chi_{p+r}^{\vee}(\Delta^{\vee}) \leq c) \right] \end{aligned}$$

$$E \left[\chi_{p+r}^{-\mathcal{F}}(\Delta^{\vee}) \right] = E_R \left[[(p + \vee R)(p + \vee r - \vee)]^{-1} \right]$$

$$\begin{aligned} E \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right] &= E_R \left[(p + \nu R - \nu)^{-1} \right] \\ E \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right] &= E_R \left[(p + \nu R)(p + \nu R + \nu) \right]^{-1} \\ E \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] &= E_R \left[(p + \nu R - \nu)^{-1} H_{p+\nu R}(c; \circ) \right] \\ E \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] &= E_R \left[(p + \nu R)^{-1} (p + \nu R + \nu)^{-1} H_{p-\nu+\nu R}(c; \circ) \right] \end{aligned}$$

برهان. طبق تعریف ADR، رابطه (۱۸.۲)، می توان نوشت

$$\begin{aligned} \text{ADR}(\hat{\beta}^{*M}) &= \lim_{n \rightarrow \infty} E \left[n \left(\hat{\beta}^{*M} - \beta \right)^{\top} \left(\hat{\beta}^{*M} - \beta \right) \right] \\ &= \lim_{n \rightarrow \infty} E \left[n \left(\hat{\beta}^{\text{uM}} - g(\mathcal{L}_n^M) \hat{\beta}^{\text{uM}} - \beta \right)^{\top} \left(\hat{\beta}^{\text{uM}} - g(\mathcal{L}_n^M) \hat{\beta}^{\text{uM}} - \beta \right) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\left(n \hat{\beta}^{\text{uM}} - \beta \right)^{\top} \left(\hat{\beta}^{\text{uM}} - \beta \right) \right] \\ &\quad - \nu \lim_{n \rightarrow \infty} E \left[n \left(\hat{\beta}^{\text{uM}} - \beta \right)^{\top} \hat{\beta}^{\text{uM}} g(\mathcal{L}_n^M) \right] \\ &\quad + \lim_{n \rightarrow \infty} E \left[n [\hat{\beta}^{\text{uM}}]^{\top} \hat{\beta}^{\text{uM}} g([\mathcal{L}_n^M]^{\nu}) \right] \\ &= \text{ADR}(\hat{\beta}_n^{\text{uM}}) - \nu \lim_{n \rightarrow \infty} E \left[n \left(\hat{\beta}^{\text{uM}} \right)^{\top} \hat{\beta}^{\text{uM}} g(\mathcal{L}_n^M) \right] \\ &\quad + \nu \delta^{\top} \lim_{n \rightarrow \infty} E \left[\sqrt{n} \hat{\beta}^{\text{uM}} g(\mathcal{L}_n^M) \right] + \lim_{n \rightarrow \infty} E \left[n [\hat{\beta}^{\text{uM}}]^{\top} \hat{\beta}^{\text{uM}} g([\mathcal{L}_n^M]^{\nu}) \right]. \quad (21.2) \end{aligned}$$

رابطه های (۱۸.۲) و (۵.۲) منجر به رابطه زیر می شوند

$$\text{ADR}(\hat{\beta}^{\text{uM}}) = \eta^{\nu} \text{tr}(C^{-1}) \quad (22.2)$$

طبق قضیه های ۵.۲.آ و ۶.۲.آ و رابطه (۲۲.۲)، رابطه (۲۱.۲) پس از انجام محاسبات برابر خواهد بود با

$$\begin{aligned} \text{ADR}(\hat{\beta}^{*M}) &= \eta^{\nu} \text{tr}(C^{-1}) \left\{ 1 - \nu E \left[g(\chi_{p+\nu}^{\nu}(\Delta^{\nu})) \right] + E \left[g(\chi_{p+\nu}^{\nu}(\Delta^{\nu})) \right] \right\} \\ &\quad + \delta^{\top} \delta \left\{ \nu E \left[\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \right] - \nu E \left[\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \right] + E \left[g(\chi_{p+\nu}^{\nu}(\Delta^{\nu})) \right] \right\}. \quad (23.2) \end{aligned}$$

با جانشین کردن هر یک از توابع g متناظر با هر برآوردگر در رابطه های (۲۳.۲) اثبات کامل می شود. □

نتایج جانبی برای برآوردگرهای بهبودیافته M ریج در قضیه های ۳.۳.۲ و ۴.۳.۲ ارائه شده است.

قضیه ۳.۳.۲. تحت شرایط نظم و کلاس فرضیه های مقابل موضعی $\mathcal{K}(n)$ ، اربیی جانبی برآوردگرهای SRM، PTRM، RRM و PRSRM به ترتیب برابر است با

$$\begin{aligned} \text{ADB}(\hat{\beta}^{\text{uM}}(\lambda)) &= -C^{-1}(\lambda) \delta \quad C(\lambda) = C + \lambda I_p, \\ \text{ADB}(\hat{\beta}^{\text{ptM}}(\lambda)) &= -\lambda C^{-1}(\lambda) \delta - R(\lambda) \delta H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}), \\ \text{ADB}(\hat{\beta}^{\text{SM}}(\lambda)) &= -\lambda C^{-1}(\lambda) \delta - c R(\lambda) \delta E \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right], \end{aligned}$$

$$\begin{aligned} \text{ADB}(\hat{\beta}^{\text{sM}+}(\lambda)) &= -\lambda C^{-1}(\lambda)\delta - \mathbf{R}(\lambda)\delta \left\{ cE \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right] \right. \\ &\quad \left. + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu})) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) < c) \right] \right\} \end{aligned}$$

برهان. ابتدا، باید رابطه زیر اثبات شود

$$\lambda C^{-1}(\lambda) = \mathbf{I}_p - \mathbf{R}(\lambda). \quad (24.2)$$

به ازای هر $\lambda > 0$ ، $\mathbf{R}(\lambda) = (\mathbf{I}_p + \lambda C^{-1})^{-1}$ ، بنابراین $\mathbf{R}^{-1}(\lambda) = \mathbf{I}_p + \lambda C^{-1}$. در نتیجه

$$\begin{aligned} \lambda C^{-1} = \mathbf{R}^{-1}(\lambda) - \mathbf{I}_p &\implies \lambda \mathbf{R}(\lambda) C^{-1} = \mathbf{I}_p - \mathbf{R}(\lambda) \\ &\implies \lambda (\mathbf{I}_p + \lambda C^{-1})^{-1} C^{-1} = \mathbf{I}_p - \mathbf{R}(\lambda) \\ &\implies \lambda C (\mathbf{I}_p + \lambda C^{-1})^{-1} = \mathbf{I}_p - \mathbf{R}(\lambda) \\ &\implies \lambda (C + \lambda \mathbf{I}_p)^{-1} = \mathbf{I}_p - \mathbf{R}(\lambda) \\ &\implies -\lambda C^{-1}(\lambda) = \mathbf{R}(\lambda) - \mathbf{I}_p. \end{aligned}$$

اکنون، فرض کنید $\hat{\beta}^{*M}(\lambda)$ برآوردگر بهبودیافته M ریح دلخواه باشد. طبق تعریف (16.2)، ADB هر برآوردگر برابر است با

$$\begin{aligned} \text{ADB}(\hat{\beta}^{*M}(\lambda)) &= \lim_{n \rightarrow \infty} E \left[\sqrt{n}(\hat{\beta}^{*M}(\lambda) - \beta) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\sqrt{n}(\mathbf{R}_n(\lambda)\hat{\beta}^{*M} - \beta) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\sqrt{n}(\mathbf{R}_n(\lambda)\hat{\beta}^{*M} - \mathbf{R}_n(\lambda)\beta + \mathbf{R}_n(\lambda)\beta - \beta) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\mathbf{R}_n(\lambda)\sqrt{n}(\hat{\beta}^{*M} - \beta) \right] + \lim_{n \rightarrow \infty} E \left[(\mathbf{R}_n(\lambda) - \mathbf{I}_p)\sqrt{n}\beta \right] \\ &= \mathbf{R}(\lambda)\text{ADB}(\hat{\beta}^{*M}) - \lambda C^{-1}(\lambda)\delta \end{aligned} \quad (25.2)$$

خط آخر از قضیه 3.3.1 و رابطه (24.2) نتیجه شده است. از جایگذاری ADB های بیان شده در قضیه 1.3.2 در رابطه (25.2) اثبات کامل می شود. □

نتیجه 2.3.2. تحت شرایط نظم و کلاس فرضیه های مقابل موضعی $\mathcal{K}(n)$ ، توان دوم مجانبی برآوردگرهای URM، PTRM، SRM و PRSRM به ترتیب برابر است با

$$\begin{aligned} \text{ADQB}(\hat{\beta}^{\text{RM}}(\lambda)) &= \lambda^{\nu} \delta^{\top} C^{-\nu}(\lambda)\delta, \\ \text{ADQB}(\hat{\beta}^{\text{ptM}}(\lambda)) &= \lambda^{\nu} \delta^{\top} C^{-\nu}(\lambda)\delta + \nu \lambda \delta^{\top} C^{-1}(\lambda)\mathbf{R}(\lambda)\delta H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) \\ &\quad + \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda)\delta \left\{ H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) \right\}^{\nu}, \\ \text{ADQB}(\hat{\beta}^{\text{sM}}(\lambda)) &= \lambda^{\nu} \delta^{\top} C^{-\nu}(\lambda)\delta + c^{\nu} \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda)\delta \left\{ E \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right] \right\}^{\nu} \\ &\quad + \nu c \lambda \delta^{\top} C^{-1}(\lambda)\mathbf{R}(\lambda)\delta E \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right], \\ \text{ADQB}(\hat{\beta}^{\text{sM}+}(\lambda)) &= \lambda^{\nu} \delta^{\top} C^{-\nu}(\lambda)\delta + \nu \lambda \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda)\delta F(\Delta^{\nu}) \\ &\quad + \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda)\delta \{ F(\Delta^{\nu}) \}^{\nu} \end{aligned}$$

که در آن

$$F(\Delta^{\nu}) = cE \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right] + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu})) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right].$$

قضیه ۴.۳.۲. تحت شرایط نظم بیان شده و کلاس فرضیه‌های موضعی $\mathcal{K}_{(n)}$ ، عبارت‌های مخاطره جانبی برآوردگرهای URM، PTRM، SRM و PRSRM به ترتیب برابر است با

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{uM}}(\lambda)) &= \eta^\gamma \text{tr} \left([\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) + \lambda^\gamma \delta^\top \mathbf{C}^{-\gamma}(\lambda) \delta \\ \text{ADR}(\hat{\beta}^{\text{ptM}}(\lambda)) &= \eta^\gamma \text{tr} \left([\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) (1 - H_{p+\gamma}(\chi_{p,\alpha}^\gamma; \Delta^\gamma)) \\ &\quad + \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{R}(\lambda) \delta Z(\alpha, \Delta^\gamma) + \lambda^\gamma \delta^\top \mathbf{C}^{-\gamma}(\lambda) \delta \\ &\quad + \lambda \left[\delta^\top \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta + \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1}(\lambda) \delta \right] H_{p+\gamma}(\chi_{p,\alpha}^\gamma; \Delta^\gamma) \\ \text{ADR}(\hat{\beta}^{\text{sM}}(\lambda)) &= \eta^\gamma \text{tr} \left([\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) (1 - cX(\Delta^\gamma)) \\ &\quad + c(c + \mathbf{f}) \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{R}(\lambda) \delta E \left[\chi_{p+\mathbf{f}}^{-\mathbf{f}}(\Delta^\gamma) \right] \\ &\quad + \lambda^\gamma \delta^\top \mathbf{C}^{-1}(\lambda) \mathbf{C}^{-1}(\lambda) \delta + c\lambda \left[\delta^\top \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta \right. \\ &\quad \left. + \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1}(\lambda) \delta \right] E \left[\chi_{p+\gamma}^{-\gamma}(\Delta^\gamma) \right] \\ \text{ADR}(\hat{\beta}^{\text{sM}+}(\lambda)) &= \text{ADR}(\hat{\beta}^{\text{sM}}(\lambda)) \\ &\quad - \eta^\gamma \text{tr} \left([\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) E \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\Delta^\gamma))^\gamma I(\chi_{p+\gamma}^\gamma(\Delta^\gamma) \leq c) \right] \\ &\quad + \lambda \{ \delta^\top \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta + \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1}(\lambda) \delta \} \\ &\quad \times E \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\Delta^\gamma)) I(\chi_{p+\gamma}^\gamma(\Delta^\gamma) < c) \right] + \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{R}(\lambda) \delta Q(\Delta^\gamma) \end{aligned}$$

برهان. فرض کنید $\hat{\beta}^{\text{sM}}(\lambda)$ برآوردگر بهبودیافته M ریج دلخواه باشد. ADR آن برابر است با

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{sM}}) &= \lim_{n \rightarrow \infty} E \left[n(\hat{\beta}^{\text{sM}}(\lambda) - \beta)^\top (\hat{\beta}^{\text{sM}}(\lambda) - \beta) \right] \\ &= \lim_{n \rightarrow \infty} E \left[n(\mathbf{R}_n(\lambda) \hat{\beta}^{\text{sM}} - \beta)^\top (\mathbf{R}_n(\lambda) \hat{\beta}^{\text{sM}} - \beta) \right] \\ &= \lim_{n \rightarrow \infty} E \left[n(\mathbf{R}_n(\lambda) \hat{\beta}^{\text{sM}} - \mathbf{R}_n(\lambda) \beta + \mathbf{R}_n(\lambda) \beta - \beta)^\top \right. \\ &\quad \left. \times (\mathbf{R}_n(\lambda) \hat{\beta}^{\text{sM}} - \mathbf{R}_n(\lambda) \beta + \mathbf{R}_n(\lambda) \beta - \beta) \right] \\ &= \lim_{n \rightarrow \infty} E \left[n(\hat{\beta}^{\text{sM}} - \beta)^\top [\mathbf{R}_n(\lambda)]^\top \mathbf{R}_n(\lambda) (\hat{\beta}^{\text{sM}} - \beta) \right] \\ &\quad + \lim_{n \rightarrow \infty} E \left[n\beta^\top (\mathbf{R}_n(\lambda) - \mathbf{I}_p)^\top \mathbf{R}_n(\lambda) (\hat{\beta}^{\text{sM}} - \beta) \right] \\ &\quad + \lim_{n \rightarrow \infty} E \left[n(\hat{\beta}^{\text{sM}} - \beta)^\top [\mathbf{R}_n(\lambda)]^\top (\mathbf{R}_n(\lambda) - \mathbf{I}_p) \beta \right] \\ &\quad + \lim_{n \rightarrow \infty} E \left[n\beta^\top (\mathbf{R}_n(\lambda) - \mathbf{I}_p)^\top (\mathbf{R}_n(\lambda) - \mathbf{I}_p) \beta \right]. \end{aligned}$$

بنا به تعریف ADR و رابطه (۲۴.۲)، نتیجه می‌شود

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{sM}}) &= \text{ADR}(\hat{\beta}^{\text{sM}}) - \lambda \left[\delta^\top \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \text{ADB}(\hat{\beta}^{\text{sM}}) \right. \\ &\quad \left. + [\text{ADB}(\hat{\beta}^{\text{sM}})]^\top [\mathbf{R}_n(\lambda)]^\top \mathbf{C}^{-1}(\lambda) \delta \right] + \lambda^\gamma \delta^\top \mathbf{C}^{-\gamma}(\lambda) \delta \quad (26.2) \end{aligned}$$

□ با جایگذاری ADB‌های بیان شده در قضیه ۱.۳.۲ در عبارت بالا، اثبات کامل می‌شود.

۲.۳.۲ کارایی برآوردگرها: تجزیه و تحلیل مخاطره جانبی

در این بخش، کارایی برآوردگرهای پیشنهادی بر اساس مخاطره جانبی مقایسه شده است. از آنجایی که ماتریس C نیمه‌معین مثبت است، بر اساس تجزیه طیفی ماتریس متعامدی مانند

ماتریس Γ وجود دارد به طوری که

$$\Gamma^T C \Gamma = \Omega = \text{diag}(\omega_1, \dots, \omega_p),$$

که $\omega_1 \geq \dots \geq \omega_p > 0$ مقادیر ویژه ماتریس C هستند.

بدیهی است مقادیر ویژه $R(\lambda) = (I_p + \lambda C^{-1})^{-1}$ و $C(\lambda) = (C + \lambda I_p)$ به ترتیب $(\frac{\omega_1}{\omega_1 + \lambda}, \dots, \frac{\omega_p}{\omega_p + \lambda})$ و $(\omega_1 + \lambda, \dots, \omega_p + \lambda)$ می باشند. بنابراین، تساوی های

$$\begin{aligned} \text{tr} \left([R(\lambda)]^T C^{-1} R(\lambda) \right) &= \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^2}, \\ \delta^T C^{-2}(\lambda) \delta &= \sum_{i=1}^p \frac{\theta_i^2}{(\omega_i + \lambda)^2}; \quad \theta = \Gamma^T \delta, \\ \delta^T [R(\lambda)]^T R(\lambda) \delta &= \sum_{i=1}^p \frac{\omega_i^2 \theta_i^2}{(\omega_i + \lambda)^2}, \\ \delta^T [R(\lambda)]^T C^{-1}(\lambda) \delta &= \sum_{i=1}^p \frac{\omega_i \theta_i^2}{(\omega_i + \lambda)^2}. \end{aligned}$$

برقرار هستند.

قضیه ۵.۳.۲ (مقایسه بین URME و UME). با در نظر گرفتن فرضیه مقابل موضعی، شرط کافی برای برتری $\hat{\beta}^{\text{UM}}(\lambda)$ بر $\hat{\beta}^{\text{UM}}$ این است که زوج مرتب (Δ^2, λ) وجود داشته باشد به طوری که $(\Delta^2, \lambda) \in (0, \Delta^2(\lambda)] \times (0, \lambda_1)$ که در آن

$$\Delta^2(\lambda) = \frac{\text{tr}(C^{-1} - [R(\lambda)]^T C^{-1} R(\lambda))}{\lambda^2 \text{Ch}_{\min}(C^{-2}(\lambda) C^{-1})} \quad \text{و} \quad \lambda_1 = \frac{\eta^2}{\max_{1 \leq i \leq p} \theta_i^2} = \frac{\eta^2}{\theta_{\max}^2}$$

که در آن θ_{\max} بزرگترین مولفه θ است.

برهان. تابع مخاطره جانبی $\hat{\beta}^{\text{UM}}(\lambda)$ بر حسب مقادیر ویژه به صورت زیر است:

$$\text{ADR}(\hat{\beta}^{\text{UM}}(\lambda)) = \eta^2 \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^2} + \lambda^2 \sum_{i=1}^p \frac{\theta_i^2}{(\omega_i + \lambda)^2}. \quad (27.2)$$

اگر از رابطه (۲۷.۲) نسبت به λ مشتق گرفته شود، آن گاه

$$\frac{\partial \text{ADR}(\hat{\beta}^{\text{UM}}(\lambda))}{\partial(\lambda)} = 2 \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^3} (\lambda \theta_i^2 - \eta^2). \quad (28.2)$$

شرط کافی برای اینکه رابطه (۲۸.۲) منفی باشد، این است که $\lambda \in (0, \lambda_1)$ وجود داشته باشد به طوری که URME مخاطره کمتری از UME داشته باشد که λ_1 به صورت زیر تعریف می شود.

$$\lambda_1 = \frac{\eta^2}{\max_{1 \leq i \leq p} \theta_i^2} = \frac{\eta^2}{\theta_{\max}^2}.$$

همچنین اختلاف بین دو مخاطره برابر است با

$$\text{ADR}(\hat{\beta}^{\text{uM}}) - \text{ADR}(\hat{\beta}^{\text{uM}}(\lambda)) = \eta^{\nu} \text{tr}(C^{-1} - [R(\lambda)]^{\top} C^{-1} R(\lambda)) - \lambda^{\nu} \delta^{\top} C^{-\nu}(\lambda) \delta. \quad (29.2)$$

شرط کافی برای اینکه $\text{ADR}(\hat{\beta}^{\text{uM}}(\lambda))$ کوچکتر از $\text{ADR}(\hat{\beta}^{\text{uM}}) = \eta^{\nu} \text{tr}(C^{-1})$ باشد، این است که رابطه (29.2) نامنفی باشد. به عبارت دیگر،

$$\delta^{\top} C^{-\nu}(\lambda) \delta \leq \frac{\eta^{\nu} \text{tr}(C^{-1} - [R(\lambda)]^{\top} C^{-1} R(\lambda))}{\lambda^{\nu}}. \quad (30.2)$$

طبق نتیجه 2.1.آ،

$$\text{Ch}_{\min}(C^{-\nu}(\lambda)C^{-1}) \leq \frac{\delta^{\top} C^{-\nu}(\lambda) \delta^{\top}}{\delta^{\top} \delta^{\top}} = \frac{\delta^{\top} C^{\frac{1}{\nu}} C^{-\nu}(\lambda) C^{-1} C^{\frac{1}{\nu}} \delta^{\top}}{\delta^{\top} C^{\frac{1}{\nu}} C^{\frac{1}{\nu}} \delta^{\top}} \leq \text{Ch}_{\max}(C^{-\nu}(\lambda)C^{-1}),$$

و بنابراین

$$\frac{\delta^{\top} C^{-\nu}(\lambda) \delta^{\top}}{\text{Ch}_{\max}(C^{-\nu}(\lambda)C^{-1})} \leq \delta^{\top} C \delta^{\top} \leq \frac{\delta^{\top} C^{-\nu}(\lambda) \delta^{\top}}{\text{Ch}_{\min}(C^{-\nu}(\lambda)C^{-1})}. \quad (31.2)$$

از جایگذاری رابطه (30.2) در (31.2) نتیجه زیر به دست می آید:

$$\Delta^{\nu} \leq \frac{\text{tr}(C^{-1} - [R(\lambda)]^{\top} C^{-1} R(\lambda))}{\lambda^{\nu} \text{Ch}_{\min}(C^{-\nu}(\lambda)C^{-1})} = \Delta_1(\lambda).$$

□

قضیه 6.3.2 (مقایسه بین PTRME و PTME). (الف) تحت فرضیه صفر، به ازای هر λ و α ، $\hat{\beta}^{\text{ptM}}(\lambda)$ به طور یکنواخت بر $\hat{\beta}^{\text{ptM}}$ برتری دارد. (ب) تحت فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ ، یک شرط کافی برای اینکه $\hat{\beta}^{\text{ptM}}(\lambda)$ بر برآوردگر $\hat{\beta}^{\text{ptM}}$ برتری داشته باشد این است که زوج مرتب (Δ^{ν}, λ) وجود دارد به طوری که $(\Delta^{\nu}, \lambda) \in (\circ, \Delta_{\nu}^{\nu}(\lambda, \alpha)) \times (\circ, \lambda_{\nu}(\alpha, \Delta^{\nu}))$ که در آن،

$$\Delta_{\nu}^{\nu}(\lambda, \alpha) = \frac{\text{tr}(C^{-1} - [R(\lambda)]^{\top} C^{-1} R(\lambda))(1 - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}))}{\text{Ch}_{\max}(D_1(\alpha, \Delta^{\nu})C^{-1})},$$

که در آن

$$D_1(\alpha, \Delta^{\nu}) = \lambda^{\nu} C^{-\nu}(\lambda) + (\lambda)(C^{-1}(\lambda)R(\lambda) + [R(\lambda)]^{\top} C^{-1}(\lambda))H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) - (I_p - [R(\lambda)]^{\top} R(\lambda))Z(\alpha; \Delta^{\nu})$$

9

$$\lambda_{\nu}(\alpha, \Delta^{\nu}) = \frac{D_{\nu}}{\max_{1 \leq i \leq p} \left\{ \theta_i^{\nu} (1 - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu})) \right\}}$$

به طوری که

$$D_{\nu} = \min_{1 \leq i \leq p} \left\{ \eta^{\nu} (1 - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu})) + \omega_i \theta_i^{\nu} \left(H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) \right) \right\}$$

برهان. حالت ۱: فرضیه صفر \mathcal{H}_0 . اختلاف بین دو مخاطره در این مورد، برابر است با

$$\eta^{\gamma} \text{tr}(C^{-1} - \mathbf{R}^{\top}(\lambda)C^{-1}\mathbf{R}(\lambda))(1 - H_{p+\gamma}(\chi_p^{\gamma}(\alpha); \circ)) \geq \circ.$$

حالت ۲: فرضیه مقابل، $\mathcal{K}_{(n)}$. اختلاف بین دو مخاطره برابر است با

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{ptM}}) - \text{ADR}(\hat{\beta}^{\text{ptM}}(\lambda)) = & \\ & \eta^{\gamma} \text{tr}(C^{-1} - [\mathbf{R}(\lambda)]^{\top}C^{-1}\mathbf{R}(\lambda))(1 - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma})) \\ & - \lambda \left\{ \delta^{\top}C^{-1}(\lambda)\mathbf{R}(\lambda)\delta + \delta[\mathbf{R}(\lambda)]^{\top}C^{-1}(\lambda)\delta \right\} H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \\ & + \delta^{\top}(\mathbf{I}_p - [\mathbf{R}(\lambda)]^{\top}\mathbf{R}(\lambda))\delta Z(\alpha; \Delta^{\gamma}) - \lambda^{\gamma} \delta^{\top}C^{-\gamma}(\lambda)\delta. \end{aligned} \quad (32.2)$$

عبارت (۳۲.۲) نامنفی است اگر و فقط اگر

$$\begin{aligned} & \eta^{\gamma} \text{tr}(C^{-1} - [\mathbf{R}(\lambda)]^{\top}C^{-1}\mathbf{R}(\lambda))(1 - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma})) \geq \\ & \delta^{\top} \left[\lambda^{\gamma}C^{-\gamma}(\lambda) + \lambda \left(C^{-1}(\lambda)\mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^{\top}C^{-1}(\lambda) \right) H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \right. \\ & \left. - (\mathbf{I}_p - [\mathbf{R}(\lambda)]^{\top}\mathbf{R}(\lambda))Z(\alpha; \Delta^{\gamma}) \right] \delta, \end{aligned}$$

با استفاده از قضیه ۲.۱.۱ و کمی محاسبات، نتیجه به دست می آید.

اگر عبارت $\text{ADR}(\hat{\beta}^{\text{ptM}}(\lambda))$ بر حسب مقادیر ویژه و λ نوشته شود، آن گاه

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{ptM}}(\lambda)) = & \sum_{i=1}^p \frac{1}{(\omega_i + (\lambda))^{\gamma}} \left\{ \eta^{\gamma} \omega_i (1 - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma})) + \omega_i^{\gamma} \theta_i^{\gamma} Z(\alpha; \Delta^{\gamma}) \right. \\ & \left. + \lambda^{\gamma} \theta_i^{\gamma} + \gamma \lambda \omega_i \theta_i^{\gamma} H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \right\}. \end{aligned}$$

برای اثبات کفایت از مخاطره PTRME نسبت به λ مشتق گرفته و شرط کافی برای منفی بودن عبارت پیدا شود. \square

قضیه ۷.۳.۲ (مقایسه بین PTRME و URME). با در نظر گرفتن فرضیه مقابل موضعی، شرط کافی برای برتری $\hat{\beta}^{\text{ptM}}(\lambda)$ بر $\hat{\beta}^{\text{uM}}(\lambda)$ این است که زوج مرتب $(\Delta^{\gamma}, \lambda)$ وجود داشته باشد به طوری که $(\Delta^{\gamma}, \lambda) \in (\circ, \Delta_{\gamma}^{\gamma}(\lambda, \alpha)) \times (\circ, \lambda_{\gamma}(\alpha, \Delta^{\gamma}))$ که در آن

$$\Delta_{\gamma}^{\gamma}(\lambda, \alpha) = \frac{\text{tr}([\mathbf{R}(\lambda)]^{\top}C^{-1}\mathbf{R}(\lambda))H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma})}{\text{Ch}_{\min}(E_1(\alpha, \Delta^{\gamma})C^{-1})},$$

که در آن

$$E_1(\alpha, \Delta^{\gamma}) = [\mathbf{R}(\lambda)]^{\top}\mathbf{R}(\lambda)Z(\alpha; \Delta^{\gamma}) + \lambda(C^{-1}(\lambda)\mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^{\top}C^{-1}(\lambda))H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma})$$

و

$$\lambda_{\gamma}(\alpha, \Delta^{\gamma}) = \frac{\min_{1 \leq i \leq p} \left\{ \eta^{\gamma} H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) - \omega_i \theta_i^{\gamma} Z(\alpha; \Delta^{\gamma}) \right\}}{\max_{1 \leq i \leq p} \left\{ \gamma \theta_i^{\gamma} H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \right\}}$$

برهان. اختلاف بین مخاطره PTRME و URME برابر است با

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{rM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{ptM}}(\lambda)) &= \eta^{\gamma} \text{tr}([\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda)) H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \\ &\quad - \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda) \delta Z(\alpha; \Delta^{\gamma}) - \lambda (\delta^{\top} \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta \\ &\quad + \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda) \delta) H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}). \end{aligned} \quad (33.2)$$

عبارت نامنفی است وقتی که

$$\begin{aligned} \delta^{\top} \left[[\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda) Z(\alpha; \Delta^{\gamma}) + \lambda (\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda)) H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \right] \delta \\ \leq \eta^{\gamma} \text{tr}([\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda)) H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}). \end{aligned} \quad (34.2)$$

دوباره، با استفاده از نتیجه آ.۲.۱، رابطه (۳۳.۲) نامنفی خواهد بود اگر $\Delta^{\gamma} < \Delta_{\gamma}^{\lambda}(\alpha)$. بنابراین، وقتی $\Delta^{\gamma} \in (\circ, \Delta_{\gamma}^{\lambda}(\alpha))$ ، $\hat{\beta}^{\text{ptM}}(\lambda)$ بر $\hat{\beta}^{\text{rM}}(\lambda)$ برتری دارد. دوباره نویسی رابطه (۳۳.۲) بر حسب مقادیر ویژه و λ ، نتیجه می شود

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{rM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{ptM}}(\lambda)) &= \\ &= \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^{\gamma}} \left\{ \eta^{\gamma} H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) - \omega_i \theta_i^{\gamma} Z(\alpha; \Delta^{\gamma}) - \gamma \lambda \theta_i^{\gamma} H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \right\}. \end{aligned}$$

وقتی $\lambda \leq \lambda_{\gamma}(\alpha, \Delta^{\gamma})$ ، سمت راست نامنفی خواهد بود. هنگامی که $\lambda \in (\circ, \lambda_{\gamma}(\alpha, \Delta^{\gamma}))$ PTRME، URME تسلط دارد و وقتی $\lambda > \lambda_{\gamma}(\alpha, \Delta^{\gamma})$ ، URME بر PTRME تسلط دارد. □

قضیه ۸.۳.۲ (مقایسه بین SRME و SME). (الف) تحت فرضیه صفر \mathcal{H}_0 ، به ازای هر λ ، به شرطی که $\circ < c < \gamma(p - \gamma)$ برآوردگر SRM بر برآوردگر SME بر حسب معیار مخاطره کمتر، برتری دارد. (ب) تحت فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ ، برآوردگر SRM بر SME برتری دارد اگر زوج مرتب $(\Delta^{\gamma}, \lambda)$ وجود دارد به طوری که $(\Delta^{\gamma}, \lambda) \in (\circ, \Delta_{\gamma}^{\lambda}(\alpha)) \times (\circ, \lambda_{\gamma}(\Delta^{\gamma}))$ که در آن

$$\Delta_{\gamma}^{\lambda}(\alpha) = \frac{\text{tr}(\mathbf{C}^{-1} - [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda)) (1 - cX(\Delta^{\gamma}))}{\text{Ch}_{\min}(G_{\gamma}(\Delta^{\gamma}) \mathbf{C}^{-1})}, \quad (35.2)$$

و

$$\begin{aligned} G_{\gamma}(\Delta^{\gamma}) &= \lambda^{\gamma} \mathbf{C}^{-1}(\lambda) + \lambda (\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda)) \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] \\ &\quad - c(c + \gamma) (\mathbf{I}_p - [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda)) \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] \end{aligned}$$

و

$$\lambda_{\gamma}(\Delta^{\gamma}) = \frac{\min_{1 \leq i \leq p} \left\{ \eta^{\gamma} (1 - cX(\Delta^{\gamma})) + c(c + \gamma) \omega_i \theta_i^{\gamma} \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] - c \omega_i \theta_i^{\gamma} \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] \right\}}{\max_{1 \leq i \leq p} \left\{ \theta_i^{\gamma} \left(1 - c \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] \right) \right\}}$$

برهان. تحت فرضیه صفر، اختلاف بین مخاطره ها برابر است با

$$\text{ADR}(\hat{\beta}^{\text{sM}}) - \text{ADR}(\hat{\beta}^{\text{sM}}(\lambda)) = \eta^{\gamma} \text{tr}(\mathbf{C}^{-1} - [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda)) (1 - cX(\circ)),$$

که از آن جایی که $X(\circ) = (2(p-2) - c)/(p(p-2))$ ، به ازای هر λ ، به شرطی که $0 < c < 2(p-2)$ باشد، نامنفی است.

تحت فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ ، اختلاف بین مخاطره‌ها به صورت زیر است

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{SM}}) - \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda)) &= \eta^2 \text{tr}(C^{-1} - [R(\lambda)]^T C^{-1} R(\lambda))(1 - cX(\Delta^2)) \\ &+ c(c + 4)\delta^T (I_p - [R(\lambda)]^T R(\lambda))\delta E \left[\chi_{p+4}^{-4}(\Delta^2) \right] - \lambda^2 \delta^T C^{-2}(\lambda)\delta \\ &- \lambda \delta^T \left[C^{-1}(\lambda)R(\lambda) + [R(\lambda)]^T C^{-1}(\lambda) \right] \delta E \left[\chi_{p+2}^{-2}(\Delta^2) \right]. \end{aligned} \quad (36.2)$$

با استفاده از نتیجه ۲.۱.۲، SRME بر SME برتری دارد اگر و فقط اگر $\Delta^2 \leq \Delta_{\text{F}}^2(\lambda)$ در این حالت، تابع مخاطره بر حسب مقادیر ویژه به صورت زیر است

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda)) &= \sum_{i=1}^p \frac{1}{(\omega_i + \lambda)^2} \left\{ \omega_i \eta^2 (1 - cX(\Delta^2)) + c(c + 4)\omega_i^2 \theta_i^2 E \left[\chi_{p+4}^{-4}(\Delta^2) \right] \right. \\ &\left. + \lambda^2 \theta_i^2 + 2c\lambda \omega_i \theta_i^2 E \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right\}, \end{aligned} \quad (37.2)$$

اگر از (۳۷.۲) نسبت به λ مشتق گرفته شود، آن گاه

$$\begin{aligned} \frac{\partial \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda))}{\partial \lambda} &= -2 \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^3} \left\{ \eta^2 (1 - cX(\Delta^2)) \right. \\ &\quad \left. + c(c + 4)\omega_i^2 \theta_i^2 E \left[\chi_{p+4}^{-4}(\Delta^2) \right] - c\omega_i \theta_i^2 E \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right. \\ &\quad \left. - \lambda \theta_i^2 \left(1 - cE \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right) \right\}. \end{aligned} \quad (38.2)$$

بنابراین، شرط کافی برای اینکه رابطه (۳۸.۲) نامنفی باشد، این است که λ ای وجود داشته باشد که $0 < \lambda < \lambda_{\text{F}}(\Delta^2)$. □

قضیه ۹.۳.۲ (مقایسه بین SRME و URME). (الف) تحت فرضیه \mathcal{H}_0 ، SRME بر URME به ازای $c \leq 2(p-2)$ برتری دارد. (ب) تحت فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ ، SRME بر URME برتری دارد اگر زوج مرتب (Δ^2, λ) وجود دارد به طوری که $(\Delta^2, \lambda) \in (\circ, \Delta_{\text{F}}^2(\lambda)) \times (\circ, \lambda_{\text{F}}(\Delta^2))$ که در آن

$$\Delta_{\text{F}}^2(\lambda) = \frac{\text{tr} \left([R(\lambda)]^T C^{-1} R(\lambda) \right) X(\Delta^2)}{\text{Ch}_{\min} \left(J_1(\Delta^2) C^{-1} \right)}$$

که

$$J_1(\Delta^2) = (c+4)[R(\lambda)]^T R(\lambda) E \left[\chi_{p+4}^{-4}(\Delta^2) \right] + \lambda \left(C^{-1}(\lambda)R(\lambda) + [R(\lambda)]^T C^{-1}(\lambda) \right) E \left[\chi_{p+2}^{-2}(\Delta^2) \right]$$

9

$$\lambda_{\text{F}}(\Delta^2) = \frac{\min_{1 \leq i \leq p} \left\{ \eta^2 X(\Delta^2) - (c + 4)\omega_i^2 \theta_i^2 E \left[\chi_{p+4}^{-4}(\Delta^2) \right] \right\}}{\max_{1 \leq i \leq p} \left\{ 2\theta_i^2 E \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right\}}$$

برهان. حالت ۱: فرضیه صفر \mathcal{H}_0 . در این حالت، اختلاف بین مخاطره‌ها برابر است با

$$c\eta^2 \operatorname{tr} \left([\mathbf{R}(\lambda)]^T \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) X(\circ) \geq \circ.$$

بنابراین به ازای $c \leq 2(p-2)$ نتیجه به دست می‌آید.

حالت ۲: فرضیه مقابل موضعی $\mathcal{K}_{(n)}$. اختلاف بین مخاطره‌های SRME و URME به صورت زیر است:

$$\begin{aligned} \operatorname{ADR}(\hat{\beta}^{\text{FM}}(\lambda)) - \operatorname{ADR}(\hat{\beta}^{\text{SM}}(\lambda)) &= c\eta^2 \operatorname{tr}([\mathbf{R}(\lambda)]^T \mathbf{C}^{-1} \mathbf{R}(\lambda)) X(\Delta^2) \\ &\quad - c(c+4)\delta^T [\mathbf{R}(\lambda)]^T \mathbf{R}(\lambda) \delta \mathbb{E} \left[\chi_{p+4}^{-4}(\Delta^2) \right] \\ &\quad - c\lambda(\delta^T \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta) \\ &\quad + \delta^T [\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda) \delta \mathbb{E} \left[\chi_{p+2}^{-2}(\Delta^2) \right], \quad (39.2) \end{aligned}$$

اختلاف بالا نامنفی خواهد بود وقتی که

$$\begin{aligned} \delta^T \left[(c+4)[\mathbf{R}(\lambda)]^T \mathbf{R}(\lambda) \mathbb{E} \left[\chi_{p+4}^{-4}(\Delta^2) \right] + \lambda(\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \right. \\ \left. + [\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda) \mathbb{E} \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right] \delta \leq \eta^2 \operatorname{tr} \left([\mathbf{R}(\lambda)]^T \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) X(\Delta^2). \end{aligned}$$

با استفاده از نتیجه ۲.۱.۱، SRME بر URME برتری دارد زمانی که $\Delta^2 \leq \Delta_{\delta}^2(\lambda)$. برای پیدا کردن شرط کافی روی λ ، رابطه (۳۹.۲) بر حسب مقادیر ویژه به صورت زیر در نظر گرفته می‌شود

$$\sum_{i=1}^p \frac{c\omega_i}{(\omega_i + \lambda)^2} \left\{ \eta^2 X(\Delta^2) - (c+4)\omega_i \theta_i^2 \mathbb{E} \left[\chi_{p+4}^{-4}(\Delta^2) \right] - 2\lambda \theta_i^2 \mathbb{E} \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right\}.$$

وقتی مقداری مانند λ وجود داشته باشد که $\lambda \in (\circ, \lambda_{\delta}(\Delta^2))$ ، رابطه فوق نامنفی خواهد بود. □

قضیه ۱۰.۳.۲ (مقایسه بین SRME و PTRME). (الف) تحت فرضیه صفر \mathcal{H}_0 ، SRME بر PTRME به ازای $\chi_{p,\alpha}^2 \leq H_{p+2}^{-1} \left(\frac{2c(p-2)-c^2}{p(p-2)} \right)$ برتری دارد. (ب) تحت $\mathcal{K}_{(n)}$ ، SRME بر PTRME برتری دارد اگر زوج مرتب (Δ^2, λ) وجود دارد به طوری که $(\Delta^2, \lambda) \in (\circ, \Delta_{\epsilon}^2(\lambda)) \times (\circ, \lambda_{\epsilon}(\Delta^2))$ که در آن

$$\Delta_{\epsilon}^2(\lambda) = \frac{\eta^2 \operatorname{tr} \left([\mathbf{R}(\lambda)]^T \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) \left(cX(\Delta^2) - H_{p+2}(\chi_{p,\alpha}^2; \Delta^2) \right)}{\operatorname{Ch}_{\min} \left(N_1(\alpha, \Delta^2) [\mathbf{R}(\lambda)]^T \mathbf{R}(\lambda) - N_2(\alpha, \Delta^2) (\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda)) \mathbf{C}^{-1} \right)}$$

9

$$\begin{aligned} N_1(\alpha, \Delta^2) &= \left(Z(\alpha; \Delta^2) - c(c+4)\mathbb{E} \left[\chi_{p+4}^{-4}(\Delta^2) \right] \right) \\ N_2(\alpha, \Delta^2) &= c\mathbb{E} \left[\chi_{p+2}^{-2}(\Delta^2) \right] - H_{p+2}(\chi_{p,\alpha}^2; \Delta^2). \end{aligned}$$

9

$$\lambda_{\epsilon}(\Delta^2) = \frac{\min_{1 \leq i \leq p} \left\{ \eta^2 \left(cX(\Delta^2) - H_{p+2}(\chi_{p,\alpha}^2; \Delta^2) \right) - \omega_i \theta_i^2 N_1(\alpha, \Delta^2) \right\}}{\max_{1 \leq i \leq p} \left\{ 2\theta_i^2 N_2(\alpha, \Delta^2) \right\}}.$$

برهان. ابتدا حالت ۲ فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ اثبات می‌شود. اختلاف بین دو مخاطره در این حالت برابر است با

$$\begin{aligned} & \text{ADR} \left(\hat{\beta}^{\text{ptM}}(\lambda) \right) - \text{ADR} \left(\hat{\beta}^{\text{sM}}(\lambda) \right) = \\ & \eta^{\gamma} \text{tr} \left([\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) \left(cX(\Delta^{\gamma}) - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \right) \\ & - \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda) \delta \left(Z(\alpha; \Delta^{\gamma}) - c(c + \mathbf{F}) \text{E} \left[\chi_{p+\mathbf{F}}^{-\mathbf{F}}(\Delta^{\gamma}) \right] \right) \\ & - \lambda (\delta^{\top} \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta + \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda) \delta) (c \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma})). \end{aligned} \quad (۴۰.۲)$$

سمت راست رابطه (۴۰.۲) نامنفی می‌شود اگر

$$\delta^{\top} \left[N_{\gamma}(\alpha, \Delta^{\gamma}) [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda) - \lambda_{\gamma}(\alpha, \Delta^{\gamma}) (\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda)) \right] \delta \leq \eta^{\gamma} \text{tr} \left([\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) \left(cX(\Delta^{\gamma}) - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \right),$$

که در آن،

$$\begin{aligned} N_{\gamma}(\alpha, \Delta^{\gamma}) &= Z(\alpha; \Delta^{\gamma}) - c(c + \mathbf{F}) \text{E} \left[\chi_{p+\mathbf{F}}^{-\mathbf{F}}(\Delta^{\gamma}) \right] \\ N_{\gamma}(\alpha, \Delta^{\gamma}) &= c \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}). \end{aligned}$$

با استفاده از نتیجه ۲.۱.۱، SRME بر PTRME برتری دارد وقتی که $\Delta^{\gamma} \leq \Delta_{\mathbf{F}}^{\gamma}(\lambda)$. اکنون، اختلاف مخاطره‌ها در رابطه (۴۰.۲) به صورت تابعی از مقادیر ویژه در نظر بگیرید

$$\begin{aligned} \text{ADR} \left(\hat{\beta}^{\text{ptM}}(\lambda) \right) - \text{ADR} \left(\hat{\beta}^{\text{sM}}(\lambda) \right) &= \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^{\gamma}} \\ &\times \left\{ \eta^{\gamma} \left(cX(\Delta^{\gamma}) - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) \right) - \omega_i \theta_i^{\gamma} N_{\gamma}(\alpha, \Delta^{\gamma}) - \gamma \lambda \theta_i^{\gamma} N_{\gamma}(\alpha, \Delta^{\gamma}) \right\}. \end{aligned}$$

عبارت بالا نامنفی خواهد بود وقتی $\lambda \in (\circ, \lambda_{\mathbf{F}}(\Delta^{\gamma})]$. دقت کنید تحت فرضیه صفر، رابطه (۴۰.۲)، به $\left(cX(\circ) - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \circ) \right)$ کاهش می‌یابد. بنابراین، اگر $\left(cX(\circ) - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \circ) \right) \geq \circ$ ، آن‌گاه SRME بر PTRME برتری دارد و در نتیجه رابطه

$$\chi_{p,\alpha}^{\gamma} \leq H_{p+\gamma}^{-1} \left(\frac{\gamma c(p - \gamma) - c^{\gamma}}{p(p - \gamma)} \right)$$

□

برقرار است.

قضیه ۱۱.۳.۲ (مقایسه بین PRRME و PRSME). (الف) تحت \mathcal{H}_o ، به‌ازای $p \geq 3$ برآوردگر PRRM بر PRSME برتری دارد. (ب) تحت $\mathcal{K}_{(n)}$ ، شرط کافی برای اینکه PRRME بر PRSME برتری داشته باشد این است که زوج مرتب $(\Delta^{\gamma}, \lambda) \in (\circ, \Delta_{\gamma}^{\gamma}(\lambda)) \times (\circ, \lambda_{\gamma}(\Delta^{\gamma}))$ وجود دارد که

$$\Delta_{\gamma}^{\gamma}(\lambda) = \frac{\text{tr}(\mathbf{C}^{-1} - [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda)) (1 - cX(\Delta^{\gamma}) + \text{E} \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}))^{\gamma} I(\chi_{p+\gamma}^{\gamma}(\Delta^{\gamma}) \leq c) \right])}{\text{Ch}_{\min} \left((\lambda^{\gamma} \mathbf{C}^{-\gamma}(\lambda) + \lambda [\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda)] A_{\gamma}(\Delta^{\gamma}) - (\mathbf{I}_p - [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda)) A_{\gamma}(\Delta^{\gamma}) \mathbf{C}^{-1} \right)},$$

$$\begin{aligned} A_1(\Delta^2) &= \left(E \left[\chi_{p+2}^{-2}(\Delta^2) \right] + E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2)) I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \right) \\ A_2(\Delta^2) &= c(c+4) \left(E \left[\chi_{p+4}^{-4}(\Delta^2) \right] + 2E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2)) I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \right. \\ &\quad \left. - E \left[(1 - c\chi_{p+4}^{-4}(\Delta^2))^2 I(\chi_{p+4}^4(\Delta^2) \leq c) \right] \right). \end{aligned}$$

همچنین

$$\lambda_V(\Delta^2) = \frac{f(\Delta^2)}{\max_{1 \leq i \leq p} \left\{ \theta_i^2 \left(1 - cE \left[\chi_{p+2}^{-2}(\Delta^2) \right] - E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2)) I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \right) \right\}},$$

که

$$\begin{aligned} f(\Delta^2) &= \min_{1 \leq i \leq p} \left\{ \eta^2 \left(1 - cX(\Delta^2) - E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2))^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \right) \right. \\ &\quad \left. + \omega_i \theta_i^2 \left(c(c+4)E \left[\chi_{p+4}^{-4}(\Delta^2) \right] - cE \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right. \right. \\ &\quad \left. \left. - E \left[(1 - c\chi_{p+4}^{-4}(\Delta^2))^2 I(\chi_{p+4}^4(\Delta^2) \leq c) \right] \right) \right\}. \end{aligned}$$

برهان. تحت فرضیه صفر، اختلاف بین دو برآوردگر برابر است با

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{sM}+}) - \text{ADR}(\hat{\beta}^{\text{sM}+}(\lambda)) &= \eta^2 \text{tr}(\mathbf{C}^{-1} - [\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1} \mathbf{R}(\lambda)) \\ &\quad \times \left(1 - cX(\circ) + E \left[(1 - c\chi_{p+2}^{-2}(\circ))^2 I(\chi_{p+2}^2(\circ) \leq c) \right] \right). \end{aligned}$$

وقتی $\circ \geq 1 - cX(\circ) + E \left[(1 - c\chi_{p+2}^{-2}(\circ))^2 I(\chi_{p+2}^2(\circ) \leq c) \right] \geq 0$ ، سمت راست عبارت فوق نامنفی است. به عبارت دیگر،

$$cX(\circ) - 1 \leq E \left[(1 - c\chi_{p+2}^{-2}(\circ))^2 I(\chi_{p+2}^2(\circ) \leq c) \right] \leq E \left[(1 - c\chi_{p+2}^{-2}(\circ))^2 \right].$$

بنابراین،

$$2cE \left[\chi_{p+2}^{-2}(\Delta^2) \right] - c^2 E \left[\chi_{p+2}^{-4}(\Delta^2) \right] - 1 \leq 1 - 2cE \left[\chi_{p+2}^{-2}(\Delta^2) \right] + c^2 E \left[\chi_{p+2}^{-4}(\Delta^2) \right].$$

پس از انجام محاسبات ساده، $c^2 - 2c(p-2) + p(p-2) \geq 0$ ، به دست می آید که همواره به ازای $p \geq 3$ درست است. بنابراین، با این شرط تحت فرضیه \mathcal{H}_0 ، PRRME به طور یکنواخت بر PRME برتری دارد.

تحت فرضیه مقابل موضعی، اختلاف مخاطره PRSRME و PRME به صورت زیر است:

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{sM}+}) - \text{ADR}(\hat{\beta}^{\text{sM}+}(\lambda)) &= \text{ADR}(\hat{\beta}^{\text{sM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{sM}}(\lambda)) \\ &\quad + \eta^2 \text{tr} \left(\mathbf{C}^{-1} - [\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2))^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \\ &\quad + \delta^\top (\mathbf{I}_p - [\mathbf{R}(\lambda)]^\top \mathbf{R}(\lambda)) \delta Q(\Delta^2) - \lambda \delta^\top (\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1}(\lambda)) \delta \\ &\quad \times E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2)) I(\chi_{p+2}^2(\Delta^2) \leq c) \right]. \end{aligned} \quad (41.2)$$

از جانشین کردن رابطه (۳۶.۲) در رابطه (۴۱.۲)، رابطه زیر نتیجه می‌شود.

$$\begin{aligned} & \eta^{\nu} \text{tr}(C^{-1} - [R(\lambda)]^T C^{-1} R(\lambda)) \\ & \times \left(1 - cX(\Delta^{\nu}) + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right) \\ & + c(c + \nu) \delta^T (I_p - [R(\lambda)]^T R(\lambda)) \delta (E [\chi_{p+\nu}^{-\nu}(\Delta^{\nu})] + Q(\Delta^{\nu})) - \lambda^{\nu} \delta^T C^{-\nu}(\lambda) \delta \\ & - \lambda \delta^T \left[C^{-1}(\lambda) R(\lambda) + [R(\lambda)]^T C^{-1}(\lambda) \right] \delta \\ & \times \left(E [\chi_{p+\nu}^{-\nu}(\Delta^{\nu})] + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right). \end{aligned} \quad (۴۲.۲)$$

رابطه (۴۲.۲) نامنفی خواهد بود اگر

$$\begin{aligned} & \delta^T \left[\lambda^{\nu} C^{-\nu}(\lambda) + \lambda \left[C^{-1}(\lambda) R(\lambda) + [R(\lambda)]^T C^{-1}(\lambda) \right] A_1(\Delta^{\nu}) - (I_p - [R(\lambda)]^T R(\lambda)) A_{\nu}(\Delta^{\nu}) \right] \delta \\ & \leq \eta^{\nu} \text{tr}(C^{-1} - [R(\lambda)]^T C^{-1} R(\lambda)) (1 - cX(\Delta^{\nu}) + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right]). \end{aligned}$$

طبق نتیجه ۲.۱.۱، PRRME بر PRME برتری دارد وقتی $\Delta^{\nu} \leq \Delta_{\nu}^{\nu}$. برای به دست آوردن شرط روی λ ، تابع مخاطره PRRME بر حسب مقادیر ویژه به صورت زیر دوباره نوشته می‌شود:

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{SM}+}(\lambda)) &= \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda)) \\ & - \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^{\nu}} \left\{ \eta^{\nu} E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right. \\ & \left. - \omega_i \theta_i^{\nu} Q(\Delta^{\nu}) - \nu \lambda \theta_i^{\nu} E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu})) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right\}, \end{aligned} \quad (۴۳.۲)$$

که $\text{ADR}(\hat{\beta}^{\text{SM}}(\lambda))$ در رابطه (۳۷.۲) داده شده است. از رابطه (۴۳.۲) نسبت به λ مشتق گرفته می‌شود، در این صورت

$$\begin{aligned} \frac{\partial \text{ADR}(\hat{\beta}^{\text{SM}+}(\lambda))}{\partial \lambda} &= \frac{\partial \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda))}{\partial \lambda} \\ & + \nu \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^{\nu}} \left\{ \eta^{\nu} E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right. \\ & \left. - \omega_i \theta_i^{\nu} Q(\Delta^{\nu}) - \lambda \theta_i^{\nu} E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu})) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right. \\ & \left. + \omega_i \theta_i^{\nu} E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu})) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right\}, \end{aligned} \quad (۴۴.۲)$$

که $\partial \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda)) / \partial \lambda$ در رابطه (۳۸.۲) داده شده است. برای هر عدد مثبت λ به ازای \square $\lambda \leq \lambda_1(\Delta^{\nu})$ ، مخاطره PRRME کمتر از مخاطره PRME است.

قضیه ۱۲.۳.۲ (مقایسه بین PRRME و URME). (الف) تحت فرضیه \mathcal{H}_0 ، PRRME بر URME برتری دارد. (ب) تحت فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ ، شرط کافی برای اینکه PRRME بر URME برتری داشته باشد این است که زوج مرتب $(\circ, \lambda_{\lambda}(\Delta^{\nu})) \times (\circ, \Delta_{\lambda}^{\nu}(\lambda)) \in (\Delta^{\nu}, \lambda)$ وجود داشته باشد که

$$\Delta_{\lambda}^{\nu} = \frac{\text{tr}([R(\lambda)]^T C^{-1} R(\lambda)) g_1(\Delta^{\nu})}{\text{Ch}_{\min} \left(([R(\lambda)]^T R(\lambda) g_2(\Delta^{\nu}) + \lambda (C^{-1}(\lambda) R(\lambda) + [R(\lambda)]^T C^{-1}(\lambda)) g_3(\Delta^{\nu})) C^{-1} \right)}$$

$$\begin{aligned} g_1(\Delta^r) &= cX(\Delta^r) + E \left[(\mathbb{1} - c\chi_{p+r}^{-r}(\Delta^r))^r I(\chi_{p+r}^r(\Delta^r) \leq c) \right], \\ g_2(\Delta^r) &= c(c + \mathfrak{F})E \left[\chi_{p+\mathfrak{F}}^{-\mathfrak{F}}(\Delta^r) \right] + Q(\Delta^r), \\ g_3(\Delta^r) &= cE \left[\chi_{p+r}^{-r}(\Delta^r) \right] + E \left[(\mathbb{1} - c\chi_{p+r}^r(\Delta^r))I(\chi_{p+r}^r(\Delta^r) \leq c) \right]. \end{aligned}$$

همچنین

$$\lambda_\lambda(\Delta^r) = \frac{g_4(\Delta^r)}{\max_{1 \leq i \leq p} \left\{ \theta_i^r \left(cE \left[\chi_{p+r}^{-r}(\Delta^r) \right] - E \left[(\mathbb{1} - c\chi_{p+r}^{-r}(\Delta^r))^r I(\chi_{p+r}^r(\Delta^r) \leq c) \right] \right) \right\}}$$

که

$$g_4(\Delta^r) = \min_{1 \leq i \leq p} \left\{ \eta^r \left(cX(\Delta^r) - E \left[(\mathbb{1} - c\chi_{p+r}^{-r}(\Delta^r))^r I(\chi_{p+r}^r(\Delta^r) \leq c) \right] \right) - \omega_i \theta_i^r \left(c(c + \mathfrak{F})E \left[\chi_{p+\mathfrak{F}}^{-\mathfrak{F}}(\Delta^r) \right] - Q(\Delta^r) \right) \right\}.$$

برهان. تحت فرضیه صفر \mathcal{H}_0 ,

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{uM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{SM}+}(\lambda)) &= \\ \eta^r \text{tr} \left(cX(\circ) + E \left[(\mathbb{1} - c\chi_{p+r}^{-r}(\circ))^r I(\chi_{p+r}^r(\circ) \leq c) \right] \right) &\geq \circ. \end{aligned}$$

از آن جایی که

$$-cX(\circ) \leq E \left[(\mathbb{1} - c\chi_{p+r}^{-r}(\circ))^r I(\chi_{p+r}^r(\circ) \leq c) \right] \leq E \left[(\mathbb{1} - c\chi_{p+r}^{-r}(\circ))^r \right].$$

تحت فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ ، اختلاف مخاطره‌ها برابر است با

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{uM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{SM}+}(\lambda)) &= \text{ADR}(\hat{\beta}^{\text{uM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda)) \\ &+ \eta^r \text{tr} \left([\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) E \left[(\mathbb{1} - c\chi_{p+r}^{-r}(\Delta^r))^r I(\chi_{p+r}^r(\Delta^r) \leq c) \right] \\ &- \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{R}(\lambda) \delta Q(\Delta^r) - \lambda \left(\delta^\top \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta + \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1}(\lambda) \delta \right) \\ &\times E \left[(\mathbb{1} - c\chi_{p+r}^r(\Delta^r))I(\chi_{p+r}^r(\Delta^r) \leq c) \right] \\ &= \eta^r \text{tr}([\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1} \mathbf{R}(\lambda))g_1(\Delta^r) - \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{R}(\lambda) \delta g_2(\Delta^r) \\ &- \lambda(\delta^\top \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta + \delta^\top [\mathbf{R}(\lambda)]^\top \mathbf{C}^{-1}(\lambda) \delta)g_3(\Delta^r), \end{aligned} \quad (45.2)$$

که در آن

$$\begin{aligned} g_1(\Delta^r) &= cX(\Delta^r) + E \left[(\mathbb{1} - c\chi_{p+r}^{-r}(\Delta^r))^r I(\chi_{p+r}^r(\Delta^r) \leq c) \right] \\ g_2(\Delta^r) &= c(c + \mathfrak{F})E \left[\chi_{p+\mathfrak{F}}^{-\mathfrak{F}}(\Delta^r) \right] + Q(\Delta^r) \\ g_3(\Delta^r) &= cE \left[\chi_{p+r}^{-r}(\Delta^r) \right] + E \left[(\mathbb{1} - c\chi_{p+r}^r(\Delta^r))I(\chi_{p+r}^r(\Delta^r) \leq c) \right]. \end{aligned}$$

سمت راست رابطه (۴۵.۲) نامنفی است وقتی

$$\delta^T \left[[\mathbf{R}(\lambda)]^T \mathbf{R}(\lambda) g_{\gamma}(\Delta^{\gamma}) + \lambda (\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda)) g_{\gamma}(\Delta^{\gamma}) \right] \delta \leq \eta^{\gamma} \text{tr}([\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda)) g_1(\Delta^{\gamma}).$$

با استفاده از نتیجه ۲.۱.آ، PRRME بر URME برتری دارد وقتی $\Delta^{\gamma} \leq \lambda_{\gamma}(\lambda)$. برای یافتن شرط روی λ ، اختلاف مخاطره PRRME و URME بر حسب مقادیر ویژه برابر است با

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{uM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{sM}+}(\lambda)) &= \text{ADR}(\hat{\beta}^{\text{uM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{sM}}(\lambda)) \\ &- \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^{\gamma}} \left\{ \eta^{\gamma} \text{E} \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}))^{\gamma} I(\chi_{p+\gamma}^{\gamma}(\Delta^{\gamma}) \leq c) \right] \right. \\ &\left. - \omega_i \theta_i^{\gamma} Q(\Delta^{\gamma}) - \gamma \lambda \theta_i^{\gamma} \text{E} \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}))^{\gamma} I(\chi_{p+\gamma}^{\gamma}(\Delta^{\gamma}) \leq c) \right] \right\} \\ &= \sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^{\gamma}} \left\{ \eta^{\gamma} \left(cX(\Delta^{\gamma}) - \text{E} \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}))^{\gamma} I(\chi_{p+\gamma}^{\gamma}(\Delta^{\gamma}) \leq c) \right] \right) \right. \\ &- \omega_i \theta_i^{\gamma} \left(c(c + \gamma) \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] - Q(\Delta^{\gamma}) \right) \\ &\left. - \gamma \lambda \theta_i^{\gamma} \left(c \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] - \text{E} \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}))^{\gamma} I(\chi_{p+\gamma}^{\gamma}(\Delta^{\gamma}) \leq c) \right] \right) \right\}. \end{aligned}$$

سمت راست عبارت بالا نامنفی خواهد بود اگر $\lambda \leq \lambda_{\gamma}(\Delta^{\gamma})$. در این صورت، PRRME مخاطره کمتری از URME دارد. □

قضیه ۱۳.۳.۲ (مقایسه بین PRRME و PTRME). (الف) تحت فرضیه H_0 ، برای همه α هایی که در شرط

$$\left\{ \alpha : \chi_{p,\alpha}^{\gamma} \leq H_{p+\gamma}^{-1} \left(\frac{\gamma c(p - \gamma) - c^{\gamma}}{p(p - \gamma)} + \text{E} \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\circ))^{\gamma} I(\chi_{p+\gamma}^{\gamma}(\circ) \leq c) \right] \right) \right\},$$

صدق می کند، PRRME بر PTRME برتری دارد.

(ب) تحت فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ ، شرط کافی برای برتری برآوردگر جزء مثبت انقباضی نوع استاین M ریج بر PTRME این است که زوج مرتب $(\Delta^{\gamma}, \lambda) \in (\circ, \Delta_{\alpha}^{\gamma}(\lambda, \alpha)) \times (\circ, \lambda_{\alpha}(\Delta^{\gamma}, \alpha))$ به ازای $\alpha \in (\circ, 1)$ وجود داشته باشد که

$$\Delta_{\alpha}^{\gamma}(\lambda; \alpha) = \frac{\text{tr}([\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda)) \left(cX(\Delta^{\gamma}) - H_{p+\gamma}(\chi_{p,\alpha}^{\gamma}; \Delta^{\gamma}) + \text{E} \left[(1 - c\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}))^{\gamma} I(\chi_{p+\gamma}^{\gamma}(\Delta^{\gamma}) \leq c) \right] \right)}{\text{Ch}_{\min} \left(([\mathbf{R}(\lambda)]^T \mathbf{R}(\lambda) B_1(\Delta^{\gamma}) - \lambda (\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda)) B_{\gamma}(\Delta^{\gamma})) \mathbf{C}^{-1} \right)}$$

9

$$\lambda_{\alpha}(\Delta^{\gamma}) = \frac{h(\Delta^{\gamma})}{\max_{1 \leq i \leq p} \left\{ \gamma \theta_i^{\gamma} B_{\gamma}(\Delta^{\gamma}) \right\}}$$

که

$$B_1(\Delta^{\gamma}) = Z(\alpha; \Delta^{\gamma}) - c(c + \gamma) \text{E} \left[\chi_{p+\gamma}^{-\gamma}(\Delta^{\gamma}) \right] + Q(\Delta^{\gamma})$$

$$B_{\nu}(\Delta^{\nu}) = cE \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right] - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) \\ + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu})) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right]$$

9

$$h(\Delta^{\nu}) = \min_{1 \leq i \leq p} \left\{ \eta^{\nu} \left(cX(\Delta^{\nu}) - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right) \right. \\ \left. - \omega_i \theta_i^{\nu} B_1(\Delta^{\nu}) \right\}.$$

برهان. تحت فرضیه صفر \mathcal{H}_0 ، اختلاف بین مخاطره‌ها برابر است با

$$\eta^{\nu} \text{tr} \left([\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) \left(cX(\circ) - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \circ) + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\circ))^{\nu} I(\chi_{p+\nu}^{\nu}(\circ) \leq c) \right] \right) \geq \circ.$$

به‌ازای هر α در شرط زیر صدق می‌کند:

$$\left\{ \alpha : \chi_{p,\alpha}^{\nu} \leq H_{p+\nu}^{-1} \left(\frac{\nu c(p - \nu) - c^{\nu}}{p(p - \nu)} + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\circ))^{\nu} I(\chi_{p+\nu}^{\nu}(\circ) \leq c) \right] \right) \right\}. \quad (46.2)$$

بنابراین وقتی $\chi_{p,\alpha}^{\nu}$ در رابطه (46.2) صدق می‌کند، مخاطره PRRME، کوچکتر از مخاطره‌ی PTRME است.

تحت فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ ، اختلاف بین مخاطره‌ها برابر است با

$$\text{ADR}(\hat{\beta}^{\text{ptM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{SM+}}(\lambda)) = \text{ADR}(\hat{\beta}^{\text{ptM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda)) \\ + \eta^{\nu} \text{tr} \left([\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \\ - \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda) \delta Q(\Delta^{\nu}) - \lambda \delta^{\top} \left(\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda) \right) \delta \\ \times E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu})) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \\ = \eta^{\nu} \text{tr} \left([\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) \left(cX(\Delta^{\nu}) - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) \right. \\ \left. + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right) \\ - \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda) \delta \left(Z(\alpha; \Delta^{\nu}) - c(c + \nu) E \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right] + Q(\Delta^{\nu}) \right) \\ - \lambda (\delta^{\top} \mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) \delta + \delta^{\top} [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda) \delta) \left(cE \left[\chi_{p+\nu}^{-\nu}(\Delta^{\nu}) \right] \right. \\ \left. - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu})) I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right). \quad (47.2)$$

سمت راست رابطه (47.2) نامنفی خواهد بود وقتی

$$\delta^{\top} \left([\mathbf{R}(\lambda)]^{\top} \mathbf{R}(\lambda) B_1(\Delta^{\nu}) - \lambda (\mathbf{C}^{-1}(\lambda) \mathbf{R}(\lambda) + [\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1}(\lambda)) B_{\nu}(\Delta^{\nu}) \right) \delta \\ \leq \eta^{\nu} \text{tr} \left([\mathbf{R}(\lambda)]^{\top} \mathbf{C}^{-1} \mathbf{R}(\lambda) \right) \left(cX(\Delta^{\nu}) - H_{p+\nu}(\chi_{p,\alpha}^{\nu}; \Delta^{\nu}) \right. \\ \left. + E \left[(1 - c\chi_{p+\nu}^{-\nu}(\Delta^{\nu}))^{\nu} I(\chi_{p+\nu}^{\nu}(\Delta^{\nu}) \leq c) \right] \right). \quad (48.2)$$

با استفاده از نتیجه 2.1.آ در رابطه (48.2)، هنگامی که $\Delta^{\nu} \leq \Delta_q^{\nu}(\lambda; \alpha)$ ، PRRME بر PTRME مسلط است.

اکنون، اختلاف مخاطره‌های بین PRRME و PTRME بر حسب تابعی از مقادیر ویژه به صورت زیر نوشته می‌شود

$$\sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^2} \left\{ \eta^2 \left(cX(\Delta^2) - H_{p+2}(\chi_{p,\alpha}^2; \Delta^2) \right) + E \left[(1 - c\chi_{p+2}^2(\Delta^2))^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] - \omega_i \theta_i^2 F^*(\Delta^2) - 2\lambda \theta_i^2 G^*(\Delta^2) \right\}.$$

به عبارت دیگر، وقتی $\lambda \in (0, \lambda_0(\Delta^2))$ وجود داشته باشد، اختلاف مخاطره نامنفی است. □

قضیه ۱۴.۳.۲ (مقایسه بین PRRME و SRME). شرط کافی برای اینکه PRRME مخاطره کمتری نسبت به SRME داشته باشد این است که زوج مرتب $(\Delta^2, \lambda) \in (0, \infty) \times (0, \lambda_0(\Delta^2))$ وجود داشته باشد که

$$\lambda_{1_0}(\Delta^2) = \frac{\min_{1 \leq i \leq p} \left\{ \eta^2 E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2))^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] - \omega_i \theta_i^2 Q(\Delta^2) \right\}}{\max_{1 \leq i \leq p} \left\{ 2\theta_i^2 E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2)) I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \right\}}.$$

برهان. اختلاف بین دو مخاطره به صورت زیر به دست می‌آید:

$$\begin{aligned} \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda)) - \text{ADR}(\hat{\beta}^{\text{SM}+}(\lambda)) &= \\ & \eta^2 \text{tr}([\mathbf{R}(\lambda)]^T \mathbf{C}^{-1} \mathbf{R}(\lambda)) E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2))^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \\ & + 2\lambda \delta^T [\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda) \delta E \left[(c\chi_{p+2}^{-2}(\Delta^2) - 1) I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \\ & - \delta^T [\mathbf{R}(\lambda)]^T \mathbf{R}(\lambda) \delta Q(\Delta^2) \\ & = \eta^2 \text{tr}([\mathbf{R}(\lambda)]^T \mathbf{C}^{-1} \mathbf{R}(\lambda)) E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2))^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \\ & + \delta^T [\mathbf{R}(\lambda)]^T \mathbf{R}(\lambda) \delta E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2))^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \\ & + 2\lambda \delta^T [\mathbf{R}(\lambda)]^T \mathbf{R}(\lambda) \delta E \left[(c\chi_{p+2}^{-2}(\Delta^2) - 1) I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \\ & + 2\lambda \delta^T [\mathbf{R}(\lambda)]^T \mathbf{C}^{-1}(\lambda) \delta E \left[(c\chi_{p+2}^{-2}(\Delta^2) - 1) I(\chi_{p+2}^2(\Delta^2) \leq c) \right]. \end{aligned} \quad (۴۹.۲)$$

با توجه به اینکه مقدار مورد انتظار یک متغیر تصادفی مثبت است، سمت راست رابطه (۴۹.۲) مثبت است.

بر اساس تعریف تابع نشان‌گر، $0 < \chi_{p+2}^2(\Delta^2) \leq c$ ، بنابراین، $c\chi_{p+2}^{-2}(\Delta^2) - 1 \geq 0$. در نهایت $0 \leq E \left[(c\chi_{p+2}^{-2}(\Delta^2) - 1) I(\chi_{p+2}^2(\Delta^2) \leq c) \right]$ در نتیجه، به ازای هر Δ^2 ، رابطه $\text{ADR}(\hat{\beta}^{\text{SM}+}(\lambda)) \leq \text{ADR}(\hat{\beta}^{\text{SM}}(\lambda))$ برقرار است. PRRME نه تنها غیرمجاز بودن SRME را تایید می‌کند بلکه برآوردگر بهتری را ارائه می‌کند. عبارت (۴۹.۲) بر حسب مقادیر ویژه برابر است با

$$\sum_{i=1}^p \frac{\omega_i}{(\omega_i + \lambda)^2} \left\{ \eta^2 E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2))^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] - \omega_i \theta_i^2 Q(\Delta^2) - 2\lambda \theta_i^2 E \left[(1 - c\chi_{p+2}^{-2}(\Delta^2)) I(\chi_{p+2}^2(\Delta^2) \leq c) \right] \right\}.$$

به ازای $\lambda \in [0, \lambda_{1_0}(\Delta^2)]$ ، PRRME بر SRME برتری دارد. □

۳.۳.۲ انتخاب پارامتر ریج

بنا به ماهیت غیرخطی برآوردگر رگرسیون ریج نسبت به پارامتر λ ، انتخاب یا روش برآورد پارامتر ریج از اهمیت بسیاری برخوردار است. روش‌های متفاوتی برای تعیین پارامتر ریج توسط نویسندگان مختلف پیشنهاد شده است.

استوار λ به‌عنوان مثال، سیلواپول (۱۹۹۱) و ال-سلام (۲۰۱۳) را ببینید. سامکار و آلیو (۲۰۱۰)، مقدار λ را به‌صورت زیر پیشنهاد کردند:

$$\lambda = \frac{ps_M^2}{[\hat{\beta}^{uM}]^T \hat{\beta}^{uM}}. \quad (50.2)$$

۴.۳.۲ نمایش گرافیکی

با توجه به اینکه روابط بین برآوردهای پیشنهادی پیچیده هستند، در حالت رگرسیون متعامد ($C = I_p$) نمودارهای ADQB و ADR برآوردهای محدود نشده، محدود شده و بهبودیافته (۱۲.۲)–(۱۴.۲) بر حسب Δ^2 و λ برای مقادیر مختلف α ، λ ، η^2 و p در راستای درک بهتر کارایی برآوردهای بهبودیافته M ارائه شده است.

در رگرسیون متعامد، $R(\lambda) = 1/(1 + \lambda)I_p$ ، $C(\lambda) = (1 + \lambda)I_p$ و $\delta^T \delta = \eta^2 \Delta^2$ اکنون، بر اساس رابطه (۱۷.۲)، عبارت‌های ADQB برآوردها برابر است با:

$$\begin{aligned} \text{ADQB}(\hat{\beta}^{uM}) &= \frac{\lambda^2}{(1 + \lambda)^2} \eta^2 \Delta^2 \\ \text{ADQB}(\hat{\beta}^{ptM}) &= \frac{1}{(1 + \lambda)^2} \eta^2 \Delta^2 \left\{ \lambda^2 + 2\lambda H_{p+2}(\chi_{p,\alpha}^2; \Delta^2) + [H_{p+2}(\chi_{p,\alpha}^2; \Delta^2)]^2 \right\} \\ \text{ADQB}(\hat{\beta}^{sM}) &= \frac{1}{(1 + \lambda)^2} \eta^2 \Delta^2 \left\{ \lambda^2 + 2c\lambda E[\chi_{p+2}^{-2}(\Delta^2)] + c^2 (E[\chi_{p+2}^{-2}(\Delta^2)])^2 \right\} \\ \text{ADQB}(\hat{\beta}^{sM+}) &= \frac{1}{(1 + \lambda)^2} \eta^2 \Delta^2 \left\{ \lambda^2 + 2\lambda X(\Delta) + (X(\Delta^2))^2 \right\}. \end{aligned}$$

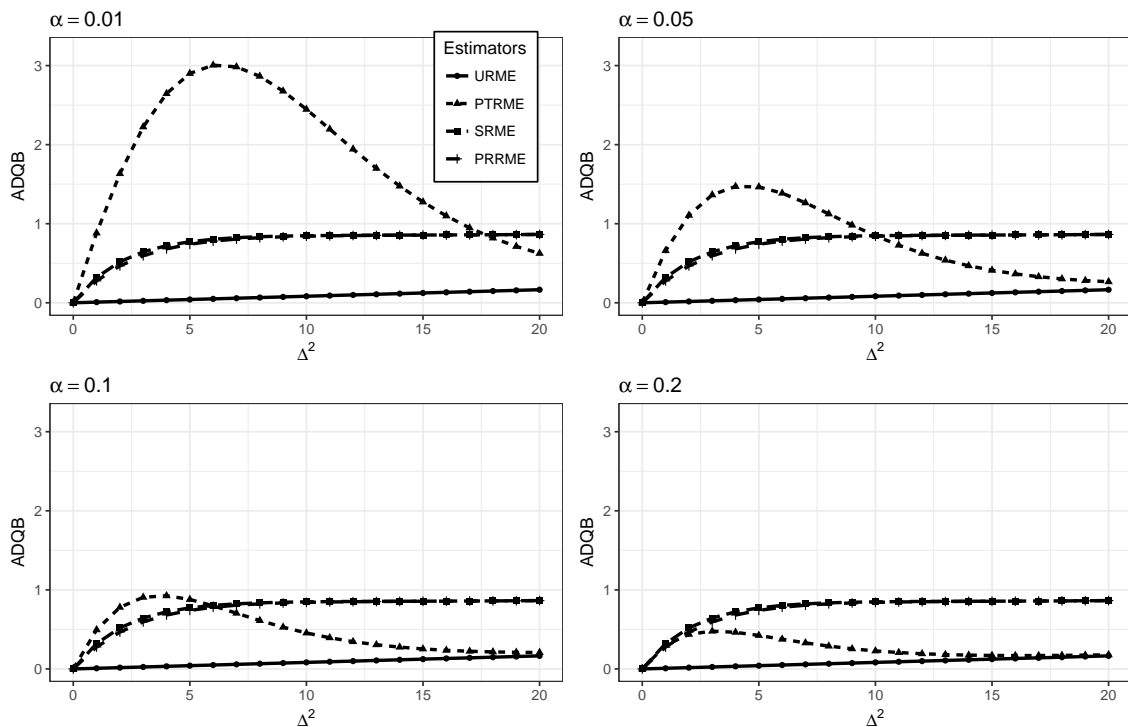
قضیه ۳.۳.۲ را برای رگرسیون متعامد، می‌توان به‌صورت زیر نوشت:

$$\begin{aligned} \text{ADR}(\hat{\beta}^{uM}) &= \frac{\eta^2}{(1 + \lambda)^2} \{p + \lambda^2 \Delta^2\} \\ \text{ADR}(\hat{\beta}^{ptM}) &= \frac{\eta^2}{(1 + \lambda)^2} \left\{ p \left(1 - H_{p+2}(\chi_{p,\alpha}^2; \Delta^2) \right) + \Delta^2 X(\alpha; \Delta^2) \right. \\ &\quad \left. + \lambda^2 \Delta^2 + 2\lambda \Delta^2 H_{p+2}(\chi_{p,\alpha}^2; \Delta^2) \right\} \\ \text{ADR}(\hat{\beta}^{sM}) &= \frac{\eta^2}{(1 + \lambda)^2} \left\{ p \left(1 - cX(\Delta^2) \right) + c(c + 4)\Delta^2 E[\chi_{p+4}^{-4}(\Delta^2)] \right. \\ &\quad \left. + \lambda^2 \Delta^2 + 2c\lambda \Delta^2 E[\chi_{p+2}^{-2}(\Delta^2)] \right\} \end{aligned}$$

(51.2)

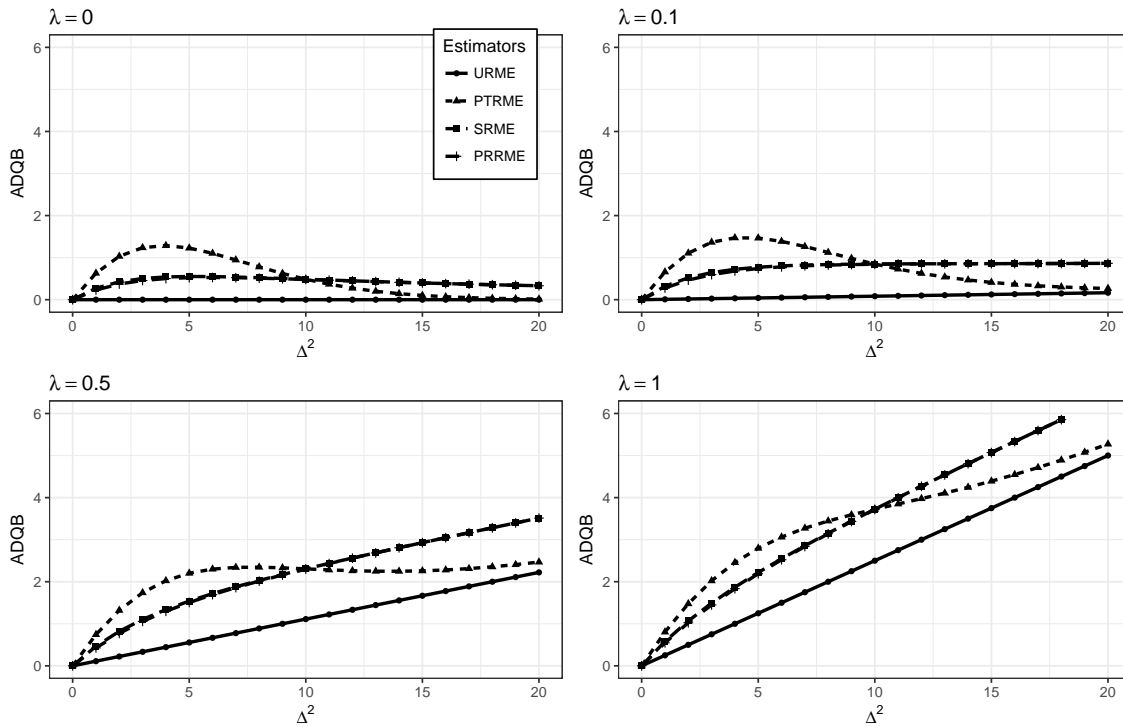
$$ADR(\hat{\beta}^{SM+}) = ADR(\hat{\beta}^{SM+}) - \frac{\eta^2}{(1+\lambda)^2} \left\{ pE \left[\left(1 - c\chi_{p+2}^{-2}(\Delta^2)\right)^2 I(\chi_{p+2}^2(\Delta^2) \leq c) \right] + 2\lambda\Delta^2 E \left[\left(1 - c\chi_{p+2}^{-2}(\Delta^2)\right) I(\chi_{p+2}^2(\Delta^2) \leq c) \right] + \Delta^2 Q(\Delta^2) \right\}.$$

شکل ۱.۲، ADQB را برای برآوردهای بهبودیافته M ریح به ازای $\lambda = 0.1$ ، $p = 5$ و برای مقادیر مختلف α نشان می‌دهد. با افزایش پارامتر α ، ADQB برآوردهای PTRM کاهش می‌یابد زیرا وقتی مقدار α افزایش پیدا کند، با احتمال بیشتری فرضیه صفر رد می‌شود.



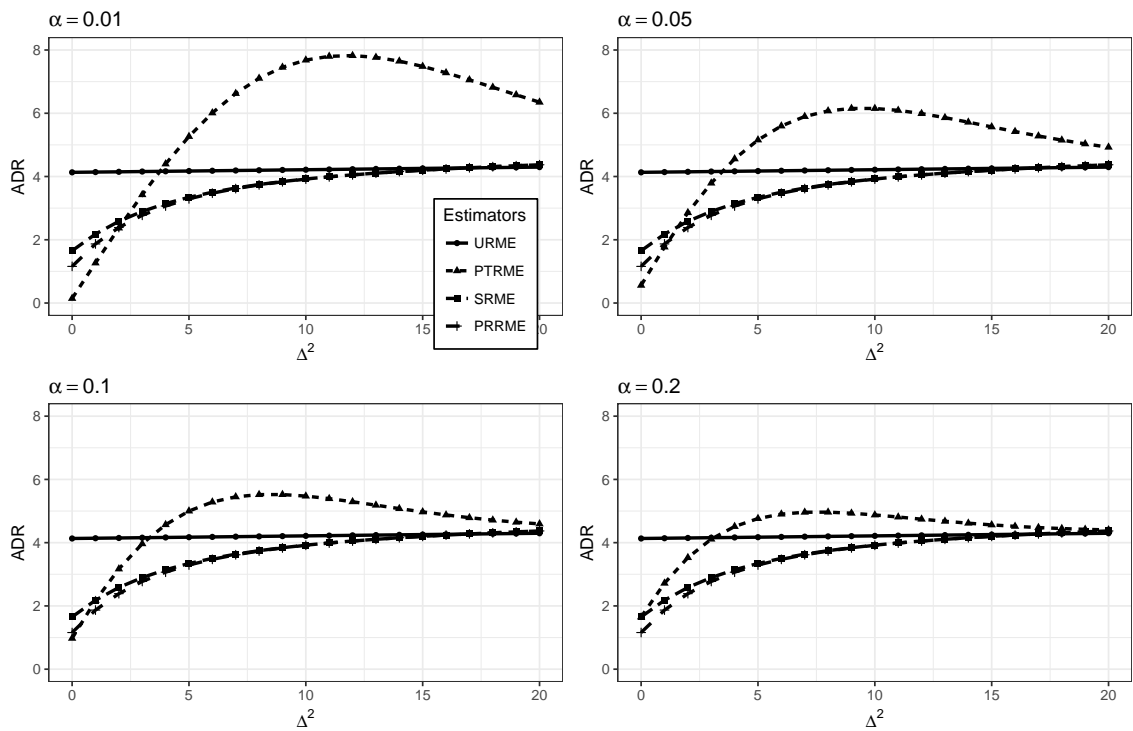
شکل ۱.۲: رفتار ADQB برآوردها برای $\eta^2 = 1$ ، $\lambda = 0.1$ ، $p = 5$ و مقادیر مختلف α .

شکل ۲.۲، ADQB را برای مقادیر مختلف λ به ازای $p = 5$ ، $\eta^2 = 1$ و $\alpha = 0.1$ نشان می‌دهد. وقتی $\lambda = 0$ ، برآوردهای پیشنهاد شده همان برآوردهای بهبودیافته M هستند. در این شکل، ADQB برآوردهای URM صفر می‌باشد. همان طور که λ افزایش می‌یابد، برآوردهای بهبودیافته M ریح به برآوردهای اریب تبدیل و ADQB آنها نیز بیشتر می‌شود. همچنین وقتی Δ^2 افزایش پیدا می‌کند، مقدار ADQB برآوردهای PTRM به ADQB برآوردهای URM نزدیک می‌شود. لازم به یادآوری است که Δ^2 میزان دوری از فرضیه صفر H_0 را نشان می‌دهد.

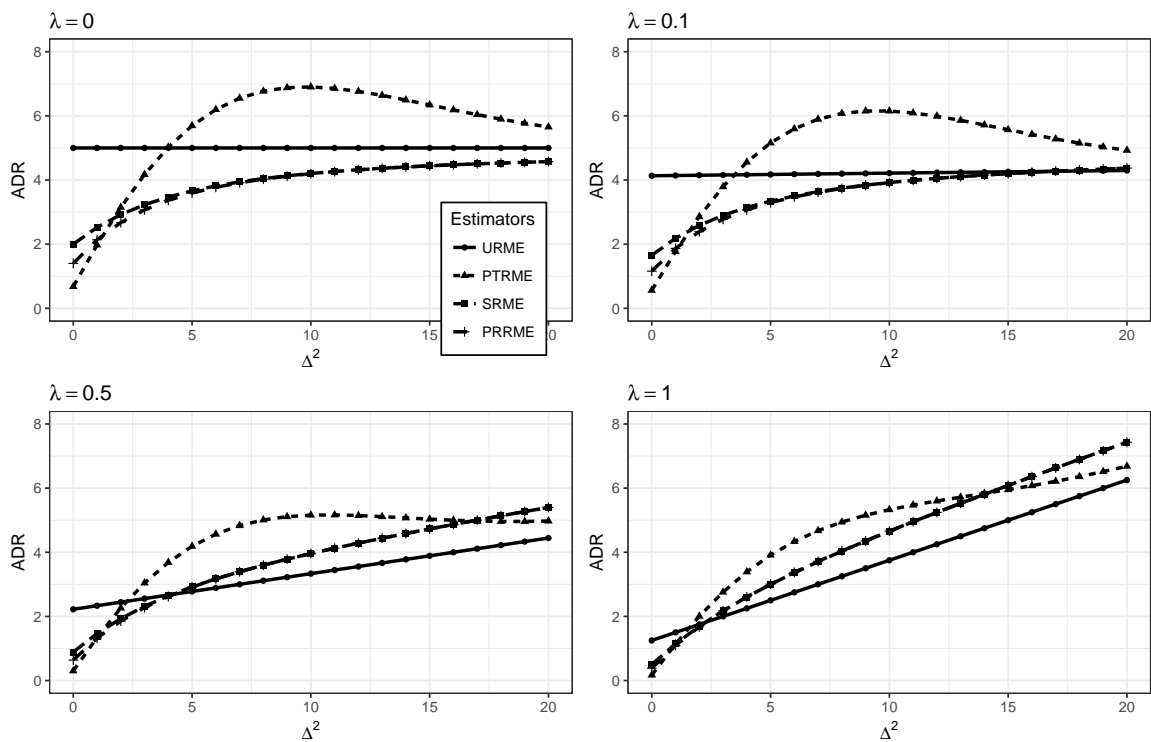


شکل ۲.۲: رفتار ADQB برآوردگرها برای $p = 5$, $\alpha = 0.05$, $\eta^2 = 1$ و مقادیر مختلف λ .

شکل‌های ۳.۲ و ۴.۲، مقادیر ADR را برای برآوردگرهای پیشنهاد شده نشان می‌دهد. برای مقادیر بزرگتر α ، PTRME، بهتر عمل می‌کند. ADR برآوردگرهای بهبودیافته M ریج کمتر از ADR برآوردگرهای بهبودیافته M دارد (نمودار بالا – چپ شکل ۴.۲ را ببینید). مخاطره PRRME همواره کمتر از مخاطره SRME است اما کارایی آن وقتی که λ به یک نزدیک می‌شود، به‌ازای $p = 5$ ، $\eta^2 = 1$ و $\alpha = 0.05$ کاهش می‌یابد. هم‌چنین مشاهده می‌شود که نزدیک فرضیه صفر، PTRME بر URME، SRME و PRRME برتری دارد. مخاطره برآوردگرهای پیشنهادی به‌ازای مقادیر بزرگتر Δ^2 ، به مقدار مشترک مخاطره URME نزدیک می‌شود. برای مشاهده شکل‌های بیشتر، **نوروزی‌راد و آرشی (۲۰۱۷c)** را ببینید.



شکل ۳.۲: رفتار ADR برآوردگرها برای $p = 5$ ، $\lambda = 0.1$ ، $\eta^2 = 1$ و مقادیر مختلف α .



شکل ۴.۲: رفتار ADR برآوردگرها برای $p = 5$ ، $\alpha = 0.05$ ، $\eta^2 = 1$ و مقادیر مختلف λ .

برای رگرسیون متعامد، کارایی برآوردگرها به طور عددی با استفاده از کارایی نسبی (REff) که به صورت زیر تعریف می شود محاسبه شده است.

$$\text{REff}(\hat{\beta}^{*M}(\lambda)) = \frac{\text{ADR}(\hat{\beta}^{uM}(\lambda))}{\text{ADR}(\hat{\beta}^{*M}(\lambda))}. \quad (52.2)$$

برآوردگر $\hat{\beta}^{*M}(\lambda)$ می تواند هر یک از برآوردگرهای PTRM، SRM و PRRM باشد. بنابراین، مخاطره نسبی بیشتر از یک، برتری برآوردگرها از دیدگاه مخاطره کمتر بر URME را نشان می دهد.

جدول های ۱.۲ و ۲.۲ مقادیر مخاطره های نسبی را برای برآوردگرهای پیشنهادی به ترتیب در دو حالت $p = 5$ و $p = 20$ نشان می دهد. برآوردگرهای بهبودیافته M ریج به ازای مقادیر کوچکتر λ و نزدیک فرضیه صفر در هر دو حالت بهتر عمل می کنند. وقتی Δ^2 از صفر دور می شود، رفتار PTRME رضایت بخش است.

بر اساس جدول های ۱.۲ و ۲.۲، نتایج زیر به دست می آید:

(۱) انتخاب پارامتر ریج λ در این مطالعه از اهمیت بسزایی برخوردار است. این نتیجه نقش کلیدی در یافتن برآوردگر بهتر بر اساس معیار مخاطره کمتر بازی می کند. تمام برآوردگرهای پیشنهادی، کارایی بهتری نزدیک فرضیه صفر دارند. وقتی Δ^2 از صفر فاصله می گیرد، رفتار PTRME هنوز برای مقادیر کوچک λ قابل قبول است.

(۲) همان طور که در قضیه ۱۲.۳.۲ اشاره شد، تحت فرضیه صفر H_0 ، همواره PRRME بر URME برتری دارد. اگر $\Delta^2 > 0$ ، آن گاه $\Delta^2_\lambda(\lambda)$ و $\lambda_\lambda(\Delta^2)$ وجود دارند به طوری که $(\Delta^2, \lambda) \in (0, \Delta^2_\lambda(\lambda)) \times (0, \lambda_\lambda(\Delta^2))$ ، مخاطره PRRME کمتر از URME است. در تایید این نکته کفایت جدول ۱.۲ و جدول ۲.۲ را به ترتیب برای $\lambda \geq 0.5$ و $\lambda \geq 0.1$ ببینید.

(۳) قضیه ۱۳.۳.۲ کارایی PRRME و PTRME را مقایسه می کند. شرط برتری PRRME بر برآوردگر PTRM، وجود $\Delta^2_\lambda(\lambda, \alpha)$ و $\lambda_\alpha(\Delta^2)$ است به طوری که $0 \leq \Delta^2 \leq \Delta^2_\lambda(\lambda, \alpha)$ و $0 \leq \lambda \leq \lambda_\alpha(\Delta^2)$ در جدول های ۱.۲ و ۲.۲، موارد زیادی وجود دارند که لزوم این شرط را تایید می کنند. به عنوان مثال، برای $\lambda = 0.5$ ، برآوردگر PTRM بهتر از PRRME به ازای $\Delta^2 = 20$ و $\Delta^2 \approx 27$ ، به ترتیب در جدول های ۱.۲ و ۲.۲، رفتار می کند. تحت فرضیه صفر، برتری به α وابسته است. در اینجا، مقدار 0.5 برای α در نظر گرفته شده است که در شرط برتری صدق می کند و همواره PRRME بهتر از PTRME رفتار می کند.

(۴) از آن جایی که کارایی های نسبی مستقل از λ می باشند، به ازای $\Delta^2 = 0$ ثابت هستند.

جدول ۱.۲: کارایی‌های نسبی برآوردگرها برای $p = ۵$ ، $\eta^2 = ۱$ ، $\alpha = ۰/۰۵$ و مقادیر مختلف λ .

Δ^2							برآوردگرها	λ
۲۰/۰۰	۱۶/۶۷	۱۳/۳۳	۱۰/۰۰	۶/۶۷	۳/۳۳	۰/۰۰		
۰/۸۸	۰/۸۲	۰/۷۵	۰/۷۲	۰/۷۷	۱/۱۱	۷/۳۸	PTRME	
۱/۰۸	۱/۱۰	۱/۱۳	۱/۱۸	۱/۲۷	۱/۵۰	۲/۵۰	SRME	۰/۰۱
۱/۰۸	۱/۱۰	۱/۱۳	۱/۲۰	۱/۳۲	۱/۶۵	۳/۵۶	PRRME	
۰/۸۸	۰/۸۱	۰/۷۴	۰/۷۰	۰/۷۴	۱/۰۷	۷/۳۸	PTRME	
۱/۰۳	۱/۰۵	۱/۰۸	۱/۱۳	۱/۲۲	۱/۴۵	۲/۵۰	SRME	۰/۰۵
۱/۰۳	۱/۰۵	۱/۰۹	۱/۴۸	۱/۲۷	۱/۵۹	۳/۵۷	PRRME	
۰/۸۷	۰/۸۰	۰/۷۳	۰/۶۸	۰/۷۲	۱/۰۲	۷/۳۸	PTRME	
۰/۹۸	۱/۰۰	۱/۰۳	۱/۰۷	۱/۱۶	۱/۳۹	۲/۵۰	SRME	۰/۱
۰/۹۸	۱/۰۰	۱/۰۳	۱/۰۹	۱/۲۱	۱/۵۳	۳/۵۷	PRRME	
۰/۸۷	۰/۷۹	۰/۷۱	۰/۶۶	۰/۶۸	۰/۹۴	۷/۳۸	PTRME	
۰/۹۱	۰/۹۲	۰/۹۴	۰/۹۸	۱/۰۷	۱/۲۹	۲/۵۰	SRME	۰/۲
۰/۹۱	۰/۹۲	۰/۹۵	۱/۰۰	۱/۱۱	۱/۴۱	۳/۵۷	PRRME	
۰/۸۹	۰/۸۲	۰/۷۲	۰/۶۴	۰/۶۲	۰/۷۹	۷/۳۸	PTRME	
۰/۸۲	۰/۸۲	۰/۸۲	۰/۸۴	۰/۸۹	۱/۰۷	۲/۵۰	SRME	۰/۵
۰/۸۲	۰/۸۲	۰/۸۳	۰/۸۵	۰/۹۲	۱/۱۴	۳/۵۷	PRRME	
۰/۹۳	۰/۸۸	۰/۸۰	۰/۷۰	۰/۶۴	۰/۷۰	۷/۳۸	PTRME	
۰/۸۴	۰/۸۳	۰/۸۲	۰/۸۱	۰/۸۱	۰/۹۰	۲/۵۰	SRME	۱
۰/۸۴	۰/۸۳	۰/۸۲	۰/۸۱	۰/۸۳	۰/۹۴	۳/۵۷	PRRME	
۰/۹۹	۰/۹۷	۰/۹۵	۰/۹۱	۰/۸۶	۰/۸۱	۷/۳۸	PTRME	
۰/۹۵	۰/۹۴	۰/۹۴	۰/۹۲	۰/۹۰	۰/۸۸	۲/۵۰	SRME	۵
۰/۹۴	۰/۹۴	۰/۹۲	۰/۹۱	۰/۸۳	۰/۸۹	۳/۵۷	PRRME	
۰/۹۹	۰/۹۹	۰/۹۸	۰/۹۶	۰/۹۳	۰/۸۹	۷/۳۸	PTRME	
۰/۹۷	۰/۹۷	۰/۹۷	۰/۹۶	۰/۹۵	۰/۹۳	۲/۵۰	SRME	۱۰
۰/۹۷	۰/۹۷	۰/۹۷	۰/۹۶	۰/۹۵	۰/۹۴	۳/۵۷	PRRME	
۱/۰۰	۰/۹۹	۰/۹۹	۰/۹۸	۰/۹۶	۰/۹۴	۷/۳۸	PTRME	
۰/۹۹	۰/۹۸	۰/۹۸	۰/۹۸	۰/۹۷	۰/۹۶	۲/۵۰	SRME	۲۰
۰/۹۹	۰/۹۸	۰/۹۸	۰/۹۸	۰/۹۷	۰/۹۶	۳/۵۷	PRRME	

(۵) جدول‌های ۱.۲ و ۲.۲ نشان می‌دهند که کارایی نسبی PRRME به‌ازای تمام مقادیر Δ^2 بهتر از/ برابر با کارایی نسبی SRME است که نتایج قضیه ۱۴.۳.۲ را تایید می‌کند. این قضیه بیان می‌کند که برای کلاس $\mathcal{K}_{(n)}$ ، این برآوردگر برای تمام مقادیر Δ^2 بهتر رفتار می‌کند اما (Δ^2) ای وجود دارد که اگر λ در بازه (Δ^2, λ_1) قرار گیرد، PRRME مخاطره کمتری از SRME دارد.

(۶) اگر جدول ۱.۲ با ۲.۲ مقایسه شود، ملاحظه می‌شود که وقتی p افزایش پیدا می‌کند، کارایی‌های نسبی بهتر می‌شود. بنابراین، برآوردگرهای پیشنهادی M رنج با افزایش p بهتر عمل می‌کند.

(۷) در جدول ۱.۲، PTRME و در جدول ۲.۲، PRRME بهتر از سایر برآوردگرها رفتار می‌کنند.

جدول ۲.۲: کارایی‌های نسبی برآوردگرها برای $p = 2$ ، $\eta^2 = 1$ ، $\alpha = 0.05$ و مقادیر مختلف λ .

Δ^2							برآوردگرها	λ
۲۰/۰۰	۱۶/۶۷	۱۳/۳۳	۱۰/۰۰	۶/۶۷	۳/۳۳	۰/۰۰		
۰/۹۵	۰/۹۹	۱/۰۷	۱/۲۶	۱/۶۸	۲/۸۷	۱۱/۳۶	PTRME	۰/۰۱
۱/۷۵	۱/۸۹	۲/۱۰	۲/۴۲	۳/۰۰	۴/۳۰	۱۰/۰۰	SRME	
۱/۷۹	۱/۹۸	۲/۲۹	۲/۹۰	۴/۲۴	۷/۸۵	۱۷/۰۲	PRRME	
۰/۹۴	۰/۹۷	۱/۰۵	۱/۲۳	۱/۶۳	۲/۷۸	۱۱/۳۶	PTRME	۰/۰۵
۱/۶۴	۱/۷۸	۱/۹۷	۲/۲۸	۲/۸۷	۴/۱۱	۱۰/۰۰	SRME	
۱/۶۷	۱/۸۵	۲/۱۵	۲/۷۰	۳/۹۵	۷/۲۷	۱۷/۰۲	PRRME	
۰/۹۳	۰/۹۵	۱/۰۲	۱/۱۹	۱/۵۷	۲/۶۸	۱۱/۳۶	PTRME	۰/۱
۱/۵۲	۱/۶۵	۱/۸۳	۲/۱۳	۲/۶۵	۳/۸۹	۱۰/۰۰	SRME	
۱/۵۵	۱/۷۱	۱/۹۸	۲/۴۹	۳/۶۲	۶/۶۴	۱۷/۰۲	PRRME	
۰/۹۰	۰/۹۲	۰/۹۷	۱/۱۲	۱/۴۶	۲/۴۸	۱۱/۳۶	PTRME	۰/۲
۱/۳۲	۱/۴۳	۱/۶۰	۱/۸۶	۲/۳۳	۳/۴۹	۱۰/۰۰	SRME	
۱/۳۴	۱/۴۸	۱/۷۱	۲/۱۴	۳/۰۸	۵/۵۹	۱۷/۰۲	PRRME	
۰/۸۵	۰/۸۴	۰/۸۷	۰/۹۶	۱/۲۱	۲/۰۰	۱۱/۳۶	PTRME	۰/۵
۰/۹۸	۱/۰۵	۱/۱۶	۱/۳۴	۱/۶۹	۲/۵۹	۱۰/۰۰	SRME	
۰/۹۹	۱/۰۸	۱/۲۲	۱/۴۹	۲/۰۶	۳/۶۲	۱۷/۰۲	PRRME	
۰/۸۹	۰/۸۱	۰/۸۰	۰/۸۳	۰/۹۸	۱/۵۰	۱۱/۳۶	PTRME	۱
۰/۸۰	۰/۸۴	۰/۸۰	۰/۹۹	۱/۱۹	۱/۷۸	۱۰/۰۰	SRME	
۰/۸۱	۰/۸۵	۰/۹۲	۱/۰۶	۱/۳۵	۲/۱۸	۱۷/۰۲	PRRME	
۰/۹۴	۰/۹۱	۰/۸۸	۰/۸۵	۰/۸۴	۰/۸۸	۱۱/۳۶	PTRME	۵
۰/۸۶	۰/۸۵	۰/۸۵	۰/۸۴	۰/۸۵	۰/۹۱	۱۰/۰۰	SRME	
۰/۸۶	۰/۸۶	۰/۸۵	۰/۸۶	۰/۸۷	۰/۹۴	۱۷/۰۲	PRRME	
۰/۹۷	۰/۹۵	۰/۹۳	۰/۹۱	۰/۹۰	۰/۹۰	۱۱/۳۶	PTRME	۱۰
۰/۹۲	۰/۹۱	۰/۹۱	۰/۹۰	۰/۹۰	۰/۹۱	۱۰/۰۰	SRME	
۰/۹۲	۰/۹۲	۰/۹۱	۰/۹۱	۰/۹۱	۰/۹۲	۱۷/۰۲	PRRME	
۰/۹۸	۰/۹۷	۰/۹۶	۰/۹۵	۰/۹۴	۰/۹۳	۱۱/۳۶	PTRME	۲۰
۰/۹۶	۰/۹۵	۰/۹۵	۰/۹۴	۰/۹۴	۰/۹۴	۱۰/۰۰	SRME	
۰/۹۶	۰/۹۵	۰/۹۵	۰/۹۵	۰/۹۵	۰/۹۴	۱۷/۰۲	PRRME	

۵.۳.۲ مطالعه شبیه‌سازی

در این بخش، برای حالت غیرمتعامد، مطالعه شبیه‌سازی به‌کار گرفته شده است تا کارایی برآوردگرهای پیشنهادی بررسی شود.

برای به‌دست آوردن برآوردگر M، از تابع هوبر به‌صورت زیر استفاده کردیم

$$\rho(\varepsilon) = \begin{cases} \varepsilon^2; & -k \leq \varepsilon \leq k \\ 2k|\varepsilon| - k^2 & \varepsilon < -k \text{ یا } \varepsilon > k \end{cases}$$

که به‌طور معادل

$$\psi(\varepsilon) = \begin{cases} \varepsilon & |\varepsilon| \leq k \\ k \operatorname{sgn}(\varepsilon) - \frac{1}{2}k^2 & |\varepsilon| > k \end{cases}$$

مقادیر مختلفی را می‌توان برای k انتخاب نمود. به ازای مقادیر بزرگ k ، برای بیشتر باقیمانده‌های مشاهده شده، $\rho(\varepsilon) = \varepsilon^2$ است و هم‌چنین برآوردگرهای M نزدیک به برآوردگرهای OLS هستند.

به ازای مقادیر کوچک k ، برای بیشتر باقیمانده‌های مشاهده شده، $\rho(\varepsilon) = 2k|\varepsilon| - k^2$ است و با توجه به اینکه کمینه کردن عبارت $\sum_{i=1}^n (2k|\varepsilon_i| - k^2)$ با کمینه‌سازی عبارت $\sum_{i=1}^n |\varepsilon_i|$ معادل است، برآوردگرهای M متناظر نزدیک به برآوردگرهای LAD هستند. مقدار k ای که بهترین برآوردها را برای ضرایب رگرسیونی نتیجه می‌دهد وابسته به توزیع خطاهای تصادفی است. اگر این توزیع، نقاط دورافتاده زیادی را تولید کند، آن‌گاه مقدار کوچکی برای k باید انتخاب شود در حالی که اگر توزیع، نقاط دورافتاده اندکی را نتیجه دهد (شبیه توزیع نرمال) در این صورت بهتر است مقدار بزرگی برای k در نظر گرفته شود.

هوبر با احتیاط کردن در مقابل انتخاب دو مقدار ممکن، مقدار $k = 1/5\hat{\sigma}$ را پیشنهاد می‌کند که $\hat{\sigma}$ برآوردگر انحراف استاندارد جامعه خطای تصادفی است. **هالند و ولش (۱۹۷۷)** مقدار $k = 1/345\hat{\sigma}$ را پیشنهاد نموده‌اند. آنها مقدار k را به گونه‌ای انتخاب کرده‌اند که اگر توزیع این خطاها نرمال باشد، برآوردگرهای M دارای کارایی جانبی ۹۵٪ می‌باشند. معمولاً فرض می‌شود $k = 1/345$.

در ابتدا بر اساس روش پیشنهادی **مک‌دونالد و گالارنو (۱۹۷۵)** برای ایجاد هم‌خطی، ماتریس طرح به صورت زیر تولید می‌شود:

$$X_{ij} = \sqrt{1 - \rho^2} Z_{ij} + \rho Z_{ip}, \quad i = 1, \dots, n; j = 1, \dots, p - 1, \quad (53.2)$$

که در آن Z_{ij} ها متغیرهای شبه‌تصادفی نرمال استاندارد هستند و ρ همبستگی بین دو متغیر توضیحی دلخواه است. هرچه مقدار ρ بیشتر باشد، وابستگی بین متغیرهای توضیحی نیز بیشتر می‌شود (هم‌خطی بیشتر).

در این جا، پارامتر β بردار p تایی و تابعی از Δ^2 است و به گونه‌ای تعریف می‌شود که $\Delta^2 = 0$ نشان می‌دهد مجموعه داده بر اساس فرضیه صفر ($\beta = 0$) تولید شده‌اند. به عنوان مثال، فرض کنید برای به دست آوردن متغیر پاسخ از $\beta = (0, 0, 0, 0, \Delta^2)^T$ استفاده شده است. به وضوح مشاهده می‌شود که Δ^2 مشابه درجه بی‌اعتمادی به اطلاع پیشین (انحراف از فرضیه صفر) عمل می‌کند. هرچه Δ^2 افزایش یابد، درجه انحراف از فرضیه صفر نیز افزایش می‌یابد. کارایی برآوردگرهای پیشنهادی بر اساس تغییرات Δ^2 مطالعه شده است.

متغیر پاسخ بر اساس مدل (۱.۱) شبیه‌سازی شده است که در آن خطاها از یک توزیع نرمال آمیخته به صورت زیر آمده است:

$$\rho \mathcal{N}_n(0, c^* \sigma^2) + (1 - \rho) \mathcal{N}_n(0, \sigma^2), \quad c^* > 1$$

در نتیجه باقیمانده‌ها با احتمال $1 - \rho$ از توزیع نرمال با واریانس $\sigma^2 = 1$ تولید شده‌اند و با احتمال ρ از توزیع نرمال با واریانس $c^* \sigma^2$ تولید می‌شود.

با انتخاب مقدار بزرگ برای c^* و مقدار کوچک برای ρ ، بیشتر مشاهده‌ها از توزیع نرمال با واریانس مشخص σ^2 و درصد کمی از توزیع نرمال با واریانس بزرگ $c^*\sigma^2$ تولید می‌شوند. ρ ، درصد داده‌های پرت تفسیر می‌شود.

در فرآیند شبیه‌سازی، MSE برای هر یک از برآوردگرهای پیشنهادی به دست آمده است و سپس کارایی نسبی آنها با تقسیم MSE برآوردگر محدود نشده M ریج بر MSE برآوردگرهای پیشنهاد شده محاسبه می‌شود.

برای N تکرار، MSE نسبی به صورت زیر به دست می‌آید:

$$\text{RMSE}(\hat{\beta}^{*M}(\lambda)) = \frac{\text{MSE}(\hat{\beta}^{\text{uM}}(\lambda))}{\text{MSE}(\hat{\beta}^{*M}(\lambda))} = \frac{\frac{1}{N} \sum_{i=1}^N (\hat{\beta}^{\text{uM}^{(i)}}(\lambda) - \beta)^\top (\hat{\beta}^{\text{uM}^{(i)}}(\lambda) - \beta)}{\frac{1}{N} \sum_{i=1}^N (\hat{\beta}^{*M^{(i)}}(\lambda) - \beta)^\top (\hat{\beta}^{*M^{(i)}}(\lambda) - \beta)},$$

که در آن $\hat{\beta}^{\text{uM}^{(i)}}(\lambda)$ همان برآوردگر URM در مرحله i ام است و $\hat{\beta}^{*M}(\lambda)$ می‌تواند هر یک از برآوردگرهای پیشنهادی باشد. MSE نسبی بزرگتر از یک، برتری برآوردگر پیشنهادی را نسبت به برآوردگر محدود نشده نشان می‌دهد. به عبارت دیگر، RMSE کمتر یا مساوی یک، می‌تواند نشان دهد که برآوردگر پیشنهادی نسبت به برآوردگر URM رفتار بهتری ندارد.

فرآیند شبیه‌سازی مطرح شده $N = 1000$ بار به‌ازای حجم نمونه $n = 50$ ، تعداد پارامترهای $p = 5, 10, 20$ ، همبستگی $\rho = 0.8, 0.9, 0.99, 0.999$ ، سطح معنی‌داری $\alpha = 0.1$ و $c^* = 100$ در توزیع نرمال آمیخته انجام می‌شود. برای هر ρ کارایی نسبی به‌ازای درصد داده‌های پرت $\rho = 0.1, 0.2$ برای ده مقدار مختلف از Δ^2 ، $0, 0.01, 0.1, 0.5, 1, 2, 5, 10, 20, 30$ ، محاسبه شده است که نتایج در جدول‌های ۳.۲-۵.۲ گزارش شده است.

نتایج کلی از جدول‌های ۳.۲-۵.۲ به صورت زیر خلاصه شده است:

(۱) رفتار برآوردگرهای PRRM و SRM کاملاً پایدار است و معمولاً بر اساس معیار RMSE، به URME برتری دارد.

(۲) هر چه p افزایش یابد، کارایی برآوردگرهای SRM و PRRM نسبت به URME بهتر می‌شود.

(۳) با افزایش مقدار ρ ، معمولاً برآوردگرهای بهبودیافته M ریج کارایی بهتری دارند.

(۴) وقتی ρ کوچک است، با افزایش درجه انحراف از فرضیه صفر، ملاحظه می‌شود که URME بهترین برآوردگر خواهد بود. با این وجود، اگر ρ بزرگتر شود، رفتار برآوردگرهای SRM و PRRM بهتر می‌شود.

(۵) هنگامی که از فرضیه صفر فاصله گرفته شود، برتری برآوردگرهای RRM و PTRM نسبت به URME کاهش می‌یابد، با این وجود، رفتار برآوردگرهای SRM و PRRM پایدار است.

(۶) از جدول‌های ۳.۲-۵.۲ چنین برداشت می‌شود که وقتی n ثابت است و p زیاد می‌شود، کارایی برآوردگرها بهتر می‌شود.

جدول ۳.۲: کارایی‌های نسبی برآوردگرهای پیشنهادی به‌ازای $n = 50$ ، $p = 10$ و $\alpha = 0.1$

۲۰٪ داده‌ی پرت			۱۰٪ داده‌ی پرت			Δ^2	ρ^2
PRSRME	SRME	PTRME	PRSRME	SRME	PTRME		
۷/۱۹	۶/۲۲	۵/۱۸	۶/۶۴	۵/۹۷	۴/۵۵	۰	
۷/۰۳	۶/۰۶	۴/۷۱	۶/۳۶	۵/۸۰	۳/۹۶	۰/۰۱	
۷/۱۵	۶/۱۶	۴/۹۰	۶/۶۰	۵/۹۰	۴/۴۸	۰/۱	
۷/۰۹	۶/۲۱	۵/۱۲	۶/۲۷	۵/۷۳	۴/۰۸	۰/۵	
۷/۰۸	۶/۱۴	۴/۸۸	۶/۱۰	۵/۵۶	۳/۸۰	۱	۰/۸
۶/۹۹	۶/۰۶	۴/۸۵	۵/۹۰	۵/۴۰	۳/۸۱	۲	
۶/۹۲	۶/۰۴	۴/۸۲	۵/۳۱	۴/۸۹	۳/۵۴	۵	
۶/۷۲	۵/۹۲	۴/۴۲	۴/۶۱	۴/۳۱	۳/۰۷	۱۰	
۶/۲۱	۵/۵۱	۴/۲۶	۳/۶۳	۳/۴۶	۲/۳۵	۲۰	
۵/۸۲	۵/۱۸	۳/۸۲	۳/۰۶	۲/۹۴	۱/۹۸	۳۰	
۶/۹۶	۶/۱۰	۴/۸۳	۶/۵۰	۵/۸۳	۴/۲۲	۰	
۷/۱۲	۶/۱۸	۵/۰۲	۶/۵۷	۵/۹۳	۴/۴۵	۰/۰۱	
۷/۰۷	۶/۱۶	۴/۸۷	۶/۵۹	۵/۹۲	۴/۴۱	۰/۱	
۷/۰۹	۶/۱۳	۴/۹۰	۶/۴۹	۵/۹۵	۴/۴۴	۰/۵	
۶/۹۴	۶/۰۱	۴/۷۰	۶/۵۲	۵/۸۷	۴/۴۰	۱	۰/۹
۶/۹۹	۶/۰۹	۴/۷۲	۶/۳۲	۵/۷۰	۴/۲۰	۲	
۶/۸۲	۶/۰۱	۴/۶۷	۵/۷۵	۵/۲۷	۳/۵۹	۵	
۶/۹۶	۶/۰۴	۴/۸۴	۵/۱۷	۴/۸۴	۳/۲۶	۱۰	
۶/۴۸	۵/۸۲	۴/۳۹	۴/۲۹	۴/۱۰	۲/۷۰	۲۰	
۶/۲۷	۵/۶۲	۴/۲۳	۳/۷۶	۳/۶۳	۲/۳۴	۳۰	
۷/۰۱	۶/۱۰	۴/۷۰	۶/۳۰	۵/۶۹	۳/۸۷	۰	
۶/۷۱	۵/۸۸	۴/۳۷	۶/۴۷	۵/۸۵	۴/۳۷	۰/۰۱	
۶/۸۲	۵/۹۸	۴/۵۲	۶/۵۵	۵/۹۴	۴/۳۱	۰/۱	
۶/۸۸	۵/۹۸	۴/۶۵	۶/۵۵	۵/۸۸	۴/۳۳	۰/۵	
۷/۰۰	۶/۰۳	۴/۷۵	۶/۵۲	۵/۹۴	۴/۲۸	۱	۰/۹۹
۷/۳۴	۶/۳۳	۵/۲۴	۶/۲۰	۵/۶۹	۳/۸۷	۲	
۶/۹۶	۶/۱۵	۴/۸۱	۶/۰۵	۵/۵۷	۳/۷۹	۵	
۷/۰۵	۶/۰۲	۴/۷۷	۵/۸۷	۵/۴۷	۳/۶۲	۱۰	
۶/۶۵	۵/۹۴	۴/۴۱	۵/۲۱	۴/۹۷	۳/۰۰	۲۰	
۶/۵۴	۵/۸۶	۴/۳۱	۴/۹۳	۴/۷۶	۲/۸۰	۳۰	
۶/۸۸	۶/۰۷	۴/۷۴	۶/۶۲	۶/۰۱	۴/۴۱	۰	
۶/۹۲	۶/۰۷	۴/۶۱	۶/۴۵	۵/۸۸	۴/۱۷	۰/۰۱	
۶/۸۱	۵/۹۸	۴/۵۲	۶/۴۹	۵/۸۲	۴/۲۶	۰/۱	
۷/۰۸	۶/۲۲	۴/۷۵	۶/۴۹	۵/۸۹	۴/۱۰	۰/۵	
۷/۲۵	۶/۱۶	۵/۲۰	۶/۵۱	۵/۸۹	۴/۲۵	۱	۰/۹۹۹
۷/۰۵	۶/۱۶	۴/۸۴	۶/۲۵	۵/۷۲	۳/۸۲	۲	
۶/۹۷	۶/۰۸	۴/۹۶	۶/۱۶	۵/۶۷	۳/۹۲	۵	
۷/۱۱	۶/۲۵	۵/۰۴	۵/۷۵	۵/۳۷	۳/۳۷	۱۰	
۶/۶۸	۵/۹۷	۴/۲۱	۵/۴۱	۵/۱۳	۳/۱۲	۲۰	
۶/۳۷	۵/۷۵	۴/۰۷	۵/۱۳	۴/۹۳	۲/۸۵	۳۰	

جدول ۴.۲: کارایی‌های نسبی برآوردگرهای پیشنهادی به‌ازای $n = 50$ ، $p = 20$ و $\alpha = 0.1$

۲۰٪ داده‌ی پرت			۱۰٪ داده‌ی پرت			Δ^2	ρ^2
PRSRME	SRME	PTRME	PRSRME	SRME	PTRME		
۲۰/۲۶	۱۵/۴۳	۱۱/۰۳	۱۹/۳۵	۱۵/۸۱	۱۰/۵۶	۰	
۱۹/۹۵	۱۵/۵۱	۱۰/۳۱	۱۹/۶۰	۱۵/۵۲	۱۰/۸۴	۰/۰۱	
۲۰/۸۹	۱۵/۶۳	۱۱/۲۱	۱۹/۰۵	۱۵/۵۹	۱۰/۴۷	۰/۱	
۲۰/۲۶	۱۶/۰۹	۱۱/۵۸	۱۹/۸۸	۱۵/۷۱	۱۰/۶۱	۰/۵	
۱۹/۵۶	۱۵/۳۷	۱۰/۴۹	۱۸/۳۳	۱۴/۸۰	۱۰/۳۲	۱	۰/۸
۱۹/۵۳	۱۵/۳۷	۱۰/۲۷	۱۶/۱۸	۱۳/۶۷	۸/۶۰	۲	
۱۷/۱۹	۱۳/۹۷	۹/۱۸	۹/۵۳	۸/۶۳	۵/۶۱	۵	
۱۲/۰۳	۱۰/۶۸	۶/۶۸	۴/۱۴	۴/۰۶	۲/۲۷	۱۰	
۵/۷۸	۵/۶۵	۲/۷۵	۱/۸۲	۱/۸۲	۱/۰۵	۲۰	
۳/۴۹	۳/۴۸	۱/۵۰	۱/۲۹	۱/۲۹	۱/۰۰	۳۰	
۲۰/۱۱	۱۵/۸۳	۱۰/۲۹	۱۸/۲۴	۱۴/۸۰	۸/۶۲	۰	
۲۰/۳۳	۱۵/۷۲	۱۰/۷۶	۱۹/۰۱	۱۵/۵۰	۹/۶۲	۰/۰۱	
۲۰/۷۱	۱۶/۲۳	۱۲/۲۸	۱۸/۸۶	۱۵/۳۸	۱۰/۱۶	۰/۱	
۲۱/۰۷	۱۵/۷۹	۱۱/۶۲	۱۹/۱۶	۱۵/۵۹	۹/۷۳	۰/۵	
۲۰/۵۲	۱۵/۲۲	۱۰/۹۴	۱۸/۶۷	۱۵/۱۸	۹/۶۷	۱	۰/۹
۲۰/۳۵	۱۵/۵۵	۱۱/۵۱	۱۷/۶۸	۱۴/۷۰	۹/۶۶	۲	
۱۷/۸۴	۱۴/۱۷	۹/۰۱	۱۱/۶۵	۱۰/۵۰	۵/۸۳	۵	
۱۴/۱۵	۱۲/۱۳	۷/۲۳	۵/۸۵	۵/۷۳	۲/۷۳	۱۰	
۷/۴۹	۷/۳۰	۳/۰۳	۲/۵۱	۲/۵۱	۱/۱۲	۲۰	
۴/۶۸	۴/۶۷	۱/۶۹	۱/۷۲	۱/۷۲	۱/۰۰	۳۰	
۲۰/۶۹	۱۵/۵۸	۱۰/۹۸	۱۸/۶۹	۱۵/۱۸	۱۰/۲۵	۰	
۱۹/۸۱	۱۵/۵۳	۱۰/۷۶	۱۹/۲۳	۱۵/۲۶	۹/۹۱	۰/۰۱	
۲۰/۳۰	۱۵/۳۲	۱۰/۴۴	۱۸/۹۵	۱۵/۳۱	۹/۶۲	۰/۱	
۲۱/۰۰	۱۵/۹۰	۱۱/۵۱	۱۹/۲۳	۱۵/۱۴	۹/۸۸	۰/۵	
۲۱/۰۰	۱۵/۳۰	۱۰/۹۸	۱۹/۳۰	۱۵/۵۱	۹/۸۴	۱	۰/۹۹
۲۰/۴۲	۱۵/۸۹	۱۰/۹۲	۱۸/۲۱	۱۵/۱۱	۹/۳۳	۲	
۱۸/۹۲	۱۵/۰۸	۹/۸۸	۱۵/۱۷	۱۳/۴۱	۶/۷۱	۵	
۱۶/۸۱	۱۴/۳۱	۸/۷۵	۹/۸۵	۹/۶۰	۳/۴۴	۱۰	
۱۰/۴۷	۱۰/۱۵	۳/۷۶	۴/۶۹	۴/۶۹	۱/۲۲	۲۰	
۶/۹۹	۶/۹۷	۱/۹۴	۳/۲۸	۳/۲۸	۱/۰۱	۳۰	
۲۰/۰۹	۱۵/۴۹	۱۰/۱۳	۱۹/۷۱	۱۵/۷۳	۱۰/۸۸	۰	
۲۰/۶۹	۱۵/۵۵	۱۱/۶۵	۱۹/۳۰	۱۵/۵۲	۱۰/۵۶	۰/۰۱	
۱۹/۵۵	۱۵/۳۲	۱۰/۵۲	۱۹/۶۰	۱۵/۳۴	۱۱/۳۸	۰/۱	
۲۰/۹۷	۱۶/۱۹	۱۲/۸۹	۱۹/۸۱	۱۵/۴۸	۱۱/۰۰	۰/۵	
۲۰/۶۸	۱۵/۸۶	۱۱/۸۷	۱۹/۴۷	۱۵/۶۰	۱۰/۰۹	۱	۰/۹۹۹
۲۰/۷۴	۱۵/۹۰	۱۱/۰۱	۱۸/۹۰	۱۵/۴۶	۹/۵۰	۲	
۱۹/۰۴	۱۵/۲۴	۹/۳۴	۱۵/۷۲	۱۴/۰۶	۶/۷۱	۵	
۱۷/۰۱	۱۴/۵۱	۸/۳۰	۱۰/۳۳	۱۰/۱۲	۳/۳۲	۱۰	
۱۰/۹۹	۱۰/۶۹	۳/۶۴	۵/۲۳	۵/۲۳	۱/۲۲	۲۰	
۷/۳۶	۷/۳۴	۱/۹۳	۳/۷۸	۳/۷۸	۱/۰۱	۳۰	

جدول ۵.۲: کارایی‌های نسبی برآوردگرهای پیشنهادی به‌ازای $n = 5$ ، $p = 5$ و $\alpha = 0.1$

۱۰٪ داده‌ی پرت			۵٪ داده‌ی پرت			Δ^2	ρ^2
PRSRME	SRME	PTRME	PRSRME	SRME	PTRME		
۲/۹۴	۲/۷۹	۳/۱۷	۲/۷۴	۲/۶۵	۲/۷۵	۰	
۲/۹۲	۲/۷۹	۳/۱۰	۲/۷۴	۲/۶۴	۲/۷۴	۰/۰۱	
۲/۹۲	۲/۷۷	۳/۲۳	۲/۶۸	۲/۶۰	۲/۵۹	۰/۱	
۲/۹۷	۲/۸۳	۳/۲۷	۲/۷۸	۲/۶۸	۲/۸۹	۰/۵	
۲/۹۰	۲/۷۵	۳/۱۶	۲/۶۲	۲/۵۳	۲/۵۵	۱	۰/۸
۲/۹۲	۲/۷۷	۳/۳۲	۲/۴۲	۲/۳۵	۲/۲۱	۲	
۲/۵۳	۲/۴۴	۲/۳۹	۱/۶۷	۱/۶۴	۱/۳۲	۵	
۱/۹۹	۱/۹۵	۱/۵۸	۱/۱۷	۱/۱۷	۰/۹۲	۱۰	
۱/۳۶	۱/۳۶	۰/۹۹	۱/۰۱	۱/۰۱	۰/۹۹	۲۰	
۱/۱۷	۱/۱۷	۰/۹۶	۰/۹۷	۰/۹۷	۱/۰۰	۳۰	
۲/۸۸	۲/۷۵	۳/۱۱	۲/۷۲	۲/۶۴	۲/۶۶	۰	
۲/۹۳	۲/۷۹	۳/۱۸	۲/۷۱	۲/۶۳	۲/۷۸	۰/۰۱	
۲/۹۱	۲/۷۴	۳/۱۰	۲/۷۵	۲/۶۵	۲/۷۹	۰/۱	
۲/۹۲	۲/۷۹	۳/۱۴	۲/۶۷	۲/۵۵	۲/۶۵	۰/۵	
۲/۸۹	۲/۷۵	۳/۰۸	۲/۶۸	۲/۶۱	۲/۷۶	۱	۰/۹
۲/۸۲	۲/۶۹	۲/۹۰	۲/۵۰	۲/۴۲	۲/۳۴	۲	
۲/۶۱	۲/۵۲	۲/۵۷	۱/۹۱	۱/۸۸	۱/۵۷	۵	
۲/۱۷	۲/۱۳	۱/۸۰	۱/۳۳	۱/۳۲	۱/۰۰	۱۰	
۱/۵۲	۱/۵۲	۱/۱۰	۱/۰۶	۱/۰۶	۱/۰۰	۲۰	
۱/۲۶	۱/۲۶	۰/۹۹	۱/۰۱	۱/۰۱	۱/۰۰	۳۰	
۲/۸۲	۲/۶۸	۲/۹۸	۲/۷۲	۲/۶۳	۲/۸۲	۰	
۲/۸۹	۲/۷۸	۳/۱۸	۲/۶۹	۲/۶۲	۲/۵۶	۰/۰۱	
۲/۸۵	۲/۷۴	۳/۰۲	۲/۷۰	۲/۶۲	۲/۷۵	۰/۱	
۲/۸۸	۲/۷۷	۳/۱۰	۲/۷۰	۲/۶۲	۲/۶۰	۰/۵	
۲/۸۹	۲/۷۸	۳/۱۶	۲/۶۸	۲/۶۰	۲/۶۴	۱	۰/۹۹
۲/۸۷	۲/۷۵	۳/۰۱	۲/۶۸	۲/۶۲	۲/۶۹	۲	
۲/۷۷	۲/۶۹	۲/۸۹	۲/۲۵	۲/۲۳	۱/۸۶	۵	
۲/۴۰	۲/۳۴	۲/۰۸	۱/۶۸	۱/۶۸	۱/۱۸	۱۰	
۱/۷۸	۱/۷۸	۱/۲۴	۱/۲۷	۱/۲۷	۱/۰۰	۲۰	
۱/۴۵	۱/۴۵	۱/۰۳	۱/۱۵	۱/۱۵	۱/۰۰	۳۰	
۲/۸۸	۲/۷۶	۳/۱۷	۲/۷۰	۲/۶۱	۲/۶۷	۰	
۲/۹۰	۲/۷۸	۳/۱۲	۲/۶۹	۲/۶۲	۲/۶۷	۰/۰۱	
۲/۸۷	۲/۷۶	۲/۹۶	۲/۶۹	۲/۶۰	۲/۶۸	۰/۱	
۲/۸۸	۲/۷۷	۳/۰۷	۲/۶۲	۲/۵۵	۲/۴۷	۰/۵	
۲/۸۶	۲/۷۲	۳/۰۱	۲/۶۷	۲/۶۱	۲/۶۷	۱	۰/۹۹۹
۲/۹۰	۲/۷۹	۳/۱۳	۲/۶۷	۲/۶۰	۲/۶۷	۲	
۲/۷۳	۲/۶۴	۲/۷۶	۲/۳۰	۲/۲۸	۱/۹۵	۵	
۲/۴۱	۲/۳۷	۲/۰۷	۱/۷۶	۱/۷۶	۱/۲۲	۱۰	
۱/۸۰	۱/۷۹	۱/۲۷	۱/۳۳	۱/۳۳	۱/۰۰	۲۰	
۱/۴۸	۱/۴۸	۱/۰۴	۱/۲۳	۱/۲۳	۱/۰۰	۳۰	

(۷) برآوردگر PRRM بر SRME به طور یکنواخت برتری دارد.

(۸) می توان به ازای n و ρ های کوچک، تحت فرضیه صفر، ترتیب برتری برآوردگرها را به صورت زیر خلاصه نمود

$$RRME \succ PTRME \succ PRSRME \succ SRME$$

به عبارت دیگر، تحت فرضیه مقابل، این رابطه به صورت زیر خواهد بود

$$PRSRME \succ SRME \succ PTRME \succ RRME$$

که \succ برتری را نشان می دهد.

(۹) به ازای ρ در نظر گرفته شده، کارایی نسبی برآوردگرهای بهبودیافته M ریج یک تابع نزولی از Δ^2 است. در این حالت، هرگاه Δ^2 افزایش یابد، کارایی نسبی کاهش می یابد.

۶.۳.۲ تحلیل مثال های کاربردی

در مثال های کاربردی، با توجه به اینکه مقدار واقعی پارامتر مجهول است، باید از چندین معیار برای مقایسه برآوردگرهای پیشنهادی استفاده گردد. در این جا دو معیار به کار گرفته شده اند: (الف) خطای پیشگویی اعتبارسنجی متقابل و (ب) خطای استاندارد بوت استرپ.

در اعتبارسنجی متقابل، خطای پیشگویی (PE) برآورد شده در اجراهای مختلف تغییر می کند. بنابراین فرآیند ۱۰۰۰ بار تکرار می شود، میانگین خطاهای پیشگویی محاسبه می شوند و خطای پیشگویی نسبی (RPE) به صورت

$$RPE(\hat{\beta}^{SM+}(\lambda)) = \frac{\text{Ave}(\text{PE}(\hat{\beta}_n^{uM}(\lambda)))}{\text{Ave}(\text{PE}(\hat{\beta}^{SM+}(\lambda)))}$$

تعریف می شود که در آن Ave میانگین را نشان می دهد. تعداد تکرار ۱۰۰۰ به اندازه کافی بزرگ است تا انحراف های استاندارد را ثابت نگه دارد و تغییرات قابل توجهی مشاهده نشود.

ایده اصلی روش های بوت استرپ بر این اصل بنا شده است که استنباط درباره یک جامعه بر اساس نمونه را می توان با استفاده از باز نمونه گیری از نمونه و استنباط روی مشاهدات باز نمونه گیری مدل بندی کرد.

فرض کنید $\hat{\beta}_b^{*M}(\lambda)$ برآوردگر دلخواه پیشنهادی برای b امین نمونه بوت استرپی باشد. برآوردگر بوت استرپ از B نمونه بوت استرپی به صورت زیر محاسبه می شود:

$$\hat{\beta}_B^{*M}(\lambda) = \frac{1}{B} \sum_{b=1}^B \hat{\beta}_b^{*M}(\lambda) \quad (54.2)$$

انحراف استاندارد بوت استرپ برآوردگر $\hat{\beta}_B^{*M}(\lambda)$ به صورت زیر به دست می آید.

$$\widehat{s.e.}_B = \sqrt{\frac{1}{B-1} \left(\hat{\beta}_B^{*M}(\lambda) - \hat{\beta}^{*M}(\lambda) \right)^T \left(\hat{\beta}_B^{*M}(\lambda) - \hat{\beta}^{*M}(\lambda) \right)}$$

در اینجا، برای $B = 1000$ تکرار، $\widehat{s.e.}_B$ محاسبه شده است تا به برآوردگرهای بهتر دست یافت.

تحلیل داده‌های V02

مجموعه داده‌های V02^۱ مربوط به مطالعه تعیین تناسب اندام فیزیکی است. داده‌ها شامل اطلاعات مربوط به ۲۳۳ مشاهده است. در این مطالعه، در مجموع ۷ متغیر وجود دارد که یکی از آنها به‌عنوان متغیر وابسته و ۶ متغیر توضیحی با وجود هم‌خطی و داده‌های پرت در نظر گرفته شده است. در این مجموعه داده، متغیر وابسته: V02، بیشینه مصرف اکسیژنی است که با آزمون روی تردمیل اندازه‌گیری شده است (ml/kg/min)،

متغیرهای توضیحی: AGE، سن (سال)، MHR، بیشینه ضربان قلب (ضربه/دقیقه)، DUR، مدت زمان آزمون تردمیل (ثانیه)، MXS، بیشینه فشار خون سیاستولیک روی آزمون تردمیل (mmHg)، MXD، بیشینه فشارخون دیاستولیک روی آزمون تردمیل (mmHg)، FAI، اختلال هوایی کاربردی (درصد نسبی به سن و جنسیت).

جدول ۶.۲ با در نظر گرفتن ماتریس همبستگی متغیرهای توضیحی به‌صورت شکل ۵.۲، آمار توصیفی متغیرها را نشان می‌دهد.

جدول ۶.۲: آماره‌های توصیفی برای مجموعه داده V02

متغیرها	کمینه	چارک اول	میانه	میانگین	چارک سوم	بیشینه
AGE	۳۴/۰۰	۴۵/۰۰	۵۲/۰۰	۵۱/۷۶	۵۸/۰۰	۷۳/۰۰
MHR	۹۴/۰	۱۴۰/۰	۱۶۲/۰	۱۵۸/۳	۱۷۸/۰	۲۰۰/۰
DUR	۳۶/۰	۲۸۹/۰	۴۷۸/۰	۴۴۴/۸	۵۹۹/۰	۷۹۸/۰
MXS	۱۰۵/۰	۱۵۰/۰	۱۷۵/۰	۱۷۵/۶	۱۹۶/۰	۲۹۰/۰
MXD	۳۰/۰	۷۰/۰	۸۰/۰۰	۷۸/۶	۹۰/۰	۱۵۲/۰
FAI	-۲۲/۰۰	-۱/۰۰	۱۲/۰۰	۱۷/۹۹	۳۷/۰۰	۷۱/۰۰
V02	۴/۷۰	۲۰/۱۰	۳۰/۶۰	۲۹/۱۷	۳۷/۳۰	۵۹/۷۰

از شکل ۵.۲، می‌توان حدس زد که هم‌خطی وجود دارد. مقادیر VIF در جدول ۷.۲ نشان داده شده است.

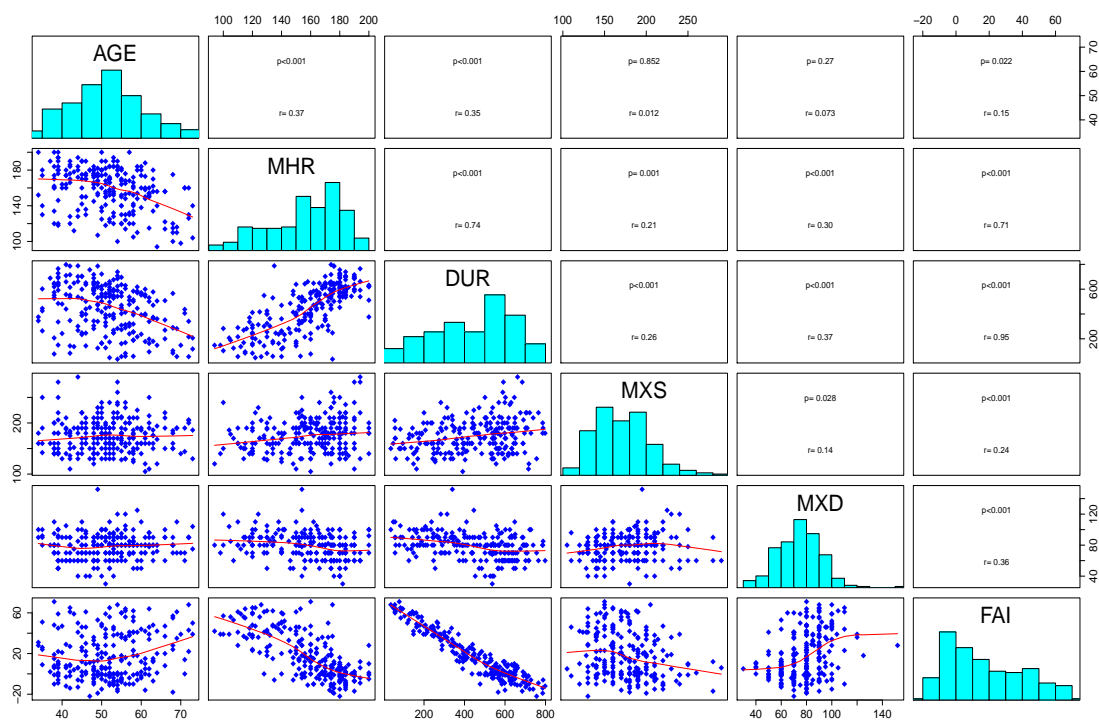
جدول ۷.۲: مقادیر VIF برای مجموعه داده V02

متغیرها	AGE	MHR	DUR	MXS	MXD	FAI
VIF	۲/۰۷۰۹	۲/۴۰۱۳	۱۸/۸۴۸۸	۱/۲۰۸۲	۱/۲۷۰۲	۱۶/۹۶۱۱

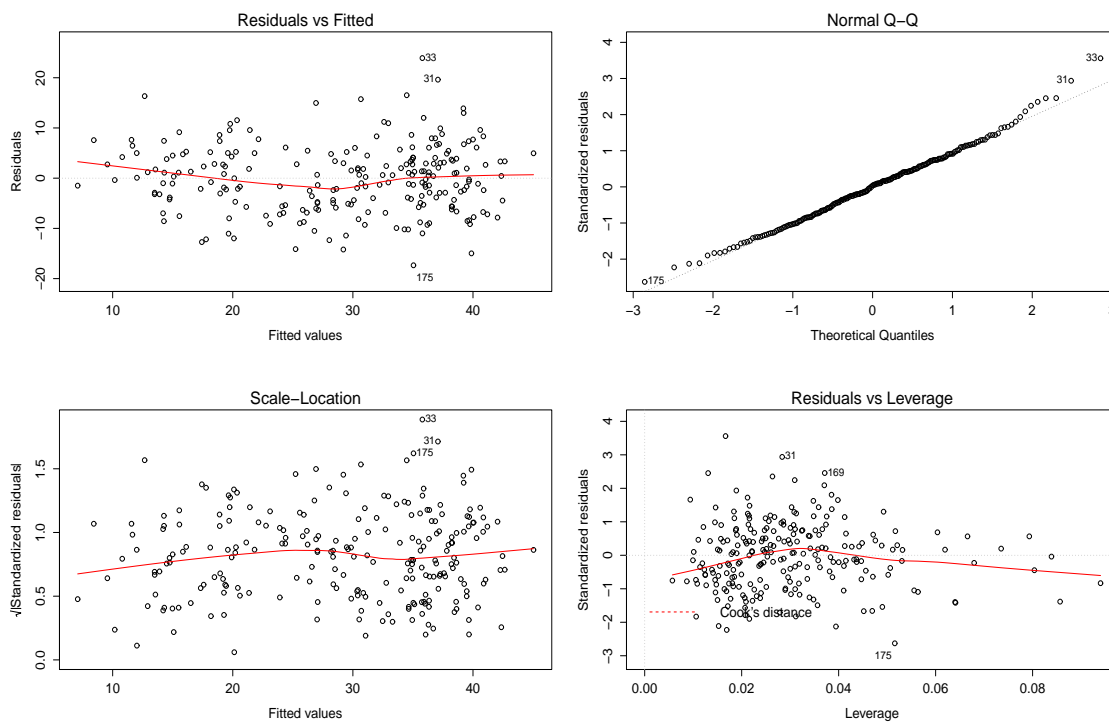
مقادیر جدول ۷.۲ و مقدار شرطی $\kappa = ۳۴۵۱/۷۲۸$ نشان می‌دهد که بین متغیرهای توضیحی، هم‌خطی وجود دارد.

شکل ۶.۲ نشان می‌دهد مشاهده‌های ۳۱، ۳۳، ۷۵ ممکن است داده پرت باشند. نتایج به‌کاربردن آزمون داده‌های پرت در بسته نرم‌افزاری car در نرم‌افزار R، تایید می‌کند که مشاهده ۳۳، داده پرت است. بنابراین، هم‌خطی و داده پرت به‌طور همزمان در داده‌ها وجود دارد.

^۱ این مجموعه داده در وبسایت <http://people.umass.edu/be640/yr2004/resources/data2002/VO2.txt> در دسترس است.



شکل ۵.۲: نمودار ماتریسی - نقطه‌ای از متغیرهای مستقل مجموعه داده V02



شکل ۶.۲: نمودار شناسایی داده‌های پرت در مجموعه داده V02

مقدار برآوردگرها برای پارامتر ريج $\lambda = 307/6858$ (بر اساس رابطه ۵۰.۲)، در جدول ۸.۲ گزارش شده است.

جدول ۸.۲: برآوردگرهای OLS، M، M، ريج و جزء مثبت انقباضی M ريج برای مجموعه داده V02

متغیرها	OLS	RME	URME	PRSRME
AGE	-۰/۱۴۹۳	-۰/۱۵۲۱	-۰/۰۵۸۶	-۰/۰۵۷۶
MHR	۰/۰۶۹۴	۰/۰۶۲۴	-۰/۰۲۳۶	-۰/۰۲۳۲
DUR	۰/۰۰۰۶	۰/۰۰۱۲	۰/۰۳۱۶	۰/۰۳۱۱
DBP	-۰/۰۰۹۶	-۰/۰۰۳۰	-۰/۰۳۷۳	-۰/۰۳۶۷
MXS	-۰/۰۱۹۷	-۰/۰۲۹۱	-۰/۰۵۸۰	-۰/۰۵۷۰
MXD	-۰/۳۲۱۶	-۰/۳۱۲۹	-۰/۰۹۷۸	-۰/۰۹۶۲

مقدار $RPE(\hat{\beta}^{sM+}(\lambda)) = 1/154686$ بهبود پیشگویی ۱۵ درصدی را در برآورد M ريج نشان می‌دهد. در صورتی که از روش بوت‌استرپ استاندارد استفاده شود، با مقایسه مقادیر انحراف استاندارد $\widehat{s.e.B}(\hat{\beta}^{sM+}(\lambda)) = 0/01448573$ و $\widehat{s.e.B}(\hat{\beta}^{uM}(\lambda)) = 0/07049829$ ، کارایی خوب برآوردگر جزء مثبت انقباضی نوع استاین M ريج تایید می‌شود.

تحلیل داده‌های چربی بدن

در این بخش، برآوردگرهای پیشنهادی را در مجموعه داده‌های چربی بدن^۲ به کار برده شده است. این داده‌ها ابتدا توسط **پنروز و همکاران (۱۹۸۵)** معرفی شده‌اند اما اولین بار در مقاله **جانسون (۱۹۹۶)** به طور کامل بررسی گردید. در این داده‌ها، درصد چربی بدن برای ۲۵۲ مرد بررسی شده است. متغیرها به صورت زیر هستند:

متغیر پاسخ: SIRI، درصد چربی بدن با استفاده از معادله سیری،
 متغیرهای توضیحی: DENSITY، چگالی تعیین شده از وزن یک مشخص (gm/cm^3)، این چگالی،
 نسبت جرم بدن به حجم بدن است. حجم بدن با استفاده از تکنیک وزن در زیر آب اندازه‌گیری شده است (**هوتینگ و همکاران، ۱۹۹۹**). AGE، سن (سال)، WEIGHT، وزن (lbs)، HEIGHT،
 قد (اینچ)، NECK، دور گردن (سانتی‌متر)، CHEST، دور سینه (سانتی‌متر)، ABDOMEN، دور شکم (سانتی‌متر)، HIP، دور باسن (سانتی‌متر)، THIGH، دور ران (سانتی‌متر)، KNEE، دور زانو (سانتی‌متر)، ANKLE، دور مچ پا (سانتی‌متر)، BICEPS، دور عضله سر (سانتی‌متر)، FOREARM، دور ساعد (سانتی‌متر) WRIST، دور کمر (سانتی‌متر).

^۲ این مجموعه داده در وبسایت http://www.stat.wisc.edu/gvludwig/fall_2012/bodyfat.csv موجود است.

توصیفی از متغیرها را می‌توان در جدول ۹.۲ و نمودار ۷.۲، مشاهده کرد.

جدول ۹.۲: آماره‌های توصیفی از متغیرها در مجموعه داده چربی بدن

متغیرها	کمینه	چارک اول	میانه	میانگین	چارک سوم	بیشینه
DENSITY	۰/۹۹۵	۱/۰۴۱	۱/۰۵۵	۱/۰۵۶	۱/۰۷۰	۱/۱۰۹
AGE	۲۲/۰۰	۳۵/۷۵	۴۳/۰۰	۴۴/۸۸	۵۴/۰۰	۸۱/۰۰
WEIGHT	۱۱۸/۵	۱۵۹/۰	۱۷۶/۵	۱۷۸/۹	۱۹۷/۰	۳۶۳/۱
HEIGHT	۲۹/۵۰	۶۸/۲۵	۷۰/۰۰	۷۰/۱۵	۷۲/۲۵	۷۷/۷۵
NECK	۳۱/۱۰	۳۶/۴۰	۳۸/۰۰	۳۷/۹۹	۳۹/۴۲	۵۱/۲۰
CHEST	۷۹/۳۰	۹۴/۳۵	۹۹/۶۵	۱۰۰/۸۲	۱۰۵/۳۸	۱۳۶/۲۰
ABDOMEN	۶۹/۴۰	۸۴/۵۸	۹۰/۹۵	۹۲/۵۶	۹۹/۳۳	۱۴۸/۱۰
HIP	۸۵/۰	۹۵/۵	۹۹/۳	۹۹/۹	۱۰۳/۵	۱۴۷/۷
THIGH	۴۷/۲۰	۵۶/۰۰	۵۹/۰۰	۵۹/۴۱	۶۲/۳۵	۸۷/۳۰
KNEE	۳۳/۰۰	۳۶/۹۸	۳۸/۵۰	۳۸/۵۹	۳۹/۹۲	۴۹/۱۰
ANKLE	۱۹/۱۰	۲۲/۰۰	۲۲/۸۰	۲۳/۱۰	۲۴/۰۰	۳۳/۹۰
BICEPS	۲۴/۸۰	۳۰/۲۰	۳۲/۰۵	۳۲/۲۷	۳۴/۳۳	۴۵/۰۰
FOREARM	۲۱/۰۰	۲۷/۳۰	۲۸/۷۰	۲۸/۶۶	۳۰/۰۰	۳۴/۹۰
WRIST	۱۵/۸۰	۱۷/۶۰	۱۸/۳۰	۱۸/۲۳	۱۸/۸۰	۲۱/۴۰
SIRI	۰/۰۰	۱۲/۴۷	۱۹/۲۰	۱۹/۱۵	۲۵/۳۰	۴۷/۵۰

مقادیر VIF در جدول ۱۰.۲ نشان داده شده است.

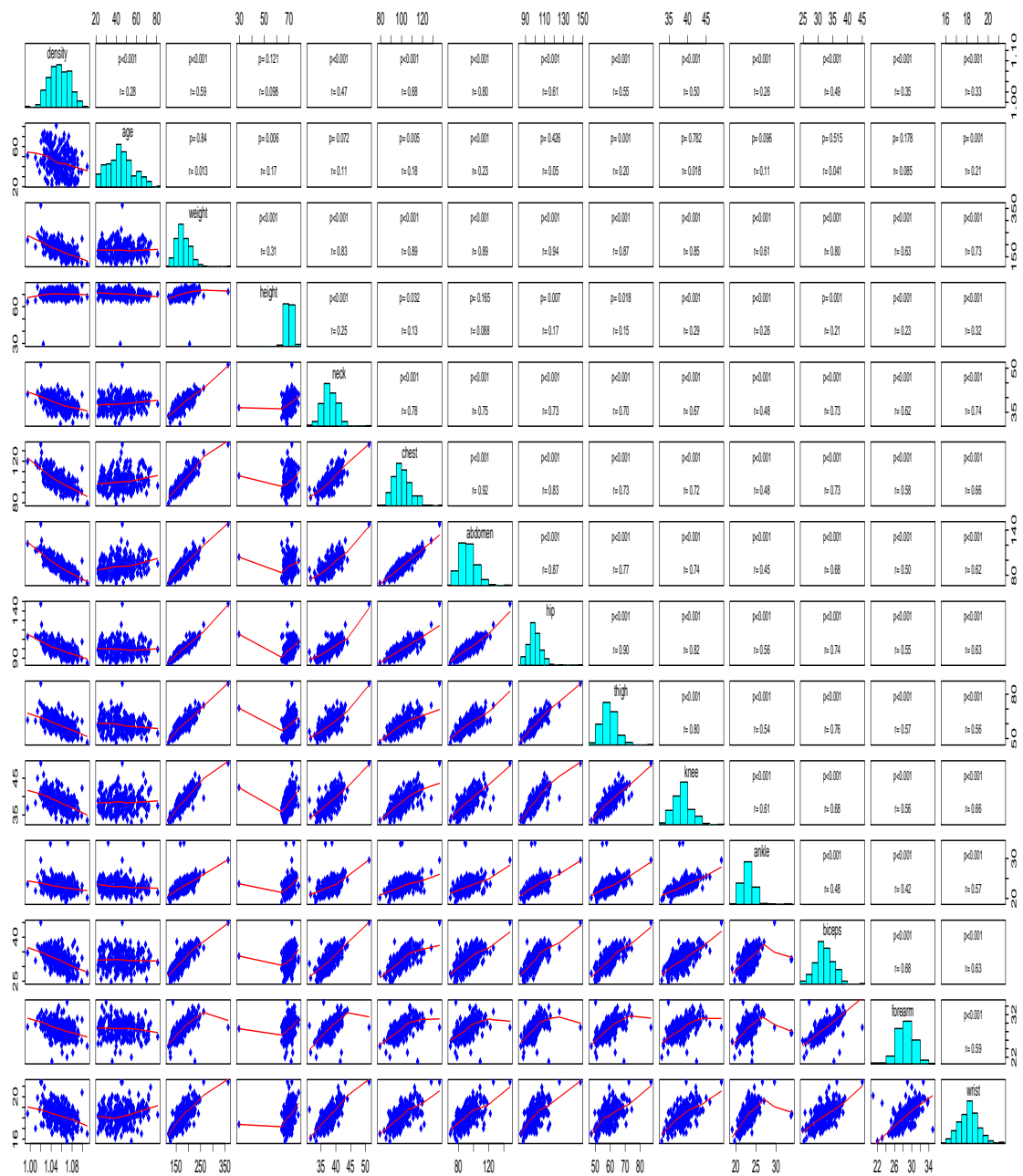
جدول ۱۰.۲: مقادیر VIF برای مجموعه داده چربی بدن

متغیرها	density	age	weight	height	neck	chest	abdomen
VIF	۳/۸۱۸۲	۲/۲۷۴۷	۳۴/۰۳۱۷	۱/۶۷۷۸	۴/۳۹۶۵	۹/۴۷۲۲	۱۸/۱۱۹۹
متغیرها	hip	thigh	knee	ankle	biceps	forearm	wrist
VIF	۱۴/۹۶۰۹	۷/۸۸۷۷	۴/۶۱۲۳	۱/۹۲۰۰	۳/۶۵۱۶	۲/۲۳۷۰	۳/۵۲۱۵

مقادیر جدول ۱۰.۲ و عدد شرطی $\kappa = ۱۳۷۵۱/۱۹$ نشان می‌دهد که هم‌خطی در بین متغیرهای مستقل وجود دارد.

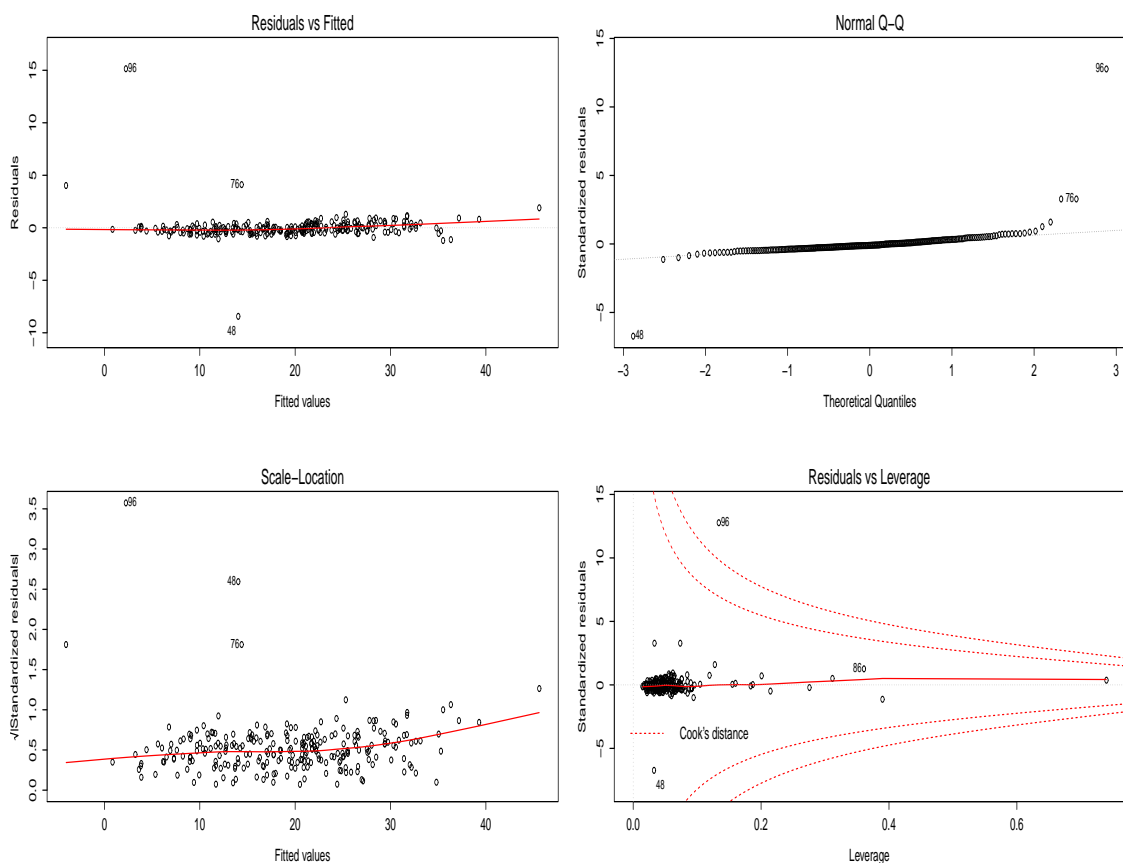
شکل ۸.۲ نشان می‌دهد که مشاهده‌های ۴۸، ۷۶، ۸۶ و ۹۶ ممکن است داده پرت باشند. به کار بردن آزمون داده پرت در بسته نرم‌افزاری car در نرم‌افزار R، منجر به شناسایی مشاهده‌های ۴۸ و ۹۶ به عنوان داده پرت می‌شود. بنابراین، در این مجموعه داده، مشکل هم‌خطی و داده پرت وجود دارند.

مقدار برآوردگرها برای پارامتر ریج با استفاده از رابطه (۵.۲) به صورت $\lambda = ۰/۰۰۰۰۰۰۹۶۳$ برآورد شده است و در جدول ۱۱.۲ گزارش شده است.



شکل ۷.۲: نمودار ماتریسی نقطه‌ای برای متغیرها در مجموعه داده چربی بدن

مقدار $RPE(\hat{\beta}^{SM+}(\lambda)) = 1/110996$ ، بهبود در پیشگویی ۱۱ درصدی را نشان می‌دهد. هم‌چنین در صورتی که از روش بوت‌استرپ استاندارد استفاده شود، با مقایسه مقادیر انحراف استاندارد $\hat{s.e.B}(\hat{\beta}^{SM+}(\lambda)) = 1/75082$ و $\hat{s.e.B}(\hat{\beta}^{uM}(\lambda)) = 2/35263$ که کارایی خوب PRSRME را تایید می‌کند.



شکل ۸.۲: نمودارهای شناسایی داده پرت در مجموعه داده چربی بدن

جدول ۱۱.۲: برآوردگرهای OLS، M_eM ريج و جزء مثبت انقباضی M ريج برای داده‌های چربی بدن

متغیرها	OLS	UME	URME	PRRME
DENSITY	-۴۱۱/۲۳۷۸	-۴۴۲/۰۴۵۵	-۴۳۱/۹۳۱۰	-۴۱۱/۲۴۲۵
AGE	۰/۰۱۲۵۹	۰/۰۰۰۷۴	۰/۰۰۲۸۴	۰/۰۰۲۷۰
WEIGHT	۰/۰۱۰۰۵	۰/۰۰۱۸۴	۰/۰۳۱۳۱	۰/۰۲۹۸۱
HEIGHT	-۰/۰۰۷۹۸	-۰/۰۰۹۳۶	-۰/۰۴۸۳۱	-۰/۰۴۵۹۹
NECK	-۰/۰۲۸۴۶	۰/۰۰۲۵۷	-۰/۰۵۴۴۴	-۰/۰۵۱۸۳
CHEST	۰/۰۲۶۷۸	۰/۰۰۱۰۸	-۰/۰۲۸۳۹	-۰/۰۲۷۰۳
ABDOMEN	۰/۰۱۸۵۶	۰/۰۰۰۵۳	۰/۰۲۳۶	۰/۰۲۲۵۰
HIP	۰/۰۱۹۱۷	۰/۰۰۹۳۵	-۰/۰۵۴۶۴	-۰/۰۵۲۰۶
THIGH	-۰/۰۱۶۷۶	-۰/۰۱۱۳۳	-۰/۰۱۲۳۳	-۰/۰۱۱۷۴
KNEE	-۰/۰۰۴۶۴	-۰/۰۰۶۸۱	-۰/۰۴۳۵۶	-۰/۰۴۱۴۸
ANKLE	-۰/۰۸۵۶۸	۰/۰۰۳۳۷	-۰/۰۱۶۵۷	-۰/۰۱۵۷۷
BICEPS	-۰/۰۵۵۰۵	-۰/۰۰۴۹۲	-۰/۰۰۶۷۵	-۰/۰۰۶۴۳
FOREARM	۰/۰۳۳۸۶	-۰/۰۰۰۰۵	-۰/۰۰۰۰۶۳	-۰/۰۰۰۰۶۰
WRIST	۰/۰۰۷۳۴	-۰/۰۱۴۳۶	-۰/۰۱۱۸۶۰	-۰/۰۱۱۲۹۲

۴.۲ مدل افراز شده

بدون از دست دادن کلیت مسئله مورد بررسی، افرازی به صورت زیر در نظر بگیرید.

$$\beta^T = \left(\beta_1^T \quad \beta_2^T \right) \quad \text{و} \quad X_n = \left(X_{n1} \quad X_{n2} \right)$$

به طوری که X_{ni} زیرماتریس در اندازه $n \times p_i$ و β_i زیربردار در اندازه $1 \times p_i$ ، $i = 1, 2$ بوده و $p_2 \geq 0$ و $p_1 \geq 0$ ، $p = p_1 + p_2$.

مدل خطی (۱.۱) می تواند به صورت زیر نوشته شود

$$y_n = X_{n1}\beta_1 + X_{n2}\beta_2 + \varepsilon_n. \quad (55.2)$$

به مدل رگرسیونی (55.2)، مدل افراز شده گویند.

در اینجا، β_1 پارامتر مورد علاقه است و فرض می شود β_2 به مقدار ثابت β_2^0 نزدیک باشد. بدون اینکه کلیت مسئله از دست برود، فرض کنید $\beta_2^0 = 0$. در این حالت کاهش دادن بعد فضای پارامتر در رویه برآورد، منجر به بهبود برآوردگر β_1 می شود. معمولاً، اطلاعات غیرنمونه ای به صورت

$$H_0 : \beta_2 = 0 \quad \text{v.s.} \quad H_A : \beta_2 \neq 0, \quad (56.2)$$

فرمول بندی می شود. به عنوان یک مثال، در طرح آزمایش های چندعاملی، β_1 و β_2 به ترتیب اثرهای اصلی و اثرهای متقابل را نشان می دهد. ممکن است علاقه مند باشید که همه اثرهای متقابل معنی دار نباشند و فقط به دنبال برآورد اثرهای اصلی باشید.

برای مدل رگرسیونی (۱.۱)، برآوردگر محدود نشده M ، پاسخ معادله $M_n(b) = 0$ در رابطه (۱.۲) است که بر اساس مدل افراز شده (55.2) به صورت زیر $\hat{\beta}^{uM} = (\hat{\beta}_1^{uM}, \hat{\beta}_2^{uM})^T$ افراز می شود. جرکووا (۱۹۷۷) و سینگر و سن (۱۹۸۵) ویژگی های مختلف برآوردگر $\hat{\beta}_1^{uM}$ را در مدل افراز شده (55.2) بررسی نموده اند.

فرض کنید

$$C_n = X_n^T X_n = \begin{bmatrix} X_{n1}^T X_{n1} & X_{n1}^T X_{n2} \\ X_{n2}^T X_{n1} & X_{n2}^T X_{n2} \end{bmatrix} = \begin{bmatrix} C_{n11} & C_{n12} \\ C_{n21} & C_{n22} \end{bmatrix}. \quad (57.2)$$

مشابه مدل (۱.۱)، برآوردگر OLS پارامتر β_1 در مدل افراز شده (55.2) برابر است با

$$\hat{\beta}_1^{Ols} = C_{n11}^{-1} X_{n1}^T y_n.$$

به منظور به دست آوردن برآوردگر ریج معمولی برای β_1 ، کافی است عامل ریج به صورت

$$R_{n1}(\lambda_n) = [I_{p_1} + \lambda C_{n1}^{-1}]^{-1}$$

تعریف شود که منجر به برآوردگر زیر می شود:

$$\hat{\beta}_1^{\text{Rdg}}(\lambda) = R_{n1}(\lambda) \hat{\beta}_1^{\text{Ols}}.$$

برآوردگر استوار نسبت به هم خطی پارامتر β_1 از ضرب کردن عامل ریج در UME، برآوردگر محدود نشده M ریج برای پارامتر β_1 به صورت زیر به دست می آید:

$$\hat{\beta}_1^{\text{uM}}(\lambda) = R_{n1}(\lambda) \hat{\beta}_1^{\text{uM}}.$$

این استراتژی برای تعریف کردن تمامی برآوردگرهای بهبودیافته M ریج مفید خواهد بود. مدل (۵۵.۲) محدود شده، $y_n = X_{n1} \beta_1 + \epsilon_n$ است (به عبارت دیگر، $\beta_2 = 0$). در این حالت، $\hat{\beta}_1^{\text{M}}$ ، برآوردگر M پارامتر β_1 که آن را RME پارامتر β_1 گویند، پاسخ معادله زیر است:

$$M_{n1}(\beta_1) = (M_1(\beta), \dots, M_{p_1}(\beta))^T = \sum_{i=1}^n x_{1i} \psi(y_i - x_{1i}^T \beta_1) = 0, \quad (58.2)$$

که در آن x_{1i} ، i امین سطر ماتریس X_{n1} است. از نماد $\hat{\beta}_1^{\text{rM}}$ برای نشان دادن RME پارامتر β_1 استفاده می شود. لازم به ذکر است در مقایسه رابطه (۵۸.۲) با رابطه (۴.۲)، لزومی ندارد که آماره آزمون هم پایای مقیاس باشد و بنابراین در این قسمت از رابطه (۵۸.۲) در انجام محاسبات استفاده می کنیم تا نشان دهیم حتی اگر آماره آزمون هم پایای مقیاس نباشد، باز هم بهبود برآوردگرها صورت گرفته و به نتایج مطلوبی دست پیدا می کنیم. بر اساس احمد و همکاران (۲۰۰۶)، برآوردگر محدود شده M پارامتر β_1 می تواند به صورت زیر نوشته شود:

$$\hat{\beta}_1^{\text{rM}} = \hat{\beta}_1^{\text{uM}} + C_{n1}^{-1} C_{n12} \hat{\beta}_2^{\text{uM}}.$$

اکنون، فرض کنید

$$\widehat{M}_{n2}^{\text{R}} = \sum_{i=1}^n x_{2i} \psi(y_i - x_{1i}^T \hat{\beta}_1^{\text{rM}})$$

که در آن x_{2i} ، i امین سطر ماتریس X_{n2} است. اگر

$$\hat{\sigma}_\psi^2 = \frac{1}{n - p_1} \sum_{i=1}^n \psi^2(y_i - x_{1i}^T \hat{\beta}_1^{\text{rM}}),$$

با استفاده از جرکووا و سن (۱۹۹۶)، آماره آزمون مناسب برای آزمون کردن فرضیه $\mathcal{H}_0: \beta_2 = 0$ در مقابل $\mathcal{H}_1: \beta \neq 0$ می تواند به صورت زیر نوشته شود

$$\mathcal{L}_{n1}^{\text{M}} = \frac{[\widehat{M}_{n2}^{\text{R}}]^T C_{n22}^{-1} [\widehat{M}_{n2}^{\text{R}}]}{\hat{\sigma}_\psi^2}.$$

برآوردگر انقباضی نوع استاین M پارامتر β_1 که آن را با نماد $\hat{\beta}_1^{sM}$ نشان می‌دهند، به صورت زیر تعریف می‌شود:

$$\begin{aligned}\hat{\beta}_1^{sM} &= \hat{\beta}_1^{rM} + (1 - c[\mathcal{L}_n^M]^{-1})(\hat{\beta}_1^{uM} - \hat{\beta}_1^{rM}) \\ &= \hat{\beta}_1^{uM} - c[\mathcal{L}_n^M]^{-1}(\hat{\beta}_1^{uM} - \hat{\beta}_1^{rM})\end{aligned}$$

که در آن c پارامتر انقباضی است و دامنه آن در بازه‌ی $(0, 2(p_2 - 2))$ تغییر کند و اغلب برابر با $c = p_2 - 2$ در نظر گرفته می‌شود. بنابراین، فرض می‌شود $p_2 \geq 3$. با به کار بردن تکنیک **سالو و همکاران (۱۹۷۲)**، می‌توان SME را با PRSME بهبود بخشید. این برآوردگر برای پارامتر β_1 به صورت زیر تعریف می‌شود:

$$\begin{aligned}\hat{\beta}_1^{sM+} &= \hat{\beta}_1^{rM} + (1 - c[\mathcal{L}_n^M]^{-1})^+(\hat{\beta}_1^{uM} - \hat{\beta}_1^{rM}) \\ &= \hat{\beta}_1^{sM} - (1 - c[\mathcal{L}_n^M]^{-1})I(\mathcal{L}_n^M < c)(\hat{\beta}_1^{uM} - \hat{\beta}_1^{rM})\end{aligned}$$

وقتی $\beta_2 = 0$ برقرار باشد، RME انحراف استاندارد جانبی کوچکتری از UME دارد. با این وجود، برای $\beta_2 \neq 0$ ، RME ممکن است در بسیاری از حالت‌ها اریب و ناسازگار باشد. به این دلیل، معمولاً PTME پارامتر β_1 را که با نماد $\hat{\beta}_1^{ptM}$ نشان می‌دهند، مشابه **سن و صالح (۱۹۸۷)** به صورت زیر تعریف می‌کنند:

$$\begin{aligned}\hat{\beta}_1^{ptM} &= \hat{\beta}_1^{uM} I(\mathcal{L}_n^M > \mathcal{L}_{n, \alpha}) + \hat{\beta}_1^{rM} I(\mathcal{L}_n^M \leq \mathcal{L}_{n, \alpha}) \\ &= \hat{\beta}_1^{uM} - (\hat{\beta}_1^{uM} - \hat{\beta}_1^{rM})I(\mathcal{L}_n^M < \mathcal{L}_{n, \alpha})\end{aligned} \quad (59.2)$$

که در آن $\mathcal{L}_{n, \alpha}$ چندک $100\alpha\%$ بالایی از توزیع صفر \mathcal{L}_n^M است. به طور جانبی، به صورت توزیع کی دو با p_2 درجه آزادی تحت فرضیه صفر می‌باشد. دقت کنید برای تعریف PTME، کافی است $p_2 \geq 1$. با این وجود، اگر $p_2 \geq 3$ ، مشابه **سالو و همکاران (۱۹۷۲)**، می‌توان برآوردگر آزمون اولیه بهبودیافته M ریبج را IPTRME که با نماد $\hat{\beta}_1^{iptM}$ نشان داده می‌شود، به شکل زیر تعریف نمود:

$$\hat{\beta}_1^{iptM} = \hat{\beta}_1^{sM} I(\mathcal{L}_n^M > \mathcal{L}_{n, \alpha}) + \hat{\beta}_1^{rM} I(\mathcal{L}_n^M \leq \mathcal{L}_{n, \alpha}).$$

از آن جایی که بازه موثر از مخاطره کمتر برآوردگر نوع استاین نسبت به برآوردگر M، همسایگی کوچکی از $\beta_2 = 0$ است. همان طور که n بزرگ‌تر می‌شود این بازه موثر، کوچکتر می‌شود. معمولاً برای فرضیه $\beta_2 = 0$ ، دنباله‌ای از فرضیه‌های مقابل موضعی به صورت $\mathcal{K}_{(n)}$ تعریف می‌گردد:

$$\mathcal{K}_{(n)} : \beta_2 = \beta_{2(n)} = n^{-\frac{1}{2}} \xi, \quad \xi = (\xi_{p_1+1}, \dots, \xi_p)^T \in \mathbb{R}^{p_2}.$$

بنابراین فرضیه \mathcal{H}_0 به $\xi = 0$ کاهش می‌یابد. در ابتدا برای یافتن توزیع آماره آزمون، بر اساس **سینگر و سن (۱۹۸۵)**، چندین لم و قضیه مورد نیاز است. ابتدا، فرضیه‌های لازم بیان می‌شوند:

(الف-۱) متغیرهای تصادفی خطا، تابع توزیع مطلقاً پیوسته F با تابع چگالی متقارن f دارند به طوری که مشتق‌های آن موجود است.

(الف-۲) تابع توزیع F ، دارای ماتریس معین مثبت و متناهی اطلاع فیشر به صورت زیر است:

$$I(f) = (I_{ij}(f)) = \left(\int \left[\frac{f'_i(\varepsilon)f'_j(\varepsilon)}{f(\varepsilon)} \right]^2 f(\varepsilon) d\varepsilon \right).$$

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \mathbf{x}_{ni}^\top (\mathbf{X}_n^\top \mathbf{X}_n)^{-1} \mathbf{x}_{ni} = \lim_{n \rightarrow \infty} \mathbf{x}_{ni}^\top \mathbf{C}_n^{-1} \mathbf{x}_{ni} = o(1) \quad (\text{ب-۱})$$

$$\lim_{n \rightarrow \infty} n^{-1} \mathbf{C}_n = \mathbf{C} \quad (\text{ب-۲}) \quad \text{که } \mathbf{C} \text{ ماتریس معین مثبت است.}$$

(ج) تابع امتیاز ψ ، تابع غیرثابت به شکل $\psi(\mathbf{x}) = \sum_{j=1}^s \psi_j(\mathbf{x})$ می‌باشد که ψ_j ها تابع یکنوا،

پله‌ای و مطلقاً پیوسته روی بازه کراندار \mathbb{R} در مشتق ψ'_j می‌باشد. همچنین

$$\begin{aligned} \int \psi(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} &= 0 \\ \sigma_\psi^2 &= \int \psi^2(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} < \infty \\ \gamma &= \int \phi'(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = - \int \psi(\mathbf{x}) f'(\mathbf{x}) d\mathbf{x} < \infty. \end{aligned}$$

لم ۱.۴.۲. تحت فرضیه‌های (ب-۱) و (ب-۲)،

$$n^{-\frac{1}{2}} \mathbf{M}_n(\mathbf{b}) \xrightarrow{D} \mathcal{N}_p(o, \sigma_\psi^2 \mathbf{C}). \quad (۶۰.۲)$$

برهان. بر اساس تعریف (۴.۲)، کفایت از قضیه‌های ۱.۳.آ و ۲.۳.آ استفاده شود. □

لم ۲.۴.۲. (جرکووا، ۱۹۷۷) عبارت $n^{\frac{1}{2}} \|\hat{\beta}_n^{\text{uM}} - \beta\|$ در احتمال کراندار است.

لم ۳.۴.۲. (سینگر و سن، ۱۹۸۵) تحت فرض‌های (الف-۱)، (الف-۲)، (ب-۱)، (ب-۲) و (ج)،

$$P \left(\sup_{\sqrt{n} \|\hat{\beta}_n^{\text{uM}} - \beta\| \leq K} n^{-\frac{1}{2}} \|\mathbf{M}_n(\hat{\beta}_n^{\text{uM}}) - \mathbf{M}_n(\beta) - n\gamma \mathbf{C}(\hat{\beta}_n^{\text{uM}} - \beta)\| \geq \varepsilon \right) \rightarrow 0.$$

قضیه ۱.۴.۲. تحت فرض‌های (الف-۱)، (الف-۲) و (ب-۱) و (ب-۲) و (ج)،

$$n^{\frac{1}{2}} (\hat{\beta}_n^{\text{uM}} - \beta) \xrightarrow{D} \mathcal{N}_p(o, \eta^2 \mathbf{C}^{-1}), \quad (۶۱.۲)$$

که در آن $\eta^2 = \sigma_\psi^2 / \gamma^2$.

برهان. بر اساس لم‌های ۲.۴.۲ و ۳.۴.۲، عبارت $n^{\frac{1}{2}} (\gamma \mathbf{C})(\hat{\beta}_n^{\text{uM}} - \beta)$ توزیع مجانبی مشابهی با $n^{-\frac{1}{2}} \mathbf{M}_n(\mathbf{b})$ دارد که همان $\mathcal{N}_p(o, \sigma_\psi^2 \mathbf{C})$ است. بنابراین $n^{\frac{1}{2}} (\hat{\beta}_n^{\text{uM}} - \beta)$ دارای توزیع p متغیره نرمال با میانگین o و ماتریس واریانس-کوواریانس $\sigma_\psi^2 \gamma^{-2} \mathbf{C}^{-2} \mathbf{C} = \eta^2 \mathbf{C}^{-1}$ می‌باشد. □

نتیجه ۱.۴.۲. تحت فرض‌های (الف-۱)، (الف-۲) و (ب-۱) و (ب-۲) و (ج)،

$$n^{\frac{1}{2}} \left(\hat{\beta}_{n1}^{\text{uM}} - \beta \right) \xrightarrow{D} \mathcal{N}_{p_1} \left(o, \eta^2 C_{11}^{-1} \right), \quad \eta^2 = \sigma_\psi^2 / \gamma^2. \quad (۶۲.۲)$$

قضیه ۲.۴.۲. تحت فرض‌های (الف-۱)، (الف-۲) و (ب-۱) و (ب-۲) و (ج)،

$$n^{\frac{1}{2}} \left(\hat{\beta}_{n1}^{\text{rM}} - \beta_1 \right) \xrightarrow{D} \mathcal{N}_{p_1} \left(o, \eta^2 C_{11}^{-1} \right), \quad (۶۳.۲)$$

برهان. طبق لم‌های ۳.۴.۲ و ۲.۴.۲، توزیع مجانبی عبارت $n^{\frac{1}{2}} (\hat{\beta}_{n1}^{\text{uM}} - \beta_1)$ مشابه $n^{-\frac{1}{2}} M_{n1}(b)$ می‌باشد که همان $\mathcal{N}_{p_1}(o, \sigma_\psi^2 C_{11})$ است. □

قضیه ۳.۴.۲. تحت برقراری شرایط نظم و کلاس $\mathcal{K}_{(n)}$ ، آماره آزمون \mathcal{L}_{n1}^M دارای توزیع مجانبی کی دو غیرمرکزی با درجه آزادی و پارامتر غیرمرکزی δ $\Delta^2 = \eta^{-2} \delta^\top C_{22.1} \delta$ است که در آن $\delta = C_{11}^{-1} C_{12} \epsilon$ و

$$\eta^2 = \frac{\int_{\mathbb{R}} \psi^2(y) dF(y)}{\int_{\mathbb{R}} \psi(x) \{-f'(x)/f(x)\} dF(x)}, \quad f'(x) = \frac{df(x)}{dx}.$$

برهان. بردار $M_n(b)$ را می‌توان به صورت

$$\begin{aligned} M_n(b) &= M_n(b_1, b_2) = \sum_{i=1}^n x_i \psi \left(y_i - x_{1i}^\top b_1 - x_{2i}^\top b_2 \right) \\ &= (M_{n1}(b_1, b_2), M_{n2}(b_1, b_2)), \end{aligned}$$

بازنویسی نمود که در آن

$$\begin{aligned} M_{n1}(b_1, b_2) &= \sum_{i=1}^n x_{1i} \psi \left(y_i - x_{1i}^\top b_1 - x_{2i}^\top b_2 \right) \\ M_{n2}(b_1, b_2) &= \sum_{i=1}^n x_{2i} \psi \left(y_i - x_{1i}^\top b_1 - x_{2i}^\top b_2 \right). \end{aligned}$$

بر اساس لم ۱.۴.۲،

$$n^{-\frac{1}{2}} \begin{pmatrix} M_{n1}(b_1, b_2) \\ M_{n2}(b_1, b_2) \end{pmatrix} \xrightarrow{D} \mathcal{N}_p \left(\begin{pmatrix} o \\ o \end{pmatrix}, \eta^2 \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \right). \quad (۶۴.۲)$$

فرض کنید $\hat{\beta}_1^{\text{rM}}$ برآوردگر محدودشده M پارامتر β_1 تحت فرضیه \mathcal{H}_0 باشد که پاسخ معادله $M_{n1}(\hat{\beta}_1^{\text{rM}}, o) = o$ است. تعریف کنید $M_{n2}(\hat{\beta}_1^{\text{rM}}, o) = M_{n2}(\hat{\beta}_1^{\text{rM}})$. اکنون، ملاحظه می‌شود

$$\begin{aligned} n^{-\frac{1}{2}} \widehat{M}_{n2}(\hat{\beta}_1^{\text{rM}}) &= n^{-\frac{1}{2}} \left\{ M_{n2}(\beta_1 + (\hat{\beta}_1^{\text{rM}} - \beta_1), o) \right. \\ &\quad \left. - M_{n2}(\beta_1, o) - n\gamma C_{21} (\hat{\beta}_1^{\text{rM}} - \beta_1) \right\} + n^{-\frac{1}{2}} \left\{ M_{n2}(\beta_1, o) \right. \\ &\quad \left. + n\gamma C_{21} (\hat{\beta}_1^{\text{rM}} - \beta_1) \right\}. \quad (۶۵.۲) \end{aligned}$$

بر اساس لم ۳.۴.۲، اولین عبارت در (۶۵.۲) تحت هر یک از فرضیه‌های \mathcal{H}_0 یا \mathcal{H}_A در احتمال به صفر همگراست. علاوه براین، لم ۳.۴.۲ نتیجه می‌دهد که $\hat{\beta}_{n1}^{rM} - \beta_1$ همان توزیع مجانبی $n^{-\frac{1}{2}}(\gamma^{-1}C_{11}^{-1})M_{n1}(\beta_1, o)$ را دارد. بنابراین،

$$n^{-\frac{1}{2}}\hat{M}_{n2}(\hat{\beta}_{n1}^{rM}) = n^{-\frac{1}{2}}\{M_{n2}(\beta_1, o) + C_{21}C_{11}^{-1}M_{n1}(\beta_1, o)\} + o_p(1). \quad (۶۶.۲)$$

طبق رابطه (۶۴.۲)، بدیهی است که توزیع $n^{-\frac{1}{2}}\hat{M}_{n2}(\hat{\beta}_{n1}^{rM})$ نرمال p_2 متغیره با بردار میانگین o و ماتریس واریانس-کوواریانس $C_{22} - C_{21}C_{11}^{-1}C_{12} = C_{22.1}$ است. بنابراین

$$\frac{n\hat{M}_{n2}(\hat{\beta}_{n1}^{rM})C_{22.1}^{-1}\hat{M}_{n2}(\hat{\beta}_{n1}^{rM})}{\eta^2} \xrightarrow{D} \chi_{p_2}^2. \quad (۶۷.۲)$$

با استفاده از قضیه ۳.۳.آ، \mathcal{L}_{n1}^M دارای توزیع χ^2 با p_2 درجه آزادی است. تحت فرضیه‌های موضعی مقابل،

$$n^{\frac{1}{2}}(\hat{\beta}_2^{uM} - \beta_2) \xrightarrow{D} N_{p_2}(C_{11}^{-1}C_{12}\delta, \eta^2 C_{22}^{-1}).$$

توزیع $n^{\frac{1}{2}}M_{n2}(b_1, b_2)$ مشابه $n^{\frac{1}{2}}(\hat{\beta}_2^{uM} - \beta_2)$ است، می‌توان نتیجه گرفت که

$$n^{-\frac{1}{2}}\hat{M}_{n2}(\hat{\beta}_1^{rM}) \sim N_{p_2}(C_{11}^{-1}C_{12}\xi, \eta^2(C_{22} - C_{21}C_{11}^{-1}C_{12})).$$

بنابراین،

$$\frac{n\hat{M}_{n2}(\hat{\beta}_1^{rM})C_{22.1}^{-1}\hat{M}_{n2}(\hat{\beta}_1^{rM})}{\eta^2} \xrightarrow{D} \chi_{p_2}^2(\Delta^2),$$

که در آن $\Delta^2 = \eta^{-2}\xi^T C_{11.2}^{-1}\xi$. \mathcal{L}_{n1}^M دارای توزیع χ^2 با p_2 درجه آزادی و پارامتر غیرمرکزی Δ^2 است. □

۱.۴.۲ برآوردگرهای بهبودیافته M ریج

مشابه صالح و کبیریا (۲۰۱۱)، برآوردگرهای PRSRM و SRM، IPTRM، PTRM، RRM، URM در برآوردگرهای $R_1(\lambda)$ به ترتیب با ضرب $R_1(\lambda)$ در برآوردگرهای SM، IPTM، PTM، RM، UM به صورت

$$\begin{aligned} \hat{\beta}_1^{uM}(\lambda) &= R_1(\lambda)\hat{\beta}_1^{uM}, \\ \hat{\beta}_1^{rM}(\lambda) &= R_1(\lambda)\hat{\beta}_1^{rM}, \\ \hat{\beta}_1^{ptM}(\lambda) &= R_1(\lambda)\hat{\beta}_1^{ptM}, \\ \hat{\beta}_1^{iptM}(\lambda) &= R_1(\lambda)\hat{\beta}_1^{iptM}, \\ \hat{\beta}_1^{sM}(\lambda) &= R_1(\lambda)\hat{\beta}_1^{sM}, \\ \hat{\beta}_1^{sM+}(\lambda) &= R_1(\lambda)\hat{\beta}_1^{sM+}. \end{aligned}$$

تعریف می‌شوند. با توجه به قضیه‌های ۹.۳.۲ - ۱۴.۳.۲ نتیجه می‌شود که این برآوردگر بهتر رفتار می‌کند. برای جزییات بیشتر، مقاله نوروزی‌راد و همکاران (۲۰۱۷b) را می‌توان ملاحظه نمود.

۲.۴.۲ انتخاب پارامتر ریح

انتخاب پارامتر ریح، یکی از مهم‌ترین موضوعات در این رگرسیون است. مشابه روش پیشنهادی هورل و همکاران (۱۹۷۵)، روش تکراری زیر برای تعیین پارامتر ریح استوار، λ ، پیشنهاد می‌گردد. بر اساس این روش، تکرار از نقطه λ_0 به صورت $\lambda_0 = ps^2 \left[\left(\hat{\beta}_1^{uM} \right)^\top \left(\hat{\beta}_1^{uM} \right) \right]^{-1}$ شروع می‌شود و λ_i ها بر اساس

$$\lambda_i = ps^2 \left[\left(\hat{\beta}_1^{uM}(\lambda_{i-1}) \right)^\top \left(\hat{\beta}_1^{uM}(\lambda_{i-1}) \right) \right]^{-1}$$

محاسبه می‌شوند. تغییرات در مقادیر λ_i ، تعداد فرآیند تکرار را مشخص می‌کند. تکرار متوقف می‌شود اگر

$$\frac{\lambda_i - \lambda_{i-1}}{\lambda_{i-1}} < \epsilon,$$

که ثابت مثبت ϵ به اندازه کافی کوچک می‌باشد. در پایان فرآیند، $\hat{\beta}_1^{uM}(\lambda_i)$ می‌تواند برای برآورد β به کار گرفته شود.

۳.۴.۲ مطالعه شبیه‌سازی

در طرح شبیه‌سازی، $\beta = (\beta_1^\top, \beta_2^\top)^\top$ بردار p بعدی است که $\beta_1 = (1, \dots, 1)^\top$ بردار p_1 تایی از یک‌ها و $\beta_2 = (0, \dots, 0, \Delta^2)^\top$ بردار p_2 تایی است که به جز آخرین مولفه‌اش، سایر مولفه‌های آن صفر است. وقتی $\Delta^2 = 0$ باشد، β_2 بردار صفر است. $\Delta^2 > 0$ معادل با انحراف از فرضیه صفر است. مقادیر Δ^2 را به صورت $0, 0/5, 1, 1/5, 2, 2/5, 3, 5$ و 10 در نظر گرفته شده است. $\Delta^2 = 0$ نشان می‌دهد که داده‌ها از فرضیه صفر تولید شده‌اند و $\Delta^2 > 0$ نشان می‌دهد که داده‌ها بر اساس فرضیه مقابل به دست آمده‌اند. پس از 1000 بار تکرار، مخاطره پارامترهای رگرسیونی برآورد شده محاسبه شده‌اند. بنابراین، مخاطره‌ها برای برآوردگرهای M ریح پیشنهاد شده به دست آمده‌اند.

مدل پیشنهادی زیر برای تولید n مشاهده از متغیر پاسخ به کار گرفته شد.

$$y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i \quad (68.2)$$

که در آن β_i یک بردار p_i بعدی به ازای $i = 1, 2$ است.

متغیر خطای ϵ_i از توزیع نرمال آمیزه $\varrho \mathcal{N}(0, 9) + (1 - \varrho) \mathcal{N}(0, 1)$ زیر تولید شده است که ϱ درصد داده‌های پرت است (سیمپسون، ۱۹۹۵). درجه‌های مختلف هم‌خطی با تولید متغیرهای مستقل از مدل داده‌شده در رابطه (۶۸.۲) برای مقادیر ρ به دست می‌آید. هرچه مقدار ρ بزرگتر باشد، وابستگی (نرخ هم‌خطی) بین متغیرهای توضیحی بیشتر است.

بر اساس روش شرادر و هتمانسیرگر (۱۹۸۰) به منظور برآورد s_n ، می‌توان از برآوردگر

$$s_n = 2/1 \times \text{med}(|r_i^*|)$$

استفاده کرد که $|r_i^*|$ ، $n-p+1$ تای بزرگترین قدرمطلق باقیماندهها بر اساس باقیمانده به دست آمده از برازش برآوردگر LAD هستند و $\text{med}(\cdot)$ میانه دادهها است.

برای N تکرار، MSE نسبی برآوردگرها به صورت

$$\text{RMSE}(\hat{\beta}_1^{*M}(\lambda)) = \frac{\text{MSE}(\hat{\beta}_1^{\text{uM}}(\lambda))}{\text{MSE}(\hat{\beta}_1^{*M}(\lambda))} = \frac{\frac{1}{N} \sum_{i=1}^N (\hat{\beta}_1^{\text{uM}^{(i)}}(\lambda) - \beta_1)^\top (\hat{\beta}_1^{\text{uM}^{(i)}}(\lambda) - \beta_1)}{\frac{1}{N} \sum_{i=1}^N (\hat{\beta}_1^{*M^{(i)}}(\lambda) - \beta_1)^\top (\hat{\beta}_1^{*M^{(i)}}(\lambda) - \beta_1)}$$

بیان می شود که در آن $\hat{\beta}_1^{\text{uM}^{(i)}}(\lambda)$ برآوردگر M ریج در i امین مرحله است و $\hat{\beta}_1^{M^{*(i)}}(\lambda)$ می تواند هر یک از برآوردگرهای پیشنهادی باشد. MSE نسبی بیشتر از یک، برتری برآوردگر پیشنهادی را بر URME نشان می دهد.

فرآیند مطرح شده برای $\rho = 0/8, 0/9, 0/99, 0/999$ ، حجم نمونه $n = 50, 100$ و هم چنین واریانس خطای $\sigma^2 = 1$ ، 1000 بار تکرار شده است. برای ترکیبهای مختلف n و ρ ، کارایی نسبی به ازای $\rho = 0/2$ ، در حالت های $(p_1, p_2) = (2, 3), (2, 5)$ به دست آمده اند.

جدول ۱۲.۲: کارایی نسبی برآوردگرهای پیشنهادی در حالت $n = 50, \sigma^2 = 1, \alpha = 0/05, \rho = 0/2$ ، $(p_1, p_2) = (2, 3), \lambda = 0/5$

برآوردگرها							Δ^2	ρ
PRSRME	SRME	IPTRME	PTRME	RRME	URME			
1/2079	1/0441	1/4850	1/4618	1/5917	1	0	0/8	
1/1939	1/0439	1/4318	1/4133	1/5151	1	0/5		
1/2039	1/0502	1/4160	1/3965	1/4914	1	1		
1/1465	1/0219	1/2170	1/1919	1/3007	1	2		
1/0313	1/0169	0/7790	0/7501	0/6832	1	5		
1/0057	1/0057	0/8523	0/8254	0/2455	1	10		
1/2353	1/0378	1/5266	1/4969	1/6818	1	0	0/9	
1/2359	1/0729	1/5170	1/4864	1/6481	1	0/5		
1/2275	0/9835	1/4890	1/4604	1/6290	1	1		
1/2039	1/0582	1/3882	1/3534	1/5494	1	2		
1/2197	0/8399	1/4684	1/4439	1/5967	1	5		
1/0020	1/0012	0/6879	0/6656	0/3715	1	10		
1/2641	0/8478	1/5712	1/5254	1/8269	1	0	0/95	
1/2616	1/1522	1/5784	1/5345	1/8101	1	0/5		
1/2658	0/2855	1/5518	1/5070	1/7690	1	1		
1/2527	0/7968	1/5356	1/5061	1/6730	1	2		
1/1484	0/8298	1/1471	1/1068	1/2617	1	5		
1/0333	0/8394	0/7411	0/7148	0/6157	1	10		
1/2972	0/5103	1/6122	1/5330	2/2334	1	0	0/99	
1/3156	0/3376	1/6535	1/5862	2/1881	1	0/5		
1/2860	0/1905	1/5743	1/5031	2/1594	1	1		
1/3190	0/1687	1/5952	1/5270	2/1397	1	2		
1/2522	0/6549	1/4378	1/3805	1/8573	1	5		
1/2039	0/7027	1/2628	1/2091	1/4388	1	10		

جدول ۱۳.۲: کارایی نسبی برآوردگرهای پیشنهادی در حالت $n = 100$ ، $\sigma^2 = 1$ ، $\alpha = 0.05$ ، $(p_1, p_2) = (2, 3)$ ، $\lambda = 0.5$ ، $\rho = 0.2$.

برآوردگرها						Δ^2	ρ
PRSRME	SRME	IPTRME	PTRME	RRME	URME		
۱/۲۰۴۴	۱/۰۲۶۰	۱/۴۸۹۱	۱/۴۶۹۳	۱/۵۹۰۹	۱	۰	۰/۸
۱/۱۹۳۰	۱/۰۷۹۸	۱/۴۰۴۸	۱/۳۷۷۹	۱/۵۳۵۱	۱	۰/۵	
۱/۱۴۴۱	۰/۸۵۱۸	۱/۲۷۸۷	۱/۲۵۵۰	۱/۳۶۱۳	۱	۱	
۱/۰۶۷۰	۱/۰۱۸۴	۰/۹۵۹۳	۰/۹۳۳۴	۰/۹۹۱۳	۱	۲	
۱/۰۰۶۵	۰/۹۹۳۱	۰/۸۲۹۷	۰/۸۰۶۷	۰/۳۵۴۲	۱	۵	
۱/۰۰۳۳	۱/۰۰۳۳	۱/۰۰۳۳	۱/۰۰۰۰	۰/۱۰۲۳	۱	۱۰	
۱/۲۳۲۴	۰/۸۴۱۹	۱/۵۳۸۴	۱/۵۱۰۶	۱/۶۷۹۵	۱	۰	۰/۹
۱/۲۲۴۳	۰/۶۵۲۲	۱/۴۸۲۳	۱/۴۴۷۳	۱/۶۵۲۵	۱	۰/۵	
۱/۲۰۸۶	۱/۰۴۸۱	۱/۴۲۲۲	۱/۳۹۳۵	۱/۵۴۱۹	۱	۱	
۱/۱۳۳۴	۱/۰۴۹۱	۱/۱۳۷۲	۱/۱۱۰۱	۱/۲۱۰۶	۱	۲	
۱/۰۰۸۶	۰/۹۹۴۲	۰/۷۶۲۶	۰/۷۳۴۱	۰/۵۵۸۲	۱	۵	
۱/۰۰۶۱	۱/۰۰۶۱	۰/۹۸۷۳	۰/۹۷۹۲	۰/۱۷۴۶	۱	۱۰	
۱/۲۵۱۷	۰/۵۰۹۳	۱/۵۶۵۱	۱/۵۲۵۷	۱/۷۶۱۶	۱	۰	۰/۹۵
۱/۲۱۸۴	۰/۷۹۶۱	۱/۴۶۴۱	۱/۴۲۹۶	۱/۶۶۱۴	۱	۰/۵	
۱/۲۱۵۴	۰/۹۸۱۰	۱/۴۵۸۴	۱/۴۲۹۵	۱/۶۰۱۲	۱	۱	
۱/۱۷۲۵	۰/۴۳۴۷	۱/۲۷۸۱	۱/۲۴۲۹	۱/۴۲۱۹	۱	۲	
۱/۰۶۵۵	۰/۴۴۹۳	۰/۹۰۷۳	۰/۸۷۸۹	۰/۹۰۴۶	۱	۵	
۱/۰۰۸۴	۱/۰۰۲۸	۰/۷۳۲۹	۰/۷۰۹۶	۰/۳۵۸۹	۱	۱۰	
۱/۲۶۲۵	۰/۶۴۴۷	۱/۵۲۸۵	۱/۴۶۷۷	۲/۰۱۲۸	۱	۰	۰/۹۹
۱/۲۵۹۱	۰/۵۴۹۹	۱/۵۱۶۷	۱/۴۶۸۷	۱/۸۹۶۳	۱	۰/۵	
۱/۲۴۸۸	۰/۸۴۵۸	۱/۴۸۳۳	۱/۴۳۱۸	۱/۸۴۵۱	۱	۱	
۱/۲۲۸۶	۰/۸۸۶۲	۱/۴۳۶۷	۱/۳۸۶۶	۱/۷۷۹۶	۱	۲	
۱/۱۷۰۹	۰/۸۳۹۵۳	۱/۲۶۳۰	۱/۲۱۵۳	۱/۵۱۶۹	۱	۵	
۱/۰۸۲۱	۰/۹۱۷۶	۰/۹۲۰۹	۰/۸۸۹۵	۰/۸۸۴۵	۱	۱۰	

بر اساس جدول‌های ۱۲.۲ - ۱۵.۲، نتایج زیر به دست آورده می‌شود.

(۱) رفتار SRME کاملاً ناپایدار است و مخاطره آن همیشه بر برآوردگر URM برای حالت‌های $(p_1, p_2) = (2, 3), (2, 5)$ برتری ندارد و بر URME تسلط می‌یابد. سایر برآوردگرها (به عبارت دیگر، برآوردگرهای IPTRM، PTRM، RRM و PRRM) بهتر از URME حتی برای مقادیر کوچک p_1 و p_2 پیرامون فرضیه H_0 رفتار می‌کند. همیشه IPTRME بر PTRME برتری دارد. وقتی Δ^2 بزرگتر می‌شود، هر دوی این برآوردگرها به طور ضعیفی رفتار می‌کنند.

(۲) هرچه بعدهای پارامترهای (p_1, p_2) بیشتر شود، مخاطره برآوردگرهای انقباضی نسبت به URME کمتر است. SRME تسلط خود را از نظر مخاطره نزدیک کمیت مخاطره باقی می‌گذارد و پایدار خواهد بود که بزرگی برتری مخاطره PTRME و IPTRME نسبت به URME کوچکتر می‌شود.

جدول ۱۴.۲: کارایی نسبی برآوردگرهای پیشنهادی در حالت $\sigma^2 = 1, n = 5, \alpha = 0.05, \rho = 0.2$ ، $(p_1, p_2) = (2, 5), \lambda = 0.5$

برآوردگرها							Δ^2	ρ
PRSRME	SRME	IPTRME	PTRME	RRME	URME			
۱/۱۷۶۵	۱/۱۷۲۳	۱/۴۸۳۱	۱/۴۳۸۳	۱/۶۹۶۸	۱	۰	۰/۸	
۱/۱۶۹۸	۱/۱۶۴۳	۱/۴۳۱۹	۱/۳۸۵۹	۱/۶۷۳۷	۱	۰/۵		
۱/۱۵۶۸	۱/۱۵۱۶	۱/۴۰۴۲	۱/۳۵۶۱	۱/۶۵۷۳	۱	۱		
۱/۱۴۲۰	۱/۱۳۳۵	۱/۲۱۳۱	۱/۱۵۷۶	۱/۴۳۳۱	۱	۲		
۱/۰۵۶۵	۱/۰۵۶۵	۰/۸۷۹۸	۰/۸۳۹۴	۰/۶۶۱۹	۱	۵		
۱/۰۲۰۷	۱/۰۲۰۷	۰/۹۸۵۵	۰/۹۶۲۰	۰/۲۴۱۰	۱	۱۰		
۱/۲۱۲۵	۱/۲۰۸۳	۱/۵۶۲۹	۱/۴۹۲۰	۱/۹۷۷۶	۱	۰	۰/۹	
۱/۲۰۰۹	۱/۱۹۸۰	۱/۵۶۶۳	۱/۵۰۰۸	۱/۹۴۴۶	۱	۰/۵		
۱/۱۸۲۹	۱/۱۷۲۹	۱/۳۸۹۹	۱/۳۲۸۰	۱/۷۷۱۷	۱	۱		
۱/۱۶۶۴	۱/۱۶۲۳	۱/۳۴۳۳	۱/۲۷۶۵	۱/۶۷۸۵	۱	۲		
۱/۱۰۳۰	۱/۰۹۹۱	۰/۹۶۱۸	۰/۹۱۳۱	۰/۹۴۶۷	۱	۵		
۱/۰۴۲۰	۱/۰۴۲۰	۰/۸۷۱۶	۰/۸۳۲۶	۰/۳۶۳۷	۱	۱۰		
۱/۱۹۸۷	۱/۱۸۹۰	۱/۴۹۳۰	۱/۴۱۵۲	۲/۰۲۸۶	۱	۰	۰/۹۵	
۱/۲۰۸۶	۱/۱۴۷۵	۱/۵۲۱۶	۱/۴۴۶۶	۱/۹۸۲۵	۱	۰/۵		
۱/۱۹۰۷	۱/۱۸۸۱	۱/۴۸۳۸	۱/۴۱۴۷	۱/۹۲۵۷	۱	۱		
۱/۱۸۰۹	۱/۱۶۹۴	۱/۳۹۳۶	۱/۳۲۸۴	۱/۷۷۴۴	۱	۲		
۱/۱۳۶۴	۱/۱۱۸۵	۱/۱۴۸۰	۱/۰۸۷۵	۱/۳۱۸۷	۱	۵		
۱/۰۶۷۵	۱/۰۶۶۴	۰/۸۵۷۰	۰/۸۱۴۵۸	۰/۶۱۸۴	۱	۱۰		
۱/۱۷۱۵	۱/۱۶۱۷	۱/۴۳۸۸	۱/۳۷۶۱	۱/۷۸۳۳	۱	۰	۰/۹۹	
۱/۱۶۵۶	۱/۱۶۲۵	۱/۴۳۸۷	۱/۳۹۱۱	۱/۷۰۶۳	۱	۰/۵		
۱/۱۵۹۸	۱/۱۵۸۲	۱/۳۷۶۹	۱/۳۲۰۵	۱/۶۹۹۴	۱	۱		
۱/۱۵۹۰	۱/۰۷۷۲	۱/۳۱۹۷	۱/۲۶۴۲	۱/۵۵۵۱	۱	۲		
۱/۰۷۴۱	۱/۰۵۹۸	۰/۹۰۷۶	۰/۸۶۴۳	۰/۸۴۲۹	۱	۵		
۱/۰۲۵۵	۱/۰۲۵۵	۰/۹۲۱۰	۰/۸۹۳۵	۰/۳۲۳۴	۱	۱۰		

(۳) در تمامی جدول‌های تولید شده، وقتی درجه هم‌خطی، ρ افزایش یابد، تمامی برآوردگرها به جز نوع انقباضی، مخاطره را کاهش و برآوردگر را بهبود می‌دهد.

(۴) تحت فرضیه صفر، $\Delta^2 = 0$ ، تمام برآوردگرهای انقباضی کارایی بهتری را دارد. با این وجود، وقتی Δ^2 بزرگتر شود، از کارایی برآوردگر کاسته می‌گردد.

(۵) هرچه حجم نمونه n افزایش یابد، حتی برای مقادیر کوچک p_1, p_2 و ρ ، همان‌طور که انتظار می‌رفت کارایی برآوردگرها بهتر شود.

(۶) در بین همه برآوردگرها، PRRME بر URME برتری دارد.

جدول ۱۵.۲: کارایی نسبی برآوردگرهای پیشنهادی در حالت $n = 100$ ، $\sigma^2 = 1$ ، $\alpha = 0.05$ ، $\rho = 0.2$ ، $\lambda = 0.5$ ، $(p_1, p_2) = (2, 5)$.

برآوردگرها						Δ^2	ρ
PRSRME	SRME	IPTRME	PTRME	RRME	URME		
۱/۱۵۵۰	۱/۱۵۱۷	۱/۳۹۶۸	۱/۳۴۳۰	۱/۶۹۵۰	۱	۰	۰/۸
۱/۱۵۲۰	۱/۱۴۷۵	۱/۴۰۵۵	۱/۳۶۶۸	۱/۵۹۸۷	۱	۰/۵	
۱/۱۴۲۳	۱/۱۴۱۸	۱/۳۰۵۲	۱/۲۶۱۵	۱/۵۱۸۷	۱	۱	
۱/۰۸۶۷	۱/۰۸۲۴	۱/۰۴۷۰	۱/۰۰۶۷	۱/۰۹۴۹	۱	۲	
۱/۰۲۲۸	۱/۰۲۲۸	۰/۹۲۴۷	۰/۸۹۶۹	۰/۳۹۵۵	۱	۵	
۱/۰۰۳۸	۱/۰۰۳۸	۱/۰۰۳۸	۱/۰۰۰۰	۰/۱۱۲۴	۱	۱۰	
۱/۱۹۴۴	۱/۱۸۹۱	۱/۴۷۳۳	۱/۴۱۸۱	۱/۸۰۳۹	۱	۰	۰/۹
۱/۱۷۸۹	۱/۱۷۵۳	۱/۴۷۳۴	۱/۴۲۲۷	۱/۷۴۵۲	۱	۰/۵	
۱/۱۴۹۳	۱/۱۳۶۶	۱/۲۸۱۷۳	۱/۲۴۰۶	۱/۵۰۷۷	۱	۱	
۱/۱۲۳۲	۱/۱۱۲۴	۱/۱۶۱۶	۱/۱۱۳۹	۱/۳۲۴۸	۱	۲	
۱/۰۳۳۸	۱/۰۳۳۴	۰/۸۵۱۰	۰/۸۱۸۳	۰/۵۹۹۷	۱	۵	
۱/۰۰۸۴	۱/۰۰۸۴	۰/۹۷۵۷	۰/۹۶۳۴	۰/۲۰۲۳	۱	۱۰	
۱/۱۸۰۵	۱/۱۵۷۳	۱/۴۲۵۶	۱/۳۵۵۴	۱/۹۳۷۴	۱	۰	۰/۹۵
۱/۱۹۴۳	۱/۱۸۵۸	۱/۴۷۵۵	۱/۴۱۵۰	۱/۸۵۵۴	۱	۰/۵	
۱/۱۷۷۵	۱/۱۶۳۰	۱/۴۴۸۲	۱/۳۸۹۵	۱/۸۱۱۳	۱	۱	
۱/۱۵۷۴	۱/۱۵۲۱	۱/۳۰۰۲	۱/۲۴۷۴	۱/۵۸۴۰	۱	۲	
۱/۰۷۵۹	۱/۰۶۲۵	۰/۹۳۲۲	۰/۸۹۳۹	۰/۸۵۲۸	۱	۵	
۱/۰۱۵۲	۱/۰۱۵۲	۰/۸۴۴۱	۰/۸۲۰۸	۰/۳۳۱۹	۱	۱۰	
۱/۱۸۹۶	۱/۰۶۴۰	۱/۴۰۳۱	۱/۳۳۲۷	۲/۰۰۷۴	۱	۰	۰/۹۹
۱/۱۷۸۵	۱/۱۱۴۶	۱/۴۱۵۵	۱/۳۴۵۵	۱/۹۷۹۹	۱	۰/۵	
۱/۱۷۵۱	۱/۰۵۷۷	۱/۳۷۹۵	۱/۳۱۰۹	۱/۹۶۲۷	۱	۱	
۱/۱۶۶۶	۱/۱۶۴۱	۱/۳۶۸۱	۱/۳۰۴۳	۱/۸۸۷۵	۱	۲	
۱/۱۵۰۹	۱/۱۴۸۶	۱/۲۳۴۴	۱/۱۷۸۶	۱/۵۸۸۴	۱	۵	
۱/۱۱۴۰	۱/۱۰۰۲	۱/۰۳۵۲	۰/۹۹۰۶	۱/۰۲۱۱۸	۱	۱۰	

به‌عنوان یک نتیجه، در حالتی که هم‌خطی وجود دارد و توزیع خطاها غیر نرمال باشد، پیشنهاد می‌شود که از برآوردگرهای PRRM و IPTRM برای برآورد β_1 استفاده گردد. در واقع، SRME فقط به‌عنوان یک ابزار برای ساختن این برآوردها مورد استفاده قرار می‌گیرد. نمی‌توان به‌طور صریحی بیان کرد که کدام‌یک از این برآوردگرها بهتر است. گاهی اوقات IPTRME بهتر رفتار می‌کند و گاهی اوقات PRSRME. وقتی از فرضیه صفر فاصله گرفته شود، به‌نظر می‌رسد که برآوردگر PRSRM بهتر رفتار می‌کند.

۴.۴.۲ تحلیل مجدد داده‌های چربی بدن

در اینجا مجدداً داده‌های چربی بدن را تحلیل می‌کنیم که در صفحه ۶۷ به‌طور کامل معرفی شدند. برآوردگرهای M و M ریح به‌ازای $\lambda = 0.22$ که با استفاده از روش تکراری مطرح شده در بخش ۲.۴.۲ به‌دست آمده است، در جدول ۱۶.۲ گزارش شده است.

جدول ۱۶.۲: برآوردگرهای M و M ریج برای مجموعه داده‌های چربی بدن

نماد	متغیرها	M	M ریج
β_1	DENSITY	-۴۴۲/۰۴	-۴۳۱/۴۴
β_2	AGE	۰/۰۰۰۷۴	۰/۰۰۰۲۹
β_3	WEIGHT	۰/۰۰۰۱۸۴	۰/۰۰۰۳۲۷
β_4	HEIGHT	-۰/۰۰۰۹۳۶	-۰/۰۰۰۵۰۲
β_5	NECK	۰/۰۰۰۲۵۷	-۰/۰۰۰۵۷۲
β_6	CHEST	۰/۰۰۰۱۰۸	-۰/۰۰۰۲۹۸
β_7	ABDOMEN	۰/۰۰۰۰۵۳	۰/۰۰۰۲۴۷
β_8	HIP	۰/۰۰۰۰۹۳۵	-۰/۰۰۰۵۷۸
β_9	THIGH	-۰/۰۰۰۱۱۳۳	-۰/۰۰۰۱۲۳
β_{10}	KNEE	-۰/۰۰۰۰۶۸۱	-۰/۰۰۰۰۴۵۳
β_{11}	ANKLE	۰/۰۰۰۰۳۳۷	-۰/۰۰۰۰۱۷۵
β_{12}	BICEPS	-۰/۰۰۰۰۴۹۲	-۰/۰۰۰۰۰۶
β_{13}	FOREARM	-۰/۰۰۰۰۰۰۵	-۰/۰۰۰۰۰۶
β_{14}	WRIST	-۰/۰۰۰۱۴۳۶	-۰/۰۰۰۰۱۲۴

بر اساس نتایج جدول ۱۶.۲، می‌توان حدس زد که نیازی نیست همه متغیرها در اندازه‌گیری درصد چربی بدن دخیل باشند و هشت متغیر اضافی خواهد بود. (البته فرض می‌کنیم) می‌توان این حدس را که طبق NSI به‌دست آمده‌اند به کار برد و فرضیه صفر $H_0: \beta_2 = 0$ را در نظر گرفت که

$$\beta = \left(\beta_1 : \beta_2 \right)^T = \left(\beta_9, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14} : \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8 \right)^T.$$

دقت کنید که ابتدا باید چینش متغیرها در ساختار کوواریانس تغییر یابد که استراتژی‌های ما به کار برده شود.

مدل رگرسیون خطی برای پیشگویی متغیرهای مورد علاقه برآزش داده می‌شود. برآوردگرهای UM، URM، PTRM، IPTRM، SRM و PRSRM برای پارامترهای رگرسیونی محاسبه شده و کارایی برآوردگرها بر اساس میانگین خطای اعتبارسنجی متقابل^{۱۰} تایی به‌دست آمده است. در اعتبارسنجی متقابل، خطای پیشگویی برآورد شده در اجراهای مختلف تغییر می‌کند. بنابراین، فرآیند ۱۰۰۰ بار تکرار و میانگین خطاهای پیشگویی محاسبه شده است. تعداد ۱۰۰۰ اجرا به اندازه کافی بزرگ است تا انحراف استانداردها پایدار شود بدین معنی که هیچ تغییر قابل توجهی برای مقادیر بزرگ مشاهده نشود. مقادیر خطای پیشگویی نسبی در جدول ۱۷.۲ بیان شده است.

جدول ۱۷.۲: مقادیر RPE برآوردگرهای پیشنهادی مجموعه داده چربی بدن

برآوردگرها	RRME	PTRME	IPTRME	SRME	PRSRME
RPE	۱/۰۰۰۰۹۷۹	۱/۰۰۰۰۹۷۳	۱/۰۰۰۱۱۰۴	۱/۰۰۰۱۱۰۱	۱/۰۰۰۱۱۰۱

همان‌طور که ملاحظه می‌شود، برآوردگرهای SRME، PRSRME و IPTRME بهتر از URME و RRME رفتار می‌کنند.

۵.۲ نتیجه‌گیری

در این فصل، مسئله‌ای مورد بررسی قرار گرفت که همزمان هم خطی و داده پرت در مجموعه متغیرهای توضیحی در مدل بندی وجود داشت. هم‌چنین فرض کردیم اطلاع پیشین غیرقطعی درباره پارامترها نیز در این داده‌ها وجود دارد. در چنین موقعیتی، برآوردگر انقباضی نوع استاین ضعیف رفتار می‌کند. به عنوان یک راه‌حل، پیشنهاد شد از برآوردگر نوع استاین فقط به عنوان یک ابزار برای ساختن دو کلاس از برآوردگرهای انقباضی شامل IPTME و PRRME استفاده شود که بهتر از URME رفتار می‌کند.

طبق نتایج شبیه‌سازی، تصویری از برتری برآوردگرهای پیشنهادی تحت درستی اطلاع پیشین $\beta = 0$ (مشابه فرضیه صفر، $H_0: \beta = 0$)

$$\hat{\beta}^{rM}(\lambda) > \hat{\beta}^{iptM}(\lambda) > \hat{\beta}^{ptM}(\lambda) > \hat{\beta}^{sM+}(\lambda) > \hat{\beta}^{sM}(\lambda) > \hat{\beta}^{uM}$$

تحت $\beta \neq 0$ (مشابه فرضیه مقابل $H_A: \beta \neq 0$)

$$\hat{\beta}^{sM+}(\lambda) > \hat{\beta}^{sM}(\lambda) > \hat{\beta}^{uM}(\lambda) > \hat{\beta}^{iptM}(\lambda) > \hat{\beta}^{ptM}(\lambda) > \hat{\beta}^{rM}(\lambda)$$

که $>$ تسلط را از منظر کمتر بودن میزان مخاطره نشان می‌دهد.

لازم به ذکر است اگر فرضیه صفر $H_0: \beta = 0$ درست باشد، آن‌گاه مدل رگرسیونی معنی‌دار نبوده و هیچ‌کدام از برآوردگرهایی که در این مجموعه معرفی نشده‌اند کارایی ندارند اما نوآوری این مجموعه (که در تمامی فصل‌ها به آن پرداخته شده است) این است که از اطلاع (نادرست) $\beta = 0$ استفاده کردیم و برآوردگرهایی را ارائه دادیم که از منظر کمتر بودن میزان مخاطره، کارایی بیشتر نسبت به برآوردگرهای معمول داشتند. این ادعای افزایش کارایی برای اطلاع $\beta_2 = 0$ نیز برقرار است. در صورتی که فرضیه صفر $H_0: \beta_1 = 0$ نشان نمی‌دهد که مدل رگرسیون معنی‌دار نیست بلکه کاربرد زیادی در انتخاب متغیر و تحلیل مسایل رگرسیونی در بعد بالا دارد.

فصل ۳

برآوردگرهای بهبودیافته بر پایه برآوردگر لاسو

در این فصل، ایده دوباره انقباضی در برآوردگر لاسو مطرح شده است. برآوردگرهای پیشنهادی معرفی، ویژگی‌های آن‌ها بررسی و در یک مطالعه شبیه‌سازی و مثال واقعی تحلیل شده‌اند. از این فصل، مقاله‌های [نوروزی‌راد و آرشی \(۱۳۹۶\)](#)، [نوروزی‌راد و آرشی \(۲۰۱۷a\)](#) و [نوروزی‌راد و همکاران \(۲۰۱۵b\)](#) نتیجه شده است.

۱.۳ مقدمه

روش‌های استاندارد برای برآورد پارامتر β در مدل رگرسیون خطی (۱.۱) بر اساس این فرضیه است که ماتریس $C_n = X_n^T X_n$ نامنفرد و وارون‌پذیر باشد. در صورتی که وارون C_n وجود نداشته باشد، پارامترها به‌طور منحصر به‌فرد برآورد نمی‌شوند. روش رگرسیونی OLS، ترکیب خطی ناریب از X_n را نتیجه می‌دهد که مجموع توان دوم باقیمانده‌ها را کمینه می‌کند. با این وجود، اگر تعداد متغیرها، p زیاد باشد یا به‌شدت همبسته باشند (هم‌خطی وجود داشته باشد)، برآوردگر OLS منجر به برآوردهایی با واریانس بزرگ می‌شود که دقت پیشگویی را کاهش می‌دهد. در این حالت روش‌های پیشنهادی عبارتند از:

- (۱) انتخاب متغیر، مانند روش انتخاب بهترین زیرمجموعه (میلر، ۲۰۰۲)،
- (۲) روش‌های کاهش بعد، مانند رگرسیون مولفه‌های اصلی یا رگرسیون خطی جزئی،
- (۳) روش‌های هموارسازی مانند ریج (هورل و کنارد، ۱۹۷۰)، لاسو (تیبشیرانی، ۱۹۹۶)

اگر از بین مدل‌های ممکن برای p متغیر، هدف یافتن مدل بهینه باشد، می‌توان از روش انتخاب متغیر استفاده نمود. رایج‌ترین رهیافت مورد استفاده، پیش‌انتخاب یا انتخاب زیرمجموعه است. در این روش، به‌سادگی یک مجموعه کوچکتر از متغیرها طبق یک معیار انتخاب می‌شود و مدلی (استاندارد) به این متغیرهای انتخابی برازش داده می‌شود. وقتی p بزرگ (بزرگتر از ۱۰۰) باشد، این رهیافت از لحاظ محاسباتی پیچیده می‌شود. روش‌های متفاوتی برای مواجهه با این مسئله وجود دارد. برخی از این روش‌ها عبارتند از: انتخاب پیش‌رو، انتخاب پس‌رو یا روش گام به گام. اما این روش‌ها ناپایدار هستند. بدین معنی که تغییرات اندکی در داده‌ها منجر به تغییرات بزرگی در برآورد می‌شود (برایمن، ۱۹۹۶). این روش از یک قاعده تصمیم سخت (متغیر می‌میرد یا زنده می‌ماند) استفاده می‌کند.

رهیافت دوم از روش‌هایی مانند رگرسیون مولفه‌های اصلی یا رگرسیون خطی جزئی استفاده می‌کند. این روش‌ها تعداد اندکی از ترکیب‌های خطی از متغیرهای توضیحی را انتخاب می‌کند و از این ترکیب‌های خطی به عنوان جایگزین متغیرهای اصلی استفاده می‌کند. استفاده از این مدل‌ها به منظور پیشگویی، منطقی است اما اغلب تفسیرپذیری این مدل‌ها دشوار است (هیستی و همکاران، ۲۰۰۹).

در روش‌های هموارسازی، ضرایب در مقایسه با برآوردهای به‌دست آمده از روش انتخاب بهترین زیرمجموعه، منقبض می‌شوند. این رهیافت، پیوسته است و بنابراین، برآوردی با واریانس کمتری از برآوردهای انتخاب بهترین زیرمجموعه به‌دست می‌آید و همچنین خطای پیشگویی مدل کلی را کاهش می‌دهد. روش‌های انقباضی، اغلب دقت پیشگویی را افزایش می‌دهد.

برآوردگر لاسو، ضرایب را به سمت صفر منقبض کرده و متغیرهای اضافی را دقیقاً صفر می‌کند. هنس (۲۰۱۶) مقایسه‌ای بین برآوردگرهای لاسو، انقباضی نوع استاین و بهترین انتخاب بر اساس مخاطره کمترین توان‌های دوم انجام داد. وی نتیجه گرفت که برآوردگر لاسو و برآوردگر نوع استاین به‌طور یکنواخت بر دیگری برتری ندارد. همچنین در مطالعه شبیه‌سازی نشان داد که برآوردگر لاسو به پارامترهای مدل نسبتاً حساس است و در بخشی از فضای پارامتر، برآوردگر لاسو مخاطره نسبتاً بیشتری نسبت به برآوردگر OLS دارد. لازم به ذکر است علاوه بر برآوردگر لاسو، می‌توان از برآوردگرهای جایگزین دیگری مانند لاسوی تطبیقی (ژو، ۲۰۰۶)، SCAD (فن و لی، ۲۰۰۱)، Enet (ژو و هیستی، ۲۰۰۵) و ... استفاده کرد.

هنگامی که علاوه بر اطلاعات به‌دست آمده از نمونه، از سایر اطلاعات در برآورد پارامترها استفاده شده است، معمولاً نتایج پرباری به‌همراه دارد. در این حالت، موضوع دوباره انقباضی مطرح می‌شود و برآوردگر لاسو، به‌سمت برآوردگری که بر اساس NSI یا UPI به‌دست آمده است، منقبض می‌شود. در حقیقت در این فصل برآورد لاسو را به سمت یک زیرفضا که همان محدودیت است، منقبض می‌کنیم.

۱.۱.۳ برآوردگرهای انقباضی بر پایه لاسو

از آن جایی که برآوردگر لاسو وابسته به UPI نیست، از این پس، آن را برآوردگر محدود نشده لاسو (ULE) نامیده و از نماد $\hat{\beta}^{uLso}$ برای نمایش آن استفاده خواهد شد. در برخی موارد ممکن است که NSI روی پارامترهای مدل به عنوان محدودیت وجود داشته باشند. فرض کنید اطلاعات غیرنمونه‌ای را بتوان به صورت زیر بیان نمود:

$$H\beta = h \quad (1.3)$$

که H ماتریس $q \times p$ با رتبه کامل سطری q ، $(q \leq p)$ و بردار h بردار $q \times 1$ از ثابت‌های معلوم است. چنانچه برآوردگر لاسو در رابطه (۱.۳) صدق کند، برآوردگر محدود شده لاسو (RLE) نامیده شده و با $\hat{\beta}^{rLso}$ نشان داده می‌شود. مشابه برآوردگر محدود شده OLS پارامتر β ، نسبت به محدودیت $H\beta = h$ (صالح، ۲۰۰۶)، برآوردگر محدود شده لاسو به صورت زیر پیشنهاد می‌شود:

$$\hat{\beta}^{rLso} = \hat{\beta}^{uLso} - C_n^{-1} H^T (HC_n^{-1} H^T)^{-1} (H\hat{\beta}^{uLso} - h).$$

بر اساس استراتژی بیان شده، برآوردگر آزمون اولیه لاسو (PTLE)، برآوردگر انقباضی نوع استاین لاسو (SLE) و برآوردگر جزء مثبت نوع استاین لاسو (PRLE) به ترتیب به صورت زیر خواهند بود:

$$\hat{\beta}^{ptLso} = \hat{\beta}^{uLso} - (\hat{\beta}^{uLso} - \hat{\beta}^{rLso}) I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha}) \quad (2.3)$$

$$\hat{\beta}^{sLso} = \hat{\beta}^{uLso} - c_n \mathcal{L}_n^{-1} (\hat{\beta}^{uLso} - \hat{\beta}^{rLso}), \quad (3.3)$$

$$\hat{\beta}^{sLso+} = \hat{\beta}^{sLso} - (1 - c_n \mathcal{L}_n^{-1}) I(\mathcal{L}_n \leq c_n) (\hat{\beta}^{uLso} - \hat{\beta}^{rLso}) \quad (4.3)$$

که $\mathcal{L}_{n,\alpha}$ چندک α ام توزیع آماره آزمون \mathcal{L}_n ، برای آزمون $H\beta = h$ ، تحت فرضیه H_0 است. همین‌طور، ثابت انقباضی است و در یک بازه به گونه‌ای انتخاب می‌شود که در آن $\hat{\beta}^{sLso}$ بر برآوردگر $\hat{\beta}^{uLso}$ برتری داشته باشد. پیشنهاد می‌شود مقدار $m(q-2)/(m+2)$ برای آن انتخاب شود که $m = n - p$.

در برآوردگر آزمون اولیه، $\hat{\beta}^{uLso}$ را می‌توان با $\hat{\beta}^{sLso}$ جایگذاری نمود که منجر به بهبودی این برآوردگر می‌شود. **سالو و همکاران (۱۹۷۲)** بر این اساس برآوردگر آزمون اولیه بهبودیافته را مشابه برآوردگر OLS پیشنهاد دادند. برآوردگر آزمون اولیه بهبودیافته لاسو (IPTLE) به صورت زیر می‌توان تعریف نمود:

$$\begin{aligned} \hat{\beta}^{iptLso} &= \hat{\beta}^{sLso} - (\hat{\beta}^{sLso} - \hat{\beta}^{rLso}) I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha}) \\ &= \hat{\beta}^{ptLso} - c_n \mathcal{L}_n^{-1} I(\mathcal{L}_n > \mathcal{L}_{n,\alpha}) (\hat{\beta}^{uLso} - \hat{\beta}^{rLso}) \end{aligned}$$

۲.۳ ویژگی برآوردگرها

برای به دست آوردن خواص مجانبی برآوردگرهای بهبودیافته پیشنهادی مشابه بخش ۱.۳.۲، مجموعه فرضیه‌های مقابل موضعی را به صورت زیر در نظر بگیرید.

$$\mathcal{K}_{(n)} : H\beta = h + \frac{\xi}{\sqrt{n}}, \quad \xi = (\xi_1, \dots, \xi_p)^\top \neq o \quad (5.3)$$

از آن جا که انتخاب $\xi = o$ نتیجه می‌دهد $H\beta = h$ ، می‌توان (۱.۳) را حالت خاصی از کلاس $\mathcal{K}_{(n)}$ دانست. برای محاسبه ADB و ADR به ترتیب از رابطه‌های (۱۶.۲) و (۱۸.۲) استفاده می‌شود. ADQB برآوردگرها به صورت زیر تعریف می‌شود:

$$\text{ADQB}(\hat{\beta}^*) = \sigma^{-2} [\text{ADB}(\hat{\beta}^*)]^\top C [\text{ADB}(\hat{\beta}^*)] \quad (6.3)$$

فرض کنید شرایط نظم (۱) و (۲) در بخش ۳.۲ برقرار باشند. هم‌چنین می‌توان آماره آزمون را به صورت

$$\mathcal{L}_n = \frac{(H\hat{\beta}^{\text{uLso}} - h)^\top (HC_n^{-1}H^\top)^{-1} (H\hat{\beta}^{\text{uLso}} - h)}{s_L^2} \quad (7.3)$$

در نظر گرفت که در آن

$$s_L^2 = \frac{1}{m} (y - X\hat{\beta}^{\text{uLso}})^\top (y - X\hat{\beta}^{\text{uLso}}) \quad (8.3)$$

قضیه ۱.۲.۳. با برقراری شرایط نظم، در کلاس $\mathcal{K}_{(n)}$ هنگامی که $n \rightarrow \infty$ ، اگر C یک ماتریس وارون پذیر باشد، به ازای $\lambda_n = o(n^{1/2})$ نتایج زیر برقرار است:

- (۱) $\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) \xrightarrow{\mathcal{D}} W$
- (۲) $\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta) \xrightarrow{\mathcal{D}} W - C^{-1}H^\top (HC^{-1}H^\top)^{-1}(HW + \xi)$
- (۳) $\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \xrightarrow{\mathcal{D}} C^{-1}H^\top (HC^{-1}H^\top)^{-1}(HW + \xi)$
- (۴) $\lim_{n \rightarrow \infty} P(\mathcal{L}_n \leq x) = H_q(x; \Delta^2)$
- (۵) $\sqrt{n}(\hat{\beta}^{\text{ptLso}} - \beta) \xrightarrow{\mathcal{D}} W - C^{-1}H^\top (HC^{-1}H^\top)^{-1}(HW + \xi)I(\mathcal{L} \leq \mathcal{L}_\alpha)$
- (۶) $\sqrt{n}(\hat{\beta}^{\text{sLso}} - \beta) \xrightarrow{\mathcal{D}} W - cC^{-1}H^\top (HC^{-1}H^\top)^{-1}(HW + \xi)\mathcal{L}^{-1}$
- (۷) $\sqrt{n}(\hat{\beta}^{\text{iptLso}} - \beta) \xrightarrow{\mathcal{D}} W - (c\mathcal{L}^{-1} + (1 - c\mathcal{L}^{-1})I(\mathcal{L} \leq \mathcal{L}_\alpha)) \times C^{-1}H^\top (HC^{-1}H^\top)^{-1}(HW + \xi)$
- (۸) $\sqrt{n}(\hat{\beta}^{\text{sLso}+} - \beta) \xrightarrow{\mathcal{D}} W - (c\mathcal{L}^{-1} + (1 - c\mathcal{L}^{-1})I(\mathcal{L} \leq c)) \times C^{-1}H^\top (HC^{-1}H^\top)^{-1}(HW + \xi)$

که در آن $\Delta^2 = \sigma^{-2}\xi^\top (HC^{-1}H^\top)\xi = \sigma^{-2}\delta^\top \delta$ ، $\delta = C^{-1}H^\top (HC^{-1}H^\top)^{-1}\xi$

$$\mathcal{L} = \frac{(HW + \xi)^\top (HC^{-1}H^\top)^{-1}(HW + \xi)}{\sigma^2}$$

و $W \sim \mathcal{N}_p(o, \sigma^2 C^{-1})$

برهان. به منظور اثبات قسمت (۱) این قضیه، نایت و فو (۲۰۰۰) نشان دادند که به ازای بردار دلخواه $\mathbf{u} = (u_1, \dots, u_p)^\top$ ، اگر $\lambda_n / \sqrt{n} \rightarrow \lambda_0 > 0$ ، آن گاه

$$\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) \xrightarrow{D} \operatorname{argmin}_{\mathbf{u}} (V(\mathbf{u})),$$

که در آن

$$V(\mathbf{u}) = -\mathbf{2u}^\top \mathbf{V} + \mathbf{u}^\top \mathbf{C} \mathbf{u} + \lambda_0 \sum_{j=1}^p [u_j \operatorname{sgn}(\beta_j) I(\beta_j \neq 0) + |u_j| I(\beta_j = 0)],$$

و $\mathbf{V} \sim \mathcal{N}_p(0, \sigma^2 \mathbf{C})$.

در حالتی که $\lambda_0 = 0$ ، $\lambda_n = o(n^{1/2})$ و بنابراین $\operatorname{argmin}_{\mathbf{u}} V(\mathbf{u}) = \mathbf{C}^{-1} \mathbf{V}$ که هم توزیع با $\mathbf{W} \sim \mathcal{N}_p(0, \sigma^2 \mathbf{C}^{-1})$ می باشد. به منظور اثبات (۲)، طبق رابطه (۲.۳)،

$$\begin{aligned} \sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta) &= \sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) \\ &\quad - \sqrt{n} \mathbf{C}_n^{-1} \mathbf{H}^\top (\mathbf{H} \mathbf{C}_n^{-1} \mathbf{H}^\top)^{-1} [\mathbf{H}(\hat{\beta}^{\text{uLso}} - \beta) + (\mathbf{H}\beta - \mathbf{h})] \\ &= \sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) \\ &\quad - n \mathbf{C}_n^{-1} \mathbf{H}^\top (n \mathbf{H} \mathbf{C}_n^{-1} \mathbf{H}^\top)^{-1} [\mathbf{H} \sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) + \sqrt{n}(\mathbf{H}\beta - \mathbf{h})], \end{aligned}$$

با استفاده از قضیه ۳.۳.آ و قسمت (۱) این قضیه،

$$\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta) \xrightarrow{D} \mathbf{W} - \mathbf{C}^{-1} \mathbf{H}^\top (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}^\top)^{-1} [\mathbf{H} \mathbf{W} + \delta].$$

برای اثبات (۳) می توان نوشت

$$\begin{aligned} \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) &= \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{uLso}} + \mathbf{C}_n^{-1} \mathbf{H}^\top (\mathbf{H} \mathbf{C}_n^{-1} \mathbf{H}^\top)^{-1} (\mathbf{H} \hat{\beta}^{\text{uLso}} - \mathbf{h})) \\ &= \mathbf{C}_n^{-1} \mathbf{H}^\top (\mathbf{H} \mathbf{C}_n^{-1} \mathbf{H}^\top)^{-1} (\mathbf{H} \sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) + \delta). \end{aligned}$$

نتیجه بر اساس قضیه ۳.۳.آ و قسمت (۲) برقرار است.

برای اثبات (۴)، می توان آماره آزمون \mathcal{L}_n را به صورت زیر نوشت:

$$\frac{(\mathbf{H}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) + \sqrt{n}(\mathbf{H}\beta - \mathbf{h})))^\top (\mathbf{H}(n \mathbf{C}_n^{-1}) \mathbf{H}^\top)^{-1} (\mathbf{H}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) + \sqrt{n}(\mathbf{H}\beta - \mathbf{h})))}{s_L^2} \quad (۹.۳)$$

بر اساس قسمت (۱)، $\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta)$ توزیع مجانبی مشابه متغیر \mathbf{W} دارد. بنابراین، تحت کلاس $\mathcal{K}_{(n)}$ و شرایط نظم، صورت کسر (۹.۳) با عبارت زیر هم توزیع است.

$$(\mathbf{H} \mathbf{W} + \xi)^\top (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}^\top)^{-1} (\mathbf{H} \mathbf{W} + \xi).$$

از طرفی بدیهی است که s_L^2 به σ^2 همگراست. طبق قضیه ۳.۳.آ، آماره آزمون \mathcal{L}_n به \mathcal{L} همگراست که به سادگی می توان نشان داد \mathcal{L} ، توزیع کی دو با q درجه آزادی و پارامتر غیرمرکزی Δ^2 می باشد.

برای اثبات (۵)، از این حقیقت استفاده می‌شود که $I(\mathcal{L} \leq \mathcal{L}_\alpha) \xrightarrow{\mathcal{D}} I(\mathcal{L}_n \leq \mathcal{L}_n(\alpha))$ ، طبق قسمت (۱) و (۳) و قضیه ۳.۳.آ، $\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta)$ در توزیع به رابطه زیر همگرا است

$$W - C^{-1}H^T(HC^{-1}H^T)^{-1}(HW + \delta)I(\mathcal{L} \leq \mathcal{L}_\alpha).$$

برای اثبات (۶)،

$$\sqrt{n}(\hat{\beta}^{\text{sLso}} - \beta) = \sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) - c_n \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})\mathcal{L}_n^{-1}.$$

با توجه به قسمت (۱) و (۳)، $\mathcal{L}_n^{-1} \xrightarrow{\mathcal{D}} \mathcal{L}$ ، $c_n \xrightarrow{\mathcal{D}} c$ و به کار بردن قضیه ۳.۳.آ، می‌توان نوشت

$$\sqrt{n}(\hat{\beta}^{\text{sLso}} - \beta) \xrightarrow{\mathcal{D}} W - c \left[C^{-1}H^T(HC^{-1}H^T)^{-1}(HW + \delta) \right] \mathcal{L}^{-1}.$$

برای اثبات (۷)،

$$\sqrt{n}(\hat{\beta}^{\text{iptLso}} - \beta) = \sqrt{n} \left(\hat{\beta}^{\text{ptLso}} - \beta - c_n \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})\mathcal{L}_n^{-1}I(\mathcal{L}_n > \mathcal{L}_n(\alpha)) \right).$$

مشابه قسمت (۵)، $I(\mathcal{L}_n > \mathcal{L}_n(\alpha)) \xrightarrow{\mathcal{D}} I(\mathcal{L} > \mathcal{L}_\alpha)$ ، اثبات $\sqrt{n}(\hat{\beta}^{\text{iptLso}} - \beta)$ ساده است. می‌توان PRLE را به صورت زیر بازنویسی نمود:

$$\hat{\beta}^{\text{sLso}+} = \hat{\beta}^{\text{uLso}} - (c_n \mathcal{L}_n + (1 - c_n \mathcal{L}_n)I(\mathcal{L}_n \leq c_n))(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}).$$

از آن جایی که $\mathcal{L}_n \rightarrow \mathcal{L}$ ، $c_n \rightarrow c$ و بنا به قسمت (۱) و به کار بردن قضیه ۳.۳.آ، قسمت (۸) به راحتی اثبات می‌شود. \square

لم ۱.۲.۳. تحت مجموعه فرضیه‌های مقابل موضعی $\mathcal{K}_{(n)}$

$$(۱) \quad \sqrt{n}(\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h}) \xrightarrow{\mathcal{D}} \mathcal{N}_q(\boldsymbol{\xi}, \sigma^2(\mathbf{H}C^{-1}\mathbf{H}^T))$$

$$(۲) \quad \begin{bmatrix} \sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) \\ \sqrt{n}(\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h}) \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_{p+q} \left(\begin{bmatrix} \circ \\ \boldsymbol{\xi} \end{bmatrix}, \sigma^2 \begin{bmatrix} C^{-1} & C^{-1}\mathbf{H}^T \\ \mathbf{H}C^{-1} & \mathbf{H}C^{-1}\mathbf{H}^T \end{bmatrix} \right)$$

$$(۳) \quad \sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}_p(-\delta, \sigma^2 \mathbf{A})$$

$$(۴) \quad \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\delta, \sigma^2(C^{-1} - \mathbf{A}))$$

$$(۵) \quad \begin{bmatrix} \sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta) \\ \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_{2p} \left(\begin{bmatrix} -\delta \\ \delta \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{A} & \circ \\ \circ & C^{-1} - \mathbf{A} \end{bmatrix} \right)$$

برهان. برای اثبات قسمت (۱)، می توان نوشت:

$$\begin{aligned}\sqrt{n}(\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h}) &= \sqrt{n}(\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{H}\beta + \mathbf{H}\beta - \mathbf{h}) \\ &= \mathbf{H} \left(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) \right) + \sqrt{n}(\mathbf{H}\beta - \mathbf{h}).\end{aligned}\quad (10.3)$$

طبق قسمت (۱) قضیه ۱.۲.۳، عبارت $\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta)$ هم توزیع با متغیر W است که همان توزیع $\mathcal{N}_p(o, \sigma^2 C^{-1})$ می باشد. بنا به ویژگی های توزیع نرمال، عبارت (۱۰.۳) هم توزیع با $\xi + HW$ است و بنابراین

$$\sqrt{n}(\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h}) \xrightarrow{D} \mathcal{N}_q(\xi, \sigma^2(\mathbf{H}C^{-1}\mathbf{H}^\top)).\quad (11.3)$$

در اثبات قسمت (۲)، کفایت $\text{Cov}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta), \sqrt{n}(\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h}))$ محاسبه گردد. بدین منظور می توان نوشت

$$\begin{aligned}\text{Cov}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta), \sqrt{n}(\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h})) \\ &= \text{Cov}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta), \sqrt{n}(\mathbf{H}(\hat{\beta}^{\text{uLso}} - \beta) + \mathbf{H}\beta - \mathbf{h})) \\ &= \text{Var}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta))\mathbf{H}^\top + \text{Cov}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta), \xi) = \sigma^2 C^{-1}\mathbf{H}.\end{aligned}$$

بنا به رابطه (۲.۳)، عبارت $\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta)$ به طور مجانبی توزیع نرمال با میانگین

$$o - C^{-1}\mathbf{H}(\mathbf{H}C^{-1}\mathbf{H}^\top)^{-1}\xi = -\delta,$$

و ماتریس واریانس-کوواریانس $A = \sigma^2 C^{-1}\mathbf{H}^\top(\mathbf{H}C^{-1}\mathbf{H}^\top)^{-1}\mathbf{H}C^{-1}$ دارد که به راحتی از قسمت (۲) همین قضیه نتیجه می شود و اثبات قسمت (۳) کامل می شود. برای اثبات قسمت (۴)، می توان عبارت $\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})$ را بر اساس رابطه (۲.۳) به صورت $\sqrt{n}\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h}$ نوشت که بنا به ویژگی های توزیع نرمال و قسمت (۱) این قضیه، به طور مجانبی دارای توزیع نرمال با میانگین δ و ماتریس واریانس-کوواریانس

$$\sigma^2 C^{-1}\mathbf{H}^\top(\mathbf{H}C^{-1}\mathbf{H}^\top)^{-1}(\mathbf{H}C^{-1}\mathbf{H}^\top)(\mathbf{H}C^{-1}\mathbf{H}^\top)^{-1}\mathbf{H}C^{-1} = \sigma^2(C^{-1} - A)$$

می باشد. به منظور اثبات قسمت (۵)، فقط باید نشان داد

$$\text{Cov}(\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta), \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})) = o.$$

بدین منظور

$$\begin{aligned}\text{Cov}(\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta), \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})) \\ &= \text{Cov}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta) - (\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})), \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})) \\ &= \text{Cov}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta), \sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})) - \text{Var}(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}))\end{aligned}$$

$$= \text{Cov} \left(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \beta), C_n^{-1} H^\top (HC^{-1} H^\top)^{-1} \sqrt{n}(H\hat{\beta}^{\text{uLso}} - h) \right) \\ - \text{Var} \left(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \right),$$

که وقتی $n \rightarrow \infty$ ، به

$$\sigma^2 C_n^{-1} H^\top (HC^{-1} H^\top)^{-1} HC^{-1} - \sigma^2 (C^{-1} - A) = \sigma^2 (C^{-1} - A) - \sigma^2 (C^{-1} - A) = 0$$

□ میل می کند و اثبات کامل می شود.

قضیه ۲.۲.۳. تحت فرض های قضیه ۱.۲.۳، عبارتهای ADB برآوردگرها به صورت زیر به دست می آید:

$$\begin{aligned} \text{ADB}(\hat{\beta}^{\text{uLso}}) &= 0, \\ \text{ADB}(\hat{\beta}^{\text{rLso}}) &= -\delta, \\ \text{ADB}(\hat{\beta}^{\text{ptLso}}) &= -\delta H_{q+2} \left(\mathcal{L}_\alpha; \Delta^2 \right), \\ \text{ADB}(\hat{\beta}^{\text{sLso}}) &= -c\delta E \left[\chi_{q+2}^{-2}(\Delta^2) \right], \\ \text{ADB}(\hat{\beta}^{\text{sLso}+}) &= \text{ADB}(\hat{\beta}^{\text{sLso}}) - \delta E \left[(1 - c\chi_{q+2}^{-2}(\Delta^2)) I(\chi_{q+2}^2(\Delta^2) \leq c) \right], \\ \text{ADB}(\hat{\beta}^{\text{iptLso}}) &= \text{ADB}(\hat{\beta}^{\text{ptLso}}) - c\delta E \left[\chi_{q+2}^{-2}(\Delta^2) \right] \\ &\quad - c\delta E \left[\chi_{q+2}^{-2}(\Delta^2) I(\chi_{q+2}^2(\Delta^2) \leq \mathcal{L}_\alpha) \right] \end{aligned}$$

که در آن $c = \lim_{n \rightarrow \infty} c_n = q - 2$

برهان. مشابه اثبات ویژگی های برآوردگرهای بهبودیافته M-ریج در فصل ۲، دوباره شکل کلی از برآوردگرها را که تمامی برآوردهای پیشنهادی حالت خاصی از آن باشند را به صورت زیر در نظر بگیرید:

$$\begin{aligned} \hat{\beta}^{*\text{Lso}} &= \hat{\beta}^{\text{uLso}} - g(\mathcal{L}_n)(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \\ &= \hat{\beta}^{\text{uLso}} - C^{-1} H^\top (HC^{-1} H^\top)^{-1} (H^{\text{uLso}} - h). \end{aligned}$$

(الف) اگر $g(\mathcal{L}_n) = 0$ ، آن گاه $\hat{\beta}^{*\text{Lso}} = \hat{\beta}^{\text{uLso}}$ ، ULE است.

(ب) اگر $g(\mathcal{L}_n) = 1$ ، آن گاه $\hat{\beta}^{*\text{Lso}} = 0 = \hat{\beta}^{\text{rLso}}$ ، RLE است.

(ج) اگر $g(\mathcal{L}_n) = I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha})$ ، آن گاه $\hat{\beta}^{*\text{Lso}} = \hat{\beta}^{\text{ptLso}}$ ، PTLE است.

(د) اگر $g(\mathcal{L}_n) = c_n \mathcal{L}_n^{-1}$ ، آن گاه $\hat{\beta}^{*\text{Lso}} = \hat{\beta}^{\text{sLso}}$ ، SLE است.

(ه) اگر $g(\mathcal{L}_n) = c \mathcal{L}_n^{-1} + (1 - c_n \mathcal{L}_n^{-1}) I(\mathcal{L}_n \leq c_n)$ ، آن گاه $\hat{\beta}^{*\text{Lso}} = \hat{\beta}^{\text{sLso}+}$ ، PRSLE است.

(ه) اگر $g(\mathcal{L}_n) = c \mathcal{L}_n^{-1} + (1 - c_n \mathcal{L}_n^{-1}) I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha})$ ، آن گاه $\hat{\beta}^{*\text{Lso}} = \hat{\beta}^{\text{iptLso}}$ ، IPTLE است.

و \mathcal{L}_n به صورت زیر تعریف می شود:

$$\mathcal{L}_n = \frac{(\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h})^\top (\mathbf{H}\mathbf{C}_n^{-1}\mathbf{H}^\top)^{-1} (\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h})}{s_L^2},$$

که در آن

$$s_L^2 = \frac{1}{m} (\mathbf{y} - \mathbf{X}\hat{\beta}^{\text{uLso}})^\top (\mathbf{y} - \mathbf{X}\hat{\beta}^{\text{uLso}}),$$

و طبق قسمت (۴) قضیه ۱.۲.۳، دارای توزیع کی دو با q درجه آزادی و پارامتر غیرمرکزی Δ^2 است.

اکنون، برای محاسبه ADBهای برآوردگرها می توان نوشت:

$$\begin{aligned} \text{ADB}(\hat{\beta}^{*\text{Lso}}) &= \lim_{n \rightarrow \infty} \text{E} \left[\sqrt{n} (\hat{\beta}^{*\text{Lso}} - \beta) \right] \\ &= \lim_{n \rightarrow \infty} \text{E} \left[\sqrt{n} (\hat{\beta}^{\text{uLso}} - g(\mathcal{L}_n)(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) - \beta) \right] \\ &= \lim_{n \rightarrow \infty} \text{E} \left[\sqrt{n} (\hat{\beta}^{\text{uLso}} - \beta) \right] - \lim_{n \rightarrow \infty} \text{E} \left[\sqrt{n} g(\mathcal{L}_n) (\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \right] \\ &= \text{ADB}(\hat{\beta}^{\text{uLso}}) - \lim_{n \rightarrow \infty} \text{E} \left[\sqrt{n} g(\mathcal{L}_n) (\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \right] \\ &= \text{ADB}(\hat{\beta}^{\text{uLso}}) \\ &\quad - \lim_{n \rightarrow \infty} \mathbf{C}_n^{-1} \mathbf{H} (\mathbf{H}\mathbf{C}_n^{-1} \mathbf{H}^\top)^{-\frac{1}{2}} \\ &\quad \times \text{E} \left[\sqrt{n} (\mathbf{H}\mathbf{C}_n^{-1} \mathbf{H}^\top)^{-\frac{1}{2}} (\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h}) g(\mathcal{L}_n) \right]. \end{aligned} \quad (12.3)$$

طبق قسمت (۱) قضیه ۱.۲.۳، $\text{E} \left[\sqrt{n} (\hat{\beta}^{\text{uLso}} - \beta) \right] = 0$ لذا $\text{ADB}(\hat{\beta}^{\text{uLso}}) = 0$. از طرفی با در نظر گرفتن $\mathbf{Z} = \sqrt{n} (\mathbf{H}\mathbf{C}_n^{-1} \mathbf{H}^\top)^{-\frac{1}{2}} (\mathbf{H}\hat{\beta}^{\text{uLso}} - \mathbf{h})$ بر اساس رابطه (۱۱.۳) و قضیه ۵.۲.آ، رابطه (۱۲.۳) برابر می شود با

$$\text{ADB}(\hat{\beta}^{*\text{Lso}}) = 0 - \delta \text{E} \left[g(\chi_{p+2}^2(\Delta^2)) \right]. \quad (13.3)$$

با جایگذاری هر یک از توابع g متناظر با هر برآوردگر در رابطه های (۱۳.۳) اثبات کامل می شود. \square

نتیجه ۱.۲.۳. تحت فرض های قضیه ۱.۲.۳، عبارت های ADQB برآوردگرها به صورت زیر به دست می آید:

$$\begin{aligned} \text{ADQB}(\hat{\beta}^{\text{uLso}}) &= 0, \\ \text{ADQB}(\hat{\beta}^{\text{rLso}}) &= \Delta^2, \\ \text{ADQB}(\hat{\beta}^{\text{ptLso}}) &= \Delta^2 \{ H_{q+2}(\mathcal{L}_\alpha; \Delta^2) \}^2, \\ \text{ADQB}(\hat{\beta}^{\text{sLso}}) &= c^2 \Delta^2 \left\{ \text{E} \left[\chi_{q+2}^{-2}(\Delta^2) \right] \right\}^2, \\ \text{ADQB}(\hat{\beta}^{\text{sLso}+}) &= \Delta^2 \left\{ c \text{E} \left[\chi_{q+2}^{-2}(\Delta^2) \right] - \text{E} \left[(1 - c \chi_{q+2}^{-2}(\Delta^2)) I(\chi_{q+2}^2(\Delta^2) \leq c) \right] \right\}^2, \\ \text{ADQB}(\hat{\beta}^{\text{iptLso}}) &= \Delta^2 \left\{ H_{q+2}(\mathcal{L}_\alpha; \Delta^2) - c \text{E} \left[\chi_{q+2}^{-2}(\Delta^2) \right] \right. \\ &\quad \left. - c \text{E} \left[\chi_{q+2}^{-2}(\Delta^2) I(\chi_{q+2}^2(\Delta^2) \leq \mathcal{L}_\alpha) \right] \right\}^2 \end{aligned}$$

قضیه ۳.۲.۳. تحت فرض‌های قضیه ۱.۲.۳، ADR برآوردگرها به صورت زیر هستند:

$$\begin{aligned}
 \text{ADR}(\hat{\beta}^{\text{uLso}}) &= \sigma^2 \text{tr}(\mathbf{C}^{-1}), \\
 \text{ADR}(\hat{\beta}^{\text{rLso}}) &= \sigma^2 \text{tr}(\mathbf{C}^{-1} - \mathbf{A}) + \delta^\top \delta, \\
 \text{ADR}(\hat{\beta}^{\text{ptLso}}) &= \sigma^2 \text{tr}(\mathbf{C}^{-1}) - \sigma^2 \text{tr}(\mathbf{C}^{-1} - \mathbf{A}) H_{q+\gamma}(\mathcal{L}_\alpha; \Delta^2) + \delta^\top \delta Z(\alpha; \Delta^2), \\
 \text{ADR}(\hat{\beta}^{\text{sLso}}) &= \sigma^2 \text{tr}(\mathbf{C}^{-1}) - c\sigma^2 \text{tr}(\mathbf{C}^{-1} - \mathbf{A}) X(\Delta^2) + c(c + \epsilon) \delta^\top \delta, \\
 \text{ADR}(\hat{\beta}^{\text{sLso}+}) &= \text{ADR}(\hat{\beta}^{\text{sLso}}) - \delta^\top \delta Q(\Delta^2) - \sigma^2 \text{tr}(\mathbf{C}^{-1} - \mathbf{A}) \\
 &\quad \times \text{E} \left[(\lambda - c\chi_{q+\gamma}^{-\gamma}(\Delta^2))^\gamma I(\chi_{q+\gamma}^\gamma(\Delta^2) \leq c) \right] \\
 \text{ADR}(\hat{\beta}^{\text{iptLso}}) &= \text{ADR}(\hat{\beta}^{\text{ptLso}}) \\
 &\quad - \sigma^2 \text{tr}(\mathbf{C}^{-1} - \mathbf{A}) \left\{ cX(\Delta^2) + c^\gamma \text{E} \left[\chi_{q+\gamma}^{-\epsilon}(\Delta^2) I(\chi_{q+\gamma}^\gamma(\Delta^2) \leq \mathcal{L}_\alpha) \right] \right\} \\
 &\quad + c\delta^\top \delta \left\{ c\text{E} \left[\chi_{q+\epsilon}^{-\epsilon}(\Delta^2) \right] + \gamma \text{E} \left[\chi_{q+\gamma}^{-\gamma}(\Delta^2) \right] - \gamma \text{E} \left[\chi_{q+\epsilon}^{-\gamma}(\Delta^2) \right] \right. \\
 &\quad \left. + c\text{E} \left[\chi_{q+\epsilon}^{-\epsilon}(\Delta^2) I(\chi_{q+\epsilon}^\gamma(\Delta^2) \leq \mathcal{L}_\alpha) \right] - \gamma \text{E} \left[\chi_{q+\gamma}^{-\gamma}(\Delta^2) I(\chi_{q+\gamma}^\gamma(\Delta^2) \leq \mathcal{L}_\alpha) \right] \right\}
 \end{aligned}$$

که در آن

$$\mathbf{A} = \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{H}^\top (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}^\top)^{-1} \mathbf{H} \mathbf{C}^{-1}$$

9

$$\begin{aligned}
 Z(\alpha; \Delta^2) &= \gamma H_{q+\gamma}(\mathcal{L}_\alpha; \Delta^2) - H_{q+\epsilon}(\mathcal{L}_\alpha; \Delta^2), \\
 X(\Delta^2) &= \gamma \text{E} \left[\chi_{q+\gamma}^{-\gamma}(\Delta^2) \right] - c \text{E} \left[\chi_{q+\epsilon}^{-\epsilon}(\Delta^2) \right], \\
 Q(\Delta^2) &= \gamma \text{E} \left[(\lambda - c\chi_{q+\gamma}^{-\gamma}(\Delta^2))^\gamma I(\chi_{q+\gamma}^\gamma(\Delta^2) \leq c) \right] \\
 &\quad - \text{E} \left[(\lambda - c\chi_{q+\epsilon}^{-\epsilon}(\Delta^2))^\gamma I(\chi_{q+\epsilon}^\gamma(\Delta^2) \leq c) \right].
 \end{aligned}$$

برهان. طبق تعریف ADR، رابطه (۱۸.۲)، می‌توان نوشت

$$\begin{aligned}
 \text{ADR}(\hat{\beta}^{*\text{Lso}}) &= \lim_{n \rightarrow \infty} \text{E} \left[n \left(\hat{\beta}^{*\text{Lso}} - \beta \right)^\top \left(\hat{\beta}^{*\text{Lso}} - \beta \right) \right] \\
 &= \lim_{n \rightarrow \infty} \text{E} \left[n \left(\hat{\beta}^{\text{uLso}} - g(\mathcal{L}_n) \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right) - \beta \right)^\top \right. \\
 &\quad \left. \left(\hat{\beta}^{\text{uLso}} - g(\mathcal{L}_n) \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right) - \beta \right) \right] \\
 &= \lim_{n \rightarrow \infty} \text{E} \left[\left(n \hat{\beta}^{\text{uLso}} - \beta \right)^\top \left(\hat{\beta}^{\text{uLso}} - \beta \right) \right] \\
 &\quad - \gamma \lim_{n \rightarrow \infty} \text{E} \left[n \left(\hat{\beta}^{\text{uLso}} - \beta \right)^\top \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right) g(\mathcal{L}_n) \right] \\
 &\quad + \lim_{n \rightarrow \infty} \text{E} \left[n \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right)^\top \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right) g(\mathcal{L}_n) \right] \\
 &= \text{ADR}(\hat{\beta}_n^{\text{uLso}}) \\
 &\quad - \gamma \lim_{n \rightarrow \infty} \text{E} \left[n \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} + \hat{\beta}^{\text{rLso}} - \beta \right)^\top \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right) g(\mathcal{L}_n) \right] \\
 &\quad + \lim_{n \rightarrow \infty} \text{E} \left[n \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right)^\top \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right) g(\mathcal{L}_n) \right]. \quad (۱۴.۳)
 \end{aligned}$$

اکنون هر یک از عبارتهای رابطه (۱۴.۳) جداگانه حساب می‌شود و در نهایت جایگزین می‌گردند. بدین منظور،

$$\begin{aligned} \lim_{n \rightarrow \infty} E \left[n \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} + \hat{\beta}^{\text{rLso}} - \beta \right)^\top \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right) g(\mathcal{L}_n) \right] = \\ \lim_{n \rightarrow \infty} E \left[n \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right)^\top \left(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} \right) g(\mathcal{L}_n) \right] \\ + \lim_{n \rightarrow \infty} E \left[\left(\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta) \right)^\top \left(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \right) g(\mathcal{L}_n) \right]. \quad (15.3) \end{aligned}$$

از آن جایی که

$$\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}} = C_n^{-1} H^\top (HC^{-1}H^\top)^{-1} (H\hat{\beta}^{\text{uLso}} - h),$$

کافیست $Z^\top = (HC^{-1}H^\top)^{-1/2} (H\hat{\beta}^{\text{uLso}} - h)$ در نظر گرفته شود، در این صورت طبق قضیه ۶.۲.آ عبارت اول رابطه (۱۵.۳) برابر است با

$$\begin{aligned} \text{tr} \left(C^{-1} H^\top (HC^{-1}H^\top)^{-1/2} (HC^{-1}H^\top)^{-1/2} HC^{-1} \right) E \left[g(\chi_{q+2}^2(\Delta^2)) \right] \\ + \delta^\top \delta E \left[g(\chi_{q+4}^2(\Delta^2)) \right]. \end{aligned}$$

بنا به ویژگی‌های توزیع نرمال، بنا به قسمت (۵) لم ۱.۲.۳، و صفر بودن کوواریانس بین متغیرهای $\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})$ و $\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta)$ ، نتیجه می‌شود که این متغیرها مستقل از یکدیگرند و می‌توان عبارت دوم رابطه (۱۵.۳) را به صورت

$$\begin{aligned} \lim_{n \rightarrow \infty} E \left[\left(\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta) \right) \left(\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \right) g(\mathcal{L}_n) \right] = \\ \lim_{n \rightarrow \infty} E \left[\sqrt{n}(\hat{\beta}^{\text{rLso}} - \beta)^\top \right] \lim_{n \rightarrow \infty} E \left[\sqrt{n}(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) g(\mathcal{L}_n) \right] = \\ - \delta^\top \lim_{n \rightarrow \infty} C^{-1} H^\top (HC^{-1}H^\top)^{-1/2} E \left[(HC^{-1}H^\top)^{-1/2} (H\hat{\beta}^{\text{uLso}} - h) g(\mathcal{L}_n) \right] = \\ - \delta^\top \delta E \left[g(\chi_{q+2}^2(\Delta^2)) \right] \end{aligned}$$

نوشت. خط آخر از انتخاب $Z = (HC^{-1}H^\top)^{-1/2} (H\hat{\beta}^{\text{uLso}} - h)$ در قضیه ۵.۲.آ نتیجه می‌شود. بنابراین رابطه (۱۵.۳) برابر است با

$$\text{tr}(C^{-1} - A) E \left[g(\chi_{q+2}^2(\Delta^2)) \right] + \delta^\top \delta \left\{ E \left[g(\chi_{q+4}^2(\Delta^2)) \right] - E \left[g(\chi_{q+2}^2(\Delta^2)) \right] \right\}. \quad (16.3)$$

از طرف دیگر، عبارت $\lim_{n \rightarrow \infty} E \left[n(\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}})^\top (\hat{\beta}^{\text{uLso}} - \hat{\beta}^{\text{rLso}}) \right]$ را می‌توان به صورت

$$\begin{aligned} \lim_{n \rightarrow \infty} (HC_n^{-1}H^\top)^{-1/2} H^\top C_n^{-1} \\ \times E \left[n((HC_n^{-1}H^\top)^{-1/2} (H\hat{\beta}^{\text{uLso}} - h))^\top ((HC_n^{-1}H^\top)^{-1/2} (H\hat{\beta}^{\text{uLso}} - h)) g(\mathcal{L}_n) \right] \\ \times C_n^{-1} H (HC_n^{-1}H^\top)^{-1/2} \quad (17.3) \end{aligned}$$

نوشت. دوباره کافیست $Z^\top = (HC^{-1}H^\top)^{-1/2} (H\hat{\beta}^{\text{uLso}} - h)$ در نظر گرفته شود، بنابراین بر اساس قضیه ۵.۲.آ می‌توان (۱۷.۳) را به صورت

$$\text{tr}(C^{-1} - A) E \left[g(\chi_{q+2}^2) \right] + \delta^\top \delta E \left[g(\chi_{q+4}^2) \right] \quad (18.3)$$

نوشت. از جایگذاری (۱۶.۳) و (۱۸.۳) در رابطه (۱۴.۳) می‌توان ADR برآوردگرها را به صورت

$$\begin{aligned} \text{ADR}(\hat{\beta}^{*Lso}) &= \text{ADR}(\hat{\beta}^{uLso}) - 2 \text{tr}(C^{-1} - A) E[g(\chi_{q+2}^2)] \\ &\quad - 2\delta^T \delta \left\{ E[g(\chi_{q+4}^2)] - E[g(\chi_{q+2}^2)] \right\} \\ &\quad + \text{tr}(C^{-1} - A) E[g(\chi_{q+2}^4)] + \delta^T \delta E[g(\chi_{q+4}^4)] \\ &= \text{ADR}(\hat{\beta}^{uLso}) - \text{tr}(C^{-1} - A) \left\{ 2E[g(\chi_{q+2}^2)] - E[g(\chi_{q+2}^4)] \right\} \\ &\quad + \delta^T \delta \left\{ E[g(\chi_{q+4}^4)] - 2E[g(\chi_{q+4}^2)] + 2E[g(\chi_{q+2}^2)] \right\}. \quad (19.3) \end{aligned}$$

به دست آورد. با جانشین کردن هر یک از توابع g متناظر با هر برآوردگر در رابطه‌های (۱۹.۳) اثبات کامل می‌شود. □

در این فصل می‌توان همانند قضیه‌های ۹.۳.۲ – ۱۴.۳.۲، قضیه‌های مرتبط را بیان و اثبات کرد و برای عدم تکرار، می‌توان به مقاله **نوروزی‌راد و آرشی (۱۳۹۶)** و **نوروزی‌راد و آرشی (۲۰۱۷ا)** اشاره کرد.

۳.۳ مطالعه شبیه‌سازی

در این بخش، با استفاده از یک شبیه‌سازی مونت‌کارلو، کارایی نسبی برآوردگرهای پیشنهاد شده مورد بررسی قرار می‌گیرد.

ماتریس X از توزیع نرمال p متغیره با بردار میانگین $\mu = o$ و ماتریس واریانس-کوواریانس Σ ایجاد می‌شود، که در آن عناصر غیرقطری ماتریس Σ برابر با ρ در نظر گرفته شده‌اند که مقادیر ۰، ۰/۲ و ۰/۹ را پذیرفته است و عناصر روی قطر اصلی همگی ۱ هستند. فرض می‌شود تعداد مشاهده‌ها، $n = 100$ و تعداد پارامترها، p مقادیر ۱۰، ۱۵ و ۲۰ را بپذیرد.

انتخاب‌های متفاوتی می‌توان برای H و h در نظر گرفت. یکی از مهم‌ترین آن‌ها انتخابی است که منجر به انتخاب متغیر می‌شود. گاهی یک متخصص ادعا می‌کند که بعضی از متغیرها، مدل رگرسیونی را تحت تاثیر قرار نمی‌دهد.

اگر $\beta = (\beta_1^T, \beta_p^T)^T$ ، آن‌گاه $\beta_p = o$ معادل این است که بعضی از متغیرها در پیشگویی مدل نادیده گرفته می‌شود. فرض کنید $\beta = (\beta_1^T, \beta_p^T)^T = (1_{p-q}^T, o_q^T)^T$ که در آن 1_{p-q} و o_q بردارهای ۱ و صفر به ترتیب با ابعاد $p - q$ و q هستند.

به منظور بررسی رفتار برآوردگرهای پیشنهادی، میزان انحراف از فرضیه صفر به صورت $\|\beta - \beta_o\|^2 = \Delta^2$ تعریف می‌شود، که در آن $\beta_o = (1_{p-q}^T, o_q^T)^T$. اگر $\Delta^2 = o$ ، آن‌گاه $\beta = \beta_o$. به‌ازای $\Delta^2 > o$ ، بردار $\beta = (1_{p-q}^T, \Delta^T)^T$ در نظر گرفته می‌شود که $\Delta = (\Delta^1, \dots, \Delta^q)^T$ بردار q تایی از مقادیر Δ^2 است. هنگامی که مقدار Δ^2 افزایش می‌یابد، درجه انحراف از فرضیه صفر بیشتر می‌شود. پنج مقدار ۰، ۰/۱، ۰/۵، ۱ و ۵ برای Δ^2 در نظر گرفته شده است.

بدون از دست دادن کلیت مسئله، فرض کنید β یک بردار p تایی است که $p - q$ مولفه اول آن عدد یک و بقیه q مولفه، صفر است. متغیر پاسخ از مدل (۱.۱) با خطای نرمال استاندارد تولید شده است.

معیار MSE نسبی برآوردگر پیشنهادی دلخواه $\hat{\beta}^{*Lso}$ به صورت

$$RMSE(\hat{\beta}^*) = \frac{MSE(\hat{\beta}^{uLso})}{MSE(\hat{\beta}^{*Lso})}$$

محاسبه می‌شود. چنانچه مقدار کارایی بزرگتر از یک باشد، مخاطره برآوردگر پیشنهادی کمتر از مخاطره برآوردگر لاسو است.

در جداول ۱.۳-۳.۳ مقادیر عددی کارایی نسبی برآوردگرهای ارائه شده نسبت به برآوردگر لاسو برای ۱۰۰۰ بار تکرار گزارش شده است.

جدول ۱.۳: کارایی‌های نسبی برآوردگرها برای Δ^2 های ثابت، $\rho = 0$ ، $q = p - 6$ و مقادیر مختلف p

Δ^2	p	ULE	RLE	PTLE	SLE	PRSLE	IPTLE
۰	۱۰	۱	۱/۶۳	۱/۵۲	۱/۲۴	۱/۳۵	۱/۵۵
	۱۵	۱	۲/۳۲	۲/۱۹	۱/۷۵	۲/۰۵	۲/۲۷
	۲۰	۱	۳/۵۱	۲/۹۹	۲/۴۴	۲/۹۹	۲/۳۲
۰/۱	۱۰	۱	۱/۵۷	۱/۴۶	۱/۲۱	۱/۳۲	۱/۴۹
	۱۵	۱	۲/۱۹	۲/۰۲	۱/۶۶	۱/۹۵	۲/۱۱
	۲۰	۱	۳/۴۸	۲/۹۲	۲/۴۰	۲/۹۹	۳/۲۸
۰/۵	۱۰	۱	۰/۸۵	۰/۸۶	۱/۰۵	۱/۰۷	۰/۹۰
	۱۵	۱	۱/۵۶	۱/۴۲	۱/۳۷	۱/۵۴	۱/۵۲
	۲۰	۱	۲/۸۶	۲/۳۵	۲/۱۹	۲/۵۸	۲/۷۱
۱	۱۰	۱	۰/۳۷	۰/۹۰	۱/۰۰	۱/۰۰	۰/۹۲
	۱۵	۱	۰/۸۳	۰/۸۷	۱/۱۲	۱/۱۴	۰/۹۶
	۲۰	۱	۱/۸۱	۱/۴۳	۱/۸۱	۱/۹۰	۱/۷۷
۵	۱۰	۱	۰/۰۲	۱/۰۰	۰/۹۸	۰/۹۸	۰/۹۸
	۱۵	۱	۰/۰۶	۱/۰۰	۰/۹۸	۰/۹۸	۰/۹۸
	۲۰	۱	۰/۱۳	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰

جدول ۲.۳: کارایی‌های نسبی برآوردگرها به ازای Δ^2 های ثابت، $\rho = 0.2$ ، $q = p - 6$ و مقادیر مختلف p

Δ^2	p	ULE	RLE	PTLE	SLE	PRSLE	IPTLE
0	10	1	1/95	1/71	1/34	1/44	1/78
	15	1	2/70	2/37	1/90	2/28	2/56
	20	1	4/72	3/55	2/98	3/63	4/27
0.1	10	1	1/93	1/61	1/33	1/40	1/69
	15	1	2/69	2/30	1/92	2/21	2/51
	20	1	4/72	3/39	2/97	3/62	4/22
0.5	10	1	0.97	0.92	1/17	1/18	1/03
	15	1	1/87	1/30	1/70	1/74	1/68
	20	1	3/74	1/93	2/75	2/91	3/00
1	10	1	0.37	0.99	1/08	1/08	1/07
	15	1	0.84	0.97	1/37	1/37	1/30
	20	1	1/85	1/07	2/10	2/10	1/92
5	10	1	0.02	1/00	0.99	0.99	0.99
	15	1	0.04	1/00	1/02	1/02	1/02
	20	1	0.85	1/00	1/13	1/13	1/13

جدول ۳.۳: کارایی‌های نسبی برآوردگرها به ازای Δ^2 های ثابت، $\rho = 0.9$ ، $q = p - 6$ و مقادیر مختلف p

Δ^2	p	ULE	RLE	PTLE	SLE	PRSLE	IPTLE
0	10	1	6/76	3/34	1/86	1/86	4/00
	15	1	8/48	6/44	4/35	4/99	7/55
	20	1	14/74	14/48	6/14	11/32	14/68
0.1	10	1	6/35	2/95	1/82	1/82	3/60
	15	1	8/32	5/77	4/28	4/68	7/10
	20	1	14/12	12/56	5/96	10/69	13/68
0.5	10	1	3/28	1/39	1/69	1/69	1/88
	15	1	5/41	2/71	3/86	3/93	4/22
	20	1	10/40	7/10	6/77	8/68	9/52
1	10	1	1/19	0.97	1/55	1/55	1/37
	15	1	2/50	1/26	3/14	3/14	2/46
	20	1	5/29	2/57	6/34	6/58	4/80
5	10	1	0.02	1/00	0.87	0.87	0.87
	15	1	0.07	1/00	1/58	1/58	1/58
	20	1	0.19	1/00	2/69	2/69	2/69

- با نگاهی دقیق‌تر به مقادیر جداول می‌توان نتیجه‌های زیر را به‌دست آورد.
- (۱) هنگامی که فرضیه صفر برقرار است ($\Delta^2 = 0$)، RLE بهتر از سایر برآوردگرها رفتار می‌کند. هر چه از فرضیه صفر دور شویم، از کارایی این برآوردگر کاسته می‌شود.
 - (۲) با افزایش مقدار Δ^2 عدد کارایی نسبی کاهش می‌یابد. به عبارت دقیق‌تر هرچه از فرضیه صفر دور شده ایم، میزان برتری برآوردگرهای پیشنهادی به ULE کاهش یافته است. حتی در مواردی برآوردگر محدود نشده لاسو بهتر رفتار می‌کند.
 - (۳) در ارتباط با نتیجه (۲)، میزان برتری برآوردگرها متأثر از درجه همبستگی نیز می‌باشد. بدین معنی که هرچه درجه همبستگی بیشتر شود تأثیر دور شدن از فرضیه صفر بر برتری برآوردگرهای پیشنهادی کاهش می‌یابد. به عنوان مثال، در حالت $\rho = 0$ ، برآوردگر لاسو به ازای $\Delta^2 = 5$ بر برآوردگرهای پیشنهادی برتری دارد در صورتی که به ازای $\rho = 0.9$ این برتری تغییر می‌کند.
 - (۴) با افزایش اندازه ρ ، عدد کارایی نسبی افزایش می‌یابد. به عبارت دیگر، با افزایش درجه همبستگی، مخاطره برآوردگر پیشنهادی نسبت به لاسو بیشتر می‌شود که نشان از مطلوبیت استراتژی به کار رفته می‌باشد.
 - (۵) با زیاد شدن بعد پارامتر، در تمامی موارد، میزان برتری برآوردگرهای پیشنهادی بر برآوردگر لاسو افزایش می‌یابد.
 - (۶) سطح معنی‌داری α تأثیر مستقیمی بر میزان برتری برآوردگرهای PTL و IPTL دارد. با افزایش α ، عدد کارایی نسبی کاهش می‌یابد.
 - (۷) هیچ یک از برآوردگرهای آزمون اولیه لاسو و انقباضی نوع استاین لاسو بر دیگری برتری ندارند. برآوردگر جزء مثبت همواره بر برآوردگر نوع استاین لاسو برتری دارد.
 - (۸) برآوردگر بهبودیافته آزمون اولیه لاسو نزدیک فرضیه صفر بهتر از سایر برآوردگرها رفتار می‌کند. هر چه مقدار Δ^2 بیشتر شود، برآوردگر جزء مثبت نوع استاین لاسو بهتر از IPTLE رفتار می‌کند و در نهایت به یک مقدار مشترک میل می‌کنند.

۴.۳ تحلیل داده‌های آمار و ارقام ایالات آمریکا

از داده‌های آمار و ارقام ایالات آمریکا (بکر و همکاران، ۱۹۸۸) که شامل $n = 50$ ایالت آمریکا و $p = 9$ متغیر است، برای نشان دادن کارایی برآوردگرهای پیشنهادی بهبودیافته بر پایه لاسو استفاده شده است.

مدل رگرسیون خطی به داده‌ها برازش داده شده است تا متغیر پاسخ، LIFE EXP، امید به زندگی، را در حضور متغیرهای توضیحی، POPULATION، جمعیت (هزار نفر)؛ INCOME، درآمد سرانه (دلار)؛ ILLITERACY، نرخ بی‌سوادی (درصد ناتوانی برای خواندن و نوشتن)؛ MURDER، نرخ خودکشی به ازای هر ۱۰۰۰۰۰ نفر؛ HS GRAD، درصد باسوادانی که تحصیلات دانشگاهی ندارند؛ FROST، میانگین تعداد روزها در هر سال با دمای زیر صفر؛ AREA، مساحت زمین (مایل مربع) و DENSITY، شاخصی برحسب جمعیت، بتوان مدل‌بندی کرد. آماره‌های توصیفی متغیرها در جدول ۴.۳ نشان داده شده است.

جدول ۴.۳: آماره‌های توصیفی از متغیرهای داده‌های آمار و ارقام ایالات آمریکا

نماد	متغیرها	کمینه	چارک اول	میانه	میانگین	چارک سوم	بیشینه
	LIFE.EXP	۶۷/۹۶	۰/۶۲	۷۰/۶۷	۷۰/۸۸	۷۱/۸۹	۷۳/۶۰
β_1	POPULATION	۳۶۵	۱۰۸۰	۲۸۳۸	۴۲۴۶	۴۹۶۸	۲۱۱۹۸
β_2	INCOME	۳۰۹۸	۳۹۹۳	۴۵۱۹	۴۴۳۶	۴۸۱۴	۶۳۱۵
β_3	ILLITERACY	۰/۵۰	۰/۶۲۵	۰/۹۵۰	۱/۱۷	۱/۵۷	۲/۸۰
β_4	MURDER	۶۷/۹۶	۴/۳۵	۶/۸۵۰	۷/۳۸	۱۰/۶۷	۱۵/۱۰
β_5	HS.GRAD	۳۷/۸۰	۴۸/۰۵	۵۳/۲۵	۵۳/۱۱	۵۹/۱۵	۶۷/۳۰
β_6	FROST	۰/۰۰	۶۶/۲۵	۱۱۴/۵۰	۱۰۴/۴۶	۱۳۹/۷۵	۱۸۸/۰۰
β_7	AREA	۱۰۴۹	۳۶۹۸۵	۵۴۲۷۷	۷۰۷۳۶	۸۱۱۶۳	۵۶۶۴۳۲
β_8	DENSITY	۰/۶۴	۲۵/۳۳	۷۳/۰۱	۱۴۹/۲۲	۱۴۴/۲۸	۹۷۵/۰۱

از آن جایی که یکی از چالش‌های استراتژی‌های برآوردگرهای پیشنهاد شده، تعیین ماتریس H و بردار h است و نیاز به نظر متخصص یا مطالعات قبلی دارد، در اینجا به منظور سادگی و دوری از خطای ناشی از انتخاب نادرست این پارامترها، فرض می‌شود ماتریس H یک ماتریس همانی 7×7 باشد.

به منظور نشان دادن تاثیر درستی یا نادرستی فرضیه صفر، دو حالت فرضی برای بردار h در نظر گرفته می‌شود:

حالت اول: فرض کنید $h = (0, 0, 10, 0, 0, 0, 0, 0)^T$. در این حالت، فرضیه صفر، $H_0: \beta = h$ و بدین معنی است که حضور POPULATION، INCOME و AREA در مدل معنی‌دار نیستند.

حالت دوم: فرض کنید $h = (0, 0, 0, 0, 0, 0, 0, 0)^T$. که در این صورت فرضیه صفر عبارت است از $H_0: \beta = 0$ و بدین معنی است که هیچ‌یک از متغیرهای بیان شده برای برازش مدل مناسب نیستند.

کارایی برآوردگرها با استفاده از میانگین خطای 10% تایی اعتبارسنجی متقابل به دست آمده است. خطای پیشگویی، توان دوم اختلاف بین مقدار واقعی و پیشگویی متغیر پاسخ در مجموعه داده‌های آزمون است. این فرآیند، 1000 بار تکرار می‌شود. میانگین خطاهای پیشگویی، توان دوم اختلاف بین مقدار واقعی و پیشگویی متغیر پاسخ محاسبه می‌شوند و RPE به صورت زیر تعریف می‌شود:

$$RPE(\hat{\beta}^{*Lso}) = \frac{Ave(PE(\hat{\beta}^{tLso}))}{Ave(PE(\hat{\beta}^{*Lso}))}$$

که در آن Ave میانگین را نشان می‌دهد و $\hat{\beta}^{*Lso}$ یکی از برآوردگرهای RL، SL، PRSL، و PTL و IPTL به ازای $\alpha = 0.05$ است. جدول ۵.۳ نتایج را نشان می‌دهد.

جدول ۵.۳: جدول خطاهای پیشگویی نسبی برآوردگرها در داده آمار و ارقام ایالت‌های آمریکا

متغیرها	حالت اول	حالت دوم
RLE	۲۲/۲۳۹۹۷۰۲	۱/۰۰۰۹۰۹۶
PTLE	۱/۰۰۰۰۵۹۱	۱/۰۰۰۰۵۲۸۲
SLE	۱/۰۱۸۰۰۹۱	۱/۰۰۰۲۳۹۴
PRSLE	۱/۰۱۸۰۷۰۲	۱/۰۰۰۰۵۵۰۱
IPTLE	۱/۰۱۷۹۹۴۰	۱/۰۰۰۳۹۵۷

از آنجایی که اعداد جدول ۵.۳ همگی بزرگتر از یک هستند، می‌توان نتیجه گرفت که برآوردگرهای پیشنهادی میانگین خطای پیشگویی کمتری نسبت به برآوردگر لاسو دارند. اما با نگاهی دقیق‌تر، می‌توان نتیجه گرفت در بین همه برآوردگرها، در هر دو حالت برآوردگر محدود شده لاسو، بهترین رفتار را دارد زیرا در این حالت، فرضیه صفر تقریباً درست است اما با این وجود، برآوردگر جزء مثبت انقباضی نوع استاتین لاسو در مقایسه با سایر برآوردگرها، بهتر از سایر برآوردگرها بردار پارامترهای مدل را پیشگویی می‌کند. بنا به نتایج جدول ۵.۳ حدس زده می‌شود که ادعای مطرح شده در حالت اول درست‌تر است اما با این وجود برآوردگرهای پیشنهادی در هر دو حالت بهتر از برآوردگر لاسو در پیشگویی متغیر پاسخ رفتار می‌کنند و هنگامی که تردید در مورد ادعای مطرح شده وجود دارد، می‌توان با اطمینان از این برآوردگرها استفاده نمود.

ممکن است این گمان پیش بیاید که بر اساس نتایج جدول ۵.۳، کارایی برآوردگرهای بهبودیافته در مقایسه با برآوردگر لاسو برتری چندانی ندارد. توجه به این نکته ضروری است که برآوردگر لاسو به دلیل انتخاب متغیر و برآورد پارامترها به‌طور همزمان کارایی بسیار بالایی در پیشگویی دارد و لذا نباید انتظار داشت که برآوردگری بهتر از آن بر اساس دوباره منقبض کردن به سمت یک زیرفضا حاصل شود. از طرفی انتخاب درست فرضیه صفر جهت دوباره منقبض کردن لاسو از اهمیت بسزایی برخوردار است و در این مثال بدون استفاده از نظر متخصص و تنها برای نمایش چگونگی رفتار برآوردگرهای ارائه شده زیرفضای خاص را فرض

کردیم. بدیهی است انتخاب درست و مناسب زیرفضا در افزایش کارایی برآوردگرهای پیشنهادی تاثیر زیادی دارد.

۵.۳ نتیجه‌گیری

به کار بردن UPI در فرآیند برآورد لاسو منجر به دستیابی برآوردگرهای بهتر بر حسب مخاطره کمتر می‌شود. از آنجایی که در مورد این اطلاعات تردید وجود دارد، آماره آزمون فرضیه صفر، $H_0: H\beta = h$ از اهمیت بسزایی در معرفی برآوردگرهای جدید برخوردار است. برآوردگرهای جزء مثبت انقباضی نوع استاین لاسو و بهبودیافته آزمون اولیه لاسو بهتر از سایر برآوردگرها رفتار می‌کنند. در واقع این برآوردگرها، دوباره برآوردگر لاسو را به سمت اطلاعات غیرنمونه‌ای منقبض می‌کند. در یک مثال کاربردی، کارایی این برآوردگرها از نظر پیشگویی بررسی گردید که مشاهده شد خطای پیشگویی برآوردگرهای پیشنهادی کمتر از برآوردگر لاسو است.

فصل ۴

برآوردگرهای بهبودیافته در رگرسیون LAD

در این فصل، از برآوردگرهای جریمه‌شده استوار برای دستیابی به اطلاعات لازم برای ساختن برآوردگرهای انقباضی استوار در حالتی که داده‌های پرت (الف) در راستای متغیر پاسخ و (ب) در راستای متغیرهای توضیحی یا همان نقاط اهرمی وجود داشته باشد، استفاده شده است. در هر دو حالت، برآوردگرها معرفی، ویژگی‌های آن‌ها بررسی و در یک مطالعه شبیه‌سازی و مثال واقعی تحلیل شده‌اند. از این فصل، مقاله‌های نوروزی‌راد و همکاران (۲۰۱۸)، نوروزی‌راد و همکاران (۲۰۱۷a) و نوروزی‌راد و آرشی (۲۰۱۶) نتیجه شده است.

۱.۴ مقدمه

ساده‌ترین نوع برآوردگر که گاهی اوقات پایدارتر از برآوردگر OLS است، برآوردگر کمترین قدر مطلق انحراف‌ها (LAD) می‌باشد. برآوردها در این روش با کمینه کردن مجموع قدر مطلق مقادیر باقیمانده‌ها به دست می‌آیند. به عبارت دیگر

$$\hat{\beta}^{\text{LAD}} = \operatorname{argmin}_{\beta} \left\{ \sum_{i=1}^n |\varepsilon_i| \right\} = \operatorname{argmin}_{\beta} \left\{ \sum_{i=1}^n |y_i - \mathbf{x}_i^{\top} \beta| \right\} \quad (1.4)$$

رگرسیون LAD نسبت به مشاهدات با مقادیر غیرعادی متغیر پاسخ y استوار است. نقطه فروریزش آن، $1/\sqrt{2} \approx 0.707$ و بنابراین در حدود ۲۹٪ از داده‌های پرت را می‌تواند تحمل کند. کارایی برآوردگر LAD، ۲۶٪ بیشتر از برآوردگر OLS و \sqrt{n} سازگار است و بدون اینکه پذیره خاصی برای توزیع خطاها وجود داشته باشد، به‌طور مجانبی نرمال است (پولارد، ۱۹۹۱).

اگر خطاهای ε_i متغیرهای مستقل و هم‌توزیع از یک توزیع نمایی دوگانه (لاپلاس) باشد، آن‌گاه $\hat{\beta}^{LAD}$ ، MLE پارامتر β است. در حالت کلی، میانه y در نقطه $x_i^T \beta$ است که رگرسیون میانه، حالت خاصی برای رگرسیون چندکی نیز نامیده می‌شود.

اگرچه، روش OLS، مشهورترین روش در بین روش‌های رگرسیونی است، اما برازش منحنی LAD تقریباً نیم قرن زودتر از برآوردگر OLS ظاهر شد. لژاندر مقاله‌اش را با عنوان «اصل کمترین توان‌های دوم» در سال ۱۸۰۵ منتشر کرد اما بوسکوویچ (۱۷۵۷)، از تکنیک LAD برای برازش یک خط به چند نقطه استفاده کرد و آن نمودار، یکی از مشهورترین نمودارهای غیرعادی در قرن هجدهم شد (آیزنهارت، ۱۹۶۱). در واقع، تاریخچه برآوردگرهای استوار به قرن نوزدهم برمی‌گردد.

رگرسیون LAD (آمسترانگ و فرام، ۱۹۷۶؛ جنتل و همکاران، ۱۹۷۷؛ گراس و استیگر، ۱۹۷۹؛ آن و چن، ۱۹۷۹) برآوردگری است که لاپلاس (۱۸۱۸) آن را «روش موقعیت‌ها» نامید و با اسامی مختلف به کار گرفته شده است (استیگر، ۱۹۷۳).

برآوردگر LAD، گاهی اوقات به رگرسیون L_1 نیز شناخته می‌شود زیرا L_1 -نرم (به عبارت دیگر، مجموع قدر مطلق خطاها) را کمینه می‌کند.

برآوردگر LAD به وجود داده‌های پرت حساس است و البته باید برای به‌دست آوردن این برآوردگرها، هزینه پیچیدگی محاسبات را پرداخت نمود. با توجه به اینکه برای روش‌های عددی مشخصی برای حل کردن مسئله کمینه‌سازی رابطه (۱.۴) وجود ندارد، به یک الگوریتم اولیه نیاز است. روش رایج، الگوریتم سیمپلکس بارودالا-رابرتس است. این الگوریتم را می‌توان با استفاده از نرم‌افزار R و تابع rq بسته $quantreg$ (باروداله و رابرتس، ۱۹۷۳) یا تابع lad در بسته Blossom (تالبرت و همکاران، ۲۰۱۵) یا تابع lad در بسته $L1pack$ (اسوریو، ۲۰۱۵) اجرا کرد.

برآوردگر LAD برای توزیع خطاهای دم پهن بسیار مناسب است. با این وجود، اگر نقاط اهرمی در متغیرهای توضیحی مشاهده شود، این برآوردگر بدتر از برآوردگر OLS بوده و شدیداً به وجود نقاط اهرمی حساس است. در حالتی که نقاط اهرمی در متغیرهای توضیحی وجود دارد، روش برآورد LAD موزون پیشنهاد شده است. برای جزئیات بیشتر در مورد این روش، ایس و مورگنهالر (۱۹۹۲)؛ هوبرت و روسو (۱۹۹۷)؛ گیلونی و همکاران (۲۰۰۶) را ببینید. ایده اصلی در این روش، کاستن از وزن مشاهداتی است که به‌طور افراطی بزرگ هستند (به عبارت دیگر، نقاط اهرمی) می‌باشد و بنابراین، تاثیر داده‌های پرت بر فرآیند برآورد کاهش می‌یابد.

لذا، از آن جایی که برآوردگر LAD نسبت به حضور داده‌های پرت در متغیرهای پاسخ استوار است اما وجود داده‌های پرت در بین متغیرهای پیشگو، به شدت این برآوردگر را تحت تاثیر قرار می‌دهد. به منظور یافتن راه‌حلی برای این مسئله، برآوردگر رگرسیونی WLAD پیشنهاد شده است که رابطه

$$\sum_{i=1}^n w_i |y_i - \mathbf{x}_i^T \boldsymbol{\beta}|,$$

را کمینه می‌کند که در آن w_i ها ($0 < w_i < 1$) وزن‌هایی هستند که با استفاده از یک معیار استوار روی \mathbf{x}_i تعیین و به گونه‌ای انتخاب می‌شود که وزن نقاط اهرمی را کاهش دهد. رگرسیون WLAD می‌تواند به عنوان رگرسیون LAD روی داده‌های تغییر یافته در نظر گرفت. تعیین نقطه فروریزش رگرسیون LAD با وزن‌های مشخص، با تعیین نقطه فروریزش از رگرسیون LAD با داده‌های $(\tilde{x}_i, \tilde{y}_i)$ متناظر است که $\tilde{x}_i = w_i \mathbf{x}_i$ و $\tilde{y}_i = w_i y_i$. از این رو، بر اساس داده‌های تغییر یافته، برآوردگر WLAD به راحتی محاسبه می‌شود.

معمولاً در رهیافت برآوردگرهای انقباضی UPI یا NSI با قرار دادن محدودیت روی پارامترهای مدل، زیرمدل‌ها را ایجاد می‌کنند. معمولاً فرض می‌شود محدودیت‌ها از طریق نظر کارشناس (اطلاعات غیرنمونه‌ای) یا آزمایش‌های گذشته و داده‌های قبلی (اطلاعات پیشین) به دست می‌آیند. ایده‌ای که اخیراً مطرح می‌شود، استفاده از اطلاعات فرعی در رهیافت برآوردگرهای انقباضی است. به عنوان مثال، تکنیک‌های انتخاب متغیر می‌توانند زیرمدل‌های مناسب را در مدل‌های تنک شناسایی کنند.

فرض کنید بردار پارامترهای رگرسیونی به صورت $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_p^T)^T$ تجزیه شده است که در آن $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_{p_1})^T$ بردار p_1 بعدی مربوط به متغیرهای اصلی و بردار $\boldsymbol{\beta}_p = (\beta_{p_1+1}, \dots, \beta_p)^T = (\beta_1, \dots, \beta_{p_1})^T$ بردار p_2 بعدی مدل‌های تنک به آن نگریسته و برابر صفر در نظر گرفته می‌شود. مشخص کردن بردار پارامتر $\boldsymbol{\beta}_p$ به کمک تکنیک‌های انتخاب متغیر، اطلاعات فرعی را به برآوردهای انقباضی تزریق می‌کند. معیارهای اطلاع (مانند AIC، BIC و ...) که برای انتخاب زیرمدل استفاده می‌شوند بر اساس برآوردگرهای OLS به دست می‌آیند در حالی که خطاهای مدل دارای توزیع‌های دم‌پهن هستند، کارایی این معیارها به شدت کاهش می‌یابد. **هارویش و تسای (۱۹۹۰)** چندین روش کاربردی برای انتخاب مدل بر اساس برآوردگرهای LAD پیشنهاد کردند اما آن روش‌ها محدودیت‌هایی نیز دارند. اگر تعداد متغیرها، نسبتاً زیاد باشد، یافتن بهترین زیرمجموعه از بین تمام مدل‌های پیشنهادی بسیار سخت خواهد بود (مثلاً، هزینه محاسباتی، وقت‌گیر بودن و ...)

وانگ و همکاران (۲۰۰۷) یک روش انتخاب متغیر و انقباضی را پیشنهاد کردند که می‌تواند مشابه رگرسیون لاسو، ضرایب را منقبض و انتخاب متغیر انجام دهد و هم‌چنین مانند برآوردگر LAD نسبت به وجود داده‌های پرت در پاسخ‌ها یا خطاها با دم‌های پهن استوار است. در واقع، آن‌ها رهیافت LAD و جریمه لاسو را با هم ترکیب می‌کنند تا برآوردگر LAD-LASSO را بسازند.

برآوردگر LAD-LASSO کمینه‌کننده تابع LAD جریمه‌شده به صورت

$$\sum_{i=1}^n |y_i - \mathbf{x}_i^T \boldsymbol{\beta}| + n\lambda \sum_{j=1}^p |\beta_j|, \quad (2.4)$$

است، که در آن $\lambda > 0$ پارامتر تنظیم‌کننده است. از آنجایی که لاسو برای همه ضرایب رگرسیونی فقط از پارامتر λ استفاده می‌کند، نتیجه به دست آمده ممکن است اریبی قابل توجهی داشته باشند (فن و لی، ۲۰۰۱). با استفاده از این حقیقت، وانگ و همکاران (۲۰۰۷) تابع هدف در رابطه (۲.۴) را تغییر دادند و معیار LAD-LASSO را به صورت

$$\sum_{i=1}^p |y_i - \mathbf{x}_i^T \boldsymbol{\beta}| + n \sum_{j=1}^p \lambda_j |\beta_j| \quad (3.4)$$

تعریف کردند که برای هر ضریب، امکان استفاده از پارامترهای تنظیم‌کننده متفاوتی را می‌دهد. برآوردگر نتیجه‌شده، کاراتر از کمینه رابطه (۲.۴) است.

الگوریتم محاسبه برآوردگر LAD-LASSO (وانگ و همکاران، ۲۰۰۷):

گام ۱: برآوردگر LAD، $\hat{\boldsymbol{\beta}}^{\text{LAD}}$ ، را بر اساس $\{(\mathbf{x}_i^T, y_i)\}$ ، $i = 1, 2, \dots, n$ به دست آورید.

گام ۲: قرار دهید $(\lambda_j = \log(n)/(n|\hat{\beta}_j^{\text{LAD}}|))$ ، $j = 1, \dots, n$.

گام ۳: به ازای $1 \leq j \leq p$ ، بردار p بعدی بسازید که j امین مولفه آن برابر ۱ و سایر مولفه‌ها صفر باشد و آن را e_j بنامید.

گام ۴: مجموعه داده تقویت‌شده $\{(\mathbf{x}_i^{*T}, y_i)\}$ را به ازای $i = 1, \dots, n + p$ در نظر بگیرید به طوری که

$$(الف) \text{ به ازای هر } 1 \leq i \leq n, (\mathbf{x}_i^{*T}, y_i^*) = (\mathbf{x}_i, y_i)$$

$$(ب) \text{ برای } 1 \leq j \leq p, (\mathbf{x}_{n+j}^{*T}, y_{n+j}^*) = (n\lambda_j e_j, 0).$$

گام ۵: برآوردگر LAD-LASSO همان برآوردگر LAD برای مجموعه داده جدید $\{(\mathbf{x}_i^*, y_i)\}$ است.

با استفاده از این الگوریتم هر نرم‌افزار آماری می‌تواند برای یافتن برآوردگر LAD-LASSO مورد استفاده قرار گیرد. اخیراً، امیرگلو (۲۰۱۴) مسئله یافتن برآوردگر LAD-LASSO را به مسئله برنامه‌ریزی ریاضی تبدیل کرد و نشان داد که با روش سیمپلکس می‌تواند حل شود.

اگر نقاط اهرمی در متغیرهای پیشگو وجود داشته باشد، کارایی برآوردگر LAD بهتر از برآوردگر OLS است. بنابراین، کارایی برآوردگر LAD-LASSO نیز بهتر از لاسو نیست و به حضور نقاط اهرمی استوار نمی‌باشد (گیلونی و همکاران، ۲۰۰۶). آرسلان (۲۰۱۲) روش LAD-LASSO را با ترکیب کردن معیار WLAD با تابع جریمه L_1 به روش دیگری به نام WLAD-LASSO برای برآورد پارامترهای رگرسیونی تعمیم داده است. معیار WLAD وزن مشاهداتی که در فضای متغیرهای پیشگو دورافتاده هستند را کاهش می‌دهد و به حضور داده‌های پرت در متغیر پاسخ

حساس است بنابراین برآوردگر نتیجه شده، کمتر تحت تاثیر داده‌های پرت در متغیر پاسخ و نقاط اهرمی در متغیرهای توضیحی قرار می‌گیرد. برآوردگر WLAD-LASSO به صورت

$$\hat{\beta}^{\text{WLAD-LASSO}} = \operatorname{argmin}_{\beta} \left(\sum_{i=1}^n w_i |y_i - \mathbf{x}_i^T \beta| + n \sum_{j=1}^p \lambda_j |\beta_j| \right),$$

تعریف می‌شود که در آن w_i ها وزن‌هایی هستند که بر اساس \mathbf{x}_i ها برای $i = 1, \dots, n$ به دست آمده‌اند. همچنین λ_j ها، به ازای $j = 1, \dots, p$ پارامترهای تنظیم‌کننده در تابع هدف لاسوی تطبیقی است که با استفاده از داده‌ها برآورد می‌شوند. مشاهده‌های پرت در متغیرهای توضیحی فاصله‌های زیادی دارند، از این رو وزن‌های متناظر آن‌ها کوچک خواهند بود. بنابراین، انتظار می‌رود که برآورد به دست آمده در برابر داده‌های پرت در متغیر پاسخ و نقاط اهرمی استوار باشد. برآوردگر WLAD-LASSO به راحتی با استفاده از داده‌های تقویت‌شده $\{(\tilde{\mathbf{x}}_i, \tilde{y}_i) = (w_i \mathbf{x}_i, w_i y_i)\}$ در الگوریتم پیشنهادی **وانگ و همکاران (۲۰۰۷)** برای محاسبه LAD-LASSO به دست می‌آید. در مسئله رگرسیون معمولی، متغیرهایی که معنی‌دار نیستند، ادیده گرفته می‌شوند و از مدلی بر پایه فقط متغیرهای انتخاب شده استفاده می‌شود. در برآورد انقباضی، اطلاعات متغیرهای بی‌معنی نیز با شرکت دادن آن‌ها در فرآیند برآورد استفاده می‌شود و تمامی ضرایب را به سمت متغیرهای مهم منقبض می‌کند. در پایان، یک مدل کامل با همه متغیرها و یک زیرمدل با مجموعه متغیرهای مهم وجود خواهد داشت. برآوردهای محدودنشده و محدودشده به ترتیب از مدل کامل و زیرمدل به دست می‌آیند که ترکیب آن‌ها، برآوردهای انقباضی را می‌سازد.

در ادامه، برآوردگرهای انقباضی در مدل‌های رگرسیونی LAD و WLAD با به کار بردن ایده اطلاعات فرعی مطرح شده و ویژگی‌های آن‌ها از زوایای مختلف مورد بررسی قرار می‌گیرند.

۲.۴ برآوردگرهای انقباضی LAD

مدل رگرسیون خطی زیر را در نظر بگیرید

$$\mathbf{y}_n = \mathbf{X}_n \beta + \varepsilon_n = \mathbf{X}_{n1} \beta_1 + \mathbf{X}_{n2} \beta_2 + \varepsilon_n \quad (۴.۴)$$

که در آن به ترتیب ماتریس‌های \mathbf{X}_{n1} و \mathbf{X}_{n2} دارای ابعاد $n \times p_1$ و $n \times p_2$ بوده و تابع توزیع تجمعی بردار خطاهای تصادفی، $F(\varepsilon)$ میانه صفر دارد. برآوردگر محدودنشده LAD (ULAD) کمینه‌کننده تابع هدف زیر است.

$$\|\mathbf{y}_n - \mathbf{X}_n \beta\|_1. \quad (۵.۴)$$

همچنین، برآوردگر محدودنشده LAD پارامتر β با $(\hat{\beta}_1^{\text{uLAD}}, \hat{\beta}_2^{\text{uLAD}})^T$ نشان داده می‌شود. بنابراین $\hat{\beta}_1^{\text{uLAD}}$ برآوردگر ULAD پارامتر β_1 است.

با در نظر گرفتن محدودیت $\beta_2 = 0$ ، مدل رابطه (۴.۴) به $y_n = X_{n1}\beta_1 + \varepsilon$ کاهش می‌یابد. بنابراین، برآوردگر محدودشده LAD پارامتر β_1 (RLAD)، $\hat{\beta}_1^{\text{rLAD}}$ است که تابع هدف جدید $\|y_n - X_{n1}\beta_1\|_1$ را کمینه می‌کند.

استفاده از مدل‌های محدودشده فوق به درستی فرضیه $H_0: \beta_2 = 0$ بستگی دارد. شرایط نظم زیر را در نظر بگیرید:

(۱) $F(\varepsilon)$ پیوسته است و چگالی مثبت پیوسته $f(\varepsilon)$ در میانه دارد،

(۲) فرض کنید $C_n = X_n^\top X_n$ به‌ازای ماتریس معین مثبت دلخواه $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$

$$\lim_{n \rightarrow \infty} n^{-1} C_n^{-1} = C$$

$$\max_{1 \leq i \leq n} x_i^\top C_n^{-1} x_i = o(1)$$

برای آگاهی بیشتر، پولار (۱۹۹۱) و نایت (۱۹۹۸) را ببینید.

لم ۱.۲.۴. (باست و کرونگر، ۱۹۷۸) تحت شرایط نظم، اگر $n \rightarrow \infty$ ، آن‌گاه $\sqrt{n}(\hat{\beta}^{\text{uLAD}} - \beta)$ دارای توزیع مجانبی نرمال p متغیره با بردار میانگین صفر و ماتریس واریانس-کوواریانس $\tau^2 C^{-1}$ است که در آن τ^2 واریانس مجانبی میانه نمونه‌ای از توزیع دلخواه F است، به‌عبارت دیگر، $\tau = [2f(0)]^{-1}$.

$$\sqrt{n}(\hat{\beta}^{\text{uLAD}} - \beta) \xrightarrow{D} \mathcal{N}_p(o, \tau^2 C^{-1}) \quad (۶.۴)$$

نتیجه ۱.۲.۴. تحت درستی شرایط نظم، داریم

$$\sqrt{n}(\hat{\beta}_1^{\text{uLAD}} - \beta_1) \xrightarrow{D} \mathcal{N}_{p_1}(o, \tau^2 C_{11,2}^{-1}) \quad (۷.۴)$$

$$\sqrt{n}(\hat{\beta}_2^{\text{uLAD}} - \beta_2) \xrightarrow{D} \mathcal{N}_{p_2}(o, \tau^2 C_{22,1}^{-1}) \quad (۸.۴)$$

برهان. فرض کنید که ماتریس X_n به‌ترتیب به ماتریس‌های X_{n1} و X_{n2} با اندازه‌های $n \times p_1$ و $n \times p_2$ افزایش شود. بر این اساس، μ و Σ به‌صورت زیر افزایش می‌شوند:

$$\mu = \begin{bmatrix} \mu_1(p_1 \times 1) \\ \mu_2(p_2 \times 1) \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11}(p_1 \times p_1) & \Sigma_{12}(p_1 \times p_2) \\ \Sigma_{21}(p_2 \times p_1) & \Sigma_{22}(p_2 \times p_2) \end{bmatrix}.$$

اکنون، فرض کنید که ماتریس واریانس-کوواریانس به‌صورت $\tau^2 C^{-1}$ باشد. بنابراین، بر اساس متمم شور، معکوس ماتریس‌های افزایش شده به‌صورت زیر است:

$$\tau^2 \begin{bmatrix} C_{11,2}^{-1} & -C_{11,2}^{-1} C_{12} C_{22}^{-1} \\ -C_{22,1}^{-1} C_{21} C_{11}^{-1} & C_{22,1}^{-1} \end{bmatrix}. \quad (۹.۴)$$

□

طبق لم ۱.۲.۴، نتیجه دلخواه حاصل می‌شود.

قضیه ۱.۲.۴. تحت شرایط نظم، آماره آزمون فرضیه صفر $\mathcal{H}_0: \beta_{\tau} = 0$ برابر است با

$$\mathcal{L}_n = \frac{(\hat{\beta}_{\tau}^{\text{uLAD}})^{\top} C_{\tau\tau,1} (\hat{\beta}_{\tau}^{\text{uLAD}})}{\tau^2}, \quad \tau = [2f(\circ)]^{-1}, \quad (10.4)$$

که دارای توزیع χ^2 مجانبی با p_{τ} درجه آزادی است.

برهان. بر اساس نتیجه ۱.۲.۴،

$$\sqrt{n}(\hat{\beta}_{\tau}^{\text{uLAD}} - \beta_{\tau}) \sim \mathcal{N}_{p_{\tau}}(0, \tau^2 C_{\tau\tau,1}^{-1}). \quad (11.4)$$

بنابراین، طبق ویژگی‌های توزیع نرمال و هم‌چنین فرض ب، نتیجه می‌شود

$$\frac{(\hat{\beta}_{\tau}^{\text{uLAD}} - \beta_{\tau})^{\top} C_{\tau\tau,1} (\hat{\beta}_{\tau}^{\text{uLAD}} - \beta_{\tau})}{\tau^2}, \quad (12.4)$$

که دارای توزیع χ^2 با p_{τ} درجه آزادی است. تحت فرضیه صفر، $\beta_{\tau} = 0$ و می‌توان رابطه (۱۰.۴) را نتیجه گرفت و اثبات کامل می‌شود. \square

بدیهی است که آماره آزمون به توزیع خطاها از طریق τ^2 وابسته است. با توجه به مجهول بودن این توزیع، می‌توان از برآوردگر ناپارامتری به صورت زیر استفاده نمود:

$$\hat{f}(e) = \frac{1}{h} \mathbb{1}_{\{|e_i| \leq h/2\}},$$

که در آن $h = h_n$ پهنای بانندی است که وقتی $n \rightarrow \infty$ ، $h \rightarrow 0$ در حالت کلی، فرض کنید

$$\hat{f}(e) = \frac{1}{h} K\left(\frac{e - e_i}{h}\right),$$

که تابع هسته $K(\cdot)$ در رابطه $\int K(u) du = 1$ صدق می‌کند.

وقتی $\mathcal{H}_0: \beta_{\tau} = 0$ برقرار باشد، برآوردگر RLAD واریانس مجانبی کمتری نسبت به برآوردگر ULAD دارد. با این وجود، به‌ازای $\beta_{\tau} \neq 0$ ، برآوردگر RLAD ممکن است اریب و در بعضی موارد، ناسازگار باشد. به‌این دلیل، برآوردگر آزمون اولیه LAD (PTLAD) به‌صورت زیر تعریف می‌شود:

$$\hat{\beta}_{\tau}^{\text{ptLAD}} = \hat{\beta}_{\tau}^{\text{uLAD}} - (\hat{\beta}_{\tau}^{\text{uLAD}} - \hat{\beta}_{\tau}^{\text{rLAD}}) I(\mathcal{L}_n \leq \chi_{p_{\tau}, \alpha}^2), \quad p_{\tau} \geq 1,$$

برآوردگر PTLAD یک برآوردگر گسسته است و بر اساس نتیجه آزمون که شدیداً به مقدار α وابسته است، همواره به یکی از دو برآوردگر محدودنشده یا محدودشده LAD ساده می‌شود. برای غلبه بر این مسئله، برآوردگر انقباضی LAD که نوع همواری از برآوردگر آزمون اولیه WLAD است، تعریف می‌شود.

برآوردگر انقباضی نوع استاین LAD (SLAD) پارامتر β_1 را با نماد $\hat{\beta}_1^{\text{sLAD}}$ نشان می‌دهند و به صورت

$$\begin{aligned}\hat{\beta}_1^{\text{sLAD}} &= \hat{\beta}_1^{\text{uLAD}} - c\mathcal{L}_n^{-1} (\hat{\beta}_1^{\text{uLAD}} - \hat{\beta}_1^{\text{rLAD}}) \\ &= \hat{\beta}_1^{\text{rLAD}} + (1 - c\mathcal{L}_n^{-1}) (\hat{\beta}_1^{\text{uLAD}} - \hat{\beta}_1^{\text{rLAD}}),\end{aligned}$$

تعریف می‌شود که در آن c یک ثابت انقباضی است که مخاطره این برآوردگر را کمینه می‌کند و معمولاً آن را برابر با $2 - p_2$ به ازای $p_2 \geq 3$ در نظر می‌گیرند.

به منظور عدم رویارویی با پدیده بیش انقباضی در برآوردگر SLAD، برآوردگر جزء مثبت انقباضی نوع-استاین LAD، (PRSLAD)، به صورت زیر تعریف می‌شود:

$$\begin{aligned}\hat{\beta}_1^{\text{sLAD}+} &= \hat{\beta}_1^{\text{rLAD}} + (1 - c\mathcal{L}_n^{-1})^+ (\hat{\beta}_1^{\text{uLAD}} - \hat{\beta}_1^{\text{rLAD}}), \\ &= \hat{\beta}_1^{\text{sLAD}} - (1 - c\mathcal{L}_n^{-1}) (\hat{\beta}_1^{\text{uLAD}} - \hat{\beta}_1^{\text{rLAD}}) I(\mathcal{L}_n \leq c). \\ \hat{\beta}_1^{\text{sLAD}+} &= \hat{\beta}_1^{\text{sLAD}} - (1 - c\mathcal{L}_n^{-1}) (\hat{\beta}_1^{\text{uLAD}} - \hat{\beta}_1^{\text{rLAD}}) I(\mathcal{L}_n \leq c).\end{aligned}$$

۱.۲.۴ ویژگی برآوردگرها

دنباله‌ای از فرضیه‌های مقابل به صورت $\mathcal{K}_{(n)}$ در نظر بگیرید:

$$\mathcal{K}_{(n)} : \beta_2 = n^{-\frac{1}{2}} \xi, \quad \xi = (\xi_{p_1+1}, \dots, \xi_p)^\top \in \mathbb{R}^{p_2} \quad (13.4)$$

لم ۲.۲.۴. با در نظر گرفتن $\mathcal{K}_{(n)}$ ،

$$\hat{\beta}_1^{\text{rLAD}} = \hat{\beta}_1^{\text{uLAD}} + C_{11}^{-1} C_{12} \hat{\beta}_2^{\text{uLAD}}$$

قضیه ۲.۲.۴. فرض کنید شرایط نظم برقرار باشند، آن‌گاه

$$\sqrt{n} (\hat{\beta}_1^{\text{rLAD}} - \beta_1) \xrightarrow{D} \mathcal{N}_{p_1} (0, \tau^2 C_{11}^{-1})$$

برهان. از آن جایی که

$$\sqrt{n} (\hat{\beta}_1^{\text{uLAD}} - \beta_1) \sim \mathcal{N}_{p-q} (0, \tau^2 C_{11.2}^{-1}).$$

برای به دست آوردن توزیع مجانبی برآوردگر RLAD، باید از توزیع شرطی $(\hat{\beta}_1^{\text{uLAD}} | \hat{\beta}_2^{\text{uLAD}} = \beta_2)$ استفاده شود. از آن جایی که توزیع $X_1 | X_2 = x_2$ برابر است با

$$\mathcal{N}_{p-q} (\mu_1 + C_{12} C_{22}^{-1} (x_2 - \mu_2), C_{11} - C_{12} C_{22}^{-1} C_{21}). \quad (14.4)$$

□ اکنون، با جایگذاری رابطه (۹.۴) در رابطه (۱۴.۴) نتیجه مورد نظر به دست می‌آید.

فرض کنید شکل کلی برآوردگر را بتوان به صورت

$$\hat{\beta}_1^{*LAD} = \hat{\beta}_1^{uLAD} - g(\mathcal{L}_n) (\hat{\beta}_1^{uLAD} - \hat{\beta}_1^{rLAD}). \quad (15.4)$$

نوشت.

(الف) اگر $g(\mathcal{L}_n) = 0$ ، آن گاه برآوردگر ULAD برابر است با $\hat{\beta}_1^{*LAD} = \hat{\beta}_1^{uLAD}$ ،

(ب) اگر $g(\mathcal{L}_n) = 1$ ، آن گاه $\hat{\beta}_1^{*LAD} = o = \hat{\beta}_1^{rLAD}$ برآوردگر RLAD است.

(ج) اگر $g(\mathcal{L}_n) = I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha})$ ، آن گاه $\hat{\beta}_1^{*LAD} = \hat{\beta}_1^{ptLAD}$ همان برآوردگر PTLAD است.

(د) اگر $g(\mathcal{L}_n) = c\mathcal{L}_n^{-1}$ ، آن گاه $\hat{\beta}_1^{*LAD} = \hat{\beta}_1^{sLAD}$ همان برآوردگر SLAD است.

(ه) اگر $g(\mathcal{L}_n) = c\mathcal{L}_n^{-1} + (1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \leq c)$ ، آن گاه $\hat{\beta}_1^{*LAD} = \hat{\beta}_1^{sLAD+}$ PRSLADE است.

و \mathcal{L}_n در رابطه (۱۰.۴) تعریف شده است و دارای توزیع χ^2 با p_2 درجه آزادی باشد.

قضیه ۳.۲.۴. تحت شرایط نظم در مجموعه فرضیه‌های مقابل موضعی $\mathcal{K}_{(n)}$ ، ADB برآوردگرها برابر است با

$$\begin{aligned} \text{ADB}(\hat{\beta}_1^{uLAD}) &= o \\ \text{ADB}(\hat{\beta}_1^{rLAD}) &= -C_{11}^{-1}C_{12}\xi = -\delta \\ \text{ADB}(\hat{\beta}_1^{ptLAD}) &= -\delta H_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^2), \quad \Delta^2 = \tau^{-2}\xi^T C_{22.1}\xi \\ \text{ADB}(\hat{\beta}_1^{sLAD}) &= -(p_2 - 2)\delta E \left[\chi_{p_2+2}^{-2}(\Delta^2) \right] \\ \text{ADB}(\hat{\beta}_1^{sLAD+}) &= \text{ADB}(\hat{\beta}_1^{sLAD}) \\ &\quad - \delta E \left[\left(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2) \right) I \left(\chi_{p_2+2}^2(\Delta^2) \leq (p_2 - 2) \right) \right], \end{aligned}$$

برهان. با تعریف ADB، رابطه (۱۶.۲)، تحت فرضیه مقابل موضعی، $\{\mathcal{K}_{(n)}\}$ نتیجه می‌شود

$$\begin{aligned} \text{ADB}(\hat{\beta}_n^{*LAD}) &= \lim_{n \rightarrow \infty} E \left[\sqrt{n} (\hat{\beta}_1^{*LAD} - \beta_1) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\sqrt{n} (\hat{\beta}_1^{uLAD} - g(\mathcal{L}_n) (\hat{\beta}_1^{uLAD} - \hat{\beta}_1^{rLAD}) - \beta_1) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\sqrt{n} (\hat{\beta}_1^{uLAD} - \beta_1) \right] - \lim_{n \rightarrow \infty} E \left[\sqrt{n} g(\mathcal{L}_n) (\hat{\beta}_1^{uLAD} - \hat{\beta}_1^{rLAD}) \right] \\ &= \text{ADB}(\hat{\beta}_1^{uLAD}) - \lim_{n \rightarrow \infty} E \left[\sqrt{n} g(\mathcal{L}_n) C_{11}^{-1} C_{12} \hat{\beta}_2^{uLAD} \right] \\ &= \text{ADB}(\hat{\beta}_1^{uLAD}) - C_{11}^{-1} C_{12} \lim_{n \rightarrow \infty} E \left[\sqrt{n} g(\mathcal{L}_n) \hat{\beta}_2^{uLAD} \right] = -\delta E \left[g(\chi_{p_2+2}^2(\Delta^2)) \right]. \end{aligned}$$

□

خط آخر از تعریف δ و قضیه ۵.۲.آ نتیجه می‌شود.

نتیجه ۲.۲.۴. تحت شرایط نظم در مجموعه فرضیه‌های مقابل موضعی $\mathcal{K}_{(n)}$ ، ADQB برآوردگرها برابر است با

$$\begin{aligned} \text{ADQB}(\hat{\beta}_1^{\text{uLAD}}) &= 0 \\ \text{ADQB}(\hat{\beta}_1^{\text{rLAD}}) &= \delta^\top C_{11,2} \delta \\ \text{ADQB}(\hat{\beta}_1^{\text{ptLAD}}) &= \delta^\top C_{11,2} \delta \left\{ H_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^2) \right\}^2 \\ \text{ADQB}(\hat{\beta}_1^{\text{sLAD}}) &= (p_2 - 2) \delta^\top C_{11,2} \delta \left\{ E \left[\chi_{p_2+2}^{-2}(\Delta^2) \right] \right\}^2 \\ \text{ADQB}(\hat{\beta}_1^{\text{sLAD}+}) &= \delta^\top C_{11,2} \delta \left\{ (p_2 - 2) E \left[\chi_{p_2+2}^{-2}(\Delta^2) \right] \right. \\ &\quad \left. - E \left[\left(1 - (p_2 - 2) \chi_{p_2+2}^{-2}(\Delta^2) \right) I \left(\chi_{p_2+2}^2(\Delta^2) < (p_2 - 2) \right) \right] \right\}^2 \end{aligned}$$

به‌ازای $C_{12} = 0$ ، $C_{21} C_{11}^{-1} C_{11,2} C_{11}^{-1} C_{12} = 0$ ، بنابراین ADQB‌ها برای همه ξ ها به مقدار مشترک صفر کاهش پیدا می‌کنند. همه این تغییرات، باعث می‌شوند برآوردگرها ADQB-معادل شوند. از این رو، در ادامه فرض می‌شود که $C_{12} \neq 0$. مقدار ADQB برآوردگر $\hat{\beta}_1^{\text{rLAD}}$ تابع بی‌کران از $\tau^2 \Delta^2$ است. به‌منظور بررسی کردن برآوردگرهای $\text{ADQB}(\hat{\beta}_1^{\text{sLAD}})$ و $\text{ADQB}(\hat{\beta}_1^{\text{sLAD}+})$ ، از نتایج ماتریسی زیر استفاده می‌شود.

$$\begin{aligned} \text{Ch}_{\min}(\tau^2 \Delta^2) &\leq \frac{\tau^2 \xi^\top C_{21} C_{11}^{-1} C_{11,2} C_{11}^{-1} C_{12} \xi}{\xi^\top C_{22,1} \xi} \\ &\leq \text{Ch}_{\max}(\tau^2 C_{21} C_{11}^{-1} C_{11,2} C_{11}^{-1} C_{12} C_{22,1}^{-1}) \end{aligned}$$

بنابراین، $\text{ADQB}(\hat{\beta}_1^{\text{sLAD}})$ در نقطه $\xi^\top C_{21} C_{11}^{-1} C_{11,2} C_{11}^{-1} C_{12} \xi = 0$ از صفر شروع می‌شود، تا یک نقطه افزایش می‌یابد و با توجه به اینکه $E \left[\chi_{p_2+2}^{-2}(\Delta^2) \right]$ تابعی لگاریتم-محدب نزولی از Δ^2 است، به سمت صفر نزول می‌یابد. رفتار $\hat{\beta}_1^{\text{sLAD}+}$ مشابه با $\hat{\beta}_1^{\text{sLAD}}$ است. با این وجود، به‌ازای همه مقادیر Δ^2 ، منحنی توان دوم اریبی $\hat{\beta}_1^{\text{sLAD}+}$ همچنان زیر منحنی $\hat{\beta}_1^{\text{sLAD}}$ قرار می‌گیرد.

قضیه ۴.۲.۴. تحت شرایط نظم در مجموعه فرضیه‌های مقابل موضعی $\mathcal{K}_{(n)}$ ، ADQR مخاطره‌ها برابر است با

$$\begin{aligned} \text{ADR}(\hat{\beta}_1^{\text{uLAD}}) &= \tau^2 \text{tr}(C_{11,2}^{-1}) \\ \text{ADR}(\hat{\beta}_1^{\text{rLAD}}) &= \tau^2 \text{tr}(C_{11}^{-1}) + \delta^\top \delta \\ \text{ADR}(\hat{\beta}_1^{\text{ptLAD}}) &= \tau^2 \text{tr}(C_{11,2}^{-1}) - \tau^2 \text{tr}(C_{21} C_{11}^{-2} C_{12}) H_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^2) \\ &\quad + \delta^\top \delta \left\{ 2 H_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^2) - H_{p_2+4}(\chi_{p_2,\alpha}^2; \Delta^2) \right\} \\ \text{ADR}(\hat{\beta}_1^{\text{sLAD}}) &= \tau^2 \left\{ \text{tr}(C_{11,2}^{-1}) - (p_2 - 2) \text{tr}(C_{21} C_{11}^{-2} C_{12} C_{22,1}^{-1}) \right. \\ &\quad \left. \times \left\{ 2 E \left[\chi_{p_2+2}^{-2}(\Delta^2) \right] - (p_2 - 2) E \left[\chi_{p_2+2}^{-4}(\Delta^2) \right] \right\} \right\} \\ &\quad + (p_2 - 4) \delta^\top \delta E \left[\chi_{p_2+4}^{-4}(\Delta^2) \right] \end{aligned}$$

$$\begin{aligned} \text{ADR}(\hat{\beta}_1^{\text{sLAD}+}) &= \text{ADR}(\hat{\beta}_1^{\text{sLAD}}) \\ &+ \tau^\top (p_\tau - \nu) \text{tr}(C_{\tau 1} C_{11}^{-1} C_{11}^{-\nu} C_{12} C_{\tau \tau}^{-1}) \\ &\times \left\{ \nu E \left[\chi_{p_\tau + \nu}^{-\nu}(\Delta^\nu) I(\chi_{p_\tau + \nu}^\nu(\Delta^\nu) \leq p_\tau - \nu) \right] \right. \\ &- (p_\tau - \nu) E \left[\chi_{p_\tau + \nu}^{-\nu}(\Delta^\nu) I(\chi_{p_\tau + \nu}^\nu(\Delta^\nu) \leq p_\tau - \nu) \right] \left. \right\} \\ &- \tau^\top \text{tr}(C_{\tau 1} C_{11}^{-1} C_{11}^{-1} C_{12} C_{\tau \tau}^{-1}) H_{p_\tau + \nu}(p_\tau - \nu; \Delta^\nu) \\ &+ \delta^\top \delta \left\{ \nu H_{p_\tau + \nu}(p_\tau - \nu; \Delta^\nu) - H_{p_\tau + \nu}(p_\tau - \nu; \Delta^\nu) \right\} \\ &- (p_\tau - \nu) \delta^\top \delta \left\{ \nu E \left[\chi_{p_\tau + \nu}^{-\nu}(\Delta^\nu) I(\chi_{p_\tau + \nu}^\nu(\Delta^\nu) \leq p_\tau - \nu) \right] \right. \\ &- \nu E \left[\chi_{p_\tau + \nu}^{-\nu}(\Delta^\nu) I(\chi_{p_\tau + \nu}^\nu(\Delta^\nu) \leq p_\tau - \nu) \right] \left. \right\} \\ &+ (p_\tau - \nu) E \left[\chi_{p_\tau + \nu}^{-\nu}(\Delta^\nu) I(\chi_{p_\tau + \nu}^\nu(\Delta^\nu) \leq p_\tau - \nu) \right] \end{aligned}$$

برهان. بنا بر تعریف ADR،

$$\begin{aligned} \text{ADR}(\hat{\beta}_1^{*\text{LAD}}) &= \lim_{n \rightarrow \infty} E \left[\left(\hat{\beta}_1^{*\text{LAD}} - \beta_1 \right)^\top \left(\hat{\beta}_1^{*\text{LAD}} - \beta_1 \right) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\left(\hat{\beta}_1^{\text{uLAD}} - g(\mathcal{L}_n) \hat{\beta}_1^{\text{uLAD}} - \beta_1 \right)^\top \left(\hat{\beta}_1^{\text{uLAD}} - g(\mathcal{L}_n) \hat{\beta}_1^{\text{uLAD}} - \beta_1 \right) \right] \\ &= \lim_{n \rightarrow \infty} E \left[\left(\hat{\beta}_1^{\text{uLAD}} - \beta_1 \right)^\top \left(\hat{\beta}_1^{\text{uLAD}} - \beta_1 \right) \right] \\ &- \nu \lim_{n \rightarrow \infty} E \left[n \left(\hat{\beta}_1^{\text{uLAD}} - \beta_1 \right)^\top \hat{\beta}_1^{\text{uLAD}} g(\mathcal{L}_n) \right] \\ &+ \lim_{n \rightarrow \infty} E \left[n \left[\hat{\beta}_1^{\text{uLAD}} \right]^\top \hat{\beta}_1^{\text{uLAD}} g(\mathcal{L}_n^\nu) \right] \\ &= \text{ADR}(\hat{\beta}_1^{\text{uLAD}}) - \nu \lim_{n \rightarrow \infty} E \left[n \left[\hat{\beta}_1^{\text{uLAD}} \right]^\top \hat{\beta}_1^{\text{uLAD}} g(\mathcal{L}_n) \right] \\ &+ \nu \delta^\top \lim_{n \rightarrow \infty} E \left[\sqrt{n} \hat{\beta}_1^{\text{uLAD}} g(\mathcal{L}_n) \right] + \lim_{n \rightarrow \infty} E \left[n \left[\hat{\beta}_1^{\text{uLAD}} \right]^\top \hat{\beta}_1^{\text{uLAD}} g(\mathcal{L}_n^\nu) \right]. \end{aligned} \tag{۱۶.۴}$$

رابطه (۱۸.۲) منجر به نتیجه زیر می‌شود:

$$\text{ADR}(\hat{\beta}_1^{\text{uLAD}}) = \eta^\top \text{tr}(C^{-1}). \tag{۱۷.۴}$$

بر اساس قضیه‌های ۵.۲.آ و ۶.۲.آ و همچنین رابطه‌های (۱۶.۴)، (۱۷.۴) بعد از کمی محاسبات جبری نتیجه می‌شود

$$\begin{aligned} \text{ADR}(\hat{\beta}_n^{*\text{LAD}}) &= \eta^\top \text{tr}(C^{-1}) \left\{ 1 - \nu E \left[g(\chi_{p_\tau + \nu}^\nu(\Delta^\nu)) \right] + E \left[g^\nu(\chi_{p_\tau + \nu}^\nu(\Delta^\nu)) \right] \right\} \\ &+ \delta^\top \delta \left\{ \nu E \left[\chi_{p_\tau + \nu}^\nu(\Delta^\nu) \right] - \nu E \left[\chi_{p_\tau + \nu}^\nu(\Delta^\nu) \right] + E \left[g^\nu(\chi_{p_\tau + \nu}^\nu(\Delta^\nu)) \right] \right\}. \end{aligned} \tag{۱۸.۴}$$

با جایگذاری هر یک از روابط توابع g در رابطه‌های (۲۰.۲) و (۱۸.۴) نتیجه به دست می‌آید. □

طبق قضیه ۴.۲.۴ برای $C_{12} = 0$ ، $C_{\tau 1} C_{11}^{-\nu} C_{12} = 0$ و $C_{11,2} = C_{11}$ و بنابراین، همه ADQRها برای هر ξ ، به مقدار مشترک $\tau^\top \text{tr}(C_{11}^{-1})$ کاهش پیدا می‌کنند. بنابراین، همه این تغییرات، ADQR معادل خواهند شد. در ادامه، فرض می‌شود که $C_{12} \neq 0$.

با مقایسه $ADR(\hat{\beta}_1^{sLAD+})$ با $ADR(\hat{\beta}_1^{sLAD})$ ، مشاهده می‌شود که برآوردگر PTLAD بر برآوردگر SLAD به‌ازای همه مقادیر ξ برتری دارد. برای برخی از مقادیرهای ξ ، نامساوی اکید برقرار است. بنابراین، ADQR برآوردگر $\hat{\beta}_1^{sLAD+}$ کوچکتر از مخاطره جانبی $\hat{\beta}_1^{uLAD}$ در فضای پارامتر است و وقتی Δ^2 به ∞ نزدیک شود، کران بالا به‌دست می‌آید. ADQR برآوردگر PTLAD وقتی Δ^2 از \circ دور می‌شود، به‌طور یکنوا به سمت $ADR(\hat{\beta}_1^{uLAD})$ نزدیک می‌شود. این نتیجه می‌دهد که

$$ADR(\hat{\beta}_1^{sLAD+}) \leq ADR(\hat{\beta}_1^{sLAD}) \leq ADR(\hat{\beta}_1^{uLAD}), \quad \forall \xi \in \mathbb{R}^p$$

به‌ازای برخی مقادیر ξ ، نامساوی اکید برقرار است. بنابراین، $\hat{\beta}_1^{sLAD}$ و $\hat{\beta}_1^{sLAD+}$ بهتر از برآوردگر uLAD در فضای پارامتر شامل Δ^2 رفتار می‌کند. وقتی $\Delta^2 = \circ$ یا نزدیک آن است، مقدار مخاطره $\hat{\beta}_1^{uLAD}$ زیاد می‌شود.

در این فصل می‌توان همانند قضیه‌های ۹.۳.۲ – ۱۴.۳.۲، قضیه‌های مرتبط را بیان و اثبات کرد و برای عدم تکرار، می‌توان به مقاله **نوروزی‌راد و همکاران (۲۰۱۷a)** اشاره کرد.

۲.۲.۴ مطالعه شبیه‌سازی

از یک فرآیند شبیه‌سازی مونت کارلو برای مقایسه کارایی برآوردگرهای انقباضی با برآوردگر LAD-LASSO استفاده می‌شود. متغیرهای پاسخ طبق مدل (۱.۱) تولید می‌شود که خطای آن از توزیع دم پهن نمایی استاندارد دوگانه (توزیع لاپلاس) تولید شده است. نمودار ۱.۴ نشان می‌دهد که احتمال وجود داده‌های پرت در توزیع لاپلاس بیشتر از توزیع نرمال است.

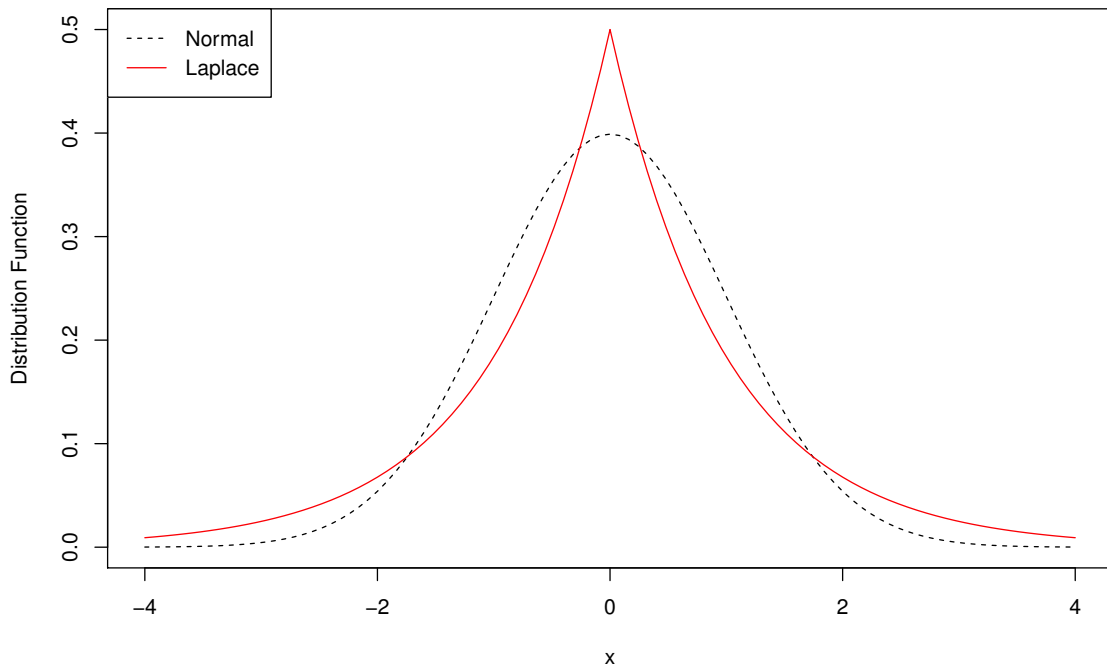
متغیر x_i از توزیع نرمال استاندارد p بعدی تولید شده است. فرض کنید $p_1 = p - p_2$. به‌ازای هر $i = p_1 + 1, \dots, p$ علاقه‌مندیم فرضیه $H_0: \beta_i = \circ$ را آزمون کنیم. بدین منظور، بردار پارامترها به‌صورت $\beta = (\beta_1^\top, \beta_2^\top)^\top = (\beta_1^\top, \mathbf{o}^\top)^\top$ افراز می‌شود که $\beta_1 = (1, \dots, 1)^\top$ یک بردار p_1 تایی از ۱ها و \mathbf{o} یک بردار p_2 تایی از ۰هاست. فرآیند شبیه‌سازی ۱۰۰۰ بار تکرار می‌شود تا نتایج پایداری به‌دست آید.

به منظور یافتن درجه عدم قطعیت درباره فرضیه صفر، فرض کنید $\|\beta - \beta^{(\circ)}\|_2^2 = \Delta^2$ ، که در آن $\beta^{(\circ)} = (\beta_1^\top, \mathbf{o}^\top)$ پارامتر β به‌صورت زیر تعریف می‌شود.

$$(\beta_1^\top, \sqrt{\Delta^2/p_2}, \dots, \sqrt{\Delta^2/p_2}).$$

انتخاب $\Delta^2 = \circ$ نتیجه می‌دهد که مدل پیشنهادی تحت فرضیه صفر، درست است در حالی که $\Delta^2 > \circ$ بیان می‌کند که از مدل فرضی، چه‌قدر فاصله گرفته شده است. مقادیر مختلف Δ^2 در بازه $[0, 4]$ در نظر گرفته شده است.

دقت کنید که برای محاسبه آماره آزمون در رابطه (۱۰.۴)، نیاز به یافتن τ^2 است. در این قسمت، فرض می‌کنیم که خطاها از توزیع استاندارد لاپلاس تولید شده باشند و با توجه به اینکه میانه این توزیع، \circ است، τ^2 برابر با $1/(4 \times (\circ/5)^2) = 1$ خواهد بود.



شکل ۱.۴: توزیع‌های لاپلاس و نرمال

کارایی مخاطره با معیار MSE نسبی محاسبه می‌شود. RMSE برآوردگر پیشنهادی $\hat{\beta}_1^*$ به صورت زیر تعریف می‌شود:

$$\text{RMSE}(\hat{\beta}_1^{\text{uLAD}}) = \frac{\text{MSE}(\hat{\beta}_1^{\text{uLAD}})}{\text{MSE}(\hat{\beta}_1^*)} \quad (19.4)$$

که $\hat{\beta}_1^*$ یکی از برآوردگرهای RLAD، PTLAD، SLAD، PRLAD و LAD-LASSO است. مقدار به دست آمده از RMSE، چنانچه بزرگتر از یک باشد، نشان دهنده درجه برتری برآوردگر $\hat{\beta}_1^*$ بر $\hat{\beta}_1^{\text{uLAD}}$ است.

مقادیر RMSE برای برآوردگرهای RLAD، PTLAD، SLAD، PRLAD و LAD-LASSO برای مقادیر مختلف n ، p_1 و p_2 محاسبه شده است. با توجه به اینکه نتایج در تمامی حالت‌های ترکیبی (n, p, p_1) مشابه است، تنها نتایج برای $n = 50$ و $(p_1, p_2) = (3, 5)$ در جدول ۱.۴ نشان داده شده است.

بر اساس جدول ۱.۴، نتایج زیر به دست می‌آید:

حالت اول، $\Delta^2 = 0$: بدیهی است، برآوردگر RLAD بهتر از سایر برآوردگر در همه حالت‌های در نظر گرفته شده در مطالعه شبیه‌سازی عمل می‌کند.

جدول ۱.۴: مقادیر RMSE شبیه‌سازی شده برای برآوردگرهای PTLAD، RLAD، SLAD و PRLAD برای $(p_1, p_2) = (3, 5)$ ، $\alpha = 0.5$ و $n = 50$

برآوردگرها				
PRLAD	SLAD	PTLAD	RLAD	Δ^2
۲/۰۳۷	۱/۲۹۵	۲/۱۵۸	۳/۵۴۳	۰/۰۰
۱/۴۶۲	۱/۲۶۷	۱/۲۲۷	۱/۹۹۲	۰/۲۵
۱/۱۵۵	۱/۱۵۲	۰/۸۱۷	۰/۸۱۰	۰/۵۰
۱/۰۷۴	۱/۰۷۴	۰/۸۹۳	۰/۴۲۹	۰/۷۵
۱/۰۳۹	۱/۰۳۹	۰/۹۷۵	۰/۲۷۰	۱/۰۰
۱/۰۳۰	۱/۰۳۰	۱/۰۰۰	۰/۱۷۳	۱/۲۵
۱/۰۱۹	۱/۰۱۹	۱/۰۰۰	۰/۱۲۳	۱/۵۰
۱/۰۱۴	۱/۰۱۴	۱/۰۰۰	۰/۰۹۴	۱/۷۵
۱/۰۱۲	۱/۰۱۲	۱/۰۰۰	۰/۰۷۳	۲/۰۰
۱/۰۱۱	۱/۰۱۱	۱/۰۰۰	۰/۰۵۶	۲/۲۵
۱/۰۰۶	۱/۰۰۶	۱/۰۰۰	۰/۰۴۷	۲/۵۰
۱/۰۰۵	۱/۰۰۵	۱/۰۰۰	۰/۰۳۷	۲/۷۵
۱/۰۰۶	۱/۰۰۶	۱/۰۰۰	۰/۰۳۱	۳/۰۰
۱/۰۰۵	۱/۰۰۵	۱/۰۰۰	۰/۰۲۷	۳/۲۵
۱/۰۰۵	۱/۰۰۵	۱/۰۰۰	۰/۰۲۳	۳/۵۰
۱/۰۰۴	۱/۰۰۴	۱/۰۰۰	۰/۰۲۱	۳/۷۵
۱/۰۰۳	۱/۰۰۳	۱/۰۰۰	۰/۰۱۸	۴/۰۰

حالت دوم، $\Delta^2 \neq 0$: هر چه قدر که محدودیت از فرضیه صفر دور می‌شود، مقدار RMSE برآوردگر RLAD شدیداً کاهش می‌یابد و به سرعت مقدارش، کمتر از یک می‌شود. مقدار RMSE برآوردگر PRLAD با کمترین سرعت در دامنه Δ^2 با دور شدن از $\Delta^2 = 0$ به ۱ نزدیک می‌شود. این موضوع نشان می‌دهد حتی در صورتی که اطلاع دقیقی در مورد زیرفضای پارامتر وجود نداشته باشد (به عبارت دیگر، $\beta_2 \neq 0$)، برآوردگر جزء مثبت انقباضی نوع استاین LAD برتری را نشان می‌دهد.

هم‌چنین برآوردگرهای پیشنهادی با برآوردگر LAD-LASSO، به‌ازای مقادیر متفاوت n و p است. نتایج در جدول ۲.۴ آمده است. به‌آسانی می‌توان نتیجه گرفت که برآوردگرهای پیشنهادی وقتی تعداد پارامترهای مزاحم، p_2 ، بیشتر است، بهتر رفتار می‌کند. هم‌چنین برآوردگر LAD-LASSO با وجودی که بهتر از برآوردگر محدود نشده LAD رفتار می‌کند، در مقایسه با برآوردگرهای ارائه شده مخاطره بیشتری دارد.

۳.۲.۴ تحلیل داده‌های نرخ رشد

مجموعه داده‌های نرخ رشد (بارو و لی، ۱۹۹۴) شامل ۱۶۱ مشاهده از نرخ‌های رشد اقتصادی ملیت‌ها در سال‌های ۱۹۶۰، ۱۹۶۵، ۱۹۷۰، ۱۹۷۵، ۱۹۸۰ و ۱۹۸۵ است. این مجموعه داده در بسته نرم‌افزاری quantreg در نرم‌افزار R وجود دارد.

جدول ۲.۴: مقادیر RMSE شبیه‌سازی شده برآوردگرها نسبت به برآوردگر uLAD به‌ازای $p_1 = 3$ و $\Delta^2 = 0$

LAD-LASSO	PRLAD	SLAD	PTLAD	RLAD	p_2	n
۱/۰۴۳	۲/۱۱۹	۱/۰۳۸	۲/۲۹۹	۳/۶۱۱	۵	۵۰
۱/۱۰۶	۳/۳۶۶	۱/۴۷۶	۲/۷۸۸	۶/۱۷۹	۱۰	
۱/۱۰۰	۶/۹۳۸	۱/۴۳۴	۵/۷۱۰	۱۱/۵۱۸	۱۵	
۱/۲۳۱	۱۵/۲۸۰	۱/۵۶۷	۱۱/۸۳۴	۲۰/۸۷۴	۲۰	
۱/۰۹۶	۱/۸۲۹	۱/۰۳۲	۱/۹۷۴	۲/۹۷۷	۵	۸۰
۱/۱۱۷	۳/۱۱۶	۲/۱۲۲	۲/۶۱۰	۵/۶۵۳	۱۰	
۱/۱۳۹	۴/۴۴۰	۱/۹۰۴	۳/۰۲۷	۹/۵۱۴	۱۵	
۱/۱۹۳	۶/۸۰۶	۲/۷۵۵	۴/۰۱۲	۱۴/۹۹۶	۲۰	
۱/۰۵۹	۱/۹۰۳	۱/۴۴۴	۲/۰۵۸	۳/۱۰۹	۵	۱۰۰
۱/۱۳۹	۳/۲۴۸	۱/۹۹۵	۲/۷۵۷	۵/۸۲۷	۱۰	
۱/۱۸۰	۴/۴۴۵	۲/۲۲۴	۳/۱۲۹	۸/۴۵۶	۱۵	
۱/۲۱۶	۶/۳۳۰	۳/۴۳۲	۳/۵۸۰	۱۳/۹۵۳	۲۰	
۱/۱۱۳	۱/۸۵۷	۱/۳۰۳	۲/۰۲۰	۳/۱۷۳	۵	۱۲۵
۱/۱۵۳	۳/۵۳۶	۲/۴۷۸	۳/۱۴۵	۶/۱۳۴	۱۰	
۱/۱۷۳	۴/۰۸۹	۲/۴۵۱	۲/۷۴۹	۸/۳۹۹	۱۵	
۱/۱۸۴	۵/۴۴۸	۲/۸۳۳	۳/۱۵۳	۱۲/۰۷۰	۲۰	

کوئکر و ماچادو (۱۹۹۹) در مجموعه داده‌های نرخ رشد، توصیفی از رشد اقتصادی بین‌المللی را با رگرسیون چندکی ارائه دادند. اخیراً **یوزباشی و همکاران (۲۰۱۷)** برآورد آزمون اولیه و انقباضی نوع-استاین را در این مجموعه داده برای مدل‌های رگرسیونی چندکی بررسی و مقایسه‌ای بین این برآوردگرها و روش‌های جریمه‌شده چندکی انجام داده‌اند.

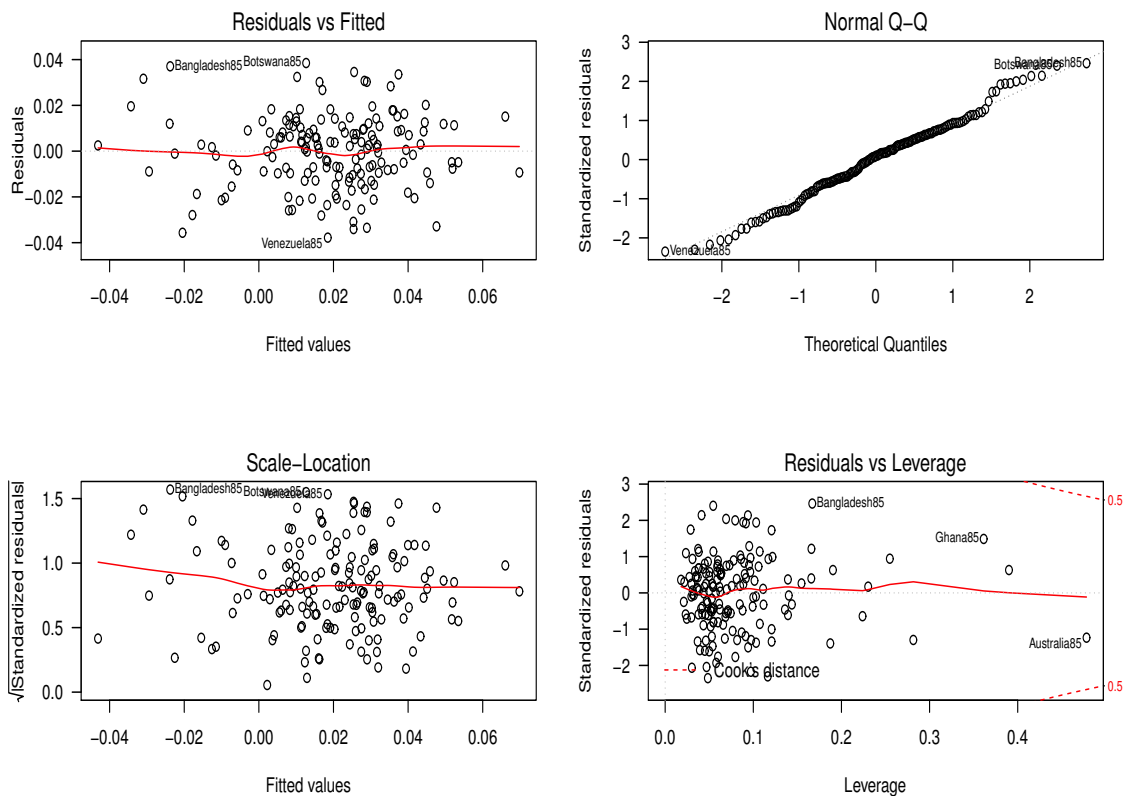
مجموعه داده بارو برای به‌دست آوردن یک مدل بین تغییرات سالیانه تولید ناخالص دولتی (GDP)، Y_{NET} و متغیرهای توضیحی: تحصیلات متوسطه مردان (MSE2)، تحصیلات متوسطه زنان (FSE2)، تحصیلات دانشگاهی زنان (FHE2)، تحصیلات دانشگاهی زنان (MHE2)، امید به زندگی (LEXP2)، سرمایه انسانی (INTR2)، نسبت تحصیلات به GDP (GEDY2)، نسبت سرمایه‌گذاری به GDP (IY2)، نسبت مصرف عمومی به GDP (GCONY2)، بازار سیاه حق بیمه (LBLAK2)، بی‌ثباتی سیاسی (POL2) و نرخ رشد تجارت (TTRAD2) به‌کار برده شده است. توصیفی از متغیرها در جدول ۳.۴ نشان داده شده است.

شکل ۲.۴ نشان می‌دهد که مشاهده‌های استرالیا، بنگلادش، بوستوانا، غنا و ونزوئلا در سال ۱۹۸۵ ممکن است داده پرت باشند. نتایج به‌کار بردن آزمون بنفرونی (**کوک و ویسبرگ، ۱۹۸۴**) تایید می‌کند این مشاهدات و همچنین بنگلادش در سال ۱۹۸۰ داده پرت هستند. بنابراین، در این مجموعه داده، داده‌های پرت وجود دارد و برای انجام رگرسیون LAD مناسب است.

جدول ۳.۴: آماره‌های توصیفی برای مجموعه داده نرخ رشد

پارامتر	متغیرها	کمینه	چارک اول	میانه	میانگین	چارک سوم	بیشینه
	Y.NET	-۰/۰۶	۰/۰۱	۰/۰۲	۰/۰۲	۰/۰۳	۰/۰۸
β_1	LGDP2	۵/۸۲	۶/۹۹	۷/۷۴	۷/۷۹	۸/۵۹	۹/۵۱
β_2	MSE2	۰/۰۲	۰/۳۲	۰/۷۲	۰/۹۷	۱/۲۵	۴/۲۳
β_3	FSE2	۰/۰۰	۰/۱۳	۰/۴۲	۰/۷۱	۰/۹۷	۴/۲۰
β_4	FHE2	۰/۰۰	۰/۰۱	۰/۰۳	۰/۰۸	۰/۰۹	۰/۷۱
β_5	MHE2	۰/۰۰	۰/۰۴	۰/۱۱	۰/۱۶	۰/۲۱	۱/۰۱
β_6	LEXP2	۳/۵۵	۳/۸۷	۴/۰۶	۴/۰۴	۴/۲۴	۴/۳۱
β_7	INTR2	-۱/۱۲	۰/۰۰	۰/۵۶	۱/۴۶	۱/۸۷	۱۵/۰۴
β_8	GEDY2	۰/۰۱	۰/۰۲	۰/۰۳	۰/۰۴	۰/۰۵	۰/۰۸
β_9	IY2	۰/۰۲	۰/۱۴	۰/۲۰	۰/۲۰	۰/۲۷	۰/۴۰
β_{10}	GCONY2	۰/۰۰	۰/۰۵	۰/۰۸	۰/۱۰	۰/۱۳	۰/۲۸
β_{11}	LBLAK2	۰/۰۰	۰/۰۰	۰/۰۷	۰/۱۹	۰/۲۴	۲/۱۲
β_{12}	POL2	۰/۰۰	۰/۰۰	۰/۰۵	۰/۱۷	۰/۲۳	۰/۸۰
β_{13}	TTRAD2	-۰/۱۲	-۰/۰۲	-۰/۰۱	۰/۰۰	۰/۰۱	۰/۱۴

lm(y.net ~ . - 1)



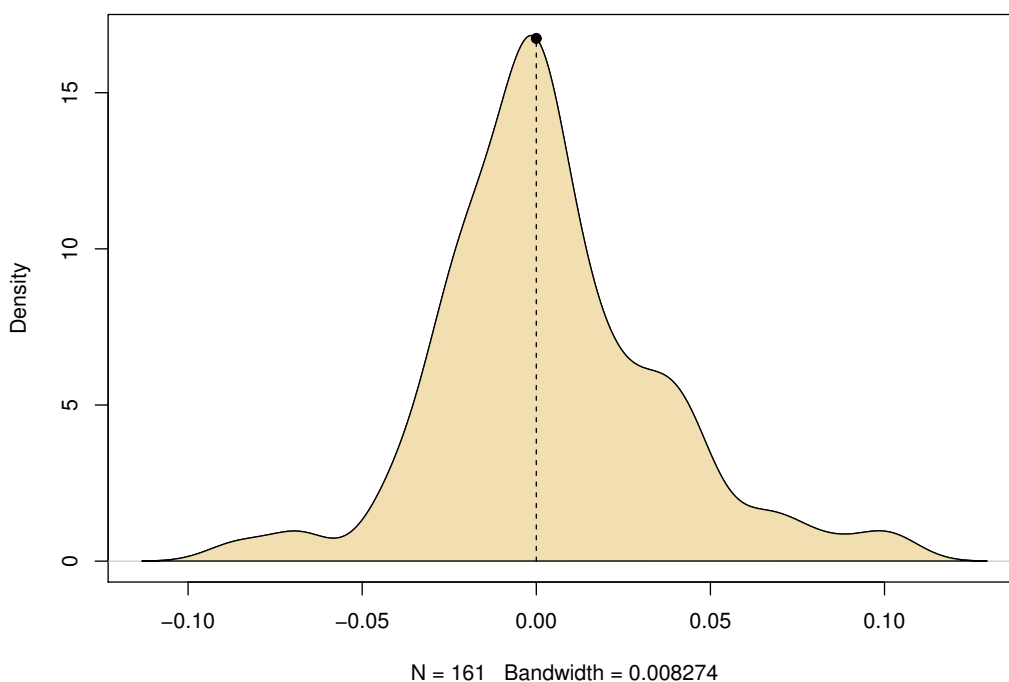
شکل ۲.۴: نمودارها برای شناسایی داده‌های پرت در مجموعه داده نرخ رشد

برآوردگر LAD-LASSO متغیرهای MSE2، FHE2، INTR2، GEDY2، TTRAD2 را به عنوان بهترین متغیرهایی که باید در مدل باشند، در نظر می‌گیرد. متغیرهای حذف شده یعنی LGDP، FSE2، POL2، LBLAK2، GCONY2، IY2، LEXP2، MHE2 به عنوان متغیرهای مزاحم نامیده می‌شوند و در قالب این فرضیه که ضریب همه متغیرهای مزاحم صفر است آزمون می‌شوند. این فرضیه صفر به صورت زیر در نظر گرفته می‌شود.

$$\beta_2 = (\beta_1, \beta_3, \beta_5, \beta_6, \beta_9, \beta_{10}, \beta_{11}, \beta_{12})^T = (0, 0, 0, 0, 0, 0, 0, 0)^T$$

دو مدل وجود دارد: مدل کامل و یک زیر مدل که متغیرهای مزاحم را نادیده می‌گیرد. برای مطالعه برآوردگرهای انقباضی، به آماره آزمون نیاز است که برای تعیین مقدار انقباض به کار گرفته می‌شود. در رابطه (۱۰.۴)، آماره آزمون معرفی شده که به خطاها وابسته است. بنابراین لازم است چگالی در نقطه صفر را محاسبه کرده و بر اساس آن τ^2 را برآورد کنیم. در شکل ۳.۴، می‌توان برآورد توزیع باقیمانده‌ها که توزیعی دم‌پهن با مقدار تقریبی $16/73$ در صفر است را مشاهده کرد. بنابراین، τ^2 ممکن است با 0.0597 تقریب زده شود.

Density estimate of the residuals



شکل ۳.۴: چگالی باقیمانده‌های برآوردگرهای LAD در مجموعه داده بارو

برآوردهای انقباضی در دو مرحله به دست می‌آیند: ابتدا، زیر مدلی انتخاب می‌گردد که متغیرهای آن بر اساس نتیجه برآوردگر LAD-LASSO به دست آمده است. در دومین مرحله، زیر مدل به دست آمده در مرحله اول به عنوان مدل محدود شده در نظر گرفته شده و برآوردهای انقباضی به دست می‌آیند.

در ادامه، برای به دست آوردن خطای پیشگویی مدل از روش 10° تایی اعتبارسنجی متقابل استفاده می‌شود. جدول ۴.۴ برآورد ضرایب رگرسیونی و RPE اعتبارسنجی متقابل 10° تایی را برای برآوردگرهای محدود نشده، محدود شده، جزء مثبت انقباضی نوع استاین LAD و نیز LAD-LASSO در ۱۰۰۰ بار تکرار نشان می‌دهد.

جدول ۴.۴: برآورد ضرایب و کارایی آن‌ها برای مجموعه داده بارو

پارامتر	ULAD	RLAD	PRLAD	LAD-LASSO
β_2	۰/۰۲۶۹۱	۰/۰۱۳۸۰	۰/۰۲۳۸۲	۰/۰۱۳۸۹
β_4	-۰/۰۲۶۳۹	-۰/۰۱۷۷۵	-۰/۰۲۴۳۵	-۰/۰۱۶۱۷
β_7	-۰/۰۵۳۱۷	-۰/۰۲۶۶۷	-۰/۰۴۶۹۲	-۰/۰۲۷۲۱
β_8	۰/۰۰۴۳۷	۰/۰۰۲۷۶	۰/۰۰۳۹۹	۰/۰۰۲۴۲
β_{13}	۰/۰۰۱۰۴	-۰/۰۰۷۳۵	۰/۰۰۰۹۳	-۰/۰۰۱۱۳
RPE	۱/۰۰	۱/۱۴	۱/۰۶	۱/۰۷

بر اساس جدول ۴.۴، برآوردگر RLAD، بیشترین کارایی را دارد زیرا بر اساس این فرضیه به دست آمده است که زیر-مدل، یک مدل درست است. برآوردگر PTLAD بر برآوردگر جزء مثبت نوع استاین LAD برتری دارد. به عبارت دیگر، برآوردگرهای پیشنهادی، کارایی بیشتری نسبت به برآوردگر LAD-LASSO دارند.

۳.۴ برآوردگرهای انقباضی LAD موزون

برآوردگر WLAD در مدل کامل، برآوردگر محدود نشده WLAD، (UWLAD) نامیده می‌شود که در حالت افراز شده با نماد $\hat{\beta}^{uWLAD} = (\hat{\beta}_1^{uWLAD}, \hat{\beta}_2^{uWLAD})^T$ نشان می‌دهند. تحت فرضیه $H_0: \beta_2 = 0$ ، مدل (۴.۴) به $y_n = X_n \beta_1 + \varepsilon_n$ تقلیل می‌یابد. بنابراین، برآوردگر محدود شده WLAD، (RWLAD) که با نماد $\hat{\beta}_1^{rWLAD}$ نشان داده می‌شود، کمینه کننده تابع هدف زیر است:

$$\sum_{i=1}^n w_i |y_i - x_i^T \beta_1|.$$

شرایط نظم زیر را در نظر بگیرید:

(۱) خطاهای ε_i مستقل هستند و با تابع توزیع F هم توزیع است. فرض کنید که F در نقطه \circ دو بار مشتق پذیر است و $f(\circ) = F'(\circ) > 0$.

(۲) فرض کنید $W = \text{diag}(w_1, \dots, w_n)$ و همچنین w_i مقدارهای مثبت معلومی هستند که در شرط $\max\{w_i\} = O(1)$ صدق می‌کنند.

(۳) به ازای ماتریس معین مثبت $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ ، $\lim_{n \rightarrow \infty} n^{-1} X_n^T W X_n = C$.

تحت شرایط نظم، وقتی $n \rightarrow \infty$ ،

$$\sqrt{n}(\hat{\beta}^{\text{uWLAD}} - \beta) \xrightarrow{D} \mathcal{N}_p(\mathbf{o}, \tau^2 \mathbf{C}^*),$$

که در آن $\tau = [\mathcal{J}f(\mathbf{o})]^{-1}$ و $\mathbf{C}^* = \mathbf{C}^{-1}(\mathbf{X}^\top \mathbf{W}^\lambda \mathbf{X})\mathbf{C}^{-1}$

بنابراین آماره آزمون فرضیه صفر $\mathcal{H}_0: \beta_\gamma = \mathbf{o}$ به صورت زیر است:

$$\mathcal{L}_n = \frac{(\hat{\beta}_\gamma^{\text{uWLAD}})^\top \mathbf{C}_{\gamma\gamma}^* (\hat{\beta}_\gamma^{\text{uWLAD}})}{\tau^2}, \quad \mathbf{C}_{\gamma\gamma}^* = \mathbf{C}_{\gamma\gamma}^* - \mathbf{C}_{\gamma 1}^* \mathbf{C}_{11}^{*-1} \mathbf{C}_{1\gamma}^*.$$

تحت فرضیه صفر \mathcal{H}_0 و شرایط نظم، \mathcal{L}_n دارای توزیع مجانبی $\chi_{p_\gamma}^2$ است. وقتی $\mathcal{H}_0: \beta_\gamma = \mathbf{o}$ برقرار باشد، برآوردگر RWLAD واریانس مجانبی کمتری نسبت به برآوردگر UWLAD دارد. با این وجود، به ازای $\beta_\gamma \neq \mathbf{o}$ ، برآوردگر RWLAD ممکن است اریب و در بعضی موارد، ناسازگار باشد. به این دلیل، برآوردگر آزمون اولیه WLAD (PTWLAD) به صورت زیر تعریف می شود:

$$\hat{\beta}_1^{\text{ptWLAD}} = \hat{\beta}_1^{\text{uWLAD}} - (\hat{\beta}_1^{\text{uWLAD}} - \hat{\beta}_1^{\text{rWLAD}})I(\mathcal{L}_n \leq \chi_{p_\gamma, \alpha}^2), \quad p_\gamma \geq 1,$$

برآوردگر PTWLAD یک برآوردگر دوضابطه ای است و بر اساس نتیجه آزمون که شدیداً به مقدار α وابسته است، فقط دو برآوردگر محدود نشده یا محدود شده WLAD را می پذیرد. برای غلبه بر این مسئله، برآوردگر انقباضی WLAD که نوع همواری از برآوردگر آزمون اولیه WLAD است، تعریف می شود.

برآوردگر انقباضی نوع استاین WLAD (SWLAD) پارامتر β_1 را با نماد $\hat{\beta}_1^{\text{sWLAD}}$ نشان می دهند و به صورت زیر تعریف می شود:

$$\begin{aligned} \hat{\beta}_1^{\text{sWLAD}} &= \hat{\beta}_1^{\text{uWLAD}} - c\mathcal{L}_n^{-1}(\hat{\beta}_1^{\text{uWLAD}} - \hat{\beta}_1^{\text{rWLAD}}) \\ &= \hat{\beta}_1^{\text{rWLAD}} + (1 - c\mathcal{L}_n^{-1})(\hat{\beta}_1^{\text{uWLAD}} - \hat{\beta}_1^{\text{rWLAD}}), \quad p_\gamma \geq 3, \end{aligned}$$

که $c = p_\gamma - 2$ ثابت انقباضی است.

برآوردگر جزء مثبت انقباضی نوع استاین WLAD (PRSWLAD) برآوردگر SWLAD را بهبود می بخشد و به صورت زیر تعریف می شود:

$$\begin{aligned} \hat{\beta}_1^{\text{sWLAD}+} &= \hat{\beta}_1^{\text{rWLAD}} + (1 - c\mathcal{L}_n^{-1})^+(\hat{\beta}_1^{\text{uWLAD}} - \hat{\beta}_1^{\text{rWLAD}}) \\ &= \hat{\beta}_1^{\text{sWLAD}} - (1 - c\mathcal{L}_n^{-1})(\hat{\beta}_1^{\text{uWLAD}} - \hat{\beta}_1^{\text{rWLAD}})I(\mathcal{L}_n \leq c). \end{aligned}$$

اگر به ویژگی‌های مشترک بین برآوردگرهای انقباضی WLAD و WLAD-LASSO توجه شود، ملاحظه می‌شود که نتیجه برآوردگر WLAD-LASSO از یک دیدگاه مشابه روش‌های انقباضی است، آن‌ها هر دو برآوردگر WLAD را به سمت برآوردگر صفر تحت خطاهای دم‌پهن منقبض می‌کند. اما از جهت دیگر، آن‌ها در موارد زیر متفاوت هستند:

(۱) برآوردگر WLAD-LASSO برآورد و انتخاب متغیر را به‌طور خودکار با رفتار کردن همه متغیرها به صورت یکسان انجام می‌دهد و به مقدار پارامتر تنظیم‌کننده بستگی دارد. زیرمجموعه‌ای از پیشگوها را انتخاب می‌کند.

(۲) روش برآوردگر WLAD انقباضی یک زیرمجموعه بر اساس اطلاعات اضافی مانند تجربه‌های قبلی یا ماهیت متغیرها انتخاب می‌کند.

۱.۳.۴ ویژگی برآوردگرها

در حالت کلی، تحت شرایط نظم، برای فرضیه مقابل ثابت، $\beta_2 = \xi$ که $\xi \neq 0$ ، آماره آزمون \mathcal{L}_n در احتمال به بی‌نهایت میل می‌کند. بنابراین، برای فرضیه ثابت $\xi \neq 0$ با افزایش n ، $\hat{\beta}_1^{\text{ptWLAD}}$ ، $\hat{\beta}_1^{\text{sWLAD}}$ و $\hat{\beta}_1^{\text{sWLAD+}}$ در احتمال با $\hat{\beta}_1^{\text{uWLAD}}$ برابر هستند. با این وجود، وقتی β_2 در یک همسایگی انقباضی از 0 قرار داشته باشند، می‌توان دنباله‌ای از فرضیه مقابل موضعی $\mathcal{K}_{(n)}$ را به صورت زیر در نظر بگیرید:

$$\mathcal{K}_{(n)} : \beta_2 = n^{-1/2} \xi, \quad \xi = (\xi_1, \dots, \xi_{p_2})^\top \in \mathbb{R}^{p_2}. \quad (20.4)$$

ADQB یک برآوردگر دلخواه $\hat{\beta}_1^*$ از پارامتر β_1 به صورت زیر تعریف می‌شود:

$$\text{ADQB}(\hat{\beta}_1^*) = \tau^{-2} \left(\text{ADB}(\hat{\beta}_1^*) \right)^\top C_{11,2}^* \left(\text{ADB}(\hat{\beta}_1^*) \right). \quad (21.4)$$

برای به دست آوردن ویژگی‌های جانبی برآوردگرها قضیه زیر بیان می‌شود.

قضیه ۱.۳.۴. (کوئوکر و باست، ۱۹۸۲) تحت مجموعه فرضیه‌های مقابل موضعی $\mathcal{K}_{(n)}$ و شرایط نظم،

$$\hat{\beta}_1^{\text{rWLAD}} = \hat{\beta}_1^{\text{uWLAD}} + C_{11}^{*-1} C_{12}^* \hat{\beta}_2^{\text{uWLAD}} + o_P(n^{-1/2}).$$

و

$$\begin{bmatrix} \sqrt{n}(\hat{\beta}_2^{\text{uWLAD}} - \beta_2) \\ \sqrt{n}(\hat{\beta}_1^{\text{rWLAD}} - \beta_1^{\text{rWLAD}}) \end{bmatrix} \xrightarrow{D} \mathcal{N}_p \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^2 C_{22,1}^{*-1} & 0 \\ 0 & \tau^2 \Sigma^{**} \end{bmatrix} \right),$$

که در آن

$$\Sigma^{**} = \Sigma_{12} = \Sigma_{21}^\top = C_{11,2}^{*-1} - C_{11}^{*-1} C_{12}^* C_{22,1}^{*-1} C_{21}^* C_{11}^{*-1}.$$

بنابراین،

$$\begin{bmatrix} \sqrt{n}(\hat{\beta}_1^{uWLAD} - \beta_1) \\ \sqrt{n}(\hat{\beta}_1^{uWLAD} - \hat{\beta}_1^{rWLAD}) \\ \sqrt{n}(\hat{\beta}_1^{rWLAD} - \beta_1) \end{bmatrix} \xrightarrow{D} \mathcal{N}_{3p_1} \left(\begin{bmatrix} 0 \\ \gamma \\ -\gamma \end{bmatrix}, \tau^2 \begin{bmatrix} C_{11,2}^{*-1} & \Omega_{12} & \Sigma_{12} \\ \Omega_{21} & \Omega^{**} & 0 \\ \Sigma_{12} & 0 & \Sigma^{**} \end{bmatrix} \right).$$

که در آن

$$\begin{aligned} \gamma &= -C_{11}^{*-1} C_{12}^* \xi \\ \Omega^{**} &= \Omega_{12} = \Omega_{21}^\top = C_{11}^{*-1} C_{12}^* C_{22,1}^{*-1} C_{21}^* C_{11}^{*-1} \end{aligned}$$

برای سادگی محاسبات، فرض کنید $g(\mathcal{L}_n)$ تابعی از \mathcal{L}_n باشد، آن گاه شکل کلی برآوردگرها به صورت

$$\hat{\beta}_1^{*WLAD} = \hat{\beta}_1^{uWLAD} - g(\mathcal{L}_n) (\hat{\beta}_1^{uWLAD} - \hat{\beta}_1^{uWLAD}),$$

است.

$$(1) \text{ اگر } g(\mathcal{L}_n) = 0, \text{ آن گاه } \hat{\beta}_1^{*WLAD} = \hat{\beta}_1^{uWLAD}$$

$$(2) \text{ اگر } g(\mathcal{L}_n) = 1, \text{ آن گاه } \hat{\beta}_1^{*WLAD} = \hat{\beta}_1^{uWLAD}$$

$$(3) \text{ اگر } g(\mathcal{L}_n) = I(\mathcal{L}_n \leq \chi_{p_r, \alpha}^2), \text{ آن گاه } \hat{\beta}_1^{*WLAD} = \hat{\beta}_1^{ptWLAD}$$

$$(4) \text{ اگر } g(\mathcal{L}_n) = c\mathcal{L}_n^{-1}, \text{ آن گاه } \hat{\beta}_1^{*WLAD} = \hat{\beta}_1^{sWLAD}$$

$$(5) \text{ اگر } g(\mathcal{L}_n) = c\mathcal{L}_n^{-1} + (1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \leq c), \text{ آن گاه } \hat{\beta}_1^{*WLAD} = \hat{\beta}_1^{prWLAD}$$

قضیه ۲.۳.۴. تحت فرض های قضیه ۱.۳.۴، داریم

$$\begin{aligned} \text{ADB}(\hat{\beta}_1^{uWLAD}) &= 0 \\ \text{ADB}(\hat{\beta}_1^{rWLAD}) &= -C_{11}^{*-1} C_{12}^* \xi = \gamma \\ \text{ADB}(\hat{\beta}_1^{ptWLAD}) &= H_{p_r+2}(\chi_{p_r, \alpha}^2, \Delta^2) \gamma \\ \text{ADB}(\hat{\beta}_1^{sWLAD}) &= (p_r - 2) \gamma E \left[\chi_{p_r+2}^{-2}(\Delta^2) \right], \quad \Delta^2 = \tau^{-2} \xi^\top C_{22,1}^* \xi \\ \text{ADB}(\hat{\beta}_1^{sWLAD+}) &= \text{ADB}(\hat{\beta}_1^{sWLAD}) \\ &\quad + \gamma E \left[\left(1 - (p_r - 2) \chi_{p_r+2}^{-2}(\Delta^2) \right) I \left(\chi_{p_r+2}^2(\Delta^2) < (p_r - 2) \right) \right], \end{aligned}$$

برهان. عبارتهای ADB برآوردگر دلخواه $\hat{\beta}_1^*$ به صورت زیر محاسبه می شود:

$$\begin{aligned} \text{ADB}(\hat{\beta}_1^{*WLAD}) &= \lim_{n \rightarrow \infty} E \left(\sqrt{n} (\hat{\beta}_1^{*WLAD} - \beta_1) \right) \\ &= \lim_{n \rightarrow \infty} E \left(\sqrt{n} (\hat{\beta}_1^{uWLAD} - g(\mathcal{L}_n) (\hat{\beta}_1^{uWLAD} - \hat{\beta}_1^{uWLAD}) - \beta_1) \right) \end{aligned}$$

$$\begin{aligned}
 &= \lim_{n \rightarrow \infty} E \left(\sqrt{n} \left(\hat{\beta}_1^{uWLAD} - \beta_1 \right) \right) \\
 &\quad - \lim_{n \rightarrow \infty} E \left(\sqrt{n} g(\mathcal{L}_n) \left(\hat{\beta}_1^{uWLAD} - \hat{\beta}_1^{uWLAD} \right) \right) \\
 &= ADB(\hat{\beta}_1^{uWLAD}) - \lim_{n \rightarrow \infty} E \left(\sqrt{n} g(\mathcal{L}_n) C_{11}^{*-1} C_{12}^* \hat{\beta}_2^{uWLAD} \right) \\
 &= ADB(\hat{\beta}_1^{uWLAD}) - C_{11}^{*-1} C_{12}^* \lim_{n \rightarrow \infty} E \left(\sqrt{n} g(\mathcal{L}_n) \hat{\beta}_2^{uWLAD} \right) \\
 &= \gamma E \left(g(\chi_{p+2}^2(\Delta^2)) \right).
 \end{aligned}$$

□

خط آخر از قضیه ۶.۲.آ نتیجه می شود.

نتیجه ۱.۳.۴. تحت فرض های قضیه ۱.۳.۴، ADQB برآوردگرها به صورت زیر هستند:

$$\begin{aligned}
 ADQB(\hat{\beta}_1^{uWLAD}) &= 0 \\
 ADQB(\hat{\beta}_1^{rWLAD}) &= \xi^T \Upsilon \xi \\
 ADQB(\hat{\beta}_1^{ptWLAD}) &= \xi^T \Upsilon \xi \left\{ H_{p+2}(\chi_{p+2}^2, \Delta^2) \right\}^2 \\
 ADQB(\hat{\beta}_1^{sWLAD}) &= (p-2)^2 \xi^T \Upsilon \xi \left\{ E \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right\}^2 \\
 ADQB(\hat{\beta}_1^{sWLAD+}) &= \xi^T \Upsilon \xi \times \left\{ (p-2) E \left[\chi_{p+2}^{-2}(\Delta^2) \right] \right. \\
 &\quad \left. + E \left[\left(1 - (p-2) \chi_{p+2}^{-2}(\Delta^2) \right) I \left(\chi_{p+2}^2(\Delta^2) < (p-2) \right) \right] \right\}^2,
 \end{aligned}$$

که در آن $\Upsilon = \tau^{-2} C_{21}^* C_{11}^{*-1} C_{11.2}^* C_{11}^{*-1} C_{12}^*$

به ازای $C_{12}^* = 0$ ، $C_{21}^* C_{11}^{*-1} C_{11.2}^* C_{11}^{*-1} C_{12}^* = 0$ و $C_{11.2}^* = C_{11}^*$ بنابراین همه ADQB به مقدار مشترک صفر برای همه ξ هاست. در نتیجه، فرض می شود که $C_{12}^* \neq 0$. مقدار ADQB برآوردگر $\hat{\beta}_1^{rWLAD}$ یک تابع بی کران از $\xi^T \Upsilon \xi$ می باشد. به منظور بررسی $ADQB(\hat{\beta}_1^{sWLAD+})$ و $ADQB(\hat{\beta}_1^{sWLAD})$ از نتیجه

$$\text{Ch}_{\min}(\tau^2 \Upsilon C_{22.1}^{*-1}) \leq \frac{\tau^2 \xi^T \Upsilon \xi}{\xi^T C_{22.1}^* \xi} \leq \text{Ch}_{\max}(\tau^2 \Upsilon C_{22.1}^{*-1}),$$

استفاده می شود که $\text{Ch}_{\min}(\tau^2 \Upsilon C_{22.1}^{*-1})$ و $\text{Ch}_{\max}(\tau^2 \Upsilon C_{22.1}^{*-1})$ به ترتیب کوچکترین و بیشترین مقدار ویژه $\tau^2 \Upsilon C_{22.1}^{*-1}$ می باشد. بنابراین، $ADQB(\hat{\beta}_1^{sWLAD})$ از صفر در نقطه $\xi^T \Upsilon \xi = 0$ شروع می شود، به یک نقطه افزایش می یابد سپس از آن جایی که تابع $E \left[\chi_{p+2}^{-2}(\Delta^2) \right]$ تابع لگاریتم-محدب نزولی از Δ^2 است. رفتار $\hat{\beta}_1^{sWLAD+}$ مشابه $\hat{\beta}_1^{sWLAD}$ می باشد. منحنی توان دوم اربیبی به ازای تمام مقادیر Δ^2 ، برآوردگر $\hat{\beta}_1^{sWLAD+}$ زیر منحنی $\hat{\beta}_1^{sWLAD}$ است.

قضیه ۳.۳.۴. اگر فرض های قضیه ۱.۳.۴ برقرار باشد، آن گاه ADR برآوردگرها برابر است با:

$$\begin{aligned}
 ADR(\hat{\beta}_1^{uWLAD}) &= \tau^2 \text{tr}(C_{11.2}^{*-1}) \\
 ADR(\hat{\beta}_1^{rWLAD}) &= \tau^2 \text{tr}(C_{11}^{*-1}) + \gamma^T \gamma
 \end{aligned}$$

$$\begin{aligned}
 \text{ADR}(\hat{\beta}_1^{\text{ptWLAD}}) &= \text{ADR}(\hat{\beta}_1^{\text{uWLAD}}) - \tau^2 H_{p_2+2}(\chi_{p_2,\alpha}^2, \Delta^2) \text{tr}(\Omega^{**}) \\
 &\quad + \left(2 H_{p_2+2}(\chi_{p_2,\alpha}^2, \Delta^2) - H_{p_2+4}(\chi_{p_2,\alpha}^2, \Delta^2) \right) \text{tr}(\gamma^T \gamma) \\
 \text{ADR}(\hat{\beta}_1^{\text{sWLAD}}) &= \text{ADR}(\hat{\beta}_1^{\text{uWLAD}}) - \tau^2 \left((p_2 - 2) \text{tr}(\Omega^{**}) \left(2 E \left[\chi_{p_2+2}^{-2}(\Delta^2) \right] \right. \right. \\
 &\quad \left. \left. - (p_2 - 2) E \left[\chi_{p_2+2}^{-4}(\Delta^2) \right] \right) \right) + (p_2 - 4) \gamma^T \gamma E \left[\chi_{p_2+4}^{-4}(\Delta^2) \right] \\
 \text{ADR}(\hat{\beta}_1^{\text{sWLAD}+}) &= \text{ADR}(\hat{\beta}_1^{\text{sWLAD}}) \\
 &\quad + \tau^2 (p_2 - 2) \text{tr}(\mathbf{W} \Omega^{**}) \left(2 E \left[\chi_{p_2+2}^{-2}(\Delta^2) I(\chi_{p_2+2}^2(\Delta^2) \leq p_2 - 2) \right] \right. \\
 &\quad \left. - (p_2 - 2) E \left[\chi_{p_2+2}^{-4}(\Delta^2) I(\chi_{p_2+2}^2(\Delta^2) \leq p_2 - 2) \right] \right) \\
 &\quad - \tau^2 \text{tr}(\Omega^{**}) H_{p_2+2}(p_2 - 2, \Delta^2) \\
 &\quad + \gamma^T \gamma \left(2 H_{p_2+2}(p_2 - 2, \Delta^2) - H_{p_2+4}(p_2 - 2, \Delta^2) \right) \\
 &\quad - (p_2 - 2) \gamma^T \gamma \left(2 E \left[\chi_{p_2+2}^{-2}(\Delta^2) I(\chi_{p_2+2}^2(\Delta^2) \leq p_2 - 2) \right] \right. \\
 &\quad \left. - 2 E \left[\chi_{p_2+2}^{-4}(\Delta^2) I(\chi_{p_2+2}^2(\Delta^2) \leq p_2 - 2) \right] \right) \\
 &\quad + (p_2 - 2) E \left[\chi_{p_2+2}^{-4}(\Delta^2) I(\chi_{p_2+2}^2(\Delta^2) \leq p_2 - 2) \right].
 \end{aligned}$$

برهان. عبارت‌های ADR برای برآوردگرهای پیشنهادی به صورت زیر محاسبه می‌شود.

$$\begin{aligned}
 \text{ADR}(\hat{\beta}_1^{\text{sWLAD}}) &= \lim_{n \rightarrow \infty} E \left(n \left(\hat{\beta}_1^{\text{sWLAD}} - \beta_1 \right)^T \left(\hat{\beta}_1^{\text{sWLAD}} - \beta_1 \right) \right) \\
 &= \lim_{n \rightarrow \infty} E \left(n \left(\hat{\beta}_1^{\text{uWLAD}} - g(\mathcal{L}_n) \hat{\beta}_1^{\text{uWLAD}} - \beta_1 \right)^T \right. \\
 &\quad \left. \left(\hat{\beta}_1^{\text{uWLAD}} - g(\mathcal{L}_n) \hat{\beta}_1^{\text{uWLAD}} - \beta_1 \right) \right) \\
 &= \lim_{n \rightarrow \infty} E \left(n \left(\hat{\beta}_1^{\text{uWLAD}} - \beta_1 \right)^T \left(\hat{\beta}_1^{\text{uWLAD}} - \beta_1 \right) \right) \\
 &\quad - 2 \lim_{n \rightarrow \infty} E \left(n \left(\hat{\beta}_1^{\text{uWLAD}} - \beta_1 \right)^T \hat{\beta}_1^{\text{uWLAD}} g(\mathcal{L}_n) \right) \\
 &\quad + \lim_{n \rightarrow \infty} E \left(n \left(\hat{\beta}_1^{\text{uWLAD}} \right)^T \hat{\beta}_1^{\text{uWLAD}} g(\mathcal{L}_n) \right) \\
 &= \text{ADR}(\hat{\beta}_1^{\text{uWLAD}}) - 2 \lim_{n \rightarrow \infty} E \left(n \left(\hat{\beta}_1^{\text{uWLAD}} \right)^T \hat{\beta}_1^{\text{uWLAD}} g(\mathcal{L}_n) \right) \\
 &\quad + 2 \xi^T \mathbf{W} \lim_{n \rightarrow \infty} E \left(\sqrt{n} \hat{\beta}_1^{\text{uWLAD}} g(\mathcal{L}_n) \right) \\
 &\quad + \lim_{n \rightarrow \infty} E \left(n \left(\hat{\beta}_1^{\text{uWLAD}} \right)^T \hat{\beta}_1^{\text{uWLAD}} g(\mathcal{L}_n) \right),
 \end{aligned}$$

که در آن $\text{ADR}(\hat{\beta}_1^{\text{uWLAD}}) = \tau^2 \text{tr}(C_{11,2}^{*-1})$.

با استفاده از قضیه ۵.۲.۱ و ۶.۲.۱ و عبارت ADR برآوردگر $\hat{\beta}_1^{\text{uWLAD}}$ می‌توان ADR برآوردگر

دلخواه $\hat{\beta}_1^*$ را به صورت

$$\begin{aligned}
 \text{ADR}(\hat{\beta}_1^{\text{sWLAD}}) &= \tau^2 \text{tr}(C_{11,2}^{*-1}) \left\{ 1 - 2 E \left(g(\chi_{p_2+2}^2(\Delta^2)) \right) + E \left(g^2(\chi_{p_2+2}^2(\Delta^2)) \right) \right\} \\
 &\quad + \gamma^T \gamma \left\{ 2 E \left(\chi_{p_2+2}^2(\Delta^2) \right) - 2 E \left(\chi_{p_2+4}^2(\Delta^2) \right) + E \left(g^2(\chi_{p_2+4}^2(\Delta^2)) \right) \right\}.
 \end{aligned}$$

نوشت. اثبات قضیه با جایگذاری هر یک از توابع g در رابطه بالا کامل می‌شود. □

وقتی $C_{12}^* = 0$ ، $C_{21}^* C_{11}^{*-1} C_{11}^{*-1} C_{12}^* = 0$ و بنابراین $C_{11,2}^* = C_{11}^*$ و بنابراین ADR به ازای همه مقادیر ξ ، به مقدار مشترک $\tau^2 \text{tr}(C_{11}^{*-1})$ کاهش پیدا می‌کند. همه برآوردگرها به طور مجانبی مخاطره‌های برابر و کران‌دار هستند. بنابراین، فرض کنید $C_{12}^* \neq 0$.

اگر $C_{1\gamma}^* \neq 0$ و تحت فرضیه H_0 ، اختلاف ADR برای $ADR(\hat{\beta}_1^{uWLAD})$ و $ADR(\hat{\beta}_1^{ptWLAD})$ به صورت $\text{tr}(\Omega^{**}) \chi_{p\gamma+2}^2(\chi_{p\gamma+2}^2, \alpha, \circ)$ می باشد. بنابراین، به دست آوردن مخاطره یک مسئله اساسی است وقتی H_0 صحیح است. وقتی Δ^2 از صفر دور می شود، مخاطره $\hat{\beta}_1^{ptWLAD}$ به طور یکنوا به مخاطره برآوردگر $\hat{\beta}_1^{uWLAD}$ بعد از رسیدن به مقدار بیشینه می رسد. محاسبه مقدار برتری برآوردگر $\hat{\beta}_1^{ptWLAD}$ بر $\hat{\beta}_1^{uWLAD}$ دشوار است.

با مقایسه کردن $ADR(\hat{\beta}_1^{sWLAD+})$ با $ADR(\hat{\beta}_1^{sWLAD})$ ، مشاهده می شود که به ازای همه مقادیر ξ ، برآوردگر $\hat{\beta}_1^{sWLAD+}$ بر برآوردگر $\hat{\beta}_1^{sWLAD}$ برتری دارد و برای بعضی از مقادیر ξ ، نامساوی اکید برقرار است. بنابراین، مخاطره $\hat{\beta}_1^{sWLAD+}$ در کل فضای پارامتر کمتر از مخاطره برآوردگر $\hat{\beta}_1^{uWLAD}$ است و وقتی Δ^2 به سمت بی نهایت میل کند، حد بالایی برای این مخاطره پیدا می شود. وقتی Δ^2 از صفر دور می شود، مقدار ADR برآوردگر $\hat{\beta}_1^{sWLAD+}$ به طور یکنوا به $ADR(\hat{\beta}_1^{uWLAD})$ میل می کند. این نتیجه می دهد که

$$ADR(\hat{\beta}_1^{sWLAD+}) \leq ADR(\hat{\beta}_1^{sWLAD}) \leq ADR(\hat{\beta}_1^{uWLAD}),$$

نامساوی اکید برای برخی از ξ ها برقرار است. بنابراین، برآوردگرهای $\hat{\beta}_1^{sWLAD}$ و $\hat{\beta}_1^{sWLAD+}$ در کل فضای پارامتر که با Δ^2 ایجاد شده است، بهتر از برآوردگر $\hat{\beta}_1^{uWLAD}$ رفتار می کند. در اینجا می توان همانند قضیه های $9.3.2 - 14.3.2$ ، قضیه های مرتبط را بیان و اثبات کرد و برای عدم تکرار، می توان به مقاله **نوروزی راد و همکاران (۲۰۱۸)** اشاره کرد.

۲.۳.۴ مطالعه شبیه سازی

فرض کنید ρ نرخ آلودگی با مقدار $\rho/2$ باشد. مقادیر $10, 15, 20, 25$ را در نظر بگیرید. $n - m$ تا مولفه اول (x_{1i}, y_{1i}) به صورت زیر تولید می شوند:

(۱) ماتریس طرح $(n - m) \times p$ بعدی X_1 را از توزیع نرمال p متغیره تولید کنید.

(۲) فرضیه صفر $H_0: \beta_2 = 0$ را در نظر بگیرید. هدف، برآورد پارامتر β_1 است. مقدار واقعی بردار پارامتر $\beta = (\beta_1^T, \beta_2^T)^T$ را برابر با $\beta_1 = (1/8, -2/2, 1/3)^T$ و $\beta_2 = (1/8, -2/2, 1/3, 1, 1, 1)^T$ و هم چنین $\beta_2 = 0$ در نظر بگیرید که 0 بردار $1 \times p_2$ از صفرهاست. نمونه های بیشتری برای مقادیر متفاوت Δ^2 که در بازه بین 0 و 2 در نظر گرفته شده است، تولید می شود.

(۳) خطای ε را از توزیع کوشی (توزیع t با 1 درجه آزادی) تولید کنید. با این کار داده های پرت در جهت y ایجاد می شود.

(۴) متغیر پاسخ را از مدل $y_{1i} = x_{1i}^T \beta + \varepsilon_i$ به ازای $i = 1, \dots, n - m$ تولید کنید.

مرحله بعدی، m داده آلوده شده، $(x_{\gamma_i}, y_{\gamma_i})$ را به صورت زیر تولید می کند.

(۱) بردار متغیرهای توضیحی x_{γ_i} را از توزیع $\mathcal{N}_p(\delta, \gamma \circ I_{p \times p})$ تولید کنید.

(۲) به ازای $y_{\gamma_i}, i = 1, \dots, m$ از مدل $y_{\gamma_i} = x_{\gamma_i}^T \beta_{\gamma} + \varepsilon_i$ تولید می شود که در آن توزیع ε_i مشابه مرحله (۳) است.

دو مجموعه داده بالا با هم ترکیب می شود تا مجموعه داده ای برای $i = 1, \dots, n$ ، به صورت (x_i, y_i) به دست آید. برای هر حالت، فرآیند شبیه سازی 5000 بار تکرار می شود. هدف، در اینجا بررسی رفتار برآوردگرهای انقباضی برای $\Delta^2 \geq 0$ و برآوردگر جریمه شده برای $\Delta^2 = 0$ است زیرا این برآوردگر نمی تواند پارامترهای مدل را برای $\Delta^2 > 0$ به درستی برآورد کند (احمد و رحیم، ۲۰۱۲). به منظور بررسی رفتار برآوردگرهای پیشنهادی، فاصله بین مدل شبیه سازی شده و مدل تحت فرضیه \mathcal{H}_0 به صورت $\Delta^2 = \|\beta - \beta^{(0)}\|_2^2$ تعریف می شود که $\beta^{(0)} = (\beta_{\gamma}^T, 0^T)^T$. معیار مقایسه رفتار مخاطره برآوردگر دلخواه $\hat{\beta}_{\gamma}^*$ در این مطالعه، MSE است. کارایی برآوردگرهای پیشنهادی از نسبت MSE برآوردگر $\hat{\beta}_{\gamma}^*$ به MSE برآوردگر $\hat{\beta}_{\gamma}^{uWLAD}$ به دست می آید. به عبارت دیگر،

$$RMSE(\hat{\beta}_{\gamma}^*) = \frac{MSE(\hat{\beta}_{\gamma}^{uWLAD})}{MSE(\hat{\beta}_{\gamma}^*)}$$

که در آن $\hat{\beta}_{\gamma}^*$ می تواند هر یک از برآوردگرهای PRSWLAD، SWLAD، PTWLAD، RWLAD و WLAD-LASSO باشد. مشاهده می شود که مقدار RMSE بزرگتر از ۱، درجه برتری برآوردگر دلخواه $\hat{\beta}_{\gamma}^*$ را نسبت به برآوردگر $\hat{\beta}_{\gamma}^{uWLAD}$ نشان می دهد.

در جدول ۵.۴ مقدارهای RMSE برآوردگرهای PRSWLAD، SWLAD، PTWLAD، RWLAD و WLAD-LASSO نسبت به برآوردگر UWLAD به ازای $\Delta^2 = 0$ برای حجم های نمونه $n = 50, 75, 100, 150$ و $p_1 = 3, 6$ گزارش شده است. نتایج برای خطاهای با توزیع کوشی و ۲۰٪ نقاط اهرمی ($\rho = 0.2$) به ازای $\Delta^2 > 0$ به صورت زیر خلاصه می شود:

(۱) بیشینه RMSE تمامی برآوردگرها نسبت به برآوردگر UWLAD به ازای $\Delta^2 = 0$ اتفاق می افتد. از جدول ۵.۴ نتیجه می شود که RWLAD بر سایر برآوردگرها در حالت $\Delta^2 = 0$ برتری دارد زیرا این برآوردگر در این حالت، ناریب است. به عبارت دیگر، وقتی Δ^2 افزایش پیدا می کند، مخاطره برآوردگر RWLAD نیز بیشتر و بی کران می شود. در نتیجه، RMSE برآوردگر RWLAD کاهش پیدا کرده و به سمت صفر میل می کند که این نتیجه در جدول ۶.۴ مشهود است.

جدول ۵.۴: مقادیر RMSE برآوردگرها نسبت به برآوردگر UWLAD به ازای $p_1 = 3$ و $\Delta^2 = 0$

WLAB-LASSO	PRSWLAD	SWLAD	PTWLAD	RWLAD	p_2	n	p_1
۱/۱۳	۱/۲۹	۱/۲۰	۱/۲۳	۱/۸۸	۵		
۱/۳۸	۱/۳۲	۱/۲۹	۱/۱۲	۱/۹۴	۱۰	۵۰	
۲/۱۶	۱/۶۶	۱/۶۰	۱/۲۲	۲/۹۷	۱۵		
۱/۲۰	۱/۲۹	۱/۰۴	۱/۲۹	۱/۵۱	۵		
۱/۲۵	۱/۳۱	۱/۲۵	۱/۱۶	۱/۵۵	۱۰	۷۵	
۱/۷۱	۱/۵۶	۱/۵۱	۱/۲۱	۲/۰۸	۱۵		
۱/۱۸	۱/۳۰	۱/۱۱	۱/۳۲	۱/۴۲	۵		۳
۱/۲۴	۱/۳۷	۱/۱۸	۱/۳۰	۱/۴۶	۱۰	۱۰۰	
۱/۶۱	۱/۵۱	۱/۴۳	۱/۲۷	۱/۸۰	۱۵		
۱/۱۳	۱/۲۵	۱/۰۰	۱/۲۷	۱/۳۱	۵		
۱/۲۷	۱/۳۷	۱/۱۴	۱/۳۷	۱/۴۰	۱۰	۱۵۰	
۱/۵۲	۱/۵۲	۱/۴۱	۱/۳۶	۱/۶۸	۱۵		
۰/۸۹	۱/۲۲	۱/۲۱	۱/۰۶	۱/۸۶	۹		
۱/۷۲	۱/۵۶	۱/۵۳	۱/۱۴	۳/۶۹	۱۴	۵۰	
۱/۵۱	۱/۲۹	۱/۲۹	۱/۱۸	۶/۴۱	۱۹		
۱/۰۲	۱/۴۲	۱/۱۹	۱/۲۷	۱/۷۰	۹		
۱/۲۱	۱/۵۳	۱/۴۹	۱/۱۷	۲/۱۸	۱۴	۷۵	
۱/۸۱	۱/۶۴	۱/۶۰	۱/۱۸	۲/۵۷	۱۹		
۱/۰۱	۱/۲۸	۱/۱۵	۱/۲۱	۱/۳۶	۹		۶
۱/۲۶	۱/۵۰	۱/۴۲	۱/۲۴	۱/۷۳	۱۴	۱۰۰	
۱/۵۲	۱/۷۳	۱/۵۲	۱/۴۶	۱/۹۶	۱۹		
۱/۰۶	۱/۲۶	۱/۰۳	۱/۲۴	۱/۲۹	۹		
۱/۱۶	۱/۴۲	۱/۳۱	۱/۳۰	۱/۵۱	۱۴	۱۵۰	
۱/۲۷	۱/۴۵	۱/۳۹	۱/۲۴	۱/۵۸	۱۹		

جدول ۶.۴: مقادیر RMSE برآوردگرهای پیشنهادی نسبت به برآوردگر UWLAD به ازای $\Delta^2 \geq 0$

$p_2 = 15$ و $p_1 = 3$

PRSWLAD	SWLAD	PTWLAD	RWLAD	Δ^2
۱/۵۱	۱/۴۳	۱/۲۶	۱/۸۰	۰
۱/۴۲	۱/۳۶	۱/۱۷	۱/۶۲	۰/۰۳
۱/۳۷	۱/۳۱	۱/۱۲	۱/۵۴	۰/۰۷
۱/۳۳	۱/۳۰	۱/۰۸	۱/۴۲	۰/۱
۱/۲۷	۱/۲۴	۱/۰۲	۱/۲۸	۰/۱۵
۱/۲۳	۱/۲۲	۱/۰۰	۱/۱۹	۰/۲
۱/۱۸	۱/۱۸	۰/۹۶	۱/۰۳	۰/۳
۱/۱۵	۱/۱۵	۰/۹۵	۰/۹۳	۰/۴
۱/۱۱	۱/۱۱	۰/۹۸	۰/۷۲	۰/۷
۱/۰۷	۱/۰۷	۰/۹۹	۰/۵۷	۱/۰
۱/۰۴	۱/۰۴	۱/۰۰	۰/۳۵	۲/۰

(۲) ملاحظه می‌شود که برآوردگرهای SWLAD و PRSWLAD برتری خودشان را بر برآوردگر UWLAD و در بازه بزرگی از Δ^2 بر برآوردگر PTWLAD حفظ می‌کنند. به خصوص، وقتی $n = 75$ و $p_1 = 3$ ، کارایی برآوردگرهای SWLAD و PRSWLAD بهتر از برآوردگر PTWLAD

برای $\Delta^2 \geq 0.7\%$ است. برای مقادیر بزرگتر p_2 ، کارایی برآوردگر SWLAD پایدار و بهتر از برآوردگر PTWLAD است.

(۳) برآوردگر WLAD-LASSO هنگامی که n و p_2 هر دو کوچک است، با برآوردگرهای دیگر قابل مقایسه است. برآوردگر WLAD-LASSO در مقایسه با برآوردگرهای SWLAD و PRSWLAD زمانی که n و p_2 افزایش پیدا می‌کند، بدتر رفتار می‌کند. اگر پیچیدگی مدل در نظر گرفته شود، برآوردگر WLAD-LASSO در مقایسه با برآوردگر SWLAD برای مدل‌های تنک، انتخاب بهتری نیست.

(۴) برآوردگر PTWLAD وقتی فرضیه صفر تقریباً درست باشد، یعنی برای مقادیر کوچک Δ^2 ، بر برآوردگرهای SWLAD و PRSWLAD برتری دارد. برای مقادیر بزرگتر Δ^2 ، مقدار RMSE این برآوردگر به سرعت کاهش می‌یابد و زیر خط یک قرار می‌گیرد. مشاهده می‌شود که RMSE برآوردگر PTWLAD سریع‌تر از برآوردگرهای SWLAD و PRSWLAD به یک همگراست. هنگامی که p_2 افزایش پیدا می‌کند، برتری برآوردگر PTWLAD کاهش می‌یابد. این موضوع تایید می‌کند در حالی که اطلاعات زیرفضای نادقیقی وجود دارد (به عبارت دیگر، $\Delta^2 > 0$)، برآوردگر PTWLAD گاهی اوقات برتری خود را حفظ می‌کند.

(۵) همان‌طور که انتظار می‌رفت، برآوردگر PRSWLAD بهتر از برآوردگر SWLAD رفتار می‌کند.

جدول ۷.۴ نتایج شبیه‌سازی را به‌ازای $p_2 \geq 30$ نشان می‌دهد. وقتی داده‌ها با متغیرهای توضیحی زیاد، نقاط اهرمی دارند، برآوردگر WLAD-LASSO بدترین کارایی را در مقایسه با برآوردگرهای SWLAD و PRSWLAD دارد. این نتیجه یکی از مزیت‌های استفاده از برآوردگرهای انقباضی پیشنهادی را نشان می‌دهد که برای بعد بالا نیز کارا هستند.

جدول ۷.۴: مقادیر RMSE برآوردگرهای پیشنهادی نسبت به برآوردگر UWLAD به‌ازای $\Delta^2 = 0$ و $p_1 = 6$.

WLAD-LASSO	PRSWLAD	SWLAD	PTWLAD	RWLAD	p_2
۱/۷۶	۱/۶۳	۱/۶۲	۱/۱۳	۲/۱۴	۳۰
۳/۰۹	۲/۳۱	۲/۲۶	۱/۲۸	۳/۶۹	۴۰
۳/۹۶	۱/۴۱	۱/۴۱	۱/۰۰	۵/۰۴	۶۰
۱۰/۲۷	۱/۱۹	۱/۱۹	۱/۰۰	۱۵/۵۱	۸۰

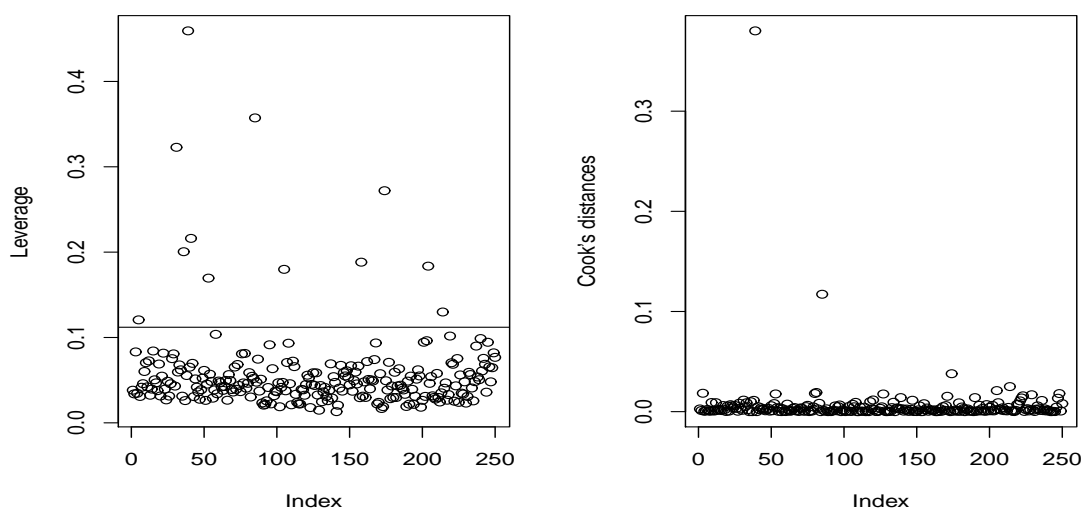
به‌طور خلاصه، برآوردگرهای SWLAD و PRSWLAD بر برآوردگر UWLAD در کل فضای پارامتر برتری دارند.

۳.۳.۴ تحلیل مجدد داده‌های چربی بدن

مجموعه داده‌های چربی بدن در صفحه ۶۷ را در نظر بگیرید.

از آن جایی که این مثال خیلی جلوتر بررسی شده بود، در این جا جزئیات بیشتری آورده می‌شود. دو مورد از مشاهدات (مشاهده‌های ۴۲ و ۱۸۳) حذف شده‌اند. قد در مشاهده ۴۲ به‌طور عجیبی ۲۹/۵ اینچ و چربی بدن در مشاهده ۱۸۳، صفر ثبت شده است. همچنین فرض کنید β_j به‌ازای $j = 1, \dots, 13$ ، ضریب متناظر با متغیر پیشگو باشد. تجزیه و تحلیل با برازش یک مدل رگرسیون خطی به مجموعه داده‌ها آغاز می‌شود. ابتدا با استفاده از برآوردگر OLS مدل برازش شده است. برای این مدل، از یک بررسی تشخیصی (مانند فاصله کوک و نقاط اهرمی) بهره گرفته شده تا نشان دهد مشاهده‌های ۳۹ و ۸۵ مشاهده‌های پرت هستند زیرا فاصله‌های کوک آن‌ها مقدار زیادی را نشان می‌دهد و مشاهده‌های ۳۱، ۳۶، ۳۹، ۴۱، ۵۳، ۸۵، ۱۰۵، ۱۵۸، ۱۷۴ و ۲۰۴ را به‌عنوان نقاط اهرمی تشخیص می‌دهد (شکل ۴.۴ را ببینید). بنابراین، اعتبار استفاده از برآوردگر OLS و روش‌های معمولی انتخاب مدل برای این داده‌ها چالش برانگیز است در حالی که روش‌های انتخاب مدل استوار (مانند برآوردگر LAD موزون) دقیق‌تر است. برآوردگر LAD-LASSO نسبت به داده‌های پرت در متغیر پاسخ استوار است، اما در برابر نقاط اهرمی حساس نیست، بنابراین از روش WLAD-LASSO برای انتخاب متغیرهای مناسب استفاده می‌شود. به‌کار بردن این روش، نشان می‌دهد که متغیرهای chest، neck، height، age و abdomen مهم‌ترین متغیرها برای درصد چربی بدن هستند. بنابراین، زیر فضای محدود شده برابر است با

$$\beta_{\mathcal{P}} = (\beta_{\mathcal{P}_1}, \beta_{\mathcal{P}_2}, \beta_{\mathcal{P}_3}, \beta_{\mathcal{P}_4}, \beta_{\mathcal{P}_5}, \beta_{\mathcal{P}_6}, \beta_{\mathcal{P}_7}, \beta_{\mathcal{P}_8}, \beta_{\mathcal{P}_9})^T = (0, 0, 0, 0, 0, 0, 0, 0, 0)^T.$$



شکل ۴.۴: فاصله کوک و نقاط اهرمی برای داده‌های چربی بدن

از روش باز نمونه‌گیری بوت‌استرپ برای محاسبه برآوردها، خطاهای استاندارد و کارایی‌های نسبی برآوردگرهای پیشنهاد شده استفاده شده است.

جدول ۸.۴: برآوردها (خطای استاندارد) برای ضرایب و RMSE بر اساس شبیه‌سازی بوت‌استرپ نسبت به برآوردگر UWLAD

برآوردگرها	UWLAD	RWLAD	PTWLAD	SWLAD	PRSWLAD	WLAD-LASSO
β_1	۰/۰۹۵ (۰/۰۹۲)	۰/۰۵۹ (۰/۰۵۶)	۰/۰۶۶ (۰/۰۶۹)	۰/۰۶۸ (۰/۰۷۳)	۰/۰۷۲ (۰/۰۶۳)	۰/۰۶۵ (۰/۰۷۴)
β_3	-۰/۵۰۳ (۰/۴۵۵)	-۰/۶۴۶ (۰/۲۱۸)	-۰/۶۱۶ (۰/۳۱۶)	-۰/۶۲۷ (۰/۳۴۵)	-۰/۶۰۰ (۰/۲۹۵)	-۰/۴۸۰ (۰/۳۷۱)
β_4	-۰/۵۹۷ (۰/۸۳۷)	-۰/۵۲۳ (۰/۵۹۳)	-۰/۵۴۲ (۰/۶۴۱)	-۰/۵۷۹ (۰/۶۹۴)	-۰/۵۶۸ (۰/۶۲۳)	-۰/۴۱۳ (۰/۶۴۵)
β_5	۰/۰۶۰ (۰/۲۹۷)	۰/۰۵۶ (۰/۲۶۲)	۰/۰۵۰ (۰/۲۷۴)	۰/۰۴۱ (۰/۲۸۸)	۰/۰۵۱ (۰/۲۵۹)	۰/۰۳۵ (۰/۲۰۵)
β_6	۰/۸۸۸ (۰/۲۶۹)	۰/۸۱۴ (۰/۲۰۵)	۰/۸۲۶ (۰/۲۲۵)	۰/۸۲۱ (۰/۲۴۷)	۰/۸۳۵ (۰/۲۱۳)	۰/۸۸۸ (۰/۲۲۵)
RMSE	۱/۰۰	۲/۰۰	۱/۶۴	۱/۴۱	۱/۷۷	۱/۵۹

فرض کنید که $D = \{(y_i, \mathbf{x}_i); i = 1, \dots, 250\}$ مجموعه داده‌ها باشد. تعداد $B = 1000$ نمونه بوت‌استرپ مستقل با جایگذاری $(y_i^{*b}, \mathbf{x}_i^{*b})$ ، به ازای $b = 1, \dots, B$ از مجموعه داده‌ها شامل ۲۵۰ مورد انتخاب می‌شود و برآوردها، خطاهای استاندارد و RMSE متناظر با هر نمونه بوت‌استرپ محاسبه می‌شود. برآوردهای UWLAD، RWLAD، PTWLAD، SWLAD، PRSWLAD و برآوردهای WLAD-LASSO بر اساس روش‌های بیان شده به دست می‌آیند. نتایج ارائه شده در جدول ۸.۴ با نتایج شبیه‌سازی (جدول‌های ۵.۴-۷.۴) همخوانی دارد. برآوردگر جزء مثبت انقباضی نوع استاین LAD بهتر از برآوردگر WLAD-LASSO رفتار می‌کند. برآوردگر RWLAD بر همه برآوردها برتری دارد زیرا متغیرهای توضیحی بی‌اهمیت در مدل کامل، واقعا نامرتب هستند.

۴.۴ نتیجه‌گیری

در رگرسیون استوار، وانگ و همکاران (۲۰۰۷) رگرسیونی را معرفی کردند که مشابه لاسو، همزمان انقباض ضرایب و انتخاب متغیر انجام می‌دهد. در این فصل، اطلاعات به دست آمده از این مدل را با مدل کامل رگرسیونی ترکیب کرده تا برآوردگر جدیدی به نام برآوردگر جزء مثبت انقباضی نوع استاین LAD به دست آید. مطالعه شبیه‌سازی نشان داد که کارایی برآوردگر پیشنهادی بیشتر از سایر برآوردها در حالت مجانبی است. خطای پیش‌گویی برآوردگر پیشنهادی در تحلیل یک مثال واقعی کمتر از برآوردهای LAD و LAD-LASSO بود.

در حالتی که در داده‌ها نقاط اهرمی وجود دارد، برآوردهای جزء مثبت انقباضی نوع استاین PRSWLAD، WLAD پیشنهاد شده‌اند که بهتر از برآوردگر انقباضی نوع استاین WLAD رفتار می‌کند. برآوردگر آزمون اولیه WLAD در نزدیکی فرضیه صفر، $H_0: \beta_2 = 0$ ، بهتر از

برآوردگرهای SWLAD و PRSWLAD رفتار می کند. هر دو برآوردگر SWLAD و PTWLAD بهتر از برآوردگر UWLAD رفتار می کند به ویژه زمانی که تعداد متغیرهای توضیحی بی معنی در مدل زیاد باشد. به عبارت دیگر، برآوردگر WLAD-LASSO کارایی بیشتری از برآوردگرهای SWLAD و PRSWLAD دارد وقتی تعداد متغیرهای توضیحی بی معنی کم باشد.

فصل ۵

پیشنهادات برای پژوهش‌های آینده

این فصل، به مرور مقاله‌های **آرشی و نوروزی‌راد (۲۰۱۶)**، **نوروزی‌راد و آرشی (۲۰۱۶)** و **نوروزی‌راد و همکاران (۲۰۱۵a)** که هنوز به صورت یک ایده اولیه هستند، اختصاص دارد.

۱.۵ مقدمه

در این رساله، برآوردگرهای انقباضی در مدل‌های مختلف رگرسیون خطی با پذیره‌های مختلف بررسی شده است اما هنوز مسائل متعددی وجود دارند که می‌توان از تکنیک انقباضی در آنها استفاده نمود که منجر به برآوردگرهای بهتر بر حسب مخاطره توان دوم می‌گردد. از دیدگاه‌های دیگری می‌توان به مدل‌های رگرسیونی که در این رساله مورد مطالعه قرار گرفته است، نگریست.

در فصل ۳ برآوردگرهای انقباضی لاسو معرفی شدند. اگر به این مسئله از دیدگاه بهینه‌سازی نگاه کرد، آن‌گاه رهیافت دیگری برای دستیابی به برآوردگرهای انقباضی مطرح می‌گردد. این ایده اولیه در مقاله **نوروزی‌راد و همکاران (۲۰۱۵a)** ارائه شد.

در فصل ۴، برآوردگرهای انقباضی LAD با در نظر گرفتن اطلاعات فرعی به دست آمده از برآوردگر LAD-LASSO معرفی شدند. یافتن آماره آزمون همواره دشواری‌هایی دارد. می‌توان با استفاده از روش‌های عددی، آماره آزمون مناسب را پیدا و بر اساس آن، برآوردگرهای انقباضی را مطرح نمود. این ایده، در مقاله **نوروزی‌راد و آرشی (۲۰۱۶)** ارائه شده است.

اخیرا با پیشرفت تکنولوژی، امکان ذخیره‌سازی داده‌هایی با بعد بالا وجود دارد. برآوردهای انقباضی در مدل رگرسیونی بعد بالا از جذابیت بسیاری برخوردار است. در مقاله‌ای که در سیزدهمین کنفرانس آمار ایران ارائه شده است، این برآوردها مطرح شده‌اند. در این مقاله از برآوردهای ریج به‌عنوان برآوردهای محدود نشده و برآوردهای لاسو به‌عنوان برآوردهای محدود شده استفاده شده است و برآوردهای بهتری با ترکیب این دو برآوردها به دست آمده است.

۲.۵ برآوردهای انقباضی لاسو از دیدگاه بهینه‌سازی

می‌توان به روش دیگری برآوردهای لاسو را در حضور برخی محدودیت‌های خطی فرمول‌بندی کرد. در این راستا، می‌توان به صورت زیر نوشت:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 \right\} \quad \text{s.t.} \quad \begin{cases} \sum_{j=1}^p |\beta_j| \leq t \\ \mathbf{H}\beta = \mathbf{h} \end{cases}$$

که معادل با مسئله هموارسازی محدود شده زیر است

$$\min_{\beta} f(\beta) \quad \text{s.t.} \quad \mathbf{H}\beta = \mathbf{h} \quad (1.5)$$

که در آن $f(\beta) = \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1$. پاسخ رابطه بالا، برآوردهای محدود شده لاسو نامیده می‌شود.

برای ماتریس دلخواه \mathbf{Z} ، فرض کنید \mathbf{Z}_A ، m ستون از ماتریس \mathbf{Z} متناظر با مجموعه اندیس‌گذار A با اندازه m را نشان دهد. به همین ترتیب $\mathbf{Z}_{\bar{A}}$ متناظر با ستون‌های باقیمانده ماتریس \mathbf{Z} همخوانی دارد. برای کاهش دادن نمادهای مورد استفاده، بدون از دست دادن کلیت مسئله فرض کنید که A متناظر با اولین m ستون از ماتریس \mathbf{Z} است به طوری که $\mathbf{Z} = [\mathbf{Z}_A, \mathbf{Z}_{\bar{A}}]$.

لم ۱.۲.۵. (جیمز و همکاران، ۲۰۱۲) برای هر مجموعه اندیس‌گذار A ، بردار β_A را به صورت زیر تعریف کنید

$$\beta_A = \operatorname{argmin}_{\beta} \|\mathbf{y}^* - \mathbf{X}^* \beta\|_2 + \lambda \|\beta\|_1 + \lambda \|\mathbf{H}_A^{-1}(\mathbf{h} - \mathbf{H}_{\bar{A}} \beta_A)\|_1.$$

که $\mathbf{y}^* = \mathbf{y} - \mathbf{X}_A \mathbf{H}_A^{-1} \beta$ و $\mathbf{X}^* = \mathbf{X}_{\bar{A}} - \mathbf{X}_A \mathbf{H}_A^{-1} \mathbf{H}_{\bar{A}}$. آن‌گاه، برای هر مجموعه اندیس‌گذار دلخواه A ، به شرط اینکه C_A معکوس‌پذیر باشد، پاسخ مسئله (۱.۵) به صورت زیر است.

$$\beta = \begin{bmatrix} \mathbf{H}_A^{-1}(\mathbf{h} - \mathbf{H}_{\bar{A}} \beta_A) \\ \beta_A \end{bmatrix}$$

در ادامه روش بهتر و ساده‌تری برای محاسبه بردار پارامتر β_A بیان می‌شود.

لم ۲.۲.۵. برای هر مجموعه اندیس‌گذار A ، و بردار s ، بردار پارامتر $\beta_{A,s}$ به صورت زیر بیان می‌شود.

$$\beta_{A,s} = \operatorname{argmin}_{\beta} \|\tilde{\beta} - X^* \beta\|_2 + \lambda \|\beta\|_1, \quad (2.5)$$

که $\tilde{\beta} = y^* + \lambda X^- (H_A^{-1} H_A)^T s$ و X^- ماتریسی است که در این رابطه صدق می‌کند. $X^{*T} X^- = I$. آن گاه برای هر مجموعه اندیس‌گذار A ، اگر $s = \operatorname{sgn}(\tilde{\beta}_{A,s})$ ، آن گاه $\beta_A = \beta_{A,s}$ که $\tilde{\beta}_{A,s} = H_A^{-1}(h - H_{\bar{A},s})$ و $\operatorname{sgn}(a)$ یک بردار از همان بعد بردار دلخواه a است که i امین مولفه آن، بر اساس علامت a_i یکی از مقادیر ۱ یا -۱ را می‌پذیرد.

نتیجه ۱.۲.۵. با استفاده از لم‌های ۱.۲.۵ و ۲.۲.۵، اگر A و s به گونه‌ای انتخاب شوند که ماتریس H_A معکوس پذیر باشد و $s = \operatorname{sgn}(\tilde{\beta}_{A,s})$ ، آن گاه پاسخ مسئله بهینه‌سازی (۱.۵) به صورت $\hat{\beta}^{rL} = [\tilde{\beta}_{A,s}, \beta_{A,s}]^T$

در واقعیت محدودیت خطی $H\beta = h$ ممکن است برقرار نباشد. تردید درباره این شرط می‌تواند با آزمون کردن فرضیه $\mathcal{H}_o: H\beta = h$ در مقابل فرضیه $\mathcal{H}_A: H\beta \neq h$ از بین رود. به عنوان یک نتیجه یکی از برآوردگرهای $\hat{\beta}^L$ یا $\hat{\beta}^{rL}$ ، بر اساس رد یا قبول \mathcal{H}_o انتخاب شود. برای آزمون کردن \mathcal{H}_o ، آماره آزمون \mathcal{L}_n به صورت زیر تعریف شده است.

$$\mathcal{L}_n = \frac{(H\hat{\beta}^{Ols} - h)^T (HC^{-1}H^T)^{-1} (H\hat{\beta}^{Ols} - h)}{s^2} \quad (3.5)$$

که s^2 واریانس برآوردگر OLS و $\sigma^2 \xrightarrow{a.s.} s^2$. تحت فرضیه \mathcal{H}_o ، آماره آزمون \mathcal{L}_n توزیع مجانبی کی دو با q درجه آزادی دارد.

با ترکیب برآوردگرهای محدود نشده و محدود شده لاسو برآوردگر آزمون اولیه لاسو به صورت زیر تعریف می‌شود:

$$\begin{aligned} \hat{\beta}^{ptL} &= \hat{\beta}^{rL} I(\mathcal{L}_n < \chi_{q,\alpha}^2) + \hat{\beta}^L I(\mathcal{L}_n > \chi_{q,\alpha}^2) \\ &= \hat{\beta}^L - (\hat{\beta}^L - \hat{\beta}^{rL}) I(\mathcal{L}_n < \chi_{q,\alpha}^2). \end{aligned} \quad (4.5)$$

که $\chi_{q,\alpha}^2$ چندک بالایی توزیع χ_q^2 است.

برآوردگر آزمون اولیه لاسو شدیداً به سطح معناداری α وابسته است. به منظور رهایی از وابستگی این برآوردگر به پارامتر α برآوردگر انقباضی زیر بر پایه $\hat{\beta}^L$ ، $\hat{\beta}^{rL}$ و \mathcal{L}_n پیشنهاد می‌شود برآوردگر انقباضی نوع استاین لاسو به صورت زیر بیان می‌شود.

$$\hat{\beta}^{sL} = \hat{\beta}^L - c(\hat{\beta}^L - \hat{\beta}^{rL})\mathcal{L}_n^{-1}, \quad c = \frac{(q-2)m}{m+2}, \quad m = n-p, q \leq p \quad (5.5)$$

برآوردگر $\hat{\beta}^{sL}$ ممکن است از برآوردگر محدود شده لاسو $\hat{\beta}^{rL}$ عبور کند. بنابراین برآوردگر جزء مثبت انقباضی نوع استاین لاسو برابر خواهد بود با

$$\begin{aligned}\hat{\beta}^{sL+} &= \hat{\beta}^{rL} + \{1 - c\mathcal{L}_n^{-1}\}I(\mathcal{L}_n > c)(\hat{\beta}^L - \hat{\beta}^{rL}) \\ &= \hat{\beta}^{sL} - (1 - c\mathcal{L}_n^{-1})I(\mathcal{L}_n \leq c)(\hat{\beta}^L - \hat{\beta}^{rL})\end{aligned}\quad (۶.۵)$$

۳.۵ برآوردگر انقباضی LAD از دیدگاه روش‌های عددی

در مدل‌های رگرسیونی LAD، **چن و همکاران (۲۰۰۸)** آماره آزمون زیر را پیشنهاد کرد

$$\mathcal{L}_n = \|\mathbf{y}_n - \mathbf{X}_n \hat{\beta}^{rLAD}\|_1 - \|\mathbf{y}_n - \mathbf{X} \hat{\beta}^{uLAD}\|_1. \quad (۷.۵)$$

دوباره پذیره‌های نظم را در نظر بگیرید.

(۱) $F(\varepsilon)$ پیوسته است و چگالی مثبت پیوسته $f(\varepsilon)$ در میانه دارد،

(۲) به‌ازای ماتریس معین مثبت دلخواه $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$

$$\begin{aligned}\lim_{n \rightarrow \infty} n^{-1} C_n^{-1} &= C \\ \max_{1 \leq i \leq n} \mathbf{x}_i^\top C_n^{-1} \mathbf{x}_i &= o(1)\end{aligned}$$

(۳) وزن‌های تصادفی $\omega_1, \omega_2, \dots$ متغیرهای تصادفی نامنفی مستقل و هم‌توزیع هستند به‌طوری‌که $E[\omega_1] = \text{Var}(\omega_1) = 1$ و دنباله‌های $\{\omega_i\}$ و $\{\mathbf{x}_i, e_i\}$ مستقل هستند.

قضیه ۱.۳.۵ (چن و همکاران، ۲۰۰۸) فرض کنید پذیره‌های (۱) و (۲) برقرار باشد. تحت فرضیه صفر،

$$\mathcal{L}_n \xrightarrow{D} \chi_{p-q}^2 / \{4f(0)\} \quad (۸.۵)$$

به منظور عدم برآورد چگالی، روش جدید تقریب توزیع بیان می‌شود. الگوریتم در سه مرحله بیان می‌شود:

مرحله ۱: مجموعه داده $\{(x_i, y_i) : i = 1, 2, \dots, n\}$ را در اختیار دارید. نمونه تصادفی به اندازه n ، مستقل از مجموعه داده از یک توزیع با میانگین و واریانس یک تولید کنید، به‌عنوان مثال، از توزیع نمایی استاندارد. این نمونه تصادفی با نماد $\omega = (\omega_1, \dots, \omega_n)$ نشان می‌دهند. این نمونه تصادفی در مجموعه داده جاسازی و داده-نمونه (DES) نامیده می‌شود که به‌ازای (x_i, y_i, ω_i) ، $i \in \{1, \dots, n\}$.

مرحله ۲: تحت مدل‌های محدود شده (فرضیه صفر) و کامل، به ترتیب فرض کنید

$$\begin{aligned}\hat{\beta}_1^{w-rLAD} &= \operatorname{argmin}_{\beta} \|\omega(\mathbf{y}_n - \mathbf{X}_n \beta)\|_1 \\ \hat{\beta}_1^{w-uLAD} &= \operatorname{argmin}_{\beta} \|\omega(\mathbf{y}_n - \mathbf{X}_n \beta)\|_1\end{aligned}$$

مرحله ۳: تعریف کنید

$$\begin{aligned}\mathcal{L}_n^* &= \sum_{i=1}^n w_i |y_i - \mathbf{x}_i \hat{\beta}_1^{w-rLAD}| - \sum_{i=1}^n w_i |y_i - \mathbf{x}_i \hat{\beta}_1^{w-uLAD}| \\ &\quad - \left(\sum_{i=1}^n w_i |y_i - \mathbf{x}_i \hat{\beta}_1^{rLAD}| - \sum_{i=1}^n w_i |y_i - \mathbf{x}_i \hat{\beta}_1^{uLAD}| \right) \quad (9.5)\end{aligned}$$

آماره آزمون \mathcal{L}_n^* برای تقریب توزیع \mathcal{L}_n استفاده می‌شود.

چن و همکاران (۲۰۰۸) بیان کردند که اگر پذیره (۳) برقرار باشد، اعتبار استفاده از این روش را برای تقریب توزیع \mathcal{L}_n مورد تایید است.

ملاحظه ۱.۳.۵. دقت کنید در طرح DES، هر بار یک مولفه تصادفی از نمونه $\{\omega_1, \dots, \omega_n\}$ در مجموعه داده‌ها جاسازی شد. این روش، نه یک روش باز نمونه‌گیری است و نه یک روش بوت‌استرپ.

وقتی فرضیه $\beta_2 = 0$ درست باشد، برآوردگر rLAD واریانس جانبی کمتری از برآوردگر uLAD دارد. با این وجود، به ازای $\beta_2 \neq 0$ ، در بسیاری از حالت‌ها $\hat{\beta}_1^{rLAD}$ ممکن است اریب و ناسازگار باشد. به این دلیل، مطلوب است که یک برآوردگر آزمون اولیه LAD (ptLAD) با انتخاب یکی از دو برآوردگر $\hat{\beta}_1^{uLAD}$ یا $\hat{\beta}_1^{rLAD}$ بر اساس قبول یا رد فرضیه صفر، \mathcal{H}_0 ساخته شود. این برآوردگر با نماد $\hat{\beta}_1^{ptLAD}$ نشان داده می‌شود. ptLAD پارامتر β_1 برابر است با

$$\begin{aligned}\hat{\beta}_1^{ptLAD} &= \hat{\beta}_1^{uLAD} I(\mathcal{L}_n > \mathcal{L}_\alpha) + \hat{\beta}_1^{rLAD} I(\mathcal{L}_n \leq \mathcal{L}_\alpha), \\ &= \hat{\beta}_1^{uLAD} - (\hat{\beta}_1^{uLAD} - \hat{\beta}_1^{rLAD}) I(\mathcal{L}_n \leq \mathcal{L}_\alpha), \quad (10.5)\end{aligned}$$

که \mathcal{L}_α چندک $100\alpha\%$ بالایی توزیع \mathcal{L}_n تحت فرضیه صفر است که می‌تواند با چندک $(1 - \alpha)$ ام از 10000 مقدار تولید شده آماره آزمون \mathcal{L}_n^* رابطه‌ی (۹.۵) تقریب زده می‌شود.

۴.۵ برآوردهای انقباضی در بعد بالا

مدل رگرسیونی (۱.۱) را در نظر بگیرید که در آن $\varepsilon_n \sim N_p(\mathbf{0}, \sigma^2 I_n)$. برآوردهای پارامتر β معمولاً OLS در نظر گرفته می‌شود. اگرچه، در رگرسیون کلاسیک $p < n$ و p ثابت در نظر گرفته شده‌اند)، در مثال‌های واقعی، معمولاً موقعیت‌هایی وجود دارد که p با n رشد کند و حتی p به بی‌نهایت سریع‌تر از n میل کند. اخیراً پیشرفت‌ها در روش‌های آماری به آماردانان اجازه می‌دهد تا مجموعه داده‌ها با بعد بالا را جمع‌آوری کنند که در آن، تعداد متغیرها، p به‌طور قابل ملاحظه‌ای نسبت به تعداد نمونه‌ها، n بیشتر است، به عبارت دیگر، $n \gg p$. از این نوع داده‌ها در بسیاری از علوم مانند فن‌آوری اطلاعات، پردازش تصویر، ژنومیک، نجوم، امور مالی، بازاریابی و ... استفاده می‌شود. با توجه به اینکه ابزارهای آمار کلاسیک نمی‌تواند برای مسائل با بعد بالا استفاده شود، مثلاً معکوس‌پذیری ماتریس C سوال برانگیز است، روش‌های جدید توسط نویسندگان بسیاری مانند **بولمن و ون‌دگیر (۲۰۱۱)** و **وانگ و شن (۲۰۱۱)** در دهه‌های اخیر معرفی شده‌اند.

حالت بعد بالا $p \gg n$ به‌ازای p بزرگ، n ثابت پیچیده‌تر است. برای مثال، در داده‌های میکروسکوپی، تعداد ژن‌ها (p) هزاران و تعداد نمونه (n) معمولاً کمتر از ۵۰ است. در حالی که در بیشتر مطالعات بعد بالا، فرض می‌شود که n تغییر می‌کند و p با n افزایش می‌یابد و یا از مرتبه n است. در این فصل، حالتی را در نظر بگیرید که n ثابت است.

برای برطرف کردن معکوس‌پذیری C_n در مدل (۱.۱) هنگامی که $p > n$ است، برآوردهای ریب پیشنهادی **هورل و کنارد (۱۹۷۰)**، جایگزین خوبی برای برآوردهای OLS است که در آن پارامتر ریب به p وابسته است.

در رگرسیون بعد بالا، محققین معمولاً فرض می‌کنند که مدل، تنک است. تنکی، بدین معنی است که برخی از متغیرها در متغیر پاسخ مشارکتی ندارند و بنابراین ضریب آن‌ها دقیقاً صفر است. در نتیجه، برای مواجهه با داده‌های بعد بالا، ابتدا از روش‌های کاهش بعد استفاده می‌شود. یکی از روش‌های کاهش بعد، استفاده از برآوردهای لاسو است (**تیبشیرانی، ۱۹۹۶**). اما، در واقعیت، برخی از متغیرها هنوز بر روی متغیر پاسخ تاثیر دارند، اگر چه ممکن است مقدار تاثیر آن به‌بزرگی تاثیر متغیرهای مهم نباشد. به عبارت دیگر، یک روش انتخاب متغیر نسبتاً پذیرفته شده، اغلب متغیرهای بی‌اهمیت را کلاً نادیده می‌گیرد. نادیده گرفتن کلی متغیرهای بی‌اهمیت یا اجبار آن‌ها به صفر شدن، چندان ارزشی ندارد. با این وجود، اگر این متغیرها در مدل وجود داشته باشند، ممکن است خطای پیش‌گویی کلی را در بسیاری از آزمایش‌ها مانند تجزیه و تحلیل داده‌های میکروسکوپی کاهش دهد.

فلاح‌پور و همکاران (۲۰۱۲)، **رحیم و همکاران (۲۰۱۲)** و **یوزباشی و احمد (۲۰۱۵)** روش‌های انقباضی نوع استاین را که اجازه می‌دهد متغیرهای بااهمیت و بی‌اهمیت، هر دو در مدل باقی بمانند، را در مطالعات خود به کار برده‌اند.

با این اطلاعات، استراتژی ما به کار گیری اطلاعات فرعی است که یک روش انتخاب متغیر در روش‌های انقباضی نوع استاین برای بهبود پیش‌گویی استفاده می‌کند. در ادامه، با استفاده از اطلاعات پیشین در مورد بردار پارامترهای مجهول β در مدل رگرسیونی بر اساس برآوردگر لاسو در حالت بعد بالای $p > n$ مسئله برآورد بهبود بخشیده می‌شود.

مدل پیشنهادی برای داده‌ها که فقط متغیرهای مهم در پیشگویی متغیر پاسخ را دربرمی‌گیرد، مدل محدودشده نامیده می‌شود. هم‌چنین مدلی که همه متغیرها را در نظر می‌گیرد، مدل محدودنشده می‌نامند. روش انقباضی نوع استاین برآوردهای محدودنشده و محدودشده را با هم ترکیب می‌کند.

فرض کنید β_0 بردار ضرایب متغیرهای مستقل است که می‌تواند با سایر روش‌ها مانند اطلاعات پیشین به دست آید. این مورد ممکن است به صورت محدودیت $\beta = \beta_0$ روی β نوشته شود. برآوردگری که بدون در نظر گرفتن اطلاعات پیشین و فقط بر اساس داده‌ها به دست آید، برآوردگر محدودنشده نامیده می‌شود. در این مدل رگرسیونی، برآوردگر محدودنشده، برآوردگر ریج است و آن را با نماد $\hat{\beta}^U = \hat{\beta}^{Rdg}$ نشان می‌دهند.

برآوردهای لاسو متغیرهای مهم را انتخاب می‌کند و سایر متغیرهای بی‌اهمیت را کنار می‌گذارد. بنابراین، اطلاعات فرعی می‌تواند به صورت زیر به دست آید.

$$\beta_0 = \hat{\beta}^{Lso} = \hat{\beta}^R \quad (11.5)$$

که برآوردگر محدودشده (RE) نامیده می‌شود.

برای بررسی فرضیه $H_0: \beta = \beta_0$ آزمون زیر پیشنهاد می‌شود (لو و ژو، ۲۰۱۱).

$$G_{n,p} = \frac{(y_n - X_n \hat{\beta}^{Rdg})^\top (y_n - X_n \hat{\beta}^{Rdg})}{\lambda_p^2 (\hat{\beta}^{Rdg} - \beta_0)^\top (\text{diag}(C_n))^{-1} (\hat{\beta}^{Rdg} - \beta_0) / p} \quad (12.5)$$

که در آن $\text{diag}(C_n)$ یک ماتریس قطری با عناصر قطری C_n است. لو و ژو (۲۰۱۱) نشان دادند با در نظر گرفتن پذیره‌های زیر، آماره آزمون $G_{n,p}$ وقتی $p \rightarrow \infty$ و حجم نمونه n ثابت است، به توزیع χ_n^2 همگرا است. بنابراین، در سطح معنی‌داری α ، آماره آزمون $G_{n,p}$ فرضیه H_0 را رد می‌کند اگر $G_{n,p} > \chi_{n,\alpha}^2$ ، چندک α ام توزیع χ_n^2 است.

(الف) $1/\lambda_p = o(1)$. به ازای p به اندازه کافی بزرگ، یک بردار $b_{p \times 1}$ وجود دارد به طوری که $\beta = X_n^\top X_n b$. علاوه بر این، ثابت $\epsilon > 0$ وجود دارد که هر مولفه b از مرتبه $1/p^{1+\epsilon}$ است.

(ب) σ_p و λ_p به گونه‌ای انتخاب می‌شوند که $\lambda_p = o(1)$ و $\sigma_p = o(\lambda_p^{\epsilon/5})$.

(ج) همه p متغیرهای مستقل، ناهمبسته هستند.

(د) فرض کنید ω_i ، i امین مقدار ویژه ناصفر C_n باشد و فرض کنید به ازای $i = 1, \dots, n$ ،

$\lambda_p \cdot \omega_i > 0$ طوری انتخاب می‌شود که به ازای هر $i = 1, \dots, n$ ، $p^{-\epsilon/2} \lambda_p / \sigma_p = o(1)$ و

$$\omega_i = o(\lambda_p)$$

برآوردگر انقباضی نوع استاین، $\hat{\beta}^S$ پارامتر β به صورت زیر تعریف می‌شود.

$$\begin{aligned}\hat{\beta}^S &= \beta_0 + \left\{1 - \frac{c}{G_{n,p}}\right\} (\hat{\beta}^U - \beta_0) \\ &= \hat{\beta}^U - \frac{c}{G_{n,p}} (\hat{\beta}^U - \beta_0),\end{aligned}\quad (13.5)$$

که در آن c مقدار ثابت بهینه است که مخاطره توان دوم برآوردگر را کمینه می‌کند. به ازای $n \geq 3$ ، فرض می‌شود $c = n - 2$.

مقدار انقباض که بر اساس آماره آزمون $G_{n,p}$ و ثابت c تعیین می‌شود، باعث می‌شود برخی از ضرایب صفر شوند. اگر $cG_{n,p}^{-1} > 1$ ، عامل انقباضی $\{1 - kG_{n,p}^{-1}\}$ منفی می‌شود که منجر به معکوس شدن علامت برخی از ضرایب می‌شود که از آن به بیش‌انقباضی تعبیر می‌شود. برای اینکه از بیش‌انقباضی جلوگیری کنیم، ترکیب محدبی از $\hat{\beta}^U$ و $\hat{\beta}^S$ در نظر گرفته می‌شود و آن را برآوردگر جزء مثبت انقباضی نوع استاین (PSS) می‌نامند که به صورت زیر تعریف می‌گردد:

$$\hat{\beta}^{S+} = \hat{\beta}^S - \left(1 - \frac{c}{G_{n,p}}\right) (\hat{\beta}^U - \beta_0) I(G_{n,p} \leq k) \quad (14.5)$$

که در آن $s^+ = \max\{0, s\}$. دوباره، تاکید می‌شود $\hat{\beta}^{S+}$ برای کنترل کردن بیش‌انقباضی که در ذات برآوردگر نوع استاین وجود دارد، به کار می‌رود. در ادامه، بر روی کارایی برآوردگر PSS بر برآوردگر محدود نشده $\hat{\beta}^U$ تمرکز می‌گردد.

۵.۵ ایده‌های دیگر

- در مدل رگرسیونی (۱.۱) ماتریس X_n می‌تواند ثابت یا تصادفی در نظر گرفته شود، یک مدل رگرسیونی چندگانه با هر دو ضرایب ثابت و تصادفی، یک مدل آمیخته نامیده می‌شود. برآوردگرهای انقباضی در مدل‌های آمیخته جریمه‌شده می‌تواند موضوع پژوهشی در آینده باشد.
- در این رساله، فرض شده است که واریانس X_n ‌ها مشخص و ثابت است. می‌توان آن را به ایده ناهم‌واریانسی تعمیم داد. در مدل‌های هم‌واریانس فرض شده است که واریانس خطاها برای همه مقادیر X_n ثابت باشد. ناهم‌واریانسی اجازه می‌دهد که واریانس به شدت به X_n وابسته باشد که دقت بیشتری در بسیاری از مسائل کاربردی دارد. واریانس هزینه‌ها اغلب برای کسانی که درآمد بالاتری دارند، بیشتر از افراد با درآمد پایین است. **تفالیس (۲۰۰۸)** پیشنهاد می‌کند روش کمترین توان‌های دوم را روی درصدی از خطاها به کار برد. این روش، تاثیر مقادیر بزرگتر متغیر پاسخ را در مقایسه با روش کمترین توان‌های دوم معمولی کاهش می‌دهد. ارائه برآوردگرهای انقباضی و نیز برآوردگرهای جریمه‌شده در این مدل می‌تواند موضوع پژوهش دیگری باشد.

- مدل رگرسیون خطی چندگانه، مدلی انعطاف‌پذیر است. به عبارت دیگر، بررسی برآوردگرهای ضرایب رگرسیونی در این مدل، تعمیم آن‌ها را به آسانی در سایر مدل‌های رگرسیونی از قبیل مدل‌های رگرسیونی به ظاهر نامرتبط (SUR)، مدل همبستگی ترتیبی، مدل ناهمپراش، و مدل با همبستگی برابر امکان‌پذیر می‌نماید. برای آگاهی بیشتر درباره این مدل‌ها به ترتیب **زلنر (۱۹۶۲)**؛ **اندرسون (۱۹۴۸)**؛ **تیلور (۱۹۷۸)**؛ **سریواستاوا و جایلز (۱۹۸۷)** و **کوراتا (۱۹۹۹)** را ببینید.
- می‌توان به‌جای تابع هوبر از توابع دیگری مانند هامپل (**همپل و همکاران، ۱۹۸۶**) و ... نیز در فصل ۲ استفاده کرد و نتایج آن‌ها را مورد مقایسه قرار داد.
- می‌توان توزیع خطا را توزیع‌های چوله در نظر گرفت.

پیوست آ

تعاریف، لم‌ها و قضایای اساسی

در این بخش تعدادی از تعاریف و قضایا که در اثبات بعضی از مطالب این رساله مورد استفاده قرار گرفته اند، بیان شده است. این مطالب را می‌توان در **سیرل (۱۹۸۲)** مشاهده کرد.

آ.۱ جبر ماتریس‌ها

قضیه آ.۱.۱. اگر A یک ماتریس $n \times n$ با مقادیر ویژه $\lambda_1, \dots, \lambda_n$ و بردارهای ویژه $\gamma_1, \dots, \gamma_n$ باشد، به طوری که $\lambda_n = \min \{\lambda_i\}$ و $\lambda_1 = \max \{\lambda_i\}$ و $i = 1, \dots, n$ آن‌گاه

$$A = \lambda_1 \gamma_1 \gamma_1^T + \dots + \lambda_n \gamma_n \gamma_n^T, \quad I_n = \gamma_1 \gamma_1^T + \dots + \gamma_n \gamma_n^T.$$

همچنین

$$\sup_x \left(\frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \right) = \lambda_1, \quad \inf_x \left(\frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \right) = \lambda_n.$$

بنابراین

$$\lambda_n = \min \{\lambda_i\} \leq \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} < \max \{\lambda_i\} = \lambda_1.$$

قضیه آ.۲.۱. اگر C و D دو ماتریس نیمه‌معین مثبت باشند که ماتریس D معکوس پذیر است و هر دو $(m \times m)$ هستند. آن‌گاه،

$$\text{Ch}_{\min}(CD^{-1}) \leq \frac{\mathbf{x}^T C \mathbf{x}}{\mathbf{x}^T D \mathbf{x}} \leq \text{Ch}_{\max}(CD^{-1}).$$

که $\text{Ch}_{\min}(A)$ و $\text{Ch}_{\max}(A)$ به ترتیب، کوچکترین و بزرگترین مقدار ویژه ماتریس دلخواه A و بردار ستونی m بعدی است.

قضیه ۳.۱.۱. فرض کنید $\lambda_1, \dots, \lambda_p$ مقادیر ویژه ماتریس $A_{p \times p}$ باشند، آن‌گاه

$$(۱) \det(A) = \prod_{i=1}^p \lambda_i.$$

$$(۲) \text{tr}(A) = \sum_{i=1}^p \lambda_i.$$

(۳) رتبه A برابر تعداد مقادیر ویژه مخالف صفر ماتریس A است.

(۴) ماتریس A نامنفرد است اگر و تنها اگر همه مقادیر ویژه آن غیر صفر باشند.

(۵) اگر ماتریس A متقارن باشد، آن‌گاه مقادیر ویژه آن حقیقی هستند.

(۶) ماتریس A معین مثبت (نیمه معین مثبت) است اگر و تنها اگر همه مقادیر ویژه آن مثبت (نامنفی) باشند.

(۷) اگر P یک ماتریس نامنفرد باشد، آن‌گاه A و $P^{-1}AP$ دارای مقادیر ویژه یکسان می‌باشند.

(۸) اگر H یک ماتریس متعامد باشد، آن‌گاه A و C^TAC دارای مقادیر ویژه یکسان می‌باشند.

قضیه ۴.۱.۱. فرض کنید A و B دو ماتریس نامنفرد باشند. آن‌گاه

$$(۱) (kA)^{-1} = k^{-1}A^{-1} \quad (k \text{ اسکالر است}).$$

$$(۲) (AB)^{-1} = B^{-1}A^{-1}$$

(۳) اگر $A = \text{diag}(a_{11}, \dots, a_{pp})$ آن‌گاه $A^{-1} = \text{diag}(a_{11}^{-1}, \dots, a_{pp}^{-1})$.

(۴) $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$ که در آن $D_{q \times p}$ ، $B_{p \times p}$ و $C_{q \times q}$ ماتریس هستند و $C_{q \times q}$ ماتریس نامنفرد است.

(۵) برای ماتریس متقارن $A(p \times p)$ که به صورت $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ افراز شده است،

$$\begin{aligned} A^{-1} &= \begin{pmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22.1}^{-1}A_{21}A_{11.2}^{-1} & A_{22.1}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} A_{11.2}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22.1}^{-1}A_{21}A_{11}^{-1} & A_{22.1}^{-1} \end{pmatrix}, \end{aligned} \quad (۱.آ)$$

که $A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ ، $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ (فرض می‌شود ماتریس‌های A_{11}^{-1} ، A_{22}^{-1} ، $A_{11.2}^{-1}$ و $A_{22.1}^{-1}$ موجود باشند).

$A_{22.1}$ متمم شور A_{11} و $A_{11.2}$ متمم شور A_{22} نامیده می‌شود.

قضیه ۵.۱.آ. فرض کنید ماتریس معین مثبت A به صورت زیر افراز شده باشد

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

که A_{11} ماتریس $q \times q$ است. اگر $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ ، آن گاه

$$A_{11.2}^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22}^{-1}A_{21}A_{11}^{-1}.$$

قضیه ۶.۱.آ. (تجزیه طیفی). فرض کنید $A_{p \times p}$ یک ماتریس متقارن باشد. آن گاه ماتریس متعامد H وجود دارد به طوری که

$$A = HDH^T,$$

و D ماتریس قطری است که درایه های روی قطر آن مقادیر ویژه ماتریس A است.

قضیه ۷.۱.آ. فرض کنید ماتریس های $A_{k \times p}$ و $B_{m \times p}$ ثابت و $Y_{p \times 1}$ بردار تصادفی با ماتریس کوواریانس Σ باشد، همچنین $z = AY$ و $w = BY$ ، آن گاه

$$\text{Cov}(z) = \text{Cov}(AY) = A\Sigma A^T,$$

$$\text{Cov}(z, w) = \text{Cov}(AY, BY) = A\Sigma B^T.$$

قضیه ۸.۱.آ. اگر A ماتریس متقارن ثابت، Y بردار تصادفی با بردار میانگین μ و ماتریس واریانس-کوواریانس Σ باشند، آن گاه

$$E(Y^T AY) = \text{tr}(A\Sigma) + \mu^T A\mu.$$

۲.آ توابع ریاضی و توزیع ها

در این بخش شکل تابع چگالی احتمال چند توزیع آماری که در این رساله به آنها اشاره شده است، بیان می شود.

تعریف ۱.۲.آ. (تابع گاما). به ازای هر مقدار حقیقی t بزرگتر از $(n-1)/2$ ، تابع گاما به صورت زیر تعریف می شود

$$\Gamma(a) = \int_0^{\infty} y^{a-1} \exp\{-y\} dy = (a-1)!$$

تعریف ۲.۲.آ. (تابع بتا). به ازای هر دو مقدار حقیقی a و b ، تابع بتا به صورت زیر تعریف می شود

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

تعریف ۳.۲.آ (تابع فوق هندسی تعمیم یافته). به ازای مقادیر حقیقی $a_1, \dots, a_p, b_1, \dots, b_q, \dots$ ، b_q تابع فوق هندسی تعمیم یافته به صورت زیر تعریف می شود

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k \dots (a_p)_k}{(b_1)_k \dots (b_q)_k} \times \frac{z^k}{k!},$$

که در آن $(a)_k = a(a+1) \dots (a+k-1)$ می باشد.

تعریف ۴.۲.آ (توزیع نرمال چند متغیره). بردار تصادفی X دارای توزیع نرمال p متغیره با بردار میانگین μ و ماتریس واریانس-کوواریانس Σ است و آن را با نماد $X \sim \mathcal{N}_p(\mu, \Sigma)$ نشان می دهند، هرگاه تابع چگالی احتمال آن به صورت زیر باشد

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{(x - \mu)^T \Sigma^{-1} (x - \mu)}{2} \right\}, \quad x, \mu \in \mathbb{R}^p, \Sigma \in \mathbb{R}^{p \times p}.$$

قضیه ۱.۲.آ. اگر ماتریس $B_{k \times p}$ ثابت، ماتریس $A_{p \times p}$ متقارن ثابت و $Y \sim \mathcal{N}_p(\mu, \Sigma)$ باشد، آن گاه $Y^T A Y$ و $B Y$ از هم مستقلند اگر و فقط اگر $B \Sigma A = 0$.

قضیه ۲.۲.آ. اگر A و B ماتریس های متقارن ثابت و $Y \sim \mathcal{N}_p(\mu, \Sigma)$ باشد، آن گاه $Y^T A Y$ و $Y^T B Y$ از هم مستقلند اگر و فقط اگر $A \Sigma B = 0$.

تعریف ۵.۲.آ (توزیع کی دو مرکزی). اگر $X \sim \mathcal{N}_p(0, I)$ باشد، آن گاه $U = X^T X$ دارای توزیع کی دو مرکزی با p درجه آزادی است که آن را با نماد $U \sim \chi_p^2$ نشان می دهند و تابع چگالی احتمال آن به صورت زیر است:

$$f(u) = \frac{u^{\frac{p}{2}-1}}{2^{\frac{p}{2}} \Gamma(\frac{p}{2})} \exp \left\{ -\frac{u}{2} \right\}, \quad u > 0.$$

تعریف ۶.۲.آ (توزیع کی دو غیر مرکزی). اگر $X \sim \mathcal{N}_p(\mu, I_p)$ باشد، آن گاه متغیر $U = X^T X$ توزیع کی دو غیر مرکزی با p درجه آزادی و پارامتر غیر مرکزی $\Delta^2 = \frac{1}{2} \mu^T \mu$ دارد و با نماد $U \sim \chi_p^2(\Delta^2)$ نشان می دهند. تابع چگالی احتمال آن به صورت زیر می باشد:

$$f(u) = \sum_{k=0}^{\infty} P(K = k) g_{p+2k}(u),$$

که در آن K یک متغیر تصادفی از توزیع پواسن با میانگین $\frac{\Delta^2}{2}$ و $g_{p+2k}(u)$ تابع چگالی احتمال کی دو مرکزی با $p+2k$ درجه آزادی است.

قضیه ۳.۲.آ. اگر $Y \sim \mathcal{N}_p(\mu, \sigma^2 I_p)$ باشد، آن گاه $Y^T A Y \sim \chi_r^2(\Delta^2)$ است اگر و فقط اگر A خودتوان با رتبه r باشد که در آن $\Delta^2 = \frac{\mu^T A \mu}{\sigma^2}$.

قضیه ۴.۲.آ. فرض کنید ماتریس متقارن ثابت با رتبه r و $Y \sim \mathcal{N}_p(\mu, \Sigma)$ باشد آن گاه شرط لازم و کافی برای آن که $Y^T A Y \sim \chi_r^2(\Delta^2)$ باشد آن است که $A \Sigma$ خودتوان باشد که در آن $\Delta = \frac{\mu^T A \mu}{\sigma^2}$.

قضیه آ.۵.۲. (جاج و باک، ۱۹۷۸) فرض کنید $Z = (Z_1, \dots, Z_p)^\top$ بردار p بعدی از توزیع $\mathcal{N}_p(\mu, I_p)$ باشد. به‌ازای هر تابع اندازه پذیر ϕ ,

$$E [Z \phi(Z^\top Z)] = \mu E [\phi(\chi_{p+2}^2(\Delta^2))].$$

قضیه آ.۶.۲. (جاج و باک، ۱۹۷۸) فرض کنید $Z = (Z_1, \dots, Z_p)^\top$ بردار p بعدی از توزیع $\mathcal{N}_p(\mu, I_p)$ باشد. به‌ازای هر تابع اندازه پذیر ϕ ,

$$E [ZZ^\top \phi(Z^\top Z)] = I_p E [\phi(\chi_{p+2}^2(\Delta^2))] + \mu \mu^\top E [\phi(\chi_{p+4}^2(\Delta^2))].$$

تعریف آ.۷.۲. (توزیع فیشر مرکزی). اگر $U_1 \sim \chi_{p_1}^2$ و $U_2 \sim \chi_{p_2}^2$ مستقل از یکدیگر باشند، آن‌گاه $V = \frac{U_1/p_1}{U_2/p_2} \sim \mathcal{F}_{p_1, p_2}$ دارای توزیع فیشر مرکزی با p_1 و p_2 درجه آزادی است که آن را با نماد $V \sim \mathcal{F}_{p_1, p_2}$ نشان می‌دهند و تابع چگالی احتمال آن به‌صورت زیر می‌باشد

$$f(v) = \frac{1}{B(\frac{p_1}{2}, \frac{p_2}{2})} \times \frac{p_1^{\frac{p_1}{2}} p_2^{\frac{p_2}{2}} v^{\frac{p_1}{2}-1}}{(p_2 + p_1 v)^{\frac{p_1+p_2}{2}}}, \quad v > 0.$$

تعریف آ.۸.۲. (توزیع فیشر غیر مرکزی). اگر $U_1 \sim \chi_{p_1}^2(\Delta^2)$ و $U_2 \sim \chi_{p_2}^2$ مستقل از یکدیگر باشند، آن‌گاه $V = \frac{U_1/p_1}{U_2/p_2} \sim \mathcal{F}_{p_1, p_2}(\Delta^2)$ دارای توزیع فیشر مرکزی با p_1 و p_2 درجه آزادی و پارامتر غیر مرکزی Δ^2 است که آن را با نماد $V \sim \mathcal{F}_{p_1, p_2}(\Delta^2)$ نشان می‌دهند و تابع چگالی احتمال آن به‌صورت زیر می‌باشد

$$f(v) = {}_1F_1 \left(\frac{(p_1 + p_2)}{2}; \frac{p_1}{2}; \frac{\frac{p_1 \Delta^2}{2 p_2} v}{1 + \frac{p_1}{p_2} v} \right) \times \frac{v^{\frac{p_1}{2}-1} \left(\frac{p_1}{p_2}\right)^{\frac{p_1}{2}}}{B(\frac{p_1}{2}, \frac{p_2}{2}) \left(1 + \frac{p_1}{p_2} v\right)^{\frac{(p_1+p_2)}{2}}} \exp\{-\Delta^2\}.$$

آ.۳ همگرایی‌ها

همگرایی‌های تصادفی این ایده را که می‌توان انتظار داشت یک دنباله از پیشامدهای غیرقابل پیش‌بینی یا ذاتاً تصادفی در یک الگو قرار گیرند، فرمول‌بندی می‌کنند.

تعریف آ.۱.۳. (همگرایی در توزیع). دنباله متغیرهای تصادفی حقیقی مقدار X_1, X_2, \dots به متغیر تصادفی X همگرا در توزیع است اگر

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \forall x \in \mathbb{R},$$

که در آن F_n و F به ترتیب تابع توزیع تجمعی از متغیرهای تصادفی X_n و X پیوسته است. از نماد $X_n \xrightarrow{D} X$ برای نشان دادن همگرایی در توزیع استفاده می‌شود.

تعریف آ.۲.۳. (همگرایی در احتمال). دنباله $\{X_n\}$ متغیرهای تصادفی در احتمال به متغیر تصادفی X همگرا است اگر

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

از نماد $X_n \xrightarrow{P} X$ برای نشان دادن همگرایی در احتمال استفاده می‌شود. گاهی اوقات این همگرایی را به صورت $\text{plim}_{n \rightarrow \infty} X_n = X$ نشان می‌دهند.

ایده اساسی تعریف همگرایی در احتمال این است که احتمال یک مشاهده غیرمعمول در یک دنباله با افزایش جمله‌های آن، کوچک و کوچکتر می‌شود. مفهوم همگرایی در احتمال در آمار بسیار استفاده می‌شود. برای مثال، یک برآوردگر سازگار است اگر در احتمال به کمیت برآورد شده همگرا باشد.

تعریف ۳.۳.آ (همگرایی a.s.). دنباله متغیرهای تصادفی $\{X_n\}$ به متغیر تصادفی X ، همگرایی تقریباً کامل (a.s.) است اگر

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

از نماد $X_n \xrightarrow{\text{a.s.}} X$ برای نشان دادن همگرایی در توزیع استفاده می‌شود. گاهی اوقات این همگرایی را به صورت $\text{plim}_{n \rightarrow \infty} X_n = X$ نشان می‌دهند.

نماد O بزرگ یک نماد ریاضی است که رفتار حدی یک تابع را وقتی آرگومان آن به یک مقدار ثابت یا بی‌نهایت میل می‌کند، نشان می‌دهد و معمولاً برای مقایسه رفتار جانبی توابع به کار می‌رود.

تعریف ۴.۳.آ. فرض کنید f و g توابعی باشند که روی زیرمجموعه‌ای از اعداد حقیقی تعریف شده‌اند. وقتی $x \rightarrow \infty$ ، $f(x) = O(g(x))$ اگر و فقط اگر ثابت مثبت M وجود داشته باشد به طوری که برای مقادیر به اندازه کافی بزرگ x ، قدر مطلق $f(x)$ حداکثر M برابر قدر مطلق $g(x)$ باشد. به عبارت دیگر، $f(x) = O(g(x))$ اگر عدد حقیقی مثبت M و عدد حقیقی x_0 وجود داشته باشد به طوری که

$$\forall x \geq x_0 \quad |f(x)| \leq M|g(x)|.$$

معمولاً پذیره $x \rightarrow \infty$ در نظر گرفته نمی‌شود و فقط به سادگی بیان می‌شود که

$$f(x) = O(g(x)).$$

تعریف ۵.۳.آ. فرض کنید f و g توابعی باشند که روی زیرمجموعه‌ای از اعداد حقیقی تعریف شده است. وقتی $x \rightarrow \infty$ ، $f(x) = o(g(x))$ بدین معنی است که برای ثابت مثبت ε ، ثابت N وجود دارد به طوری که

$$\forall x \geq N \quad |f(x)| \leq \varepsilon|g(x)|.$$

عبارت $f(x) = o(g(x))$ بیان می‌کند که $g(x)$ بسیار سریعتر از $f(x)$ افزایش پیدا می‌کند. دقت کنید تفاوت O و o در این است که O برای حداقل یک ثابت M برقرار است و o برای هر ثابت مثبت کوچک ε برقرار باشد. در این راستا، نماد o عبارت قوی‌تری از نماد O بیان می‌کند. هر تابعی که $o(g)$ باشد، $O(g)$ نیز است اما عکس آن برقرار نیست.

اگر $g(x)$ غیرصفر باشد، یا حداقل در همسایگی یک نقطه مشخص، غیرصفر باشد، رابطه $f(x) = o(g(x))$ معادل است با

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0.$$

تعریف ۶.۳.آ. برای مجموعه متغیرهای تصادفی X_n و مجموعه متناظر از ثابت‌های a_n ، نماد $X_n = o_p(a_n)$ بدین معنی است که هنگامی که n به حد مناسبی میل کند، مقادیر X_n/a_n به صفر همگرا باشد. به طور معادل، $X_n = o_p(a_n)$ می‌تواند به صورت $X_n/a_n = o_p(1)$ نوشته شود که $X_n = o_p(1)$ به صورت زیر تعریف می‌شود:

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|X_n| \geq \varepsilon) = 0.$$

تعریف ۷.۳.آ. برای مجموعه متغیرهای تصادفی X_n و مجموعه متناظر از ثابت‌های a_n ، نماد $X_n = O_p(a_n)$ بدین معنی است که مقادیر X_n/a_n به طور تصادفی کران‌دار هستند. به عبارت دیگر، برای هر $\varepsilon > 0$ ، اعداد متناهی $M, N > 0$ وجود دارند به طوری که

$$\forall n > N \quad P(|X_n/a_n| > M) < \varepsilon.$$

تفاوت بین تعریف‌های o_p و O_p جزئی است. اگر از تعریف حد استفاده شود آن‌گاه:

$$\forall \varepsilon \quad \exists N_\varepsilon, \delta_\varepsilon \quad P(|X_n| \geq \delta_\varepsilon) \leq \varepsilon \quad \forall n > N_\varepsilon : O_p(1)$$

$$\forall \varepsilon, \delta \quad \exists N_{\varepsilon, \delta}, \delta_\varepsilon \quad P(|X_n| \geq \delta) \leq \varepsilon \quad \forall n > N_{\varepsilon, \delta} : o_p(1)$$

تفاوت در δ قرار دارد. برای کران‌داری تصادفی، باید یک δ (نسبتاً بزرگ) وجود داشته باشد که در نامساوی صدق کند و δ می‌تواند وابسته به ε (اینجا، δ_ε) باشد. به عبارت دیگر، برای همگرایی عبارت نه فقط برای یک مقدار بلکه باید برای هر مقدار کوچک دلخواه δ برقرار باشد. در این حالت، بدین معنی است که باید کران‌دار باشد که وقتی حجم نمونه افزایش پیدا می‌کند، کران آن کوچکتر می‌شود. در این راستا، اگر یک دنباله $o_p(1)$ باشد، آن‌گاه $O_p(1)$ به عبارت دیگر، همگرایی در احتمال، کران‌داری تصادفی را نتیجه می‌دهد اما عکس آن برقرار نیست.

قضیه ۱.۳.آ. (قضیه کرامر-والد). فرض کنید $X_n = (X_{n1}, \dots, X_{nk})$ و $X = (X_1, \dots, X_k)$ بردارهای تصادفی k بعدی باشند. آن‌گاه X_n در توزیع به X به طور مجانبی همگراست اگر و فقط اگر به‌ازای بردار دلخواه (t_1, \dots, t_k) ،

$$\sum_{i=1}^k t_i X_{ni} \xrightarrow{D} \sum_{i=1}^k t_i X_i.$$

قضیه ۲.۳.آ. (قضیه هاجک-سیداک). فرض کنید X_1, X_2, \dots متغیرهای تصادفی مستقل و هم‌توزیع با میانگین μ و واریانس متناهی σ^2 باشند. اگر $a_n = (a_{n1}, \dots, a_{nn})$ بردار ثابت‌ها باشد به طوری که وقتی n به سمت بی‌نهایت میل می‌کند،

$$\max_{1 \leq i \leq n} \frac{a_{ni}^2}{\sum_{j=1}^n a_{nj}^2} \rightarrow 0,$$

آن‌گاه

$$\frac{\sum_{i=1}^n a_{ni}(X_i - \mu)}{\sigma \sqrt{\sum_{j=1}^n a_{nj}^2}} \xrightarrow{D} \mathcal{N}(0, 1).$$

قضیه ۳.۳.آ (قضیه اسلاسکی). فرض کنید $\{X_n\}$ و $\{Y_n\}$ دنباله‌ای از ماتریس‌های تصادفی باشد. اگر X_n در توزیع به متغیر تصادفی X و Y_n در احتمال به ثابت c همگرا باشد، آن‌گاه

$$\begin{aligned} X_n + Y_n &\xrightarrow{D} X + c \\ X_n Y_n &\xrightarrow{D} cX \\ X_n / Y_n &\xrightarrow{D} X/c, \quad c \neq 0. \end{aligned}$$

۴.۰ برخی توابع خاص

تعریف ۱.۴.آ (تابع چوله مقارن). هر تابع n متغیره $F : \mathbb{R}^n \rightarrow \mathbb{R}$ یک تابع چوله مقارن است، اگر علامت مولفه‌ها را تحت هر تبدیل دلخواهی تغییر دهد. به عبارت دیگر

$$F(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = -F(x_1, \dots, x_j, \dots, x_i, \dots, x_n).$$

تعریف ۲.۴.آ (تابع دلتای کرونکر). در ریاضیات، تابع دلتای کرونکر تابعی از دو متغیر است. تابع مقدار ۱ را می‌پذیرد، اگر متغیرها برابر باشند و در غیر این صورت برابر با صفر خواهد بود.

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

که دلتای کرونکر δ_{ij} تابع تکه‌ای-خطی از متغیرهاست.

به عنوان مثال، $\delta_{12} = 0$ و $\delta_{22} = 1$. در جبر خطی، ماتریس همانی I_n مولفه‌هایی برابر با دلتای کرونکر دارد ($I_{ij} = \delta_{ij}$) که i و j مقادیر $1, \dots, n$ را می‌پذیرد).

تعریف ۳.۴.آ. تابع f که روی یک زیرمجموعه محدب از فضای بردارهای حقیقی تعریف شده است و فقط مقادیر مثبت را می‌پذیرد، تابع لگاریتم-محدب یا سوپرمحدب نامیده می‌شود اگر $\log \circ f = \log(f(x))$ ، ترکیب تابع لگاریتم با تابع f یک تابع محدب باشد.

تابع لگاریتم، به طور قابل توجهی از تحدب تابع اولیه می‌کاهد، بنابراین اگر ترکیب همچنان ویژگی محدب بودن را حفظ کند، بدین معنی است که تابع اولیه f از ابتدا یک تابع محدب بوده است و بنابراین از واژه سوپرمحدب استفاده شده است.

تابع لگاریتم-محدب f یک تابع محدب است زیرا ترکیبی از تابع صعودی \exp و $\log \circ f$ است که فرض شده است یک تابع محدب است. اما عکس این مطلب درست نیست. مثلاً فرض کنید $g : x \rightarrow x^2$ یک تابع محدب است اما $\log \circ g : x \rightarrow \log x^2 = 2 \log |x|$ تابع محدب

نیست و بنابراین g تابع لگاریتم-محدب نیست. از طرف دیگر، $x \rightarrow \exp\{x^2\}$ تابع لگاریتم-محدب است زیرا $x \rightarrow \log \exp\{x^2\} = x^2$ تابع محدب است. معروفترین تابع لگاریتم-محدب، تابع گاما روی اعداد حقیقی می‌باشد.

پیوست ب

گزیده‌ای از برنامه‌های رایانه‌ای

نمودار ۱.۲

```
#Fucntion E[\chi_{p+2}^{-2} (\Delta^2)]
Echi2.2<-function(p,D){
r=seq(0,50,1); exp(-(D/2))*sum((1/factorial(r) )*((D/2)^r)*(1/(p+2*r))))}
#Fucntion E[\chi_{p+2}^{-4} (\Delta^2)]
Echi2.4<-function(p,D){
r=seq(0,50,1);
exp(-(D/2))*sum((1/factorial(r) )*((D/2)^r)*(1/(p+2*r))*(1/(p+2*r-2))))}
#Fucntion E[\chi_{p+4}^{-2} (\Delta^2)]
Echi4.2<-function(p,D){
r=seq(0,50,1);
exp(-(D/2))*sum((1/factorial(r) )*((D/2)^r)*(1/(p+2*r-2))))}
#Fucntion E[\chi_{p+4}^{-4} (\Delta^2)]
Echi4.4<-function(p,D){
r=seq(0,50,1);
exp(-(D/2))*sum((1/factorial(r) )*((D/2)^r)*(1/(p+2*r))*(1/(p+2*r+2))))}
```

```

#Fucntion E[\chi_{p+2}^{-2} (\Delta^2) I(\chi_{p+2}^2(D)<c)]
EchiI2.2<-function(p,D){
c=p-2
r=seq(0,50,1);
  exp(-(D/2))*sum((1/factorial(r) )*((D/2)^r)*(1/(p+2*r))*pchisq(c,df=p+2*r))}
#Fucntion E[\chi_{p+2}^{-4} (\Delta^2) I(\chi_{p+2}^2(D)<c)]
EchiI2.4<-function(p,D){
c=p-2
r=seq(0,50,1)
exp(-(D/2))*sum((1/factorial(r) )*((D/2)^r)*(1/(p+2*r))
  *(1/(p-2+2*r))*pchisq(c,df=p-2+2*r))}
#Fucntion E[\chi_{p+4}^{-2} (\Delta^2) I(\chi_{p+4}^2(D)<c)]
EchiI4.2<-function(p,D){
c=p-2
r=seq(0,50,1)
exp(-(D/2))*sum((1/factorial(r) )*((D/2)^r)*(1/(p+2*r+2))
  *pchisq(c,df=p+2*r+2))}
#Fucntion E[\chi_{p+4}^{-4} (\Delta^2) I(\chi_{p+4}^2(D)<c)]
EchiI4.4<-function(p,D){
c=p-2
r=seq(0,50,1)
exp(-(D/2))*sum((1/factorial(r) )*((D/2)^r)*(1/(p+2*r+2))*(1/(p+2*r))
  *pchisq(c,df=p+2*r))}
Echi2.2<-Vectorize(EchiI2.2)
Echi2.4<-Vectorize(EchiI2.4)
Echi4.2<-Vectorize(EchiI4.2)
Echi4.4<-Vectorize(EchiI4.4)
EchiI2.2<-Vectorize(EchiI2.2)
EchiI2.4<-Vectorize(EchiI2.4)
EchiI4.2<-Vectorize(EchiI4.2)
EchiI4.4<-Vectorize(EchiI4.4)
H=function(p,nu,alpha,D){ # H_{nu}(\chi_p^2(\alpha);Delta2)
a=qchisq(alpha,df=p,lower.tail=FALSE)
pchisq(a,df=nu,ncp=D,lower.tail=TRUE)}

```

```
h1 <- function(p,D){
c=p-2;
c * Echi2.2(p,D) + pchisq(c,df=p+2,ncp=D,lower.tail=TRUE) - c * EchiI2.2(p,D)}
h2 <- function(p,alpha,D){
2*H(p,p+2,alpha,D)-H(p,p+4,alpha,D)}
h3 <- function(p,D){
c=p-2; 2 * Echi2.2(p,D) - c*Echi2.4(p,D)}
h4 <- function(p,D){
c=p-2; s1 = 2*pchisq(c,df=p+2,ncp=D,lower.tail=TRUE) - 2 * c * EchiI2.2(p,D)
s2 = pchisq(c,df=p+4,ncp=D,lower.tail=TRUE) - 2 * c * EchiI4.2(p,D)
+ (c^2) * EchiI4.4(p,D)
return(s1 - s2)}
B1=function(p,D){
c=p-2; pchisq(c,df=p+2,ncp=D,lower.tail=TRUE)-c*EchiI2.2(p,D) }
B2=function(p,D){
c=p-2
pchisq(c,df=p+2,ncp=D,lower.tail=TRUE)-2*c*EchiI2.2(p,D)
+(c^2)*EchiI2.4(p,D)}
ADB_U=function(p,k,eta,alpha,D) ((k^2)*eta*D)/((1+k)^2)
ADB_PT=function(p,k,eta,alpha,D){
com = (eta*D)/((1+k)^2)
dif = (k^2)+((H(p,p+2,alpha,D))^2)+(2*k*H(p,p+2,alpha,D))
return(com * dif)}
ADB_S=function(p,k,eta,alpha,D){
c=p-2; com = (eta*D)/((1+k)^2)
dif = (k^2)+(2*c*k*Echi2.2(p,D))+((c^2)*((Echi2.2(p,D))^2))
return(com * dif)}
ADB_PR=function(p,k,eta,alpha,D){
c=p-2; com = (eta*D)/((1+k)^2)
dif = (k^2)+(2*k*h1(p,D))+((h1(p,D))^2)
return(com * dif)}

#####
# Figure 1 #
#####
```

```
rm(list=ls())
set.seed(2231346)
library(reshape)
library(ggplot2)
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)
  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)
  numPlots = length(plots)
  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
      ncol = cols, nrow = ceiling(numPlots/cols)) }
  if (numPlots==1) {
    print(plots[[1]])
  } else {
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))
    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))
      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
        layout.pos.col = matchidx$col))}}
    D.vec=seq(0,20,1)
    p=5;k=0.1;eta=1;alpha=0.01
    ADQB1.a <- matrix(0,nr = length(D.vec), nc = 4)
    colnames(ADQB1.a) <- c("URME", "PTRME" , "SRME" , "PRRME")
    ADQB1.a[,1] <- ADB_U(p,k,eta,alpha,D.vec)
    ADQB1.a[,2] <- ADB_PT(p,k,eta,alpha,D.vec)
    ADQB1.a[,3] <- ADB_S(p,k,eta,alpha,D.vec)
    ADQB1.a[,4] <- ADB_PR(p,k,eta,alpha,D.vec)
    ADQB1.a <- as.data.frame(cbind(D.vec,ADQB1.a))
    ADQB1.a <- melt(ADQB1.a, id = "D.vec", variable_name = "Estimators")
```

```
# Plot
a1 = ggplot(ADQB1.a, aes(x = D.vec , y = value )) +
geom_line(aes(linetype = Estimators),size = 0.1) +
geom_point(aes(shape = Estimators))+
labs(title = bquote( paste(alpha ==.(alpha))),x=expression(Delta^2),
y= "ADQB")+
theme_classic()+
theme(legend.background = element_rect(colour = "black"))+
theme(legend.position = c(0.9, .75))+
ylim(0,3.2)
D.vec=seq(0,20,1)
p=5;k=0.1;eta=1;alpha=0.05
ADQB2.a <- matrix(0,nr = length(D.vec), nc = 4)
colnames(ADQB2.a) <- c("URME", "PTRME" , "SRME" , "PRRME")
ADQB2.a[,1] <- ADB_U(p,k,eta,alpha,D.vec)
ADQB2.a[,2] <- ADB_PT(p,k,eta,alpha,D.vec)
ADQB2.a[,3] <- ADB_S(p,k,eta,alpha,D.vec)
ADQB2.a[,4] <- ADB_PR(p,k,eta,alpha,D.vec)
ADQB2.a <- as.data.frame(cbind(D.vec,ADQB2.a))
ADQB2.a <- melt(ADQB2.a, id = "D.vec", variable_name = "Estimators")
# Plot
a2 = ggplot(ADQB2.a, aes(x = D.vec , y = value )) +
geom_line(aes(linetype = Estimators),size = 0.1) +
geom_point(aes(shape = Estimators))+
labs(title = bquote( paste(alpha ==.(alpha))),x=expression(Delta^2),
y= "ADQB")+
theme_classic()+
theme(legend.background = element_rect(colour = "black"))+
theme(legend.position = "none")+
ylim(0,3.2)
D.vec=seq(0,20,1)
p=5;k=0.1;eta=1;alpha=0.1
ADQB3.a <- matrix(0,nr = length(D.vec), nc = 4)
colnames(ADQB3.a) <- c("URME", "PTRME" , "SRME" , "PRRME")
```

```

ADQB3.a[,1] <- ADB_U(p,k,eta,alpha,D.vec)
ADQB3.a[,2] <- ADB_PT(p,k,eta,alpha,D.vec)
ADQB3.a[,3] <- ADB_S(p,k,eta,alpha,D.vec)
ADQB3.a[,4] <- ADB_PR(p,k,eta,alpha,D.vec)
ADQB3.a <- as.data.frame(cbind(D.vec,ADQB3.a))
ADQB3.a <- melt(ADQB3.a, id = "D.vec", variable_name = "Estimators")
# Plot
a3 = ggplot(ADQB3.a, aes(x = D.vec , y = value )) +
geom_line(aes(linetype = Estimators),size = 0.1) +
geom_point(aes(shape = Estimators))+
labs(title = bquote( paste(alpha ==.(alpha))),x=expression(Delta^2),
y= "ADQB")+
theme_classic()+
theme(legend.background = element_rect(colour = "black"))+
theme(legend.position ="none")+
ylim(0,3.2)
D.vec=seq(0,20,1)
p=5;k=0.1;eta=1;alpha=0.2
ADQB4.a <- matrix(0,nr = length(D.vec), nc = 4)
colnames(ADQB4.a) <- c("URME", "PTRME" , "SRME" , "PRRME")
ADQB4.a[,1] <- ADB_U(p,k,eta,alpha,D.vec)
ADQB4.a[,2] <- ADB_PT(p,k,eta,alpha,D.vec)
ADQB4.a[,3] <- ADB_S(p,k,eta,alpha,D.vec)
ADQB4.a[,4] <- ADB_PR(p,k,eta,alpha,D.vec)
ADQB4.a <- as.data.frame(cbind(D.vec,ADQB4.a))
ADQB4.a <- melt(ADQB4.a, id = "D.vec", variable_name = "Estimators")
# Plot
a4 = ggplot(ADQB4.a, aes(x = D.vec , y = value )) +
geom_line(aes(linetype = Estimators),size = 0.1) +
geom_point(aes(shape = Estimators))+
labs(title = bquote( paste(alpha ==.(alpha))),x=expression(Delta^2),
y= "ADQB")+
theme_classic()+
theme(legend.background = element_rect(colour = "black"))+

```

```

theme(legend.position = "none")+
ylim(0,3.2)
multiplot(a1, a3, a2, a4, cols=2)

```

جدول ٣.٢

```

require(MASS);require(robustlmm);require(mnormt);library(formattable)
rm(list=ls()) #clear the environment
set.seed(12345432)
N=1000
n=50;
rho2<-c(0.85,0.999)
lambda<-c(0.1,0.2)
D2<-c(0,0.01,0.1,0.5,1,2,5,10,20,30)
length(D2)
p=30; sigma=1; alpha=0.1;n.res<-n-p+1
mat.UM<-mat.RUM<-mat.RRM<-mat.RPTM<-mat.RSM<-mat.PRM<-
mat.RIPM<-matrix(0,nc=N,nr=p)
r.UM<-r.RUM<-r.RRM<-r.RPTM<-r.RSM<-r.PRM<-r.RIPM<-matrix(0,nc=1,nr=N)
Result.mat<-matrix(0,nrow = (length(D2)*length(rho2)),ncol = 8)
for(iter.rho in 1:length(rho2)){
Z=matrix(rnorm(n*(p+1)),nr=n,nc=p+1)
X=matrix(0,nr=n,nc=p)
for(i in 1:n){
for(j in 1:p){
X[i,j]=(sqrt(1-rho2[iter.rho]))*Z[i,j]+sqrt(rho2[iter.rho])*Z[i,p+1]}}
for(iter.D2 in 1:length(D2)){
beta=matrix(c(0,0,0,0,D2[iter.D2]),nr=p,nc=1)
for(iter.lam in 1:length(lambda)){
for(iter in 1:N){
e=matrix(n*lambda[iter.lam]*rnorm(n,mean=0,sd=5)+
(1-n*lambda[iter.lam])*rnorm(n,mean=0,sd=1),nr=n,nc=1)
Y=X%*%beta+e
data=cbind.data.frame(X,Y)
fit=lm(Y~.,data=data)

```

```

res.order<-sort(abs(fit$residuals))
s.n<-2.1*median(res.order[((n-n.res)+1):n])
fit.m=rlm(X,Y,maxit = 10000)
U.M<-as.matrix(fit.m$coefficients)
C=(t(X)%*%X)/n
M<-function(b) apply(huberPsi@psi((Y-X)%*%b)/s.n, 1.345)*X,2,sum)/sqrt(n)
#fit.rm<-rlm(X,Y,maxit = 100000)
R.M<-rep(0,p)#as.matrix(fit.rm$coefficients)
S2<-sum(((huberPsi@psi((Y-X)%*%R.M)/s.n, 1.345))^2)/n
Te.M<-(t(M(R.M))%*%solve(C)%*%M(R.M))/S2; Te.M<-drop(Te.M)
U.Te.M<-qchisq(alpha,df=p,ncp=0,lower.tail = FALSE)
D=U.M-R.M
PT.M<-U.M*as.integer(Te.M>U.Te.M)+R.M*as.integer(Te.M<=U.Te.M)
c=(p-2)
s<-fit.m$s
S.M<-R.M+(1-(c/Te.M))*D
#IPT.M<-S.M*as.integer(Te.M>U.Te.M)+R.M*as.integer(Te.M<=U.Te.M)
PR.S.M<-R.M+max((1-(c/Te.M)),0)*D
k1<-drop((p*s)/(t(U.M)%*%U.M))
R<-function(k) solve(diag(1,p)+k*solve(C))
R.U.M<-R(k1)%*%U.M
R.R.M<-R(k1)%*%R.M
R.PT.M<-R(k1)%*%PT.M
R.S.M<-R(k1)%*%S.M
R.PR.S.M<-R(k1)%*%PR.S.M
# R.IPT.M<-R(k1)%*%IPT.M
mat.RUM[,iter]<-R.U.M
mat.RRM[,iter]<-R.R.M
mat.RPTM[,iter]<-R.PT.M
mat.RSM[,iter]<-R.S.M
mat.PRM[,iter]<-R.PR.S.M
# mat.RIPM[,iter]<-R.IPT.M
r.RUM[iter,]<-t(mat.RUM[,iter]-beta)%*%(mat.RUM[,iter]-beta)
r.RRM[iter,]<-t(mat.RRM[,iter]-beta)%*%(mat.RRM[,iter]-beta)

```

```
r.RPTM[iter,]<-t(mat.RPTM[,iter]-beta)%*(mat.RPTM[,iter]-beta)
r.RSM[iter,]<-t(mat.RSM[,iter]-beta)%*(mat.RSM[,iter]-beta)
r.PRM[iter,]<-t(mat.PRM[,iter]-beta)%*(mat.PRM[,iter]-beta)
r.RIPM[iter,]<-t(mat.RIPM[,iter]-beta)%*(mat.RIPM[,iter]-beta) }
risk.RUM=apply(r.RUM,2,mean)
risk.RRM=apply(r.RRM,2,mean)
risk.RPTM=apply(r.RPTM,2,mean)
risk.RSM=apply(r.RSM,2,mean)
risk.PRM=apply(r.PRM,2,mean)
result=cbind(mean(risk.RUM),mean(risk.RRM),mean(risk.RPTM),
mean(risk.RSM),mean(risk.PRM))
#plot(result[1,])
R.E.1=result[,1]/result[,1]
R.E.2=result[,1]/result[,2]
R.E.3=result[,1]/result[,3]
R.E.4=result[,1]/result[,4]
R.E.5=result[,1]/result[,5]
if(iter.rho==1){
  if(iter.lam==1){
    Result.mat[iter.D2,iter.lam]<-R.E.2
    Result.mat[iter.D2,iter.lam+1]<-R.E.3
    Result.mat[iter.D2,iter.lam+2]<-R.E.4
    Result.mat[iter.D2,iter.lam+3]<-R.E.5}
  else{
    Result.mat[iter.D2,iter.lam+3]<-R.E.2
    Result.mat[iter.D2,iter.lam+4]<-R.E.3
    Result.mat[iter.D2,iter.lam+5]<-R.E.4
    Result.mat[iter.D2,iter.lam+6]<-R.E.5}}
if(iter.rho==2){
  if(iter.lam==1){
    Result.mat[iter.D2+10,iter.lam]<-R.E.2
    Result.mat[iter.D2+10,iter.lam+1]<-R.E.3
    Result.mat[iter.D2+10,iter.lam+2]<-R.E.4
    Result.mat[iter.D2+10,iter.lam+3]<-R.E.5}}
```

```
else{
Result.mat[iter.D2+10,iter.lam+3]<-R.E.2
Result.mat[iter.D2+10,iter.lam+4]<-R.E.3
Result.mat[iter.D2+10,iter.lam+5]<-R.E.4
Result.mat[iter.D2+10,iter.lam+6]<-R.E.5}}
if(iter.rho==3){
if(iter.lam==1){
Result.mat[iter.D2+20,iter.lam]<-R.E.2
Result.mat[iter.D2+20,iter.lam+1]<-R.E.3
Result.mat[iter.D2+20,iter.lam+2]<-R.E.4
Result.mat[iter.D2+20,iter.lam+3]<-R.E.5}
else{
Result.mat[iter.D2+20,iter.lam+3]<-R.E.2
Result.mat[iter.D2+20,iter.lam+4]<-R.E.3
Result.mat[iter.D2+20,iter.lam+5]<-R.E.4
Result.mat[iter.D2+20,iter.lam+6]<-R.E.5}}
if(iter.rho==4){
if(iter.lam==1){
Result.mat[iter.D2+30,iter.lam]<-R.E.2
Result.mat[iter.D2+30,iter.lam+1]<-R.E.3
Result.mat[iter.D2+30,iter.lam+2]<-R.E.4
Result.mat[iter.D2+30,iter.lam+3]<-R.E.5}
else{
Result.mat[iter.D2+30,iter.lam+3]<-R.E.2
Result.mat[iter.D2+30,iter.lam+4]<-R.E.3
Result.mat[iter.D2+30,iter.lam+5]<-R.E.4
Result.mat[iter.D2+30,iter.lam+6]<-R.E.5}}
if(iter.rho==5){
if(iter.lam==1){
Result.mat[iter.D2+40,iter.lam]<-R.E.2
Result.mat[iter.D2+40,iter.lam+1]<-R.E.3
Result.mat[iter.D2+40,iter.lam+2]<-R.E.4
Result.mat[iter.D2+40,iter.lam+3]<-R.E.5}
else{
```



```

Result.mat[iter.D2+40,iter.lam+3]<-R.E.2
Result.mat[iter.D2+40,iter.lam+4]<-R.E.3
Result.mat[iter.D2+40,iter.lam+5]<-R.E.4
Result.mat[iter.D2+40,iter.lam+6]<-R.E.5}} ]}}
Result.mat

```

جدول ٥.٣

```

library(datasets)
data(state)
data <- data.frame(state.x77,row.names=state.abb,check.names=T)
summary(data)
p=7
X=as.matrix(data[,-4])
Y=as.matrix(data[,4])
H=diag(1,7,7)
#First
#h=matrix(c(0,0,10,0.2,0.7,0.06,0),nc=1)
#Second
h=matrix(c(0,0,0,0,0,0,0),nc=1)
q=nrow(H)
alpha1=0.01;alpha2=0.05;alpha3=0.10
EST<-function(data){
X=as.matrix(data[,-c(4,9)])
Y=as.matrix(data[,4])
n=nrow(X)
p=ncol(X)
m=n-p
K=((q-2)*m)/((m+2))
C=t(X)%*%X
C.Inv=solve(C)
ols<-C.Inv%*%t(X)%*%Y
s=cv.glmnet(X,Y)$lambda.min
fit=glmnet(X,Y,alpha=1)
UL=coef(fit,s=s)

```

```

UL<-as.matrix(UL) [-1]
RL<-UL-C.Inv%%t(H)%%solve(H%%C.Inv%%t(H)%%(H%%UL-h)
s2=(t(Y-X%%ols)%%(Y-X%%ols))/m
t.S<-((t(H%%ols-h)%%solve(H%%C.Inv%%t(H)%%(H%%ols-h))/s2)
t.s<-drop(t.S)
U.L.1=qchisq(alpha1,q,lower.tail=FALSE)
U.L.2=qchisq(alpha2,q,lower.tail=FALSE)
U.L.3=qchisq(alpha3,q,lower.tail=FALSE)
if(as.integer(t.S <= U.L.1)==1) PTL1=RL else PTL1=UL
if(as.integer(t.S <= U.L.2)==1) PTL2=RL else PTL2=UL
if(as.integer(t.S <= U.L.3)==1) PTL3=RL else PTL3=UL
a=as.numeric(K/t.S)
SL=UL-a*(UL-RL)
PRL=SL-(1-a)*as.integer(t.S <= K)*(UL-RL)
if(as.integer(t.S > U.L.1)==1) IPTL1=PTL1-a*(UL-RL) else IPTL1=PTL1
if(as.integer(t.S > U.L.2)==1) IPTL2=PTL2-a*(UL-RL) else IPTL2=PTL2
if(as.integer(t.S > U.L.3)==1) IPTL3=PTL3-a*(UL-RL) else IPTL3=PTL3
result=cbind(ols,UL,RL,PTL1,PTL2,PTL3,SL,PRL,IPTL1,IPTL2,IPTL3)
colnames(result)<-c("OLS","UL","RL","PTL1","PTL2","PTL3","SL","PRL","IPTL1",
"IPTL2","IPTL3")
return(result)}
k=5#fold
# sample from 1 to k, nrow times (the number of observations in the data)
p=7
N.BOOT=200
mat.ols<-mat.UL<-mat.RL<-mat.PTL1<-mat.PTL2<-mat.PTL3<-mat.SL<-mat.PRL
<-mat.IPTL1<-mat.IPTL2<-mat.IPTL3<-matrix(0,nc=k,nr=p)
r.ols<-r.UL<-r.RL<-r.PTL1<-r.PTL2<-r.PTL3<-r.SL<-r.PRL<-r.IPTL1<-r.IPTL2
<-r.IPTL3<-matrix(0,nc=k,nr=N.BOOT)
for(j in 1:N.BOOT){
data$id <- sample(1:k, nrow(data), replace = TRUE)
list <- 1:k
for(i in 1:k){
trainingset <- subset(data, id %in% list[-i])

```

```
testset <- as.matrix(subset(data, id %in% list[i]))
testset.x<-testset[,-c(4,8)]
testset.y<-testset[,4]
mat.ols[,i]<-EST(trainingset)[,1]
mat.UL[,i]<-EST(trainingset)[,2]
mat.RL[,i]<-EST(trainingset)[,3]
mat.PTL1[,i]<-EST(trainingset)[,4]
mat.PTL2[,i]<-EST(trainingset)[,5]
mat.PTL3[,i]<-EST(trainingset)[,6]
mat.SL[,i]<-EST(trainingset)[,7]
mat.PRL[,i]<-EST(trainingset)[,8]
mat.IPTL1[,i]<-EST(trainingset)[,9]
mat.IPTL2[,i]<-EST(trainingset)[,10]
mat.IPTL3[,i]<-EST(trainingset)[,11]
r.ols[j,i]<-t(testset.x%%mat.ols[,i]-testset.y)%%
(testset.x%%mat.ols[,i]-testset.y)
r.UL[j,i]<-t(testset.x%%mat.UL[,i]-testset.y)%%
(testset.x%%mat.UL[,i]-testset.y)
r.RL[j,i]<-t(testset.x%%mat.RL[,i]-testset.y)%%
(testset.x%%mat.RL[,i]-testset.y)
r.PTL1[j,i]<-t(testset.x%%mat.PTL1[,i]-testset.y)%%
(testset.x%%mat.PTL1[,i]-testset.y)
r.PTL2[j,i]<-t(testset.x%%mat.PTL2[,i]-testset.y)%%
(testset.x%%mat.PTL2[,i]-testset.y)
r.PTL3[j,i]<-t(testset.x%%mat.PTL3[,i]-testset.y)%%
(testset.x%%mat.PTL3[,i]-testset.y)
r.SL[j,i]<-t(testset.x%%mat.SL[,i]-testset.y)%%
(testset.x%%mat.SL[,i]-testset.y)
r.PRL[j,i]<-t(testset.x%%mat.PRL[,i]-testset.y)%%
(testset.x%%mat.PRL[,i]-testset.y)
r.IPTL1[j,i]<-t(testset.x%%mat.IPTL1[,i]-testset.y)%%
(testset.x%%mat.IPTL1[,i]-testset.y)
r.IPTL2[j,i]<-t(testset.x%%mat.IPTL2[,i]-testset.y)%%
(testset.x%%mat.IPTL2[,i]-testset.y)
```

```

r.IPTL3[j,i]<-t(testset.x%%mat.IPTL3[,i]-testset.y)%%
(testset.x%%mat.IPTL3[,i]-testset.y)} }
risk.ols=apply(r.ols,1,mean)
risk.UL=apply(r.UL,1,mean)
risk.RL=apply(r.RL,1,mean)
risk.PTL1=apply(r.PTL1,1,mean)
risk.PTL2=apply(r.PTL2,1,mean)
risk.PTL3=apply(r.PTL3,1,mean)
risk.SL=apply(r.SL,1,mean)
risk.PRL=apply(r.PRL,1,mean)
risk.IPTL1=apply(r.IPTL1,1,mean)
risk.IPTL2=apply(r.IPTL2,1,mean)
risk.IPTL3=apply(r.IPTL3,1,mean)
result=cbind(mean(risk.ols),mean(risk.UL),mean(risk.RL),mean(risk.PTL1),
mean(risk.PTL2),mean(risk.PTL3),mean(risk.SL),mean(risk.PRL),mean(risk.IPTL1),
mean(risk.IPTL2),mean(risk.IPTL3))
colnames(result)<-c("OLS","UL","RL","PTL1(0.01)","PTL2(0.05)","PTL3(0.10)",
"SL","PRL","IPTL1(0.01)","IPTL2(0.05)","IPTL3(0.10)")
rownames(result)<-c("mean")
result
r1<-c(result[1,2]/result[1,3],result[1,2]/result[1,5],
result[1,2]/result[1,7],result[1,2]/result[1,8])
r2<-c(result[1,2]/result[1,3],result[1,2]/result[1,4],
result[1,2]/result[1,7],result[1,2]/result[1,8])
r3<-c(result[1,2]/result[1,3],result[1,2]/result[1,6],
result[1,2]/result[1,7],result[1,2]/result[1,8])
r1;r2;r3

```

۵.۴ جدول

```

rm(list=ls())
#set.seed(12242)
set.seed(2231346)
# fitting LAD regression model
library(quantreg)

```

```
library(mvtnorm)
library(smoothmest)
library(MASS)
library(xtable)
#Values
n=50 #Number of observations
p1=3 #Number of nonzero Covariates
#####
tau2=(1/(2*dt(0, df=1, ncp=0)))^2 # Cauchy-distribution
#####
# Calculating Weight
weights = function(X){
  n = nrow(X)
  w = RD = numeric(n)
  for(i in 1:n){
    RD[i] = t(X[i,]-cov.mve(X)$center)%*%solve(cov.mve(X)$cov)%*%
    (X[i,]-cov.mve(X)$center)
    w[i]= min(1,p/RD[i]) }
  w }
p2.vec=c(5,10,15) #Number of zeros
mse.unres= rep(0,length(p2.vec));
mse.res= rep(0,length(p2.vec));
mse.pt=rep(0,length(p2.vec));
mse.shr= rep(0,length(p2.vec));
mse.pshr=rep(0,length(p2.vec));
mse.lasso=rep(0,length(p2.vec));
delta=0
#####
for (k1 in 1:length(p2.vec)){
  print(k1)
  p=p1+p2.vec[k1]
  # generating covariates
  n1 = floor(n*0.8); n2 = n - n1
  X1=matrix(rmvnorm(n1,rep(0,p),diag(1,p,p)),nr=n1,nc=p)
```

```

X2=matrix(rmvnorm(n2,rep(5,p),diag(25,p,p)),nr=n2,nc=p)
X = rbind(X1,X2)
W=weights(X)
beta= c(rep(1,p1), rep(0,p-p1))
m=5000
bias.un.res=matrix(0,m,p1)
bias.res=matrix(0,m,p1)
un.res=matrix(0,m,p1)
res=matrix(0,m,p1)
pt.est=matrix(0,m,p1)
bias.pt= matrix(0,m,p1)
shr.est=matrix(0,m,p1)
bias.shr.est= matrix(0,m,p1)
positive.shr.est=matrix(0,m,p1)
bias.positive.shr.est=matrix(0,m,p1)
lasso.est=matrix(0,m,p1)
bias.lasso= matrix(0,m,p1)
for ( k in 1 : m){
#generating error
e=matrix(rt(n, df=1, ncp=0),nr=n,nc=1) # Cauchy-distribution
Y=X%*%beta+e
mdata= data.frame(Y,X)
#unrestricted LAD estimate
upfit= rq(Y ~ X-1,data=mdata,tau=0.5,weights = W)
unres.coef=c(upfit$coef[1:p1])
un.res[k,]= unres.coef
bias.un.res[k,]= (unres.coef-beta[1:p1])^2
#restricted LAD estimate
uprfit=rq(Y ~ X[,1:p1]-1,data=mdata,tau=0.5,weights = W)
res.coef=c(uprfit$coef,delta[k1],rep(0,p2.vec[k1]-1))[1:p1] #delta[k1]
res[k,]= res.coef
bias.res[k,]= (res.coef-beta[1:p1])^2
# Calculating C and its partition matrices
C=(t(X)%*%diag(W)%*%X)/n

```

```
W.d = diag(W)
C.s = solve(C)%*(t(X)%*W.d%*W.d%*X)%*solve(C)
index=c(1:p1)
C11=C.s[index,index]
C12=C.s[index,-index]
C21=C.s[-index,index]
C22=C.s[-index,-index]
C22.1=C22-C21%*solve(C11)%*C12
L.test= t(c(upfit$coef[(p1+1):p]))%*C22.1%*c(upfit$coef[(p1+1):p])/tau2
#####
alpha=0.05
# finding priliminary test estimator
KK2= ifelse(L.test<qchisq(alpha,p2.vec[k1],lower.tail = FALSE),1,0)
pt.est1=unres.coef- (KK2)*(unres.coef-res.coef)
pt.est[k,]= pt.est1
bias.pt[k,]=(pt.est1 - beta[1:p1])^2
#####
# finding shrinkage estimate for regression model
q=p2.vec[k1]
shr.est1= res.coef+(1-(q-2)/L.test)*(unres.coef-res.coef)
shr.est[k,]= shr.est1
bias.shr.est[k,]=(shr.est1 - beta[1:p1])^2
#####
# finding positive shrinkage estimate for regression model
positive.shr.est1= res.coef+ max(0, (1-(q-2)/L.test))*(unres.coef-res.coef)
positive.shr.est[k,]=positive.shr.est1
bias.positive.shr.est[k,]=(positive.shr.est1-beta[1:p1])^2
#####
# finding Weighted LAD-LASSO
tau=0.5
lam=numeric(p)
e.mat=matrix(0, nr=p, nc=p)
lad=upfit$coef
for(j in 1:p){
```

```
lam[j]= log(n)/(n*abs(lad[j]))
e.mat[j,j]=1 }
Xw = W*X
Yw = W*Y
Xnew=rbind(Xw, matrix(n*lam*e.mat,nr=p))
Ynew=rbind(Yw,matrix(rep(0,p),nr=p))
datanew=as.data.frame(cbind(Xnew,Ynew))
Out=rq(Ynew ~ Xnew-1,data=datanew,tau)
lad.lasso1=coefficients(Out)[1:p1]
lasso.est[k,]=lad.lasso1
bias.lasso[k,]=(lad.lasso1-beta[1:p1])^2 }
mse.unres[k1]=sum((apply(bias.un.res,2,mean)))+ sum(diag(var(un.res)))
mse.res[k1]= sum((apply(bias.res,2,mean)))+ sum(diag(var(res)))
mse.pt[k1]= sum((apply(bias.pt,2,mean)))+ sum(diag(var(pt.est)))
mse.shr[k1]=sum((apply(bias.shr.est,2,mean)))+ sum(diag(var(shr.est)))
mse.pshr[k1]= sum((apply(bias.positive.shr.est,2,mean)))+
sum(diag(var(positive.shr.est)))
mse.lasso[k1]= sum((apply(bias.lasso,2,mean)))+ sum(diag(var(lasso.est)))
}
RMSE.RES=mse.unres/mse.res
RMSE.PT=mse.unres/mse.pt
RMSE.SE=mse.unres/mse.shr
RMSE.PSE=mse.unres/mse.pshr
RMSE.LASSO= mse.unres/mse.lasso
result<-round(cbind(RMSE.RES,RMSE.PT,RMSE.SE,RMSE.PSE,RMSE.LASSO),3)
rownames(result) <- c("5","10","15")
xtable(result)
```


مراجع

- آرشی، محمد (۱۳۸۷). برآوردگرهای بهبودیافته در رگرسیون خطی چندگانه با خطای توزیع بیضی‌گون. پایان‌نامه دکترا، دانشگاه فردوسی مشهد.
- ربیعی، محمدرضا (۱۳۸۰). رگرسیون استوار و مقایسه‌ی آن با روش کمترین توان‌های دوم. پایان‌نامه کارشناسی‌ارشد، دانشگاه شهید بهشتی تهران.
- روزبه، مهدی (۱۳۹۰). برآورد در مدل‌های خطی جزئی. پایان‌نامه دکترا، دانشگاه فردوسی مشهد.
- عربی بلاغی، رضا (۱۳۹۳). برآوردگرهای بهبودیافته در رگردها. پایان‌نامه دکترا، دانشگاه فردوسی مشهد.
- کربلایی، محمدحسین (۱۳۹۶). بررسی عملکرد برآوردگر آزمون مقدماتی در مواجهه با برخی آزمون‌های آماری. پایان‌نامه دکترا، دانشگاه فردوسی مشهد.
- فلاح، رضا (۱۳۹۶). بررسی رفتار برخی برآوردگرهای انقباضی مدل رگرسیون خطی چندگانه. پایان‌نامه دکترا، دانشگاه فردوسی مشهد.
- نوروزی‌راد، م. و م. آرشی (۱۳۹۶). مطالعه رفتار حدی برآوردگرهای انقباضی در مدل رگرسیونی توانانیده با نرم مستطیلی. *مجله علوم آماری* ۱۱(۱۱)، ۱۴۹-۱۷۴.
- Ahmed, S. E. (1992). Shrinkage preliminary test estimation in multivariate normal distributions. *Journal of Statistical Computation and Simulation* 43, 177-195.
- Ahmed, S. E. (1994). Improved estimation of the coefficient of variation. *Journal of Applied Statistics* 21(6), 565-573.
- Ahmed, S. E. (1995). A pooling methodology for coefficient of variation. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 57(1), 57-75.

- Ahmed, S. E. (2001). Shrinkage estimation of regression coefficients from censored data with multiple observations. In S. E. Ahmed and N. Reid (Eds.), *Empirical Bayes and likelihood inference, lecture notes in statistics*, Volume 148, pp. 103–120. Springer.
- Ahmed, S. E. (2002). Simultaneous estimation of coefficients of variation. *Journal of Statistical Planning and Inference* 104(1), 31–51.
- Ahmed, S. E. (2014). *Penalty, Shrinkage and pretest estimation strategies: variable selection and estimation*. Springer.
- Ahmed, S. E., K. A. Doksum, S. Hossain, and J. You (2007). Shrinkage, pretest and absolute penalty estimators in partially linear models. *Australian and New Zealand Journal of Statistics* 49, 435–454.
- Ahmed, S. E., A. A. Hussein, and P. K. Sen (2006). Risk comparison of some shrinkage m-estimators in linear models. *Journal of Nonparametric Statistics* 18(4-6), 401–415.
- Ahmed, S. E. and C. J. Nicol (2012). An application of shrinkage estimation to the nonlinear regression model. *Computational Statistics and Data Analysis* 56(11), 3309–3321.
- Ahmed, S. E. and S. M. E. Raheem (2012). Shrinkage and absolute penalty estimation in linear models. *WIREs Computational Statistics* 4, 541–553.
- Ahmed, S. E. and A. K. M. E. Saleh (1989). Pooling multivariate data. *Journal of Statistical Computation and Simulation* 35, 209–226.
- Ahmed, S. E. and A. K. M. E. Saleh (1990). Estimation strategies for the intercept vector in a simple linear multivariate normal regression model. *Computational Statistics and Data Analysis* 10, 143–206.
- Ahmed, S. E. and A. K. M. E. Saleh (1993). Improved estimation for the component mean-vector. *Japan Journal of Statistics* 10, 177–195.
- Ahmed, S. E. and A. K. M. E. Saleh (1999). Improved nonparametric estimation of location vectors in multivariate regression models. *Journal of Nonparametric Statistics* 11(1-3), 51–78.
- Ahmed, S. E. and B. Ullah (1999). Improved biased estimation in an anova model. *Linear Algebra and its Applications* 289(1), 3–24.

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, pp. 267–281.
- Akdeniz, F. (2002). More on the pre-test estimator in ridge regression. *Communications in Statistics-Theory and Methods* 31, 987–994.
- Alheety, M. and B. M. G. Kibria (2014). A generalized stochastic restricted ridge regression estimator. *Communications in Statistics-Theory and Methods* 43, 4415–4427.
- Ali, A. M. and A. K. M. E. Saleh (1991). Preliminary test and empirical bayes to shrinkage estimation of regression parameters. *Journal of Japan Statistical Society* 21, 401–416.
- An, H.-Z. and Z.-G. Chen (1979). On convergence of lad estimates in autoregression with infinite variance. *Journal of Multivariate Analysis* 12, 335–345.
- Anderson, T. W. (1948). On the theory of testing serial correlation. *Skandinavisk Aktuarietidskrift* 31, 88–1162.
- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.
- Angrist, J. D. and J. S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Arabi, R., M. Arashi, and S. M. M. Tabatabaey (2015). Improved estimators of the distribution function based on lower record values. *Statistical Papers* 65, 453–477.
- Arashi, M. (2012). Preliminary test and stein estimators in simultaneous linear equations. *Linear Algebra and Its Applications* 436, 1195–1211.
- Arashi, M., S. Khan, S. M. M. Tabatabaey, and H. Soleimani (2013). Shrinkage estimation under multivariate elliptic models. *Communication in Statistics-Theory and Methods* 42, 2084–2103.
- Arashi, M., B. M. G. Kibria, M. Norouzirad, and S. Nadarajah (2014). Improved preliminary test and stein-rule liu estimators for ill-conditioned elliptical linear regression model. *Journal of Multivariate Analysis* 124, 53–74.
- Arashi, M., B. M. G. Kibria, and M. A. Tajadod (2015). On shrinkage estimators in matrix variate elliptical models. *Metrika* 78, 29–44.
- Arashi, M. and S. Nadarajah (2017). On singular elliptical models. *Communication in Statistics - Theory and Methods* 46(1), 247–258.

-
- Arashi, M. and M. Norouzirad (2015). Improved ridge m-estimators. In *46th Annual Iranian Mathematics Conference*, Yazd University, Yazd, Iran.
- Arashi, M. and M. Norouzirad (2016). Steinian shrinkage estimation in high dimensional regression. In *13th Iranian Statistical Inference Seminar on Nonparametric Statistics and its Application*, Shahid Bahonar University of Kerman, Kerman, Iran.
- Arashi, M. and M. Roozbeh (2015). Shrinkage estimation in system regression model. *Computational Statistics* 30, 359–376.
- Arashi, M., M. Roozbeh, and H. A. Niroumand (2012). A note on Stein-type shrinkage estimator in partial linear models. *Statistics* 65(5), 673–685.
- Arashi, M., A. K. M. E. Saleh, and S. M. M. Tabatabaey (2011). Estimation of parameter of parallelism model with elliptically distributed errors. *Metrika* 71(1), 79–100.
- Arashi, M. and S. M. M. Tabatabaey (2009). Improved variance estimation under sub-space restriction. *Journal of Multivariate Analysis* 100, 1752–1760.
- Arashi, M. and S. M. M. Tabatabaey (2010). A note on classical Stein-type estimators in elliptically contoured models. *Journal of Statistical Planning and Inference* 140, 1206–1213.
- Arashi, M., S. M. M. Tabatabaey, and M. Hassanzadeh Bastian (2014). Shrinkage ridge estimators in linear regression. *Communication in Statistics-Simulation and Computation* 43, 871–904.
- Armstrong, R. D. and E. L. Frome (1976). A comparison of two algorithms for absolute deviation curve fitting. *Journal of American Statistical Association* 71, 328–330.
- Arslan, O. (2012). Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression. *Computational Statistics and Data Analysis* 56, 1952–1965.
- Askin, R. G. and D. C. Montgomery (1980). Augmented robust estimation. *Technometrics* 22, 333–341.
- Bancraft, T. A. (1944). On biases in estimating due to use of preliminary tests of significance. *The Annals of Mathematical Statistics* 15, 195–204.
- Bancroft, T. A. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. *Biometrics* 20, 427–442.

- Bancroft, T. A. and C.-P. Han (1977). Inference based on conditional specification: A note and a bibliography. *International Statistical Review / Revue Internationale de Statistique* 45(2), 117–127.
- Bancroft, T. A. and C. P. Han (1983). A note of pooling variances. *Journal of the American Statistical Association* 78, 981–983.
- Barro, R. and J. W. Lee (1994). *Dataset for a panel of 138 countries*. Discussion paper.
- Barrodale, L. and F. D. K. Roberts (1973). An improved algorithm for discrete l1 linear approximation. *SIAM Journal on Numerical Analysis* 10(5), 839–848.
- Bassett, G. and R. Koenker (1978). Asymptotic theory of least absolute error regression. *Journal of American Statistical Association* 78, 618–622.
- Becker, R. A., J. M. Chambers, and A. R. Wilks (1988). *The New S Language*. Wadsworth & Brooks/Cole.
- Benda, N. (1996). Pre-test estimation and design in linear model. *Journal of Statistical Planning and Inference* 52(2), 225–240.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2 ed.). New York: Springer-Verlag.
- Bollen, K. A. and R. Jackman (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In J. Fox and J. S. Long (Eds.), *Modern Methods of Data Analysis*, pp. 257–291. Sage.
- Boscovich, R. (1757). De literaria expeditione per pontificiam ditionem et synopsis amplioris operis. *Bononiensi Scientiarum et Artum Instituto atqua Academia Commentarii* 4, 353–396.
- Brandwein, A. C. and W. E. Strawderman (1990, 08). Stein estimation: The spherically symmetric case. *Statistical Science* 5(3), 356–369.
- Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics* 24(6), 2350–2383.
- Brekson, J. (1942). Test of significance considered as evidence. *Journal of American Statistical Association* 37, 235–335.
- Brikes, D. and Y. Dodge (1993). *Alternative Methods of Regression*. John Wiley & Sons.

-
- Buhlmann, P. and S. van de Geer (2011). *Statistics for High Dimensional Data Methods: Methods, Theory and Applications*. Berlin: Springer Series in Statistics. Springer.
- Cai, T. and X. Shen (2011). *High-Dimensional Data Analysis*. Higher Education Press.
- Casella, G. (1985). An introduction to empirical bayes data analysis. *The American Statistician* 39, 83–87.
- Chen, K., Z. Ying, H. Zhang, and L. Zhao (2008). Analysis of least absolute deviation. *Biometrika* 95, 107–122.
- Clarke, J. A., D. E. A. Giles, and D. Wallace (1987). Preliminary-test estimation of the error variance in linear regression. *Economic Theory* 3, 299–304.
- Coakley, C. W. and T. P. Hettmansperger (1993). A bounded influence, high breakdown, efficient regression estimator. *Journal of American Statistical Association* 88, 872–880.
- Cook, R. D. and S. Weisberg (1984). *Residuals and Influence in Regression*. New York: Chapman and Hall.
- Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society. Series B (Methodological)* 45(3), 311–354.
- Croux, C., P. J. Rousseeuw, and O. Hossjer (1994). Generalized S-estimators. *Journal of American Statistical Association* 89, 1271–1281.
- de Laplace, P. S. (1818). *Deuxieme Supplement d La Throm Analytique des Probabilites*. Courcier.
- Don, F. J. (1982). Restrictions on variables. *Journal of Econometrics* 18, 369–393.
- Donoho, D. L. and P. J. Huber (1983). *The Notation of Break-down Point*. Wadsworth: in A Festchrift for E. L. Lehmann.
- Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *Annals of Statistics* 32, 407–499.
- Efron, B. and C. Morris (1975). Data analysis using stein's estimator and its generalizations. *Journal of the American Statistical Association* 70(350), 311–319.
- Eisenhart, C. (1961). Boscovitch and comination of observations. In L. L. Whyte (Ed.), *Roger Joseph Boscovich*. Forgham University Press.

- El-Salam, M. (2013). The efficiency of some robust ridge regression for handling multicollinearity and non-normals errors problems. *Applied Mathematical Sciences* 7, 3831–3846.
- Ellis, S. and S. Morgenthaler (1992). Leverage and breakdown in l_1 regression. *Journal of the American Statistical Association* 87, 143–148.
- Emiroglu, E. (2014). *Estimation of regression parameters as optimization application*. Ph.D. disseration , Istanbul University.
- Fallahpour, S., S. E. Ahmed, and K. A. Doksum (2012). L1 penalty and shrinkage estimation in partially linear models with random coefficient autoregressive errors. *Applied Stochastic Models in Business and Industry* 28(3), 236–250.
- Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348–1360.
- Finney, D. J. (1996). A note on the history of regression. *Journal of Applied Statistics* 23(5), 555–558.
- Fisher, R. A. (1956). *Statistical Methods in Scientific Inference*. Oliver and Boyd.
- Galton, F. (1877). Typical laws of heredity. *Nature* 15, 492–495, 512–514, 532–533.
- Gentle, J. E., W. J. Kennedy, and V. A. Sposito (1977). On least absolute deviations estimators. *Communication in Statistics, A* 6, 839–845.
- Gervini, D. and V. J. Yohai (2002). A class of robust and fully efficient regression estimators. *The Annals of Statistics* 30, 583–616.
- Gibbons, D. G. A. (1981). Simulation study of some ridge estimators. *Journal of the American Statistical Association* 76, 131–139.
- Giles, A. J. (1991). Pretesting for linear restrictions in a regression model with spherically symmetric distributions. *Journal of Econometrics* 50, 377–398.
- Giles, J. A. (1992). Estimation of the error variance after a preliminary-test of homogeneity in a regression model with spherically disturbances. *Journal of Econometrics* 53, 345–361.
- Giloni, A., J. Simonoff, and B. Sengupta (2006). Robust weighted LAD regression. *Computational Statistics and Data Analysis* 50, 3124–3140.

- Gross, S. and W. L. Steiger (1979). Least absolute deviation estimates in autoregression with infinite variance. *Journal of Applied Probability* 16, 104–116.
- Gruber, M. H. J. (1998). *Improving Efficiency by Shrinkage the James-Stein and Ridge Regression Estimators*. Springer Verlag.
- Gujrati, D. N. (2004). *Basic Econometrics* (4th ed.). Tata McGraw- Hill.
- Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley.
- Han, C. P. and T. A. Bancroft (1968). On pooling means when variance is unknown. *Journal of the American Statistical Association* 63, 1333–1342.
- Handschin, E., J. Kohlas, A. Fiechter, and F. Schweppe (1975). Bad data analysis for power system state estimation. *IEEE Transactions on Power Apparatus and Systems* 2, 329–337.
- Hansen, B. E. (2016). The risk of james stein and lasso shrinkage. *Econometric Reviews* 35, 1456–1470.
- HassanzadehBashtian, M., M. Arashi, and S. M. M. Tabatabaey (2011a). Ridge estimation under the stochastic restriction. *Communication in Statistics-Theory and Methods* 40, 3711–3747.
- HassanzadehBashtian, M., M. Arashi, and S. M. M. Tabatabaey (2011b). Using improved estimation strategies to combat multicollinearity. *Journal of Statistical Computation and Simulation* 81(12), 1773 – 1797.
- Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference and prediction (2nd ed)*. Springer.
- Hettmansperger, T. and J. McKean (1998). *Robust Nonparametric Statistical Methods*. London: Arnold.
- Hoaglin, D. C. and R. E. Welsch (1978). The hat matrix in regression and anova. *The Annals of Statistics* 32, 17–22.
- Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* 12, 55–67.
- Hoerl, A. E., R. W. Kennard, and K. Baldwin (1975). Ridge regression: some simulations. *Communications in Statistics* 4, 105–123.

- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: a tutorial. *Statistical Science* 14, 382–417.
- Holland, P. W. and R. E. Welsch (1977). Robust regression using iteratively re-weighted least-squares. *Communications in Statistics. Theory and Methods* 6(9), 813–827.
- Hossain, S., M. Norouzirad, and M. Arashi (2017). Estimation strategies for weighted least absolute deviations regression models. In *45th Annual Meeting of the Statistical Society of Canada*, Winnipeg, Canada.
- Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 35(1), 73–101.
- Huber, P. J. (1968). Robust confidence limits. *Z. Wahrscheinlichk. verw. Geb* 10, 269–278.
- Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *The Annals of Statistics* 1, 799–821.
- Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons.
- Hubert, M. and P. Rousseeuw (1997). Robust regression with both continuous and binary regressors. *Journal of Statistical Planning and Inference* 57, 153–163.
- Hurvish, C. M. and C. L. Tsai (1990). Model selection for least absolute deviations regression in small samples. *Statistics and Probability letters* 9, 259–265.
- Inoue, T. (2001). Improving the "hkb" ordinary type ridge estimators. *Journal of Japan Statistical Society* 31, 67–83.
- Jadhav, N. H. and D. N. Kashid (2011). A jackknifed ridge m-estimator for regression model with multicollinearity and outliers. *J. Stat. Theory Pract.* 5, 659–673.
- Jadhav, N. H. and D. N. Kashid (2014). Robust winsorized shrinkage estimators for linear regression models. *Journal of Modern Applied Statistical Methods* 13, 131–150.
- Jaeckel, L. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *The Annals of Mathematical Statistics* 43, 1449–1458.
- James, G. M., C. Paulson, and P. Rusmevichientong (2012). The constrained lasso. Manuscript.

- James, W. and C. Stein (1961). Estimation with quadratic loss. In *the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 361–379. CA: University of California Press.
- Johnson, R. W. (1996). Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education* 4, 1.
- Judge, G. G. and M. E. Bock (1978). *The statistical implications of pre-test and Stein-rule estimators in econometrics*. North-Holland Publishing Company.
- Jureckova, J. (1977). Asymptotic relations of M-estimators and R-estimators in linear model. *Annals of Statistics* 5, 664–672.
- Jureckova, J. and P. K. Sen (1996). *Robust Statistical Procedures*. Wiley.
- Kariya, T. and H. Kurata (2004). *Generalized Least Squares*. New York: Wiley series in probability and statistics.
- Khan, S. and S. E. Ahmed (2003). Improved estimation of coefficient vector in a regression model. *Communications in Statistics-Simulation and Computation* 32(3), 747–769.
- Khan, S. and A. K. M. E. Saleh (1997). Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate student-*t* errors. *Biometrical Journal* 39, 1–17.
- Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics - Simulation and Computation* 32, 419–435.
- Kibria, B. M. G., K. Mansson, and G. Shukur (2015). A simulation study of some biasing parameters for the ridge type estimation of poisson regression. *Communications in Statistics-Simulation and Computation* 44, 943–957.
- Kibria, B. M. G. and A. K. M. E. Saleh (2003). Effect of w, lr, and lm tests on the performance of preliminary test ridge regression estimators. *Journal of Japan Statistical Society* 33, 119–136.
- Kibria, B. M. G. and A. K. M. E. Saleh (2004). Preliminary test ridge regression estimators with students *t* errors and conflicting test-statistics. *Metrika* 59, 105–124.
- Kleyn, J., M. Arashi, A. Bekker, and S. Millard (2017). Preliminary testing of the cobb–douglas production function and related inferential issues. *Communications in Statistics - Simulation and Computation* 46(1), 469–488.

- Knight, K. (1998). Limiting distributions for l_1 regression estimators under general conditions. *The Annals of Statistics* 26, 755–770.
- Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics* 28, 1356–1378.
- Koenker, R. and G. Bassett (1982). Tests of linear hypotheses and l_1 estimation. *Econometrica* 50, 1577–1584.
- Koenker, R. and J. A. F. Machado (1999). Goodness of fit and related inference process for quantile regression. *Journal of the American Statistical Association* 94, 1296–1310.
- Kulperger, R. J. and S. E. Ahmed (1992). A bootstrap theorem for a preliminary test estimator. *Communications in Statistics - Theory and Methods* 21(7), 2071–2082.
- Kurata, H. (1999). On the efficiencies of the several generalized least squares estimators in a seemingly unrelated regression and a heteroscedastic model. *Journal of Multivariate Analysis* 70(1), 86–94.
- Lawrence, K. D. and L. C. Marsh (1984). Robust ridge estimation methods for predicting us coal mining fatalities. *Communications in Statistics-Theory and Methods* 13, 139–149.
- Luo, J. and Y. J. Zuo (2011). A new test for large dimensional regression coefficients. *Open Journal of Statistics* 1, 212–216.
- Mallows, C. L. (1973). Some comments on cp. *Technometrics* 15, 661–675.
- Mallows, C. L. (1975). *On Some topics in robustness*. Murray Hill: Bell Telephone Laboratories.
- Malthouse, E. C. (1999). Shrinkage estimation and direct marketing scoring model. *Journal of Interactive Marketing* 13, 10–23.
- Mansson, K., G. Shukur, and B. Kibria (2010). A simulation study of some ridge regression estimators under different distributional assumptions. *Communications in Statistics-Simulation and Computation* 39, 1639–1670.
- Marrona, R. A., D. R. Martin, and V. J. Yohai (2006). *Robust Statistics, Theory and Methods*. Wiley.
- Martin, S. B. (1998). Less than the least: An alternative method of least squares linear regression. Master's thesis, McMurry University.

- McDonald, G. C. and D. I. Galarneau (1975). A monte carlo evaluation of some ridge-type estimators. *Journal American Statistics Association* 70, 407–416.
- Miller, A. (2002). *Subset Selection in Regression* (2 ed.). Abingdon: CRC Press.
- Montgomery, D. C. and R. G. Askin (1981). Problems of nonnormality and multicollinearity for forecasting methods based on least squares. *A I I E Transactions* 13, 102–115.
- Montgomery, D. C., E. A. Peck, and G. G. Vinig (2011). *Introduction to Linear Regression Analysis* (3rd ed.). John Wiley & Sons.
- Muniz, G. and B. M. G. Kibria (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics—Simulation and Computation* 38, 621–630.
- Muniz, G., B. M. G. Kibria, K. M. Mansson, and G. Shukur (2012). On developing ridge regression parameters: a graphical investigation. *SORT* 36, 115–138.
- Najarian, S., M. Arashi, and B. M. G. Kibria (2013). A simulation study on some restricted ridge regression estimators. *Communications in Statistics-Simulation and Computation* 42, 871–890.
- Norouzirad, M. and M. Arashi (2016). From Lad-LASSO to PR-LAD estimator. In *Ind Seminar on Nonparametric Statistics and its Application*, Allame Tabatabaeyi University, Tehran, Iran.
- Norouzirad, M. and M. Arashi (2017a). Preliminary test and stein-type shrinkage lasso-based estimators. *SORT-Statistics and Operations Research Transactions*, Accepted.
- Norouzirad, M. and M. Arashi (2017b). Preliminary test and stein-type shrinkage ridge estimators in robust regression. *Statistical Papers*, DOI: 10.1007/s00362-017-0899-3.
- Norouzirad, M. and M. Arashi (2017c). Supplemental materials: Preliminary test and stein-type shrinkage ridge estimators in robust regression. DOI: 10.1007s00362-017-0899-3.
- Norouzirad, M., M. Arashi, and S. E. Ahmed (2017a). Asymptotic properties of shrinkage lad estimators. In *11th Seminar on Probability and Stochastic Processes*, Qazvin, Iran.
- Norouzirad, M., M. Arashi, and S. E. Ahmed (2017b). Improved robust ridge m-estimation. *Journal of Statistical Computations and Simulation* 87(18), 3469–3490.
- Norouzirad, M., M. Arashi, and A. K. M. E. Saleh (2015a). Shrinkage estimation through convex optimization. In *8th International Conference of the Iranian Society of Operations research*, Ferdowsi University, Mashhad, Iran.

- Norouzirad, M., M. Arashi, and A. K. M. E. Saleh (2015b). Some asymptotic results on improved lasso estimators. In *10th Seminar on Probability and Stochastic Process*, Yazd University, Yazd, Iran.
- Norouzirad, M., S. Hossain, and M. Arashi (2018). Shrinkage and penalized estimators in weighted least absolute deviations regression models. *Journal of Statistical Computation and Simulation* 88(8), 1557–1575.
- Osorio, F. (2015). *Routines for L1 estimation*. R package version 0.3.
- Pearson, K. (1897). Mathematical contributions to the theory of evolution.- on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London* 60, 489–498.
- Pearson, K. (1901). On lines and planes of closest fit to system of points in space. *The London, Edinburg and Dublin Philosophical Magazine and Journal of Science* 2, 559–572.
- Penrose, K., A. Nelson, and A. Fisher (1985). Generalized body composition prediction equation for men using simple measurement techniques. *Medicine and Science in Sports and Exercise* 17, 189.
- Pfaffenberger, R. C. and T. E. Dielman (1985). A comparison of robust ridge estimators. In *Proceedings of the American Statistical Association Business and Economic Statistics Section*, pp. 631–635.
- Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
- Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
- Raheem, S. M. (2012). *Absolute Penalty and Shrinkage Estimation Strategies in Linear and Partially Linear Models*. Ph.D. dissertation , University of Windsor, Canada.
- Raheem, S. M., S. Ahmed, and K. Doksum (2012). Absolute penalty and shrinkage estimation in partially linear models. *Computational Statistics and Data Analysis* 56, 874–891.
- Raheem, S. M. E. and S. E. Ahmed (2011). Positive-shrinkage and pretest estimation in multiple regression: a monte carlo study with applications. *Journal of Iranian Statistical Society* 10, 267–289.

- Rao, R. C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* 44(1), 50–57.
- Ravishanker, N. and D. K. Dey (2002). *A First Course in Linear Model Theory*. Chapman and Hall/CRC.
- Relles, D. A. (1968). *Robust regression by modified least squares*. Ph.D. dissertation, Yale University.
- Robbins, H. (1983). Some thoughts on empirical bayes estimation. *The Annals of Statistics* 11(3), 713–723.
- Roohbeh, M. (2015). Shrinkage ridge estimator in semiparametric regression models. *Journal of Multivariate Analysis* 136, 56–74.
- Roohbeh, M. and M. Arashi (2013). Feasible ridge estimator in partially linear models. *Statistics* 116, 35–44.
- Roohbeh, M. and M. Arashi (2016). Shrinkage ridge regression in partial linear models. *Communications in Statistics Theory and Methods* 45, 6022–6044.
- Roohbeh, M., M. Arashi, and H. A. Niroumand (2011). Ridge regression methodology in partial linear models with correlated errors. *Journal of Statistical Computation and Simulation* 81, 517–528.
- Rousseeuw, P. (1984). Least median of squares regression. *Journal of American Statistical Association* 79, 871–880.
- Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, and W. Wertz (Eds.), *Mathematical Statistics and its Applications (Vol. B)*, pp. 283–297. Riedel: Dordrecht.
- Rousseeuw, P. and V. Yohai (1984). Robust regression by means of s-estimators. In *Robust and Nonlinear Time Series Analysis: Proceedings of a Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", Heidelberg 1983*, pp. 256–272. New York, NY: Springer US.
- Rousseeuw, P. J. and A. M. Leroy (1987). *Regression Analysis and outlier detection*. John Wiley & Sons.

- Rukhin, A. L. (1995). Admissibility: survey of concept in progress. *International Statistical Review* 63, 95–115.
- Saleh, A. K. M. E. (2006). *Theory of Preliminary Test and Stein-Type Estimation with Applications*. John Wiley & Sons.
- Saleh, A. K. M. E., M. Arashi, and S. M. M. Tabatabaey (2014). *Statistical Inference for Models with Multivariate t-Distributed Errors*. John Wiley.
- Saleh, A. K. M. E. and C. P. Han (1990). Shrinkage estimation in regression analysis. *Estadística* 42, 40–43.
- Saleh, A. K. M. E. and B. M. G. Kibria (1993). Performance of some new preliminary test ridge regression estimators and their properties. *Communications in Statistics-Theory and Methods* 22, 2747–2764.
- Saleh, A. K. M. E. and B. M. G. Kibria (2011). On some ridge regression estimators: a nonparametric approach. *Journal of Nonparametric Statistics* 23, 819–851.
- Saleh, A. K. M. E. and P. K. Sen (1978). Non parametric estimation of location parameter after a preliminary test on regression. *The Annals of Statistics* 6, 154–168.
- Saleh, A. K. M. E. and P. K. Sen (1985). On shrinkage estimators of location parameters. *Communications in Statistics-Theory and Methods* 14, 2313–2329.
- Saleh, A. K. M. E. and T. Shiraishi (1989). On some R and M estimators of regression parameters under uncertain restriction. *Journal of the Japan Statistical Society* 19, 129–137.
- Samkar, H. and O. Alpu (2010). Ridge regression based on some robust estimators. *Journal of Modern Applied Statistical Methods* 9, 495–501.
- Sarker, N. (1992). A new estimator combining the ridge regression and the restricted least squares method of estimation. *Communications in Statistics-Theory and Methods* 21, 1987–2000.
- Schrader, R. M. and T. P. Hettmansperger (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika* 67, 93–101.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6, 461–464.

- Sclove, S. L., C. Morris, and R. Radhakrishnan (1972). Non-optimality of preliminary test estimators for the mean of a multivariate normal distribution. *Annals of Mathematics and Statistics* 43, 1481–1490.
- Searle, S. R. (1982). *Matrix ALgebra useful for Staistics*. John Wiley.
- Sen, P. K. (1986). On the asymptotic distributional risk shrinkage and preliminary test version of maximum likelihood estimators. *Sankhya A* 48, 354–371.
- Sen, P. K. and A. K. M. E. Saleh (1987). On preliminary and shrinkage M-estimation in linear model. *The Annals of Statistics* 15, 1580–1592.
- Sengupta, D. and S. R. Jammalamadaka (2003). *Linear Models: An Integrated Approach*. World Scientific Publishing Company.
- Shalabh (1998). Improved estimation in measurement error models through stein-rule procedure. *Journal of Multivariate Analysis* 67, 35–48.
- Shalabh (2001). Pitman closeness comparison of least squares and stein-rule estimators in linear regression models with non-normal disturbances. *The American Journal of Mathematical and Management Sciences* 21, 89–100.
- Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika* 69, 242–244.
- Silvapull, M. J. (1991). Robust ridge regression based on an m estimator. *Australian Journal of Statistics* 33, 319–333.
- Simpson, J. R. (1995). *New methods and comparable evaluations for robust and biased-robust regression estimation*. Ph.D. disseration , Arizona state university.
- Singer, J. M. and P. K. Sen (1985). M-methods in multivariate linear models. *Journal of Multivariate Analysis* 17, 168–184.
- Singh, R. S. (1991). James-stein rule estimators in linear regression models with multivariate t distributed error. *Australian and New Zealand Journal of Statistics* 33, 145–158.
- Snedecor, G. W. (1938). *Statistical Methods*. Iowa: Collegiate Press.
- Srivastava, V. K. and D. E. A. Giles (1987). *Seemingly Unrelated Regression Equation Models: Estimation and Inference*. New York: Marcel Dekker.

- Stanton, J. M. (2001). Galton, pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education* 9(3), null.
- Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Berkeley, Calif., pp. 197–206. University of California Press.
- Stigler, A. M. (1973). Laplace, fisher, and the discovery of the concept of sufficiency. *Biometrika* 60, 439–445.
- Stigler, S. M. (1990, 02). The 1988 neyman memorial lecture: A galtonian perspective on shrinkage estimators. *Statistical Science* 5(1), 147–155.
- Susanti, Y., H. Pratiwi, S. Sulistijowati, and T. Liana (2014). M estimation, S estimation, and MM estimation in robust regression. *International Journal of Pure and Applied Mathematics* 91, 349–360.
- Tabakan, G. and F. Akdeniz (2010). Difference-based ridge estimator of parameters in partial linear model. *Statistical Papers* 51, 357–368.
- Tabatabaey, S. M. M. (1995). *Preliminary test approach estimation: regression model with spherically symmetric errors*. Ph.D. dissertation , Carleton University, Canada.
- Tabatabaey, S. M. M., A. K. M. E. Saleh, and B. G. M. Kibria (2004). Simultaneous estimation of regression parameters with spherically symmetric errors under possible stochastic constraints. *International Journal of Statistical Sciences* 3, 1–20.
- Talbert, M., J. Richards, P. Mielke, and B. Cade (2015). *Blossom: Statistical Comparisons with Distance-Function Based Permutation Tests*. R package version 1.3.
- Taylor, W. E. (1978). The heteroscedastic linear model: exact finite sample results. *Econometrica* 46, 663–675.
- Theil, H. (1971). *Principles of Econometrics*. New York: Wiley.
- Thompson, J. C. (1968). Some shrinkage techniques for estimating the mean. *Journal of American Statistical Association* 63, 113–122.

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58, 267–288.
- Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Doklady Akademii Nauk SSSR* 83, 501–504. Translated in *Soviet Mathematics* 4: 1035–1038.
- Tofallis, C. (2008). Least squares percentage regression. *Journal of Modern Applied Statistical Methods* 7, 526–534.
- Torehzade, S. and M. Arashi (2014). A note on shrinkage wavelet estimation bayesian analysis. *Statistics and Probability Letters* 84, 231–234.
- Touloumis, A. (2015). Nonparametric stein-type shrinkage covariance matrix estimators in high dimensional setting. *Computational Statistics & Data Analysis* 83, 251–261.
- Wang, D. Q., S. Chukova, and C. D. Lai (2004). On the relationship between regression analysis and mathematical programming. *Journal of Applied Mathematics and Decision Sciences* 8, 131–140.
- Wang, H., G. Li, and G. Jiang (2007). Robust regression shrinkage and consistent variable selection through the lad-lasso. *Journal of Business & Economic Statistics* 25, 347–355.
- Yao, Y. (2004). *Statistical Applications of Linear Programming for Feature Selection Via Regularization Methods*. Ph.D. dissertation , The Ohio State University.
- Yohai, V. J. (1987). High breakdown point and high efficiency estimates for regression. *The Annals of Statistics* 15, 642–656.
- Yu, C. and W. Yao (2017). Robust linear regression: A review and comparison. *Communications in Statistics - Simulation and Computation* 46(8), 6261–6282.
- Yuzbashi, B., S. E. Ahmed, and M. Gungor (2017). Pretest and stein-type estimators in quantile regression model. arXive preprint arXive:1707.03820.
- Yuzbasi, B. and S. E. Ahmed (2015). Shrinkage ridge regression estimators in high dimensional linear models. In *Proceeding of the ninth international conference on management science and engineering management*, pp. 793–807.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression and test for aggregation bias. *Journal of the American Statistical Association* 58, 977–992.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101, 1418–1429.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* 67(2), 301–320.

واژه‌نامه فارسی به انگلیسی

Preliminary test	آزمون اولیه
Interaction	اثرهای متقابل
Trace	اثر ماتریس
Bias	اریبی
Robust	استوار
Uncertain Prior Information	اطلاعات پیشین غیرقطعی
Nonsample Information	اطلاعات غیرنمونه‌ای
Cross Validation	اعتبارسنجی متقابل
Partition	افراز
Euclidean	اقلیدسی
Shrinkage	انقباضی
Dominant	برتری
Quadratic programming	برنامه‌ریزی درجه‌ی دوم
Improved	بهبودیافته
Distrust	بی‌اعتمادی
Empirical Bayes	بیز تجربی
Overshrinkage	بیش انقباضی
Assumption	پذیره
Outlier	پرت
Prediction	پیشگویی
Score function	تابع امتیاز
Indicator function	تابع نشان‌گر
Empirical	تجربی
Rank factorization	تجزیه رتبه‌ای
Spectral decomposition	تجزیه طیفی
Violation	تخطی

Loner order	ترتیب لونر
Dominant	تسلط
Augment	تقویت
Sparsity	تنکی
Penalized	جریمه‌شده
Pooled	جمع‌شده
Partial	جزئی
Oracle property	خاصیت پیش‌گویی
Idempotent	خودتوان
Autocorrelation	خودهمبستگی
Likelihood	درست‌نمایی
tail	دم
Full rank	رتبه کامل
Approach	رهیافت
Fisher recipe	رهیافت فیشر
Square root factorization	ریشه دوم
Characteristic roots	ریشه‌های مشخصه
Loss	زیان
Consistent	سازگار
Simulation Monte Carlo	شبیه‌سازی مونت کارلو
Penalty factor	ضریب جریمه
Inadmissible	غیرمجاز
Hypothesis	فرضیه
Auxiliary	فرعی
Quadratic form	فرم درجه دوم
Breakdown	فروریزش
Extreme	فرین
Redundancy	فزونگی
Diagonal	قطری
Efficient	کارا
Stepwise	گام به گام
log-concave	لگاریتم-محدب
Fisher information matrix	ماتریس اطلاع فیشر
Orthogonal	متعامد

Response Variable	متغیر پاسخ
Predictor Variable	متغیر پیشگو یا توضیحی
Independent Variable	متغیر مستقل
Symmetric	متقارن
Schur complement	متمم شور
Admissible	مجاز
Restricted	محدودشده
Unrestricted	محدودنشده
Restriction	محدودیت
Risk	مخاطره
Positive Definite	معین مثبت
Eigenvalue	مقدار ویژه
Regularization	منظم‌سازی
Unbiased	نااریبی
Nonparametric	ناپارامتری
Inconsistent	ناسازگار
Inefficient	ناکارا
Nonsingular	نامنفرد
Heteroscedastic	ناهمپراش
Positive semidefinite	نیمه‌معین مثبت
Serial correlation	همبستگی ترتیبی
Homoscedastic	همپراش یا هم‌واریانس
Collinearity	همخطی
Exact Collinearity	همخطی کامل
Partial Collinearity	همخطی ناقص

Abstract

In this thesis, the uncertain prior (non-sample) information has been used in penalized linear (ridge and LASSO) and robust penalized linear (M ridge) models, which are resulted in preliminary test, shrinkage Stein-type and its positive rule estimators. The characteristics of these estimators have been derived by considering some linear restrictions on the parameter space. Also, the available auxiliary information has helped the estimation procedure to obtain shrinkage estimators, when outliers and leverage points in dataset are present. In all above models, the achieved results are validated by some simulation studies and real data analyses. The results confirm the introduced shrinkage estimators outperform the natural estimators in the aforementioned specific models.

Keywords: Multiple Linear regression models, LASSO, Preliminary test estimator, Ridge, Robust estimator, Shrinkage Estimator.



Shahrood University of Technology

Faculty of Mathematical Sciences

Ph.D. Thesis in Statistics

**Improved Estimators in some penalized
linear regression models**

By: Mina Norouzirad

Supervisor

Dr. Mohammad Arashi

March 2017