

الحمد لله
الذي هدانا لهذا
الذي كنا لنهتدي لولا
هدى الله لنا



دانشکده علوم ریاضی

رشته آمار، گرایش آمار ریاضی

پایان نامه کارشناسی ارشد

مقایسه رفتار برخی برآوردهای انقباضی بریج در مدل رگرسیون چندگانه

نگارنده: محمد آرست

استادان راهنما

دکتر محمد آرشی
دکتر محمد رضا ربیعی

بهمن ۱۳۹۵

تقدیم خالصانه به پدرم

وزین ترین کلمه در دیباچه، مستی، تقدیم به او که هر چه دارم مدیون زحمات بی دریغش است. کاش می توانستم ذره ای از گذشته و فداکاری اش را جبران نمایم.

تقدیم عاشقانه به مادرم

سکفت ترین زاده می عشق الهی، کوهری بی همتا در بحری کران فداکاری ها، کسی که دامن پر مهرش پناهگاه دوران کودکی ام و قلب پاکش سرچشمه دعای خیر در زندگی ام بوده است.

سپاس‌گزاری...

با سپاس از سه وجود مقدس:
آنان که ناتوان شدند تا ما به توانایی برسیم...
موهایشان سپید شد تا ما روسفید شویم...
و عاشقانه سوختند تا گرمابخش وجود ما و روشن‌گر راهمان باشند...
پدرانمان
مادرانمان
استادانمان

برخود واجب می‌دانم مراتب سپاس و قدردانی عمیق قلبی خود را به خدمت استاد فرزانه و گران‌قدرم جناب آقای دکتر محمد آرشی که در طول دوران تحصیل و ارائه پایان‌نامه از چشمه جوشان دانش و اخلاق والایشان بهره‌مند گشته‌ام، ابراز نمایم.
و از استاد راهنمای صبور و ارجمند، جناب آقای دکتر محمد رضا ربیعی، که زحمت مشاوره و راهنمایی این پایان‌نامه را برعهده داشتند کمال تشکر را دارم.
همچنین از اساتید بزرگوار و دلسوز دکتر داود شاهسونی و دکتر حسین باغی‌شنی که زحمت داوری این رساله را متقبل شدند؛ کمال تشکر و قدردانی را دارم.
در انتها جا دارد از همه دوستان و عزیزانی که در این مدت مرا یاری و راهنمایی نمودند، قدردانی نمایم.

محمد آرست

بهمن ۱۳۹۵

تعهد نامه

اینجانب **محمد آرست** دانشجوی کارشناسی ارشد رشته **آمار علوم ریاضی** دانشگاه صنعتی شاهرود، نویسنده پایان نامه با عنوان **مقایسه رفتار برخی برآوردگرهای انقباضی بریج در مدل رگرسیون چندگانه**، 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.

محمد آرست

بهمن ۱۳۹۵

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

با وجود مشکل همخطی چندگانه برآوردگر کمترین توان‌های دوم کاربردی نبوده و اغلب از برآوردگرهای انقباضی اریب استفاده می‌شود. روش‌های متعددی برای به‌دست آوردن برآوردگرهای انقباضی ضرایب رگرسیون مطرح شده است. یکی از این روش‌ها استفاده از برآوردگر بریج است که حالت خاص آن برآوردگرهای ریج و لاسو می‌باشند.

معمولاً در مسائل کاربردی و مدل‌های اقتصادی، لازم است از اطلاعات اولیه در خصوص بعضی از پارامترهای مورد علاقه در مدل آماری تحت بررسی استفاده شود. برآوردگری که بر این اطلاعات اولیه، به‌دست می‌آید، برآوردگر محدود شده نامیده می‌شود.

در این پایان‌نامه رفتار مجانبی برآوردگر بریج و بریج محدود شده را مورد مطالعه قرار می‌دهیم. از آنجایی که این برآوردگر صورت بسته‌ای ندارد، با استفاده از الگوریتم نیوتن – رافسون برآوردگر بریج را به صورت عددی محاسبه می‌کنیم. خواص دو حالت خاص، برآوردگرهای ریج و لاسو نیز مورد بررسی قرار گرفته‌اند.

کلمات کلیدی: همخطی چندگانه، روش انقباضی، برآوردگر بریج، برآوردگر ریج، برآوردگر لاسو، محدودیت خطی

لیست مقالات مستخرج از پایان نامه

۱. آرست م.، آرشی م. و ربیعی م. (۱۳۹۵) برآوردگر انقباضی بریج محدود شده در مدل رگرسیون چندگانه، سیزدهمین کنفرانس آمار ایران – دانشگاه شهید باهنر کرمان.
۲. آرست م.، آرشی م. و ربیعی م. (۱۳۹۵) مطالعه رفتار برآوردگر انقباضی تحت یک قید خطی در مدل رگرسیون جریمه شده، مجله علوم آماری (تحت داوری).

فهرست مطالب

فهرست تصاویر

ف

فهرست جداول

ق

۱ مقدمه و تاریخچه

۱

۱.۱ مقدمه ۱

۱

۲.۱ مدل رگرسیونی چندگانه ۳

۳

۱.۲.۱ برآورگر کمترین توان‌های دوم ۳

۳

۳.۱ همخطی در رگرسیون چندگانه ۷

۷

۱.۳.۱ اثرات همخطی بر برآورد پارامترها ۸

۸

۲.۳.۱ چگونه می‌توان به وجود همخطی چندگانه پی برد؟ ۹

۹

۳.۳.۱ روش‌هایی برای برخورد با همخطی چندگانه ۱۱

۱۱

۲ برآوردگر بریج

۱۳

۱.۲ مقدمه ۱۳

۱۳

۲.۲ برآورگر انقباضی بریج ۱۴

۱۴

۱.۲.۲ حل مسئله بهینه‌سازی رگرسیون انقباضی بریج ($\gamma \geq 1$) ۱۵

۱۵

۲.۲.۲ الگوریتم محاسبه برآوردگر بریج ۱۷

۱۷

۳.۲.۲ واریانس برآوردگر بریج ۱۹

۱۹

۴.۲.۲ انتخاب پارامتر انقباضی γ و پارامتر کنترل k ۲۰

۲۰

۳.۲ مطالعه رفتار مجانبی ۲۱

۲۱

۳ حالات خاص برآوردگر بریج

۲۹

۱.۳ مقدمه ۲۹

۲۹

۲.۳ برآوردگر ریج ۳۰

۳۰

۳.۳ برآوردگر لاسو ۳۲

۳۲

۱.۳.۳ واریانس برآوردگر لاسو ۳۶

۳۶

۴.۳ محاسبات با R ۳۶

۳۶

۳۶	 بسته glmnet	۱.۴.۳
۴۱	 بسته lasso2	۲.۴.۳
۴۳	۴ برآوردگر رگرسیونی بریج محدود شده	
۴۳	 مقدمه	۱.۴
۴۴	 برآوردگر بریج با محدودیت ثابت	۲.۴
۴۶	 خطای برآوردگر انقباضی بریج محدود شده	۱.۲.۴
۵۱	 برآوردگر بریج با محدودیت تصادفی	۳.۴
۵۲	 برآوردگر بریج محدود شده	۱.۳.۴
۵۴	 خطای برآوردگر انقباضی بریج محدود شده	۲.۳.۴
۵۵	۵ مطالعات عددی	
۵۵	 شبیه‌سازی	۱.۵
۵۵	 مقدمه	۱.۱.۵
۶۷	 کاربرد	۲.۵
۶۷	 داده‌های سرطان پروستات	۱.۲.۵
۶۸	 داده‌های میزان رشد گندم	۲.۲.۵
۶۹	 مزیت‌های روش نیوتن رافسون اصلاح شده و الگوریتم پیش‌برنده لاسو	۳.۵
۷۱	آ خلاصه و پیشنهادات برای تحقیقات آینده	
۷۱	 خلاصه	۱.آ
۷۲	 پیشنهادات برای پژوهش‌های آینده (ارائه زمینه‌های تحقیق)	۲.آ
۷۳	ب	
۷۳	 جبر ماتریس‌ها	۱.ب
۷۶	 قضیه‌های حدی	۲.ب
۷۸	 همگرایی	۳.ب
۷۹	 برخی اصول آنالیز ریاضی	۴.ب
۸۰	 برخی اصول آنالیز عددی	۵.ب
۸۱	 روش دلتا	۱.۵.ب
۸۱	 روش نیوتن رافسون	۲.۵.ب
۸۳	پ	
۸۳	 عامل تورم واریانس	۱.پ
۸۵	ت	
۸۵	 عدد شرطی	۱.ت

۸۷	ث
۸۹	ج
۸۹	ج.۱ برنامه‌های کامپیوتری با نرم افزار R
۸۹	ج.۲ برنامه داده‌های سرطان پروستات
۹۱	ج.۳ برنامه مربوط به سناریو ۱.۱.۱
۹۳	ج.۴ محاسبه ضرایب ریج
۱۰۹	مراجع
۱۱۳	واژه‌نامه فارسی به انگلیسی
۱۱۵	واژه‌نامه انگلیسی به فارسی

فهرست تصاویر

۱۸	نحوه‌ی برخورد تابع $d-$ (خط پر) با خط S_j (نقطه چین) در معادله‌ی (۱۰.۲) برای $\gamma \geq 1$. . .	۱.۲
۳۳	عملکرد رگرسیون بریج در حالت $t = 1, p = 2$	۱.۳
۳۳	نواحی انقباض مختلف در رگرسیون بریج.	۲.۳
۳۵	نحوه‌ی برخورد تابع $f_2(\beta)$ (خط پر) با تابع $f_1(\beta)$ (خط چین) در معادله‌ی (۴.۳)	۳.۳
۳۸	نمودار ضرایب لاسو در مقابل $s = \frac{t}{\sum_j \beta_j^{OLS} }$ و لگاریتم λ	۴.۳
۳۹	نمودار MSE در مقابل لگاریتم λ	۵.۳
۳۹	نمودار ضرایب ریج در مقابل s و لگاریتم λ	۶.۳
۴۰	نمودار MSE در مقابل لگاریتم λ	۷.۳
۷۹	نمودار تابع محدب	۱.ب
۸۲	روش نیوتن رافسون	۲.ب

فهرست جداول

۳۸	توصیف متغیرهای مجموعه داده پروستات	۱.۳
۳۸	آماره‌های توصیفی متغیر پاسخ در مجموعه داده پروستات	۲.۳
۴۰	ضرایب داده‌های پروستات در برآوردگر ریدج	۳.۳
۴۰	ضرایب داده‌های پروستات در برآوردگر لاسو	۴.۳
۴۱	ضرایب داده‌های پروستات در برآوردگر لاسو	۵.۳
۴۲	نتایج ضرایب لاسو	۶.۳
۵۶	نتایج سناریو ۱.۱.۵ ($p = 4, n = 5, \nu = 0.7$) با نقطه شروع OLS	۱.۵
۵۶	نتایج سناریو ۱.۱.۵ ($p = 4, n = 5, \nu = 0.9$) با نقطه شروع OLS	۲.۵
۵۶	نتایج سناریو ۱.۱.۵ ($p = 4, n = 5, \nu = 0.9$) با نقطه شروع Ridge	۳.۵
۵۷	نتایج سناریو ۱.۱.۵ ($p = 4, n = 50, \nu = 0.9$) با نقطه شروع OLS	۴.۵
۵۷	نتایج سناریو ۱.۱.۵ ($p = 4, n = 50, \nu = 0.9$) با نقطه شروع Ridge	۵.۵
۵۷	نتایج سناریو ۱.۱.۵ ($p = 4, n = 20, \nu = 0.9$) با نقطه شروع Ridge	۶.۵
۵۸	نتایج MSE	۷.۵
۵۹	نتایج سناریو ۲.۱.۵ ($p = 40, n = 20, \nu = 0.7$) با نقطه شروع Ridge	۸.۵
۶۰	نتایج سناریو ۲.۱.۵ ($p = 40, n = 50, \nu = 0.7$) با نقطه شروع Ridge	۹.۵
۶۱	نتایج سناریو ۲.۱.۵ ($p = 40, n = 20, \nu = 0.9$) با نقطه شروع Ridge	۱۰.۵
۶۲	نتایج سناریو ۲.۱.۵ ($p = 40, n = 50, \nu = 0.9$) با نقطه شروع Ridge	۱۱.۵
۶۳	نتایج سناریو ۳.۱.۵ ($p = 40, n = 20, \nu = 0.7$) با نقطه شروع Ridge	۱۲.۵
۶۴	نتایج سناریو ۳.۱.۵ ($p = 40, n = 50, \nu = 0.7$) با نقطه شروع Ridge	۱۳.۵
۶۵	نتایج MSE سناریوهای ۲.۱.۵ و ۳.۱.۵	۱۴.۵
۶۶	نتایج سناریو ۴.۱.۵ ($p = 8, n = 20, \nu = 0.7$) با نقطه شروع OLS	۱۵.۵
۶۶	نتایج سناریو ۵.۱.۵	۱۶.۵
۶۷	ضرایب متغیرهای پروستات	۱۷.۵
۶۷	نتایج ضرایب لاسو به ازای مقادیر مختلف k	۱۸.۵
۶۸	توصیف متغیرهای داده‌های آیووا	۱۹.۵
۶۸	آماره‌های توصیفی متغیر پاسخ در داده‌های آیووا	۲۰.۵

۶۹	۲۱.۵ ضرایب متغیرهای داده‌های آیووا
----	--

فصل ۱

مقدمه و تاریخچه

در این فصل، ضمن یادآوری مدل رگرسیون خطی چندگانه و روش کمترین توان‌های دوم، روی پدیده همخطی تمرکز کرده و تاثیر آن را بر برآوردگر کمترین توان‌های دوم مورد بررسی قرار داده و در مورد راههای تشخیص همخطی و راهکارهای مقابله با آن صحبت می‌کنیم. مطالب این فصل عمدتاً از کتاب‌های مقدمه‌ای بر رگرسیون خطی مونت‌گمری [۸]، تحلیل رگرسیون با مثال نیرومند [۹]، پایان‌نامه نجاریان [۱۰]، روزبه [۷] و مقاله‌های کاریا و کوراتا [۲۴]، مارکوارت [۲۷] برگرفته شده‌اند.

۱.۱ مقدمه

تحلیل رگرسیونی یکی از روش‌های آماری برای تحلیل و مدل‌بندی داده‌های چندعاملی است که حوزه کاربرد آن بیش‌ترین وسعت را دارد. نتایج پرجاذبه آن از نظر مفهومی، فرآیند ساده بکارگیری یک معادله است که ارتباط بین دسته‌ای از متغیرها را بیان می‌کند. تحلیل رگرسیونی همچنین از جهت نظری، به لحاظ ظرافت و زیبایی ریاضیات، دارای جذابیت است. موفقیت در تحلیل، نیاز به درک و تیزبینی در دو مقوله نظری و مسائل عملی دارد و زمانی بروز می‌کند و خود را نشان می‌دهد که تکنیک در بکارگیری داده‌ها و اطلاعات جهان واقع قرار می‌گیرد.

تحلیل رگرسیونی فن و تکنیکی آماری برای بررسی و مدل‌بندی ارتباط بین متغیرهاست. کاربردهای رگرسیون متعدد هستند و تقریباً در هر زمینه‌ای از جمله مهندسی، فیزیک، اقتصاد، مدیریت، علوم زیستی و بیولوژی و علوم اجتماعی صورت می‌پذیرد.

کلمه «رگرسیون»^۱ در لغت به معنای «برگشت» است. دلیل این نام‌گذاری به مطالعات گالتون^۲ [۱۷] در مورد رابطه قد پدر و قد پسر و قد پسر با رسم نمودار قد پسر در مقابل قد پدر (نمودار پراکنش)^۳ براساس حجم بزرگی از داده‌ها به این نتیجه رسید که اغلب افراد بلند قد، پسرانی کوتاه قد دارند و پدران کوتاه قد، پسرانی بلند قدتر از خود دارند و این پدیده را «برگشت به میانگین» نامید و از آن پس نام رگرسیون بر روش‌های بررسی روابط بین متغیرهای آماری باقی ماند.

به عنوان مثال از یک مسئله که در آن تحلیل رگرسیونی می‌تواند مفید واقع شود، فرض می‌کنیم یک مهندس صنایع، توسط یک سازنده نوشابه استخدام شده که محصول تحویلی و عملیات سرویس و خدمت رسانی برای ماشین‌های فروش را تحلیل کند. او گمان می‌کند که زمان لازم برای این که یک فرد تحویل دهنده، این ماشین را سرویس دهد بستگی به تعداد موارد محصول تحویل شده دارد. مهندس ۲۵ فروشنده جزئی را که دارای ماشین فروش می‌باشند، به تصادف بازدید کرده است و زمان تحویل (برحسب دقیقه) و حجم محصول تحویل داده شده (بر حسب مورد) برای هر یک مشاهده می‌شود. به وسیله نمودار پراکنش، ارتباط بین زمان تحویل و حجم تحویل مشخص می‌شود. نقاطی که نشان‌دهنده داده‌ها هستند در امتداد یک خط مستقیم و نه دقیقاً روی خط، قرار می‌گیرند. اگر Y نشان‌دهنده زمان تحویل و X را برای حجم تحویل شده قرار دهیم، در این صورت معادله یک خط مستقیم که این دو متغیر را به هم مرتبط می‌سازد، چنین است

$$Y = \beta_0 + \beta_1 X \quad (1.1)$$

که β_0 عرض از مبدا و β_1 شیب خط است. از آنجایی که نقاط داده‌ها دقیقاً روی یک خط مستقیم قرار نمی‌گیرند بنابراین معادله (۱.۱) بایستی به عنوان برآورد کننده آن اصلاح و تعدیل شود. اگر اختلاف بین مقدار مشاهده شده Y و خط مستقیم $(\beta_0 + \beta_1 X)$ را خطای ϵ قرار دهیم، مناسب است که خطای ϵ به عنوان یک خطای آماری تلقی شود. بدین معنی که متغیری تصادفی است که عدم برازش مدل را بیان کرده و اندازه ناتوانی مدل، از برازش داده‌ها را برآورد می‌کند. این خطا ممکن است به لحاظ اثرات دیگر متغیرها، خطای اندازه‌گیری و غیره روی زمان تحویل صورت پذیرد. بنابراین مدلی با قابلیت پذیرش بیشتر برای داده‌های زمان تحویل عبارت است از

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (2.1)$$

معادله (۲.۱) یک مدل رگرسیون خطی نامیده می‌شود؛ که در آن X متغیر پیش‌بین یا همراه و Y متغیر پاسخ است. از آنجایی که در (۲.۱) فقط یک متغیر پیش‌بین وجود دارد، این مدل را مدل رگرسیونی ساده نامیده‌اند. برای آگاهی بیشتر کتاب مقدمه‌ای بر رگرسیون خطی [۸] را ببینید. حال فرض کنید علاوه بر حجم تولید، متغیر دیگری را نیز به عنوان متغیر پیش‌بین در نظر بگیریم، در این صورت نمی‌توان از مدل (۲.۱) استفاده کرد و باید مدل کلی‌تری که دربرگیرنده متغیرهای پیش‌بین چندتایی یا چندگانه است، در نظر گرفت، که در بخش بعد این گونه مدل را معرفی می‌کنیم. برای اطلاعات بیشتر در این زمینه به نیرومند [۹] مراجعه شود.

^۱Regression

^۲Galton

^۳Scatter plot

۲.۱ مدل رگرسیونی چندگانه

در حالت کلی ممکن است متغیر پاسخ به p متغیر پیش‌بین $X_1, X_2, \dots, X_p$ مربوط باشد، بنابراین مدل رگرسیونی به صورت زیر نوشته می‌شود

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad (3.1)$$

که در آن X_i ، i امین متغیر پیش‌بین است. چون در این مدل رگرسیونی بیش از یک متغیر درگیر هستند، رگرسیون چندگانه نام دارد. پارامترهای β_j ($j = 0, 1, \dots, p$) ضرایب رگرسیون نامیده می‌شوند. بکارگیری صفت خطی بودن در مدل فوق اشاره بر این دارد که مدل، برحسب پارامترهای $\beta_0, \beta_1, \dots, \beta_p$ خطی است و نه به دلیل اینکه Y تابعی خطی از X_i ‌هاست. این مدل یک ابر صفحه در فضای p بعدی از متغیرهای رگرسیونی X_i است. پارامتر β_j نشان‌دهنده تغییرات مورد انتظار متغیر پاسخ به ازای یک واحد تغییر در X_j است وقتی که همه‌ی متغیرهای پیش‌بین باقیمانده دیگر (X_i ($i \neq j$)) ثابت نگه‌داشته شوند. به همین جهت پارامترهای β_j ($j = 1, 2, \dots, p$) ضرایب جزئی رگرسیون نامیده می‌شوند. مدل‌های رگرسیون چندگانه، اغلب به عنوان تقریب توابع بکار می‌روند. بدین معنی که ارتباط تابعی واقعی بین $X_1, X_2, \dots, X_p$ و Y شناخته شده نیست، اما مدل رگرسیون خطی روی دامنه تغییرات متغیرهای پیش‌بین، تقریب مناسبی از ارتباط تابعی مذکور می‌باشد.

۱.۲.۱ برآورگر کمترین توان‌های دوم

فرض کنید $n > p$ مشاهده در دسترس هستند. می‌توان مدل متناظر با (۳.۱) را براساس نماد ماتریسی به صورت زیر نوشت

$$Y = X\beta + \epsilon \quad (4.1)$$

که در آن $Y = (Y_1, \dots, Y_n)^T$ یک بردار n مولفه‌ای شامل متغیر پاسخ و Y_i نمایش دهنده i امین پاسخ مشاهده شده، $X = (x_1, x_2, \dots, x_n)^T$ ماتریس طرح $n \times p$ که در آن $x_i \in \mathbb{R}^p$ و X_{ij} نشان دهنده i امین مشاهده در j امین سطح متغیر پیش‌بین باشد. فرض می‌کنیم ماتریس X دارای رتبه کامل ستونی است. با در نظر گرفتن این فرض، رتبه ماتریس X ، p بوده و p ستون مستقل خطی داریم بنابراین ماتریس p بعدی $X^T X$ معکوس پذیر است. $\beta = (\beta_1, \dots, \beta_p)^T$ یک بردار p بعدی شامل ضرایب رگرسیونی و $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ بردار n مولفه‌ای خطای تصادفی است که در آن $E(\epsilon) = 0$ و $E(\epsilon\epsilon^T) = \sigma^2 I_n$. برآوردگر کمترین توان‌های دوم معمولی^۴ (OLS) برآوردگری است که مجموع

^۴ Ordinary least square

توان‌های دوم مانده‌ها^۵ (RSS) در زیر را کمینه کند.

$$\begin{aligned} RSS(\beta) &= \epsilon^T \epsilon = \sum_{i=1}^n \epsilon_i^2 \\ &= \sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^p \beta_j X_{ij} \right)^2 \\ &= (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}^T \mathbf{Y} - 2\mathbf{Y}^T \mathbf{X}\beta + \beta^T \mathbf{X}^T \mathbf{X}\beta \end{aligned}$$

با مشتق‌گیری از $RSS(\beta)$ نسبت به β و مساوی صفر قرار دادن داریم $-2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X}\beta$ و برآوردگر ضرایب مدل به صورت $\hat{\beta}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ حاصل می‌شوند. ارتباط بردار مقادیر برازش شده $\hat{\mathbf{Y}}$ با مقادیر مشاهده شده $\mathbf{Y}$ چنین خواهد بود

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{H}\mathbf{Y}$$

که در آن $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ ماتریس هت^۶ یا برازش شده نامیده می‌شود، زیرا باعث می‌شود روی $\mathbf{Y}$ یک علامت هت گذاشته شود و از طرفی بردار مقادیر مشاهدات را به بردار مقادیر برازش شده تصویر می‌کند. ماتریس $\mathbf{H}$ ماتریس $p \times p$ ، متقارن و خودتوان^۷ است و بنا به قضیه ب.۳۳.۱ در پیوست ب، رابطه $\text{rank}(\mathbf{H}) = \text{tr}(\mathbf{H}) = p$ بین رتبه و اثر ماتریس برقرار است. ماتریس واریانس-کواریانس $\hat{\beta}_{OLS}$ به صورت زیر محاسبه می‌شود

$$\begin{aligned} \text{Cov}(\hat{\beta}_{OLS}) &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Cov}(\mathbf{Y}) (\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1})^T \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{I}_p ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)^T \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \end{aligned} \quad (۵.۱)$$

قضیه ۱.۲.۱. کاریا و کوراتا [۲۴] (قضیه گوس-مارکوف) مدل خطی (۴.۱) را در نظر بگیرید. اگر $E(\mathbf{Y}) = \mathbf{X}\beta$ و $\text{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}_n$ ، آنگاه برآوردهای کمترین توان‌های دوم $\hat{\beta}_j, j = 0, 1, \dots, p$ در میان تمام برآوردهای نااریب خطی دارای کمترین واریانس هستند.

برهان.

یک برآوردهای خطی مانند $\mathbf{AY}$ از β در نظر می‌گیریم و ماتریس $\mathbf{A}$ را به قسمی پیدا می‌کنیم که $\mathbf{AY}$ یک برآوردهای نااریب با کمترین واریانس برای β باشد. از این رو باید داشته باشیم $E(\mathbf{AY}) = \beta$. با استفاده از فرض $E(\mathbf{Y}) = \mathbf{X}\beta$ ، به ازای همه مقادیر $\beta \in \mathbb{R}$ می‌توان نوشت

$$\beta = E(\mathbf{AY}) = \mathbf{A}E(\mathbf{Y}) = \mathbf{A}\mathbf{X}\beta$$

^۵Residual sum of squares

^۶Hat matrix

^۷Idempotent

که شرط ناریبی زیر را می‌دهد

$$AX = I_n$$

ماتریس واریانس-کواریانس برای برآوردگر AY عبارت است از

$$\text{Cov}(AY) = A(\sigma^2 I_n)A^T = \sigma^2 AA^T$$

واریانس $\hat{\beta}_j$ ها روی قطر $\sigma^2 AA^T$ قرار دارند، بنابراین باید A را به قسمی انتخاب کنیم که (با توجه به شرط ناریبی) اعضای قطری AA^T کمینه شوند. برای ارتباط AY به $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$ عبارت $(X^T X)^{-1} X^T$ را اضافه و کم می‌کنیم

$$AA^T = (A - (X^T X)^{-1} X^T + (X^T X)^{-1} X^T) (A - (X^T X)^{-1} X^T + (X^T X)^{-1} X^T)^T$$

اگر این عبارت را نسبت به $A - (X^T X)^{-1} X^T$ و $(X^T X)^{-1} X^T$ بسط دهیم، چهار جمله به دست می‌آوریم که دو جمله آن بنابر $AX = I_n$ ، صفر می‌شوند. در نتیجه داریم

$$AA^T = (A - (X^T X)^{-1} X^T) (A - (X^T X)^{-1} X^T)^T + (X^T X)^{-1} \quad (۶.۱)$$

ماتریس $(A - (X^T X)^{-1} X^T) (A - (X^T X)^{-1} X^T)^T$ از سمت راست (۶.۱) نیمه معین مثبت است (قضیه ۳۵.۱.۲ را ببینید)، بنا به قسمت اول قضیه ۳۶.۱.۲ اعضای قطری بزرگتر یا مساوی صفرند. با انتخاب $A = (X^T X)^{-1} X^T$ ، این اعضای قطری را می‌توان برابر صفر قرار داد. (این مقدار A در شرط ناریب بودن $AX = I_n$ نیز صدق می‌کند.) پس $AY = (X^T X)^{-1} X^T Y$ و لذا $A = (X^T X)^{-1} X^T$ □
که برآوردگر کمترین توان‌های دوم معمولی، یعنی $\hat{\beta}_{OLS}$ را نتیجه می‌دهد.

قضیه ۲.۲.۱. فرض کنید $S^2 = \frac{RSS(\beta)}{n-p}$ ، آنگاه S^2 برآوردگر ناریب σ^2 است. **برهان.** می‌دانیم

$$\frac{(n-p)S^2}{\sigma^2} \sim \chi^2_{(n-p)}$$

□

در نتیجه S^2 برای σ^2 ناریب است.

قضیه ۳.۲.۱. کاریا و کوراتا [۲۴] مدل خطی (۴.۱) را در نظر بگیرید. اگر $\text{Cov}(Y) = \text{Cov}(\epsilon) = \sigma^2 \Sigma$ ، که در آن Σ ماتریسی معین مثبت است، آنگاه با قرار دادن

$$X^T \Sigma^{-1} X = C$$

الف) برآوردگر کمترین توان‌های دوم ضرایب رگرسیونی به صورت زیر است

$$\hat{\beta}_{OLS} = C^{-1} X^T \Sigma^{-1} Y$$

ب) ماتریس کواریانس $\hat{\beta}_{OLS}$ عبارت است از

$$Cov(\hat{\beta}_{OLS}) = \sigma^2 C^{-1}$$

ج) برآوردگر ناریب $\hat{\sigma}^2$ به صورت زیر است

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})^T \Sigma^{-1} (Y - X\hat{\beta})}{n - p}$$

برهان.

الف) از این که Σ ماتریس معین مثبت است، ماتریس ناویژه $P_{n \times n}$ وجود دارد که در آن $\Sigma = PP^T$ (قضیه ب. ۳۲.۱). با ضرب نمودن P^{-1} از سمت چپ در مدل (۴.۱) داریم $P^{-1}Y = P^{-1}X\beta + P^{-1}\epsilon$ که در آن $E(P^{-1}\epsilon) = P^{-1}E(\epsilon) = 0$ همچنین

$$\begin{aligned} Cov(P^{-1}\epsilon) &= P^{-1}Cov(\epsilon)(P^{-1})^T \\ &= P^{-1}\sigma^2\Sigma(P^{-1})^T \\ &= \sigma^2 P^{-1}PP^T(P^{-1})^T \\ &= \sigma^2 I_n \end{aligned}$$

بنابراین

$$\begin{aligned} \hat{\beta}_{OLS} &= (X^T(P^{-1})^T P^{-1}X)^{-1} X^T(P^{-1})^T P^{-1}Y \\ &= (X^T(P^T)^{-1} P^{-1}X)^{-1} X^T(P^T)^{-1} P^{-1}Y \\ &= (X^T(PP^T)^{-1}X)^{-1} X^T(PP^T)^{-1}Y \\ &= C^{-1}X^T\Sigma^{-1}Y \end{aligned}$$

ب)

$$\begin{aligned} Cov(C^{-1}X^T\Sigma^{-1}Y) &= C^{-1}X^T\Sigma^{-1}Cov(Y)\Sigma^{-1}XC^{-1} \\ &= C^{-1}X^T\Sigma^{-1}\sigma^2\Sigma\Sigma^{-1}XC^{-1} \\ &= \sigma^2 C^{-1} \end{aligned}$$

ج) با توجه به این که

$$\begin{aligned} \hat{\beta}_{OLS}^T X^T \Sigma^{-1} X \hat{\beta}_{OLS} &= \hat{\beta}_{OLS}^T X^T \Sigma^{-1} X C^{-1} X^T \Sigma^{-1} Y \\ &= \hat{\beta}_{OLS}^T C C^{-1} X \Sigma^{-1} Y \\ &= \hat{\beta}_{OLS}^T X \Sigma^{-1} Y \end{aligned}$$

می‌توان نوشت

$$\begin{aligned} (Y - X\hat{\beta}_{OLS})^T \Sigma^{-1} (Y - X\hat{\beta}_{OLS}) &= Y^T \Sigma^{-1} Y - Y^T \Sigma^{-1} X \hat{\beta}_{OLS} \\ &\quad - \hat{\beta}_{OLS}^T X^T \Sigma^{-1} Y + \hat{\beta}_{OLS}^T X^T \Sigma^{-1} X \hat{\beta}_{OLS} \\ &= Y^T \Sigma^{-1} Y - Y^T \Sigma^{-1} X C^{-1} X^T \Sigma^{-1} Y \\ &= Y^T (\Sigma^{-1} - \Sigma^{-1} X C^{-1} X^T \Sigma^{-1}) Y \end{aligned}$$

و بنا به قضیه ۳۸.۱ ب و با در نظر گرفتن $V = \sigma^2 \Sigma$ ، $A = (\Sigma^{-1} - \Sigma^{-1} X C^{-1} X^T \Sigma^{-1})$ و $\mu = X\beta$ نتیجه حاصل می‌شود.

۳.۱ همخطی در رگرسیون چندگانه

مدل‌های رگرسیونی در حوزه کاربردی به‌طور وسیعی مورد استفاده قرار می‌گیرند. یک مسئله جدی که می‌تواند استفاده از مدل رگرسیونی را با اشکال مواجه کند، همخطی چندگانه یا وابستگی خطی نزدیک بین متغیرهای رگرسیونی است. وجود وابستگی خطی نزدیک، توانایی برآورد ضرایب مدل رگرسیون را با مشکل مواجه می‌کند.

بکارگیری و تعبیر یک مدل رگرسیونی چندگانه بطور ضمنی یا بطور صریح به برآوردهای ضرایب مدل رگرسیونی بستگی دارد. بعضی از استنباط‌هایی که بطور معمول در مدل رگرسیونی صورت می‌پذیرد، شامل موارد زیر است:

۱- مشخص کردن اثرات نسبی متغیرهای پیش‌بین

۲- پیش‌بینی یا برآورد

۳- انتخاب دسته‌ای مناسب از متغیرها برای مدل

تعبیر و تفسیر معادله‌ی رگرسیون چندگانه به‌طور ضمنی به این فرض وابسته است که متغیرهای پیش‌بین وابستگی قوی با یکدیگر ندارند و به عبارتی مستقل خطی هستند. زیرا تعبیر یک ضریب رگرسیون به صورت اندازه‌گیری تغییر در متغیر پاسخ وقتی متغیر پیش‌بین یک واحد افزایش یافته و سایر متغیرهای پیش‌بین ثابت باشند امری متداول است؛ اما این تغییر در صورتی که رابطه خطی قوی بین متغیرهای پیش‌بین وجود داشته باشد معتبر نخواهد بود. وقتی هیچ ارتباط خطی بین متغیرهای پیش‌بین وجود نداشته باشد، این متغیرها را متعامد گویند.

در اکثر کاربردهای رگرسیونی متغیرهای پیش‌بین متعامد نیستند. در بعضی وضعیت‌ها متغیرهای پیش‌بین تقریباً به‌طور ناقص ارتباط خطی دارند و در این چنین مواردی نتایج رگرسیون مبهم خواهد بود.

اگر یکی از متغیرهای پیش‌بین یک تابع دقیق خطی از یک یا چند متغیر پیش‌بین دیگر باشد گوئیم رگرسیون دارای همخطی کامل^۸ است. همخطی ناقص^۹ وقتی اتفاق می‌افتد که یکی از متغیرهای پیش‌بین

^۸Exact collinearity

^۹Partial collinearity

به طور تقریبی یک تابع خطی از یک یا چند متغیر پیش‌بین دیگر باشد. به عبارت دیگر همخطی کامل زمانی رخ می‌دهد که دست‌کم به ازای یک i ($i = 1, \dots, p$) داشته باشیم $R_i^2 = 1$ و همخطی ناقص زمانی پیش می‌آید که دست‌کم به ازای یک i داشته باشیم $R_i^2 \simeq 1$ ، که در آن ضریب تعیین رگرسیون خطی متغیر X_i بر سایر متغیرهای پیش‌بین است. باید دقت داشت که ضریب تعیین R^2 حاصل تقسیم پراکندگی بیان شده به پراکندگی کل است. لذا در مدل رگرسیون (۴.۱)، ضریب تعیین Y روی X برابر است با

$$R^2 = \frac{\hat{\beta}_{OLS}^T \mathbf{X}^T \mathbf{Y} - n \bar{Y}^2}{\mathbf{Y}^T \mathbf{Y} - n \bar{Y}^2}, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

برای اطلاع از چگونگی ایجاد همخطی کامل و ناقص به روزه [۷] مراجعه کنید.

۱.۳.۱ اثرات همخطی بر برآورد پارامترها

وجود همخطی چندگانه تاثیرات جدی بر برآوردهای کمترین توان‌های دوم ضرایب رگرسیون دارد. برای درک بهتر آثار، مدلی با دو متغیر رگرسیونی را در نظر بگیرید

$$\mathbf{Y} = \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \epsilon$$

دستگاه معادلات کمترین توان‌های دوم نرمال شده عبارتند از

$$(\mathbf{X}^T \mathbf{X}) \hat{\beta} = \mathbf{X}^T \mathbf{Y}$$

یا به عبارت دیگر

$$\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} r_{1y} \\ r_{2y} \end{pmatrix}$$

که در آن r_{12} همبستگی ساده بین $\mathbf{X}_1$ و $\mathbf{X}_2$ و r_{jy} همبستگی ساده بین $\mathbf{X}_j$ و $\mathbf{Y}$ به ازای $j = 1, 2$ می‌باشد. معکوس $(\mathbf{X}^T \mathbf{X})$ به صورت زیر حاصل می‌شوند

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{(1-r_{12}^2)} & \frac{-r_{12}}{(1-r_{12}^2)} \\ \frac{-r_{12}}{(1-r_{12}^2)} & \frac{1}{(1-r_{12}^2)} \end{pmatrix} \quad (7.1)$$

و برآورد ضرایب رگرسیون عبارتند از

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12}r_{2y}}{(1 - r_{12}^2)}, \quad \hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{(1 - r_{12}^2)}$$

اگر بین $\mathbf{X}_1$ و $\mathbf{X}_2$ همخطی چندگانه موجود باشد، در این صورت ضریب همبستگی r_{12} بزرگ خواهد بود. از (۷.۱) خواهیم دید اگر $|r_{12}| \rightarrow 1$ آنگاه $\text{Var}(\hat{\beta}_j) = S_{jj} \hat{\sigma}^2 \rightarrow \infty$ و $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \rightarrow \pm \infty$ بنابراین همخطی زیاد بین $\mathbf{X}_1$ و $\mathbf{X}_2$ منجر به واریانس‌ها و کواریانس‌های بزرگ برای برآوردهای کمترین توان‌های دوم ضرایب رگرسیون خواهد شد. در نتیجه اگر از سطوح یکسان $\mathbf{X}$ نمونه‌های مختلف گرفته شود، برآوردهای بدست آمده برای هر پارامتر با مقدار واقعی پارامتر، تفاوت

زیادی خواهد داشت. وقتی که بیش از دو متغیر رگرسیونی وجود دارد، همخطی اثرات مشابهی ایجاد می‌کند. می‌توان نشان داد که اعضای قطر اصلی ماتریس $S = (X^T X)^{-1}$ عبارتند از

$$S_{jj} = \frac{1}{1 - R_j^2}, \quad j = 1, 2, \dots, p$$

که در آن R_j^2 ضریب تعیین چندگانه از رگرسیون X_j نسبت به $p - 1$ متغیر پیش‌بین باقیمانده است. اگر همخطی شدید بین X_j و هر زیرمجموعه‌ای از $p - 1$ متغیر پیش‌بین دیگر وجود داشته باشد، در این صورت مقدار R_j^2 نزدیک به واحد خواهد بود. چون واریانس $\hat{\beta}_j$ برابر است با

$$\text{Var}(\hat{\beta}_j) = S_{jj}\hat{\sigma}^2 = (1 - R_j^2)^{-1}\hat{\sigma}^2$$

بنابراین همخطی ناقص باعث می‌شود که دترمینان ماتریس $X^T X$ بسیار کوچک شود و در نتیجه ماتریس $(X^T X)^{-1}$ دارای اعضای باشند که از لحاظ قدرمطلق بسیار بزرگ هستند. همچنین، واریانس برآوردگر $\hat{\beta}_{OLS}$ بسیار بزرگ می‌شود. در این صورت، $\hat{\beta}_{OLS}$ برآورد دقیقی از β نمی‌باشد. به عبارت دیگر دقت برآورد ضرایب رگرسیون بسیار کم می‌شود. علاوه بر این $\hat{\beta}_{OLS}$ بسیار ناپایدار^{۱۰} است. یعنی با تغییر کوچکی در مقادیر متغیرهای پیش‌بین، مقادیر موجود در $\hat{\beta}_{OLS}$ بسیار تغییر می‌کنند. در نتیجه محاسبه $\hat{\beta}_{OLS}$ توسط دو رایانه متفاوت (دو روش متفاوت) ممکن است جواب‌های کاملاً متفاوتی بدست دهند. همچنین در صورت وجود همخطی کامل ماتریس $(X^T X)^{-1}$ وارون پذیر نبوده (زیرا این ماتریس پر رتبه^{۱۱} نیست) و در نتیجه برای برآورد ضرایب، جواب یکتا وجود نخواهد داشت. بنابراین مهم است بدانیم که چه وقت همخطی وجود دارد تا مواظب پیامدهای ممکن آن باشیم.

چهار منبع اولیه پیدایش همخطی وجود دارند:

۱- شیوه جمع‌آوری داده‌ها

۲- قیدهای روی مدل یا درون جامعه

۳- نوع مدل

۴- مدلی با متغیرهای پیش‌بین بیش از حد

درک اختلاف بین این منابع همخطی چندگانه اهمیت دارد، به‌طوری که توصیه برای تحلیل داده‌ها و تعبیر مدل نتیجه‌شده تا اندازه‌ای به علت مسئله همخطی بستگی دارد، برای اطلاعات بیش‌تر در مورد همخطی چندگانه رجوع کنید به میسون، گونست و وبستر [۲۸].

۲.۳.۱ چگونه می‌توان به وجود همخطی چندگانه پی برد؟

در این قسمت بعضی از روش‌هایی که برای تشخیص همخطی چندگانه بکار می‌روند، ارائه می‌شود.

محک ماتریس همبستگی

فرض کنید X یک ماتریس $n \times p$ باشد. یک اندازه بسیار ساده همخطی چندگانه بررسی اعضای غیر قطر $r_{ij} \forall i, j = 1, 2, \dots, p$ در ماتریس $X^T X$ است. اگر متغیرهای پیش‌بین X_i و X_j تقریباً وابسته

^{۱۰}Unstable

^{۱۱}Full rank

خطی باشند، در این صورت $|r_{ij}|$ نزدیک واحد خواهد بود.

عامل تورم واریانس

اعضای قطر ماتریس $S = (X^T X)^{-1}$ برای آشکار سازی همخطی چندگانه کاربرد زیادی دارند. S_{jj} ، زامین عضو قطر ماتریس S می تواند به صورت $S_{jj} = (1 - R_j^2)^{-1}$ نوشته شود، که در آن R_j^2 ضریب تعیین بدست آمده است هنگامی که X_j به $1 - p$ متغیر پیش بین باقیمانده برگشته است. اگر X_j تقریباً بر متغیرهای پیش بین باقیمانده عمود باشد، R_j^2 کوچک و S_{jj} به واحد نزدیک می شود، در حالی که اگر وابستگی تقریباً خطی به بعضی متغیرهای رگرسیونی داشته باشد R_j^2 نزدیک به واحد و S_{jj} بزرگ می شود. چون واریانس زامین ضریب رگرسیون برابر است با $S_{jj}\sigma^2$ می توان S_{jj} را به عنوان عاملی دید که به وسیله آن واریانس $\hat{\beta}_j$ به لحاظ وابستگی نزدیک بین متغیرهای پیش بین، افزایش یافته است. معیار

$$VIF_j = S_{jj} = (1 - R_j^2)^{-1}$$

را عامل تورم واریانس^{۱۲} می نامند. این نام گذاری مربوط به مارکوارت^{۱۳} [۲۷] است. برای هر جمله مدل، VIF اثر مرکب از ارتباط بین متغیرهای پیش بین روی واریانس آن جمله را اندازه گیری می کند. یک VIF بزرگ اشاره بر همخطی چندگانه دارد. تجربیات حاکی از این است که اگر هر یک از VIF ها از ۵ یا ۱۰ تجاوز کنند، خطاری مبنی بر این است که ضرایب رگرسیونی زیربط به علت همخطی چندگانه به صورت ضعیفی برآورد شده اند، برای اطلاعات کامل تر درباره عامل تورم واریانس پیوست ۱.۰ را ببینید.

عدد شرطی

ریشه های مشخصه یا مقادیر ویژه $X^T X$ یعنی $\lambda_1, \lambda_2, \dots, \lambda_p$ می توانند برای اندازه گیری وسعت همخطی چندگانه در داده ها مورد استفاده قرار گیرند. عدد شرطی^{۱۴} $X^T X$ به صورت زیر تعریف می شود

$$CN = \frac{\lambda_{max}}{\lambda_{min}}$$

که در آن $\lambda_{Max} = \max(\lambda_i)$ و $\lambda_{Min} = \min(\lambda_i)$ است. عدد شرطی یک اندازه گیری گسترده در طیف مقادیر ویژه $X^T X$ است. در حالت کلی اگر عدد شرطی کمتر از ۱۰۰ باشد، مشکل جدی در مورد همخطی چندگانه وجود ندارد. اعداد شرطی بین ۱۰۰ و ۱۰۰۰ منتج به همخطی چندگانه شدید و اگر از ۱۰۰۰ تجاوز کند، همخطی چندگانه بسیار شدید را حکایت می کند.

لازم به ذکر است که در بعضی مقالات جذر رابطه بالا مد نظر است و گفته می شود اگر جذر عدد شرطی $X^T X$ بزرگتر از ۱۰ باشد آنگاه در ماتریس طرح همخطی وجود دارد. که اگر جذر عدد شرطی بین ۳۰ تا ۱۰۰ باشد همخطی شدید و بالاتر از ۱۰۰ همخطی کاملاً جدی و مشکل ساز است (بلسلی و همکاران^{۱۵} [۱۲])، اطلاعات کامل تر در خصوص عدد شرطی در پیوست ۱.۰ آمده است.

^{۱۲}Variance inflation factor

^{۱۳}Marquardt

^{۱۴}Condition number

^{۱۵}Belsley

۳.۳.۱ روش‌هایی برای برخورد با همخطی چندگانه

روش‌های متعددی برای برخورد با مشکلات ناشی از همخطی چندگانه پیشنهاد شده‌اند. روش‌های مرسوم شامل جمع‌آوری داده‌های اضافی، تخصیص مدل مجدد و بکارگیری روش‌های برآورد غیر از کمترین توان دوم که به طور ویژه طراحی شده‌اند، برای از بین بردن مشکلات ایجاد شده به لحاظ همخطی چندگانه است.

وقتی روش کمترین توان‌های دوم در مورد متغیرهای غیر متعامد به کار می‌رود، معمولاً برآوردهای خیلی ضعیف برای ضرایب رگرسیونی به دست می‌آید. که در این صورت واریانس برآوردهای کمترین توان‌های دوم ضرایب رگرسیون، ممکن است در حد قابل توجهی افزایش یابد و همچنین قدرمطلق برآوردهای کمترین توان‌های دوم خیلی بزرگ می‌شود و لذا بسیار ناپایدار هستند، بدین معنی که با ارائه نمونه‌های متفاوت، اندازه‌ها و علامت‌ها در حد قابل توجهی تغییر خواهند کرد.

قضیه ۱.۲.۱ این اطمینان را می‌دهد که برآوردگر کمترین توان‌های دوم در رده برآوردهای خطی نارایب دارای کمترین واریانس است، اما تضمینی وجود ندارد که واریانس کوچکترین باشد. اگر واریانس $\hat{\beta}_{OLS}$ بزرگ باشد در نتیجه فاصله اطمینان برای β عریض می‌شود که در این صورت برآورد نقطه‌ای $\hat{\beta}_{OLS}$ بسیار ناپایدار است. یک راه کوچک جلوه دادن این مسئله، کنار گذاشتن این ضرورت است که بایستی برآوردگر β نارایب باشد.

فرض کنیم که بتوان یک برآوردگر اریب $\hat{\beta}^*$ برای β پیدا کرد که دارای واریانس کمتر از واریانس برآوردگر نارایب $\hat{\beta}_{OLS}$ باشد. میانگین توان دوم خطای^{۱۶} (MSE) برآوردگر $\hat{\beta}^*$ به صورت زیر است

$$\begin{aligned} \text{MSE}(\hat{\beta}^*) &= E\left((\hat{\beta}^* - \beta)^T(\hat{\beta}^* - \beta)\right) = \text{Var}(\hat{\beta}^*) + \left(E(\hat{\beta}^*) - \beta\right)^T \left(E(\hat{\beta}^*) - \beta\right) \\ &= \text{Var}(\hat{\beta}^*) + \left(\text{Bias}(\hat{\beta}^*)\right)^T \left(\text{Bias}(\hat{\beta}^*)\right) \end{aligned} \quad (۸.۱)$$

توجه داریم که MSE برابر با امیدریاضی توان دوم فاصله $\hat{\beta}^*$ تا β می‌باشد. با پذیرفتن مقدار کمی اریبی برای $\hat{\beta}^*$ ، واریانس $\hat{\beta}^*$ می‌تواند طوری کوچک شود که MSE برای $\hat{\beta}^*$ از واریانس برآوردگر نارایب $\hat{\beta}_{OLS}$ کمتر باشد. باید توجه داشت که نمی‌توان هم‌زمان هم اریبی هم واریانس را کاهش داد. با کاستن یکی، دیگری افزایش پیدا می‌کند. به این خاصیت مبادله پایایی اریبی و واریانس می‌گویند که در این موارد معمولاً مقدار MSE را کمینه می‌کنند.

با به کارگیری برآوردگر اریب برای β بایستی فاصله اطمینان برای β باریک‌تر شود. واریانس کوچک برای برآوردگر اریب نیز نتیجه می‌دهد که $\hat{\beta}^*$ برآوردگری پایدارتر از برآوردگر نارایب $\hat{\beta}_{OLS}$ برای β می‌باشد. یکی از روش‌های مقابله با همخطی روش‌های انقباضی^{۱۷} است. برآوردگر انقباضی ریدج^{۱۸}، که حالت خاصی از برآوردگر انقباضی ریدج^{۱۹} است، در سال ۱۹۷۰ توسط هورل و کنارد [۲۰] ارائه شد. تییشیرانی [۲۴] با استفاده از عملگر علامت برآوردگر انقباضی لاسو (LASSO)^{۲۰} را برای اولین بار ارائه کرد که

^{۱۶} Mean square error

^{۱۷} Shrinkage method

^{۱۸} Ridge

^{۱۹} Bridge

^{۲۰} Least absolute shrinkage and selection operator

برخلاف برآوردگر انقباضی ریج که از یک محدودیت L_2 بر روی فضای پارامتر استفاده کرده و برآوردگر را منقبض می‌کند، از یک محدودیت L_1 استفاده کرد و برآوردگر انقباضی با امکان انتخاب متغیر با نام برآوردگر لاسو را ارائه نمود. فرانک و فریدمن [۱۴] رگرسیون بریج را معرفی کردند که از محدودیت L_1 استفاده می‌کند. برآوردگر بریج واضح و روشن نیست اما آن‌ها استدلال کردند که انتخاب بهینه γ موجب افزایش دقت پیش‌گویی می‌شود. فو [۱۶] برآوردگرهای ریج، لاسو و بریج را با استفاده از معیار عملکرد پیشگویی مقایسه کرد. نایت و فو [۲۳] سازگاری برآوردگر بریج را با تعداد متغیر (p) ثابت تحت شرایط نظم بررسی کردند. برای اطلاعات بیشتر درباره‌ی برآوردگر انقباضی بریج می‌توان به فو [۱۶]، نایت و فو [۲۳] مراجعه نمود. در فصل‌های بعد با این برآوردگرها بیشتر آشنا می‌شویم.

فصل ۲

برآوردگر بریج

در این فصل، برآوردگر بریج را معرفی کرده و از آن جایی که این برآوردگر صورت بسته‌ای ندارد با استفاده از روش نیوتن رافسون اصلاح شده ضرایب برآوردگر بریج به ازای $\gamma > 1$ را به دست آورده و واریانس آن را محاسبه می‌نماییم. در ادامه رفتار مجانبی این برآوردگر را مورد مطالعه قرار می‌دهیم. مطالب این فصل عمدتاً از مقاله‌های فو [۱۶]، نایت و فو [۲۳]، تیبشیرانی [۳۴] و کراون و وهبا [۱۳] گرفته شده‌اند.

۱.۲ مقدمه

مدل رگرسیونی زیر را در نظر بگیرید

$$Y = X\beta + \epsilon \quad (1.2)$$

که در آن $Y = (Y_1, \dots, Y_n)$ یک بردار n مولفه‌ای شامل متغیر پاسخ، $X = (x_1, \dots, x_n)^T$ یک ماتریس $n \times p$ که $x_i \in \mathbb{R}^p$ ، i امین سطر ماتریس X است. فرض می‌کنیم ماتریس X دارای رتبه کامل ستونی باشد. $\beta = (\beta_1, \dots, \beta_p)^T$ یک بردار p بعدی شامل ضرایب رگرسیونی از فضای پارامتر $\mathcal{B} = \{\beta = (\beta_1, \beta_2, \dots, \beta_p) \subseteq \mathbb{R}^p\}$ و $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ بردار n مولفه‌ای خطای تصادفی می‌باشد که در آن $E(\epsilon) = 0$ و $Cov(\epsilon) = \sigma^2 I_n$.

فرض کنید $\hat{\beta}$ یک برآوردگر بردار ضرایب β باشد. در مسئله برآورد کردن پارامترها، همواره به دنبال برآوردگری هستیم که کمترین میزان خطا را داشته باشد. از آنجایی که مقدار خطای برآوردگر وابسته به دو معیار آریبی و واریانس است، بنابراین بدین منظور در بین رده‌های مختلف، برآوردگری را می‌یابیم که

میانگین توان دوم خطای کمتری داشته باشد. میانگین توان دوم خطای برآوردگر $\hat{\beta}$ در برآورد پارامتر β به صورت (۸.۱) محاسبه می شود؛ برآوردگر کمترین توان های دوم رایج ترین برآوردگر مدل رگرسیون است. این برآوردگر مجموع توان های دوم مانده ها

$$RSS(\beta) = RSS(\beta_1, \dots, \beta_p) = \epsilon^T \epsilon = (Y - X\beta)^T (Y - X\beta) \quad (2.2)$$

را کمینه می نماید. با مشتق گیری از RSS نسبت به β و برابر صفر قرار دادن آن، برآوردگر ضرایب مدل به صورت $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$ حاصل می شوند، که بنا به قضیه گوس-مارکف (۱.۲.۱) بهترین برآوردگر خطی ناریب^۱ (BLUE) بوده و لذا در رده برآوردگرهای خطی دارای کمترین MSE است. وقتی که در داده ها همخطی وجود داشته باشد، این برآوردگر کارایی لازم را ندارد. (رجوع شود به بخش ۳.۱). حال شرایطی را در نظر بگیرید که در داده ها مشکل همخطی وجود دارد. مسئله همخطی و کاربردی نبودن برآوردگر OLS، باعث شده پا را فراتر از رده برآوردگرهای ناریب گذاشته و منجر به تعریف برآوردگرهای اربب مانند برآوردگر انقباضی نوع استاین^۲، برآوردگر مولفه های اصلی^۳، برآوردگر کمترین توان های دوم جزئی^۴، برآوردگر نوع لیو^۵، برآوردگر تقریباً ناریب^۶ و برآوردگر ریج و برآوردگر لاسو شده است. برای اطلاعات بیش تر در زمینه ی برآوردگرهای مولفه های اصلی، کمترین توان های دوم جزئی و لیو می توان به [۱۱]، [۲۵] و [۳۳] و [۳] مراجعه نمود.

در این میان استفاده از برآوردگرهای انقباضی بسیار معمول است، برای آگاهی بیشتر در خصوص برآوردگرهای انقباضی به [۴] مراجعه کنید. یکی از معمول ترین روش های دستیابی به برآوردگرهای انقباضی استفاده از رگرسیون جریمه شده^۷ (PLS) می باشد. به عبارت دقیق تر حاصل جواب رگرسیون کمترین توان های دوم جریمه شده یک برآوردگر انقباضی است. تحت مفروضات مدل (۱.۲) در مسئله رگرسیون PLS به دنبال دستیابی به برآوردگر $\hat{\beta}^S$ هستیم، که جواب مسئله بهینه سازی زیر می باشد

$$\hat{\beta}^S = \underset{\beta}{\operatorname{argmin}} \left(RSS(\beta) + \sum_j P(\beta_j) \right) \quad (3.2)$$

در ادامه این فصل برآوردگر انقباضی بریج را مورد مطالعه و بررسی قرار می دهیم.

۲.۲ برآوردگر انقباضی بریج

فرانک و فریدمن [۱۴] مجموع توان دوم باقیمانده (۲.۲) را با در نظر گرفتن شرط کلی $\sum_{j=1}^p |\beta_j|^\gamma \leq t$ به ازای $\gamma \geq 0$ در نظر گرفتند و آن را به عنوان رگرسیون بریج ارائه نمودند. بنابراین چنانچه در (۳.۲) عبارت جریمه را به صورت $\sum_j P(\beta_j) = k \sum_{j=1}^p |\beta_j|^\gamma$ در نظر بگیریم، داریم

^۱Best linear unbiased estimator

^۲Stain estimator

^۳Principal component estimator

^۴Partial least squares estimator

^۵Liu type estimator

^۶Almost unbiased estimator

^۷Penalized least squares

$$\hat{\beta}_{Bridge} = \underset{\beta}{\operatorname{argmin}} \left(\operatorname{RSS}(\beta) + k \sum_{j=1}^p |\beta_j|^\gamma \right) \quad (4.2)$$

که در آن $k, t \geq 0$ پارامترهای تنظیم کننده^۸ بوده و میزان انقباض تحمیل شده به ضرایب را کنترل می‌کنند. لازم به ذکر است که انتخاب بهینه‌ی γ منجر به افزایش کارایی برآوردگر خواهد شد. اگر $\gamma = 2$ ، مسئله فوق به رگرسیون ریج تبدیل شده (بخش ۲.۳) و اگر $\gamma = 1$ آن‌گاه به رگرسیون لاسو تبدیل می‌شود (بخش ۳.۳). با توجه به ساختار برآوردگر بریج، مشخص می‌شود به ازای $\gamma \in (1, 2)$ این برآوردگر بین برآوردهای لاسو و ریج رفتار می‌کند و در واقع بین این دو برآوردگر پل برقرار می‌کند. از این رو نام آن را برآوردگر بریج نهاده‌اند (بریج در لغت^۹ به معنای پل است). برآوردگر بریج در حالت کلی (به ازای $\gamma \geq 0$) دارای صورت بسته‌ای نیست.

لازم به ذکر است فرانک و فریدمن [۱۴] فقط برآوردگر بریج را پیشنهاد دادند اما نتوانستند این مساله را در حالت کلی برای $\gamma \geq 0$ حل نمایند. در نهایت، فو [۱۶] این مساله را برای $\gamma \geq 1$ حل نمود و الگوریتم نیوتن رافسون^{۱۰} را برای برآوردگر بریج زمانی که $\gamma > 1$ است، ارائه نمود و همچنین یک الگوریتم پیش‌برنده^{۱۱} برای لاسو پیشنهاد کرد. در ادامه رفتار برآوردگر بریج برای $\gamma \geq 1$ مورد بررسی قرار گرفته است.

۱.۲.۲ حل مسئله بهینه سازی رگرسیون انقباضی بریج ($\gamma \geq 1$)

برای حل رگرسیون انقباضی بریج زمانی که $\gamma \geq 1$ است، براساس توضیحات بخش ۲.۲ بایستی مسئله بهینه سازی زیر را حل کنیم

$$\hat{\beta}_{Bridge} = \underset{\beta}{\operatorname{argmin}} \left((Y - X\beta)^T(Y - X\beta) + k \sum_{j=1}^p |\beta_j|^\gamma \right) \quad k \geq 0 \quad (5.2)$$

معادله‌ی (۵.۲) به ازای $\gamma \geq 1$ را، به صورت $\underset{\beta}{\operatorname{argmin}} G(\beta, X, Y, k, \gamma)$ باز نویسی می‌کنیم که در آن

$$G(\beta, X, Y, k, \gamma) = \operatorname{RSS}(\beta) + k \sum_{j=1}^p |\beta_j|^\gamma$$

تابع G نسبت به β محدب^{۱۲} است. اگر $\|\beta\| \rightarrow +\infty$ که در آن $\|\cdot\|$ تابع نرم اقلیدسی است، آن‌گاه $G \rightarrow +\infty$.

به ازای $\gamma \geq 1$ برای یافتن کمینه، باید از تابع $G(\beta, X, Y, k, \gamma)$ نسبت به β مشتق گرفته و برابر صفر قرار دهیم تا مقدار کمینه کننده‌ی تابع پیدا شود. تابع $\sum_{j=1}^p |\beta_j|^\gamma$ در $\beta_j = 0$ مشتق پذیر نیست و تنها

^۸Tuning parameters

^۹Word

^{۱۰} Newton-Raphson algorithm

^{۱۱}Shooting algorithm

^{۱۲}Convex

می‌توان مشتقات جزئی G را نسبت به β_j که $\beta_j \neq 0$ و $j = 1, \dots, p$ محاسبه نمود. فرض کنید

$$\begin{aligned} S_j(\beta, \mathbf{X}, \mathbf{Y}) &= \frac{\partial RSS}{\partial \beta_j} \\ &= \mathbf{x}_j^T \mathbf{x}_j \beta_j + \sum_{i \neq j} \mathbf{x}_j^T \mathbf{x}_i \beta_i - \mathbf{x}_j^T \mathbf{Y} \end{aligned} \quad (6.2)$$

که $\mathbf{x}_j^T = (x_{1j}, \dots, x_{nj})$. برای اثبات رابطه (۶.۲) به پیوست ۳ رجوع کنید. همچنین

$$d(\beta_j, k, \gamma) = k\gamma |\beta_j|^{\gamma-1} \text{Sgn}(\beta_j) \quad (7.2)$$

که در آن

$$\text{Sgn} = \begin{cases} -1 & \beta_j < 0 \\ 0 & \beta_j = 0 \\ 1 & \beta_j > 0 \end{cases}$$

با برابر صفر قرار دادن $\frac{\partial G}{\partial \beta_j}$ روابط زیر بدست می‌آیند

$$\begin{cases} S_1(\beta, \mathbf{X}, \mathbf{Y}) + d(\beta_1, k, \gamma) = 0 \\ \vdots \\ S_j(\beta, \mathbf{X}, \mathbf{Y}) + d(\beta_j, k, \gamma) = 0 \\ \vdots \\ S_p(\beta, \mathbf{X}, \mathbf{Y}) + d(\beta_p, k, \gamma) = 0 \end{cases} \quad (8.2)$$

مسأله‌ی (۵.۲) را می‌توان از طریق مسئله‌ی (۸.۲) حل کرد، که در بخش بعد به آن پرداخته خواهد شد. برای رابطه‌ی (۸.۲) قضیه زیر برای توابع S_j وجود دارد که قبل از مطرح کردن آن، پذیره‌های زیر را در نظر بگیرید.

فرض کنید β یک بردار p بعدی از فضای پارامتر $\mathcal{B}$ ، $\mathbf{X}$ ماتریس $n \times p$ و $\mathbf{Y}$ یک بردار n بعدی از فضای نمونه $\mathbb{R}^n$ باشد. برای مقادیر $\mathbf{Y}, \mathbf{X}$ و $k \geq 0$ و $\gamma \geq 1$ ثابت، توابع حقیقی مقدار $S_j(\beta, \mathbf{X}, \mathbf{Y})$ ، $S_j(\cdot, \mathbf{X}, \mathbf{Y}) : \mathcal{B} \rightarrow \mathbb{R}$ به ازای $j = 1, \dots, p$ را در نظر بگیرید و $\mathbf{S}$ را به صورت $\mathbf{S} = (S_1, S_2, \dots, S_p)^T$ نشان می‌دهیم.

قضیه ۱.۲.۲. (فو [۱۶]) برای $\gamma > 0$ و $k > 0$ اگر تابع $\mathbf{S}$ نسبت به β به‌طور پیوسته مشتق‌پذیر^{۱۳} باشد و ماتریس ژاکوبی $(\partial \mathbf{S} / \partial \beta)$ نیمه‌معین مثبت باشد^{۱۴}، آن‌گاه:

۱. معادله‌ی (۸.۲) دارای جواب منحصر به فرد $\hat{\beta}(k, \gamma)$ است که در (k, γ) پیوسته است.

۲. $\lim_{\gamma \rightarrow 1^+} \hat{\beta}(k, \gamma)$ وجود دارد و دارای جواب منحصر به فرد است.

^{۱۳}Continuously differentiable

^{۱۴}Positive-Semi-Definite

۲.۲.۲ الگوریتم محاسبه برآوردگر بریج

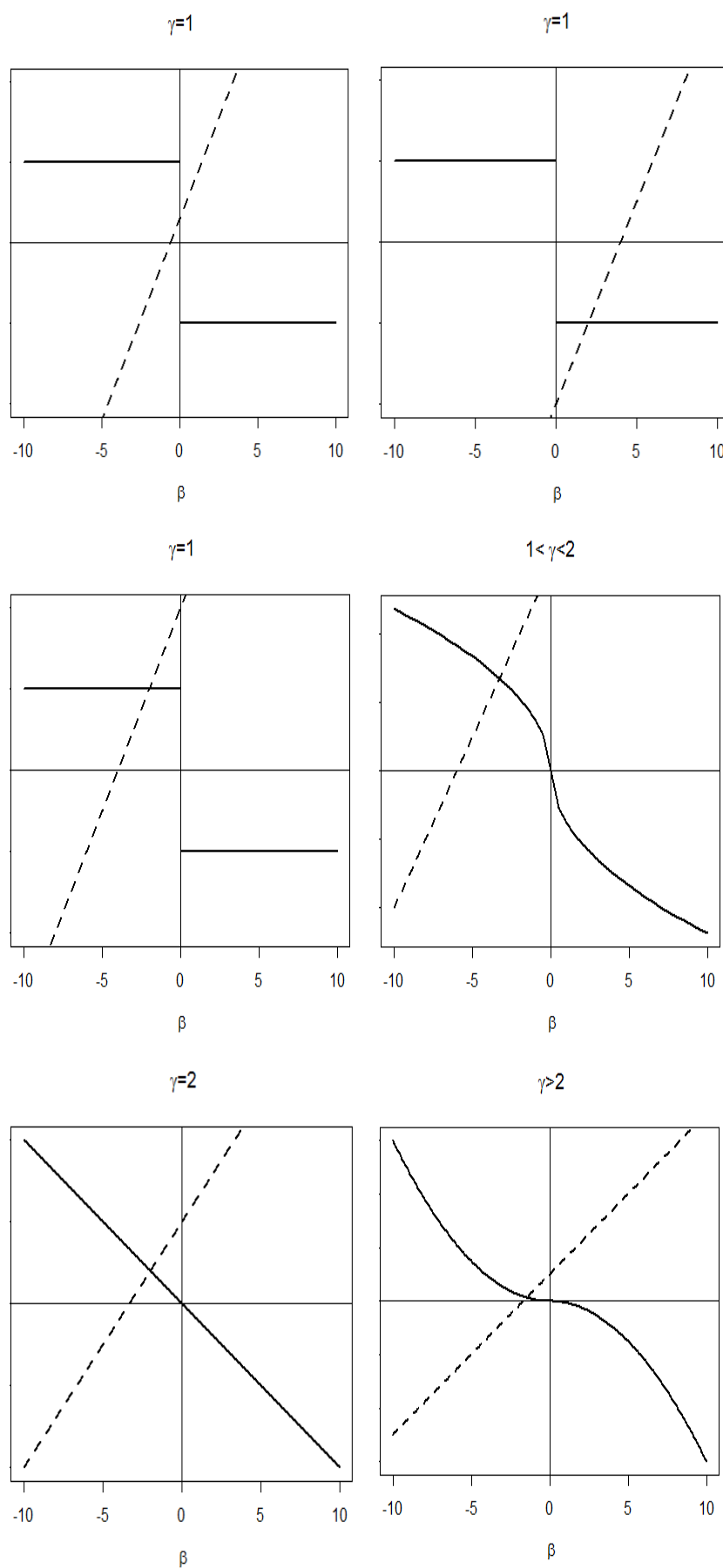
برای حل رگرسیون بریج زمانی که $\gamma \geq 1$ و $k > 0$ ، معادله‌ی (۸.۲) را در نظر بگیرید. فرض کنید بردار β را به صورت $\beta = (\beta_j, \beta^{-j})^T$ بازنویسی کنیم که در آن β^{-j} شامل $p - 1$ بردار مرکب از β_i ها به غیر از β_j می‌باشد. برای j امین معادله در (۸.۲) داریم

$$S_j = S_j(\beta_j, \beta^{-j}, \mathbf{X}, \mathbf{Y}) = -d(\beta_j, k, \gamma) = -d \quad (9.2)$$

در نتیجه با استفاده از (۶.۲) و (۷.۲) رابطه فوق به صورت زیر به دست می‌آید

$$2\mathbf{x}_j^T \mathbf{x}_j \beta_j + \sum_{i \neq j} 2\mathbf{x}_j^T \mathbf{x}_i \beta_i - 2\mathbf{x}_j^T \mathbf{Y} = -k\gamma |\beta_j|^{\gamma-1} \text{Sgn}(\beta_j) \quad (10.2)$$

معادله‌ی سمت چپ (۱۰.۲) برای β^{-j} ثابت، یک تابع خطی از β_j می‌باشد که دارای شیب مثبت $2\mathbf{x}_j^T \mathbf{x}_j$ است. اما معادله‌ی سمت راست یک تابع غیرخطی نسبت به β_j می‌باشد و این معادله برای $\gamma > 1$ به طور یکنواخت نزولی، پیوسته، مشتق پذیر به جز در نقطه $\beta_j = 0$ به ازای $\gamma = 1$ مشتق پذیر نیست. این تابع پرسی به طول $2k$ در نقطه $\beta_j = 0$ زمانی که $\gamma = 1$ است، دارد. بنابراین معادله‌ی (۱۰.۲) برای $\gamma > 1$ دارای جواب یکتا و برای $\gamma = 1$ دارای جواب منحصر به فرد یا بدون جواب می‌باشد. شکل ۱.۲ نحوه برخورد معادله $f_1(\beta) = -d$ و $f_2(\beta) = S_j$ را به ازای γ های مختلف نشان می‌دهد.



شکل ۱.۲: نحوه‌ی برخورد تابع d - (خط پر) با خط S_j (نقطه چین) در معادله‌ی (۱۰.۲) برای $\gamma \geq 1$.

از روش نیوتن رافسون برای محاسبات برآوردگر بریج زمانی که $\gamma > 1$ است می‌توان استفاده نمود. و از آنجایی که تابع $d(\beta_j, k, \gamma)$ برای $1 < \gamma < 2$ و $\beta_j = 0$ ، مشتق پذیر نیست، این روش نیازمند اصلاح برای همگرایی جواب است. در زیر روش نیوتن رافسون اصلاح شده^{۱۵}، برای $\gamma > 1$ ، با حل تکراری برای به دست آوردن جواب منحصر به فرد زامین معادله‌ی (۵.۲) ارائه شده است.

الگوریتم ۱. الگوریتم اصلاح شده نیوتن رافسون برآوردگر بریج برای $\gamma > 1$

(۱) با $\hat{\beta}_0 = \hat{\beta}_{OLS} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$ شروع می‌کنیم.

(۲) در گام m ام، برای هر $j = 1, \dots, p$ ، فرض کنید $S_j(\cdot, \hat{\beta}^{-j}, \mathbf{X}, \mathbf{Y})$ ، اگر $S_0 = 0$ باشد آن‌گاه $\hat{\beta}_j = 0$ درغیراینصورت اگر $\gamma \geq 2$ ، از روش نیوتن رافسون برای بدست آوردن جواب منحصر به فرد $\hat{\beta}_j$ از معادله‌ی (۱۰.۲) استفاده می‌کنیم، در غیر اینصورت اگر $\gamma < 2$ ، تابع $-d$ را با تانژانت نقطه‌ی حاصل از تقاطع خط مماس بر $-d$ و S_j تعویض کرده و این‌چنین روش نیوتن رافسون را اصلاح می‌کنیم. برآوردگر جدید $\hat{\beta}_m = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$ بعد از به‌روز کردن همه‌ی $\hat{\beta}_j$ ها به‌دست می‌آید.

(۳) مرحله‌ی (۲) را تا زمانی که $\hat{\beta}_m$ همگرا شود، تکرار می‌کنیم.

توجه نمایید $\hat{\beta}_{OLS}$ در اکثر اوقات وجود دارد. اما در حالت کلی می‌توان از نقاط شروع دیگری به‌جای $\hat{\beta}_{OLS}$ استفاده نمود، به عنوان مثال در صورت وجود همخطی می‌توان از برآوردگر ریدج استفاده نمود تا دقت برآوردگر افزایش یابد.

۳.۲.۲ واریانس برآوردگر بریج

در این بخش واریانس برآوردگر بریج را در قالب یک قضیه ارائه می‌کنیم.

قضیه ۲.۲.۲. (فو [۱۶]) تحت مفروضات مدل (۱.۲) و معادله (۴.۲)، واریانس برآوردگر بریج برای $\gamma > 1$ به‌صورت زیر است

$$Var(\hat{\beta}_{Bridge}) = \sigma^2 \left(\mathbf{X}^T \mathbf{X} + kD^*(\hat{\beta})|_{\mathbf{y}_0} \right)^{-1} \mathbf{X}^T \mathbf{X} \left(\mathbf{X}^T \mathbf{X} + kD^*(\hat{\beta})|_{\mathbf{y}_0} \right)^{-1}$$

که در آن $D^*(\hat{\beta}) = \text{diag} \left(\frac{\gamma(\gamma-1)}{2} |\hat{\beta}_j|^{\gamma-2} \right)$ و $\mathbf{y}_0$ نقطه‌ای ثابت در فضای نمونه است.

برهان.

فرض کنید $\mathbf{F} = (F_1, \dots, F_p)^T$ که در آن $F_j = S_j(\hat{\beta}, \mathbf{X}, \mathbf{Y}) + d(\hat{\beta}_j, k, \gamma)$ بنابراین
 $\frac{\partial \mathbf{F}}{\partial \hat{\beta}} = 2\mathbf{X}^T \mathbf{X} + 2kD^*(\hat{\beta})$ و $\frac{\partial \mathbf{F}}{\partial \mathbf{Y}} = -2\mathbf{X}^T$ که $\mathbf{F} = 2\mathbf{X}^T \mathbf{X} \hat{\beta} - 2\mathbf{X}^T \mathbf{Y} + k\gamma |\hat{\beta}|^{\gamma-1} \text{Sgn}(\hat{\beta})$ که در آن

$$D^*(\hat{\beta}) = \text{diag} \left(\frac{\gamma(\gamma-1)}{2} |\hat{\beta}_j|^{\gamma-2} \right)$$

^{۱۵}Modified Newton-Raphson Method

از آن جایی که $\frac{\partial \hat{\beta}}{\partial \mathbf{Y}} = \left(\frac{\partial \mathbf{F}}{\partial \hat{\beta}} \right)^{-1} \frac{\partial \mathbf{F}}{\partial \mathbf{Y}}$ با استفاده از لم ۱.۵.۱ می توان نتیجه گرفت

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}(f(\mathbf{Y})) = \left(\frac{\partial \hat{\beta}}{\partial \mathbf{Y}} \right)_{|\mathbf{y}_*} \text{Var}(\mathbf{Y}) \left(\frac{\partial \hat{\beta}}{\partial \mathbf{Y}} \right)^T_{|\mathbf{y}_*} \\ &= \sigma^2 \left(\mathbf{X}^T \mathbf{X} + kD^*(\hat{\beta})|_{\mathbf{y}_*} \right)^{-1} \mathbf{X}^T \mathbf{X} \left(\mathbf{X}^T \mathbf{X} + kD^*(\hat{\beta})|_{\mathbf{y}_*} \right)^{-1} \end{aligned}$$

□

۴.۲.۲ انتخاب پارامتر انقباضی γ و پارامتر کنترل k

برای انتخاب بهینه پارامترهای انقباضی γ و کنترل k از آماره اعتبار سنجی متقابل تعمیم یافته^{۱۶} (GCV) که کراون و وهبا [۱۳] و تیبشیرانی [۳۴] ارائه کردند، استفاده می کنیم. برای $k \geq 0$ و $\gamma \geq 1$ ، این معیار به صورت زیر تعریف می شود:

$$\text{GCV} = \frac{1}{n} \sum_{i=1}^n \left(\frac{e_{i,k}}{1 - \frac{p(k)}{n}} \right)^2 \quad (11.2)$$

که در آن به ازای $\gamma = 1$

$$p(k) = \text{tr} \left(\mathbf{X} (\mathbf{X}^T \mathbf{X} + k\Omega^-)^{-1} \mathbf{X}^T \right) - n.$$

و برای سایر مقادیر γ

$$p(k) = \text{tr} \left(\mathbf{X} (\mathbf{X}^T \mathbf{X} + k\Omega^{-\gamma})^{-1} \mathbf{X}^T \right)$$

که n تعداد $\hat{\beta}_j$ های برابر صفر و Ω^- معکوس تعمیم یافته ماتریس Ω است. همچنین $\Omega = \text{diag} \left(\frac{\gamma}{|\hat{\beta}_j|^{\gamma-2}} \right)$ و $e_{i,k} = y_i - \hat{y}_i^{\text{Bridge}}$ است. پارامترهای γ و k چنان انتخاب می شوند که آماره GCV مینیمم مقدار خود را اختیار کند.

^{۱۶}Generalized Cross Validation

۳.۲ مطالعه رفتار مجانبی

در این بخش رفتار مجانبی^{۱۷} برآوردگر بریج (۵.۲) را مورد بررسی قرار می‌دهیم. مدل خطی (۱.۲) را در نظر بگیرید، برای برآورد β از تابع هدف زیر استفاده می‌کنیم

$$Z_n(\phi) = \frac{1}{n} \text{RSS}(\phi) + \frac{k_n}{n} \sum_{j=1}^p |\phi_j|^\gamma \quad \forall \gamma > 0, k_n > 0, n \in \mathbb{N} \quad (12.2)$$

که در آن $\phi = (\phi_1, \dots, \phi_p)^T$ و $\text{RSS}(\phi)$ در رابطه (۲.۲) آمده است. همچنین فرض کنید شرایط نظم زیر برقرار باشند

$$C_n = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \rightarrow C, n \rightarrow \infty \quad (A1)$$

که در آن C ماتریس معلوم و نامنفی است و همچنین

$$\frac{1}{n} \max_{1 \leq i \leq n} \mathbf{x}_i^T \mathbf{x}_i \rightarrow 0, n \rightarrow \infty \quad (A2)$$

مدل خطی (۱.۲) دارای جواب منحصر به فرد است اگر ماتریس C_n نامنفرد یا ماتریس طرح (X) پر رتبه باشد. شایان ذکر است حتی اگر ماتریس C_n منفرد باشد کمینه مقدار (۱.۲) ممکن است وجود داشته و منحصر به فرد باشد. با این حال فرض کنید ماتریس C تعریف شده در شرط نظم (A1)، نامنفرد باشد. رفتار حدی برآوردگر بریج، $\hat{\beta}_n$ ، با مطالعه‌ی رفتار مجانبی تابع هدف (۱۲.۲) تعیین می‌شود. فرض کنید مقدار کمینه (۱۲.۲) به ازای $\phi = \hat{\beta}_n$ حاصل شود در قضیه زیر نشان می‌دهیم اگر $k_n = o(n)$ آن‌گاه $\hat{\beta}_n$ برای β سازگار است.

قضیه ۱.۳.۲. اگر ماتریس C تعریف شده در رابطه‌ی (A1)، نامنفرد باشد و $k_n \rightarrow k_0 \geq 0$ آن‌گاه

$$\text{argmin}(Z_n(\phi)) \xrightarrow{P} \text{argmin}(Z(\phi))$$

که در آن

$$Z(\phi) = \sigma^2 + (\phi - \beta)^T C (\phi - \beta) + k_0 \sum_{j=1}^p |\phi_j|^\gamma$$

بنابراین اگر $k_n = o(n)$ ، آن‌گاه $\text{argmin}(Z(\phi)) = \beta$ و بنابراین $\hat{\beta}_n \xrightarrow{P} \beta$.

برهان. فرض کنید

$$Z_n(\phi) = \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{x}_i^T \phi)^2 + \frac{k_n}{n} \sum_{j=1}^p |\phi_j|^\gamma \quad \gamma > 0 \quad (13.2)$$

^{۱۷}Asymptotic behavior

اثبات را برای دو حالت $\gamma \geq 1$ و $\gamma < 1$ بیان می‌کنیم. دقت کنید به ازای $\gamma \geq 1$ ، $Z_n(\phi)$ محدب بوده در صورتی که به ازای $\gamma < 1$ محدب نمی‌باشد.

با توجه به اینکه $Y_i = \mathbf{x}_i^T \beta + \epsilon_i$ ، معادله‌ی (۱۳.۲) را می‌توان به صورت زیر بازنویسی کرد

$$\begin{aligned} Z_n(\phi) &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i^T \beta + \epsilon_i - \mathbf{x}_i^T \phi)^2 + \frac{k_n}{n} \sum_{j=1}^p |\phi_j|^\gamma \\ &= \frac{1}{n} \sum_{i=1}^n (\epsilon_i - \mathbf{x}_i^T (\phi - \beta))^2 + \frac{k_n}{n} \sum_{j=1}^p |\phi_j|^\gamma \\ &= \frac{1}{n} \sum_{i=1}^n (\epsilon_i^2 - 2\epsilon_i \mathbf{x}_i^T (\phi - \beta) + (\phi - \beta)^T \mathbf{x}_i \mathbf{x}_i^T (\phi - \beta)) + \frac{k_n}{n} \sum_{j=1}^p |\phi_j|^\gamma \\ &= \frac{1}{n} \left(\sum_{i=1}^n \epsilon_i^2 - 2 \sum_{i=1}^n \epsilon_i \mathbf{x}_i^T (\phi - \beta) + (\phi - \beta)^T \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T (\phi - \beta) \right) + \frac{k_n}{n} \sum_{j=1}^p |\phi_j|^\gamma \end{aligned}$$

اگر $X_1, \dots, X_n$ در شرایط قانون ضعیف اعداد بزرگ صدق کنند (پیوست ب.۲.۱۰ را ببینید) و $S_n = \sum_{i=1}^n X_i$ ، آن‌گاه $\frac{S_n}{n} \xrightarrow{P} \mu$. با توجه به اینکه $E(\epsilon_i) = 0$ ، $Var(\epsilon_i) = \sigma^2$ ، لذا $E(\epsilon_i^2) = \sigma^2$ نتیجه می‌دهد $\frac{1}{n} \sum_{i=1}^n \epsilon_i^2 \xrightarrow{P} \sigma^2$ از طرفی، اگر $X_n \xrightarrow{P} X$ آن‌گاه $X_n \xrightarrow{D} X$ بنابراین داریم

$$\frac{1}{n} \sum_{i=1}^n \epsilon_i^2 \xrightarrow{D} \sigma^2$$

به طور مشابه از آنجایی که $E(\epsilon_i \mathbf{x}_i^T) = 0$ می‌توان نتیجه گرفت $\frac{1}{n} \sum_{i=1}^n \epsilon_i \mathbf{x}_i^T (\phi - \beta) \xrightarrow{D} 0$. تحت شرط نظم (A۱)، داریم

$$(\phi - \beta)^T \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T (\phi - \beta) \rightarrow (\phi - \beta)^T C (\phi - \beta) \quad (۱۴.۲)$$

از طرفی طبق فرض $\frac{k_n}{n} \rightarrow k_0$ ، لذا

$$\frac{k_n}{n} \sum_{j=1}^p |\phi_j|^\gamma \rightarrow k_0 \sum_{j=1}^p |\phi_j|^\gamma \quad (۱۵.۲)$$

با توجه به روابط (۱۴.۲) و (۱۵.۲) و با استفاده از قضیه‌ی اسلاتسکی ب.۲.۹ داریم

$$Z_n(\phi) \xrightarrow{P} \sigma^2 + (\phi - \beta)^T C (\phi - \beta) + k_0 \sum_{j=1}^p |\phi_j|^\gamma = Z(\phi)$$

پس $Z_n(\phi) - Z(\phi) \xrightarrow{P} 0$ و

$$\sup_{\phi \in \mathbb{E}} |Z_n(\phi) - Z(\phi)| \xrightarrow{P} 0$$

که در آن $\mathbb{E}$ یک مجموعه فشرده می‌باشد، بنابراین با استفاده از قاعده پیوستگی تابع احتمال ب.۲.۲ می‌توان نوشت

$$\forall \epsilon > 0 \quad \exists n, n_0 \in \mathbb{N} \quad s.t \quad \forall n \geq n_0, \quad \forall \phi \in \mathbb{E} \quad |Z_n(\phi) - Z(\phi)| \leq \epsilon$$

در نتیجه

$$Z_n(\phi) \xrightarrow{U} Z(\phi)$$

که منظور از نماد U همگرایی یکنواخت است و بنا به قضیه ب.۱۰.۴ کمینه $Z_n(\phi)$ نیز به کمینه $Z(\phi)$ همگراست.

از آنجایی که $\gamma \geq 1$ و $Z_n(\phi)$ محدب و دارای همگرایی یکنواخت و $\mathbb{E}$ فشرده است پس ϕ کراندار است. در نتیجه

$$\operatorname{argmin}(Z_n(\phi)) \xrightarrow{P} \operatorname{argmin}(Z(\phi))$$

اگر $\gamma < 1$ ، آن‌گاه $Z_n(\phi)$ محدب نمی‌باشد. اما چون ϕ کراندار است. و ϕ کمینه کننده‌ی $Z_n(\phi)$ است پس

$$Z_n(\phi) \geq \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{x}_i^T \beta)^2 = Z_{n,OLS} \quad (۱۶.۲)$$

و با توجه به کرانداری برآوردگر OLS طبق نابرابری (۱۶.۲)

$$\hat{\beta}_n = O_p(1)$$

بنابراین

$$\operatorname{argmin}(Z_n(\phi)) \xrightarrow{P} \operatorname{argmin}(Z(\phi))$$

در نتیجه برای هر $\gamma > 0$ ،

$$\operatorname{argmin}(Z_n(\phi)) \xrightarrow{P} \operatorname{argmin}(Z(\phi))$$

حال سازگاری $\hat{\beta}_n$ را اثبات می‌کنیم. فرض کنید $\frac{k_n}{n} \rightarrow 0$ پس $k_n = o(n)$. همچنین

$$Z(\phi) = \sigma^2 + (\phi - \beta)^T C (\phi - \beta)$$

از آنجایی که $Z(\phi)$ کمترین مقدار را دارد که این مقدار یکتاست و همچنین تابع هدف به طور یکنواخت همگراست پس اثبات کامل است. $\square$

یعنی اگر $k_n = o(n)$ و $\operatorname{argmin}(Z(\phi)) = \beta$ پس $\hat{\beta}_n$ سازگار است.

در دو قضیه‌ی زیر، توزیع مجانبی برآوردگر بریج را برای $\gamma \geq 1$ (قضیه ۲.۳.۲) و $\gamma < 1$ (قضیه ۳.۳.۲) ارائه می‌کنیم.

قضیه ۲.۳.۲. (فو [۱۶]) فرض کنید $\gamma \geq 1$ ، اگر $k_0 \geq 0$ و $\frac{k_n}{\sqrt{n}} \rightarrow k_0$ نامنفرد باشد، آنگاه

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} \operatorname{argmin}(\mathbf{V}(\mathbf{u}))$$

که در آن به ازای $\gamma > 1$

$$\mathbf{V}(\mathbf{u}) = -\mathbf{u}'\mathbf{W} + \mathbf{u}'\mathbf{C}\mathbf{u} + k_0 \sum_{j=1}^p \gamma u_j \operatorname{Sgn}(\beta_j) |\beta_j|^{\gamma-1}$$

و به ازای $\gamma = 1$

$$\mathbf{V}(\mathbf{u}) = -\mathbf{u}'\mathbf{W} + \mathbf{u}'\mathbf{C}\mathbf{u} + k_0 \sum_{j=1}^p (u_j \operatorname{Sgn}(\beta_j) I(\beta_j \neq 0) + |u_j| I(\beta_j = 0))$$

و $\mathbf{W}$ دارای توزیع $N(0, \sigma^2 \mathbf{C})$ و $\mathbf{u} = (u_1, \dots, u_p)^T$ می‌باشد.

برهان. فرض کنید

$$\begin{aligned} \mathbf{V}_n(\mathbf{u}) &= \text{مجموع توان دوم تاوانیده } \beta - \text{مجموع توان دوم تاوانیده } \phi \\ &= \sum_{i=1}^n (y_i - \mathbf{x}_i^T \phi)^2 + k_n \sum_{j=1}^p |\phi_j|^\gamma - \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2 - k_n \sum_{j=1}^p |\beta_j|^\gamma \\ &= \sum_{i=1}^n (\mathbf{x}_i^T \beta + \epsilon_i - \mathbf{x}_i^T \phi)^2 + k_n \sum_{j=1}^p |\beta_j - \beta_j + \phi_j|^\gamma \\ &\quad - \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2 - k_n \sum_{j=1}^p |\beta_j|^\gamma \\ &= \sum_{i=1}^n \left(\epsilon_i - \mathbf{x}_i^T (\phi - \beta) \frac{\sqrt{n}}{\sqrt{n}} \right)^2 + k_n \sum_{j=1}^p \left| \beta_j + (\phi_j - \beta_j) \frac{\sqrt{n}}{\sqrt{n}} \right|^\gamma \\ &\quad - \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2 - k_n \sum_{j=1}^p |\beta_j|^\gamma \end{aligned}$$

با فرض $\mathbf{u} = \sqrt{n}(\phi - \beta)$ در این صورت تابع فوق به صورت زیر مجدداً پارامتری می‌شود:

$$\begin{aligned} \mathbf{V}_n(\mathbf{u}) &= \sum_{i=1}^n \left(\epsilon_i - \frac{\mathbf{x}_i^T \mathbf{u}}{\sqrt{n}} \right)^2 + k_n \sum_{j=1}^p \left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - \sum_{i=1}^n \epsilon_i^2 - k_n \sum_{j=1}^p |\beta_j|^\gamma \\ &= \sum_{i=1}^n \left((\epsilon_i - \frac{\mathbf{x}_i^T \mathbf{u}}{\sqrt{n}})^2 - \epsilon_i^2 \right) + k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right) \\ &= \sum_{i=1}^n \left(\epsilon_i^2 - 2 \frac{\epsilon_i \mathbf{x}_i^T \mathbf{u}}{\sqrt{n}} + \frac{\mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u}}{n} - \epsilon_i^2 \right) + k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right) \\ &= -2 \sum_{i=1}^n \frac{\epsilon_i \mathbf{x}_i^T \mathbf{u}}{\sqrt{n}} + \sum_{i=1}^n \frac{\mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u}}{n} + k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right) \end{aligned}$$

با توجه به اینکه $\epsilon_i \sim N(0, \sigma^2)$ داریم

$$E(\mathbf{W}) = E\left(\sum_{i=1}^n \frac{\mathbf{x}_i \epsilon_i}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i E(\epsilon_i) = 0$$

و با توجه به شرط نظم (A۱)

$$Var(\mathbf{W}) = Var\left(\sum_{i=1}^n \frac{\mathbf{x}_i \epsilon_i}{\sqrt{n}}\right) = \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^T Var(\epsilon_i)}{\sqrt{n} \sqrt{n}} = \sigma^2 \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^T}{n} \rightarrow \sigma^2 \mathbf{C}$$

بنابراین $\mathbf{W} = \sum \frac{\mathbf{x}_i \epsilon_i}{\sqrt{n}} \sim N_p(0, \sigma^2 \mathbf{C})$ و همچنین

$$-\frac{1}{2} \sum_{i=1}^n \frac{\epsilon_i \mathbf{x}_i^T \mathbf{u}}{\sqrt{n}} + \sum_{i=1}^n \frac{\mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u}}{n} \xrightarrow{D} -\frac{1}{2} \mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u}$$

فرض کنید

$$f(x) = k_n \sum_{j=1}^p (|x_j|^\gamma - |\beta_j|^\gamma)$$

با توجه به بسط تیلور مرتبه اول [۳.۵.ب](#) تابع $f(\cdot)$ ، در نقطه $x_0 = \beta_j$ و به ازای $x = \beta_j + \frac{u_j}{\sqrt{n}}$ ، داریم

$$\begin{aligned} f(x) &= k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right) = k_n \sum_{j=1}^p (|\beta_j|^\gamma - |\beta_j|^\gamma) + k_n \sum_{j=1}^p \frac{u_j}{\sqrt{n}} \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \\ &= k_n \sum_{j=1}^p \frac{u_j}{\sqrt{n}} \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \\ &= \frac{k_n}{\sqrt{n}} \sum_{j=1}^p u_j \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \end{aligned}$$

که در آن $\beta_j^* \in (\beta_j, \beta_j + \frac{u_j}{\sqrt{n}})$. طبق فرض $\frac{k_n}{\sqrt{n}} \rightarrow k_0$ ، بنابراین می توان نتیجه گرفت

$$\frac{k_n}{\sqrt{n}} \sum_{j=1}^p u_j \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \rightarrow k_0 \sum_{j=1}^p u_j \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1}$$

با توجه به اینکه $\beta_j^* \in (\beta_j, \beta_j + \frac{u_j}{\sqrt{n}})$ با افزایش n ($n \rightarrow \infty$)، لذا $\beta_j^* \rightarrow \beta_j$.

$$\frac{k_n}{\sqrt{n}} \sum_{j=1}^p u_j \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \rightarrow k_0 \sum_{j=1}^p u_j \gamma \operatorname{Sgn}(\beta_j) |\beta_j|^{\gamma-1}$$

که نتیجه می دهد

$$\mathbf{V}_n(\mathbf{u}) \xrightarrow{D} \mathbf{V}(\mathbf{u})$$

که در آن

$$\mathbf{V}_n(\mathbf{u}) = -\frac{1}{2} \sum_{i=1}^n \frac{\epsilon_i \mathbf{x}_i^T \mathbf{u}}{\sqrt{n}} + \sum_{i=1}^n \frac{\mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u}}{n} + k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right) \quad (۱۷.۲)$$

$$\mathbf{V}(\mathbf{u}) = -\mathbf{u}'\mathbf{W} + \mathbf{u}'\mathbf{C}\mathbf{u} + k_{\circ} \sum_{j=1}^p \gamma u_j \text{Sgn}(\beta_j) |\beta_j|^{\gamma-1}$$

می باشد.

اگر $\gamma > 1$ ، آن گاه تابع $\mathbf{V}_n(\mathbf{u})$ محدب بوده و دارای یک مقدار کمینه منحصر به فرد است. پس از $\mathbf{V}_n(\mathbf{u}) \xrightarrow{D} \mathbf{V}(\mathbf{u})$ نتیجه می گیریم

$$\text{argmin}(\mathbf{V}_n(\mathbf{u})) = \sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} \text{argmin}(\mathbf{V}(\mathbf{u}))$$

برای $\gamma = 1$ ، دو حالت $\beta_j = 0$ و $\beta_j \neq 0$ در نظر گرفتن رابطه (۱۷.۲) به صورت زیر است:

$$\mathbf{V}_n(\mathbf{u}) \xrightarrow{D} -\mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u} + k_{\circ} \sum_{j=1}^p u_j \text{Sgn}(\beta_j)$$

حالت دوم، $\gamma = 1$ و $\beta_j = 0$:

$$\mathbf{V}_n(\mathbf{u}) \xrightarrow{D} -\mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u} + k_{\circ} \sum_{j=1}^p |u_j|$$

بنابراین برای $\gamma = 1$ در حالت کلی داریم

$$\mathbf{V}_n(\mathbf{u}) \xrightarrow{D} \mathbf{V}(\mathbf{u}) = -\mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u} + k_{\circ} \sum_{j=1}^p (u_j \text{Sgn}(\beta_j) I(\beta_j \neq 0) + |u_j| I(\beta_j = 0))$$

پس

$$\text{argmin}(\mathbf{V}_n(\mathbf{u})) \xrightarrow{D} \text{argmin}(\mathbf{V}(\mathbf{u}))$$

در نتیجه برای هر $\gamma \geq 1$ ،

$$\text{argmin}(\mathbf{V}_n(\mathbf{u})) \xrightarrow{D} \text{argmin}(\mathbf{V}(\mathbf{u}))$$

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} \text{argmin}(\mathbf{V}(\mathbf{u}))$$

□

اثبات کامل است.

قضیه ۳.۳.۲. فرض کنید $\gamma < 1$ اگر $k_{\circ} \geq 0$ ، $\frac{k_n}{n^{\frac{\gamma}{1-\gamma}}} \rightarrow k_{\circ}$ ، آن گاه

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} \text{argmin}(\mathbf{V}(\mathbf{u}))$$

که در آن

$$\mathbf{V}(\mathbf{u}) = -\mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u} + k_{\circ} \sum_{j=1}^p |u_j|^{\gamma} I(\beta_j = 0)$$

برهان. فرض کنید

$$\mathbf{V}_n(\mathbf{u}) = -\mathfrak{z} \sum_{i=1}^n \frac{\epsilon_i \mathbf{x}_i^T \mathbf{u}}{\sqrt{n}} + \sum_{i=1}^n \frac{\mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u}}{n} + k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right)$$

همانند اثبات قضیه ۲.۳.۲ می توان نشان داد

$$-\mathfrak{z} \sum_{i=1}^n \frac{\epsilon_i \mathbf{x}_i^T \mathbf{u}}{\sqrt{n}} + \sum_{i=1}^n \frac{\mathbf{u}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{u}}{n} \xrightarrow{D} -\mathfrak{z} \mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u}$$

برای $k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right)$ دو حالت $\beta_j \neq 0$ و $\beta_j = 0$ را در نظر می گیریم
حالت اول اگر $\beta_j \neq 0$ باشد، فرض کنید

$$f(x) = k_n \sum_{j=1}^p (|x_j|^\gamma - |\beta_j|^\gamma)$$

با توجه به بسط تیلور مرتبه اول ۳.۵.ب تابع $f(\cdot)$ ، در نقطه $x_0 = \beta_j$ و به ازای $x = \beta_j + \frac{u_j}{\sqrt{n}}$ داریم

$$\begin{aligned} f(x) &= k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right) = k_n \sum_{j=1}^p (|\beta_j|^\gamma - |\beta_j|^\gamma) + k_n \sum_{j=1}^p \frac{u_j}{\sqrt{n}} \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \\ &= k_n \sum_{j=1}^p \frac{u_j}{\sqrt{n}} \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \\ &= \frac{k_n}{\sqrt{n}} \sum_{j=1}^p u_j \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \end{aligned}$$

با توجه به اینکه $\beta_j^* \in (\beta_j, \beta_j + \frac{u_j}{\sqrt{n}})$ با افزایش n ($n \rightarrow \infty$)، $\beta_j^* \rightarrow \beta_j$ و همچنین برای $\gamma < 1$ ،
طبق فرض $\frac{k_n}{\sqrt{n}}$ به k_0 میل می کند پس $\frac{k_n}{\sqrt{n}} \rightarrow 0$ لذا

$$\frac{k_n}{\sqrt{n}} \sum_{j=1}^p u_j \gamma \operatorname{Sgn}(\beta_j^*) |\beta_j^*|^{\gamma-1} \rightarrow 0 \times \sum_{j=1}^p u_j \gamma \operatorname{Sgn}(\beta_j) |\beta_j|^{\gamma-1} = 0$$

که نتیجه می دهد

$$\mathbf{V}_n(\mathbf{u}) \xrightarrow{D} -\mathfrak{z} \mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u}$$

حالت دوم، اگر $\beta_j = 0$ باشد، آن گاه

$$k_n \sum_{j=1}^p \left(\left| \beta_j + \frac{u_j}{\sqrt{n}} \right|^\gamma - |\beta_j|^\gamma \right)_{\beta_j=0} = k_n \sum_{j=1}^p \left(\left| \frac{u_j}{\sqrt{n}} \right|^\gamma \right) = \frac{k_n}{n^{\frac{\gamma}{2}}} \sum_{j=1}^p |u_j|^\gamma \xrightarrow{D} k_0 \sum_{j=1}^p |u_j|^\gamma$$

بنابراین

$$\mathbf{V}_n(\mathbf{u}) \xrightarrow{D} \mathbf{V}(\mathbf{u})$$

که

$$\mathbf{V}(\mathbf{u}) = -\mathfrak{z} \mathbf{u}^T \mathbf{W} + \mathbf{u}^T \mathbf{C} \mathbf{u} + k_0 \sum_{j=1}^p |u_j|^\gamma I(\beta_j = 0)$$

از انجایی که همگرایی یکنواخت است و $\mathbf{u}$ متعلق مجموعه‌ای فشرده است. بنابراین

$$\operatorname{argmin}(\mathbf{V}_n(\mathbf{u})) \xrightarrow{D} \operatorname{argmin}(\mathbf{V}(\mathbf{u}))$$

و

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{D} \operatorname{argmin}(\mathbf{V})$$

□

بنابراین اثبات کامل است.

فصل ۳

حالات خاص برآوردگر بریج

در این فصل، خواص دو حالت خاص برآوردگر بریج، برآوردگرهای ریج و لاسو مورد بررسی قرار گرفته‌اند. همچنین پکیج‌های مورد استفاده از این دو برآوردگر را در زبان برنامه‌نویسی R شرح می‌دهیم. مطالب این فصل، عمدتاً از مقاله‌های هورل و کنارد [۲۰] و تیبشیرانی [۳۴] گرفته شده‌اند.

۱.۳ مقدمه

در بخش‌های قبل گفتیم که وقوع همخطی چندگانه بین متغیرهای پیش‌بین در تحلیل رگرسیون خطی ممکن است باعث ناپایداری شدید در برآورد کم‌ترین توان‌های دوم پارامترها شود. در نتیجه، برآورد کم‌ترین توان‌های دوم بدست آمده، قابل اعتماد نخواهد بود. توجه نمایید که دو نوع همخطی در داده‌ها وجود دارند: همخطی کامل و همخطی ناقص. در صورت وجود همخطی کامل، دترمینان ماتریس $X^T X$ صفر است و لذا معکوس ماتریس $X^T X$ وجود ندارد و در نتیجه جواب یکتا وجود نخواهد داشت و می‌توان به جای آن از معکوس تعمیم یافته $X^T X$ استفاده نمود. همخطی ناقص باعث می‌شود دترمینان ماتریس $X^T X$ بسیار کوچک شود و در نتیجه ماتریس $(X^T X)^{-1}$ دارای مولفه‌هایی باشد که قدر مطلق آن‌ها بسیار بزرگ هستند، از طرفی واریانس برآوردگر $\hat{\beta}$ زیاد می‌شود. در این صورت، $\hat{\beta}$ برآورد دقیقی از β نخواهد بود. به عبارت دیگر، دقت برآورد ضرایب رگرسیون کم می‌شود. همچنین رفتار ناپایداری دارد.

در واقعیت با داده‌هایی که در آن‌ها همخطی کامل وجود داشته باشد، مواجه نمی‌شویم. لذا اغلب مسائلی که با آن‌ها سروکار داریم، داده‌هایی هستند که برای آن‌ها دترمینان ماتریس $X^T X$ نزدیک به

صفر است. برای این اساس، در این پایان نامه همخطی ناقص مورد بررسی قرار گرفته است. برای حل این مشکل با توجه به اینکه روش‌های انقباضی کارایی پارامتر مکان را افزایش می‌دهند، بنابراین برآوردگر انقباضی بریج را برای بهبود مدل معرفی کردیم و آن را شرح دادیم. از آنجایی که این برآوردگر حالت کلی‌تری از برآوردگر ریج است بنابراین در این بخش به بررسی حالات خاص برآوردگر بریج، با تکیه بر برآوردگرهای ریج و لاسو پرداخته و این دو برآوردگر را به صورت دقیق‌تر مورد بررسی قرار خواهیم داد.

۲.۳ برآوردگر ریج

اگر توان دوم تفاضل β از مقدار واقعی آن را به صورت زیر در نظر بگیریم، خواهیم داشت

$$L_1^* = (\hat{\beta} - \beta)^T (\hat{\beta} - \beta)$$

در این صورت

$$\begin{aligned} E(L_1^*) &= E \left((\hat{\beta} - \beta)^T (\hat{\beta} - \beta) \right) = \sum_{j=1}^p E(\hat{\beta}_j - \beta_j)^2 \\ &= \sum_{j=1}^p \text{Var}(\hat{\beta}_j) = \sigma^2 \text{tr}((\mathbf{X}^T \mathbf{X})^{-1}) = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j} \end{aligned}$$

که در آن λ_j ها مقادیر ویژه ماتریس $\mathbf{X}^T \mathbf{X}$ می‌باشند.

در صورت وجود همخطی، حداقل یکی از مقادیر ویژه بسیار کوچک خواهد بود. در این صورت امید ریاضی توان دوم فاصله خیلی بزرگ می‌شود. بنابراین برآورد کم‌ترین توان‌های دوم به طور متوسط خیلی بزرگ خواهد شد؛ لذا از مقدار واقعی خود بسیار متفاوت خواهد بود. با توجه به اینکه $E(L_1^*) = E(\hat{\beta}^T \hat{\beta}) - \beta^T \beta$ داریم

$$E(\hat{\beta}^T \hat{\beta}) = \beta^T \beta + \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j}$$

این رابطه نشان می‌دهد که در صورت وجود همخطی، برآورد کم‌ترین توان‌های دوم β به طور متوسط دارای طولی بزرگتر از مقدار واقعی می‌باشد. یک روش برای رفع مشکل، یافتن برآوردگری اریب است که میانگین توان دوم خطای کمتری نسبت به برآوردگر نارایب $\hat{\beta}^*$ داشته باشد. اگر این برآوردگر را با $\hat{\beta}^*$ نشان دهیم، میانگین توان دوم خطای برآوردگر $\hat{\beta}^*$ به صورت زیر تعریف می‌شود:

$$\begin{aligned} \text{MSE}(\hat{\beta}^*) &= E \left((\hat{\beta}^* - \beta)^T (\hat{\beta}^* - \beta) \right) \\ &= \text{tr}(\text{Cov}(\hat{\beta}^*)) + \text{Bias}^T(\hat{\beta}^*) \text{Bias}(\hat{\beta}^*) \end{aligned}$$

که در آن $\text{Bias}(\hat{\beta}^*) = E(\hat{\beta}^* - \beta)$ مقدار اریبی برآوردگر $\hat{\beta}^*$ است. با کوچک در نظر گرفتن میزان اریبی $\hat{\beta}^*$ ، واریانس $\hat{\beta}^*$ را می‌توان کوچک کرد، به طوری که $\text{MSE}(\hat{\beta}^*)$ کمتر از واریانس برآوردگر نارایب $\text{MSE}(\hat{\beta})$ شود.

روش‌های متعددی برای غلبه با مشکل همخطی در مدل‌های خطی ارائه شده است. یکی از موثرترین روش‌ها برای حل مشکل همخطی، رگرسیون ريج است که توسط هورل و کنارد [۲۰] معرفی شد. اگر در رابطه (۲.۲)، RSS را نسبت به $\sum_{j=1}^p \beta_j^2 \leq t$ کمینه کنیم، برآوردگری که به دست می‌آید برآوردگر ريج^۱ نام دارد، برآوردگر ريج به پارامتر ريج k بستگی دارد که با افزودن این پارامتر به ماتریس $X^T X$ به صورت زیر باعث بهبود وضعیت آن می‌شود

$$\begin{aligned}\hat{\beta}_{\text{Ridge}} &= \underset{\beta}{\operatorname{argmin}} \left(\text{RSS} + k \sum_{j=1}^p \beta_j^2 \right) \\ &= (X^T X + kI_p)^{-1} X^T Y \\ &= (X^T X + kI_p)^{-1} (X^T X) \hat{\beta}_{OLS} \\ &= T(k) \hat{\beta}_{OLS}\end{aligned}\quad (۱.۳)$$

که در آن $T(k) = (k(X^T X)^{-1} + I_p)^{-1}$ و $k, t \geq 0$ پارامترهای تنظیم کننده^۲ بوده و میزان انقباض ضرایب را کنترل می‌کنند. برآوردگر ريج یک تبدیل خطی از برآوردگر کم‌ترین توان‌های دوم است. امید ریاضی $\hat{\beta}_{\text{Ridge}}$ عبارت است از

$$E(\hat{\beta}_{\text{Ridge}}) = E((X^T X + kI_p)^{-1} X^T Y) = (X^T X + kI_p)^{-1} X^T X \beta$$

بنابراین، $\hat{\beta}_{\text{Ridge}}$ برآوردگر اریب β است. همچنین ماتریس واریانس – کواریانس آن به صورت زیر می‌باشد

$$\text{Cov}(\hat{\beta}_{\text{Ridge}}) = \sigma^2 (X^T X + kI)^{-1} X^T X (X^T X + kI)^{-1}$$

میانگین توان‌های دوم خطای برآوردگر ريج عبارت است از

$$\begin{aligned}\text{MSE}(\hat{\beta}_{\text{Ridge}}) &= E((\hat{\beta}_{\text{Ridge}} - \beta)^T (\hat{\beta}_{\text{Ridge}} - \beta)) \\ &= \text{tr}(\text{Cov}(\hat{\beta}_{\text{Ridge}})) + \text{Bias}^T(\hat{\beta}_{\text{Ridge}}) \text{Bias}(\hat{\beta}_{\text{Ridge}}) \\ &= \sigma^2 \text{tr}((X^T X + kI)^{-1} X^T X (X^T X + kI)^{-1}) \\ &\quad + k^2 \beta^T (X^T X + kI)^{-1} \beta\end{aligned}\quad (۲.۳)$$

اولین جمله سمت راست (۲.۳) مجموع واریانس‌های مولفه‌های پارامتر $\hat{\beta}_{\text{Ridge}}$ و دومین عبارت، توان دوم اریبی (ترانهاد اریبی در خودش) است. بنابراین اریبی $\hat{\beta}_{\text{Ridge}}$ با افزایش k ، افزایش و واریانس آن با افزایش k ، کاهش می‌یابد و برعکس. با استفاده از تعریف ب.۳۷.۱ ماتریس $X^T X + kI_p$ را می‌توان به صورت $\Gamma(\Lambda + kI_p)\Gamma^T$ تجزیه کرد که در آن، Γ یک ماتریس متعامد است و ستون‌های آن بردارهای ویژه ماتریس $X^T X + kI_p$ است و $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_p)$ مقادیر ویژه ماتریس $X^T X$ است و

^۱Ridge Regression

^۲Tuning Parameters

می‌باشد. پس داریم $(\lambda_1 > \lambda_2 > \dots > \lambda_p \geq 0)$

$$\begin{aligned} \text{MSE}(\hat{\beta}_{\text{Ridge}}) &= \sigma^2 \text{tr} \left((\Gamma(\Lambda + k\mathbf{I}_p))^{-1} \Gamma^T \Gamma \Lambda \Gamma^T \Gamma (\Lambda + k\mathbf{I}_p)^{-1} \Gamma^T \right) \\ &\quad + k^2 \beta^T \Gamma (\Lambda + k\mathbf{I}_p)^{-1} \Gamma^T \beta \\ &= \sigma^2 \text{tr} \left((\Lambda + k\mathbf{I}_p)^{-1} \Lambda (\Lambda + k\mathbf{I}_p)^{-1} \right) + k^2 \alpha^T (\Lambda + k\mathbf{I}_p)^{-1} \alpha \\ &= \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2} \end{aligned}$$

که در آن $\alpha = \Gamma^T \beta$ است. در رگرسیون ریج، پارامتر k طوری انتخاب می‌شود که کاهش در واریانس بیشتر از افزایش توان دوم اریبی باشد. در این صورت میانگین توان دوم خطای برآورد ریج $\hat{\beta}_{\text{Ridge}}$ کمتر از واریانس برآوردگر کمترین توان‌های دوم $\hat{\beta}_{\text{OLS}}$ خواهد بود. هورل و کنارد [۲۰] ثابت کردند که اگر $\beta^T \beta$ کراندار باشند، مقدار غیرصفری از k وجود دارد به طوری که مخاطره $\hat{\beta}_{\text{Ridge}}$ از مخاطره برآوردگر کمترین توان‌های دوم $\hat{\beta}_{\text{OLS}}$ کمتر خواهد شد.

۳.۳ برآوردگر لاسو

تیبشیرانی [۲۴] پیشنهاد کرد در رابطه (۲.۲)، به جای بهینه‌سازی RSS با شرط $\sum_{j=1}^p \beta_j^2 \leq t$ مسئله‌ی بهینه‌سازی زیر حل شود

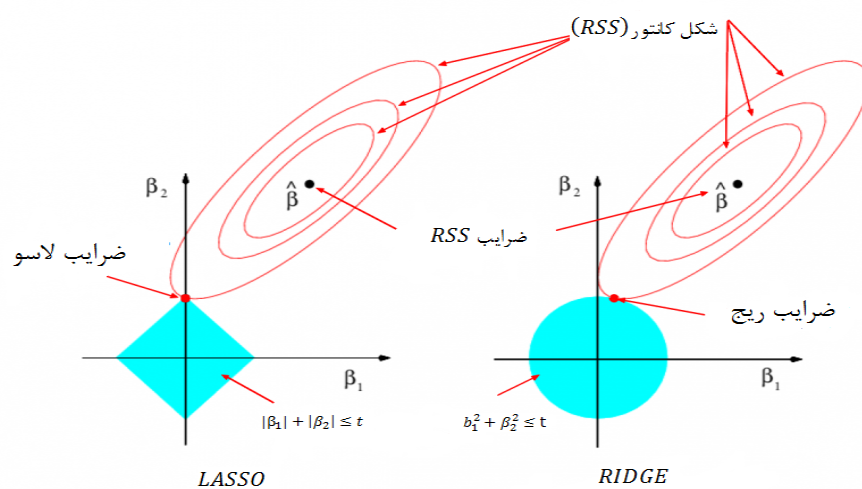
$$\min_{\beta} \left(\text{RSS} \quad \text{s.t} \quad \sum_{j=1}^p |\beta_j| \leq t \right) \quad (۳.۳)$$

$\hat{\beta}$ که از حل رابطه‌ی ۳.۳ به‌دست می‌آید، برآوردگر لاسو^۳ نام دارد. یعنی

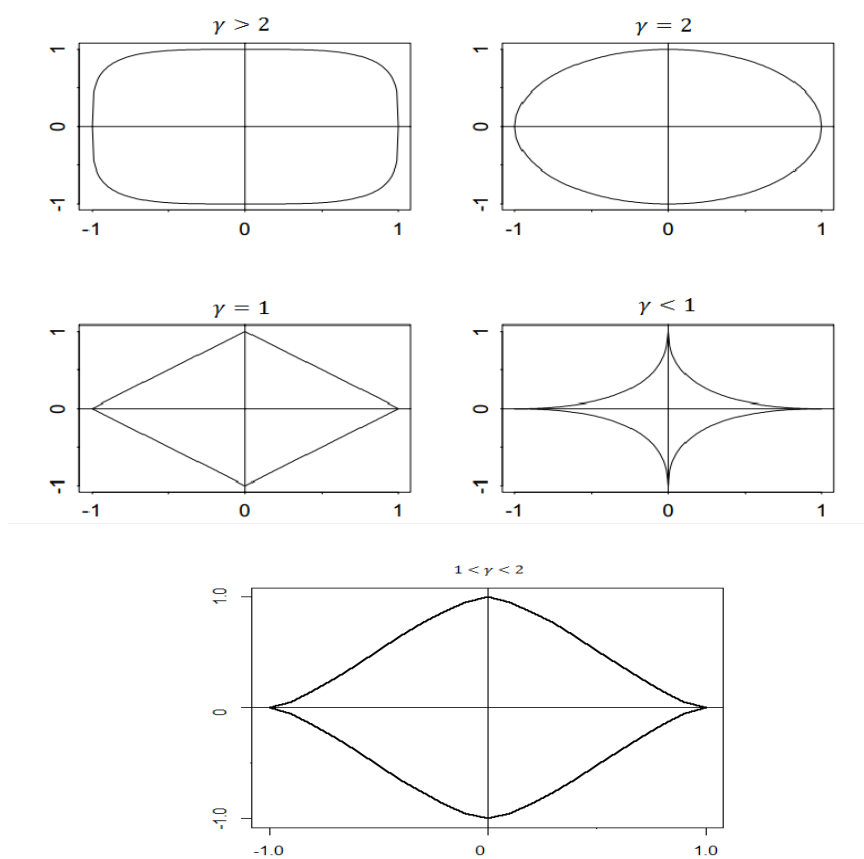
$$\hat{\beta}_{\text{LASSO}} = \underset{\beta}{\text{argmin}} \left(\text{RSS} + k \sum_{j=1}^p |\beta_j| \right)$$

روش لاسو ضرایب $\hat{\beta}$ را به سمت صفر منقبض نموده، در این صورت برخی از $\hat{\beta}_j$ ها مقدار صفر را انتخاب می‌کنند، بنابراین لاسو عملیات انتخاب متغیر را نیز انجام می‌دهد. برخلاف روش ریج، لاسو صورت بسته‌ای ندارد و برای هر مقدار k بردار ضرایب به‌صورت عددی به‌دست می‌آیند.

^۳LASSO (Least absolute shrinkage and selection operator)



شکل ۱.۳: عملکرد رگرسیون بریج در حالت $t = 1, p = 2$.



شکل ۲.۳: نواحی انقباض مختلف در رگرسیون بریج.

شکل ۱.۳ ناحیه تاوان ($|\beta_1|^\gamma + |\beta_2|^\gamma \leq 1$) و مجموع توان دوم مانده‌ها با مرکزیت برآورد کمترین توان‌های دوم معمولی را نشان می‌دهد. این شکل به وضوح ناتوانی روش بریج در حالت $\gamma \geq 2$ در برآورد کردن ضرایب صفر را نشان داده، که احتمال برخورد دو ناحیه در نقطه‌ای که در آن یکی از ضرایب صفر باشد، بسیار اندک است. اما زمانی که $\gamma \leq 1$ ، برآورد ضرایب به گونه‌ای انجام می‌پذیرد که بعضی از ضرایب، مربوط به متغیرهای بی‌اثر، دقیقاً صفر برآورد می‌شوند و به این ترتیب متغیرهای مربوط به آن ضرایب از مدل خارج خواهند شد. بنابراین در این روش، برآوردیابی و انتخاب متغیر توأم صورت می‌پذیرد، که این ویژگی استفاده از روش انقباضی بریج را برای $\gamma \leq 1$ (از جمله لاسو به ازای $\gamma = 1$) در مسائل ابعاد بالا بسیار محبوب ساخته است. در شکل ۲.۳، نواحی جریمه مختلف به ازای γ ‌های متفاوت در برآوردگر بریج نشان داده شده‌اند.

برای محاسبه‌ی برآوردگر لاسو طبق قضیه‌ی ۱.۲.۲ می‌توان حد برآوردگر بریج هنگامی که $\gamma \rightarrow 1^+$ را یافت. برای سادگی از یک الگوریتم جدید که ساده، سرراست، سرعت بالا و روشی پیش‌برنده است ارائه می‌شود.

مسئله در دو حالت $p = 1$ و $p > 1$ شرح داده می‌شود:

(۱) برای حالت $p = 1$ معادله‌ی (۸.۲) به صورت زیر است:

$$2\mathbf{x}^T\mathbf{x}\beta - 2\mathbf{x}^T\mathbf{y} + k \operatorname{Sgn}(\beta) = 0$$

با استفاده از (۱۰.۲) داریم

$$2\mathbf{x}^T\mathbf{x}\beta + S_0 = -k \operatorname{Sgn}(\beta) \quad (۴.۳)$$

جواب مسئله، نقطه‌ی تقاطع دو تابع $f_1(\beta) = 2\mathbf{x}^T\mathbf{x}\beta + S_0$ و $f_2(\beta) = -k \operatorname{Sgn}(\beta)$ است. نحوه‌ی برخورد دو معادله $f_1(\beta)$ و $f_2(\beta)$ را به ازای $\gamma = 1$ می‌توان در شکل ۱.۲ مشاهده کرد. با توجه به شکل ۲.۳ اگر $S_0 > k$ (شکل بالا چپ)، $\beta < 0$ است و داریم:

$$2\mathbf{x}^T\mathbf{x}\beta + S_0 = +k \Rightarrow \hat{\beta} = \frac{k - S_0}{2\mathbf{x}^T\mathbf{x}}$$

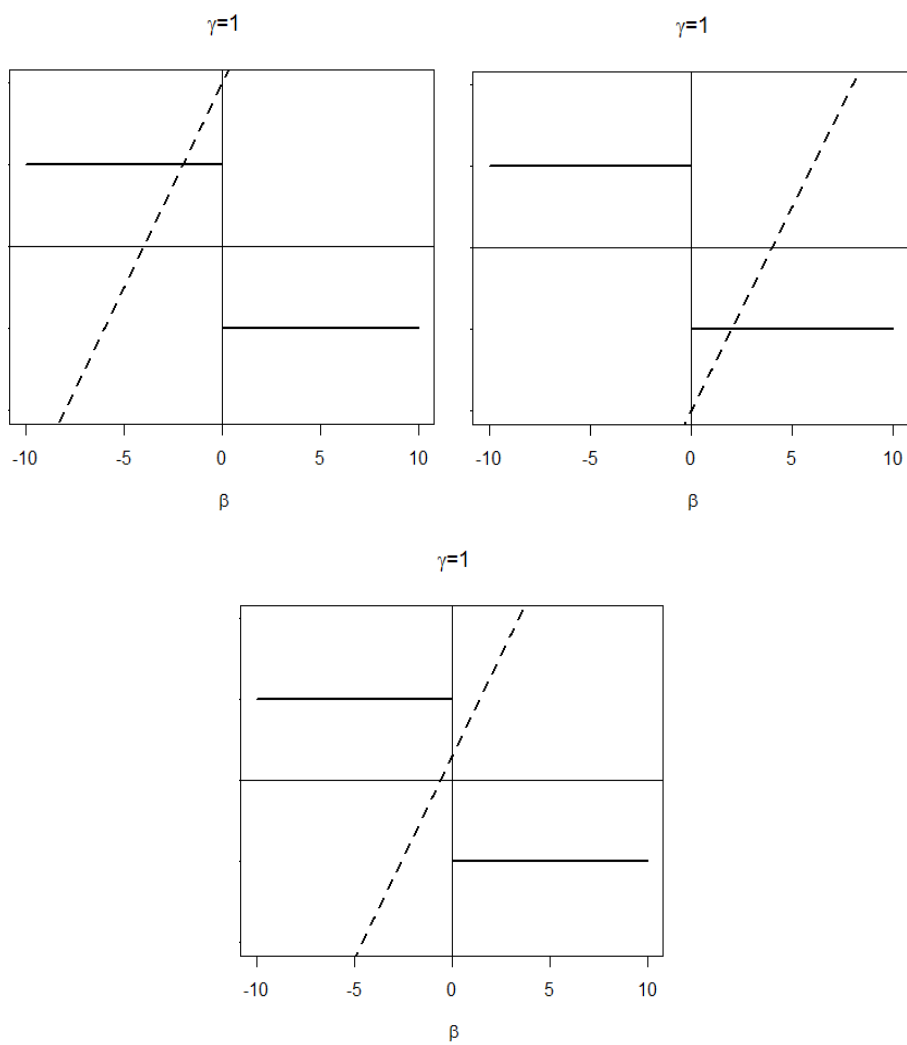
به‌طور مشابه اگر $S_0 < -k$ (شکل بالا راست) پس $\beta > 0$ ، در نتیجه:

$$2\mathbf{x}^T\mathbf{x}\beta + S_0 = -k \Rightarrow \hat{\beta} = \frac{-k - S_0}{2\mathbf{x}^T\mathbf{x}}$$

و اگر $|S_0| \leq k$ باشد $\hat{\beta}$ مقدار صفر را اختیار می‌کند.

(۲) برای حالت $p > 1$ ، ابتدا با $\hat{\beta}_{OLS} = \hat{\beta}_0$ شروع می‌کنیم. در گام m ام $\hat{\beta}_m$ را با به روز کردن $\hat{\beta}_j$ برای $\hat{\beta}^{-j}$ به ازای $j = 1, \dots, p$ ، مانند قبل ($p = 1$) محاسبه می‌کنیم. مراحل را تا زمانی که $\hat{\beta}_m$ همگرا شود تکرار خواهیم کرد.

این الگوریتم به‌طور خلاصه در زیر مطرح شده است.



شکل ۳.۳: نحوه‌ی برخورد تابع $f_2(\beta)$ (خط پر) با تابع $f_1(\beta)$ (خط چین) در معادله‌ی (۴.۲).

الگوریتم ۲. الگوریتم پیش‌برنده برای لاسو

(۱) با $\hat{\beta}_0 = \hat{\beta}_{OLS} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$ شروع می‌کنیم.(۲) در گام m ، برای هر $j = 1, \dots, p$ فرض می‌کنیم $S_j = S_j(\cdot, \hat{\beta}^{-j}, \mathbf{X}, \mathbf{Y})$ و قرار می‌دهیم:

$$\hat{\beta}_j = \begin{cases} \frac{k - S_j}{\sqrt{\mathbf{x}_j^T \mathbf{x}_j}} & \text{if } S_j > k \\ \frac{-k - S_j}{\sqrt{\mathbf{x}_j^T \mathbf{x}_j}} & \text{if } S_j < -k \\ 0 & \text{if } |S_j| \leq k \end{cases}$$

که در آن $\mathbf{x}_j$ ستون j ام ماتریس $\mathbf{X}$ می‌باشد. برآوردگر جدید $\hat{\beta}_m = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T$ بعد از به‌روز کردن همه‌ی $\hat{\beta}_j$ ها بدست می‌آید.

(۳) مرحله‌ی (۲) را تا زمانی که $\hat{\beta}_m$ همگرا شود، تکرار می‌کنیم.

۱.۳.۳ واریانس برآوردگر لاسو

در بخش ۲.۲ واریانس برآوردگر بریج را زمانی که $\gamma > 1$ بود، محاسبه کردیم اما از آنجایی که در برآوردگر لاسو برخی $\hat{\beta}_j$ ها برابر صفر هستند، قضیه ۲.۲.۲ را نمی‌توان برای محاسبه واریانس به‌کار برد. برای حل این مشکل می‌توان با استفاده از روش‌های خودگردان سازی و جک‌نایف واریانس را محاسبه نمود. همچنین تیبشیرانی [۳۴] تقریبی از واریانس را به‌صورت زیر ارائه داد

$$\text{Var}(\hat{\beta}_{\text{Lasso}}) \simeq \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X} + k\mathbf{W}^-)^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X} + k\mathbf{W}^-)^{-1}$$

که $\hat{\sigma}^2$ برآوردی از خطای واریانس و $\mathbf{W}^-$ معکوس تعمیم یافته ماتریس $\mathbf{W} = \text{diag}(|\tilde{\beta}_j|)$ است.

۴.۳ محاسبات با R

در نرم افزار R دو بسته برای محاسبه برآوردگرهای ریج و لاسو وجود دارند، که در ادامه به معرفی و نحوه‌ی استفاده از آن‌ها می‌پردازیم.

۱.۴.۳ بسته glmnet

بسته glmnet برای مدل‌های خطی تعمیم یافته، با قید جریمه

$$(1 - \alpha) \sum_{j=1}^p |\beta_j|^2 + \alpha \sum_{j=1}^p |\beta_j|, \quad 0 \leq \alpha \leq 1$$

به کار می‌رود (فریدمن، هستی و تیبشیرانی [۱۵]). این بسته را می‌توان با نوشتن دستور زیر در R دانلود و نصب کرد.

```
>install.packages("glmnet")
```

این بسته را می‌توان به صورت زیر فراخوانی کرد

```
> library(glmnet)
```

دستور

```
> glmnet.upgrade(testing=TRUE)
```

برای به روزرسانی این بسته به آخرین نسخه به کار می‌رود. این بسته را می‌توان به صورت زیر اجرا کرد

```
> glmnet(x, y, family="...", alpha, standardize=TRUE, intercept=TRUE, ...)
```

که در آن x ماتریس طرح، y متغیر پاسخ، و آرگومان `family` نوع توزیع متغیر پاسخ را مشخص می‌کند

```
family=c("gaussian", "binomial", "poisson", "multinomial", "cox")
```

همچنین α مقداری بین $[0, 1]$ می‌پذیرد. (توجه نمایید $\alpha = 1$ برآوردگر لاسو و $\alpha = 0$ برآوردگر ریج می‌باشد)، آرگومان `standardize` برای استاندارد کردن ماتریس طرح به کار می‌رود که دو مقدار `TRUE` و `FALSE` را می‌پذیرد (پیش فرض آن `TRUE` است) و همچنین آرگومان `intercept` برای مشخص کردن عرض از مبدا مدل به کار می‌رود. با دستور زیر می‌توان آرگومان‌های دیگر این تابع را مشاهده نمود.

```
> ?glmnet
```

در ادامه در قالب یک مثال این بسته را شرح می‌دهیم.

مثال ۱.۴.۳. از داده‌های سرطان پروستات که استیمی و همکاران (۱۹۸۹) ارائه داده‌اند، استفاده می‌کنیم. این داده‌ها وابستگی بین آنتیژن نوع خاص پروستات و اندازه‌ی کلینیکی مردانی که به پروستاتکومی مراجعه کرده‌اند را نشان می‌دهد. جدول ۱.۳ توصیفی از این متغیرها و همچنین خلاصه‌ی آمار توصیفی متغیر پاسخ (*lpsa*) در جدول ۲.۳ آمده است. داده‌های پروستات در بسته *lars* موجود می‌باشند.

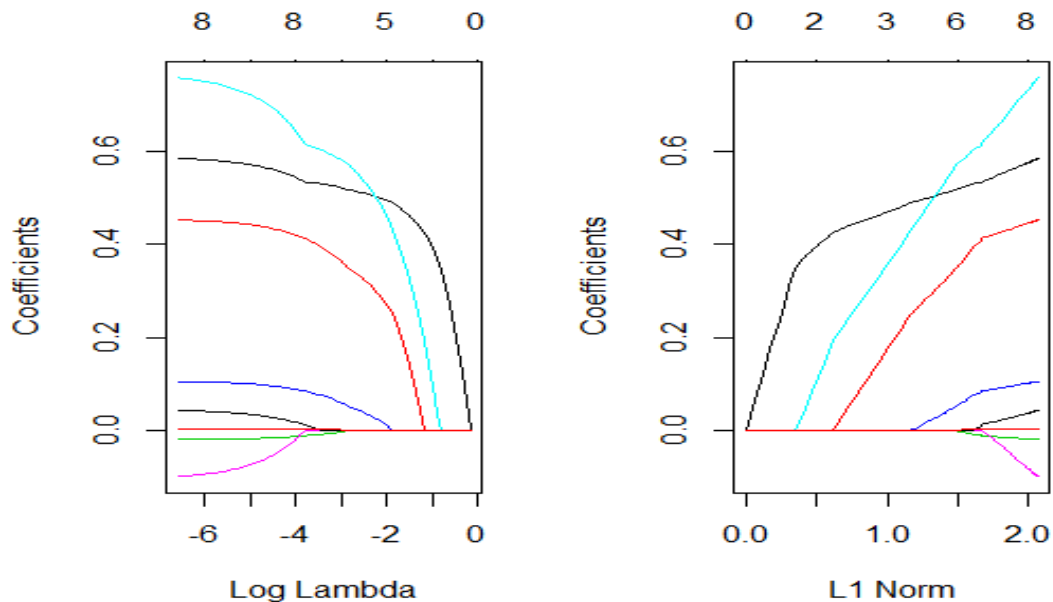
جدول ۱.۳: توصیف متغیرهای مجموعه داده پروستات

متغیرها	توصیف	توضیحات
lpsa	لگاریتم آنتی ژن پروستات	متغیر پاسخ
lcavol	لگاریتم حجم سرطان	
lweight	لگاریتم وزن سرطان	
age	سن	
Age in years. lbph	لگاریتم مقدار هیپرپلازی پروستات خوش خیم	
svi	حمله کیسه منی	
lcp	لگاریتم نفوذ کپسول	
gleason	نمره گلیسون	بردار عددی
pgg45	درصد گلیسون نمره ۴ یا ۵	

جدول ۲.۳: آماره‌های توصیفی متغیر پاسخ در مجموعه داده پروستات

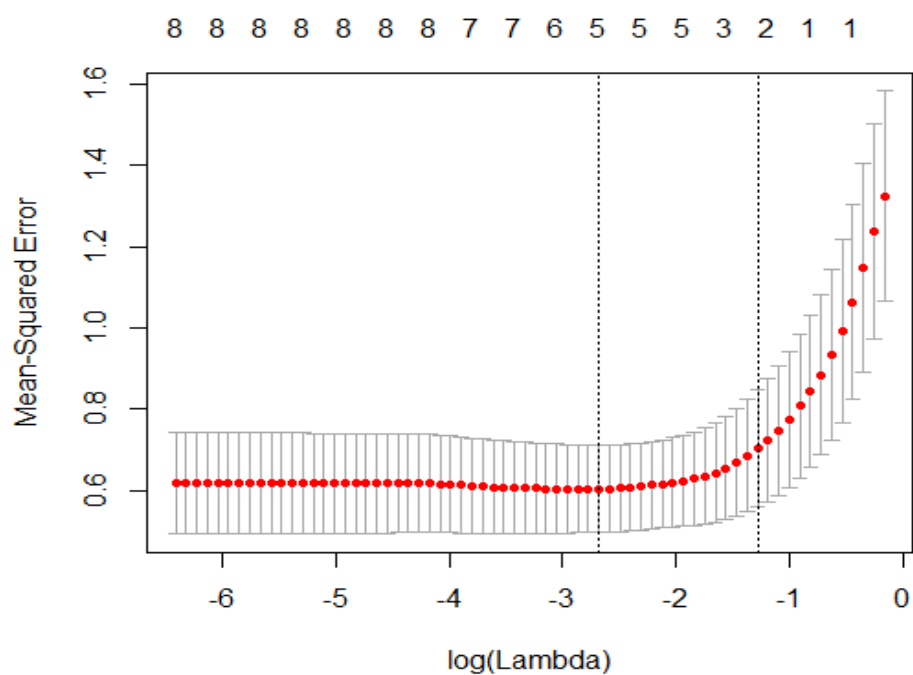
مینیمم	چارک اول	میانه	چارک سوم	ماکزیمم	میانگین	انحراف استاندارد
-۰.۴۳۰۸	۱.۷۳۲۰	۲.۵۹۲۰	۳.۰۵۶۰	۵.۵۸۳۰	۲.۴۷۸۰	۱.۱۵۴۳

برای مشاهده دستورات لازم در R به پیوست ۲.۰ مراجعه کنید.



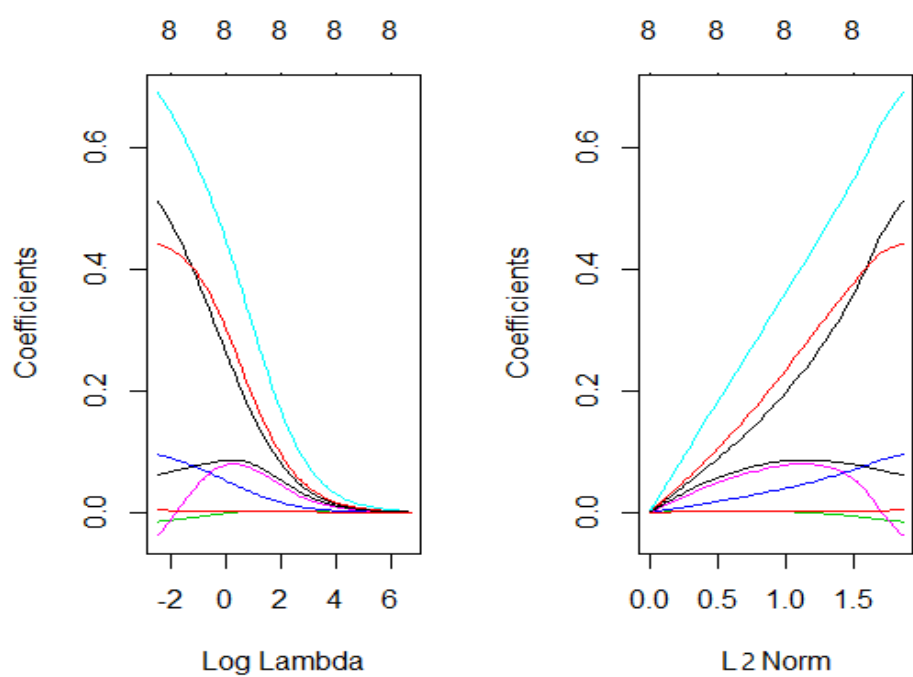
شکل ۴.۳: نمودار ضرایب لاسو در مقابل $s = \frac{t}{\sum_j |\beta_j^{OLS}|}$ و لگاریتم λ

نمودار ۴.۳ نشان می‌دهد که با افزایش s تعداد پارامترهای غیر صفر بیش‌تر می‌شود.

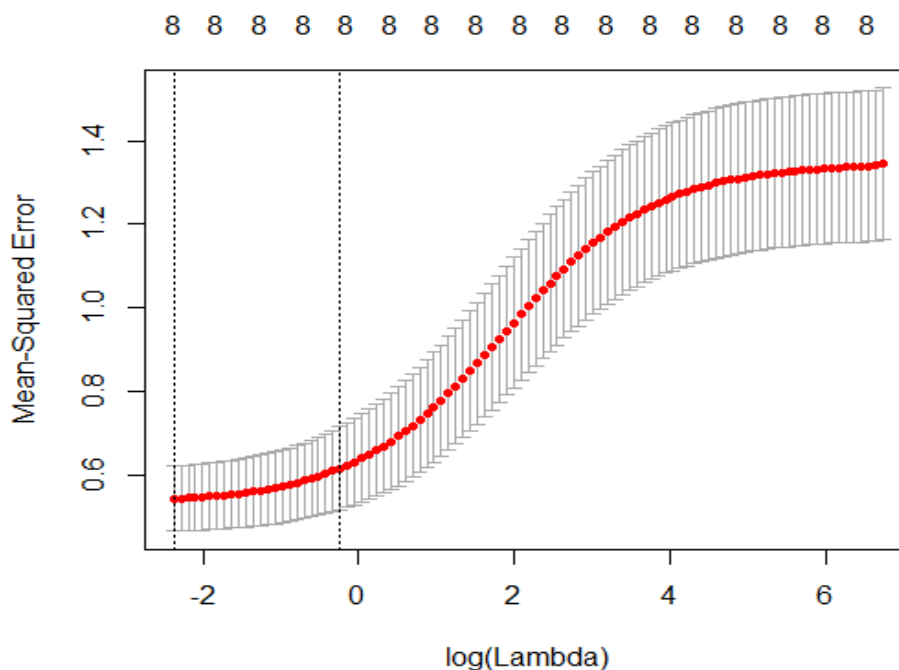


شکل ۵.۳: نمودار MSE در مقابل لگاریتم λ

شکل ۵.۳ نشان می‌دهد نمودار MSE در مقابل λ صعودی است.



شکل ۶.۳: نمودار ضرایب ريج در مقابل s و لگاریتم λ



شکل ۷.۳: نمودار MSE در مقابل لگاریتم λ

جدول ۳.۳: ضرایب داده‌های پروستات در برآوردگر ریج

متغیر	$\lambda_{\text{lse}} = 0.5$	$\lambda_{\text{min}} = 0.9$
Intercept	۰.۲۹۸۱۵۲۳۵۳	۰.۴۶۵۳۳۰۶۳۶
lcavol	۰.۳۴۷۶۶۴۸۴۱	۰.۵۰۵۷۹۰۱۹۰
lweight	۰.۳۷۰۸۰۴۵۴۳	۰.۴۴۱۰۰۹۴۴۳
age	-۰.۰۰۵۶۳۹۱۹	-۰.۰۱۴۸۵۹۸۸۱
lbph	۰.۰۶۷۶۱۸۱۱۴	-۰.۰۹۴۲۱۷۹۹۸
svi	۰.۵۳۷۰۳۸۴۱۷	۰.۶۸۴۹۷۱۸۱۱
lcp	۰.۰۵۸۸۲۶۲۲۲	-۰.۰۳۳۴۶۶۸۳۰
gleason	۰.۰۸۰۱۰۱۴۷۳	۰.۰۳۳۴۶۶۸۳۰
pgg45	۰.۰۰۲۶۵۷۳۸۰	۰.۰۰۳۳۹۶۴۶۹

جدول ۴.۳: ضرایب داده‌های پروستات در برآوردگر لاسو

متغیر	$\lambda_{\text{lse}} = 0.2$	$\lambda_{\text{min}} = 0.3$
Intercept	۱.۰۷۸۵۹۸۰	۰.۶۳۴۴۴۶۱
lcavol	۰.۴۷۲۰۲۴۰	۰.۵۲۹۳۴۶۶۰
lweight	۰.۱۸۶۹۵۰۵	۰.۳۹۲۱۰۲۴
age	۰	-۰.۰۰۷۸۲۲۸۷۱
lbph	۰	-۰.۰۷۵۵۰۴۴۲
svi	۰.۳۶۸۰۴۷۶	۰.۰۶۰۰۷۰۳۱
lcp	۰	۰
gleason	۰	۰.۰۰۰۴۱۲
pgg45	۰	۰.۰۰۲۴۱۱۵۶۸

۲.۴.۳ بسته lasso2

بسته lasso2 برای محاسبات برآوردگر لاسو به کار می‌رود (اوسبرن، پرسنل و تورلاچ [۳۲]). این بسته را می‌توان با نوشتن دستور زیر در R دانلود و نصب کرد.

```
>install.packages("lasso2")
```

و به طور مشابه با دستور زیر می‌توان این بسته را فراخونی و به روز رسانی نمود

```
> library(lasso2)
> lasso2.upgrade(testing=TRUE)
```

این بسته را می‌توان به صورت زیر اجرا کرد

```
> l1ce(formula, data, standardize = TRUE, bound = 0.5,...)
```

که در آن formula، متغیر پاسخ را با عملگر ~ به متغیرهای پیش‌بین ربط می‌دهد (برای مثال $y \sim x_1 + x_2$). آرگومان data یک ساختار داده است که شامل متغیر پاسخ و متغیرهای پیش‌بین است. و آرگومان bound عددی یا برداری است که محدودیت روی قید L_1 ($\sum_{j=1}^n |\beta_j| \leq t$) را مشخص می‌نماید.

مثال ۲.۴.۳. مثال ۱.۴.۳ را با بسته‌ی lasso2 اجرا می‌کنیم. برای مشاهده دستورات لازم به پیوست ج ۲ رجوع کنید.

جدول ۵.۳ ضرایب برآوردگر لاسو را نشان می‌دهد.

جدول ۵.۳: ضرایب داده‌های پروستات در برآوردگر لاسو	
متغیر	bound = ۰/۴۴
Intercept	۱.۰۴۳۵۸۰۳
lcavol	۰.۴۷۴۰۸۳۱
lweight	۰.۱۹۵۳۱۵۶
age	۰
lph	۰
svi	۰.۳۷۵۸۱۹۹
lcp	۰
gleason	۰
pgg45	۰

جدول ۶.۳: نتایج ضرایب لاسو			
	الگوریتم پیش برنده	glmnet	lasso2
lcavol	۰.۴۹۳۹۶۱۶۱	۰.۴۵۵۷۴۲۵	۰.۴۷۴۰۸۳۱
lweight	۰.۱۰۰۱۰۷۳۸	۰.۱۲۱۱۸۵۱	۰.۱۹۵۳۱۵۶
age	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰
lbph	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰
svi	۰.۱۵۹۳۶۱۶۵	۰.۳۰۷۰۸۲۹	۰.۳۷۵۸۱۹۹
lcp	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰
gleason	۰.۰۰۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰
pgg45	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰
MPE	۰.۴۰۴۱۹۲۳	۰.۳۸۹۹۵۸	۰.۳۸۳۷۶۵۲

جدول ۶.۳ ضرایب برآوردگر لاسو، با سه روش الگوریتم پیش برنده و بسته‌های glmnet و lasso2 را نشان می‌دهد، همان طور که می‌بینید ضرایب بسیار نزدیک می‌باشند.

فصل ۴

برآوردگر رگرسیونی بریج محدود شده

معمولاً در مسائل کاربردی و مدل‌های اقتصادی لازم است از اطلاعات اولیه در خصوص بعضی از پارامترهای مورد علاقه در مدل آماری تحت بررسی استفاده شود. برآوردگری که براساس اطلاعات اولیه به دست می‌آید برآوردگر محدود شده نامیده می‌شود. به همین جهت در این فصل، مدل رگرسیون بریج را تحت محدودیت خطی بررسی می‌کنیم. همچنین در این فصل از ایده‌ی برآوردگر ریج محدود شده که گراب [۱۹] آن را معرفی نمود استفاده کرده‌ایم و از پایان‌نامه‌های روزبه [۷] و نجاریان [۱۰] نیز استفاده شده است.

۱.۴ مقدمه

در مسائل کاربردی و مدل‌های اقتصادی، استفاده از اطلاعات اولیه در خصوص بعضی از پارامترهای مورد علاقه در مدل آماری تحت بررسی، اغلب منجر به استفاده از روش‌های پیشرفته و تعمیم یافته‌ای می‌شود که با توجه به قطعی یا غیرقطعی بودن این اطلاع، به کار برده می‌شوند. استفاده از اطلاعات اولیه معلوم در مدل‌های آماری، معمولاً به صورت یک قید مطرح است که در نهایت به مدل‌های خطی محدود شده^۱ (RLM) یا مقید منتج می‌شود. برآوردگری که براین اساس به دست می‌آید، برآوردگر محدود شده نامیده می‌شود.

همچنین در تحقیقات آماری، ممکن است علاوه بر مدل رگرسیونی، اطلاعات اولیه تجربی نیز دستخوش خطا شوند. بنابراین لازم است که برای محدودیت اعمال شده روی مدل نیز یک جمله خطای

^۱ Restricted Linear Model

تصادفی در نظر گرفته شود. لازم به ذکر است که، اعتبار و کارایی برآوردگرهای محدود شده براساس قید اعمال شده بر روی فضای پارامتر مورد سنجش و ارزیابی قرار می‌گیرند، در حالی که برآودگرهای محدود نشده، در کل فضای پارامتر بررسی می‌شوند. بنابراین نتایج به‌دست آمده برپایه مدل‌های محدود شده و محدود نشده تنها با سنجش درستی و اعتبار محدودیت‌ها و قیود اعمال شده، می‌تواند معقول و کاربردی باشد. باید توجه کرد که انتخاب برآوردگرهای محدود شده در شرایطی که اطلاعات اولیه درست نباشد، توجیه‌پذیر نیست.

در این فصل، پارامترهای مدل رگرسیونی محدود شده با محدودیت ثابت و محدودیت تصادفی را برآورد و ویژگی‌های آن‌ها را مورد بررسی قرار می‌دهیم.

۲.۴ برآوردگر بریج با محدودیت ثابت

در مدل (۴.۱) محدودیت خطی $R\beta = r$ را برای β در نظر بگیرید، که در آن ماتریس معلوم محدودیت، با $Rank(R) = q < p$ و $r_{(q \times 1)}$ بردار q -بعدی معلوم است.

به عنوان مثالی برای درک بهتر ماتریس محدودیت R ، فرض کنید براساس اطلاعات اولیه از نمونه‌ای تصادفی، دریافتیم که روابط زیر بین بردار پارامترها برقرار است. لازم به ذکر است که از ساختار زیر به عنوان ضرایب مقابله در طرح آزمایش‌ها استفاده می‌شود.

$$\begin{cases} \beta_1 + \beta_2 + \beta_3 = 1 \\ \beta_3 - 2\beta_2 = 3 \\ 2\beta_1 + 3\beta_2 = 0 \end{cases}$$

در این صورت

$$R = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -2 & 1 \\ 2 & 3 & 0 \end{bmatrix} \quad r = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$$

به منظور یافتن برآوردگر کمترین توان‌های دوم β تحت مفروضات $R\beta = r$ ، از روشی که توسط رائو در سال ۱۹۹۵ پیشنهاد داده شد، استفاده می‌کنیم. این روش که تلفیقی از روش نظریه کمترین توان‌های دوم و لاگرانژ است، تابع

$$F(\beta, \lambda) = (Y - X\beta)^T(Y - X\beta) + 2\lambda^T(R\beta - r)$$

را به ازای $\lambda = (\lambda_1, \dots, \lambda_q)^T$ ، که تابعی خطی از مجموع توان‌های دوم خطا و محدودیت اعمال شده بر روی مدل است، کمینه می‌سازد.

با مشتق‌گیری از $F(\beta, \lambda)$ نسبت به β و برابر صفر قرار دادن آن داریم

$$-2X^TY + 2X^TX\beta + 2R^T\lambda = 0$$

در نتیجه داریم

$$\begin{aligned}\hat{\beta}_{OLS}^R &= (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y} - \mathbf{R}^T \boldsymbol{\lambda}) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \boldsymbol{\lambda} \\ &= \hat{\beta}_{OLS} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \boldsymbol{\lambda}\end{aligned}\quad (۱.۴)$$

از آنجایی که $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ باید $\mathbf{R}\hat{\beta}_{OLS}^R = \mathbf{r}$ در این صورت

$$\mathbf{R}\hat{\beta}_{OLS} - \mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \boldsymbol{\lambda} = \mathbf{r} \quad (۲.۴)$$

بنابراین $\mathbf{R}\hat{\beta}_{OLS} = \mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \hat{\boldsymbol{\lambda}} + \mathbf{r}$ و داریم

$$\hat{\boldsymbol{\lambda}} = \left(\mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right)^{-1} (\mathbf{R}\hat{\beta}_{OLS} - \mathbf{r})$$

بنابراین برای RLM برآوردگر کمترین توان‌های دوم محدود شده^۲ (RLS) به صورت

$$\hat{\beta}_{OLS}^R = \hat{\beta}_{OLS} + \mathbf{S}^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}^{-1} \mathbf{R}^T)^{-1} (\mathbf{r} - \mathbf{R}\hat{\beta}_{OLS}) \quad (۳.۴)$$

به دست می‌آید، که در آن $\mathbf{S} = \mathbf{X}^T \mathbf{X}$ و $\hat{\beta}_{OLS} = \mathbf{S}^{-1} \mathbf{X}^T \mathbf{Y}$. حال اگر بخواهیم برآوردگر RLM را در مدل (۵.۲) بدست آوریم کافی است تابع زیر را نسبت به $\boldsymbol{\beta}$ و $\boldsymbol{\lambda}$ کمینه کنیم

$$F(\boldsymbol{\beta}, k, \boldsymbol{\lambda}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + k \sum_{j=1}^p |\beta_j|^\gamma + \boldsymbol{\lambda}^T (\mathbf{R}\boldsymbol{\beta} - \mathbf{r}) \quad (۴.۴)$$

در قضیه زیر برآوردگر بریج محدود شده که حاصل کمینه کردن تابع (۴.۴) می‌باشد، ارائه شده است.

قضیه ۱.۲.۴. مدل خطی (۵.۲) را در نظر بگیرید. به ازای $\gamma > ۱$ برآوردگر انقباضی بریج محدود شده به صورت زیر است:

$$\hat{\beta}_{Bridge}^R = \hat{\beta}_{Bridge} + \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} (\mathbf{r} - \mathbf{R}\hat{\beta}_{Bridge}) \quad (۵.۴)$$

که در آن $\mathbf{S}_k = \mathbf{X}^T \mathbf{X} + k D(\hat{\boldsymbol{\beta}})|_{\gamma}$ و $\hat{\beta}_{Bridge} = \mathbf{S}_k^{-1} \mathbf{X}^T \mathbf{Y}$ همچنین $D(\hat{\boldsymbol{\beta}}) = \frac{\gamma}{\gamma-1} \text{diag}(|\hat{\beta}_j|^{\gamma-1})$ است.

برهان. با مشتق گرفتن از (۴.۴) نسبت به $\boldsymbol{\beta}$ و مساوی با صفر قرار دادن این مشتق داریم:

$$\begin{aligned}\frac{\partial F}{\partial \boldsymbol{\beta}} &= -\mathbf{X}^T \mathbf{Y} + \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + k \gamma \text{diag}(|\beta_j|^{\gamma-1}) \text{Sgn}(\boldsymbol{\beta}) + \mathbf{R}^T \boldsymbol{\lambda} = 0 \\ &= -\mathbf{X}^T \mathbf{Y} + \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + k \gamma \text{diag}(|\beta_j|^{\gamma-1}) \text{diag}(|\beta_j|^{-1}) \text{diag}(|\beta_j|) \text{Sgn}(\boldsymbol{\beta}) \\ &\quad + \mathbf{R}^T \boldsymbol{\lambda} = 0 \\ &= -\mathbf{X}^T \mathbf{Y} + \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + k \gamma \text{diag}(|\beta_j|^{\gamma-1}) \text{diag}(|\beta_j|^{-1}) \boldsymbol{\beta} + \mathbf{R}^T \boldsymbol{\lambda} = 0 \\ &= -\mathbf{X}^T \mathbf{Y} + \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + k \gamma \text{diag}(|\beta_j|^{\gamma-2}) \boldsymbol{\beta} + \mathbf{R}^T \boldsymbol{\lambda} = 0 \\ &= -\mathbf{X}^T \mathbf{Y} + \left(\mathbf{X}^T \mathbf{X} + k \frac{\gamma}{\gamma-1} \text{diag}(|\beta_j|^{\gamma-2}) \right) \boldsymbol{\beta} + \mathbf{R}^T \boldsymbol{\lambda} = 0\end{aligned}$$

^۲ Restricted least square

در این صورت

$$\begin{aligned}\hat{\beta}_{Bridge}^R &= \left(\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y^0} \right)^{-1} (\mathbf{X}^T \mathbf{Y} - \mathbf{R}^T \boldsymbol{\lambda}) \\ &= \hat{\beta}_{Bridge} - \left(\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^T \boldsymbol{\lambda}\end{aligned}\quad (۶.۴)$$

حال از آنجایی که $\mathbf{R}\hat{\beta} = \mathbf{r}$ پس $\mathbf{R}\beta = \mathbf{r}$ بنابراین

$$\mathbf{R} \left(\hat{\beta}_{Bridge} - \left(\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^T \boldsymbol{\lambda} \right) = \mathbf{r}$$

در نتیجه

$$\hat{\boldsymbol{\lambda}} = \left(\mathbf{R} \left(\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^T \right)^{-1} \left(\mathbf{R} \hat{\beta}_{Bridge} - \mathbf{r} \right) \quad (۷.۴)$$

با جایگذاری (۷.۴) در (۶.۴) داریم

$$\begin{aligned}\hat{\beta}_{Bridge}^R &= \hat{\beta}_{Bridge} - \left(\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^T \\ &\quad \times \left(\mathbf{R} \left(\mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^T \right)^{-1} \left(\mathbf{R} \hat{\beta}_{Bridge} - \mathbf{r} \right)\end{aligned}$$

حال اگر قرار دهیم $\mathbf{S}_k = \mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y^0}$ و $\hat{\beta}_{Bridge} = \mathbf{S}_k^{-1} \mathbf{X}^T \mathbf{Y}$ می‌توانیم این برآوردگر را به صورت ساده‌تری بنویسیم:

$$\hat{\beta}_{Bridge}^R = \hat{\beta}_{Bridge} + \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} (\mathbf{r} - \mathbf{R} \hat{\beta}_{Bridge})$$

□

۱.۲.۴ خطای برآوردگر انقباضی بریج محدود شده

برای هر برآوردگر β^* ، میانگین توان دوم خطا (MSE) به صورت

$$MSE(\beta, \beta^*) = E(\beta^* - \beta)^T (\beta^* - \beta)$$

به دست می‌آید.

قضیه ۲.۲.۴. تحت مفروضات قضیه ۱.۲.۴ MSE برآوردگر بریج محدود شده برای $\gamma > 1$ به صورت زیر است:

$$MSE(\beta, \hat{\beta}_{Bridge}^R) = \sigma^2 \mathbf{M}_k \mathbf{S} \mathbf{M}_k + k^2 \mathbf{M}_k D^T(\hat{\beta})|_{y^0} \beta^T \beta D(\hat{\beta})|_{y^0} \mathbf{M}_k$$

که در آن $\mathbf{S}_k = \mathbf{X}^T \mathbf{X} + kD(\hat{\beta})|_{y^0}$ و $\mathbf{M}_k = \mathbf{S}_k^{-1} - \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1}$

برهان. ابتدا امید ریاضی برآوردگر بریج را محاسبه می‌کنیم:

$$\begin{aligned} E(\hat{\beta}_{Bridge}) &= E\left(\left(\mathbf{X}^T\mathbf{X} + kD(\hat{\beta})|_{y_*}\right)^{-1} \mathbf{X}^T\mathbf{Y}\right) \\ &= \left(\mathbf{X}^T\mathbf{X} + kD(\hat{\beta})|_{y_*}\right)^{-1} \mathbf{X}^T\mathbf{X}\beta \\ &= \left(\mathbf{X}^T\mathbf{X} + kD(\hat{\beta})|_{y_*}\right)^{-1} \left(\left(\mathbf{X}^T\mathbf{X} + kD(\hat{\beta})|_{y_*}\right) - kD(\hat{\beta})|_{y_*}\right) \beta \\ &= \beta - kD(\hat{\beta})|_{y_*} \left(\mathbf{X}^T\mathbf{X} + kD(\hat{\beta})|_{y_*}\right)^{-1} \beta \\ &= \left(\mathbf{I} - kD(\hat{\beta})|_{y_*} \mathbf{S}_k^{-1}\right) \beta \end{aligned}$$

فرض کنید $\mathbf{M}_k = \mathbf{S}_k^{-1} - \mathbf{S}_k^{-1}\mathbf{R}^T(\mathbf{R}\mathbf{S}_k^{-1}\mathbf{R}^T)^{-1}\mathbf{R}\mathbf{S}_k^{-1}$ در این صورت می‌توان برآوردگر بریج محدود شده (۵.۴) را به صورت زیر باز نویسی کرد:

$$\begin{aligned} \hat{\beta}_{Bridge}^R &= \mathbf{S}_k^{-1}\mathbf{S}_k\hat{\beta}_{Bridge} + \mathbf{S}_k^{-1}\mathbf{R}^T(\mathbf{R}\mathbf{S}_k^{-1}\mathbf{R}^T)^{-1}(\mathbf{r} - \mathbf{R}\mathbf{S}_k^{-1}\mathbf{S}_k\hat{\beta}_{Bridge}) \\ &= \mathbf{M}_k\mathbf{S}_k\hat{\beta}_{Bridge} + \mathbf{S}_k^{-1}\mathbf{R}^T(\mathbf{R}\mathbf{S}_k^{-1}\mathbf{R}^T)^{-1}\mathbf{r} \\ &= \mathbf{M}_k\mathbf{S}_k\hat{\beta}_{Bridge} + \mathbf{S}_k^{-1}\mathbf{R}^T(\mathbf{R}\mathbf{S}_k^{-1}\mathbf{R}^T)^{-1}\mathbf{R}\beta_* \\ &= \mathbf{M}_k\mathbf{S}_k\hat{\beta}_{Bridge} + \mathbf{S}_k^{-1}\mathbf{R}^T(\mathbf{R}\mathbf{S}_k^{-1}\mathbf{R}^T)^{-1}\mathbf{R}\beta_* - \mathbf{S}_k^{-1}\mathbf{S}_k\beta_* + \beta_* \\ &= \mathbf{M}_k\mathbf{S}_k\hat{\beta}_{Bridge} - \mathbf{M}_k\mathbf{S}_k\beta_* + \beta_* \end{aligned}$$

بنابراین داریم:

$$\begin{aligned} E(\hat{\beta}_{Bridge}^R) &= \mathbf{M}_k\mathbf{S}_k\left((\mathbf{I} - kD(\hat{\beta})|_{y_*}\mathbf{S}_k^{-1})\beta - \beta_*\right) + \beta_* \\ &= \mathbf{M}_k\mathbf{S}_k(\beta - \beta_*) + \beta_* - kD(\hat{\beta})|_{y_*}\mathbf{M}_k\beta \end{aligned}$$

با توجه به اینکه $\mathbf{R}\beta = \mathbf{r}$ و $\mathbf{R}\beta_* = \mathbf{r}$ و اینکه $\mathbf{M}_k\mathbf{S}_k(\beta - \beta_*) = \beta - \beta_*$ ، می‌توان نوشت:

$$\begin{aligned} E(\hat{\beta}_{Bridge}^R) &= \beta - kD(\hat{\beta})|_{y_*}\mathbf{M}_k\beta \\ E(\hat{\beta}_{Bridge}^R - \beta) &= -kD(\hat{\beta})|_{y_*}\mathbf{M}_k\beta \end{aligned}$$

از طرفی

$$Cov(\hat{\beta}_{Bridge}^R) = \sigma^2\mathbf{M}_k\mathbf{X}^T\mathbf{X}\mathbf{M}_k = \sigma^2\mathbf{M}_k\mathbf{S}\mathbf{M}_k$$

. لذا نتیجه مورد نظر به صورت زیر حاصل می‌شود:

$$\begin{aligned} MSE(\beta, \hat{\beta}_{Bridge}^R) &= Cov(\hat{\beta}_{Bridge}^R) + \left(E(\hat{\beta}_{Bridge}^R - \beta)\right)\left(E(\hat{\beta}_{Bridge}^R - \beta)^T\right) \\ &= \sigma^2\mathbf{M}_k\mathbf{S}\mathbf{M}_k + k^2\mathbf{M}_kD^T(\hat{\beta})|_{y_*}\beta^T\beta D^T(\hat{\beta})|_{y_*}\mathbf{M}_k \end{aligned}$$

□

قضیه ۳.۲.۴. تحت مفروضات قضیه ۱.۲.۴ ماتریس MSE ، برآوردگر بریج محدود شده به صورت زیر است

$$MSEM \left(\hat{\beta}_{Bridge}^R, \beta \right) = \sigma^2 \sum_{i=1}^p \frac{\lambda_i (\lambda_i + kD(\hat{\beta})|_{y^0} - r_{ii}^*)^2}{(\lambda_i + kD(\hat{\beta})|_{y^0})^4} + \left(kD(\hat{\beta})|_{y^0} \right)^2 \sum_{i=1}^p \left(\frac{a_i (\lambda_i + kD(\hat{\beta})|_{y^0} - r_{ii}^*)}{(\lambda_i + kD(\hat{\beta})|_{y^0})^2} \right)^2$$

برهان. با استفاده از تجزیه طیفی، برای ماتریس نیمه معین مثبت S ماتریس متعامد Γ وجود دارد، به طوری که $S = \Gamma \Lambda \Gamma^T$ و همچنین $S_k = \Gamma (\Lambda + kD(\hat{\beta})|_{y^0})^{-1} \Gamma^T$. اکنون ملاحظه می کنید که

$$\begin{aligned} M_k S M_k &= \left(S_k^{-1} - S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R S_k^{-1} \right) S \left(S_k^{-1} - S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R S_k^{-1} \right) \\ &= (S_k^{-1} S S_k^{-1}) - \left(S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R S_k^{-1} S S_k^{-1} \right) \\ &\quad - \left(S_k^{-1} S S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R S_k^{-1} \right) \\ &\quad + \left(S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R S_k^{-1} S S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R S_k^{-1} \right) \\ &= \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \Lambda \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \\ &\quad - \left[\left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} R^* \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \Lambda \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \right] \\ &\quad - \left[\left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \Lambda \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} R^* \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \right] \\ &\quad + \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} R^* \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \Lambda \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} R^* \\ &\quad \times \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \end{aligned}$$

حال با توجه به این که $S_k = \Gamma (\Lambda + kD(\hat{\beta})|_{y^0}) \Gamma^T$ داریم

$$\begin{aligned} tr (S_k^{-1} S S_k^{-1}) &= tr \left(\Gamma \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \Gamma^T \Gamma \Lambda \Gamma^T \Gamma \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \Gamma^T \right) \\ &= tr \left(\left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-2} \Lambda \right) \\ &= \sum_{i=1}^p \frac{\lambda_i}{\left(\lambda_i + kD(\hat{\beta})|_{y^0} \right)^2} \end{aligned} \tag{۸.۴}$$

همچنین

$$\begin{aligned}
 & \text{tr} \left(\mathbf{S}_k^{-1} \mathbf{S} \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1} \right) \\
 &= \text{tr} \left(\Gamma \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \Gamma \mathbf{\Lambda} \Gamma^T \Gamma \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \mathbf{R}^T \Gamma (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R})^{-1} \mathbf{R} \right. \\
 &\quad \times \Gamma \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \\
 &= \text{tr} \left(\left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{R}^* \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \right) \\
 &= \text{tr} \left(\mathbf{R}^* \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{\Lambda} \right) \\
 &= \sum_{i=1}^p \frac{r_{ii}^* \lambda_i}{\left(\lambda_i + kD(\hat{\beta})|_{y^\circ} \right)^2} \tag{۹.۴}
 \end{aligned}$$

به روش مشابه داریم

$$\begin{aligned}
 & \text{tr} \left(\mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1} \mathbf{S} \mathbf{S}_k^{-1} \right) \\
 &= \text{tr} \left(\Gamma \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R})^{-1} \mathbf{R} \Gamma \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \Gamma \mathbf{\Lambda} \Gamma^T \Gamma \right. \\
 &\quad \times \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \\
 &= \text{tr} \left(\left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{R}^* \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \right) \\
 &= \text{tr} \left(\mathbf{R}^* \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{\Lambda} \right) \\
 &= \sum_{i=1}^p \frac{r_{ii}^* \lambda_i}{\left(\lambda_i + kD(\hat{\beta})|_{y^\circ} \right)^2} \tag{۱۰.۴}
 \end{aligned}$$

از طرفی

$$\begin{aligned}
 & \text{tr} \left(\mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1} \mathbf{S} \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1} \right) \\
 &= \text{tr} \left(\Gamma \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \mathbf{R} (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \Gamma \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \Gamma \mathbf{\Lambda} \Gamma^T \Gamma \right. \\
 &\quad \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \mathbf{R} (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \Gamma \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \Gamma^T \\
 &= \text{tr} \left(\left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{R}^* \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{R}^* \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \right) \\
 &= \text{tr} \left(\left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^\circ} \right)^{-1} \mathbf{R}^* \mathbf{\Lambda} \mathbf{R}^* \right) \\
 &= \sum_{i=1}^p \frac{(r_{ii}^*)^2 \lambda_i}{\left(\lambda_i + kD(\hat{\beta})|_{y^\circ} \right)^2} \tag{۱۱.۴}
 \end{aligned}$$

که در آن

$$\mathbf{R}^* = \mathbf{\Gamma}^T \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{\Gamma}$$

و

$$Diag(\mathbf{R}^*) = (r_{11}^*, r_{22}^*, \dots, r_{pp}^*)$$

با جمع کردن (۹.۴) و (۱۰.۴) و (۱۱.۴) می توان نتیجه گرفت

$$\begin{aligned} tr(\mathbf{M}_k \mathbf{S} \mathbf{M}_k) &= \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + kD(\hat{\beta})|_{y^0})^2} - \sum_{i=1}^p \frac{r_{ii}^* \lambda_i}{(\lambda_i + kD(\hat{\beta})|_{y^0})^3} + \sum_{i=1}^p \frac{(r_{ii}^*)^2 \lambda_i}{(\lambda_i + kD(\hat{\beta})|_{y^0})^4} \\ &= \sum_{i=1}^p \frac{\lambda_i (\lambda_i + kD(\hat{\beta})|_{y^0} - r_{ii}^*)^2}{(\lambda_i + kD(\hat{\beta})|_{y^0})^4} \end{aligned} \quad (12.4)$$

در مرحله بعد دقت کنید

$$\begin{aligned} \mathbf{M}_k^* &= (\mathbf{S}_k^{-1} - \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1}) (\mathbf{S}_k^{-1} - \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1}) \\ &= \mathbf{S}_k^{-2} - \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1} - \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1} \\ &\quad + \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1} \end{aligned} \quad (13.4)$$

برای محاسبه $tr(\mathbf{M}_k^*)$ ، لازم است اثر تک تک عبارات (۱۳.۴) محاسبه شود. برای این منظور داریم

$$\begin{aligned} tr(\mathbf{S}_k^{-2}) &= tr \left(\mathbf{\Gamma} \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{\Gamma}^T \mathbf{\Gamma} \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{\Gamma}^T \right) \\ &= \sum_{i=1}^p \frac{1}{(\lambda_i + kD(\hat{\beta})|_{y^0})^2} \end{aligned} \quad (14.4)$$

و

$$\begin{aligned} tr(\mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1}) &= tr \left(\mathbf{\Gamma} \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^* \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{\Gamma}^T \right) \\ &= \sum_{i=1}^p \frac{r_{ii}^*}{(\lambda_i + kD(\hat{\beta})|_{y^0})^3} \end{aligned} \quad (15.4)$$

و

$$\begin{aligned} tr(\mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1}) &= tr \left(\mathbf{\Gamma} \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^* \left(\mathbf{\Lambda} + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{\Gamma}^T \right) \\ &= \sum_{i=1}^p \frac{r_{ii}^*}{(\lambda_i + kD(\hat{\beta})|_{y^0})^3} \end{aligned} \quad (16.4)$$

و همچنین

$$\begin{aligned}
& \text{tr} (\mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R})^{-1} \mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T (\mathbf{R} \mathbf{S}_k^{-1} \mathbf{R}^T)^{-1} \mathbf{R} \mathbf{S}_k^{-1}) \\
&= \text{tr} \left(\Gamma \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^* \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \mathbf{R}^* \left(\Lambda + kD(\hat{\beta})|_{y^0} \right)^{-1} \Gamma^T \right) \\
&= \sum_{i=1}^p \frac{(r_{ii}^*)^2}{(\lambda_i + kD(\hat{\beta})|_{y^0})^4} \quad (17.4)
\end{aligned}$$

با جمع کردن روابط (۱۴.۴) تا (۱۷.۴) می‌توان نتیجه گرفت

$$\begin{aligned}
\text{tr} (\mathbf{M}_k^2) &= \sum_{i=1}^p \left(\frac{1}{(\lambda_i + kD(\hat{\beta})|_{y^0})^2} - \frac{2r_{ii}^*}{(\lambda_i + kD(\hat{\beta})|_{y^0})^3} + \frac{(r_{ii}^*)^2}{(\lambda_i + kD(\hat{\beta})|_{y^0})^4} \right) \\
&= \sum_{i=1}^p \frac{(\lambda_i + kD(\hat{\beta})|_{y^0} - r_{ii}^*)^2}{(\lambda_i + kD(\hat{\beta})|_{y^0})^4} \quad (18.4)
\end{aligned}$$

نتیجه مورد نظر از تلفیق (۱۲.۴) و (۱۸.۴) رابطه زیر بدست می‌آید

$$tr(\alpha^T \mathbf{M}_k \alpha) = \sum_{i=1}^p \left(\frac{\alpha_i^2 (\lambda_i + kD(\hat{\beta})|_{y^0} - r_{ii}^*)}{(\lambda_i + kD(\hat{\beta})|_{y^0})^2} \right) \quad (19.4)$$

که در آن $\alpha = \Gamma^T \beta = (\alpha_1, \alpha_2, \dots, \alpha_p)$ بنابراین

$$\begin{aligned}
MSEM(\hat{\beta}_R^{Ridge}, \beta) &= \sigma^2 tr(\mathbf{M}_k \mathbf{S} \mathbf{M}_k) + \left(kD(\hat{\beta})|_{y^0} \right)^2 tr(\mathbf{M}_k \beta^T \beta \mathbf{M}_k) \\
&= \sigma^2 \sum_{i=1}^p \frac{\lambda_i (\lambda_i + kD(\hat{\beta})|_{y^0} - r_{ii}^*)^2}{(\lambda_i + kD(\hat{\beta})|_{y^0})^4} \\
&\quad + \left(kD(\hat{\beta})|_{y^0} \right)^2 \sum_{i=1}^p \left(\frac{a_i (\lambda_i + kD(\hat{\beta})|_{y^0} - r_{ii}^*)}{(\lambda_i + kD(\hat{\beta})|_{y^0})^2} \right)^2 \quad (20.4)
\end{aligned}$$

□

اثبات کامل است.

۳.۴ برآوردگر بریج با محدودیت تصادفی

در تحقیقات آماری، ممکن است علاوه بر مدل رگرسیونی، اطلاعات اولیه تجربی نیز دستخوش خطا شوند. بنابراین لازم می‌آید که برای محدودیت اعمال شده روی مدل نیز یک جمله خطای تصادفی در نظر گرفته شود و به طور کلی ما اطلاعات اولیه دقیقی مانند $\mathbf{R}\beta = \mathbf{r}$ در روابط اقتصادی، ساختارهای صنعتی، برنامه‌ریزی و ... نداریم. بنابراین می‌توانیم رویدادهای تصادفی نامعلوم را بر محدودیت‌های خطی اعمال کنیم. برای اطلاعات بیشتر در خصوص دلیل کارکردن با محدودیت‌های تصادفی و کاربردهای آن، به تیل [۳۶] مراجعه شود.

۱.۳.۴ برآوردگر بریج محدود شده

مدل خطی (۴.۱) را در نظر بگیرید. فرض کنید اطلاعات اولیه به صورت محدودیت تصادفی $\mathbf{r} = \mathbf{R}\beta + \nu$ است. که در آن ماتریس معلوم محدودیت، با $\text{Rank}(\mathbf{R}) = q < p$ و $\mathbf{r}_{(q \times 1)}$ بردار q -بعدی معلوم است و ν جمله خطای تصادفی است. فرض کنید $\epsilon \sim N_n(0, \sigma^2 \mathbf{I})$ و $\nu \sim N_q(\mu, \sigma^2 \mathbf{I})$ که μ بردار q -بعدی از مقادیر معلوم و σ^2 مجهول است. به علاوه فرض کنید ϵ و ν مستقل باشند. تیل و گولدرگر^۴ [۳۵] پیشنهاد دادند که برای به دست آوردن رابطه‌ای خطی، اطلاعات اولیه تصادفی و مدل خطی به صورت زیر ترکیب شوند.

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{r} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{R} \end{pmatrix} \beta + \begin{pmatrix} \epsilon \\ \nu \end{pmatrix} \quad (۲۱.۴)$$

در این حالت محدودیت مورد نظر عبارت است از

$$H_0 = \begin{bmatrix} -\mathbf{R} & \mathbf{I}_q \end{bmatrix} \times \begin{bmatrix} \beta \\ \mathbf{R}\beta + \mu \end{bmatrix} = \mu = 0 \quad (۲۲.۴)$$

مدل (۲۱.۴) را به صورت

$$\mathbf{Y}^* = \mathbf{Z}\Phi + \eta \quad (۲۳.۴)$$

بازنویسی می‌کنیم. که در آن

$$\mathbf{Y}^* = \begin{bmatrix} \mathbf{Y} \\ \mathbf{r} \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{X} & 0_{n \times q} \\ 0_{q \times p} & \mathbf{I}_q \end{bmatrix}, \quad \Phi = \begin{bmatrix} \beta \\ \mathbf{R}\beta \end{bmatrix}, \quad \eta = \begin{bmatrix} \epsilon \\ \nu \end{bmatrix} \quad (۲۴.۴)$$

که $\mathbf{Y}^*$ برداری $(n+q) \times 1$ بعدی از مشاهدات، $\mathbf{Z}$ ماتریس معلوم $(n+q) \times (p+q)$ بعدی با رتبه $p+q$ ، Φ برداری $(p+q) \times 1$ بعدی از پارامترهای مورد بررسی، η بردار خطای تصادفی $(n+q) \times 1$ است. بنابراین محدودیت روی مدل جدید به صورت زیر به دست می‌آید.

$$H_0 = \mathbf{H}\Phi = \mu = 0 \quad (۲۵.۴)$$

که در آن $\mathbf{H}$ ماتریس محدودیت $q \times (p+q)$ بعدی با رتبه q است. برای به دست آوردن ضرایب Φ باید تابع $F(\Phi, \lambda) = (\mathbf{Y}^* - \mathbf{Z}\Phi)^T (\mathbf{Y}^* - \mathbf{Z}\Phi) + \lambda^T \mathbf{H}\Phi$ را نسبت به Φ کمینه کرد. با مشتق‌گیری از $F(\Phi, \lambda)$ نسبت به Φ و برابر صفر قرار دادن عبارت حاصل می‌توان نتیجه گرفت

$$\begin{aligned} \hat{\Phi}_{OLS}^H &= (\mathbf{Z}^T \mathbf{Z})^{-1} (\mathbf{Z}^T \mathbf{Y}^* - \mathbf{H}^T \lambda) \\ &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y}^* - (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{H}^T \lambda \\ &= \hat{\Phi}_{OLS} - (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{H}^T \lambda \end{aligned} \quad (۲۶.۴)$$

^۴Goldberger

که در آن $\hat{\Phi}_{OLS} = (Z^T Z)^{-1} Z^T Y^*$ برآوردگر محدود نشده بردار Φ است. از آنجایی که

$$H \hat{\Phi}_{OLS}^H = H \hat{\Phi}_{OLS} - H (Z^T Z)^{-1} H^T \lambda = 0$$

بنابراین

$$H \hat{\Phi}_{OLS} = H (Z^T Z)^{-1} H^T \lambda$$

$$\hat{\lambda} = \left(H (Z^T Z)^{-1} H^T \right)^{-1} H \hat{\Phi}_{OLS} \quad (27.4)$$

با جایگذاری معادله (27.4) در (26.4) نتیجه می شود

$$\hat{\Phi}_{OLS}^H = \hat{\Phi}_{OLS} - (Z^T Z)^{-1} H^T \left(H (Z^T Z)^{-1} H^T \right)^{-1} H \hat{\Phi}_{OLS} \quad (28.4)$$

$\hat{\Phi}_{OLS}^H$ را بهترین برآوردگر نااریب خطی Φ می نامیم.

در صورت وجود همخطی، معادله (28.4) را اصلاح کرده و برآوردگر بریج محدود شده را به صورت زیر پیشنهاد می کنیم.

قضیه ۱.۳.۴. تحت مفروضات مدل (23.4)، برای $\gamma > 1$ برآوردگر انقباضی بریج محدود شده به صورت زیر است

$$\begin{aligned} \hat{\Phi}_{Bridge}^H &= \hat{\Phi}_{Bridge} - \left(Z^T Z + kD(\hat{\Phi})|_{y^\circ} \right)^{-1} H^T \left(H \left(Z^T Z + kD(\hat{\Phi})|_{y^\circ} \right)^{-1} H^T \right)^{-1} \\ &\times H \hat{\Phi}_{Bridge} \end{aligned} \quad (29.4)$$

که در آن $D(\hat{\Phi}) = \frac{2}{\gamma} diag(|\Phi_j|^{\gamma-2})$.

برهان. با گرفتن مشتق از تابع

$$F(\Phi, \lambda) = (Y^* - Z\Phi)^T (Y^* - Z\Phi) + 2\lambda^T H\Phi + k \sum_{j=1}^p |\Phi_j|^\gamma$$

نسبت به Φ و برابر صفر قرار دادن، عبارت حاصل نتیجه می شود، داریم

$$\begin{aligned} \hat{\Phi}_{Bridge}^H &= \left(Z^T Z + kD(\hat{\Phi})|_{y^\circ} \right)^{-1} (Z^T Y^* - H^T \lambda) \\ &= \left(Z^T Z + kD(\hat{\Phi})|_{y^\circ} \right)^{-1} Z^T Y^* - \left(Z^T Z + kD(\hat{\Phi})|_{y^\circ} \right)^{-1} H^T \lambda \\ &= \hat{\Phi}_{Bridge} - \left(Z^T Z + kD(\hat{\Phi})|_{y^\circ} \right)^{-1} H^T \lambda \end{aligned} \quad (30.4)$$

که در آن $\hat{\Phi}_{Bridge} = \left(Z^T Z + kD(\hat{\Phi})|_{y^\circ} \right)^{-1} Z^T Y^*$ برآوردگر محدود نشده بریج است.

همچنین

$$\mathbf{H}\hat{\Phi}_{Bridge}^H = \mathbf{H}\hat{\Phi}_{Bridge} - \mathbf{H}\left(\mathbf{Z}^T\mathbf{Z} + k\mathbf{D}(\hat{\Phi})|_{y^0}\right)^{-1}\mathbf{H}^T\boldsymbol{\lambda} = 0$$

بنابراین

$$\mathbf{H}\hat{\Phi}_{Bridge} = \mathbf{H}\left(\mathbf{Z}^T\mathbf{Z} + k\mathbf{D}(\hat{\Phi})|_{y^0}\right)^{-1}\mathbf{H}^T\boldsymbol{\lambda}$$

داریم

$$\hat{\boldsymbol{\lambda}} = \left(\mathbf{H}\left(\mathbf{Z}^T\mathbf{Z} + k\mathbf{D}(\hat{\Phi})|_{y^0}\right)^{-1}\mathbf{H}^T\right)^{-1}\mathbf{H}\hat{\Phi}_{Bridge} \quad (31.4)$$

با جایگذاری معادله (31.4) در (30.4) به دست می آوریم

$$\begin{aligned} \hat{\Phi}_{Bridge}^H &= \hat{\Phi}_{Bridge} - \left(\mathbf{Z}^T\mathbf{Z} + k\mathbf{D}(\hat{\Phi})|_{y^0}\right)^{-1}\mathbf{H}^T\left(\mathbf{H}\left(\mathbf{Z}^T\mathbf{Z} + k\mathbf{D}(\hat{\Phi})|_{y^0}\right)^{-1}\mathbf{H}^T\right)^{-1} \\ &\quad \times \mathbf{H}\hat{\Phi}_{Bridge} \end{aligned} \quad (32.4)$$

□

اثبات کامل است.

۲.۳.۴ خطای برآوردگر انقباضی بریج محدود شده

قضیه ۲.۳.۴. تحت مفروضات این بخش، MSE برآوردگر بریج برای $\gamma > 1$ به صورت زیر است:

$$MSE\left(\Phi, \hat{\Phi}_{Bridge}^H\right) = \sigma^2 M_k S M_k + k^2 M_k D^T(\hat{\Phi})|_{y^0} \Phi^T \Phi D(\hat{\Phi})|_{y^0} M_k$$

که در آن $S_k = \mathbf{X}^T \mathbf{X} + kD(\hat{\Phi})|_{y^0}$ و $M_k = S_k^{-1} - S_k^{-1} R^T (R S_k^{-1} R^T)^{-1} R S_k^{-1}$ اثبات همانند قضیه ۲.۲.۴ می باشد.

□

قضیه ۳.۳.۴. ماتریس میانگین توان های دوم خطای، برآوردگر بریج محدود شده به صورت زیر است

$$\begin{aligned} MSEM\left(\Phi, \hat{\Phi}_{Bridge}^H\right) &= \sigma^2 \sum_{i=1}^p \frac{\lambda_i(\lambda_i + kD(\hat{\Phi})|_{y^0} - r_{ii}^*)^2}{(\lambda_i + kD(\hat{\Phi})|_{y^0})^4} \\ &\quad + \left(kD(\hat{\Phi})|_{y^0}\right)^2 \sum_{i=1}^p \left(\frac{a_i(\lambda_i + kD(\hat{\Phi})|_{y^0} - r_{ii}^*)}{(\lambda_i + kD(\hat{\Phi})|_{y^0})^2}\right)^2 \end{aligned}$$

□

برهان. اثبات همانند قضیه ۲.۲.۴ است.

فصل ۵

مطالعات عددی

در این بخش برای بررسی عملکرد برآوردگر بریج و مقایسه رفتار آن با برآوردگرهای کمترین توان‌های دوم، روش شبیه‌سازی مونت کارلو را به کار می‌گیریم و دو مثال واقعی را بررسی کرده‌ایم. برای گردآوری مطالب این فصل عمدتاً از مقاله‌های مک‌دونالد [۲۶]، گیوبنز [۱۸]، نیوهوسن و اومن [۳۱]، فو [۱۶]، تیشیرانی [۳۴] و هورل و کنارد [۲۰] استفاده شده است.

۱.۵ شبیه‌سازی

۱.۱.۵ مقدمه

در این فصل با استفاده از یک فرایند شبیه‌سازی مونت کارلو به ارزیابی عملکرد و مقایسه روش کمترین توان‌های دوم و بریج در مدل رگرسیونی چندگانه در دو حالت بدون محدودیت و با محدودیت خواهیم پرداخت. از آنجایی که روش‌های انقباضی برای مقابله با همخطی چندگانه است لذا مدل‌های شبیه‌سازی شده را بیشتر در این طرح مورد مطالعه قرار می‌دهیم. برای داشتن داده‌هایی با میزان همخطی متفاوت، مک‌دونالد و گلارنیو [۲۶] و گیوبنز [۱۸] پیشنهاد دادند که داده‌ها براساس رابطه

$$x_{ij} = (1 - \nu^2)^{1/2} z_{ij} + \nu z_{ip} \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p$$

تولید شود که در آن z_{ij} اعداد شبه‌تصادفی مستقل از توزیع نرمال استاندارد و ν همبستگی بین هر دو متغیر توضیحی است. در ادامه در قالب چند سناریو، ابتدا ویژگی‌های داده‌های شبیه‌سازی شده را بیان و سپس نتایج کاربرد روش‌ها را ارائه می‌دهیم.

سناریو ۱.۱.۵. نیوهوسن و اومن [۳۱] بیان کردند در صورتی که میانگین توان دوم خطا، تابعی از β ، σ^2 و k ، و متغیرهای توضیحی ثابت باشند، اگر β بردار ویژه نرمال شده متناظر با بزرگترین مقدار ویژه ماتریس $X'X$ تحت قید $\beta'\beta = 1$ باشد، آنگاه MSE کمینه می‌شود. بنابراین از بردار ویژه متناظر با بزرگترین مقدار ویژه ماتریس $X'X$ برای برآورد ضرایب $\beta_1, \beta_2, \dots, \beta_p$ هنگامی که $n = 5, 20, 50, 100$ و $\nu = 0/7, 0/9, 0/99$ و $p = 4, 20$ ، $\sigma = 0/5$ می‌باشد در شبیه‌سازی مقادیر متغیر پاسخ استفاده می‌شود. در تمامی مثال‌های ارائه شده مقدار k براساس معیار GCV بدست آمده است. منظور از نتایج روش پیشنهادی، نتایج حاصل از روش نیوتن رافسون اصلاح شده و الگوریتم پیش‌برنده لاسو است و منظور از نتایج $glmnet$ ، نتایج حاصل از بسته $glmnet$ است.

جدول ۱.۵: نتایج سناریو ۱.۱.۵ ($p = 4, n = 5, \nu = 0/7$) با نقطه شروع OLS

β	نتایج روش پیشنهادی						نتایج $glmnet$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
-۰.۵۷۲۹	-۰.۲۱۹۵	-۰.۶۱۰۴	-۰.۱۹۸۱	-۰.۰۱۰۵	-۰.۰۹۸۳	-۰.۲۰۰۷	-۰.۱۲۳	-۰.۴۵۷۵
-۰.۳۷۰۴	-۰.۳۳۲۸	-۰.۴۶۷۵	-۰.۳۲۸۵	-۰.۰۱۵۱	-۰.۱۳۸۲	-۰.۳۳۲۰	-۰.۶۲۶۹	-۰.۶۰۷۸
-۰.۴۳۶۸	۰.۳۸۴۵	۰.۰۰۰۰	۰.۳۸۵۰	۰.۰۱۷۰	۰.۱۵۳۴	۰.۳۸۸۶	۰.۰۰۰۰	-۰.۰۴۰۰
-۰.۵۸۶۲	-۱.۶۷۶۱	-۱.۴۲۸۴	-۰.۷۰۸۰	۰.۰۰۰۰	-۰.۳۴۲۸	-۱.۵۵۱۲	-۱.۱۷۷۱	-۰.۸۲۰۷

جدول ۲.۵: نتایج سناریو ۱.۱.۵ ($p = 4, n = 5, \nu = 0/9$) با نقطه شروع OLS

β	نتایج روش پیشنهادی						نتایج $glmnet$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
-۰.۵۴۱۲	۰.۳۷۸۸	۰.۰۰۰۰	۰.۰۵۵۴	۰.۰۰۳۹	۰.۰۵۰۰	۰.۰۵۶۴	۰.۰۰۰۰	-۰.۴۷۶۵
-۰.۴۲۴۲	-۰.۳۶۲۷	-۰.۲۵۹۵	-۰.۳۵۰۴	-۰.۰۱۲۵	-۰.۲۱۷۹	-۰.۳۵۵	-۱.۰۲۲۷	-۰.۸۴۰۴
-۰.۴۶۳۴	۰.۸۸۲۱	-۰.۸۰۱۵	۰.۸۷۶۲	۰.۰۲۴۴	۰.۴۱۷۷	۰.۸۶۸۳	۰.۰۰۰۰	-۰.۰۲۶۴
-۰.۵۵۸۶	-۲.۴۱۵۳	-۲.۳۱۷۶	-۲.۳۶۱۹	۰.۰۰۰۰	-۰.۷۹۸۰	-۲.۹۳۵	-۱.۰۰۲۰	-۰.۶۵۲۹

جدول ۳.۵: نتایج سناریو ۱.۱.۵ ($p = 4, n = 5, \nu = 0/9$) با نقطه شروع Ridge

β	نتایج روش پیشنهادی						نتایج $glmnet$	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
-۰.۵۴۱۲	۰.۰۳۷	-۰.۴۲۷	-۰.۶۰۹۳	-۰.۰۳۸۵	-۰.۰۷۸۳	-۰.۰۰۰۰	۰.۰۰۰۰	-۰.۴۷۶۵
-۰.۴۲۴۴	-۰.۳۶۲۷	-۰.۶۲۲	-۰.۶۱۴۵	-۰.۰۳۷۵	-۰.۰۷۷۰	-۱.۶۴۲	-۱.۰۴۲	-۰.۸۴۰۴
-۰.۴۶۳۴	۰.۸۸۲۱	۰.۰۰۰۰	-۰.۴۲۳۱	-۰.۰۲۹۷	-۰.۰۶۴۳۵	-۱.۱۳۶۵	۰.۰۰۰	-۰.۰۲۶۴
-۰.۵۵۸۶	-۲.۴۱۵۳	۰.۰۰۰۰	-۰.۶۴۵۲	-۰.۰۳۹۷	-۰.۰۸۰۲	-۱.۵۷۱۲	-۱.۰۰۲	-۰.۶۵۲۹

جدول ۴.۵: نتایج سناریو ۱.۱.۵ ($p = ۴, n = ۵۰, \nu = ۰.۹$) (OLS با نقطه شروع)

β	نتایج روش پیشنهادی						glmnet	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
-۰.۴۴۶۴	-۰.۲۹۸۱	-۰.۲۹۷۲	-۰.۲۸۷۴	-۰.۰۴۶۵	-۰.۰۸۴۲	-۰.۲۹۲۹	-۰.۲۸۹۰	-۰.۳۸۵۸
-۰.۴۵۱۹	-۰.۲۶۶۴	-۰.۲۶۵۴	-۰.۲۶۰۵	-۰.۰۴۲۶	-۰.۰۷۸۴	-۰.۲۶۵۷	-۰.۲۵۸۲	-۰.۴۰۷۹
-۰.۴۷۵۱	-۰.۶۹۸۱	-۰.۶۹۷۲	-۰.۶۸۰۵	-۰.۰۸۵۷	-۰.۱۳۸۲	-۰.۶۹۲۳	-۰.۶۹۰۹	-۰.۶۵۲۹
-۰.۶۰۸۸	-۰.۷۲۳۲	-۰.۷۲۲۲	-۰.۷۰۷۲	-۰.۱۱۴۳	-۰.۱۷۰۸	-۰.۷۱۵۱	-۰.۷۲۹۰	۰.۵۳۸۳

جدول ۵.۵: نتایج سناریو ۱.۱.۵ ($p = ۴, n = ۵۰, \nu = ۰.۹$) (Ridge با نقطه شروع)

β	نتایج روش پیشنهادی						glmnet	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
-۰.۴۴۶۴	-۰.۲۹۸۱	-۰.۳۸۷۰	-۰.۴۴۲۹	-۰.۱۰۰۱	-۰.۵۰۳۸	-۰.۸۷۴۵	-۰.۲۸۹۰	-۰.۳۸۵۸
-۰.۴۵۱۹	-۰.۲۶۶۴	-۰.۴۱۷۸	-۰.۴۵۹۱	-۰.۱۰۱۷	-۰.۵۱۷۴	-۰.۹۲۱۵	-۰.۲۵۸۲	-۰.۴۰۷۹
-۰.۴۷۵۱	-۰.۶۹۸۱	-۰.۵۲۵۴	-۰.۵۱۲۹	-۰.۱۱۱۲	-۰.۵۵۷۹	-۱.۰۰۵۷	-۰.۶۹۰۹	-۰.۶۵۲۹
-۰.۶۰۸۸	-۰.۷۲۳۲	-۰.۶۴۴۸	-۰.۵۸۹۲	-۰.۱۴۰۱	-۰.۶۰۹۳	-۰.۹۵۳۷	-۰.۷۲۰۹	-۰.۵۳۸۳

جدول ۶.۵: نتایج سناریو ۱.۱.۵ ($p = ۴, n = ۲۰, \nu = ۰.۹$) (Ridge با نقطه شروع)

β	نتایج روش پیشنهادی						glmnet	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
-۰.۴۴۶۴	-۰.۲۹۸۱	-۰.۳۸۷۰	-۰.۴۴۲۹	-۰.۱۰۰۱	-۰.۵۰۳۸	-۰.۸۷۴۵	-۰.۲۸۹۰	-۰.۳۸۵۸
-۰.۴۵۱۹	-۰.۲۶۶۴	-۰.۴۱۷۸	-۰.۴۵۹۱	-۰.۱۰۱۷	-۰.۵۱۷۴	-۰.۹۲۱۵	-۰.۲۵۸۲	-۰.۴۰۷۹
-۰.۴۷۵۱	-۰.۶۹۸۱	-۰.۵۲۵۴	-۰.۵۱۲۹	-۰.۱۱۱۲	-۰.۵۵۷۹	-۱.۰۰۵۷	-۰.۶۹۰۹	-۰.۶۵۲۹
-۰.۶۰۸۸	-۰.۷۲۳۲	-۰.۶۴۴۸	-۰.۵۸۹۲	-۰.۱۴۰۱	-۰.۶۰۹۳	-۰.۹۵۳۷	-۰.۷۲۰۹	-۰.۵۳۸۳

جدول ۷.۵ مقدار میانگین توان دوم خطا را برای برآوردگرهای کمترین توان‌های دوم، ریدج و لاسو تحت مفروضات ۱.۱.۵ به ازای $m = ۱۰۰۰$ تکرار نشان می‌دهد. این نتایج از رابطه‌ی زیر به‌دست می‌آیند:

$$MSE(\hat{\beta}) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p (\hat{\beta}_{ji} - \beta_j)^2.$$

جدول ۷.۵: نتایج MSE

ν	MSE	$p=4$				$p=20$			
		۱۰۰	۵۰	۲۰	۵	۱۰۰	۵۰	۲۰	۵
۰.۷	OLS	۰.۰۲۳	۰.۰۴۳	۰.۱۵۰	۳۷۱۵	۰.۱۴۹	۰.۴۷۱	NA	NA
	Ridge	۰.۰۲۷	۰.۰۴۰	۰.۰۸۰	۰.۴۷۳	۰.۰۷۲	۰.۱۰۴	۰.۲۸۴	۰.۴۲۱
	Lasso	۰.۰۲۱	۰.۰۴۰	۰.۱۴۴	۰.۸۵۰	۰.۱۵۹	۰.۴۸۳	۱.۲۸۲	۳.۶۱۷
۰.۹	OLS	۰.۰۶۲	۰.۱۲۶	۰.۴۰۷	۵۱۳۵	۰.۴۶۵	۱.۲۹۱	NA	NA
	Ridge	۰.۰۴۵	۰.۰۶۱	۰.۱۰۹	۰.۳۷۱	۰.۰۶۳	۰.۰۹۱۷	۰.۲۸۲	۰.۲۵۴
	Lasso	۰.۰۶۲	۰.۱۲۵	۰.۳۴۲	۱.۳۶۳	۰.۴۷۰	۱.۰۲۵	۱.۹۴۲	۵.۳۲۲
۰.۹۹	OLS	۰.۱۳۳	۰.۲۷۴	۰.۸۷۶	۴۹۵۵	۵.۶۹۱	۱۵.۳۱	NA	NA
	Ridge	۰.۰۴۴۱	۰.۰۶۰	۰.۱۰۹	۰.۳۵۱	۰.۰۴۵	۰.۰۷۲	۰.۱۳۳	۰.۱۴۹
	Lasso	۰.۱۳۳	۰.۲۵۵	۰.۵۶۶	۱.۷۵۱	۱.۶۱۲	۲.۱۲۹	۳.۴۵۱	۷.۴۷۴

این نتایج نشان می‌دهند زمانی که تعداد متغیرها بیش از مشاهدات است یا وقتی داده‌ها دارای هم‌خطی می‌باشند، برآورگر کمترین توان‌های دوم نمی‌تواند به خوبی ضرایب را به‌دست آورد و مقدار میانگین توان دوم خطای آن بسیار زیاد بوده اما برآوردرگر لاسو بهتر از برآوردرگر کمترین توان‌های دوم ضرایب را پیدا کرده و دارای میانگین توان دوم خطای کمتری است. رگرسیون ریج دارای بهترین دقت پیش‌بینی در بین سایر روش‌هاست.

در ادامه، روش شبه‌سازی را مشابه سناریو ۱.۱.۵ در نظر می‌گیریم با این تفاوت که بردار β ساختارهای متفاوتی داشته و در یک حالت آن محدودیت خطی نیز به آن اضافه شده است.

سناریو ۲.۱.۵. بردار β را به صورت $\beta = (\underbrace{0, \dots, 0}_{1^0}, \underbrace{2, \dots, 2}_{1^0}, \underbrace{0, \dots, 0}_{1^0}, \underbrace{2, \dots, 2}_{1^0})^T$ در نظر می‌گیریم. در این مدل گروه‌هایی از متغیرهای موثر و غیرموثر وجود دارند که همبستگی بین آن‌ها متوسط و یکسان است

جدول ۸.۵: نتایج سناریو ۲.۱.۵ ($p = 40, n = 20, \nu = 0.7$) با نقطه شروع (Ridge)								
نتایج روش پیشنهادی					نتایج glmnet			
β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
۰	-۵۰.۴۶۷	۰.۰۰۰	۰.۷۱۰۷	۰.۱۶۴	۰.۲۳۳	۳.۵۳۹	۰.۰۰۰	۰.۵۷۳
۰	-۳۷.۵۸۳	۰.۰۵۳	۰.۷۳۹	۰.۱۷۱	۰.۲۴۰	۳.۳۲۷	۰.۰۰۰	۰.۵۲۱
۰	-۷.۱۵۶	۰.۰۰۰	۰.۶۳۳	۰.۱۵۳	۰.۲۲۱	۳.۰۷۴	۰.۰۰۰	۰.۵۳۴
۰	-۲۲.۷۴۶	۰.۰۱۳	۰.۷۶۴	۰.۱۷۱	۰.۲۴۱	۳.۹۵۰	۰.۰۰۰	۰.۶۹۱
۰	-۲۰.۸۹۳	۰.۰۰۰	۰.۷۲۹	۰.۱۶۹	۰.۲۳۸	۳.۳۶۱	۰.۰۰۰	۰.۶۱۶
۰	۸.۸۳۹	۰.۰۰۰	۰.۵۴۲	۰.۱۳۴	۰.۲۰۱	۳.۱۷۸	۰.۰۰۰	۰.۵۷۰
۰	-۱.۳۲۶	۰.۰۰۰	۰.۷۰۵	۰.۱۶۰	۰.۲۲۹	۴.۰۴۱	۰.۰۰۰	۰.۷۷۳
۰	۷۲.۱۷۲	۰.۰۰۰	۰.۶۳۸	۰.۱۵۱	۰.۲۱۹	۳.۴۸۶	۰.۵۶۳	۰.۶۲۵
۰	۲۸.۱۴۰	۰.۱۰۷	۰.۷۷۵	۰.۱۷۳	۰.۲۴۳	۳.۹۱۱	۰.۰۰۰	۰.۶۹۲
۰	۱۶.۸۱۹	۱.۲۳۶	۰.۹۸۶	۰.۲۱۱	۰.۲۸۱	۳.۹۹۳	۰.۰۰۰	۰.۶۳۵
۲	۳۵.۲۹۸	۰.۰۰۰	۰.۷۳۳	۰.۱۶۵	۰.۲۳۴	۴.۰۳۱	۰.۰۰۰	۰.۶۸۱
۲	۱.۳۴۰	۱.۹۰۴	۱.۰۶۹	۰.۲۱۳	۰.۲۸۵	۵.۲۷۰	۳.۷۶۳	۰.۸۹۹
۲	۳۸.۴۳۶	۰.۰۰۰	۰.۵۶۷	۰.۱۳۷	۰.۲۰۴	۳.۵۶۹	۰.۰۰۰	۰.۶۷۶
۲	۵۷.۷۰۰	۲.۱۸۷	۱.۱۳۳	۰.۲۲۳	۰.۲۹۵	۵.۲۳۱	۳.۵۴۲	۰.۸۵۸
۲	-۱۳.۰۸۷	۰.۸۳۶	۰.۹۰۲	۰.۱۹۰	۰.۲۶۱	۴.۶۰۴	۰.۸۱۴	۰.۸۶۵
۲	-۶۹.۹۸۷	۰.۰۰۰	۰.۷۲۲	۰.۱۶۰	۰.۲۳۰	۴.۴۷۲	۱.۰۰۷	۰.۸۸۹
۲	۲۰.۰۴۸	۱.۰۶۶	۰.۹۴۱	۰.۱۹۷	۰.۲۶۸	۴.۶۰۸	۰.۰۷۲	۰.۷۹۲
۲	-۳۸.۴۰۴	۰.۳۴۰	۰.۸۳۲	۰.۱۷۸	۰.۲۴۹	۴.۶۰۰	۰.۵۱۰	۰.۸۵۸
۲	۱۵.۱۸۶	۰.۰۰۰	۰.۷۰۸	۰.۱۵۹	۰.۲۲۸	۴.۳۸۳	۰.۰۰۰	۰.۸۳۴
۲	NA	۰.۸۶۶	۰.۹۰۶	۰.۱۹۴	۰.۲۶۵	۴.۲۲۶	۰.۰۰۰	۰.۷۰۸
۰	NA	۰.۰۰۰	۰.۷۴۷	۰.۱۷۰	۰.۲۴۰	۳.۶۵۷	۰.۰۰۰	۰.۶۱۶
۰	NA	۰.۰۰۰	۰.۶۴۳	۰.۱۵۳	۰.۲۲۱	۳.۲۷۹	۰.۰۰۰	۰.۶۱۱
۰	NA	۰.۸۶۵	۰.۹۰۵	۰.۱۹۷	۰.۲۶۷	۳.۹۱۳	۰.۰۰۰	۰.۶۴۲
۰	NA	۰.۰۰۰	۰.۵۹۰	۰.۱۴۰	۰.۲۰۷	۳.۸۶۴	۰.۰۰۰	۰.۸۰۹
۰	NA	۰.۰۰۰	۰.۷۱۲	۰.۱۶۴	۰.۲۳۳	۳.۵۳۳	۰.۰۰۰	۰.۶۷۱
۰	NA	۰.۰۰۰	۰.۷۳۶	۰.۱۶۶	۰.۲۳۵	۳.۹۹۳	۰.۰۰۰	۰.۷۴۸
۰	NA	۰.۰۰۰	۰.۶۳۹	۰.۱۴۸	۰.۲۱۷	۳.۹۸۹	۰.۰۰۰	۰.۷۴۰
۰	NA	۰.۰۰۰	۰.۶۳۶	۰.۱۵۲	۰.۲۲۰	۳.۲۵۴	۰.۰۰۰	۰.۵۵۳
۰	NA	۰.۰۰۰	۰.۶۶۱	۰.۱۵۸	۰.۲۲۷	۳.۰۸۴	۰.۰۰۰	۰.۵۱۱
۲	NA	۰.۱۰۵	۰.۷۹۳	۰.۱۷۳	۰.۲۴۴	۴.۳۴۸	۱.۴۹۰	۰.۷۴۸
۲	NA	۰.۷۱۸	۰.۸۸۰	۰.۱۸۹	۰.۲۶۰	۴.۲۳۲	۲.۵۵۳	۰.۷۲۴
۲	NA	۱.۷۷۹	۱.۰۸۴	۰.۲۲۰	۰.۲۹۲	۴.۶۹۱	۱.۲۰۶	۰.۷۴۲
۲	NA	۱.۸۵۵	۱.۰۹۵	۰.۲۲۱	۰.۲۹۳	۴.۷۸۴	۳.۱۲۹	۰.۷۹۳
۲	NA	۰.۰۰۰	۰.۷۲۶	۰.۱۶۷	۰.۲۳۷	۳.۴۸۰	۰.۰۰۰	۰.۵۸۲
۲	NA	۰.۵۴۲	۰.۸۵۶	۰.۱۸۳	۰.۲۵۴	۴.۴۹۰	۰.۵۱۶	۰.۷۵۶
۲	NA	۰.۷۵۹	۰.۸۸۳	۰.۱۹۳	۰.۲۶۴	۳.۸۲۱	۱.۴۸۹	۰.۶۲۰
۲	NA	۰.۳۹۱	۰.۸۲۶	۰.۱۸۰	۰.۲۵۱	۴.۱۷۰	۰.۰۰۰	۰.۶۴۸
۲	NA	۰.۰۰۰	۰.۷۴۴	۰.۱۶۷	۰.۲۳۷	۳.۹۸۱	۰.۰۰۰	۰.۷۰۴
۲	NA	۲.۹۸۵	۱.۶۰۲	۰.۳۱۸	۰.۳۸۴	۴.۴۶۹	۹.۰۶۳	۰.۵۷۳

جدول ۹.۵: نتایج سناریو ۲.۱.۵ $p = ۴۰, n = ۵۰, \nu = ۰.۷$ (Ridge با نقطه شروع)

نتایج روش پیشنهادی				نتایج glmnet				
β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
۰	-۰.۱۱۴	۰.۰۰۰	۰.۳۹۸	۰.۱۲۷	۰.۱۸۷	۱.۰۴۶	۰.۰۰۰	-۰.۱۸۷
۰	-۰.۰۴۳	۰.۰۰۰	۰.۵۷۱	۰.۱۵۷	۰.۲۲۳	۱.۶۷۹	۰.۰۰۰	-۰.۱۳۰
۰	۰.۶۴۳	۰.۰۰۰	۰.۷۷۰	۰.۱۸۳	۰.۲۵۲	۲.۸۳۲	۰.۰۰۰	۰.۶۶۲
۰	۰.۱۹۰	۰.۰۰۰	۰.۶۳۳	۰.۱۶۶	۰.۲۳۳	۱.۹۹۲	۰.۰۰۰	۰.۱۸۲
۰	-۰.۲۶۳	۰.۰۰۰	۰.۶۶۶	۰.۱۷۶	۰.۲۴۳	۱.۹۰۱	۰.۰۰۰	۰.۱۱۲
۰	-۰.۱۵۱	۰.۰۰۰	۰.۵۰۵	۰.۱۴۰	۰.۲۰۴	۱.۷۳۴	۰.۰۰۰	۰.۴۰۶
۰	۰.۲۳۵	۰.۲۸۳	۰.۹۶۵	۰.۲۱۹	۰.۲۸۸	۳.۱۲۳	۰.۱۹۱	۰.۷۶۲
۰	۰.۲۵۴	۰.۰۰۰	۰.۸۳۲	۰.۱۹۸	۰.۲۶۷	۲.۷۰۸	۰.۰۰۰	۰.۷۸۲
۰	-۰.۵۹۱	۰.۰۰۰	۰.۶۸۴	۰.۱۷۶	۰.۲۴۴	۲.۱۱۲	۰.۰۰۰	۰.۶۰۰
۰	۰.۲۶۶	۱.۰۴۴	۱.۱۸۲	۰.۲۵۸	۰.۳۲۷	۳.۴۳۸	۰.۱۹۲	۰.۹۷۱
۲	۱.۷۰۴	۱.۶۶۲	۱.۰۶۵	۰.۲۳۷	۰.۳۰۷	۳.۲۴۷	۱.۶۹۷	۱.۶۲۱
۲	۱.۷۲۷	۱.۲۸۱	۱.۲۶۲	۰.۲۸۶	۰.۳۵۳	۳.۰۷۴	۱.۸۰۷	۱.۵۸۹
۲	۱.۹۸۵	۱.۶۴۲	۱.۰۴۷	۰.۲۳۸	۰.۳۰۷	۳.۰۲۹	۱.۴۹۹	۱.۴۹۵
۲	۱.۷۹۹	۱.۲۱۹	۱.۲۴۰	۰.۲۶۸	۰.۳۳۷	۳.۴۹۲	۱.۸۳۷	۱.۶۷۸
۲	۱.۹۸۴	۱.۸۴۸	۱.۱۰۹	۰.۲۵۱	۰.۳۲۰	۳.۰۶۶	۱.۵۸۷	۱.۷۴۲
۲	۲.۷۵۸	۰.۴۸۶	۱.۰۰۳	۰.۲۲۹	۰.۲۹۹	۲.۹۹۴	۱.۹۴۱	۱.۶۷۷
۲	۲.۳۷۲	۰.۴۸۵	۱.۹۸۹	۰.۳۳۰	۰.۳۰۰	۲.۸۱۳	۱.۴۲۸	۱.۸۰۵
۲	۲.۰۷۴	۱.۰۱۶	۱.۱۷۸	۰.۲۵۳	۰.۳۲۲	۳.۶۱۴	۱.۹۶۷	۲.۲۴۲
۲	۱.۶۱۹	۰.۶۳۲	۱.۰۵۸	۰.۲۳۵	۰.۳۰۵	۳.۲۶۰	۱.۴۹۶	۱.۸۷۳
۲	۱.۸۷۴	۱.۱۱۵	۱.۲۰۵	۰.۲۶۲	۰.۳۳۱	۳.۴۳۱	۱.۹۲۱	۱.۸۷۱
۰	-۰.۶۷۴	۰.۰۰۰	۰.۶۹۷	۰.۱۷۸	۰.۲۴۶	۲.۱۴۸	۰.۰۰۰	۰.۰۶۶
۰	۰.۱۸۱	۰.۰۰۰	۰.۷۲۷	۰.۱۹۴	۰.۲۶۱	۱.۸۵۰	۰.۰۰۰	۰.۰۱۳
۰	-۰.۴۶۲	۰.۰۰۰	۰.۳۸۵	۰.۱۱۶	۰.۱۷۷	۱.۳۳۰	۰.۰۰۰	۰.۱۴۶
۰	-۰.۳۹۸	۰.۰۰۰	۰.۶۷۴	۰.۱۷۹	۰.۲۴۶	۱.۸۹۲	۰.۰۰۰	۰.۴۶۰
۰	-۰.۷۱۷	۰.۰۰۰	۰.۷۹۲	۰.۱۹۹	۰.۲۶۷	۲.۲۳۴	۰.۰۰۰	۰.۳۶۹
۰	۰.۱۰۷	۰.۰۰۰	۰.۴۴۰	۰.۱۳۰	۰.۱۹۳	۱.۳۵۵	۰.۰۰۰	۰.۲۲۶
۰	۰.۵۹۸	۰.۰۰۰	۰.۷۲۲	۰.۱۸۹	۰.۲۵۷۷	۱.۹۵۵	۰.۰۰۰	۰.۴۲۳
۰	۰.۸۴۰	۰.۰۰۰	۰.۴۸۶	۰.۱۴۱	۰.۲۰۴	۱.۴۵۰	۰.۰۰۰	-۰.۰۶۵
۰	۰.۱۶۴	۰.۰۰۰	۰.۷۷۲	۰.۱۹۱	۰.۲۶۰	۲.۳۶۴	۰.۰۰۰	۰.۴۹۳
۰	-۰.۳۳۶	۰.۰۰۰	۰.۵۴۴	۰.۱۵۲	۰.۲۱۸	۱.۵۹۴	۰.۰۰۰	-۰.۱۰۳
۲	۱.۸۲۹	۰.۰۱۳	۰.۸۹۲	۰.۲۰۶	۰.۲۷۶	۲.۹۵۵	۱.۷۱۰	۱.۵۷۴
۲	۰.۹۱۴	۱.۲۲۱	۱.۲۴۲	۰.۲۵۸	۰.۳۲۷	۴.۰۲۹	۲.۱۴۴	۲.۰۲۴
۲	۲.۳۵۷	۱.۸۷۳	۱.۱۱۸	۰.۲۵۲	۰.۳۲۱	۳.۰۷۴	۱.۶۹۷	۱.۶۱۶
۲	۲.۲۹۷	۱.۲۹۲	۱.۲۶۷	۰.۲۸۵	۰.۳۵۲	۳.۱۱۶	۱.۷۹۹	۱.۶۹۸
۲	۲.۴۴۹	۰.۰۰۰	۰.۸۴۴	۰.۱۹۳	۰.۲۶۳	۳.۱۶۷	۱.۳۱۸	۱.۷۲۸
۲	۲.۲۴۲	۱.۴۱۹	۱.۰۳۶	۰.۲۲۳	۰.۲۹۴	۳.۷۲۳	۲.۰۴۵	۱.۶۵۲
۲	۳.۰۶۳	۱.۴۴۷	۱.۰۰۰	۰.۲۲۷	۰.۲۹۶	۳.۰۷۸	۱.۸۸۶	۱.۸۵۹
۲	۱.۵۲۶	۱.۱۷۹	۰.۹۴۰	۰.۲۱۴	۰.۲۸۴	۳.۱۰۴	۰.۹۵۰	۱.۳۶۳
۲	۱.۹۲۶	۱.۱۷۰	۰.۹۷۵	۰.۲۱۴	۰.۲۸۴	۳.۵۶۴	۱.۹۲۳	۲.۲۳۳
۲	۱.۹۲۴	۲.۳۲۰	۱.۷۶۶	۰.۳۹۰	۰.۴۴۷	۳.۵۶۵	۴.۱۸۱	۱.۱۴۳

جدول ۱۰.۵: نتایج سناریو ۲.۱.۵ ($p = 40, n = 20, \nu = 0.9$) با نقطه شروع Ridge

نتایج روش پیشنهادی					نتایج glmnet			
β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
۰	-۵۸.۲۲۱	۰.۰۰۰	۰.۸۳۵	۰.۱۸۳	۰.۲۵۳	۴.۰۱۶	۰.۰۰۰	۰.۷۵۱
۰	-۴۲.۴۸۶	۰.۰۰۰	۰.۸۳۹	۰.۱۸۶	۰.۲۵۶	۳.۸۱۰	۰.۰۰۰	۰.۶۹۳
۰	-۸.۸۵۲	۰.۰۰۰	۰.۸۰۰	۰.۱۷۹	۰.۲۴۹	۳.۷۲۸	۰.۰۰۰	۰.۷۱۲
۰	-۲۲.۹۶۳	۰.۰۰۰	۰.۸۷۶	۰.۱۹۰	۰.۲۶۰	۴.۰۸۹	۰.۰۰۰	۰.۷۶۲
۰	-۲۶.۴۶۴	۰.۰۰۰	۰.۸۶۲	۰.۱۹۰	۰.۲۶۰	۳.۷۹۰	۰.۰۰۰	۰.۷۲۱
۰	۱۱.۰۹۶	۰.۰۰۰	۰.۷۴۷	۰.۱۶۹	۰.۲۳۸	۳.۸۰۲	۰.۰۰۰	۰.۷۵۲
۰	-۱.۳۰۸	۰.۰۰۰	۰.۸۴۴	۰.۱۸۵	۰.۲۵۵	۴.۰۲۷	۰.۰۰۰	۰.۷۷۶
۰	۸۳.۱۶۸	۰.۰۰۰	۰.۷۹۹	۰.۱۷۸	۰.۲۴۸	۳.۸۹۹	۰.۵۲۵	۰.۷۵۹
۰	۳۴.۹۱۱	۰.۰۰۰	۰.۸۸۱	۰.۱۹۱	۰.۲۶۲	۴.۰۵۶	۰.۰۰۰	۰.۷۵۵
۰	۱۹.۱۷۹	۰.۵۹۰	۰.۹۸۶	۰.۲۱۱	۰.۲۸۲	۳.۹۷۴	۰.۰۰۰	۰.۶۹۱
۲	۳۸.۶۷۳	۰.۰۰۰	۰.۸۲۴	۰.۱۸۰	۰.۲۵۰	۴.۱۷۵	۰.۰۰۰	۰.۸۰۱
۲	۲.۸۱۳	۳.۴۱۲	۰.۹۸۱	۰.۲۰۳	۰.۲۷۴	۴.۶۸۶	۳.۹۳۹	۰.۸۶۰
۲	۴۲.۶۵۷	۰.۰۰۰	۰.۷۶۷	۰.۱۷۱	۰.۲۴۰	۴.۰۸۴	۰.۰۰۰	۰.۸۳۵
۲	۶۳.۹۳۴	۰.۷۲۰	۱.۰۲۸	۰.۲۱۲	۰.۲۸۳	۴.۵۷۶	۳.۵۲۹	۰.۸۱۲
۲	-۱۶.۲۰۹	۰.۰۶۸	۰.۹۱۴	۰.۱۹۴	۰.۲۶۵	۴.۳۲۲	۰.۹۲۰	۰.۸۳۸
۲	-۸۰.۴۳۱	۰.۰۰۰	۰.۸۲۸	۰.۱۷۹	۰.۲۵۰	۴.۳۸۶	۰.۹۹۳	۰.۸۷۹
۲	۲۵.۴۰۶	۱.۳۷۶	۰.۹۶۰	۰.۲۰۳	۰.۲۷۴	۴.۲۹۱	۰.۴۹۵	۰.۷۸۰
۲	-۴۶.۴۲۰	۰.۰۰۰	۰.۸۹۰	۰.۱۹۱	۰.۲۶۱	۴.۲۹۱	۰.۳۰۵	۰.۸۲۶
۲	۱۶.۹۰۷	۰.۰۰۰	۰.۸۱۴	۰.۱۷۸	۰.۲۴۸	۴.۲۶۲	۰.۰۰۰	۰.۸۵۱
۲	NA	۰.۲۷۸	۰.۹۳۶	۰.۲۰۰	۰.۲۷۱	۴.۱۳۰	۰.۰۰۰	۰.۷۵۴
۰	NA	۰.۰۰۰	۰.۸۶۷	۰.۱۸۹	۰.۲۶۰	۳.۹۶۲	۰.۰۰۰	۰.۷۳۱
۰	NA	۰.۰۰۰	۰.۸۱۳	۰.۱۸۱	۰.۲۵۱	۳.۷۵۴	۰.۰۰۰	۰.۷۲۳
۰	NA	۰.۳۳۳	۰.۹۳۵	۰.۲۰۲	۰.۲۷۳	۳.۹۱۰	۰.۰۰۰	۰.۷۰۳
۰	NA	۰.۰۰۰	۰.۷۷۷	۰.۱۷۳	۰.۲۴۳	۴.۰۳۷	۰.۰۰۰	۰.۸۲۰
۰	NA	۰.۰۰۰	۰.۸۴۹	۰.۱۸۷	۰.۲۵۸	۳.۸۲۸	۰.۰۰۰	۰.۷۴۲
۰	NA	۰.۰۰۰	۰.۸۵۵	۰.۱۸۷	۰.۲۵۷	۴.۰۰۱	۰.۰۰۰	۰.۷۷۱
۰	NA	۰.۰۰۰	۰.۸۴۳	۰.۱۸۸	۰.۲۵۸	۳.۶۱۴	۰.۰۰۰	۰.۶۷۴
۰	NA	۰.۰۰۰	۰.۷۹۹	۰.۱۷۶	۰.۲۴۶	۴.۱۶۳	۰.۰۰۰	۰.۸۲۶
۰	NA	۰.۰۰۰	۰.۸۰۶	۰.۱۷۹	۰.۲۴۹	۳.۸۴۶	۰.۰۰۰	۰.۷۲۷
۰	NA	۰.۰۰۰	۰.۸۰۸	۰.۱۸۱	۰.۲۵۱	۳.۷۲۸	۰.۰۰۰	۰.۶۹۶
۲	NA	۰.۰۰۰	۰.۸۶۵	۰.۱۸۶	۰.۲۵۷	۴.۲۸۱	۱.۶۳۵	۰.۸۱۳
۲	NA	۰.۱۱۴	۰.۹۱۲	۰.۱۹۶	۰.۲۶۶	۴.۱۶۷	۲.۷۵۲	۰.۷۶۸
۲	NA	۰.۶۴۳	۱.۰۰۵	۰.۲۱۱	۰.۲۸۲	۴.۲۳۵	۰.۴۴۳	۰.۷۴۲
۲	NA	۱.۷۱۴	۱.۰۱۹	۰.۲۱۴	۰.۲۸۵	۴.۲۵۸	۲.۶۴۵	۰.۷۵۶
۲	NA	۰.۰۰۰	۰.۸۵۵	۰.۱۸۸	۰.۲۵۸	۳.۸۸۹	۰.۰۰۰	۰.۷۲۴
۲	NA	۰.۰۰۰	۰.۸۸۲	۰.۱۸۹	۰.۲۶۰	۴.۲۹۵	۰.۰۰۰	۰.۸۰۸
۲	NA	۱.۲۳۹	۰.۹۱۹	۰.۱۹۹	۰.۲۷۰	۳.۹۱۵	۱.۱۶۷	۰.۷۰۳
۲	NA	۰.۰۰۰	۰.۸۶۷	۰.۱۸۷	۰.۲۵۸	۴.۲۵۲	۰.۰۰۰	۰.۷۸۸
۲	NA	۰.۰۰۰	۰.۸۴۰	۰.۱۸۴	۰.۲۵۴	۴.۰۷۶	۰.۰۰۰	۰.۷۸۵
۲	NA	۱.۷۶۵	۱.۳۰۴	۰.۲۷۱	۰.۳۴۰	۳.۹۶۷	۱۳.۱۶۴	۰.۵۹۱

جدول ۱۱.۵: نتایج سناریو ۲.۱.۵ ($p = ۴۰, n = ۵۰, \nu = ۰/۹$) با نقطه شروع (Ridge)								
نتایج روش پیشنهادی					نتایج glmnet			
β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
۰	-۰.۱۸۶	۰.۰۰۰	۰.۶۹۳	۰.۱۸۱	۰.۲۴۹	۱.۹۶۷	۰.۰۰۰	-۰.۰۹۱
۰	-۰.۰۷۰	۰.۰۰۰	۰.۷۷۶	۰.۱۹۳	۰.۲۶۲	۲.۲۹۶	۰.۰۰۰	۰.۱۰۴
۰	۱.۰۵۴	۰.۰۰۰	۰.۸۷۴	۰.۲۰۳	۰.۲۷۲	۲.۹۴۰	۰.۰۰۰	۰.۸۷۹
۰	۰.۳۱۱	۰.۰۰۰	۰.۸۳۱	۰.۲۰۰	۰.۲۶۹	۲.۵۶۶	۰.۰۰۰	۰.۴۶۱
۰	-۰.۴۳۰	۰.۰۰۰	۰.۸۲۶	۰.۲۰۳	۰.۲۷۲	۲.۳۹۳	۰.۰۰۰	۰.۳۰۷
۰	-۰.۲۴۸	۰.۰۰۰	۰.۷۴۹	۰.۱۸۴	۰.۲۵۳	۲.۴۵۵	۰.۰۰۰	۰.۵۴۷
۰	۰.۳۸۵	۰.۲۸۰	۱.۰۱۳	۰.۲۲۹	۰.۲۹۹	۳.۰۸۵	۰.۰۶۹	۱.۱۴۱
۰	۰.۴۱۷	۰.۰۰۰	۰.۹۱۰	۰.۲۱۲	۰.۲۸۲	۲.۸۳۳	۰.۰۰۰	۱.۰۱۵
۰	-۰.۹۶۸	۰.۰۰۰	۰.۸۴۰	۰.۲۰۲	۰.۲۷۱	۲.۵۷۵	۰.۰۰۰	۰.۷۸۸
۰	۰.۴۳۶	۰.۶۰۶	۱.۰۹۷	۰.۲۴۷	۰.۳۱۶	۳.۱۱۰	۰.۴۰۹	۱.۲۰۴
۲	۱.۵۱۶	۱.۳۵۹	۱.۰۲۷	۰.۲۳۳	۰.۳۰۳	۳.۰۳۸	۱.۱۹۱	۱.۴۲۷
۲	۱.۵۵۳	۱.۷۷۲	۱.۱۳۵	۰.۲۶۰	۰.۳۲۸	۲.۹۶۸	۱.۵۳۸	۱.۴۰۸
۲	۱.۹۷۵	۱.۲۷۵	۰.۹۹۷	۰.۲۲۹	۰.۲۹۹	۲.۹۳۱	۱.۰۸۶	۱.۲۸۶
۲	۱.۶۷۱	۱.۶۱۹	۱.۰۹۹	۰.۲۴۸	۰.۳۱۷	۳.۰۹۴	۱.۷۰۵	۱.۴۶۲
۲	۱.۹۷۴	۰.۳۷۱	۱.۰۲۰	۰.۲۳۴	۰.۳۰۴	۲.۹۳۰	۱.۱۶۴	۱.۴۶۷
۲	۳.۲۴۲	۱.۱۸۸	۰.۹۸۶	۰.۲۲۵	۰.۲۹۵	۳.۰۳۰	۱.۶۵۴	۱.۵۲۹
۲	۲.۶۰۹	۰.۱۷۵	۰.۹۷۸	۰.۲۲۴	۰.۲۹۴	۲.۹۷۳	۰.۷۹۰	۱.۵۸۶
۲	۲.۱۲۱	۱.۴۷۶	۱.۰۷۶	۰.۲۳۹	۰.۳۰۸	۳.۲۷۱	۱.۸۸۵	۱.۹۵۸
۲	۱.۳۷۵	۰.۲۱۰	۱.۰۰۴	۰.۲۲۶	۰.۲۹۶	۳.۱۶۳	۰.۹۷۸	۱.۶۸۵
۲	۱.۷۹۳	۱.۵۷۳	۱.۰۹۳	۰.۲۴۵	۰.۳۱۴	۳.۱۷۰	۱.۷۹۳	۱.۵۶۳
۰	-۱.۱۰۴	۰.۰۰۰	۰.۸۵۳	۰.۲۰۵	۰.۲۷۴	۲.۵۷۱	۰.۰۰۰	۰.۲۵۱
۰	۰.۲۹۶	۰.۰۰۰	۰.۸۵۹	۰.۲۱۱	۰.۲۸۰	۲.۳۸۸	۰.۰۰۰	۰.۱۹۹
۰	-۰.۷۵۷	۰.۰۰۰	۰.۶۸۳	۰.۱۷۰	۰.۲۳۸	۲.۴۴۷	۰.۰۰۰	۰.۳۶۰
۰	-۰.۶۵۲	۰.۰۰۰	۰.۸۳۶	۰.۲۰۴	۰.۲۷۳	۲.۴۳۶	۰.۰۰۰	۰.۵۷۳
۰	-۱.۱۷۴	۰.۰۱۰	۰.۹۰۰	۰.۲۱۵	۰.۲۸۴	۲.۶۰۹	۰.۰۰۰	۰.۵۳۶
۰	۰.۱۷۵	۰.۰۰۰	۰.۷۱۱	۰.۱۸۰	۰.۲۴۹	۲.۱۸۸	۰.۰۰۰	۰.۲۴۴
۰	۰.۹۸۰	۰.۰۰۰	۰.۸۷۵	۰.۲۰۹	۰.۲۷۸	۲.۶۳۰	۰.۰۰۰	۰.۷۳۷
۰	۱.۳۷۶	۰.۰۰۰	۰.۷۵۳	۰.۱۸۹	۰.۲۵۷	۲.۲۶۷	۰.۰۰۰	۰.۱۹۹
۰	۰.۲۶۹	۰.۰۱۷	۰.۹۰۷	۰.۲۱۶	۰.۲۸۵	۲.۶۵۴	۰.۰۰۰	۰.۶۸۵
۰	-۰.۵۵۰	۰.۰۰۰	۰.۷۷۵	۰.۱۹۳	۰.۲۶۲	۲.۳۰۹	۰.۰۰۰	۰.۰۹۰
۲	۱.۷۲۰	۰.۰۰۰	۰.۹۲۶	۰.۲۱۴	۰.۲۸۴	۲.۹۱۳	۱.۲۲۵	۱.۳۱۹
۲	۰.۲۲۱	۱.۴۵۷	۱.۰۸۸	۰.۲۳۷	۰.۳۰۷	۳.۴۹۹	۲.۳۴۰	۱.۹۰۸
۲	۲.۵۸۴	۱.۴۶۷	۱.۰۴۷	۰.۲۴۰	۰.۳۰۹	۲.۹۵۳	۱.۲۲۷	۱.۳۷۱
۲	۲.۴۸۷	۱.۷۷۸	۱.۱۳۹	۰.۲۶۰	۰.۳۲۸	۲.۹۹۸	۱.۳۵۰	۱.۴۹۵
۲	۲.۷۳۶	۰.۰۰۰	۰.۸۹۲	۰.۲۰۲	۰.۲۷۲	۳.۲۲۱	۰.۸۱۵	۱.۶۲۶
۲	۲.۳۹۶	۱.۱۱۲	۱.۰۰۳	۰.۲۲۲	۰.۲۹۲	۳.۳۸۳	۲.۰۲۴	۱.۶۱۱
۲	۳.۷۴۲	۱.۱۹۳	۰.۹۸۸	۰.۲۲۵	۰.۲۹۵	۳.۰۳۵	۱.۵۴۶	۱.۶۱۴
۲	۱.۲۲۳	۰.۰۰۴	۰.۹۳۷	۰.۲۱۶	۰.۲۸۶	۲.۹۳۵	۰.۰۳۵	۱.۰۵۴
۲	۱.۸۷۹	۰.۰۰۰	۰.۹۶۸	۰.۲۱۵	۰.۲۸۵	۳.۳۵۹	۱.۷۳۴	۱.۹۴۸
۲	۱.۷۸۷	۱.۴۳۲	۱.۳۸۷	۰.۳۲۱	۰.۳۸۵	۲.۹۹۲	۹.۱۹۳	۰.۸۵۹

سناریو ۳.۱.۵. بردار β را به صورت $\beta = (\underbrace{3, \dots, 3}_{15}, \underbrace{0, \dots, 0}_{25})^T$ در نظر می‌گیریم.

جدول ۱۲.۵: نتایج سناریو ۳.۱.۵ $(p = 40, n = 20, \nu = 0.7)$ با نقطه شروع Ridge								
نتایج روش پیشنهادی					glmnet نتایج			
β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
۳	۰.۳۶۳	۰.۲۹۵	۰.۹۲۵	۰.۱۹۷	۰.۲۶۸	۴.۴۱۰	۰.۱۵۶	۰.۶۸۶
۳	۱.۱۲۸	۱.۰۷۶	۱.۰۷۵	۰.۲۲۱	۰.۲۹۲	۴.۵۳۴	۰.۰۰۰	۰.۶۷۳
۳	۱.۵۲۲	۱.۰۵۹	۱.۰۷۲	۰.۲۱۸	۰.۲۸۹	۴.۷۲۹	۱.۵۸۹	۰.۷۴۶
۳	۳.۸۲۵	۰.۶۶۹	۱.۰۰۷	۰.۲۰۶	۰.۲۷۷	۴.۸۶۷	۱.۴۱۳	۰.۸۴۷
۳	۰.۸۸۸	۰.۲۴۳	۰.۹۲۹	۰.۱۹۹	۰.۲۷۰	۴.۱۱۳	۰.۰۰۰	۰.۶۹۳
۳	-۴.۶۳۱	۰.۰۰۰	۰.۶۶۶	۰.۱۵۴	۰.۲۲۳	۳.۷۵۵	۰.۰۵۰	۰.۶۵۵
۳	۲.۶۷۴	۰.۰۰۰	۰.۷۳۴	۰.۱۶۴	۰.۲۳۴	۴.۱۶۷	۰.۰۰۰	۰.۸۲۶
۳	۷.۸۲۶	۰.۶۳۶	۱.۰۰۴	۰.۲۰۵	۰.۲۷۶	۴.۹۳۴	۵.۵۹۶	۰.۸۲۷
۳	۶.۷۱۴	۰.۰۰۰	۰.۷۹۵	۰.۱۷۶	۰.۲۴۶	۳.۹۸۸	۰.۰۰۰	۰.۷۴۹
۳	۴.۴۹۰	۲.۱۲۵	۱.۳۰۸	۰.۲۵۵	۰.۳۲۵	۵.۰۰۸	۵.۱۲۹	۰.۷۷۹
۳	۳.۵۲۳	۰.۰۰۰	۰.۶۱۷	۰.۱۴۷	۰.۲۱۵	۳.۵۲۰	۰.۰۰۰	۰.۶۳۰
۳	۴.۴۳۰	۰.۰۰۰	۰.۷۶۵	۰.۱۷۰	۰.۲۴۰	۴.۱۳۷	۰.۰۰۰	۰.۷۷۸
۳	۳.۵۹۷	۰.۰۵۹	۰.۹۳۶	۰.۱۹۲	۰.۲۶۴	۵.۱۳۱	۰.۰۸۲	۰.۹۱۹
۳	۴.۵۹۷	۱.۳۳۱	۱.۱۰۲	۰.۲۱۹	۰.۲۹۱	۵.۱۲۹	۳.۶۲۰	۰.۸۶۶
۳	۱.۴۰۲	۰.۲۵۵	۰.۹۴۳	۰.۱۹۶	۰.۲۶۷	۴.۷۵۷	۱.۰۲۵	۰.۸۷۴
۰	-۳.۹۹۸	۰.۰۰۰	۰.۷۴۰	۰.۱۶۳	۰.۲۳۳	۴.۵۵۰	۰.۲۱۷	۰.۹۷۵
۰	۲.۹۹۴	۰.۳۴۰	۰.۹۵۱	۰.۱۹۸	۰.۲۶۹	۴.۶۴۵	۰.۰۰۰	۰.۸۲۲
۰	-۴.۵۳۹	۰.۰۰۰	۰.۸۰۹	۰.۱۷۵	۰.۲۴۶	۴.۵۰۸	۰.۰۰۰	۰.۸۴۳
۰	-۰.۱۱۵	۰.۰۰۰	۰.۶۰۴	۰.۱۴۲	۰.۲۱۰	۳.۹۰۵	۰.۰۰۰	۰.۷۶۹
۰	NA	۰.۸۴۰	۱.۰۳۱	۰.۲۱۲	۰.۲۸۳	۴.۶۷۱	۰.۰۰۰	۰.۷۷۷
۰	NA	۰.۰۰۰	۰.۸۴۴	۰.۱۸۴	۰.۲۵۵	۴.۰۳۵	۰.۰۰۰	۰.۶۸۷
۰	NA	۰.۵۳۱	۰.۹۷۸	۰.۲۰۳	۰.۲۷۵	۴.۵۸۹	۰.۴۷۰	۰.۷۹۳
۰	NA	۰.۰۰۰	۰.۷۴۶	۰.۱۷۲	۰.۲۴۲	۳.۳۲۳	۰.۰۰۰	۰.۵۸۳
۰	NA	۰.۰۰۰	۰.۴۹۶	۰.۱۲۵	۰.۱۹۰	۳.۳۸۳	۰.۰۰۰	۰.۸۰۳
۰	NA	۰.۵۶۰	۰.۹۸۲	۰.۲۰۴	۰.۲۷۵	۴.۵۶۳	۰.۰۰۰	۰.۸۰۵
۰	NA	۰.۰۰۰	۰.۶۷۴	۰.۱۵۶	۰.۲۲۵	۳.۷۲۵	۰.۰۰۰	۰.۷۱۵
۰	NA	۰.۰۲۰	۰.۸۵۴	۰.۱۸۹	۰.۲۶۰	۳.۷۰۴	۰.۰۰۰	۰.۶۲۸
۰	NA	۰.۰۰۰	۰.۷۵۷	۰.۱۶۶	۰.۲۳۶	۴.۵۱۰	۰.۰۰۰	۰.۸۴۰
۰	NA	۰.۰۰۰	۰.۴۵۴	۰.۱۲۱	۰.۱۸۵	۲.۴۱۸	۰.۰۰۰	۰.۴۹۹
۰	NA	۰.۳۲۳	۰.۹۲۵	۰.۱۹۹	۰.۲۶۹	۴.۰۹۲	۰.۰۰۰	۰.۶۳۸
۰	NA	۰.۰۰۰	۰.۶۵۵	۰.۱۵۲	۰.۲۲۱	۳.۷۵۶	۰.۰۰۰	۰.۷۱۰
۰	NA	۰.۰۰۰	۰.۷۷۶	۰.۱۷۴	۰.۲۴۴	۳.۸۳۲	۱.۰۷۵	۰.۷۲۵
۰	NA	۰.۰۰۰	۰.۸۳۷	۰.۱۸۵	۰.۲۵۵	۳.۸۲۲	۰.۰۰۰	۰.۶۶۵
۰	NA	۰.۳۲۹	۰.۹۳۲	۰.۱۹۹	۰.۲۶۹	۴.۲۲۱	۰.۰۰۰	۰.۷۳۴
۰	NA	۰.۰۰۰	۰.۷۳۰	۰.۱۶۸	۰.۲۳۷	۳.۴۹۸	۰.۰۰۰	۰.۵۸۹
۰	NA	۰.۰۰۰	۰.۷۸۳	۰.۱۷۳	۰.۲۴۳	۴.۱۹۹	۰.۰۰۰	۰.۷۳۴
۰	NA	۰.۴۵۹	۰.۹۴۷	۰.۲۰۳	۰.۲۷۴	۴.۰۴۹	۰.۰۰۰	۰.۶۷۷
۰	NA	۰.۰۰۰	۰.۵۵۲	۰.۱۳۷	۰.۲۰۴	۲.۹۹۷	۰.۰۰۰	۰.۵۲۲
۰	NA	۰.۰۰۰	۰.۸۸۶	۰.۱۸۸	۰.۲۵۹	۴.۵۴۶	۱.۷۰۴	۰.۷۹۵
۰	NA	۲.۹۱۳	۱.۶۸۵	۰.۳۲۹	۰.۳۹۴	۴.۶۶۹	۹.۲۹۳	۰.۶۰۹

جدول ۱۳.۵: نتایج سناریو ۳.۱.۵ ($p = ۴۰$, $n = ۵۰$, $\nu = ۰/۷$) (Ridge با نقطه شروع)

نتایج روش پیشنهادی				نتایج glmnet				
β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
۳	۲.۹۴۳	۲.۲۵۳	۲.۷۳۰	۰.۳۴۹	۰.۴۱۱	۴.۳۰۴	۲.۶۰۲	۲.۷۳۹
۳	۲.۹۷۹	۲.۷۴۴	۲.۴۸۲	۰.۳۰۳	۰.۳۶۹	۴.۱۲۹	۲.۶۱۴	۲.۵۲۲
۳	۳.۳۲۲	۱.۳۱۷	۲.۳۴۲	۰.۲۶۶	۰.۳۳۶	۴.۶۰۴	۲.۸۸۴	۳.۱۰۹
۳	۳.۰۹۵	۲.۵۷۳	۱.۴۲۰	۰.۲۸۸	۰.۳۵۶	۴.۲۱۷	۲.۷۲۰	۲.۶۳۹
۳	۲.۸۶۹	۱.۱۰۷	۲.۲۶۶	۰.۲۷۴	۰.۳۴۲	۳.۵۰۳	۲.۵۶۴	۱.۸۳۰
۳	۲.۹۲۴	۲.۲۵۰	۲.۳۲۱	۰.۲۶۹	۰.۳۳۸	۴.۲۳۰	۲.۷۹۶	۲.۷۰۰
۳	۳.۱۱۸	۲.۳۹۹	۲.۰۶۹	۰.۲۳۴	۰.۳۰۴	۳.۴۳۰	۲.۳۹۹	۲.۱۳۲
۳	۳.۱۲۷	۲.۷۵۰	۲.۱۷۲	۰.۲۵۰	۰.۳۲۰	۳.۷۰۶	۲.۷۸۷	۲.۵۰۳
۳	۲.۷۰۵	۲.۹۴۸	۲.۲۲۲	۰.۲۶۱	۰.۳۳۰	۳.۶۳۸	۲.۵۶۵	۲.۲۹۹
۳	۳.۱۳۳	۱.۲۸۶	۲.۳۲۹	۰.۲۷۹	۰.۳۴۷	۳.۸۲۲	۲.۶۷۲	۲.۷۲۲
۳	۲.۸۵۲	۲.۹۵۹	۲.۵۴۱	۰.۳۰۵	۰.۳۷۲	۴.۵۱۱	۲.۸۱۲	۲.۷۷۹
۳	۲.۸۶۴	۲.۲۹۴	۲.۳۲۸	۰.۲۹۶	۰.۳۶۲	۳.۲۲۵	۲.۴۸۲	۱.۹۳۶
۳	۲.۹۹۲	۲.۲۵۸	۲.۹۸۶	۰.۲۲۸	۰.۲۹۸	۲.۸۶۲	۲.۳۱۵	۱.۵۱۷
۳	۲.۸۹۹	۱.۷۸۶	۲.۴۹۸	۰.۳۰۵	۰.۳۷۲	۴.۱۴۳	۲.۷۸۵	۲.۲۴۳
۳	۲.۹۹۲	۲.۵۱۵	۲.۴۰۸	۰.۲۹۵	۰.۳۶۲	۳.۸۱۸	۲.۴۸۳	۲.۲۵۸
۰	۰.۳۷۹	۰.۰۰۰	۰.۸۶۲	۰.۲۰۷	۰.۲۷۶	۲.۵۹۴	۰.۰۰۰	۰.۶۹۷
۰	۰.۱۸۶	۰.۳۲۹	۱.۰۰۲	۰.۳۳۲	۰.۳۰۲	۲.۸۴۷	۰.۰۰۰	۱.۰۴۹
۰	۰.۰۳۷	۰.۰۰۰	۰.۷۴۴	۰.۱۸۵	۰.۲۵۳	۲.۳۵۹	۰.۰۰۰	۰.۲۲۲
۰	-۰.۱۹۱	۰.۰۰۰	۰.۴۶۳	۰.۱۳۴	۰.۱۹۷	۱.۴۷۸	۰.۰۰۰	-۰.۳۰۶
۰	-۰.۰۶۳	۰.۰۰۰	۰.۷۰۶	۰.۱۸۲	۰.۲۵۰	۲.۰۶۹	۰.۰۰۰	-۰.۰۱۲
۰	-۰.۳۳۷	۰.۰۰۰	۰.۸۴۷	۰.۲۰۳	۰.۲۷۳	۲.۵۸۵	۰.۰۰۰	۰.۱۲۶
۰	۰.۰۹۰	۰.۸۴۹	۱.۱۵۵	۰.۲۶۷	۰.۳۳۴	۲.۸۹۷	۰.۰۰۰	۰.۹۱۷
۰	-۰.۲۳۱	۰.۰۰۰	۰.۳۵۳	۰.۱۰۹	۰.۱۶۹	۱.۲۱۹	۰.۰۰۰	-۰.۰۴۲
۰	-۰.۱۹۹	۰.۵۷۷	۱.۰۷۸	۰.۲۴۶	۰.۳۱۵	۲.۹۶۸	۰.۰۰۰	۰.۴۰۵
۰	-۰.۳۵۸	۰.۰۰۰	۰.۸۰۱	۰.۲۰۰	۰.۲۶۹	۲.۲۵۹	۰.۰۰۰	۰.۱۰۹
۰	۰.۰۵۳	۰.۰۰۰	۰.۷۶۹	۰.۱۹۱	۰.۲۶۰	۲.۳۳۳	۰.۰۰۰	۰.۴۷۵
۰	۰.۲۹۹	۰.۰۰۰	۰.۳۸۸	۰.۱۲۳	۰.۱۸۴	۱.۰۶۰	۰.۰۰۰	-۰.۰۵۵
۰	۰.۴۲۰	۰.۰۶۷	۰.۹۲۳	۰.۲۱۸	۰.۲۸۸	۲.۶۹۶	۰.۰۰۰	۰.۶۳۸
۰	۰.۰۸۲	۰.۰۷۶	۰.۹۴۴	۰.۲۱۹	۰.۲۸۹	۲.۸۵۸	۰.۰۰۰	۰.۳۰۷
۰	-۰.۱۶۸	۰.۰۰۰	۰.۷۱۰	۰.۱۸۳	۰.۲۵۱	۲.۰۶۵	۰.۰۰۰	۰.۰۴۳
۰	-۰.۰۸۵	۰.۰۰۰	۰.۷۹۸	۰.۱۹۱	۰.۲۶۰	۲.۶۶۴	۰.۰۰۰	۰.۳۴۶
۰	-۰.۵۴۳	۰.۰۰۰	۰.۶۷۰	۰.۱۶۹	۰.۲۳۷	۲.۲۹۲	۰.۰۰۰	۰.۱۵۱
۰	۰.۱۷۸	۰.۰۰۰	۰.۸۷۰	۰.۲۱۳	۰.۲۸۲	۲.۴۲۴	۰.۰۰۰	۰.۳۸۷
۰	۰.۱۴۹	۰.۲۱۳	۰.۹۰۱	۰.۲۲۶	۰.۲۹۴	۲.۲۴۹	۰.۰۰۰	۰.۳۴۲
۰	۰.۲۲۴	۰.۰۰۰	۰.۴۸۸	۰.۱۳۳	۰.۱۹۷	۱.۸۹۸	۰.۰۰۰	۰.۰۵۵
۰	۰.۱۲۱	۰.۰۰۰	۰.۷۱۴	۰.۱۷۳	۰.۲۴۲	۲.۶۶۲	۰.۰۰۰	۰.۴۷۴
۰	۰.۵۳۲	۰.۰۰۰	۰.۸۸۲	۰.۲۰۸	۰.۲۷۸	۲.۷۴۰	۰.۰۰۰	۰.۸۲۲
۰	-۰.۲۳۷	۰.۰۰۰	۰.۹۵۴	۰.۲۱۶	۰.۲۸۶	۳.۱۴۵	۰.۰۰۰	۰.۳۴۳
۰	-۰.۰۳۷	۰.۰۰۰	۰.۷۲۱	۰.۱۷۴	۰.۲۴۳	۲.۷۱۴	۰.۰۰۰	۰.۴۲۷
۰	-۰.۰۳۸	۲.۴۵۸	۱.۸۸۴	۰.۴۰۸	۰.۴۶۲	۳.۷۹۱	۲.۵۵۱	۰.۹۱۴

جدول ۱۴.۵ مقدار میانگین توان دوم خطا را برای برآوردهای کمترین توان‌های دوم، ریدج و لاسو تحت مفروضات ۲.۱.۵ و ۳.۱.۵ به ازای $m = 1000$ تکرار نشان می‌دهد. این نتایج از رابطه‌ی زیر به‌دست می‌آیند:

$$MSE(\hat{\beta}) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^p (\hat{\beta}_{ji} - \beta_j)^2$$

جدول ۱۴.۵: نتایج MSE سناریوهای ۲.۱.۵ و ۳.۱.۵

سناریو ۳.۱.۵				سناریو ۲.۱.۵				MSE	ν
۱۰۰	۵۰	۲۰	۵	۱۰۰	۵۰	۲۰	۵		
۱۴.۹	۴۲.۰	NA	NA	۳۶۵.۴۱۰	۲۶۳۳.۰۲۲	NA	NA	OLS	۰.۷
۷.۴۷۲	۱۶.۱۰	۸۴.۸۱	۹۷.۵۰	۲۴.۲۸۸	۳۲.۲۸۴	۴۱.۹۳	۵۱.۶۸	Ridge	
۶.۹۷۰	۱۱.۹۴	۱۷۱.۱۵	۴۱۹.۹۷	۳۲.۳۳۰	۸۵.۷۷	۱۴۱.۶	۲۸۳.۳	Lasso	
۷۳	۹۴.۴۶	NA	NA	۱۱۱۵.۲۹۴	۸۴۶۹.۳۲۶	NA	NA	OLS	۰.۹
۲۱.۳۱	۳۰.۶۱	۸۴.۹۳	۸۸.۷۳	۱۲.۶۹	۱۸.۴۱	۴۱.۳۷	۴۳.۹۶	Ridge	
۳۵.۵۴	۶۱.۰۶	۲۷۱.۲	۵۹۹.۸	۴۶.۱۵	۸۶.۲۳	۲۳۲.۹	۴۱۶.۶	Lasso	
۱۴۲.۵	۱۵۶.۲	NA	NA	۱۳۱۴.۱۶	۹۵۲۵	NA	NA	OLS	۰.۹۹
۷۵.۲۳	۷۸.۵۹	۸۶.۴۴	۸۶.۸۵	۳۸.۶۶	۳۸.۸۶	۴۱.۸۳	۴۲.۱۴	Ridge	
۳۲۷.۲	۴۶۱.۳	۵۹۹.۷	۶۹۲.۶	۴۳۲.۸۹	۸۷۱.۸	۱۰۵۰.۳	۱۱۵۳.۸	Lasso	

براساس جدول ۱۴.۵ زمانی که تعداد متغیرها بیش از مشاهدات است یا وقتی داده‌ها دارای هم‌خطی می‌باشند، برآورگر کمترین توان‌های دوم نمی‌تواند به خوبی ضرایب را به‌دست آورد و مقدار میانگین توان دوم خطای آن بسیار زیاد بوده اما برآوردر لاسو بهتر از برآوردر کمترین توان‌های دوم ضرایب را پیدا کرده و دارای میانگین توان دوم خطای کمتری است. اما رگرسیون ریدج بهترین دقت پیش‌بینی در بین سایر روش‌هاست.

سناریو ۴.۱.۵. بردار β را به صورت $\beta = (3, 1/5, 0, 1, 0, 0, 0, 0)^T$ در نظر می‌گیریم.

جدول ۱۵.۵: نتایج سناریو ۴.۱.۵ (OLS با نقطه شروع OLS) $p = 8, n = 20, \nu = 0.7$

β	نتایج روش پیشنهادی						نتایج glmnet	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
۳.۰	۳.۰۸۸	۲.۸۰۰	۳.۰۵۸	۰.۱۴۷	۰.۸۴۴	۲.۸۷۳	۳.۰۳۰	۲.۶۳۴
۱.۵	۱.۵۱۷	۱.۳۱۳	۱.۵۰۴	۰.۱۱۱	۰.۶۲۶	۱.۴۸۹	۱.۴۱۷	۱.۱۵۵
۰.۰	-۰.۰۵۰	۰.۰۰۰	-۰.۰۷۰	-۰.۰۱۲	-۰.۰۶۰	-۰.۰۷۰	۰.۰۰۰	۰.۰۱۲
۱.۰	۰.۸۴۱	۰.۴۱۶	۰.۸۲۸	۰.۰۴۶	۰.۳۲۵	۰.۸۲۶	۰.۷۲۰	۰.۶۹۵
۰.۰	-۰.۰۲۴	۰.۰۰۰	-۰.۰۵۷	-۰.۰۰۹	-۰.۰۴۹	-۰.۰۵۷	۰.۰۰۰	-۰.۱۱۲
۰.۰	-۰.۰۲۷	۰.۰۰۰	-۰.۰۴۰	-۰.۰۰۸	-۰.۰۳۶	-۰.۰۴۰	۰.۰۰۰	-۰.۰۶۱
۰.۰	۰.۰۹۸	۰.۰۰۰	۰.۰۸۸	۰.۰۱۰	۰.۰۶۷	۰.۰۸۹	۰.۰۰۰	-۰.۰۱۷
۰.۰	-۰.۱۲۹	۰.۰۰۰	-۰.۱۴۲	-۰.۰۲۳	-۰.۱۱۵	۰.۱۴۲	۰.۰۰۰	۰.۴۹۴

سناریو ۵.۱.۵. تحت مفروضات سناریو ۴.۱.۵ فرض می‌کنیم $R = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & -2 & 3 \\ 0 & -1 & 2 & 0 & 0 & 2 & 3 & 0 \\ -3 & 1 & 2 & 5 & 4 & 0 & 0 & 0 \end{bmatrix}$ ، $r = [1/9, 3/8, 2/5]^T$ ، $p = 8$ و $q = 3$ است.

جدول ۱۶.۵: نتایج سناریو ۵.۱.۵

β	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$
۳.۰	۳.۱۲۲	۱.۲۱۵	۲.۶۸۳	۲.۹۳۰	۲.۸۲۲	۲.۸۷۳
۱.۵	۱.۵۴۴	-۱.۳۹۹	۱.۳۴۹	۱.۴۹۵	۱۵.۲۳۹	۱.۴۸۹
۰.۰	-۰.۱۲	۰.۳۰۱	۰.۶۷۱	۰.۸۰۷	۱.۸۴۴	۰.۸۹۲
۱.۰	۰.۸۸۳	۱.۳۴۸	۱.۳۱۴	۱.۰۸۳	-۰.۷۸۶	۰.۸۳۳
۰.۰	۰.۰۰۳	۰.۰۵۰	۰.۳۲۲	۰.۶۹۱	-۱.۰۰۸	۰.۹۲۰
۰.۰	-۰.۰۲۶	۰.۱۵۵	۰.۴۳۴	۰.۴۰۲	۳.۰۷۵	۰.۳۳۵
۰.۰	۰.۲۲۲	۰.۴۹۷	۰.۹۸۰	۰.۹۵۸	۳.۰۶۷	۰.۹۴۵
۰.۰	-۰.۲۹۹	۰.۰۴۲	-۰.۲۹۸	-۰.۴۳۰	۱.۳۱۱	-۰.۳۹۰

۲.۵ کاربرد

در این بخش کارایی برآوردگرهای پیشنهاد شده را در قالب دو مثال واقعی بررسی می‌کنیم.

۱.۲.۵ داده‌های سرطان پروستات

در این بخش از داده‌های پروستات که استیمی و همکاران (۱۹۸۹) ارائه داده‌اند، استفاده می‌کنیم. بردار ضرایب برآوردگرهای بریج در جدول ۱۷.۵ آمده است.

جدول ۱۷.۵: ضرایب متغیرهای پروستات

	نتایج glmnet						نتایج روش پیشنهادی	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
lcavol	۰.۵۸۷۰۲۲۸۱	۰.۵۸۶۲۷۳۰۲۵	۰.۵۸۶۹۱۲۸۵۶	۰.۵۸۷۱۷۱۷۰۱	۰.۵۸۷۱۸۰۱۸۳	۰.۵۸۷۲۰۱۲۲۲	۰.۵۸۳۲۶۹۲۷۷	۰.۴۷۹۰۴۹۰۶۸
lweight	۰.۴۵۴۴۶۰۶۴۱	۰.۴۵۰۲۳۷۲۴۶	۰.۴۵۳۹۸۱۳۰۵	۰.۴۵۴۲۳۵۰۹۶	۰.۴۵۴۳۰۲۲۵۵	۰.۴۵۴۴۳۰۹۶۲	۰.۴۵۱۵۷۲۵۷۸	۰.۴۳۳۶۴۸۵۴۶
age	-۰.۰۱۹۶۳۷۲۰۸	-۰.۰۱۹۶۱۸۴۱۵	-۰.۰۱۹۶۳۷۱۱۵	-۰.۰۱۹۶۱۴۶۷۹	-۰.۰۱۹۶۱۴۶۹۲	-۰.۰۱۹۶۱۴۶۹۵	-۰.۰۱۹۰۳۰۱۱۲	-۰.۰۱۳۲۷۴۵۸۴
lbph	۰.۱۰۷۰۵۴۳۵۱	۰.۱۰۶۵۵۹۴۶۰	۰.۱۰۷۰۴۱۱۰۸	۰.۱۰۷۰۹۵۰۷۲	۰.۱۰۷۰۹۸۳۶۴	۰.۱۰۷۱۰۰۱۸۸	۰.۱۰۵۵۳۰۸۱۲	۰.۰۸۹۸۷۸۲۷۰
svi	۰.۷۶۶۱۵۵۸۸۵	۰.۷۶۰۰۷۸۱۹۰	۰.۷۶۴۹۹۳۵۳۵	۰.۷۶۵۰۹۲۷۰۱	۰.۷۶۵۰۲۹۸۳۲	۰.۷۶۵۰۰۳۲۴۳	۰.۷۵۵۳۷۱۹۶۸	۰.۶۵۹۶۱۷۰۰۹
lcp	-۰.۰۱۰۵۴۷۳۵۷۰	-۰.۰۱۰۴۹۴۰۷۷۶	-۰.۰۱۰۵۴۵۹۵۲۲	-۰.۰۱۰۵۱۵۰۶۹۶	-۰.۰۱۰۵۱۵۴۱۶۳	-۰.۰۱۰۵۱۵۶۰۶۳	-۰.۰۹۷۸۸۶۱۷۰	-۰.۰۱۳۱۴۵۵۶۹
gleason	۰.۰۴۵۱۳۵۹۶۴	۰.۰۴۳۱۳۸۴۳۶	۰.۰۴۵۱۱۳۴۳۶	۰.۰۴۵۴۰۸۵۲۳	۰.۰۴۵۴۱۳۰۱۹	۰.۰۴۵۴۱۴۴۷۰	۰.۰۴۳۱۳۷۲۸۳	۰.۰۶۶۴۲۸۹۷۴
pgg45	۰.۰۰۴۵۲۵۳۲۴	۰.۰۰۴۵۲۴۰۱۴	۰.۰۰۴۵۲۵۳۲۲	۰.۰۰۴۵۳۵۵۹۵	۰.۰۰۴۵۳۵۵۹۶	۰.۰۰۴۵۳۵۵۹۶	۰.۰۰۴۳۷۷۴۴۴	۰.۰۰۳۱۵۲۶۲۴
$\hat{k}$	۰	۰.۲	۰.۰۲۵	۰.۰۲۵	۰.۰۲۵	۰.۰۲۵	۰.۰۰۱۵۰۸۵۹۶	۰.۰۰۱۵۰۸۵۹۶
MPE	۰.۴۵۵۲۹	۰.۴۵۵۲۹۰۴	۰.۴۵۵۳۰۸۲	۰.۴۵۵۲۹۰۶	۰.۴۵۵۲۹۰۶	۰.۴۵۵۲۹۰۷	۰.۴۵۵۳۴۱۵	۰.۴۶۶۷۶۰۹

توجه نمایید با افزایش مقدار k برآوردگر لاسو تعدادی بیشتری از متغیرها را صفر می‌کند. برای مثال به ازای $k = ۱۰$ و $k = ۳۰$ ضرایب برآوردگر لاسو در جدول ۱۸.۵ آمده است.

جدول ۱۸.۵: نتایج ضرایب لاسو به ازای مقادیر مختلف k

	$k = ۳۰$	$k = ۱۰$	$k = ۰/۲$
lcavol	۰.۴۷۴۵۴۴۵۶۱	۰.۵۴۹۵۳۰۱۰۸	۰.۵۸۶۲۷۳۰۲۵
lweight	۰.۰۰۰۰۰۰۰۰۰۰	۰.۲۴۳۲۹۰۸۸۸	۰.۴۵۰۲۳۷۲۴۶
age	-۰.۰۱۶۸۱۸۳۲۵	۰.۰۱۸۶۹۷۵۸۰	-۰.۰۱۹۶۱۸۴۱۵
lbph	۰.۰۳۲۸۲۰۶۷۵	۰.۰۸۲۳۰۹۷۹۳	۰.۱۰۶۵۵۹۴۶۰
svi	۰.۰۰۰۰۰۰۰۰۰۰	۰.۴۶۲۲۷۱۱۷۳	۰.۷۶۰۰۷۸۱۹۰
lcp	-۰.۰۲۵۵۵۴۴۶۶	-۰.۰۷۸۸۳۸۶۸	-۰.۰۱۰۴۹۴۰۷۷۶
gleason	۰.۰۰۰۰۰۰۰۰۰۰	۰.۰۰۰۰۰۰۰۰۰۰	۰.۰۴۳۱۳۸۴۳۶
pgg45	۰.۰۰۴۳۲۸۸۹۸	۰.۰۰۴۴۵۹۸۴۹	۰.۰۰۴۵۲۴۰۱۴
MPE	۰.۶۶۵۷۴۱۱	۰.۴۹۳۴۹	۰.۴۵۵۳۰۸۲

۲.۲.۵ داده‌های میزان رشد گندم

داده‌های آیووا (Iowa) میزان رشد گندم برای ایالت متحده آیووا در سال‌های ۱۹۶۲-۱۹۳۰ با توجه به میزان بارندگی فصل قبل، سه ماه فصل رشد و فصل برداشت و همچنین متوسط درجه حرارت اندازه‌گیری شده برای این پنج ماه را نشان می‌دهد. در جدول ۱۹.۵ توصیفی از این متغیرها و در جدول ۲۰.۵ آماره‌های توصیفی متغیر پاسخ (Yield) آمده است.

جدول ۱۹.۵: توصیف متغیرهای داده‌های آیووا		
متغیرها	توصیف	توضیحات
Yield	میزان گندم سال داده شده	متغیر پاسخ
Year	سال اندازه‌گیری	
Rain0	بارش فصل قبل	
Temp1	متوسط دما اولین ماه رشد	
Rain1	میزان بارش اولین ماه رشد	
Temp2	متوسط دما دومین ماه رشد	
Rain2	میزان بارش دومین ماه رشد	
Temp3	متوسط دما سومین ماه رشد	
Rain3	میزان بارش سومین ماه رشد	
Temp4	متوسط دما ماه برداشت	

جدول ۲۰.۵: آماره‌های توصیفی متغیر پاسخ در داده‌های آیووا						
مینیمم	چارک اول	میانه	چارک سوم	ماکزیمم	میانگین	انحراف استاندارد
۲۰.۰	۴۳.۱	۵۲	۵۹.۰	۷۶.۰	۵۰.۰	۱۳.۱۸

بردار ضرایب برآوردهای بریج در جدول ۲۱.۵ آمده است.

جدول ۲۱.۵: ضرایب متغیرهای داده‌های آیووا

نتایج glmnet							نتایج روش پیشنهادی	
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{\gamma=1}$	$\hat{\beta}_{\gamma=2}$	$\hat{\beta}_{\gamma=2.5}$	$\hat{\beta}_{\gamma=3}$	$\hat{\beta}_{\gamma=4}$	$\hat{\beta}_{LASSO}$	$\hat{\beta}_{Ridge}$
Year	۰.۸۷۵۵۸۱۷	۰.۴۸۲۴۹۲	۰.۸۷۵۵۷۴۴۰	۰.۸۷۵۵۷۳۱۶	۰.۸۷۵۵۷۳۱۱	۰.۸۷۵۵۷۰۵	۰.۵۷۲۱۶۹۳	۰.۶۱۸۰۴۹۶
Rain0	۰.۷۸۴۵۰۱۰۳	۰.۰۰۰۰۰۰۰۰	۰.۷۸۴۴۶۵۵۳	۰.۷۸۴۴۶۱۷۲	۰.۷۸۴۴۵۹۲۵	۰.۷۸۴۴۵۷۳	۰.۰۰۰۰۰۰۰۰	۰.۴۲۹۵۰۸۶
Temp1	-۰.۴۵۹۷۸۶۹۹	۰.۰۰۰۰۰۰۰۰	-۰.۴۵۹۷۶۰۴۶	-۰.۴۵۹۷۶۴۵۱	-۰.۴۵۹۷۶۸۷۰	-۰.۴۵۹۷۷۵۸	۰.۰۰۰۰۰۰۰۰	-۰.۲۹۰۸۳۸۹
Rain1	۰.۷۷۹۴۰۶۹۵	۰.۰۰۰۰۰۰۰۰	-۰.۷۷۹۱۸۵۲۱	-۰.۷۷۹۱۶۲۳۰	-۰.۷۷۹۱۴۷۸۱	-۰.۷۷۹۱۳۷۸	۰.۰۰۰۰۰۰۰۰	-۰.۷۷۴۴۴۴۴
Temp2	۰.۴۸۲۲۰۵۳۱	۰.۰۰۰۰۰۰۰۰	۰.۴۸۲۱۶۹۷۹	۰.۴۸۲۱۷۴۴۸	۰.۴۸۲۱۷۹۶۲	۰.۴۸۲۱۸۸۸	۰.۰۰۰۰۰۰۰۰	۰.۱۱۵۶۱۲۸
Rain2	۲.۵۵۶۹۶۸۱۵	۰.۱۹۳۹۳۹۱۲	۲.۵۵۵۹۴۶۷۹	۲.۵۵۴۹۲۸۲۷	۲.۵۵۳۰۶۱۱۷	۲.۵۴۳۸۱۲۵	۰.۱۲۹۴۳۴	۲.۰۷۷۵۰۴۹
Temp3	۰.۴۸۹۰۸۶۲	-۰.۰۹۵۲۱۴۸۸	۰.۴۸۹۰۴۳۶	۰.۴۸۹۰۷۴۴	۰.۴۸۹۰۸۳۱	۰.۴۸۹۰۸۶	-۰.۰۴۹۰۷۵۱	-۰.۴۹۶۴۱۱۱
Rain3	۰.۴۰۷۸۰۳۱۵	۰.۰۰۰۰۰۰۰۰	۰.۴۰۷۶۶۸۷۱	۰.۴۰۷۶۹۵۸۴	۰.۴۰۷۷۲۰۹۲	۰.۴۰۷۷۵۸۴	۰.۰۰۰۰۰۰۰۰	۰.۶۴۱۰۷۶۱
Temp4	-۰.۶۵۶۲۹۳۶۹	-۰.۰۴۳۱۳۲۷۲	-۰.۶۵۶۲۲۶۳۲	-۰.۶۵۶۲۲۵۴۷	-۰.۶۵۶۲۲۷۳۸	-۰.۶۵۶۲۳۵۷	۰.۰۰۰۰۰۰۰۰	-۰.۵۱۰۷۵۶۸
$\hat{k}$	۰	۱.۵	۱	۱	۱	۱	۰.۰۸۶۹۸۱۷۲	۰.۰۸۶۹۸۱۷۲
MPE	۰.۴۵۵۲۹	۰.۳۸۸۶۱۶۱	۰.۲۴۴۷۹۰۳	۰.۳۱۵۶۱۶	۰.۳۱۷۷۸۵۷	۰.۳۲۰۳۴۶۹	۰.۲۹۱۰۷۹۶	۰.۲۷۳۰۷۶۶

۳.۵ مزیت‌های روش نیوتن رافسون اصلاح شده و الگوریتم پیش‌برنده لاسو

۱- بسته‌های glmnet و lasso2 در زبان برنامه‌نویسی C++ طراحی شده‌اند که امکان تغییر در ساختار آن‌ها با استفاده از دستورات R وجود ندارد، در صورتی که برنامه پیشنهادی نیوتن رافسون اصلاح شده و الگوریتم پیش‌برنده لاسو این امکان را فراهم می‌کند که هر گونه تغییری در ساختار آن بتوان ایجاد نمود. برای مثال می‌توان برآوردگرهای محدود شده ریج و لاسو را محاسبه کرد.

۲- از آنجایی که بسته‌ای در R برای محاسبه برآوردگرهای بریج به‌ازای $\gamma = 1, 2$ وجود ندارد لذا برنامه‌ی نیوتن رافسون اصلاح شده و الگوریتم پیش‌برنده لاسو این امکان را به وجود می‌آورد که بتوان جواب برآوردگرهای بریج را به ازای سایر مقادیر γ به‌دست آورد.

۳- حال سوالی که پیش می‌آید این است که چطور می‌توان به جواب‌های حاصل از برنامه‌ی نیوتن رافسون و الگوریتم پیش‌برنده لاسو در R اعتماد کرد.

ابتدا توجه نمایید که جواب‌های بسته‌های glmnet و lasso2 که هر دو تایید شده می‌باشند، با یکدیگر فرق دارند. حال برای پاسخ به سوال فوق جواب‌های برآوردگرهای بریج به ازای مقادیر مختلف γ را با خروجی بسته glmnet مقایسه کرده‌ایم. همین طور که مشاهده می‌کنید جواب‌ها تا حد زیادی به یکدیگر شبیه هستند. از طرفی میانگین خطای پیشگویی (MPE) برآوردگرهای ریج و لاسو روش‌های نیوتن رافسون اصلاح شده و الگوریتم پیش‌برنده لاسو با MPE در بسته glmnet بسیار نزدیک هم هستند.

پیوست آ

خلاصه و پیشنهادات برای تحقیقات آینده

۱.آ خلاصه

بهبود دادن برآوردگرهای معادلات رگرسیونی، زمینه کاری بسیاری از آماردانان دهه اخیر است. یکی از عواملی که منجر به تلاش برای یافتن برآوردگرهای بهبودیافته می‌شود، وجود همخطی در ماتریس طرح است. همانطور که در این پایان‌نامه شرح آن گذشت، یکی از این برآوردگرهای بهبودیافته که در سال‌های اخیر از سوی آماردانان مورد توجه زیادی قرار گرفته است، برآوردگر بریج و حالات خاص آن یعنی ریج و لاسو می‌باشد. وقتی که با مسئله همخطی مواجه هستیم، برآوردگر بریج در رده برآوردگرهای اریب، بهترین برآوردگر است و خطای آن از خطای برآوردگر کمترین توان‌های دوم کمتر است و همچنین دارای ویژگی‌های مفید و خوبی است که استفاده از این برآوردگر را توجیه می‌کنند. در این مجموعه سعی شده است، ضرایب معادلات رگرسیونی چندگانه را با در نظر گرفتن مسئله همخطی و استفاده از داده‌هایی که در آن‌ها این مشکل به چشم می‌خورد، برآورد کنیم. ابتدا برآوردگر بریج و حالات خاص آن و الگوریتم نیوتن رافسون اصلاح شده و پیش‌برنده لاسو را برای مدل رگرسیونی چندگانه خطی وقتی که ماتریس واریانس- کوواریانس خطا، $\sigma^2 I$ است ارائه کردیم و ضرایب برآوردگر بریج را یافتیم. همچنین رفتار مجانبی این برآوردگر را نیز مورد بررسی قرار دادیم. سپس اطلاعات پیشین تجربی به صورت محدودیتی روی فضای پارامترها اعمال و برآوردگر محدود شده را یافتیم و به منظور نشان دادن کارایی این برآوردگر، خطای آن را محاسبه و به کمک شبیه‌سازی و مثال عددی برتری آن را بر برآوردگر کمترین توان‌های دوم محدود شده، نشان دادیم. از آنجایی که محققان معمولاً اطلاعات پیشین دقیق ندارند و این اطلاعات

نیز می‌توانند دستخوش خطا شوند، در این پایان‌نامه حالتی کلی‌تر را برای مدل رگرسیونی محدود شده در نظر گرفتیم و ضرایب مدل رگرسیونی را در حالتی که محدودیتی روی فضای پارامتر اعمال شده بود بدست آوردیم و همچنین مسئله برآورد و ویژگی‌های آن بررسی شد.

۲.آ پیشنهادات برای پژوهش‌های آینده (ارائه زمینه‌های تحقیق)

با توجه به نتایج بدست آمده در این مجموعه و موضوعات مطرح شده، می‌توان در آینده بر روی موارد زیر تحقیق کرد.

- (۱) بدست آوردن خطای برآوردگر ارائه‌شده در فصل چهارم بر حسب مقادیر ویژه ماتریس‌ها و مقایسه این خطا با خطای برآوردگر کمترین توالی دوم محدود شده، به منظور بررسی کارایی آن.
- (۲) در مدل ارائه‌شده در فصل چهارم (مدل با محدودیت تصادفی) می‌توان وقتی که خطای مدل و همچنین خطای محدودیت، توزیع t چندمتغیره یا در حالت کلی‌تر توزیع بیضی‌گون دارند، ضرایب مدل رگرسیونی را برآورد کرد و ویژگی‌های برآوردگر را بررسی کرد.
- (۳) استفاده از معیار آکایی به جای معیار توان دوم خطا برای مقایسه برآوردگرها و مقایسه نتایج آن با نتایج بدست آمده در این مجموعه.

پیوست ب

ب.۱ جبر ماتریس‌ها

تعریف ب.۱.۱. ماتریس متعامد: ماتریس مربعی که هر جفت از ستون‌ها شامل دو بردار متعامد است.

قضیه ب.۲.۱. اگر A ماترسی متعامد باشد آنگاه $A^T = A^{-1}$.

قضیه ب.۳.۱. ماتریس A متعامد است اگر و تنها اگر $A^T A = I$.

تعریف ب.۴.۱. ماتریس وارون: ماتریس A^{-1} را وارون ماتریس A می‌نامیم اگر و تنها اگر $AA^{-1} = I$.

تعریف ب.۵.۱. وارون وارون ماتریس A ، برابر خود ماتریس A است. یعنی $(A^{-1})^{-1} = A$.

قضیه ب.۶.۱.

$$(A^T)^{-1} = (A^{-1})^T$$

قضیه ب.۷.۱.

$$(AB)^{-1} = B^{-1}A^{-1}$$

قضیه ب.۸.۱.

$$|A^{-1}| = \frac{1}{|A|}$$

قضیه ب.۹.۱. وارون یک ماتریس متقارن، متقارن است.

تعریف ب.۱۰.۱. اثر ماتریس: اگر A ماتریس مربع باشد، آنگاه اثر A ، مجموع درایه‌های قطر اصلی A است. اثر ماتریس A را با نماد $\text{tr}(A)$ نشان می‌دهیم.

قضیه ب.۱۱.۱. اگر A و B و C ماتریس‌های مربعی باشند، آنگاه $\text{tr}(AB) = \text{tr}(BA)$ و $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$ در صورتی که ضرب بین آن‌ها تعریف شده باشد.

تعریف ب.۱۲.۱. ماتریس قطری: ماتریس مربع متقارنی است که همه درایه‌های آن، به جز درایه‌های روی قطر اصلی، برابر صفر است.

تعریف ب.۱۳.۱. ماتریس اسکالر: اگر درایه‌های قطر اصلی ماتریس قطری با هم برابر باشند، آنگاه این ماتریس را اسکالر گویند.

تعریف ب.۱۴.۱. ماتریس واحد: ماتریس اسکالر که درایه‌های روی قطر اصلی آن برابر یک باشد که آن را با I نشان می‌دهیم.

تعریف ب.۱۵.۱. ماتریس معین مثبت و معین نامنفی: ماتریس A را معین مثبت گویند اگر تمام مقادیر ویژه آن مثبت باشد و اگر مقادیر ویژه A نامنفی باشد، گفته می‌شود A نیمه معین مثبت یا معین نامنفی است.

قضیه ب.۱۶.۱. ماتریس A نیمه معین مثبت است اگر و تنها اگر ماتریس نامنفرد P وجود داشته باشد به طوری که $A = P^T P$.

تعریف ب.۱۷.۱. دترمینان ماتریس: دترمینان ماتریس مربعی A با بعد $p \times p$ ، یک مقدار عددی است که شامل مجموع تعداد جملاتی است که هر جمله حاصل ضرب p درایه ماتریس A می‌باشد.

قضیه ب.۱۸.۱. اگر A و B دو ماتریس، مربعی باشند، آنگاه

$$|AB| = |BA| = |A| \times |B|$$

قضیه ب.۱۹.۱. اگر A ماتریس $p \times p$ و k یک مقدار دلخواه باشد، آنگاه $|kA| = k^p |A|$.

تعریف ب.۲۰.۱. ماتریس منفرد: اگر $|A| \neq 0$ آنگاه ماتریس A را ماتریس منفرد (تکین) می‌نامیم. ماتریسی که دترمینان آن غیر صفر باشد، ماتریس نامنفرد (ناتکین) نامیده می‌شود.

قضیه ب.۲۱.۱. فرض کنید A یک ماتریس قطری نامنفرد و به صورت زیر باشد

$$A = \text{Diag}(a_1, a_2, \dots, a_p)$$

در این صورت، وارون آن به صورت زیر است:

$$A^{-1} = \text{Diag}\left(\frac{1}{a_1}, \frac{1}{a_2}, \dots, \frac{1}{a_n}\right)$$

قضیه ب.۲۲.۱. ماتریس منفرد، وارون پذیر نیست.

تعریف ب.۲۳.۱. وابسته خطی: گوییم بردارهای $x_1, x_2, \dots, x_p$ وابسته خطی هستند اگر و تنها اگر، رابطه خطی به صورت زیر بین آن ها وجود داشته باشد:

$$c_1 x_1 + c_2 x_2 + \dots + c_p x_p = 0$$

که در آن همه ضرایب عددی $c_1, c_2, \dots, c_p$ صفر نیستند. در صورتی که یک مجموعه بردار، وابسته خطی نباشند، مستقل خطی هستند.

تعریف ب.۲۴.۱. رتبه ماتریس: رتبه ماتریس A که به صورت $rank(A)$ نوشته می شود. بعد بزرگترین زیر ماتریس مربع و نامنفرد A است. همچنین تعداد ستون ها یا سطرهاى مستقل خطی یک ماتریس، نشان دهنده رتبه آن ماتریس است.

قضیه ب.۲۵.۱. اگر A ماتریس $p \times p$ و نامنفرد باشد، آنگاه رتبه آن برابر p است.

تعریف ب.۲۶.۱. وارون تعمیم یافته یک ماتریس: اگر B ماتریس باشد به طوری که Bh جوابی برای معادلات سازگار $Ax = h$ باشد، گوییم B وارون تعمیم یافته A است و با نماد A^- نشان می دهیم.

قضیه ب.۲۷.۱. وارون تعمیم یافته یک ماتریس، یکتا نیست.

قضیه ب.۲۸.۱.

$$AA^-A = A$$

قضیه ب.۲۹.۱. فرض کنید A یک ماتریس مربع باشد، آنگاه مشتق ماتریس معکوس A نسبت به کمیت k از رابطه زیر بدست می آید

$$\frac{dA^{-1}}{dk} = -A^{-1} \frac{dA}{dk} A^{-1}$$

قضیه ب.۳۰.۱. هر ماتریس معین مثبت دارای دترمینان مثبت و وارون پذیر است.

تعریف ب.۳۱.۱. ماترس A خودتوان است اگر $A^2 = A$ باشد.

قضیه ب.۳۲.۱. ماتریس متقارن A معین مثبت است اگر و فقط اگر ماتریس ناویژه P وجود داشته باشد به طوری که $A = P^T P$

قضیه ب.۳۳.۱. اگر A یک ماتریس متقارن خودتوان با رتبه r باشد آنگاه

$$rank(A) = tr(A) = r$$

قضیه ب.۳۴.۱. مقادیر ویژه یک ماتریس خودتوان، صفر یا یک است.

قضیه ب.۳۵.۱. فرض کنید A یک ماتریس $n \times p$ است.

(۱) اگر $rank(A) = p$ آنگاه $A^T A$ معین مثبت است.

(۲) اگر $rank(A) < p$ آنگاه $A^T A$ نیمه معین مثبت است.

قضیه ب.۳۶.۱. اگر $A_{n \times n}$ معین مثبت (نیمه معین مثبت) با مقادیر ویژه $\lambda_1, \dots, \lambda_n$ باشد، آنگاه:

(۱) عناصر قطری اصلی $a_{ii} > 0$ ($a_{ii} \geq 0$) می باشد.

(۲) A ناویژه است.

(۳) A^{-1} معین مثبت است.

(۴) $P^T A P$ معین مثبت (نیمه معین مثبت) است، که در آن P یک ماتریس ناویژه است.

تعریف ب.۳۷.۱. تجزیه طیفی: ماتریس A را می توان به مقادیر و بردارهای ویژه اش به صورت زیر تجزیه کرد.

$$A = \Gamma \Lambda \Gamma^T$$

که در آن Λ یک ماتریس قطری با مقادیر ویژه ماتریس A و Γ ماتریس متعامدی است که ستون های آن بردارهای ویژه متناظر با مقادیر ویژه A می باشد.

قضیه ب.۳۸.۱. اگر A ماتریس متقارن ثابت، Y بردار تصادفی با میانگین μ و ماتریس کواریانس Σ باشد، آنگاه

$$E(Y^T A Y) = \text{tr}(A \Sigma) + \mu^T A \mu$$

□

برهان. سیرل (۱۹۷۱) را ببینید.

قضیه ب.۳۹.۱. فرض کنید $V = AY$ و $W = VY$ که در آن A ماتریس ثابت $n \times p$ و B ماتریس $p \times n$ و Y بردار تصادفی p بعدی با ماتریس واریانس-کواریانس Σ باشد، آنگاه

$$\text{Var}(V) = \text{Var}(AY) = A \Sigma A^T$$

$$\text{Cov}(v, W) = \text{Cov}(AY, BY) = A \Sigma B^T$$

ب.۲ قضیه های حدی

تعریف ب.۱.۲. دنباله $\{E_n, n \geq 1\}$ از پیشامدهای فضای نمونه ای را صعودی می گوئیم هرگاه:

$$E_1 \subseteq E_2 \subseteq E_n \subseteq E_{n+1} \dots$$

و آن را نزولی می گوئیم هرگاه:

$$E_1 \supseteq E_2 \supseteq E_n \supseteq E_{n+1} \dots$$

برای دنباله ای صعودی از پیشامدها $\{E_n, n \geq 1\}$ منظور از $\lim_{n \rightarrow \infty} E_n$ پیشامد وقوع حداقل یکی از پیشامدهای $E_i, 1 \leq i < \infty$ است. بنابراین

$$\lim_{n \rightarrow \infty} E_n = \bigcup_{n=1}^{\infty} E_n$$

برای دنباله‌ای نزولی از پیشامدها $\{E_n, n \geq 1\}$ منظور از $\lim_{n \rightarrow \infty} E_n$ پیشامد وقوع هر E_i ، $1 \leq i < \infty$ است. بنابراین

$$\lim_{n \rightarrow \infty} E_n = \bigcap_{n=1}^{\infty} E_n$$

قضیه ب.۲.۲. (پیوستگی تابع احتمال) برای هر دنباله صعودی یا نزولی از پیشامدها، یعنی $\{E_n, n \geq 1\}$:

$$\lim_{n \rightarrow \infty} P(E_n) = P(\lim_{n \rightarrow \infty} E_n)$$

تعریف ب.۳.۲. فرض کنید $\{F_n\}$ دنباله‌ای از تابع‌های توزیع است. اگر تابع توزیع F وجود داشته باشد به طوری که در هر نقطه x از نقاط پیوستگی F ،

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

آن‌گاه گوییم F_n در توزیع F همگراست.

فرض کنید $\{X_n\}$ دنباله‌ای متغیرهای تصادفی مربوط به دنباله‌ای $\{F_n\}$ باشد. گوییم X_n در توزیع به X همگراست، اگر متغیر تصادفی X با تابع توزیع F وجود داشته باشد به طوری که F_n در توزیع به F همگرا باشد. این همگرایی را با نماد $X_n \xrightarrow{D} X$ نمایش می‌دهیم و F را تابع توزیع حدی $\{X_n\}$ می‌نامیم. اگر c یک عدد ثابت باشد، منظور از $X_n \xrightarrow{D} c$ این است که X_n در توزیع به یک متغیر تصادفی ثابت در نقطه‌ی c همگرا است.

تعریف ب.۴.۲. فرض کنید $\{X_n\}$ دنباله‌ای از متغیرهای تصادفی باشد. گوییم دنباله‌ی $\{X_n\}$ در احتمال همگرا به متغیر تصادفی X است اگر برای هر $\epsilon > 0$ ،

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

این نوع همگرایی را با نماد $X_n \xrightarrow{P} X$ نمایش می‌دهیم.

اگر c یک عدد ثابت باشد منظور از $X_n \xrightarrow{D} c$ این است که X_n در احتمال همگرا به متغیر تصادفی ثابت در نقطه‌ی c است.

قضیه ب.۵.۲. فرض کنید $\{X_n\}$ دنباله‌ای از متغیرهای تصادفی باشد.

الف) اگر $X_n \xrightarrow{P} X$ و $X_n \xrightarrow{P} 0$ اگر و تنها اگر $X_n - X \xrightarrow{P} 0$

ب) اگر $X_n \xrightarrow{P} X$ و $X_n \xrightarrow{P} Y$ اگر و تنها اگر $P(X = Y) = 1$.

قضیه ب.۶.۲. فرض کنید $\{X_n\}$ و $\{Y_n\}$ دنباله‌هایی از متغیرهای تصادفی باشند.

الف) اگر $X_n \xrightarrow{P} X$ و $Y_n \xrightarrow{P} Y$ ، آن‌گاه $X_n \pm Y_n \xrightarrow{P} X \pm Y$

ب) اگر $X_n \xrightarrow{P} X$ و $Y_n \xrightarrow{P} Y$ ، آن‌گاه $X_n Y_n \xrightarrow{P} XY$

ج) اگر $X_n \xrightarrow{P} X$ و $Y_n \xrightarrow{P} Y$ ، آن‌گاه $X_n \pm Y_n \xrightarrow{P} X \pm Y$ و $P(Y_n \neq 0) = P(Y \neq 0) = 1$ ، آن‌گاه

$\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y}$ ، آن‌گاه $Y_n \xrightarrow{P} Y$ و $X_n \xrightarrow{P} X$

قضیه ب.۷.۲. فرض کنید $\{X_n\}$ دنباله‌ای از متغیرهای تصادفی باشند. اگر $X_n \xrightarrow{P} X$ آن‌گاه

$$X_n \xrightarrow{D} X$$

قضیه ب.۸.۲. فرض کنید $\{X_n\}$ دنباله‌ای از متغیرهای تصادفی باشند. $X_n \xrightarrow{P} c$ اگر و تنها اگر

$$X_n \xrightarrow{D} c$$

قضیه‌ی زیر که به قضیه اسلاتسکی معروف است می‌تواند در مواردی که برای تعیین تابع توزیع حدی به تابع توزیع دسترسی نداریم، مورد استفاده قرار بگیرد.

قضیه ب.۹.۲. (قضیه اسلاتسکی)

فرض کنید $\{X_n\}$ و $\{Y_n\}$ دنباله‌هایی از متغیرهای تصادفی باشند. اگر $X_n \xrightarrow{D} X$ و $Y_n \xrightarrow{P} c$ ، آنگاه

$$X_n \pm Y_n \xrightarrow{D} X \pm c \quad (\text{الف})$$

$$X_n Y_n \xrightarrow{D} Xc \quad (\text{ب})$$

$$X_n Y_n^{-1} \xrightarrow{D} Xc \quad (\text{ج}) \text{ و } P(Y_n \neq 0) = 1$$

قضیه ب.۱۰.۲. (قانون ضعیف اعداد بزرگ)

فرض کنید $X_1, X_2, \dots$ متغیرهای تصادفی مستقل از هم با توزیع یکسان، میانگین μ و واریانس $\sigma^2 (< \infty)$ هستند. اگر $S_n = \sum_{i=1}^n X_i$ ، آنگاه

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

ب.۳. همگرایی

تعریف ب.۱.۳. فرض کنید $\{a_n\}_{n=1}^{\infty}$ دنباله‌ای از اعداد باشد. گوییم دنباله a_n نسبت به دنباله n^{-r} دارای مرتبه کوچکتری است، هرگاه $n^r a_n$ به صفر همگرا باشد و به اختصار می‌نویسیم $a_n = o(n^{-r})$ به عبارتی $\lim_{n \rightarrow \infty} n^r a_n = 0$.

تعریف ب.۲.۳. گوییم دنباله a_n هم مرتبه با دنباله n^{-r} است، هرگاه $n^r a_n$ کراندار باشد و به اختصار می‌نویسیم $a_n = O(n^{-r})$. به عبارت دیگر، برای هر $\delta > 0$ یک مقدار ثابت $M_\delta > 0$ و یک n_δ وجود دارد به طوری که برای هر $n > n_\delta$ داشته باشیم $|n^r a_n| < M_\delta$.

تعریف ب.۳.۳. فرض کنید $\{a_n\}_{n=1}^{\infty}$ یک دنباله از متغیرهای تصادفی باشد. آنگاه می‌توان نوشت $a_n = o_p(n^{-r})$ ، هرگاه $n^r a_n$ در احتمال به صفر همگرا باشد.

تعریف ب.۴.۳. فرض کنید $\{a_n\}_{n=1}^{\infty}$ یک دنباله از متغیرهای تصادفی باشد. آنگاه می‌توان نوشت $a_n = O_p(n^{-r})$ ، هرگاه $n^r a_n$ در احتمال کراندار باشند یعنی، برای هر $\delta > 0$ یک مقدار ثابت $M_\delta > 0$ و یک n_δ وجود داشته باشد به طوری که برای هر $n > n_\delta$ داشته باشیم $P(|n^r a_n| > M_\delta) < \delta$.

لم ب.۵.۳. اگر دنباله تصادفی $n^r a_n$ در توزیع به متغیر تصادفی X با میانگین و واریانس متناهی همگرا باشد، آنگاه می‌توان نوشت $a_n = O_p(n^r)$.

برهان. طبق نامساوی مارکف می‌توان نوشت

$$P(|n^r a_n| > M_\delta) < P(|X| > M_\delta) < \frac{E(X^2)}{M_\delta^2} = \frac{\sigma_x^2 + \mu_x^2}{M_\delta^2}$$

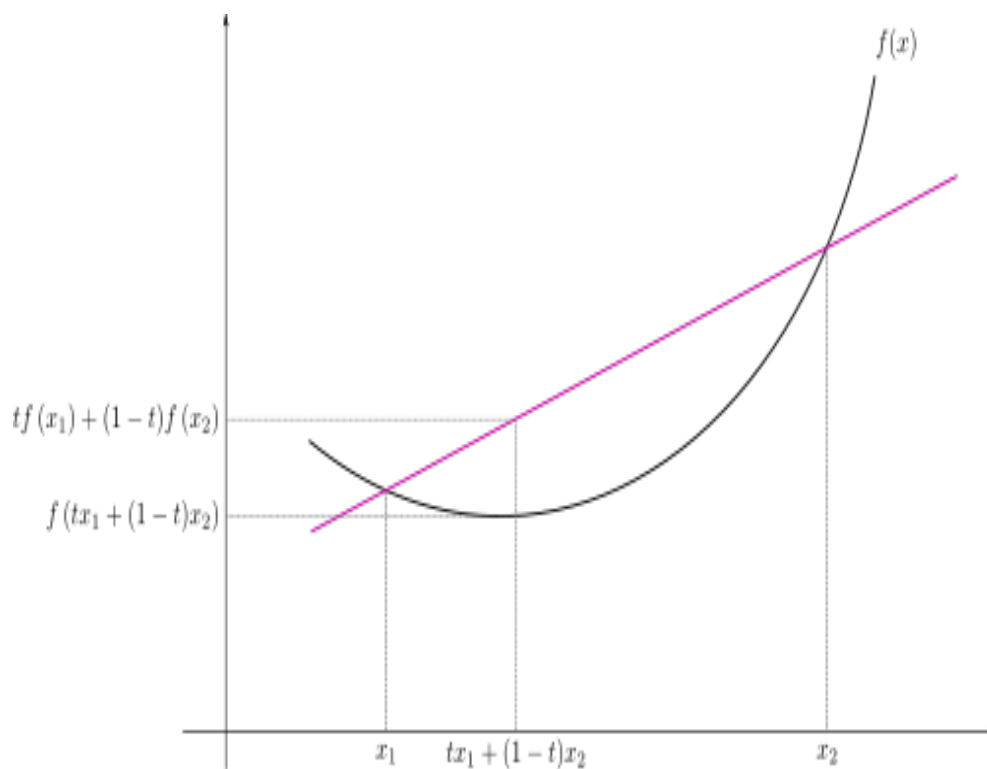
بنابراین کافی است در تعریف ب.۴.۳، δ برابر $\frac{\sigma_x^2 + \mu_x^2}{M_\delta^2}$ اختیار شود. □

ب.۴ برخی اصول آنالیز ریاضی

تعریف ب.۴.۱. منظور از قطعه (a, b) یعنی مجموعه تمام x های حقیقی که $a < x < b$. منظور از بازه $[a, b]$ یعنی مجموعه تمام x های حقیقی که $a \leq x \leq b$.

تعریف ب.۴.۲. مجموعه $E \subset \mathbb{R}^k$ را محدب خوانیم هرگاه، هر وقت $x \in E$ و $y \in E$ و $0 < \lambda < 1$ داشته باشیم $\lambda x + (1 - \lambda)y \in E$.

تعریف ب.۴.۳. تابع محدب، تابعی پیوسته است که اگر دو نقطه دلخواه بر روی این تابع در نظر بگیریم، پاره خط پیونددهنده این دو نقطه همواره بالای این نمودار جای می گیرد.



شکل ب.۱: نمودار تابع محدب

تعریف ب.۴.۴. منظور از یک پوشش باز مجموعه E در فضای متری X یعنی گردایه ای از زیر مجموعه های باز X مانند $\{G_\alpha\}$ که $E \subset \bigcup_\alpha G_\alpha$.

تعریف ب.۴.۵. زیر مجموعه K از فضای متری X را فشرده نامند هرگاه هر پوشش باز K حاوی زیر پوششی متناهی باشد.

قضیه ب.۶.۴. هرگاه مجموعه E در $\mathbb{R}^K$ یکی از سه خاصیت زیر را داشته باشد، آنگاه از دو خاصیت دیگر نیز بهره‌مند است:

(۱) E بسته و کراندار است.

(۲) E فشرده است.

(۳) هر زیر مجموعه نامتناهی E یک نقطه حدی در E دارد.

تعریف ب.۷.۴. فرض کنیم $\{f_n\}$ $n = 1, 2, \dots$ دنباله‌ای از توابع باشد که بر مجموعه E تعریف شده‌اند. و دنباله $\{f_n(x)\}$ از اعداد را به ازای هر $x \in E$ همگرا می‌گیریم. در این صورت، می‌توانیم تابع f را با

$$f(x) = \lim_{n \rightarrow \infty} f_n(x) \quad (x \in E) \quad (\text{ب.۱.})$$

تعریف می‌کنیم. $\{f_n(x)\}$ بر E نقطه به نقطه همگراست، اگر (ب.۱) برقرار باشد. و با نماد $f_n(x) \xrightarrow{P.W} f(x)$ نشان می‌دهیم.

تعریف ب.۸.۴. گوییم دنباله‌ای از توابع مانند $\{f_n\}$ و $n = 1, 2, \dots$ بر E به‌طور یکنواخت به تابع f همگراست هرگاه به ازای هر $\epsilon > 0$ عدد صحیحی مانند N باشد به طوری که $n \geq N$ نامساوی

$$|f_n(x) - f(x)| \leq \epsilon$$

را به ازای هر $x \in E$ ایجاب کند. و با نماد $f_n(x) \xrightarrow{U} f(x)$ نشان می‌دهیم.

قضیه ب.۹.۴. دنباله‌ای از توابع مانند $\{f_n\}$ تعریف شده بر E به‌طور یکنواخت بر E همگراست اگر و فقط اگر به ازای هر $\epsilon > 0$ عدد صحیحی مانند N باشد به طوری که $m \geq N$ و $n \geq N$ و $x \in E$ نامساوی

$$|f_n(x) - f_m(x)| \leq \epsilon$$

را ایجاب کند.

قضیه ب.۱۰.۴. هر دنباله به‌طور یکنواخت همگرا، نقطه به نقطه همگراست.

ب.۵. برخی اصول آنالیز عددی

تعریف ب.۱.۵. سری توانی: سری

$$a_0 + a_1x + a_2x^2 + \dots = \sum_{n=0}^{\infty} a_n x^n$$

یک سری توانی از x است، که در آن $a_0, a_1, a_2, \dots$ مقادیری ثابت هستند.

توجه ب.۲.۵. در حالت کلی، سری توانی برای $|x| < R$ همگرا و برای $|x| > R$ واگراست که R مقداری ثابت به نام شعاع همگرایی است. برای $|x| = R$ سری می‌تواند همگرا یا واگرا باشد.

تعریف ب.۳.۵. بسط تیلور: بسط تیلور تابع $f(x)$ در همسایگی x_0 به صورت سری توانی زیر است

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)(x - x_0)^2}{2} + \dots = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)(x - x_0)^n}{n!}$$

ب.۱.۵ روش دلتا

واریانس برآوردگر $\hat{\beta}$ را می‌توان طبق روش دلتا محاسبه نمود.

لم ب.۴.۵. بسط تیلور تابع $f(\hat{\beta})$ در همسایگی β_0 را در نظر بگیرید.

$$\text{Var}(f(\hat{\beta})) = f'(\beta_0) \text{Var}(\hat{\beta}) f'(\beta_0)$$

برهان.

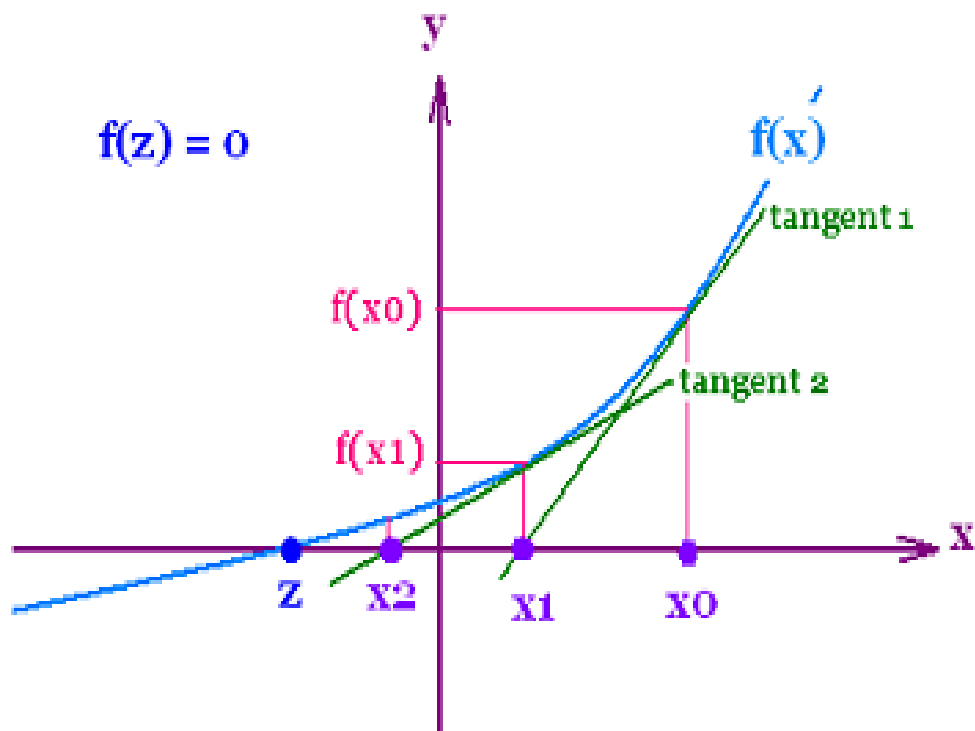
$$\begin{aligned} \text{Var}(f(\hat{\beta})) &= \text{Var}\left(f(\beta_0) + f'(\beta_0)(\hat{\beta} - \beta_0)\right) \\ &= \text{Var}\left(f(\beta_0) + f'(\beta_0)\hat{\beta} - f'(\beta_0)\beta_0\right) \\ &= \text{Var}\left(f'(\beta_0)\hat{\beta}\right) \\ &= f'(\beta_0) \text{Var}(\hat{\beta}) f'(\beta_0) \end{aligned}$$

□

برهان کامل است.

ب.۲.۵ روش نیوتن رافسون

روش نیوتن روش سریع و قدرتمندی برای پیدا کردن جواب دقیق یک معادله است. مبنای این روش استفاده از خط مماس بر منحنی به عنوان تقریبی از منحنی برای یافتن حدس بهبود یافته بعدی است. برای محاسبه شیب منحنی از مشتق منحنی در آن نقطه استفاده می‌شود. این روش تکراری به مقدار حدس اولیه بسیار وابسته است و ممکن است همیشه به جواب نرسد اما روش بسیار توانمند است که سرعت همگرایی خیلی خوبی به سمت جواب دارد و با ترکیب آن با روش‌هایی که رسیدن به جواب را تضمین می‌کنند می‌توان به روش مناسب حل معادله دست یافت.



شکل ب.۲: روش نیوتن رافسون

در هر مرحله از حل یک حدس اولیه برای جواب داریم x_n و حالا برای بهبود این حدس یعنی بدست آوردن x_{n+1} و نزدیک شدن به جواب اصلی معادله می توان از خط مماس بر منحنی در نقطه x_n استفاده نمود. شیب منحنی را با استفاده از مشتق می توان محاسبه نمود. معادله خط مماس بر منحنی در نقطه x_n به شکل زیر است:

$$y = f'(x_n)(x - x_n) + f(x_n)$$

جایی که این خط محور افقی را قطع کند، حدس بعدی یعنی x_{n+1} است. در این نقطه $y = 0$ خواهد بود بنابراین:

$$0 = f'(x_n)(x - x_n) + f(x_n)$$

حل این معادله برای یافتن x_{n+1} به شکل زیر خواهد بود

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

پیوست پ

۱. پ عامل تورم واریانس

یکی از شیوه‌های تشخیص وجود همخطی که کاربرد زیادی دارد، عامل تورم واریانس می‌باشد. این عامل نشان می‌دهد که واریانس ضرایب تا چه حد نسبت به حالتی که متغیرهای برآورد شده همبستگی خطی ندارند، متورم شده است. عامل تورم واریانس از رابطه زیر بدست می‌آید

$$VIF_i = \frac{1}{1 - R_i^2} \quad , \quad i = 1, 2, \dots, p - 1$$

که در آن R_i^2 ، ضریب تعیین مدلی است که در آن متغیر مستقل X_i روی سایر متغیرهای مستقل رگرس شده است.

در حالتی که $R_i^2 = 0$ یا به عبارت دیگر X_i به صورت خطی رابطه‌ای با سایر متغیرهای مستقل در مدل رگرسیونی نداشته باشد، VIF_i برابر یک خواهد بود. هنگامی که $R_i^2 \neq 0$ ، VIF_i بزرگتر از یک خواهد بود که نشان می‌دهد، واریانس X_i به دلیل وجود همبستگی بین متغیرهای پیش‌بین چقدر متورم شده است و مقدار آن افزایش پیدا کرده است.

هنگامی که یک متغیر پیش‌بین، وابستگی خطی کامل با سایر متغیرهای پیش‌بین مدل داشته باشد، ضریب تعیین با آن برابر یک خواهد بود و در نتیجه عامل تورم واریانس به سمت بینهایت میل خواهد کرد و از آن می‌توان نتیجه گرفت که واریانس X_i نیز بینهایت خواهد بود.

مقدار تورم واریانس برای متغیرهای پیش‌بین، اغلب به عنوان شاخصی برای اندازه‌گیری شدت همخطی بودن در مدل استفاده می‌شود. اگر بیشترین مقدار عامل تورم واریانس در مدل بزرگتر از ۱۰ باشد، معمولاً به نشانه این مسئله در نظر گرفته می‌شود که همخطی موجود در مدل، اثر نامطلوبی بر برآوردهای روش کمترین توان‌های دوم دارد.

پیوست

ت. ۱ عدد شرطی

در آنالیز عددی، عدد شرطی یک تابع، بدترین وضعیتی که مقدار آن تابع نسبت به تغییرات کوچک می‌تواند تغییر کند را اندازه‌گیری می‌کند. مسئله‌ای با عدد شرطی پایین را خوب شرطی شده و معادله‌ای با عدد شرطی بالا را ناقص شرطی شده می‌گویند. به عنوان مثال، عدد شرطی در ارتباط با معادله $AX = B$ ، حدودی را برای نادرست برآورد شدن X از راه‌حل‌های تقریبی ارائه می‌دهد. دقت کنید که عدد شرطی نسبت به تغییرات در B نیز حساس است، یعنی اگر عدد شرطی زیاد باشد، حتی خطای کوچکی در B ممکن است باعث خطای بزرگ در برآورد X شود. از سوی دیگر، اگر عدد شرطی کوچک باشد، خطا در X خیلی بزرگتر از خطا در B نخواهد بود.

عدد شرطی به صورت نسبت خطای نسبی در C به خطای نسبی در B تعریف شده است. فرض کنید که ϵ خطای نسبی در B و A ماتریس مربعی باشد، بنابراین $A^{-1}\epsilon$ خطا در معادله $A^{-1}B$ است. نسبت خطای نسبی در معادله به خطای نسبی در B به صورت زیر است

$$\frac{\|A^{-1}\epsilon\|/\|A^{-1}B\|}{\|\epsilon\|/\|B\|} \quad (1.ت)$$

که به سادگی به صورت زیر تبدیل می‌شود

$$(\|A^{-1}\epsilon\|/\|\epsilon\|) \cdot (\|B\|/\|A^{-1}B\|) \quad (2.ت)$$

بزرگترین مقدار (برای B و ϵ غیر صفر) حاصل ضرب دو نرم عملگر به صورت زیر است

$$k(A) = \|A\| \cdot \|A^{-1}\|$$

تعریف مشابهی هست که برای نرم‌های سازگار بکار می‌رود. غالباً این عدد در یک عبارت جبری خطی عددی به وجود می‌آیند که عدد شرطی ماتریس A نامیده می‌شود. البته این تعریف وابسته با انتخاب نرم است.

۱. اگر $\|\cdot\|$ نرم σ^2 باشد، بنابراین

$$k(A) = \frac{\sigma_{Max}(A)}{\sigma_{Min}(A)} \quad (3.ت)$$

که σ_{Max} و σ_{Min} به ترتیب، بزرگترین و کوچکترین مقادیر منفرد A هستند.
۲. اگر A نرمال باشد، آنگاه

$$k(A) = \left| \frac{\lambda_{Max}(A)}{\lambda_{Min}(A)} \right| \quad (4.ت)$$

که در آن $\lambda_{Max}(A)$ و $\lambda_{Min}(A)$ به ترتیب، بزرگترین و کوچکترین مقایر ویژه ماتریس A هستند.
۳. اگر A ماتریسی واحد باشد، آن‌گاه

$$k(A) = 1 \quad (5.ت)$$

۴. اگر $\|\cdot\|$ نرم بینهایت باشد و A ماتریس منفرد و پایین مثلثی باشد ($\forall i, a_{ii} \neq 0$)، بنابراین

$$k(A) = \frac{Max_i(|a_{ii}|)}{Min_i(|a_{ii}|)} \quad (6.ت)$$

پیوست

در این پیوست نشان می‌دهیم

$$\frac{\partial RSS}{\partial \beta_j} = 2\mathbf{x}_j^T \mathbf{x}_j \beta_j + \sum_{i \neq j} 2\mathbf{x}_j^T \mathbf{x}_i \beta_i - 2\mathbf{x}_j^T \mathbf{Y}$$

می‌باشد. می‌دانیم

$$\begin{aligned} RSS &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^T \mathbf{Y} - 2\mathbf{Y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \end{aligned} \quad (\text{ث.۱})$$

رابطه‌ی (ث.۱) را به صورت زیر بازنویسی می‌کنیم

$$\begin{aligned} RSS &= \begin{bmatrix} Y_1 & \dots & Y_n \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} - 2 \begin{bmatrix} Y_1 & \dots & Y_n \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{1p} \\ x_{21} & \dots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \\ &+ \begin{bmatrix} \beta_1 & \dots & \beta_p \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{n1} \\ x_{12} & \dots & x_{n2} \\ \vdots & \ddots & \vdots \\ x_{1p} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{1p} \\ x_{21} & \dots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \end{aligned}$$

که مقدار $\mathbf{Y}^T \mathbf{X} \boldsymbol{\beta}$ برابر است با

$$\begin{aligned} \mathbf{Y}^T \mathbf{X} \boldsymbol{\beta} &= \begin{bmatrix} Y_1 x_{11} + \cdots + Y_n x_{n1} & Y_1 x_{1j} + \cdots + Y_n x_{nj} & Y_1 x_{1p} + \cdots + Y_n x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \\ &= (Y_1 x_{11} + \cdots + Y_n x_{n1}) \beta_1 + \cdots + (Y_1 x_{1j} + \cdots + Y_n x_{nj}) \beta_j \\ &\quad + \cdots + (Y_1 x_{1p} + \cdots + Y_n x_{np}) \beta_p \end{aligned}$$

بنابراین

$$\frac{\partial \mathbf{Y}^T \mathbf{X} \boldsymbol{\beta}}{\partial \beta_j} = Y_1 x_{1j} + \cdots + Y_n x_{nj} = \mathbf{x}_j^T \mathbf{Y} \quad (\text{ث. ۲})$$

به طور مشابه

$$\begin{aligned} \boldsymbol{\beta}^T \mathbf{X}^T &= \begin{bmatrix} \beta_1 & \cdots & \beta_p \end{bmatrix} \begin{bmatrix} x_{11} & \cdots & x_{n1} \\ x_{12} & \cdots & x_{n2} \\ \vdots & \ddots & \vdots \\ x_{1p} & \cdots & x_{np} \end{bmatrix} \\ &= \begin{bmatrix} \beta_1 x_{11} + \cdots + \beta_p x_{1p} & \cdots & \beta_1 x_{n1} + \cdots + \beta_p x_{np} \end{bmatrix} \quad (\text{ث. ۳}) \end{aligned}$$

و

$$\begin{aligned} \mathbf{X} \boldsymbol{\beta} &= \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \\ &= \begin{bmatrix} x_{11} \beta_1 + \cdots + x_{1p} \beta_p & \cdots & x_{n1} \beta_1 + \cdots + x_{np} \beta_p \end{bmatrix} \quad (\text{ث. ۴}) \end{aligned}$$

از روابط (ث. ۳) و (ث. ۴) داریم

$$\begin{aligned} \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} &= (\beta_1 x_{11} + \cdots + \beta_p x_{1p})^2 \\ &\quad + \cdots + (\beta_1 x_{n1} + \cdots + \beta_p x_{np})^2 \end{aligned}$$

بنابراین

$$\begin{aligned} \frac{\partial \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}}{\partial \beta_j} &= 2 x_{1j} (\beta_1 x_{11} + \cdots + \beta_p x_{1p}) + \cdots + 2 x_{nj} (\beta_1 x_{n1} + \cdots + \beta_p x_{np}) \\ &= 2 \mathbf{x}_j^T \mathbf{x}_j \beta_j + \sum_{i \neq j} 2 \mathbf{x}_j^T \mathbf{x}_i \beta_i \quad (\text{ث. ۵}) \end{aligned}$$

در نتیجه

$$\frac{\partial RSS}{\partial \beta_j} = 2 \mathbf{x}_j^T \mathbf{x}_j \beta_j + \sum_{i \neq j} 2 \mathbf{x}_j^T \mathbf{x}_i \beta_i - 2 \mathbf{x}_j^T \mathbf{Y}$$

پیوست ج

ج.۱ برنامه‌های کامپیوتری با نرم افزار R

در این پیوست تمام دستورهای مورد نیاز برای رسم شکل‌ها و جداول با نرم‌افزار R-۳.۳.۲ آمده است. دستور زیر مربوط به فراخوانی بسته‌های مورد نیاز برای انجام برخی دستورات عمل‌های ماتریسی است که برای تمام برنامه‌ها به طور یکسان تکرار می‌شود.

```
>install.packages("lasso2")
>install.packages("glmnet")
>install.packages("MASS")
>library(lasso2)
>library(glmnet)
>library(MASS)
>require(stats)
```

ج.۲ برنامه داده‌های سرطان پروستات

دستورات لازم برای محاسبه ضرایب لاسو و ریج در بسته glmnet

```
> library(glmnet)
#----- data
```

```
> data(Prostate)
> Prostate
> x<-(apply(Prostate[,-9],2,as.numeric))
> y<-(Prostate[,9])
#----- lasso
> lassofit<-glmnet(x,y,alpha=1)
> par(mfrow = c(1, 2))
> plot(lassofit, xvar="lambda")
> plot(lassofit, xvar="norm")
> par(mfrow = c(1, 1))
> coef(lassofit)
#----- Cross Validation
> cvlassofit<-cv.glmnet(x,y,alpha=1)
> print(cvlassofit)
> plot(cvlassofit)
# to get the coefficients for the "optimal" lambda
> coef(cvlassofit, s = "lambda.min")
# to get the coefficients for the one SE lambda
> coef(cvlassofit, s = "lambda.1se")

#----- ridge

> ridgefit<-glmnet(x,y,alpha=0)
> par(mfrow = c(1, 2))
> plot(ridgefit, xvar="lambda")
> plot(ridgefit, xvar="norm")
> par(mfrow = c(1, 1))
> coef(ridgefit)
> cvlassofit<-cv.glmnet(x,y,alpha=0)
> print(cvlassofit)
> plot(cvlassofit)
# to get the coefficients for the "optimal" lambda
> coef(cvlassofit, s = "lambda.min")
# to get the coefficients for the one SE lambda
```

```
> coef(cvlassofit, s = "lambda.1se")
```

دستورا لازم برای محاسبه ضرایب لاسو با بسته lasso2

```
> l1c.P <- l1ce(lpsa ~ ., Prostate, bound=(1:30)/30)
>l1c.P
>coef(l1c.P)
>summary(l1c.P)
>resid(l1c.P)
>fitted(l1c.P)
>gcv(l1c.P)
>gcv=as.matrix(gcv(l1c.P),ncol=4)
>gcv
>plot(gcv[,1],gcv[,4],type="l")
>abline(v=.44,col="red")
>beta_l1ce=l1ce(lpsa~.,Prostate,bound=0.44)
>beta_l1ce=l1ce(lpsa~.,Prostate,bound=0.44)$coef
>beta_l1ce=beta_l1ce[-1]
```

ج.۳ برنامه مربوط به سناریو ۱.۱.۱

برای تولید داده‌ها از دستور زیر استفاده می‌کنیم.

```
>p=4
>n=5
>gamma=0.7
>generator=function(p,n,gamma){
+   z=matrix(c(rnorm(p*n,mean=0,sd=1)),ncol=p)
+   x=matrix(c(rep(0,n*p)),ncol=p)
+   y<-matrix(c(rep(0,p)),ncol=1)
+   for(i in 1:n){
+     for(j in 1:p){
+       x[i,j]=(1-gamma^2)^(1/2)*z[i,j]+gamma*z[i,p]
+     }
+   }
+   ww=t(x)%*%x
```

```

+ beta=eigen(wv)$vectors[,1]
+ sigma=0.5
+ error<-matrix(c(rnorm(n,mean=0,sd=sigma)),ncol=1)
+ y=x%%beta+error
+ return(list(x=x,y=y,beta=beta))
+ }

```

```

>data=generator(p,n,gamma)
>x=data$x
>y=data$y
>beta=data$beta

```

دستور زیر عدد شرطی را برای بررسی همخطی محاسبه می کند.

```

> CN=max((eigen(t(x)%*%x))$values)/min((eigen(t(x)%*%x))$values)
> CN

```

برای پیدا کردن ضرایب *OLS* داریم.

```

> betahatols=coef(lm(y~x))
> b0=betahatols=as.vector(betahatols)
> beta_ols=betahatols[-1]

```

دستور زیر داده ها را به دو قسمت آزمون (test) و آموزش (train) تقسیم می کند.

```

>splitdf <- function(data1) {
  #set.seed(0)
+   index <- 1:nrow(data1)
+   trainindex <- sample(index,trunc(length(index)*(2/3)))
+   trainset <- data1[trainindex,]
+   testset <- data1[-trainindex,]
+   return(list(trainset=trainset,testset=testset,trainindex=trainindex))
+ }
> data1=matrix(1:100,ncol=10)
> data1
> splitdf(data1)

```

توجه نمایید برای انتخاب بهینه پارامتر انقباضی باید با استفاده از داده های آموزش مدل را برازش و با داده های آزمون *GCV* را محاسبه نمود.

ج.۴ محاسبه ضرایب ریدج

برنامه‌ی زیر ضرایب ریدج را با روش نیوتن رافسون بدست می‌آورد. دقت کنید $y = \text{train}\$y$ و $x = \text{train}\$x$ می‌باشد همچنین $ytest = \text{testset}\$y$ و $xtest = \text{testset}\$x$ است.

```
#----- ridge

> lambda=c(0.1,0.5,1,10,50,100,500,1000)
> gcv_ridge_final=rep(0,length(lambda))
> l=1
> while(l<length(lambda)+1){
+   ftn <- function(x,y,b0) {           #ftn mohaseb mikonad f(x)
+     i=1
+     j=1
+     fx=rep(0,p)
+     while(j<p+1){
+       while(i<p+1){
+         fx[j]=fx[j]+2*t(x[,j])%*%x[,i]%*%b0[i]
+         i=i+1
+       }
+       fx[j]=fx[j]-2*t(x[,j])%*%y+2*lambda[l]*b0[j]
+       j=j+1
+       i=1
+     }
+     return(fx)
+   }

+   fdn <- function(x,y,b0) {           #fdn mohaseb mikonad f'(x)
+     i=1
+     j=1
+     fdx=rep(0,p)
+     i=1;j=1
+     while(j<p+1){
+       fdx[j]=2*t(x[,j])%*%x[,j]+2*lambda[l]
```

```

+     j=j+1
+   }
+   return(fdx)
+ }

+ beta_ridg=rep(0,p)
+ j=1
+ while(j<p+1){
+   newtonraphson <- function(ftn,fdx,b0, tol = 1e-3, max.iter = 100) {
+     # Newton_Raphson algorithm for solving ftn(x)[1] == 0
+     # we assume that ftn is a function of a single variable that returns
+     # the function value and the first derivative as a vector of length 2
+     # b0 is the initial guess at the root and the algorithm terminates when the

+     function value is within distance
+     # tol of 0, or the number of iterations exceeds max.iter
+     c <- b0[j]
+     fx <- ftn(x,y,b0)
+     fdx=fdn(x,y,b0)
+     iter <- 0
+     # continue iterating until stopping conditions are met
+     while ((abs(fx[j]) > tol) && (iter < max.iter)) {
+       c <- c - fx[j]/fdx[j]
+       b0[j]=c
+       fx <- ftn(x,y,b0)
+       iter <- iter + 1
+       cat("At iteration", iter, "value of x is:", x, "\n")
+     }
+     # output depends on success of algorithm
+     if (abs(fx[j]) > tol) {
+       cat("Algorithm failed to converge\n")
+       return(NULL)
+     } else {

```

```

+         cat("Algorithm converged\n")
+         return(c)
+     }
+ }
+ d=newtonraphson(ftn,fdn, b0, 1e-06,max.iter = 100)
+ beta_ridg[j]=d
+ j=j+1
+ }
+ gamma=2
+ v=gamma*(abs(beta_ridg)^(gamma-2))/2
+ w=diag(v)
+ sw=solve(w)
+ p_lambda=sum(x%*(solve(t(x)%*x+lambda[l]*sw))%*t(x))
+ yhat=xtest%*beta_ridg
+ RSS=sum((ytest-yhat)^2)
+ #RSS
+ gcv_ridg=RSS/(n*((1-(p_lambda/n))^2))
+ gcv_ridg_final[l]=gcv_ridg
+ l=l+1
+ }
> gcv_ridg_final
> which(gcv_ridg_final==min(gcv_ridg_final))
> optimal_lambda_ridg=lambda[which(gcv_ridg_final==min(gcv_ridg_final))]
> optimal_gcv_ridg=min(gcv_ridg_final)
> optimal_gcv_ridg

#----- optimal ridge

> lambda=optimal_lambda_ridg

> ftn <- function(x,y,b0) {          #ftn mohaseb mikonad f(x)
+   i=1
+   j=1
+   fx=rep(0,p)
+   while(j<p+1){

```

```

+   while(i<p+1){
+       fx[j]=fx[j]+2*t(x[,j])%*%x[,i]%*%b0[i]
+       i=i+1
+   }
+   fx[j]=fx[j]-2*t(x[,j])%*%y+2*lambda*b0[j]
+   j=j+1
+   i=1
+ }
+ return(fx)
+ }

```

```

#ftn(x,y,b0)

```

```

> fdn <- function(x,y,b0) {           #fdn mohaseb mikonad f'(x)
+   i=1
+   j=1
+   fdx=rep(0,p)
+   i=1;j=1
+   while(j<p+1){
+       fdx[j]=2*t(x[,j])%*%x[,j]+2*lambda
+       j=j+1
+   }
+   return(fdx)
+ }

```

```

> beta_ridg=rep(0,p)
> j=1
> while(j<p+1){
+   newtonraphson <- function(ftn,fdx,b0, tol = 1e-3, max.iter = 100) {
+       # Newton_Raphson algorithm for solving ftn(x)[1] == 0
+       # we assume that ftn is a function of a single variable that returns
+       # the function value and the first derivative as a vector of length 2
+       # b0 is the initial guess at the root and the algorithm terminates when the

```

```
+ function value is within distance
+   # tol of 0, or the number of iterations exceeds max.iter
+   c <- b0[j]
+   fx <- ftn(x,y,b0)
+   fdx=fdn(x,y,b0)
+   iter <- 0
+   # continue iterating until stopping conditions are met
+   while ((abs(fx[j]) > tol) && (iter < max.iter)) {
+     c <- c - fx[j]/fdx[j]
+     b0[j]=c
+     fx <- ftn(x,y,b0)
+     iter <- iter + 1
+     cat("At iteration", iter, "value of x is:", x, "\n")
+   }
+   # output depends on success of algorithm
+   if (abs(fx[j]) > tol) {
+     cat("Algorithm failed to converge\n")
+     return(NULL)
+   } else {
+     cat("Algorithm converged\n")
+     return(c)
+   }
+ }
+
+ d=newtonraphson(ftn,fdn, b0, 1e-06,max.iter = 100)
+ beta_ridg[j]=d
+ j=j+1
+ }
> beta_ridg
```

برای محاسبه ضرایب ریدج با $\gamma = ۲/۵$ کافی است در ج.۴ دستور ftn و fdn را به صورت زیر بازنویسی می‌کنیم.

```
> ftn <- function(x,y,b0) {           #ftn mohaseb mikonad f(x)
+   i=1
+   j=1
```

```

+   fx=rep(0,p)
+   while(j<p+1){
+     while(i<p+1){
+       fx[j]=fx[j]+2*t(x[,j])%%x[,i]%%b0[i]
+       i=i+1
+     }
+     fx[j]=fx[j]-2*t(x[,j])%%y+(2.5)*lambda*(abs(b0[j]))^(1.5)*sign(b0[j])
+     j=j+1
+     i=1
+   }
+   return(fx)
+ }
# ftn(x,y,b0)

>   fdn <- function(x,y,b0) {           #fdn mohaseb mikonad f'(x)
+     i=1
+     j=1
+     fdx=rep(0,p)
+     i=1;j=1
+     while(j<p+1){
+       fdx[j]=2*t(x[,j])%%x[,j]+2.5*lambda*((sqrt(abs(b0[j]))
+ 0.5*sqrt(b0[j]*sign(b0[j]))))
+       j=j+1
+     }
+     return(fdx)
+   }
  #fdn(x,y,b0)
+   gamma=2.5

```

برای محاسبه ضرایب بریج با $\gamma = 3$ ، ftn و fdn به صورت زیر می باشد.

```

>   sig=sign(b0)
>   ftn <- function(x,y,b0) {           #ftn mohaseb mikonad f(x)
+     i=1
+     j=1

```

```
+   fx=rep(0,p)
+   while(j<p+1){
+     while(i<p+1){
+       fx[j]=fx[j]+2*t(x[,j])%%x[,i]%%b0[i]
+       i=i+1
+     }
+     fx[j]=fx[j]-2*t(x[,j])%%y+3*lambda*((b0[j])^2)*sig[j]
+     j=j+1
+     i=1
+   }
+   return(fx)
+ }
#ftn(x,y,b0)

> fdn <- function(x,y,b0) {           #fdn mohaseb mikonad f'(x)
+   i=1
+   j=1
+   fdx=rep(0,p)
+   i=1;j=1
+   while(j<p+1){
+     fdx[j]=2*t(x[,j])%%x[,j]+6*lambda*b0[j]*sig[j]
+     j=j+1
+   }
+   return(fdx)
+ }
```

برای محاسبه ضرایب رید $\gamma = 4$ ، ftn و fdn به صورت زیر است.

```
> sig=sign(b0)
> ftn <- function(x,y,b0) {           #ftn mohaseb mikonad f(x)
+   i=1
+   j=1
+   fx=rep(0,p)
+   while(j<p+1){
+     while(i<p+1){
```

```

+         fx[j]=fx[j]+2*t(x[,j])%*%x[,i]%*%b0[i]
+         i=i+1
+     }
+     fx[j]=fx[j]-2*t(x[,j])%*%y+4*lambda*((b0[j])^3)
+     j=j+1
+     i=1
+ }
+ return(fx)
+ }
#ftn(x,y,b0)

> fdn <- function(x,y,b0) {          #fdn mohaseb mikonad f'(x)
+     i=1
+     j=1
+     fdx=rep(0,p)
+     i=1;j=1
+     while(j<p+1){
+         fdx[j]=2*t(x[,j])%*%x[,j]+12*lambda*(b0[j])^2
+         j=j+1
+     }
+     return(fdx)
+ }

```

برای بدست آوردن ضرایب لاسو از دستور زیر استفاده می کنیم.

```

> m=1
> while(m<1000){
#----- mohasebeh s_0

> i=1
> j=1
> s=rep(0,p)
> while(j<p+1){
+     while(i<p+1){
+         s[j]=s[j]+2*t(x[,j])%*%x[,i]%*%b0[i]
+         i=i+1

```

```

+   }
+   s[j]=s[j]-2*t(x[,j])%*%y-2*t(x[,j])%*%x[,j]%*%b0[j]
+   j=j+1
+   i=1
+ }
+ s
+ #-----
+ lambda=c(0.1,0.5,1,10,20,50)
+ gcv_lasso_final=rep(0,length(lambda))
+ beta_lasso=rep(0,p)
+ i=1
+ while(i<length(lambda)+1){
+   j=1
+   while(j<p+1){
+     if(s[j]>lambda[i]){
+       beta_lasso[j]=(lambda[i]-s[j])/(2*t(x[,j])%*%x[,j])
+     }else{
+       if(s[j]< -1*lambda[i]){
+         beta_lasso[j]=(-lambda[i]-s[j])/(2*t(x[,j])%*%x[,j])
+       }else{
+         beta_lasso[j]=0
+       }
+     }
+     j=j+1
+   }
+   i=i+1
+ }

+ gamma=1
+ v=gamma*(abs(beta_lasso)^(gamma-2))/2
+ w=diag(v)
+ sw=ginv(w)
+ #-----n0
+ n0=0
+ l=1
+ while (l<p+1) {

```

```

+   if(beta_lasso[l]==0){
+       n0=n0+1
+   }else{
+       n0=n0}
+   l=l+1
+ }
+ p_lambda=sum(x%*(solve(t(x)%*x+lambda[i]*sw))%*t(x))-n0
+ yhat=xtest%*beta_lasso
+ RSS=sum((ytest-yhat)^2)
+ #RSS
+ gcv_lasso=RSS/(n*((1-(p_lambda/n))^2))
+ gcv_lasso_final[i]=gcv_lasso
+ i=i+1
+ }
> gcv_lasso_final
> which(gcv_lasso_final==min(gcv_lasso_final))
> optimal_lambda_lasso=lambda[which(gcv_lasso_final==min(gcv_lasso_final))]
> optimal_gcv_lasso=min(gcv_lasso_final)
#-----optimal   lasso
> lambda= optimal_lambda_lasso
> beta_lasso=rep(0,p)
> j=1
> while(j<p+1){
>   if(s[j]>lambda){
>       beta_lasso[j]=(lambda-s[j])/(2*t(x[,j])%*x[,j])
>   }else{
>       if(s[j]< -1*lambda){
>           beta_lasso[j]=(-lambda-s[j])/(2*t(x[,j])%*x[,j])
>       }else{
>           beta_lasso[j]=0
>       }
>   }
>   j=j+1
> }

```

```
> b0=beta_lasso
> m=m+1
}
```

دستور زیر نقطه‌ی شروع ريج را محاسبه می‌کند.

```
#-----mohasebeh b0
#----- b0 ridge

> lambda=c(0.1,0.5,1,100)
> gcv_ridg_final=rep(0,length(lambda))
> l=1
> while(l<length(lambda)+1){

+   beta_ridg=rep(0,p)
+   j=1
+   while(j<p+1){
+     sw=solve(t(x)%*%x+lambda[l]*diag(p))
+     d=sw%*%t(x)%*%y
+     beta_ridg[j]=d
+     j=j+1
+   }
+   gamma=2
+   v=gamma*(abs(beta_ridg)^(gamma-2))/2
+   w=diag(v)
+   sw=solve(w)
+   p_lambda=sum(x%*(solve(t(x)%*%x+lambda[l]*sw))%*%t(x))
+   yhat=xtest%*%beta_ridg
+   RSS=sum((ytest-yhat)^2)
+   #RSS
+   gcv_ridg=RSS/(n*((1-(p_lambda/n))^2))
+   gcv_ridg_final[l]=gcv_ridg
+   l=l+1
+ }
```

```

> gcv_ridge_final
> which(gcv_ridge_final==min(gcv_ridge_final))
> optimal_lambda_ridge=lambda[which(gcv_ridge_final==min(gcv_ridge_final))]
> optimal_gcv_ridge=min(gcv_ridge_final)
> optimal_gcv_ridge

#-----

> lambda=optimal_lambda_ridge
> sw=solve(t(x)%*%x+lambda*diag(p))
> b0=sw%*%t(x)%*%y

```

دستور زیر ضرایب لاسو و ریج را با استفاده از پکیج glmnet محاسبه می‌کند.

```

> library(glmnet)
> lasso=glmnet(x,y,alpha=1)
> cvlasso=cv.glmnet(x,y,alpha=1)
> betahatlasso=coef(cvlasso, s = "lambda.min")
> betahatlasso=as.vector(betahatlasso)
> betahatlasso=betahatlasso[2:(p+1)]

#-----mahsebeh ridge

> ridge=glmnet(x,y,alpha=0)
> cvridge=cv.glmnet(x,y,alpha=0)
> betahatridge=coef(cvridge, s = "lambda.min")
> betahatridge=as.vector(betahatridge)
> betahatridge=betahatridge[2:(p+1)]

```

دستور زیر MSE برآوردهای ریج و لاسو و کمترین توان‌های دوم در جدول ۱۴.۵ محاسبه می‌کند.

```

#-----

> rm(list=ls(all=T))
> beta=matrix(c(0,0,0,0,0,0,0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,2,
0,0,0,0,0,0,0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,2,2),ncol=1)
> generator=function(p,n,gamma,Ro2){

```

```
+ z=matrix(c(rnorm(p*n,mean=0,sd=1)),ncol=p)
+ x=matrix(c(rep(0,n*p)),ncol=p)
+ y<-matrix(c(rep(0,n)),ncol=1)
+ for(i in 1:n){
+   for(j in 1:p){
+     x[i,j]=((1-gamma^2)^(1/2))*z[i,j]+(gamma*z[i,p])
+   }
+ }
+ error<-matrix(c(rnorm(n,mean=0,sd=sigma)),ncol=1)
+ w=t(x)%*%x
+ #beta=eigen(w)$vectors[,1]
+ y=x%*%beta+error
+ return(list(x=x,y=y,beta=beta))
+ }
>
> library(glmnet)
> m=10
> k=1
> verrorols=vector(length=m)      #vector error ols
> verrorlasso=vector(length=m)
> verrorridge=vector(length=m)
> #-----
> while(k<=m){
+   p=40
+   n=100
+   gamma=0.9
+   #Ro2=100
+   sigma=15
+   beta=matrix(c(0,0,0,0,0,0,0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,
+0,0,0,0,0,0,0,0,0,0,0,2,2,2,2,2,2,2,2,2,2,2),ncol=1)
+   #beta<-matrix(c(7,4,0,0,8,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),ncol=1)
+   #beta<-matrix(c(7,4,0,0,8,0,0,12,0,0,0,0,0,0,9,0,0,0,8,0,0,0,0,6,0,0,0,19
+0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),ncol=1) #destinated coefficient
+   #beta=matrix(c(3,1.5,0,1,0),ncol=1)
```

```
+ #beta=matrix(c(3,3,3,3,3,3,3,3,3,3,3,3,3,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),ncol=1)
+
+ #beta=matrix(c(3,0,0,1,0,7,0,0,3,6),ncol=1)
+
+ #beta
+
+ #length(beta)
+
+ data=generator(p,n,gamma,Ro2)
+
+ x=data$x
+
+ y=data$y
+
+ beta=data$beta
+
+ #sigma=(t(beta)%*%beta)/Ro2
+
+ #x;y;beta
+
+ #-----
+
+ betahatols=coef(lm(y~x))
+
+ betahatols
+
+ betahatols=as.vector(betahatols)
+
+ betahatols=betahatols[2:(p+1)]
+
+ errorols=t(betahatols-beta)%*%(betahatols-beta)
+
+ verrorols[k]=errorols
+
+ #####
+
+ lasso=glmnet(x,y,alpha=1)
+
+ cvlasso=cv.glmnet(x,y,alpha=1)
+
+ betahatlasso=coef(cvlasso, s = "lambda.min")
+
+ betahatlasso=as.vector(betahatlasso)
+
+ betahatlasso=betahatlasso[2:(p+1)]
+
+ errorlasso=t(betahatlasso-beta)%*%(betahatlasso-beta)
+
+ verrorlasso[k]=errorlasso
+
+ #-----
+
+ ridge=glmnet(x,y,alpha=0)
+
+ cvridge=cv.glmnet(x,y,alpha=0)
+
+ betahatrige=coef(cvridge, s = "lambda.min")
+
+ betahatrige=as.vector(betahatrige)
+
+ betahatrige=betahatrige[2:(p+1)]
+
+ errorridge=t(betahatrige-beta)%*%(betahatrige-beta)
+
+ verrorridge[k]=errorridge
```

```
+
+   k=k+1
+ }
>
> #-----
> mseols=mean(verrorols)
> mselasso=mean(verrorlasso)
> mseridge=mean(verrorridge)
> betahatols=as.matrix(betahatols)
> betahatlasso=as.matrix(betahatlasso)
> betahatridge=as.matrix(betahatridge)
> #-----
> coef=cbind(beta,betahatols,betahatlasso,betahatridge)
> colnames(coef)=c("beta","betahatols","betahatlasso","betahatridge")
> mse=cbind(mseols,mseridge,mselasso)
> colnames(mse)=c("mseols","mseridge","mselasso")
> coef
> mse
```

دستور زیر برآوردگر محدود شده در جدول ۱۶.۵ را محاسبه می کند.

```
%-----
> R=matrix(c(1,0,0,1,1,1,-2,3,0,-1,2,0,0,2,3,0,-3,1,2,5,4,0,0,0)
+ ,byrow = +T,ncol=8)
> R
> r=c(1.9,3.8,2.5)
> cbeta_ols=beta_ols+solve(t(x)%*%x)%*%t(R)%*%
+ solve((R)%*%(t(x)%*%x)%*%t(R))) %*%(r-R)%*%beta_ols)
> cbeta_lasso=beta_lasso+swlasso)%*%t(R)%*%
+ solve((R)%*%swlasso)%*%t(R))) %*%(r-R)%*%beta_lasso)
> cbeta_ridge=beta_ridge+swridge)%*%t(R)%*%
+ solve((R)%*%swridge)%*%t(R))) %*%(r-R)%*%beta_ridge)
> cbeta_gamma2.5=beta_gamma2.5+swbridge_gamma2.5)%*%t(R)%*%
+ solve((R)%*%swbridge_gamma2.5)%*%t(R))) %*%(r-R)%*%beta_gamma2.5)
> cbeta_gamma3=beta_gamma3+swbridge_gamma3)%*%t(R)%*%
+ solve((R)%*%swbridge_gamma3)%*%t(R))) %*%(r-R)%*%beta_gamma3)
```

```
> cbeta_gamma4=beta_gamma4+swbridge_gamma4%%t(R)%%  
+ solve((R%%swbridge_gamma4%%t(R))) %% (r-R%%beta_gamma4)  
> cbeta_final=cbind(cbeta_gamma4=cbeta_gamma4,cbeta_gamma3=  
+ cbeta_gamma3,cbeta_gamma2.5=cbeta_gamma2.5,  
+ cbeta_ridg=cbeta_ridg,cbeta_lasso=cbeta_lasso,  
+ cbeta_ols=cbeta_ols,beta=beta)  
> cbeta_final=round(cbeta_final,3)  
> cbeta_final
```

مراجع

- [۱] آرست، م. و آرشی، م. و ربیعی، م، (۱۳۹۵) برآوردگر انقباضی بریج محدود شده در مدل رگرسیون چندگانه، سیزدهمین کنفرانس آمار ایران – دانشگاه شهید باهنر کرمان.
- [۲] آرست، م. و آرشی، م. و ربیعی، م، (۱۳۹۵) مطالعه رفتار برآوردگر انقباضی تحت یک قید خطی در مدل رگرسیون جریمه شده، مجله علوم آماری (تحت داوری).
- [۳] بروزئی بیدگلی، م، (۱۳۹۳)، پایان نامه ارشد: ”بررسی رفتار برآوردگر جک نایف ریج”، دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود،
- [۴] تجدد ع، (۱۳۹۱)، پایان نامه ارشد: ”برآورد انقباضی در مدل های چندمتغیره”، دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود،
- [۵] حسن زاده بشتیپان، م، (۱۳۸۸)، پایان نامه ارشد: ”برآورد انقباضی در مدل رگرسیون ریج با خطاهای دارای توزیع بیضی گون”، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد،
- [۶] خوئی، م، (۱۳۹۵)، پایان نامه ارشد: ”رگرسیون ریج به ظاهر ناهمبسته در مدل های نیمه پارامتری”، دانشکده علوم ریاضی، دانشگاه سمنان،
- [۷] روزبه، م، (۱۳۹۰)، پایان نامه دکتری: ”برآورد در مدل های خطی جزئی”، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد،
- [۸] مونت گمری د، پک ا، (۲۰۱۲)، (ترجمه ابراهیم رضوی پاریزی) ”مقدمه ای بر تحلیل رگرسیون خطی” جلد اول، چاپ چهارم، انتشارات دانشگاه شهید باهنر کرمان
- [۹] نیرومند ح.ع، (۱۳۸۷) ”تحلیل رگرسیون خطی ابزاری برای تحقیق”، جلد اول، چاپ اول، انتشارات دانشگاه فردوسی مشهد
- [۱۰] نجاریان س، (۱۳۹۱)، پایان نامه ارشد: ”بررسی رفتار برآوردگر ریج در مدل های خطی جزئی”، دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود،
- [11] Akdeniz F. and Kaciranlar S. (1995), ”On the olmost unbiased genrelized Liu estimator and unbiased estimation of the Bias and MSE”Comm. Statist. Theo Meth, 24, 7, pp1789–1797

-
- [12] BELSLEY D. A., KUH E. and WELSCH R. E. (1980), "Regression Diagnostics" **John Wiley, New York**
- [13] Craven P. , Wahba, G. (1979)," Smoothing Noisy Data With Spline Functions" **Numerische Mathematik** 31, pp 377–403
- [14] Frank I. E. and Friedman J. H. (1993), "A statistical view of some chemometrics regression tools (with discussion)" **Technom**, 35, pp109–148
- [15] Friedman J., Hastie, T. and Tibshirani, R. (2008), "Regularization Paths for Generalized Linear Models via Coordinate Descent" **Journal of Statistical Software**, 33,1, pp 1-22
- [16] FU J. (1998), "Penalized Regressions: The Bridge Versus the Lasso" **American Statistical Association Institute of Mathematical Statistics, and Interface Foundation of North America Journal of Computational and Graphical Statistics**, 7, pp 397-416
- [17] GALTON F., (1886) "Regression towards mediocrity in hereditary stature" **The Journal of the Anthropological Institute of Great Britain and Ireland**, 15, pp 246-263
- [18] Gibbons D. G (1981), " A simulation study of some ridge estimators" **Amer. Statist. Assc**, 76, pp 1131-19
- [19] Grob J. (2003), "Restricted ridge estimation" **Statistic and probability Letters**, 65, pp 57-64
- [20] Hoerl. A. E, Kennard. R. W. (1970), " Ridge regression baised estimation for nonorthogonal problems" **Technometrics**, 12, pp 69-89
- [21] James W. , Stein C. (1961), "Estimation with quadrati loss" **Proceeding of the forth berkely symposium**, 1, pp 361-379
- [22] Jones C., Plotkin G. (1989), " A probabilistic powerdomain of evaluations" **in: Proc. 4th Annu. Symp. on Logic in Computer Science, IEEE Computer Society Press**, pp186–195
- [23] Knight K., Fu W. (2000), "ASYMPTOTICS FOR LASSO-TYPE ESTIMATORS " **The Annals of Statistics** , 28, 5, pp 1356–1378

-
- [24] KARIYA T., KURATA H. (2004), "Generalized least squares"**John Wiley and Sons, England**
- [25] Liu K. J. (2003), "Using Liu type estimator to combat multicollinearity"**Comm. Statist. Theo. Meth**, 32, 5, pp 1009-1020
- [26] McDonald G. C., Galarnaeu D. I. and Keimel K. (1975), "A Monte Carlo evaluation of some ridge-type estimators" **Amer. Statist. Assoc**, 20, pp 407- 416
- [27] MARQUART D. W. (1970), "Generalized inverse ridge regression biased linear estimation and non-linear estimation" **Technometrics**, 12, pp 591-612
- [28] Mason R. L, Gunst. R. F and Webster. J. T. (1975), "Ridge analysis and problems of multicolliearity"**Commun. Statist**, 4,3, pp 277-292
- [29] Montgomery. D. G. and Peck. A. E. (1992),**Introduction to linear regression analysis**, 2nd. ed
- [30] Montgomery. D. G. and Peck. A. E, and Vining. G. G. (2001).**Introduction to linear regression analysis**, Jhon Wiley, 3rth Ed, New York
- [31] Newhouse. J. P. and Oman. S. D. (1971), " An evaluation of ridge estimators"**Rand Corporation**. pp 716-PR
- [32] Osborne, M.R., Presnell, B. and Turlach, B.A. (2000), " On the LASSO and its Dual"**Journal of Computational and Graphical Statistics**, 9,2,pp 319–337
- [33] Swindel B. F. (1976), "Good ridge estimators based on prior information"**Comm. Statist. Theo. Meth**, 5, 2, pp 1065-1075
- [34] Tibshirani R. (1996), "Regression Shrinkage and Selection via the Lasso"**JRoyal of the Statistical Society.Series B(Methodological)**,58,1, pp 267-288
- [35] Theil H. and Goldberger A.S. (1961), "One pure and mixed statistical estimation in economics". J. Int.
- [36] Theil. H. (1963), "On the use of incomplete prior information in regression analysis"**Amer. Statist. Assoc** 58, pp 401-414

واژه‌نامه فارسی به انگلیسی

Probabilistic	احتمالی
Valuation	ارزیابی
Bias	اریبی
Trace	اثر
Regression to the mean	بازگشت به میانگین
Restricted estimator	برآوردگر محدود شده
Function Space	فضای تابع
Optimality	بهینگی
Ridge parameter	پارامتر ریج
Rank	رتبه
Multiple linear regression	رگرسیون چند متغیره خطی
Linear constrain	محدویت خطی
Scaler matrix	ماتریس اسکالر
Diagonal matrix	ماتریس طرح
Design matrix	ماتریس محدودیت
Singular matrix	ماتریس منفرد
Non-singular matrix	ماتریس نامنفرد
Unit matrix	ماتریس واحد
Orthogonal	متعامد
Response variable	متغیر پاسخ
Regressor variable	متغیر رگرسیونی
Symetric	متقارن
Singular linear model	مدل خطی منفرد
Inverse	معکوس
Generalized inverse	معکوس تعمیمیافته
Positive definite	معین مثبت
Constrast	مقابلہ

Linear dependence	وابسته خطی
Hat matrix	هت ماتریس – ماتریس برازش
Correlation	همبستگی
Multicollinearity	همخطی چندگانه
Asymptotic Distributional	توزیع مجانبی
Asymptotic Normal	نرمال مجانبی
Coefficient of Determination	ضریب تعیین
Condition Number	عدد شرطی
Cross-Validation	روش اعتبارسنجی متقابل
Determinant	دترمینان
Eigen Value	مقدار ویژه
Exact Collinearity	همخطی کامل
Full Rank	رتبه کامل
Generalized Restricted Least Squares	کمترین توان های دوم محدود شده تعمیم یافته
Idempotent	خودتوان
Ill Conditioned	بدشرطی
Independent Variable	متغیر مستقل
Indicator Function	تابع نشانگر
Lagrangian	لاگرانژ
Linear Non-Stochastic Constrains	محدودیت های خطی غیر تصادفی
Partial Collinearity	همخطی ناقص
Penalty	جریمه
Regularity Conditions	شرایط نظم
Residual	مانده
Ridge Regression	رگرسیون ریج
Bridge Regression	رگرسیون بریج
Lasso Regression	رگرسیون لاسو
Shrinkage Estimation	برآوردگر انقباضی
Spherical Restriction	محدودیت کروی
Unbiased	نااریب
Variance Inflation Factor	عامل تورم واریانس
Weighted Least Square Estimator	برآوردگر کمترین توان های دوم وزنی

واژه‌نامه انگلیسی به فارسی

Probabilistic	احتمالی
Valuation	ارزیابی
Bias	اریبی
Trace	اثر
Regression to the mean	بازگشت به میانگین
Restricted estimator	برآوردگر محدود شده
Function Space	فضای تابع
Optimality	بهینگی
Ridge parameter	پارامتر ریج
Rank	رتبه
Multiple linear regression	رگرسیون چند متغیره خطی
Linear constrain	محدویت خطی
Scaler matrix	ماتریس اسکالر
Diagonal matrix	ماتریس طرح
Design matrix	ماتریس محدودیت
Singular matrix	ماتریس منفرد
Non-singular matrix	ماتریس نامنفرد
Unit matrix	ماتریس واحد
Orthogonal	متعامد
Response variable	متغیر پاسخ
Regressor variable	متغیر رگرسیونی
Symetric	متقارن
Singular linear model	مدل خطی منفرد
Inverse	معکوس
Generalized inverse	معکوس تعمیم یافته
Positive definite	معین مثبت
Constrast	مقابلہ

Linear dependence	وابسته خطی
Hat matrix	هت ماتریس – ماتریس برازش
Correlation	همبستگی
Multicollinearity	همخطی چندگانه
Asymptotic Distributional	توزیع مجانبی
Asymptotic Normal	نرمال مجانبی
Coefficient of Determination	ضریب تعیین
Condition Number	عدد شرطی
Cross-Validation	روش اعتبارسنجی متقابل
Determinant	دترمینان
Eigen Value	مقدار ویژه
Exact Collinearity	همخطی کامل
Full Rank	رتبه کامل
Generalized Restricted Least Squares	کمترین توان های دوم محدود شده تعمیم یافته
Idempotent	خودتوان
Ill Conditioned	بدشرطی
Independent Variable	متغیر مستقل
Indicator Function	تابع نشانگر
Lagrangian	لاگرانژ
Linear Non-Stochastic Constrains	محدودیت های خطی غیر تصادفی
Partial Collinearity	همخطی ناقص
Penalty	جریمه
Regularity Conditions	شرایط نظم
Residual	مانده
Ridge Regression	رگرسیون ریج
Bridge Regression	رگرسیون بریج
Lasso Regression	رگرسیون لاسو
Shrinkage Estimation	برآوردگر انقباضی
Spherical Restriction	محدودیت کروی
Unbiased	نااریب
Variance Inflation Factor	عامل تورم واریانس
Weighted Least Square Estimator	برآوردگر کمترین توان های دوم وزنی

Aabstract

When the problem of multicollinearity is present, the least squares estimator is not applicable and mostly biased shrinkage estimators are used. Then are plenty methods to obtain shrinkage estimators. One method is the bridge estimator with special cases ridge and LASSO estimators. On the other hand, in practice and economic models, there is a need to apply some prior information on the parameters of interest. The estimator obtains under this information is said the restricted estimator. In this thesis, the asymptotic performance of the bridge and restricted bridge are investigated.

A numerical algorithm is developed to derive the bridge estimator. Properties of the two special cases ridge and LASSO are studies.

Keywords: Multicollinearity, Shrinkage method, Bridge estimator, Ridge estimator, LASSO, Linear constraint.



Shahrood University of Technology

Faculty Of Mathematical Sciences

MSc Thesis in: Mathematical Statistics

Comparing the Performance of Some Bridge Estimators in Multiple Regression Model

By: Mohammad Arast

Supervisors

Dr. Mohammad Arashi

Dr. Mohammad reza Rabie

February 2017