
بسم الله
الرحمن الرحیم

دانشکده علوم ریاضی

رشته آمار، گرایش استنباط

رساله دکتری

برآورد در مدل‌های نیمه‌پارامتری برای تحلیل داده‌های طولی

نگارنده: مژگان تعاونی

استاد راهنما

دکتر محمد آرشی

شهریور ۱۳۹۹

شماره: ۱-۳/۲-۳۳۵
تاریخ: ۹۹/۷/۲۹
ویرایش:

باسمه تعالی

مدیریت تحصیلات تکمیلی

فرم شماره ۱۱: صورت جلسه نهایی دفاع از رساله دکتری (Ph.D)

بدینوسیله گواهی می شود خانم **مژگان تعاونی** دانشجوی دکتری رشته آمار به شماره دانشجویی ۹۴۰۰۲۲۵ ورودی ۱۳۹۴ در تاریخ ۹۹/۶/۵ از رساله نظری■ / عملی□ خود با عنوان : **برآورد در مدل‌های نیمه‌پارامتری برای تحلیل داده‌های طولی**
دفاع و با اخذ نمره ۱۹,۹ به درجه : عالی نائل گردید.

الف) درجه عالی: نمره ۲۰-۱۹ ☑	ب) درجه خیلی خوب: نمره ۱۸/۹۹ – ۱۷ □
ج) درجه خوب: نمره۱۶/۹۹– ۱۵ □	د) مردود: کمتر از ۱۵ □

ردیف	هیئت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱	دکتر محمد آرشی	استاد راهنما	دانشیار	
۲	دکتر حسین باغیشنی	استاد مدعو داخلی	استادیار	
۳	دکتر محمدرضا ربیعی	استاد مدعو داخلی	استادیار	
۴	دکتر مهدی روزبه	استاد مدعو خارجی	دانشیار	
۵	دکتر سید حیدر جعفری	سرپرست (نماینده) تحصیلات تکمیلی دانشکده	استادیار	

مدیر محترم تحصیلات تکمیلی دانشگاه:

ضمن تأیید مراتب فوق مقرر فرمائید اقدامات لازم در خصوص انجام مراحل دانش آموختگی خانم مژگان تعاونی بعمل آید.

نام و نام خانوادگی رئیس دانشکده :

تاریخ و امضاء و مهر دانشکده:

تقدیم به

بزرگوار **پدرم**،

فداکار **مادرم**،

مهربان **همسرم**،

نازنین **پسرم**،

آنان که دریای بی‌کران فداکاری و عشق‌اند، وجودم برایشان همه رنج بود و وجودشان برایم همه مهر. هرچه بکوشم، قطره‌ای از دریای بی‌کران مهربانیشان را سپاس نتوانم گفت. بوسه بر دستان پرمهرتان. الها به من کمک کن تا بتوانم ادای دین کنم و به خواسته آنان جامه عمل بپوشانم .

سپاس بی‌کران پروردگار یکتا را که هستی‌مان بخشید؛
و به طریق علم و دانش رهنمونمان شد؛
و به همنشینی رهروان علم و دانش مفتخرمان نمود؛
و خوشه‌چینی از علم و معرفت را روزیمان ساخت.

نمی‌توانم معنایی بالاتر از تقدیر و تشکر بر زبانم جاری سازم و سپاس خود را در وصف استادان خویش آشکار نمایم، که هر چه گویم و سرایم، کم گفته‌ام. با این وجود بر خود واجب می‌دانم از استاد فرزانه جناب آقای دکتر محمد آرشی که به عنوان استاد راهنما در مراحل مختلف این رساله همواره با سعه‌صدر و گشاده‌رویی در کنار من بودند و در طول مدت تحصیل از راهنمایی‌های اخلاقی و حمایت‌های بی‌دریغ علمی ایشان بهره جسته‌ام، تشکر و قدردانی نمایم و برای ایشان طول عمر توام با سربلندی را آرزومندم. از اساتید گرامی جناب آقایان دکتر حسین باغیشنی، دکتر محمدرضا ربیعی و دکتر مهدی روزبه که زحمت داوری رساله اینجانب را قبول فرمودند، قدردانی می‌نمایم. خالصانه از تمامی معلمان و اساتیدی که در مقاطع مختلف تحصیلی به من علم آموخته و مرا از سرچشمه دانایی سیراب کرده‌اند، نهایت سپاس را دارم. پروردگارا حسن عاقبت، سلامت و سعادت را برای آنان مقدر نما.

مژگان تعاونی
شهریور ۱۳۹۹

تعهد نامه

اینجانب **مژگان تعاونی** دانشجوی دکتری رشته **آمار** دانشکده **علوم ریاضی** دانشگاه صنعتی شاهرود، نویسنده پایان‌نامه با عنوان **برآورد در مدل‌های نیمه‌پارامتری برای تحلیل داده‌های طولی** ، تحت راهنمایی دکتر **محمد آرشی** متعهد می‌شوم:

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

مژگان تعاونی
شهریور ۱۳۹۹

مالکیت نتایج و حق نشر

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

چکیده

این رساله به بررسی روش‌های برآورد در مدل رگرسیون نیمه‌پارامتری با اثرات آمیخته به منظور تحلیل داده‌های طولی می‌پردازد. مدل موردنظر ترکیبی از مدل اثرات آمیخته و مدل نیمه‌پارامتری است. ابتدا چند مورد از روش‌های موجود برای برآورد تابع ناپارامتری مدل شامل برآورد هسته‌ای، بازگشتی و اسپلاین معرفی شده و با درنظر گرفتن مساله هم‌خطی بین متغیرهای تبیینی، برآوردگر جدیدی به نام برآوردگر هسته‌ای ریج معرفی می‌شود. در ادامه مسئله برآورد و انتخاب متغیر همزمان برای داده‌های گاوسی و غیرگاوسی با بعد بالا درنظر گرفته شده است. مولفه ناپارامتری موجود در مدل با اسپلاین‌های رگرسیونی تقریب زده شده و از طریق بهینه‌سازی تابع هدف مبتنی بر تابع تاوان برآورد و انتخاب متغیر به طور همزمان انجام می‌گیرد. رفتار حدی برآوردگرهای حاصل در چهارچوب داده‌های بعد بالا که در آن تعداد پارامترها متناسب با افزایش حجم نمونه افزایش می‌یابد، مورد مطالعه قرار می‌گیرد. علاوه‌براین با درنظر گرفتن پاسخ‌های چندمتغیره یک روش نیرومند بر اساس توزیع تی چندمتغیره، به عنوان جایگزینی برای توزیع نرمال، برای برآوردیابی ارائه می‌شود که نسبت به مقادیر پرت یا مشاهدات غیرمعمول انعطاف‌پذیرتر است. در هر چهارچوب مطالعاتی، به منظور پیاده‌سازی روش‌های برآورد یک الگوریتم تکراری مناسب ارائه گردیده و کارایی روش‌های برآورد پیشنهادی با استفاده از مطالعات شبیه‌سازی و تحلیل داده‌های واقعی بررسی شده است. نتایج بدست آمده نشان می‌دهند که روش‌های پیشنهادی در مقایسه با سایر روش‌های موجود عملکرد مناسبی دارند.

کلمات کلیدی: اسپلاین هموارساز، انتخاب متغیر، برآوردگر ریج، بعد بالا، تاوانیده، توزیع تی چندمتغیره، توزیع مجانبی، داده‌های پرت، داده‌های طولی، دم‌کلفت، بازگشتی، هسته‌ای، معادلات برآورد تعمیم‌یافته، مدل نیمه‌پارامتری با اثرات آمیخته.

فهرست مقالات مستخرج از رساله

۱. تعاونی، م. و آرشی، م. (۱۳۹۹). انتخاب متغیر در مدل‌های خطی-جزئی با اثرات آمیخته برای داده‌های طولی با بعد بالا، مجله علوم آماری، جلد ۱۴، شماره ۲، صفحات ۳۶۷-۳۸۸.

2. Taavoni, M. and Arashi, M. (2018), Semiparametric Mixed Effect Modeling for Longitudinal Data Analysis, The Secound Seminar on Nonparametric Statistics and Its Applications, Allameh Tabataba'i University, Tehran, Iran.

3. Taavoni, M. and Arashi, M. (2018), Semiparametric Ridge Regression for Longitudinal Data, 14th Iranian Statistics Conference, Shahrood University of Technology, Shahrood, Iran.

4. Taavoni, M. and Arashi, M. (2019), Kernel Estimation in Semiparametric Mixed Effect Longitudinal Modeling, *Stat. Papers*, DOI: 10.1007/s00362-019-01125-8.

5. Taavoni, M. and Arashi, M. (2019), Regularization in Generalized Semiparametric Mixed Effects Model for Longitudinal Data, arXiv: 1909.08498v1 [stat.ME].

6. Taavoni, M., Arashi, M., Wang, W.L. and Lin, T.I (2020), Multivariate t Semiparametric Mixed-effects Model for Longitudinal Data with Multiple Characteristics, *J. Statist. Comput. Simul.*, (Revised).

7. Taavoni, M., Arashi, M., Wang, W.L. and Lin, T.I (2020), Estimation in Multivariate t Linear Mixed Models for Longitudinal Data with Multiple Outputs: Application to PBCseq Data Analysis, *Biom. J*, (Revised).

پیشگفتار

مطالعات طولی یکی از رایج‌ترین و کاربردی‌ترین نوع مطالعات در بسیاری از زمینه‌ها است. داده‌های طولی به مجموعه وسیعی از داده‌های گسسته و پیوسته اطلاق می‌شود که در آن ویژگی‌های چندین آزمودنی به عنوان نمونه‌ای از اعضای جمعیت، در طول زمان اندازه‌گیری می‌شوند. هدف اولیه مطالعات طولی، بررسی تغییرات متغیر پاسخ در طول زمان و نیز استخراج عامل‌های تاثیرگذار بر این تغییرات است. با توجه به انجام اندازه‌گیری‌های مکرر، پاسخ‌های هر آزمودنی دارای همبستگی هستند. جهت رسیدن به استنباط علمی و معتبر، درنظر گرفتن این همبستگی امری ضروری است. به دلیل وجود اندازه‌های تکراری امکان ارزشیابی تغییرات درون هر آزمودنی نیز فراهم می‌شود که این امکان در مطالعات مقطعی که آزمودنی تنها در یک زمان مشخص اندازه‌گیری می‌شود وجود ندارد.

مدل رگرسیونی یکی از پرکاربردترین ابزارهای آماری است که به دلیل تنوع زیاد مطالعات، روش‌های گوناگونی برای انجام آن پدید آمده است. از یک منظر رگرسیون را می‌توان به سه بخش پارامتری، ناپارامتری و نیمه پارامتری تقسیم کرد:

- رگرسیون پارامتری بیشتر وقتی مورداستفاده قرار می‌گیرد که اطلاعات تحقیق به قدر کافی زیاد باشد تا آنجایی‌که بتوان ارتباط بین متغیر پاسخ و تبیینی را به وسیله یک مدل خطی یا قابل تبدیل به خطی بیان کرد. در تحلیل پارامتری داده‌های طولی، رهیافت‌های مختلفی برای درنظر گرفتن همبستگی وجود دارند که شامل مدل‌های حاشیه‌ای، اثرات آمیخته و انتقالی هستند.
- رگرسیون ناپارامتری در مواقعی که نتوان نوع ارتباط بین متغیرها را با یک شکل تابعی معلوم و مشخص بیان کرد، مورد استفاده قرار می‌گیرد، که در آن می‌توان از روش‌های اسپلاین و هسته‌ای استفاده کرد.
- مدل‌های نیمه‌پارامتری، ترکیبی از مدل‌های پارامتری و ناپارامتری است، به این معنی که بعضی از متغیرها دارای روند معلوم تغییرات نسبت به متغیر پاسخ بوده و برای بعضی از متغیرها نمی‌توان شکل تابعی خاصی را درنظر گرفت.

در چند دهه اخیر، آماردانان در راستای تحلیل هر چه بهتر داده‌های طولی با خلق ترکیبی از مدل‌های ذکر شده، رده دیگری از مدل‌ها را ایجاد کردند. در این میان، مدل اثرات آمیخته از مجموعه مدل‌های پارامتری و مدل‌های نیمه‌پارامتری مبنای بسیاری از این دسته مدل‌های ترکیبی قرار گرفتند. مدل نیمه‌پارامتری با اثرات آمیخته که اساس کار این پژوهش است، به عنوان مدلی بسیار منعطف و تنومند که ویژگی‌های خوب اکثر مدل‌ها را دارد مورد توجه بسیار قرار گرفت.

از طرفی، امروزه با افزایش توان محاسباتی کامپیوترها و روش‌های نوین جمع‌آوری اطلاعات، داده‌های با بعد بالا در زمینه‌های مختلف علمی و تحقیقاتی با هزینه نسبتاً کم در اختیار محققین قرار می‌گیرد. در این نوع داده‌ها فقط تعداد اندکی از متغیرهای تبیینی واقعاً با متغیر پاسخ مرتبط هستند و سایر متغیرها تاثیری بر پاسخ ندارند. هنگامی‌که تعداد متغیرهای تبیینی در مدل زیاد است، تفسیر آن مدل بسیار مشکل بوده و همچنین حضور تعداد زیاد متغیرها باعث ایجاد هم‌خطی بین

متغیرها می‌شود؛ لذا شناسایی متغیرهای موثر بر پاسخ اهمیت بسزایی دارد. یک روش متداول برای مقابله با مشکل هم‌خطی استفاده از برآوردگرهای اریب مانند ریج است. همچنین فرآیندی که علاوه بر انتخاب متغیر و بهترین معادله رگرسیونی، برآورد ضرایب را نیز انجام دهد، روش مبتنی بر تابع تاوان است که رویکرد نوینی تحت عنوان رگرسیون تاوانیده را گسترش داد. در بسیاری از تحقیقات علمی، با داده‌های رسته‌ای یا داده‌های غیرنرمال، مانند پاسخ‌های دودویی و شمارشی مواجه هستیم که برای مدل‌بندی آن‌ها، مدل‌های خطی تعمیم‌یافته پیشنهاد می‌شوند. مطالعات طولی که متغیر پاسخ آن‌ها غیرنرمال باشد نیز در علوم مختلف جایگاه ویژه‌ای یافته‌اند. مسئله مهم کاربردی دیگر که در تحلیل داده‌های طولی مورد توجه قرار می‌گیرد، تحلیل همزمان یا چند متغیره این داده‌ها است. مدل‌های چندمتغیره آمیخته از پرکاربردترین روش‌های آماری جهت انجام استنباط اندازه‌های تکراری چندمتغیره است. پذیره معمول در برازش این مدل‌ها نرمال بودن توزیع خطاها و مولفه‌های اثرات تصادفی است. در برخورد با برخی داده‌های واقعی این پذیره ممکن است منجر به نتایج نادرست شود به‌ویژه در صورت وجود داده‌های دورافتاده، استنباط را آسیب‌پذیر ساخته و واریانس برآوردها را به شدت تحت تاثیر قرار می‌دهد. از این‌رو در سال‌های اخیر روش‌های پیشرفته آماری با جایگزین کردن توزیع‌های متقارن دم‌سنگین به جای پذیره نرمال بودن در این نوع از مدل‌بندی‌ها منجر به نتایج معتبرتری در مطالعات طولی شده است. علیرغم سازگاری این مدل‌ها با داده‌های چندمتغیره و تنومند بودن در مقابل مقادیر پرت، این مدل‌ها تنها محدود به رابطه خطی بین متغیرهای پاسخ و تبیینی است.

طبق بررسی‌های صورت گرفته، تحقیقات در زمینه انتخاب متغیر داده‌های طولی با بعد بالا در مدل رگرسیون نیمه‌پارامتری با اثرات آمیخته به ویژه در چارچوب خانواده توزیع‌های نمایی تا حد زیادی ناشناخته مانده است. همچنین استفاده از مدل‌های نیمه‌پارامتری برای داده‌های طولی چندمتغیره نسبتا نادر بوده است. چه بسا مسائلی مانند مسئله هم‌خطی و انتخاب متغیر نیز در این نوع مدل‌بندی‌ها کمتر مورد بحث قرار گرفته‌اند. لذا تحقیقات انجام شده در راستای تمرکز این رساله، که بررسی روش‌های برآورد در مدل رگرسیون نیمه‌پارامتری با اثرات آمیخته است، را در ۴ فصل گنجانده‌ایم که شرح مختصری از محتویات آن‌ها در زیر آمده است.

فصل اول: به بیان مقدمات لازم جهت آشنایی با داده‌های طولی، معرفی ویژگی‌های هر یک از مدل‌های رگرسیونی بیان‌شده و روش‌های برآورد در آن‌ها اختصاص یافته است. سپس با مبنا قراردادن مدل نیمه‌پارامتری با اثرات آمیخته، برخی از روش‌های برآورد تابع ناپارامتری مدل شامل برآوردگرهای هسته‌ای، بازگشتی و اسپلاین معرفی می‌شوند.

فصل دوم: ابتدا تاثیر وجود هم‌خطی بر عملکرد برآوردگرها بررسی شده و برآوردگر ریج، به عنوان ابزاری برای مقابله با مشکل هم‌خطی معرفی می‌شود. سپس با معرفی رایج‌ترین روش‌های انتخاب متغیر، برآورد و انتخاب متغیر در مدل نیمه پارامتری با اثرات آمیخته را مبتنی بر رهیافت توابع تاوان، در چارچوب داده‌های بعد بالا که در آن تعداد پارامترها متناسب با افزایش حجم نمونه افزایش می‌یابد، و نحوه انتخاب پارامترهای تابع تاوان مورد بررسی

قرار می‌گیرد.

فصل سوم: به معرفی اجمالی مدل‌های آمیخته تعمیم‌یافته نیمه‌پارامتری پرداخته است. سپس برآورد و انتخاب متغیر در مدل آمیخته تعمیم‌یافته نیمه‌پارامتری را مبتنی بر رهیافت توابع تاوان برای داده‌های طولی غیرنرمال با بعد بالا و بعد پایین درنظر می‌گیرد. برآوردگرهای معرفی‌شده صورت بسته ندارند؛ لذا یک الگوریتم تکراری به نام الگوریتم نیوتن رافسون مونت کارلویی تاوانیده پیشنهاد می‌شود.

فصل چهارم: با درنظر گرفتن پاسخ‌های چندمتغیره، مسئله انتخاب متغیر در مدل‌های چندمتغیره آمیخته با توزیع تی چندمتغیره به عنوان جایگزینی برای توزیع نرمال، می‌پردازد که نسبت به مقادیر پرت یا مشاهدات غیرمعمول انعطاف‌پذیرتر است. سپس نسخه نیمه‌پارامتری مدل مذکور تحت عنوان مدل آمیخته چندمتغیره نیمه‌پارامتری معرفی می‌گردد.

لازم به ذکر است که در هر چهارچوب مطالعاتی، خواص مجانبی برآوردگرهای حاصل بررسی شده و به منظور پیاده‌سازی روش‌های برآورد، یک الگوریتم تکراری مناسب ارائه شده است. کارایی روش‌های برآورد پیشنهادی با استفاده از مطالعات شبیه‌سازی و تحلیل داده‌های واقعی بررسی می‌شود. نتایج بدست آمده نشان می‌دهند که روش‌های پیشنهادی در مقایسه با سایر روش‌های موجود عملکرد مناسبی دارند. همچنین برای برآورد مولفه‌های ناپارامتری مدل، رگرسیون اسپلاین را اتخاذ کرده و در مسائل برآورد و انتخاب متغیر همزمان تابع تاوان اسکد درنظر گرفته می‌شود. این مجموعه شامل ۲ پیوست است.

پیوست آ: شامل تعاریف و ملزومات استفاده شده در رساله است.

پیوست ب: چند لم مورد نیاز و اثبات قضایای بیان‌شده در فصل‌های مختلف رساله را شامل می‌شود.

فهرست مطالب

فهرست تصاویر

فهرست جداول

فصل ۱

گذری بر مفاهیم داده‌های طولی

از آنجایی‌که آشنایی با یک موضوع جدید، مستلزم شناخت مقدمات مربوط به آن است، این فصل به معرفی مقدمات لازم جهت آشنایی با داده‌های طولی[1] اختصاص یافته است. بدین منظور، ابتدا تعریفی از داده‌های طولی، نمادگذاری متداول، اهداف و لزوم ارائه مبحث تحلیل داده‌های طولی بیان می‌گردد. سپس به معرفی و بررسی مدل‌های رگرسیونی که در تحلیل داده‌های طولی به‌کار می‌روند پرداخته شده است. مطالب این فصل عمدتاً برگرفته از کتاب تحلیل چندمتغیره‌ی گسسته در مطالعات مقطعی و طولی تالیف گنجعلی و رضایی (۱۳۸۹) و پایان نامه کارشناسی ارشد تعاونی (۱۳۹۲) است. علاقه‌مندان به منظور آشنایی بیش‌تر با داده‌های طولی و نحوه مدل‌بندی آن‌ها می‌توانند به کتاب‌های دیگل و همکاران (۲۰۰۲)، هدیکر و گیبونز (۲۰۰۶)، وو و ژانگ (۲۰۰۶)، فیتزموریس و همکاران (۲۰۰۹)، وربک و مولنبرگ (۲۰۰۹) و فیتزموریس و همکاران (۲۰۱۱) مراجعه نمایند.

۱.۱ تعریف و نمادگذاری

مسئله مهم در یک تحلیل آماری، نوع مطالعه و روش گردآوری داده‌ها است. در بسیاری از تحقیقات علمی به‌ویژه در علوم پزشکی به منظور بررسی روند تحقیقات و تاثیر روش‌های مختلف، متغیرهای

[1] Longitudinal data

موجود در مطالعه، بیش‌تر از یک بار روی یک آزمودنی[۲](انسان، حیوان و نمونه‌های آزمایشگاهی) اندازه‌گیری می‌شود؛ مشاهدات حاصل را اندازه‌گیری مکرر[۳] می‌نامند. فاکتور اندازه‌گیری براساس نوع مطالعه انتخاب می‌شود که می‌تواند زمان، مکان یا هر شرط آزمایشی دیگر باشد که اغلب به این فاکتور، فاکتور درون-گروهی[۴] یا درون- فردی[۵] گفته می‌شود. داده‌های مکرری که موقعیت‌های تکرار مشاهدات، نقاط زمانی هستند داده‌های طولی و مطالعاتی را که بر روی این داده‌ها انجام می‌گیرد، مطالعات طولی می‌نامند. می‌توان گفت مطالعات طولی یک شاخه از مطالعات با اندازه‌گیری مکرر هستند که در آن فاکتور اندازه‌گیری زمان است. به عنوان مثال در کودکان مبتلا به بیماری آسم، مشاهدات در ۴ نقطه زمانی یعنی در سنین ۷، ۸، ۹ و ۱۰ سالگی اندازه‌گیری می‌شوند. جهت آشنایی بیش‌تر با ساختار داده‌های طولی در ادامه به معرفی نمادگذاری متداول این داده‌ها می‌پردازیم.

با توجه به تعریف داده‌های طولی، نمونه‌ای تصادفی و مستقل به حجم n درنظر بگیرید به‌طوری‌که نمونه iام دارای n_i مشاهده مکرر در نقط زمانی است. متغیر پاسخ طولی[۶] به‌صورت

$$\boldsymbol{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}_{n_i \times 1}$$

تعریف می‌شود. بردار متغیرهای تبیینی[۷] وابسته به متغیر پاسخ y_{ij} به‌صورت

$$\boldsymbol{x}_{ij} = \begin{pmatrix} x_{1,ij} \\ \vdots \\ x_{p,ij} \end{pmatrix}_{p \times 1}$$

است که در آن $x_{k,ij}$ مقدار kامین متغیر تبیینی مربوط به آزمودنی iام در زمان jام است، به‌طوری‌که $i = 1, 2, \ldots, n$، $j = 1, 2, \ldots, n_i$ و $k = 1, 2, \ldots, p$ بوده و n، n_i و p به‌ترتیب معرف تعداد کل آزمودنی‌ها، تعداد دوره‌های زمانی برای آزمودنی iام و تعداد متغیرهای تبیینی هستند. در ماتریس طرح مربوط به آزمودنی iام که به‌صورت

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{i1}^\top \\ \vdots \\ \boldsymbol{x}_{in_i}^\top \end{pmatrix} = \begin{pmatrix} x_{1,i1} & \cdots & x_{p,i1} \\ \vdots & & \\ x_{1,in_i} & \cdots & x_{p,in_i} \end{pmatrix}_{n_i \times p}$$

تعریف می‌شود، ستون‌ها به مشاهدات مکرر هر متغیر در طول زمان مربوط بوده و ردیف‌ها به مشاهدات تمام متغیرها در یک زمان خاص مربوط هستند.

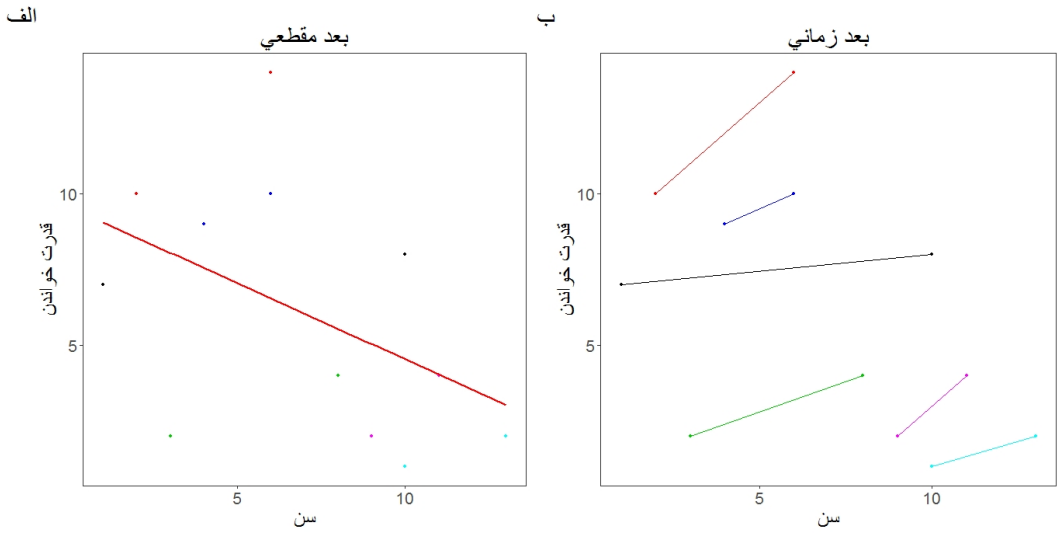
---

[2] Subject
[3] Repeated measurement
[4] Within-group
[5] Within-subject
[6] Response
[7] Explanatory

با توجه به تعریف و شیوه نمادگذاری بیان شده، این داده‌ها را می‌توان از دو بعد مورد توجه قرار داد: بعد زمانی و بعد مقطعی. به عبارت دیگر در مطالعات طولی بر خلاف مطالعات مقطعی[۸]، علاوه بر بررسی اثر متغیرهای تبیینی[۹] می‌توان اثر زمان و اثر متقابل زمان–مداخله را نیز مورد بررسی قرار داد. به عنوان مثال اگر داده‌هایی را که در زمان jام جمع‌آوری شده‌اند درنظر بگیریم، در واقع آزمودنی‌های اول تا nام را تنها در یک مقطع از زمان مورد بررسی قرار داده‌ایم؛ بنابراین برای هر دوره زمانی، داده‌ها مقطعی هستند. اما اگر داده‌های مربوط به آزمودنی iام را درنظر بگیریم، تنها یک آزمودنی را در زمان‌های متوالی مورد بررسی قرار داده‌ایم؛ بنابراین داده‌ها اثر زمان را نشان می‌دهند. لذا در تحلیل داده‌های طولی با سه هدف اصلی زیر روبرو هستیم:

۱. مطالعه تغییرات آزمودنی‌ها در طول زمان.
۲. مطالعه رابطه بین متغیرها (اثر متغیرهای تبیینی بر روی متغیر پاسخ).
۳. مطالعه تاثیر زمان بر رابطه بین متغیرها.

برای درک بهتر موضوع مثالی از دیگل و همکاران (۲۰۰۲) را مورد بررسی قرار می‌دهیم.



شکل ۱.۱: نمودار قدرت خواندن کودکان

در شکل ۱.۱، داده‌ها از یک مطالعه طولی مربوط به قدرت خواندن کودکان به دست آمده‌اند که در آن شاخص توانایی مطالعه هر کودک، دو بار اندازه‌گیری شده است. بعد مقطعی داده‌ها در قسمت (الف)، کاهش قدرت خواندن کودکان با گذشت زمان را نشان می‌دهد؛ درحالی‌که بعد زمانی داده‌ها در قسمت (ب)، بیان می‌کند که توانایی هر کودک با گذشت زمان افزایش یافته است. در مجموع می‌توان گفت، درحالی‌که توانایی خواندن در کودکان جوان‌تر سطح بالاتری دارد، همه کودکان توانایی خود را با گذشت زمان بهبود می‌بخشند. بنابراین تحلیل داده‌های طولی که با هدف مشاهده تغییر رفتار پدیده‌ها در طول زمان انجام می‌گیرد، نسبت به داده‌های صرفاً مقطعی یا صرفاً زمانی از اهمیت

[8] Cross-sectional

[۹] مداخله‌گرها

ویژه‌ای برخوردارند. این ماهیت دو بعدی، لزوم ارائه ترکیبی از ابزارهای آماری را برای تحلیل این داده‌ها آشکار می‌سازد.

۲.۱ ویژگی‌های داده‌های طولی

در مطالعات طولی هر دو متغیر پاسخ و تبیینی می‌توانند گسسته یا پیوسته و متغیرهای تبیینی می‌توانند در طول زمان ثابت یا متغیر باشند. همچنین فاصله نقاط زمانی و تعداد مشاهدات همه آزمودنی‌ها لزوماً برابر نیستند. این داده‌ها با توجه به نوع متغیرهای تبیینی، موقعیت و تعداد تکرارها به چندین دسته تقسیم‌بندی می‌شوند. شرح مختصری از این تقسیم‌بندی‌ها در زیر آمده است.

۱. اگر متغیرهای تبیینی در طول زمان ثابت باشند، یعنی به ازای هر $i = ۱, \ldots, n$، داشته باشیم

$$x_{k,i۱} = x_{k,i۲} = \ldots = x_{k,in_i},$$

به متغیر تبیینی kام، بین-گروهی[۱۰] یا زمان-ثابت[۱۱] گفته می‌شود. به عنوان مثال متغیرهای جنسیت و نژاد جزو این دسته هستند. به متغیر تبیینی kام، $k = ۱, \ldots, p$، درون-گروهی یا زمان-متغیر[۱۲] گفته می‌شود، اگر به ازای هر $i = ۱, \ldots, n$ و $j \neq j'$ داشته باشیم

$$x_{k,ij} \neq x_{k,ij'}.$$

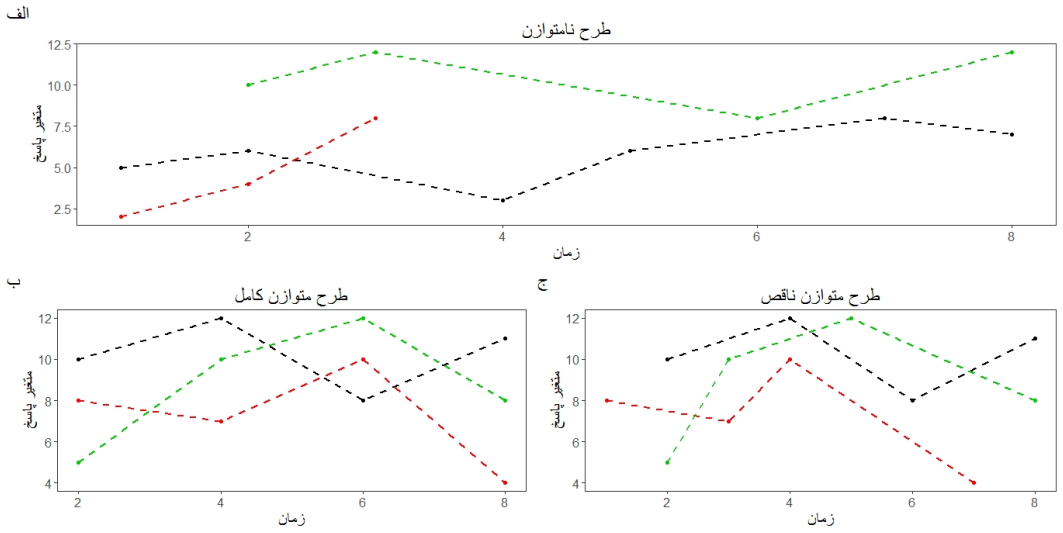
متغیرهایی مانند قد، وزن و سن شامل این موارد هستند.

متغیرهای زمان-ثابت که می‌توانند در طول دوره متغیر باشند اما بر اساس طراحی از قبل تعیین‌شده مقدار ثابت و معینی گرفته‌اند را طرح-ثابت[۱۳] می‌نامند. کد شناسایی آزمودنی‌ها و شاخص‌های درمانی شامل این مورد هستند. همچنین به متغیرهای زمان-ثابت، مانند وضعیت اولیه بیمار یا سن بیمار در زمان شروع مطالعه که مقادیر تصادفی هستند، طرح-تصادفی[۱۴] گفته می‌شود.

۲. در متون مربوط به داده‌های طولی، اگر تمام آزمودنی‌ها دارای تعداد تکرارهای برابر باشند یعنی $n_i = m$، آنگاه طرح مطالعاتی را متوازن[۱۵] و در غیر این صورت نامتوازن می‌نامند. همچنین یک طرح متوازن را کامل می‌گویند اگر همه n آزمودنی در موقعیت‌های زمانی مشابه اندازه‌گیری شده‌باشند، در غیر این صورت طرح را متوازن ناقص می‌نامند. برای درک بهتر این تقسیم‌بندی‌ها می‌توان به نمودار ساختگی در شکل ۲.۱ توجه نمود. در نمودار (الف) تعداد

[10] Between-group
[11] Time-invariant
[12] Time-varying
[13] Fixed-design
[14] Stochastic-design
[15] Balanced

مشاهدات مکرر آزمودنی‌ها متفاوت‌اند بنابراین طرح مطالعاتی نامتوازن است. در نمودار (ب) علاوه بر این که تعداد تکرارهای آزمودنی‌ها برابراند، تمام آزمودنی‌ها در موقعیت‌های زمانی ۲، ۴، ۶ و ۸ اندازه‌گیری شده‌اند؛ لذا طرح متوازن کامل است و نمودار (ج) یک طرح متوازن ناقص را نمایش می‌دهد. یکی از دلایل عمده برخورد با طرح‌های نامتوازن داده‌های گمشده هستند. داده‌های گمشده مسئله‌ای کاملاً رایج در زمینه مطالعات طولی است. روش‌ها برای بررسی داده‌های گمشده به الگوی گم‌شدن و مکانیزم مقادیر گمشده وابسته هستند. شناخت کامل این روش‌ها و بحث درباره آن‌ها، یک زمینه گسترده مطالعاتی را بوجود می‌آورد که مستلزم صرف زمان بیش‌تری است. لذا در این رساله به این موضوع پرداخته نشده‌است. علاقه‌مندان برای آگاهی بیش‌تر در خصوص مقادیر گمشده در داده‌های طولی و انواع مدل‌بندی آن‌ها می‌توانند به فیتزموریس و همکاران (۲۰۰۹) رجوع کنند. در ادامه با تاکید بر روی ساختار همبستگی در داده‌های طولی، همبستگی درون-گروهی را با جزئیات بیش‌تری بررسی می‌کنیم.



شکل ۲.۱: نمودار ساختگی جهت نمایش تقسیم‌بندی طرح‌های مطالعاتی

۱.۲.۱ همبستگی درون گروهی

علاوه بر حجیم بودن اطلاعات حاصل از داده‌های طولی و ماهیت دوبعدی آن‌ها که پیچیدگی داده‌ها را سبب می‌شوند، ویژگی بسیار مهم دیگری نیز وجود دارد که تحلیل و مدل‌بندی این داده‌ها را با چالش روبرو می‌سازد. بر خلاف فرضیات روش‌های آماری متداول که بر پایه استقلال مشاهدات طراحی شده‌اند، در این داده‌ها مجموعه مشاهدات مربوط به هر واحد دارای همبستگی درونی بوده و تنها از مشاهدات سایر واحدها مستقل هستند. با توجه به همبستگی موجود در داده‌های طولی، فرضیات یک رگرسیون خطی را می‌توان به‌صورت زیر درنظر گرفت:

الف. $(\boldsymbol{y}_۱, \boldsymbol{X}_۱), \ldots, (\boldsymbol{y}_n, \boldsymbol{X}_n)$ دوبه‌دو از هم مستقل‌اند.

ب. $\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{X}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$.

ج. $\mathrm{Cov}(\boldsymbol{y}_i|\boldsymbol{X}_i) = \boldsymbol{\Sigma}_{i_{n_i\times n_i}}$، که در آن $\boldsymbol{\Sigma}_i$ تابعی معلوم از متغیرهای تبیینی $\boldsymbol{X}_i$ است.

بدون کاستن از کلیت مسئله، می‌توان $\boldsymbol{X}_i$ها را تصادفی درنظر گرفت. اگر $\boldsymbol{X}_{ij}$ها طرح-ثابت باشند با احتمال ۱ مقادیر ثابت x_{ij} را می‌گیرند. اما هنگامی‌که $\boldsymbol{X}_{ij}$ها طرح-تصادفی باشند برای سادگی در نوشتار میانگین و کوواریانس شرطی را به‌صورت $\mathrm{E}(\boldsymbol{y}_i)$ و $\mathrm{Cov}(\boldsymbol{y}_i)$ درنظر می‌گیریم. میانگین شرطی پاسخ jام از نمونه iام به شرط $\boldsymbol{X}_{i۱}, \ldots, \boldsymbol{X}_{in_i}$ یک تابع خطی از $\boldsymbol{X}_{ij}$ها به‌صورت

$$\mathrm{E}(y_{ij}|\boldsymbol{X}_i) = \boldsymbol{X}_{ij}^\top\boldsymbol{\beta} = \beta_۱x_{۱,ij} + \ldots + \beta_px_{p,ij}$$

است. با استفاده از نماد ماتریسی می‌توان نوشت

$$\mathrm{E}\begin{pmatrix} \boldsymbol{y}_۱ \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix}_{N\times ۱} = \begin{pmatrix} \boldsymbol{X}_۱ \\ \vdots \\ \boldsymbol{X}_n \end{pmatrix}_{N\times p} \times \begin{pmatrix} \beta_۱ \\ \vdots \\ \beta_p \end{pmatrix}_{p\times ۱}$$

یا

$$\mathrm{E}(\boldsymbol{y})_{N\times ۱} = \boldsymbol{X}_{N\times p} \times \boldsymbol{\beta}_{p\times ۱}.$$

همچنین

$$\mathrm{Cov}(\boldsymbol{y})_{N\times N} = \begin{pmatrix} \boldsymbol{\Sigma}_{۱_{n_۱\times n_۱}} & \mathbf{0} & \ldots & \mathbf{0} \\ & \boldsymbol{\Sigma}_{۲_{n_۲\times n_۲}} & & \mathbf{0} \\ & & \ddots & \vdots \\ & & & \boldsymbol{\Sigma}_{n_{n_n\times n_n}} \end{pmatrix}_{N\times N},$$

که در آن $N = \sum_{i=۱}^n n_i$.

فرضیه ج این امکان را فراهم می‌کند که همبستگی بین مشاهدات هر نمونه را در مدل درنظر بگیریم. $\boldsymbol{\Sigma}_i$ می‌تواند تحت تاثیر عوامل مختلفی قرارگرفته باشد. یکی از این عوامل متغیرهای تبیینی هستند، مثلا گروه‌های جنسیتی یا نژاد می‌توانند عامل همبستگی درون گروهی باشند؛ همچنین $\boldsymbol{\Sigma}_i$ می‌تواند تابعی از زمان بوده یا تنها از طریق i به n_iها مرتبط باشد. در بسیاری از موارد تمرکز اصلی روی مطالعه تغییر رفتار متغیر پاسخ است که توسط $\boldsymbol{X}_i\boldsymbol{\beta}$ مدل‌بندی می‌شود و مطالعه روی $\boldsymbol{\Sigma}_i$ در اولویت‌های بعدی قرار گرفته، یا به عنوان عامل مزاحم درنظر گرفته می‌شود. در چنین حالتی، اگر طرح متوازن بوده یعنی $n_i = m$ و m نسبت به n کوچک باشد و همچنین فرض شود که کوواریانس به متغیرهای تبیینی وابسته نیست، آنگاه $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ را می‌توان یک ماتریس معین مثبت[۱۶] درنظر گرفت، به‌طوری‌که اعضای آن با دقت بالایی قابل برآورد است. اما اگر m نسبت به n بزرگ بوده یا طرح

[16] Positive definite

نامتوازن باشد، آنگاه لازم است که برای Σ_i مدل‌های پارامتری درنظر بگیریم. بنابراین همبستگی درون-گروهی معمولا به دو روش، شامل مدل‌های همبستگی سریالی و مدل‌های اثرات تصادفی درنظر گرفته می‌شوند. مدل‌های اثرات تصادفی در بخش‌های بعدی به تفصیل معرفی خواهند شد و در ادامه نمونه‌هایی از ساختار Σ بر اساس همبستگی سریالی با واریانس ثابت σ^2 را معرفی خواهیم کرد.

۱. همبستگی تبادل‌پذیر[17]:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}.$$

۲. همبستگی غیرساختاری[18] یا نواری[19]:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{m-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{m-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{m-1} & \rho_{m-2} & \rho_{m-3} & \dots & 1 \end{pmatrix}.$$

۳. همبستگی اتو-رگرسیو[20]:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{m-1} \\ \rho & 1 & \rho & \dots & \rho^{m-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \rho^{m-3} & \dots & 1 \end{pmatrix}.$$

همبستگی تبادل‌پذیر یکی از ساده‌ترین ساختارهاست که در آن مشاهدات هر آزمودنی همبستگی یکسانی دارند، یعنی $\mathrm{corr}(y_{ij}, y_{ik}) = \rho_{jk} \equiv \rho$. در این ساختار موقعیت مشاهدات در همبستگی آن‌ها تاثیر ندارد اما در سایر ساختارها از فاصله بین مشاهدات یا نقاط زمانی برای مدل‌سازی همبستگی استفاده می‌شود. در ساختار نواری همبستگی بین مشاهدات با ترتیب آن‌ها تعیین می‌گردد، $\mathrm{corr}(y_{ij}, y_{ij+k}) = \rho_k$. یعنی تمام جفت مشاهدات هم‌جوار دارای همبستگی یکسان به‌صورت

[17] Exchangeable
[18] Unstructured
[19] Banded
[20] Auto-regressive

$\text{corr}(y_{i۱}, y_{i۲}) = \text{corr}(y_{i۲}, y_{i۳}) = \ldots = \text{corr}(y_{im-۱}, y_{im}) = \rho_۱$ هستند. به‌طور مشابه تمام جفت مشاهداتی که به اندازه k واحد از هم فاصله دارند دارای همبستگی ρ_k می‌باشند. ساختار اتو-رگرسیو بیان‌شده برای طرح‌های متوازن مناسب است اما برای طرح‌های نامتوازن می‌توان مدل همبستگی را به‌صورت

$$\text{corr}(y_{ij}, y_{ik}) = \rho^{|t_{ij}-t_{ik}|}$$

درنظر گرفت که در آن $(t_{i۱}, t_{i۲}, \ldots, t_{in_i})$ نقاط زمانی مشاهده‌شده برای آزمودنی iام هستند. دراین ساختار از یک پارامتر همبستگی مشترک استفاده می‌شود و فاصله نقاط زمانی مشاهدات عامل تفاوت همبستگی در موقعیت‌هاست. یعنی اگر $\rho = ۰/۸$ و اختلاف نقاط زمانی مشاهدات ۱ واحد است همبستگی برابر $۰/۸$ بوده و اگر این اختلاف به اندازه ۲ واحد باشد همبستگی برابر $۰/۸^۲ = ۰/۶۴$ است. در این ساختار با افزایش فاصله بین مشاهدات میزان همبستگی تنزل می‌یابد. علاوه بر ساختارهای ذکرشده، مدل‌های متعدد دیگری نیز وجود دارند که برای آشنایی بیش‌تر می‌توان به دیگل و همکاران (۲۰۰۲) و وربک و مولنبرگ (۲۰۰۹) مراجعه نمود. در ادامه نشان می‌دهیم همبستگی تاثیر بسزایی در کارایی برآوردگرها دارد.

یک روش ساده برای تحلیل داده‌های همبسته این است که همبستگی را نادیده گرفته و با استفاده از روش‌های کلاسیک به برآورد پارامترها بپردازیم. اما این کار به‌ویژه هنگامی‌که میزان همبستگی بالاست، موجب کاهش کارایی برآوردگرها می‌شود. به عنوان مثال در یک مدل رگرسیونی خطی ساده، میانگین همه مشاهدات و واریانس آن را درنظر بگیرید.

$$\begin{aligned}
\bar{Y} &= \frac{۱}{\sum_{i=۱}^{n} n_i} \sum_{i=۱}^{n} \sum_{j=۱}^{n_i} y_{ij}, \\
\text{Var}(\bar{Y}) &= \frac{۱}{(\sum_{i=۱}^{n} n_i)^۲} \sum_{i=۱}^{n} \Big[\sum_{j=۱}^{n_i} \text{Var}(y_{ij}) + ۲ \sum_{j=۱}^{n_i-۱} \sum_{k=j+۱}^{n_i} \text{Cov}(y_{ij}, y_{ik}) \Big].
\end{aligned}$$

اگر واریانس مشاهدات ثابت باشد، $\text{Var}(y_{ij}) = \sigma^۲$، آن‌گاه

$$\text{Var}(\bar{Y}) = \frac{\sigma^۲}{(\sum_{i=۱}^{n} n_i)^۲} \sum_{i=۱}^{n} \Big[n_i + \frac{۲}{\sigma^۲} \sum_{j=۱}^{n_i-۱} \sum_{k=j+۱}^{n_i} \text{Cov}(y_{ij}, y_{ik}) \Big].$$

حال اگر فرض کنیم طرح متوازن بوده، $n_i \equiv m$، و ساختار همبستگی تبادل‌پذیر است، $\rho_{jk} = \rho$، واریانس میانگین به‌صورت

$$\text{Var}(\bar{Y}) = \frac{\sigma^۲}{nm} \Big[۱ + (m-۱)\rho \Big].$$

محاسبه می‌شود. به مقدار $\Big[۱ + (m-۱)\rho \Big]$ عامل تورم واریانس گفته می‌شود؛ زیرا این مقدار واریانس میانگین را افزایش می‌دهد که به دلیل وجود همبستگی درون-گروهی است. برای نشان‌دادن تاثیر همبستگی روی واریانس میانگین، $\Big[۱ + (m-۱)\rho \Big]$ به ازای مقادیر مختلف m و ρ در جدول ۱.۱ آمده است.

جدول ۱.۱: مقدار $\left[1+(m-1)\rho\right]$ به ازای مقادیر مختلف m و ρ

m \ ρ	۰٫۰۰۱	۰٫۰۱	۰٫۰۲	۰٫۰۵	۰٫۱
۳	۱٫۰۰۱	۱٫۰۱	۱٫۰۲	۱٫۰۵	۱٫۱۰
۶	۱٫۰۰۴	۱٫۰۴	۱٫۰۸	۱٫۲۰	۱٫۴۰
۱۱	۱٫۰۰۹	۱٫۰۹	۱٫۱۸	۱٫۴۵	۱٫۹۰
۱۰۱	۱٫۰۹۹	۱٫۹۹	۲٫۹۸	۵٫۹۵	۱۰٫۹۰
۱۰۰۱	۱٫۹۹۹	۱۰٫۹۹	۲۰٫۹۸	۵۰٫۹۵	۱۰۰٫۹۰

مقادیر جدول ۱.۱ نشان می‌دهد که حتی اگر همبستگی درون آزمودنی‌ها بسیار ناچیز بوده اما تعداد تکرارهای هر نمونه زیاد باشد، میزان همبستگی تأثیر مهمی بر خطاهای استاندارد دارد. به عنوان مثال، در حالت‌های $(\rho = ۰٫۰۰۱, n = ۱۰۰۱)$، $(\rho = ۰٫۰۱, n = ۱۰۱)$ و $(\rho = ۰٫۱, n = ۱۱)$ مقدار $\left[1+(m-1)\rho\right]$ تقریباً برابر ۲ است. با توجه به مطالب بیان‌شده، واضح است که جهت رسیدن به استنباط علمی و معتبر، درنظر گرفتن این همبستگی امری ضروری است.

۳.۱ مروری بر مدل‌بندی داده‌های طولی

در بخش‌های قبل دلایلی بیان شد که روش‌های آماری کلاسیک که عمدتاً بر پایه استقلال مشاهدات بنا شده‌اند، به تنهایی برای تحلیل داده‌های طولی مناسب نیستند. بدین منظور آماردانان روش‌های متعددی را معرفی نموده‌اند که تمرکز اصلی این روش‌ها احتساب همبستگی درون آزمودنی‌ها و ارزیابی اثرات متغیرهای تبیینی روی پاسخ‌ها به‌طور همزمان است.

گسترش روش‌های اولیه در زمینه تحلیل داده‌های طولی که متغیر پاسخ آن پیوسته است، به حدود یک قرن پیش باز می‌گردد. پایه این روش‌ها توسط یک ستاره‌شناس انگلیسی به نام آیری (۱۸۶۱) با معرفی مدل تحلیل واریانس اثرات آمیخته[21] بنا شده است. این مدل که اغلب با عنوان تحلیل واریانس اندازه‌های مکرر تک‌متغیره[22] شناخته می‌شود دارای تاریخچه‌ای طولانی در زمینه تحلیل این داده‌هاست. از آن‌جایی‌که ساختار داده‌های طولی با n آزمودنی و n_i اندازه‌گیری مکرر، مشابه داده‌های جمع‌آوری‌شده در طرح بلوک تصادفی است، طبیعی به‌نظر می‌رسد که برای تحلیل این داده‌ها از روش‌های تحلیل واریانس استفاده شود. بدین منظور در این نوع مطالعات، آزمودنی‌ها به عنوان بلوک درنظر گرفته شده و مدل به‌صورت

$$y_{ij} = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + b_i + \varepsilon_{ij}, \quad i = ۱, \ldots, n; \qquad j = ۱, \ldots, n_i,$$

[21] Mixed effects anova

[22] Univariate repeated measures anova

نوشته می‌شود. در این مدل $b_i \sim N(۰, \sigma_b^۲)$ و $\varepsilon_{ij} \sim N(۰, \sigma_\varepsilon^۲)$ به‌ترتیب عامل تصادفی و جمله خطا بوده که از هم مستقل‌اند. در این مدل، بلوک‌ها یا اثرات طرح به عنوان اثرات تصادفی مورد توجه قرار می‌گیرند نه اثرات ثابت. اثر تصادفی، b_i، نشان‌دهنده تمام عوامل نامطلوب یا مشاهده‌نشده‌ای است که باعث شده آزمودنی‌ها رفتار یا پاسخ متفاوتی را نشان دهند. نتیجه گزینش این مولفه تصادفی در مدل، اعمال همبستگی مثبت به مشاهدات مکرر است به‌طوری‌که پاسخ‌ها دارای واریانس ثابت به‌صورت $Var(y_{ij}) = \sigma_b^۲ + \sigma_\varepsilon^۲$ و کوواریانس ثابت به‌صورت $Cov(y_{ij}, y_{ik}) = \sigma_b^۲$ هستند. مشابه همه مدل‌های رگرسیونی ε_{ij} نشان‌دهنده همه مواردی است که در مدل‌بندی درنظر گرفته نشده است.

اگر چه روش معرفی‌شده سابقه‌ای طولانی و گسترده در تحلیل داده‌های طولی دارد، فرض‌های محدودکننده و بسیاری از نقص‌های آشکار آن، سودمندی آن را در برنامه‌های کاربردی محدود کرده و این شیوه تنها در مواردی که طرح مطالعاتی بسیار ساده است، می‌تواند مبنایی منطقی برای تحلیل داده‌های طولی فراهم آورد. ویژگی‌های داده‌های طولی از قبیل طرح‌های نامتوازن و ناقص، وجود داده‌های گمشده و پاسخ‌های گسسته، که موجب محدودیت در استفاده از روش‌های تحلیل واریانس شده بود، انگیزه آماردانان را در راستای ارائه تکنیک‌های همه‌جانبه‌ای که دربرگیرنده این ویژگی‌ها باشد برانگیخت. گسترش تحلیل داده‌های طولی گسسته از ۳۰ تا ۳۵ سال پیش آغاز شده و تاکنون پیشرفت‌های قابل‌توجهی داشته است. این روش‌ها عمدتاً بر روی تحلیل رگرسیونی داده‌ها متمرکز شده‌اند. تحلیل رگرسیونی روشی ساده برای بررسی روابط تابعی میان متغیرهاست. این رابطه به شکل یک معادله یا الگویی است که متغیر پاسخ را به یک یا چند متغیر تبیینی مرتبط می‌کند. رگرسیون را بر حسب نوع تابعی که رابطه بین متغیرها را تعیین می‌کند، می‌توان به سه نوع پارامتری، ناپارامتری و نیمه‌پارامتری تقسیم نمود. رگرسیون پارامتری بیش‌تر وقتی مورد استفاده قرار می‌گیرد که اطلاعات تحقیق به قدر کافی زیاد است تا آن‌جایی‌که بتوان ارتباط بین متغیرهای پاسخ و تبیینی را به وسیله یک مدل خطی یا قابل تبدیل به خطی بیان کرد. این مدل‌ها با وجود سادگی در محاسبات و تفسیر، در مدل‌سازی بسیاری از روابط پیچیده با شکست مواجه می‌شوند؛ زیرا در کاربرد مثال‌های بسیاری را می‌توان یافت که شناسایی و تشخیص اثرات رگرسیون بر اساس فرضیات بیان‌شده امکان‌پذیر نیست. در چنین شرایطی تکنیک‌های مدل‌سازی ناپارامتری مورد نیاز است. رگرسیون ناپارامتری در مواقعی که نتوان نوع ارتباط بین متغیرها را با یک صورت تابعی معلوم و مشخص بیان کرد، مورد استفاده قرار می‌گیرد. در این مدل‌ها تنها فرض می‌شود که تابع مورد نظر یک تابع هموار است که برای برآورد آن می‌توان از روش‌های موضعی مانند اسپلاین و روش‌های هسته‌ای استفاده کرد. برآوردگرهای رگرسیون ناپارامتری بسیار انعطاف‌پذیر هستند اما با افزایش تعداد متغیرهای تبیینی، دقت آماری آن‌ها به شدت کاهش می‌یابد که این مسئله به مشقت بعد[۲۳] معروف است. در نتیجه محققین سعی کرده‌اند به مدل‌ها و برآوردگرهایی بپردازند که انعطاف‌پذیری بیش‌تری نسبت به رگرسیون پارامتری دارند و در عین حال بر مشکل مشقت بعد غلبه می‌کنند. چنین روش‌هایی معمولا خصوصیات روش‌های پارامتری و ناپارامتری را با هم ترکیب کرده و به روش‌های نیمه‌پارامتری معروف هستند. در این مدل‌ها فرض می‌شود بعضی از متغیرها دارای روند معلوم تغییرات نسبت به

[23] Curse of dimension

متغیر پاسخ بوده و برای بعضی از متغیرها نمی‌توان صورت تابعی خاصی را درنظر گرفت. در چند دهه اخیر، آماردانان در راستای تحلیل هرچه بهتر داده‌های طولی با خلق ترکیبی از مدل‌های ذکرشده، رده دیگری از مدل‌ها را ایجاد نمودند. در این میان، مدل اثرات آمیخته از مجموعه مدل‌های پارامتری و مدل‌های نیمه‌پارامتری مبنای بسیاری از این دسته مدل‌های ترکیبی قرار گرفتند. مدل نیمه‌پارامتری با اثرات آمیخته که تمرکز اصلی این رساله است، به عنوان مدلی بسیار منعطف و تنومند که ویژگی‌های خوب اکثر مدل‌ها را دارد، مورد توجه بسیار قرار گرفته است.

در ادامه ویژگی‌های هر یک از مدل‌های بیان‌شده و روش‌های برآورد در آن‌ها به تفصیل شرح داده شده است.

۱.۳.۱ مدل‌های پارامتری

هنگامی‌که علاقه‌مند به تحلیل داده‌های گسسته مانند دودویی، شمارشی، اسمی و ترتیبی باشیم، مدل‌های خطی در تحلیل این داده‌ها ناکارآمد می‌باشند. در این حالت مدل‌های خطی تعمیم‌یافته[۲۴] (نلدر و ودربرن ، ۱۹۷۲)، رده جامعی از مدل‌ها برای تحلیل رگرسیونی داده‌های گسسته و پیوسته که دارای توزیع‌های نمایی باشند، فراهم آورده است؛ به این ترتیب که با اعمال یک تبدیل غیرخطی مناسب روی میانگین متغیر پاسخ، ترکیبی خطی از متغیرهای تبیینی می‌سازد. بسط مدل‌های خطی تعمیم‌یافته به داده‌های همبسته و چندمتغیره، توانایی این مدل‌ها برای تحلیل داده‌های طولی را فراهم نموده است. سه رده گسترده از این مدل‌ها شامل مدل‌های حاشیه‌ای[۲۵]، مدل‌های اثرات آمیخته[۲۶] و مدل‌های انتقالی[۲۷] هستند. پاسخ به این سوال که کدام مدل باید انتخاب شود عمدتاً به سوال یا سوالات تحقیقات بستگی دارد که باید به آن‌ها جواب داده شود. لذا لازم است که با ویژگی‌های هر مدل آشنا شویم.

در مدل‌های حاشیه‌ای، مدل‌بندی بر حسب میانگین پاسخ انجام می‌گیرد و هیچ یک از اثرهای تصادفی یا پاسخ‌های گذشته در آن سهمی ندارند. علاوه بر این میانگین متغیر پاسخ و همبستگی بین اندازه‌های مکرر ثبت‌شده از یک آزمودنی جداگانه تحلیل می‌شود. وابستگی درون-فردی بین متغیرهای پاسخ با درنظر گرفتن ساختارهای مشخصی برای ماتریس همبستگی منظور می‌شود. انتخاب ساختارهای همبستگی مانند تبادل‌پذیر، غیرساختاری و اتو-رگرسیو به پیش‌فرض‌های موجود در مشاهدات و دلایل علمی بستگی دارد. روش دیگر استفاده از مدل اثرات تصادفی است. در این مدل فرض می‌شود همبستگی بین پاسخ‌های تکرار شده برای هر آزمودنی به دلیل ناهمگونی بین آزمودنی‌ها و وجود عوامل اندازه‌گیری نشده و غیرقابل کنترلی مانند عوامل محیطی و ژنتیکی است. بنابراین لازم است این عوامل به عنوان یک عامل تصادفی که از یک آزمودنی به آزمودنی دیگر متفاوت است درنظر گرفته شوند. بنابراین مدل اثرات تصادفی به دلیل دخالت ضرایب رگرسیونی

[24] Generalized linear models
[25] Marginal models
[26] Mixed effects models
[27] Transition models

تصادفی، متفاوت از مدل حاشیه‌ای عمل می‌کند. معمولاً مدل حاشیه‌ای که از طریق روش‌هایی مرسوم به معادلات برآوردیابی تعمیم‌یافته برازش می‌شود، با عنوان میانگین-جمعیتی[۲۸] شهرت دارد و مدل اثرات تصادفی به عنوان مدل‌های ویژه-آزمودنی[۲۹] در منابع ذکر شده است. در مدل‌های انتقالی که به عنوان سومین رویکرد تحلیل داده‌های طولی از آن یاد شده، فرض می‌شود که متغیر پاسخ در یک زمان مشخص به متغیرهای پاسخ در زمان‌های قبلی وابسته است. این تحلیل توسط اصول حاکم بر زنجیرهای مارکوف انجام می‌گیرد و همبستگی‌ها از طریق ارتباط با پاسخ‌های قبلی مورد بررسی قرار می‌گیرند.

۱.۱.۳.۱ مدل‌های حاشیه‌ای

ویژگی اصلی مدل‌های حاشیه‌ای این است که، تاثیر متغیرهای تبیینی روی متغیر پاسخ به‌طور مجزا از همبستگی بین پاسخ‌ها مدل‌بندی می‌شود. به عبارت دیگر، مدل‌های جداگانه‌ای برای میانگین حاشیه‌ای، واریانس حاشیه‌ای و همبستگی حاشیه‌ای درنظر گرفته می‌شوند. به‌طور معمول یک مدل حاشیه‌ای دارای سه مشخصه زیر است:

آ) میانگین حاشیه‌ای، $\mathrm{E}(y_{ij}) = \mu_{ij}$، از طریق تابع پیوند g به متغیرهای تبیینی $\boldsymbol{x}_{ij}$ وابسته است:

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta}, \quad i = ۱, ۲, \cdots, n; \quad j = ۱, ۲, \cdots, n_i,$$

که در آن g تابعی مشخص معروف به تابع پیوند مانند لگاریتم ($g(x) = \log(x)$) یا لوجیت ($g(x) = \log(\frac{x}{x-۱})$) است.

ب) واریانس حاشیه‌ای تابعی از میانگین حاشیه‌ای و پارامتر مقیاس ϕ است:

$$var(y_{ij}|\boldsymbol{x}_{ij}) = \nu(\mu_{ij})\phi,$$

که در آن ν یک تابع مشخص است.

پ) همبستگی بین y_{ij} و y_{ik}، تابعی از میانگین‌های حاشیه‌ای و پارامترهای دیگر مانند $\boldsymbol{\alpha}$ است:

$$corr(y_{ij}, y_{ik}) = \rho(\mu_{ij}, \mu_{ik}, \boldsymbol{\alpha})$$

که در آن ρ، تابعی معلوم و $\boldsymbol{\alpha}$، برداری از پارامترها است.

روش‌های متعددی برای برآورد پارامترهای مدل $(\boldsymbol{\beta}, \phi, \boldsymbol{\alpha})$ وجود دارند، که برآورد ماکزیمم درستنمایی[۳۰] (ML) و کمترین توان‌های دوم تعمیم‌یافته[۳۱] (GLS) از جمله متداول‌ترین آن‌هاست. برای داده‌هایی

[28] Population-average
[29] Subject-specific
[30] Maximum likelihood
[31] Generalized least square

با توزیع نرمال این دو روش معادل هستند. در این حالت با تعیین امید و واریانس حاشیه‌ای پاسخ‌ها، توزیع توام پاسخ‌ها نرمال چندمتغیره به‌دست می‌آید. در نتیجه می‌توان با استفاده از روش MLE، به برآورد پارامترهای مدل پرداخت. اما هنگامی‌که با داده‌های غیرنرمال سروکار داریم، محاسبه توزیع توام پاسخ‌ها امکان‌پذیر نبوده و این روش‌ها ناکارآمد هستند. در این حالت استفاده از روش معادلات برآورد تعمیم‌یافته[۳۲] (GEE) توصیه می‌گردد. این روش که توسط لیانگ و زیگر (۱۹۸۶) پیشنهاد شده‌است، یک روش برآورد در مدل‌های خطی تعمیم‌یافته به شمار می‌آید. این روش که تعمیمی از روش شبه‌درستنمایی (ودربرن ، ۱۹۷۴) به داده‌های چندمتغیره و همبسته است، دسته‌ای از معادلات برآورد برای پارامترهای مدل رگرسیونی ارائه می‌دهد. با فرض اینکه $\boldsymbol{\mu}_i(\boldsymbol{\beta})$ و $\boldsymbol{V}_i(\boldsymbol{\alpha})$ به‌ترتیب میانگین (گشتاور مرتبه اول) و ماتریس واریانس-کوواریانس (گشتاور مرتبه دوم) برای $\boldsymbol{y}_i$ باشند، معادلات برآورد تعمیم‌یافته به‌صورت

$$S(\boldsymbol{\beta},\boldsymbol{\alpha})=\sum_{i=1}^{n}S_i(\boldsymbol{\beta},\boldsymbol{\alpha})=\sum_{i=1}^{n}\frac{\partial\boldsymbol{\mu}_i^{\top}}{\partial\boldsymbol{\beta}}\boldsymbol{V}_i^{-1}(\boldsymbol{\alpha})\big(\boldsymbol{y}_i-\mu_i(\boldsymbol{\beta})\big)=\mathbf{0}$$

تعریف می‌شوند، که در آن

$$\frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\beta}}=\begin{pmatrix}\frac{\partial g^{-1}(\boldsymbol{x}_{i1}^{\top}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\\ \frac{\partial g^{-1}(\boldsymbol{x}_{i2}^{\top}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\\ \vdots\\ \frac{\partial g^{-1}(\boldsymbol{x}_{in_i}^{\top}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\end{pmatrix}=\begin{pmatrix}\frac{x_{1,i1}}{g'(\mu_{i1})} & \cdots & \frac{x_{p,i1}}{g'(\mu_{i1})}\\ \vdots & & \vdots\\ \frac{x_{1,in_i}}{g'(\mu_{in_i})} & \cdots & \frac{x_{p,in_i}}{g'(\mu_{in_i})}\end{pmatrix}_{n_i\times p}$$

و ماتریس واریانس-کوواریانس، $\boldsymbol{V}_i(\boldsymbol{\alpha})$، به‌صورت

$$\boldsymbol{V}_i(\boldsymbol{\alpha})=\phi\boldsymbol{A}_i^{\frac{1}{2}}\boldsymbol{R}_i(\boldsymbol{\alpha})\boldsymbol{A}_i^{\frac{1}{2}}$$

است که در آن، $\boldsymbol{A}_i$ یک ماتریس قطری $n_i\times n_i$ بعدی با عناصر قطری $\nu(\mu_{ij})$ بوده و $\boldsymbol{A}_i^{\frac{1}{2}}$ ریشه دوم ماتریس $\boldsymbol{A}_i$ است. همچنین $\boldsymbol{R}_i(\boldsymbol{\alpha})$ که به عنوان ماتریس همبستگی عملی[۳۳] شناخته می‌شود، ساختارهای متعددی دارند که در بخش‌های قبل چند نمونه ذکر شد. در تعیین این معادلات، به گشتاورهای مرتبه اول و دوم پاسخ‌ها و تعیین ماتریس همبستگی برای آزمودنی‌ها نیاز است؛ و درباره توزیع توام پاسخ‌ها هیچ فرضی اختیار نمی‌شود. برای داده‌هایی با توزیع نرمال این روش معادل روش ML یا GLS است. برای حل معادلات فوق از الگوریتم تکراری نیوتن-رافسون استفاده شده و برای برآورد بردار ضرایب رگرسیونی $\boldsymbol{\beta}$، رابطه تکراری به‌صورت

$$\widehat{\boldsymbol{\beta}}^{(r+1)}=\widehat{\boldsymbol{\beta}}^{(r)}+\Big(\sum_{i=1}^{n}\frac{\partial\mu_i^{\top}}{\partial\boldsymbol{\beta}}\boldsymbol{V}_i^{-1}(\widehat{\boldsymbol{\alpha}})\frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\beta}}\Big)^{-1}\Big(\sum_{i=1}^{n}\frac{\partial\boldsymbol{\mu}_i^{\top}}{\partial\boldsymbol{\beta}}\boldsymbol{V}_i^{-1}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{y}_i-\boldsymbol{\mu}_i)\Big)$$

[32] Generalized estimating equation

[33] Working correlation matrix

تا رسیدن به همگرایی محاسبه می‌شود. برای برآورد پارامترهای همبستگی (α) می‌توان از روش‌های گشتاوری استفاده کرد. پرنتیس (۱۹۸۸)، دسته‌ای دیگر از معادلات را به‌طور مجزا برای برآورد پارامترهای همبستگی پیشنهاد داده است که به عنوان تعمیم GEE شناخته می‌شوند. زائو و پرنتیس (۱۹۹۰) به برآورد توام پارامترهای رگرسیونی و همبستگی پرداخته‌اند که بسط درجه دوم GEE[۳۴] نام گرفته‌است. هر کدام از این روش‌ها، ویژگی، معایب و محاسن منحصربه‌فرد خود را دارند. شناخت کامل این روش‌ها و بحث درباره خصوصیات آن‌ها، یک زمینه گسترده مطالعاتی را بوجود می‌آورد که مستلزم صرف زمان بیش‌تری است. لذا در این رساله، تنها به ذکر چند ویژگی برجسته این روش بسنده می‌کنیم. علاقه‌مندان برای آشنایی بیش‌تر با روش‌های GEE، تعمیم و بسط درجه دوم آن می‌توانند به تحقیقات، فیتزموریس و لرد (۱۹۹۳)، سوترادهار و داس (۱۹۹۹) و سوترادهار و فارل (۲۰۰۴) مراجعه نمایند.

این روش برآوردگرهای سازگاری برای ضرایب مدل ارائه می‌دهد که مستقل از انتخاب ساختار همبستگی بوده و $\sqrt{n}(\hat{\boldsymbol{\beta}}_{GEE}-\boldsymbol{\beta})$ دارای توزیع مجانبی نرمال چند متغیره با میانگین $\mathbf{0}$ و ماتریس واریانس- کوواریانس به‌صورت زیر است:

$$\lim_{n\longrightarrow\infty} n\Big(\sum_{i=1}^{n}\boldsymbol{D}_i^{\top}\boldsymbol{V}_i^{-1}\boldsymbol{D}_i\Big)^{-1}\Big\{\sum_{i=1}^{n}\boldsymbol{D}_i^{\top}\boldsymbol{V}_i^{-1}\operatorname{Cov}(\boldsymbol{y}_i)\boldsymbol{V}_i^{-1}\boldsymbol{D}_i\Big\}\Big(\sum_{i=1}^{n}\boldsymbol{D}_i^{\top}\boldsymbol{V}_i^{-1}\boldsymbol{D}_i\Big)^{-1}.$$

اگر $\boldsymbol{V}_i=\operatorname{Cov}(\boldsymbol{y}_i)$ به درستی تشخیص داده شود، آن‌گاه رابطه فوق به $lim_{n\longrightarrow\infty} n\Big(\sum_{i=1}^{n}\boldsymbol{D}_i^{\top}\boldsymbol{V}_i^{-1}\boldsymbol{D}_i\Big)^{-1}$ تبدیل خواهد شد. اما عدم تشخیص درست نوع ساختار همبستگی موجب کاهش کارایی برآوردگرها می‌شود که این نقص در بسط درجه دوم GEE برطرف شده است.

گسترش این شیوه باعث شده است که بسیاری از نرم افزارهای آماری ازجمله R و SAS، به محیط و کدهای اختصاصی این روش مجهز شده و اجازه انتخاب ساختارهای همبستگی مختلف را به کاربران بدهند. در پایان لازم به ذکر است که در این نرم افزارها، از معیار اطلاع تعمیم‌یافته[۳۵] (GIC) و معیار اطلاع همبستگی[۳۶] (CIC) به عنوان معیارهایی برای تشخیص ساختار همبستگی مناسب، تشخیص کارایی برآوردگرها و نیکویی برازش مدل استفاده می‌گردد. این شاخص‌ها که توسط پن (۲۰۰۱) ارائه شده است، تعمیمی از معیار اطلاع آکائیک[۳۷] (AIC) است. علاقه‌مندان جهت آشنایی بیش‌تر با این شاخص‌ها می‌توانند به هاردین و هیلب (۲۰۰۳) و زیگلر و ونس (۲۰۱۰) رجوع نمایند.

[34] Second order GEE

[35] Generalized information criterion

[36] Correlation information criterion

[37] Akaike information criterion

۲.۱.۳.۱ مدل اثرات آمیخته

مدل اثرات آمیخته یا اثرات تصادفی[۳۸]، نوع دیگری از تعمیم مدل‌های خطی تعمیم‌یافته به‌شمار می‌آید. در این مدل‌ها عموماً فرض بر آن است که برخی ضرایب رگرسیونی برای آزمودنی‌ها به دلایل متفاوت به خود آزمودنی‌ها وابسته بوده و دارای توزیعی جداگانه می‌باشند. به عبارت دیگر همبستگی بین مشاهدات در داده‌های تکراری را می‌توان ناشی از داشتن یک تاثیر تصادفی یکسان در آن‌ها دانست به‌طوری‌که در مورد یک فرد اثر تصادفی یکسان و در فرد دیگر این اثر متفاوت است. در حقیقت همین ویژگی است که موجب ایجاد همبستگی در اندازه‌های مکرر هر آزمودنی می‌گردد. به عنوان مثال رگرسیون خطی ساده برای رشد کودک را درنظر بگیرید که در آن ضرایب مدل، بیانگر تاثیر وزن تولد و شاخص رشد است. واضح است که کودکان در زمان تولد، دارای وزن و شاخص رشد متفاوت با یکدیگر هستند. بنابراین لازم است این عوامل به عنوان یک عامل تصادفی که از یک آزمودنی به آزمودنی دیگر متفاوت است درنظر گرفته شوند. در این مدل، متغیر پاسخ به‌صورت تابعی از متغیرهای تبیینی با ضرایب ثابت و همچنین زیربرداری از آن‌ها با ضرایب تصادفی ارائه می‌گردد. این مدل که اولین بار توسط هارویل (۱۹۷۷) معرفی شد، برای متغیرهای پاسخ با توزیع نرمال به‌صورت

$$y_{ij} = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i + \varepsilon_{ij}, \quad i = ۱, ۲, \cdots, n; \quad j = ۱, ۲, \cdots, n_i,$$

بیان می‌گردد، که در آن y_{ij} و ε_{ij} به‌ترتیب متغیر پاسخ و خطای آزمودنی iام در jامین تکرار است. پارامترهای $\boldsymbol{\beta} = (\beta_۱, \ldots, \beta_p)^{\top}$ و $\boldsymbol{b}_i = (b_{i۱}, \ldots, b_{iq})^{\top}$، به‌ترتیب به عنوان اثرات ثابت و تصادفی شناخته می‌شوند. $\boldsymbol{x}_{ij} = (x_{۱,ij}, \ldots, x_{p,ij})^{\top}$ بردار متغیرهای تبیینی مربوط به اثرات ثابت بوده و $\boldsymbol{z}_{ij} = (z_{۱,ij}, \ldots, z_{q,ij})^{\top}$ برداری از متغیرهای تبیینی مربوط به اثرات تصادفی می‌باشند که با دخالت ضرایب تصادفی $\boldsymbol{b}_i$، ماهیت تصادفی به این بردار اعمال شده‌ا ست. برای هر آزمودنی iام مدل به‌صورت

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \quad i = ۱, ۲, \cdots, n$$

نمایش داده می‌شود، که در آن $\boldsymbol{y}_i = (y_{i۱}, \ldots, y_{in_i})^{\top}$ و $\boldsymbol{\varepsilon}_i = (\varepsilon_{i۱}, \ldots, \varepsilon_{in_i})^{\top}$ بردارهایی n_i بعدی هستند. ماتریس طرح اثرات ثابت به‌صورت

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{x}_{i۱}^{\top} \\ \vdots \\ \boldsymbol{x}_{in_i}^{\top} \end{pmatrix} = \begin{pmatrix} x_{۱,i۱} & x_{۲,i۱} & \cdots & x_{p,i۱} \\ x_{۱,i۲} & x_{۲,i۲} & \cdots & x_{p,i۲} \\ \vdots & \vdots & \ddots & \vdots \\ x_{۱,in_i} & x_{۲,in_i} & \cdots & x_{p,in_i} \end{pmatrix}_{n_i \times p}$$

[38] Random effects model

و ماتریس طرح اثرات تصادفی به‌صورت

$$\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{z}_{i۱}^\top \\ \vdots \\ \boldsymbol{z}_{in_i}^\top \end{pmatrix} = \begin{pmatrix} z_{۱,i۱} & z_{۲,i۱} & \cdots & z_{q,i۱} \\ z_{۱,i۲} & z_{۲,i۲} & \cdots & z_{q,i۲} \\ \vdots & \vdots & \ddots & \vdots \\ z_{۱,in_i} & z_{۲,in_i} & \cdots & z_{q,in_i} \end{pmatrix}_{n_i \times q}$$

است. اگر مدل شامل عرض از مبدا ثابت باشد، ستون اول ماتریس $\boldsymbol{X}_i$ برابر با ۱ بوده و اگر مدلی با عرض از مبدا تصادفی داشته باشیم عناصر ستون اول ماتریس $\boldsymbol{Z}_i$ برابر ۱ است. توجه داشته باشید که z_{ij} زیربرداری از $\boldsymbol{x}_{ij}$ است و اغلب $\boldsymbol{Z}_i$ نسبت به ماتریس $\boldsymbol{X}_i$ ستون‌های کمتری دارد یعنی $q \leqslant p$. در این مدل، فرض می‌شود که بردارهای ε_i و $\boldsymbol{b}_i$ مستقل از هم بوده و دارای توزیع نرمال به‌صورت $\varepsilon_i \sim \mathcal{N}_{n_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$ و $\boldsymbol{b}_i \sim \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{D}_i)$ است. ماتریس کوواریانس اثرات تصادفی است که عناصر روی قطر اصلی نشان‌دهنده واریانس هر عضو از بردار $\boldsymbol{b}_i$ بوده و عناصر غیرقطری، کوواریانس بین اثرات تصادفی است. از آنجایی‌که در این مدل q اثر تصادفی موجود است، ماتریس $\boldsymbol{D}_i$ یک ماتریس $q \times q$ بعدی متقارن و معین‌مثبت است که عناصر آن به‌صورت

$$\boldsymbol{D}_i = \mathrm{Var}(\boldsymbol{b}_i) = \begin{pmatrix} \mathrm{Var}(b_{i۱}) & \mathrm{Cov}(b_{i۱}, b_{i۲}) & \dots & \mathrm{Cov}(b_{i۱}, b_{iq}) \\ \mathrm{Cov}(b_{i۲}, b_{i۱}) & \mathrm{Var}(b_{i۲}) & \dots & \mathrm{Cov}(b_{i۲}, b_{iq}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(b_{iq}, b_{iq}) & \mathrm{Cov}(b_{iq}, b_{i۲}) & \dots & \mathrm{Var}(b_{iq}) \end{pmatrix}_{q \times q}$$

است. به‌طور مشابه ماتریس کوواریانس خطاها را می‌توان به‌صورت

$$\boldsymbol{\Sigma}_i = \mathrm{Var}(\boldsymbol{\epsilon}_i) = \begin{pmatrix} \mathrm{Var}(\epsilon_{i۱}) & \mathrm{Cov}(\epsilon_{i۱}, \epsilon_{i۲}) & \dots & \mathrm{Cov}(\epsilon_{i۱}, \epsilon_{in_i}) \\ \mathrm{Cov}(\epsilon_{i۲}, \epsilon_{i۱}) & \mathrm{Var}(\epsilon_{i۲}) & \dots & \mathrm{Cov}(\epsilon_{i۲}, \epsilon_{in_i}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\epsilon_{in_i}, \epsilon_{in_i}) & \mathrm{Cov}(\epsilon_{in_i}, \epsilon_{i۲}) & \dots & \mathrm{Var}(\epsilon_{in_i}) \end{pmatrix}_{n_i \times n_i}$$

نمایش داد. نمایش ماتریسی بر پایه همه داده‌ها را می‌توان به‌صورت

$$\begin{aligned} &\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \varepsilon, \\ &\boldsymbol{b} \sim \mathcal{N}_{nq}(\boldsymbol{0}, \boldsymbol{D}), \qquad \varepsilon \sim \mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma}), \end{aligned}$$

نوشت، که در آن

$$\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_n^\top)^\top, \qquad \boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_n^\top)^\top, \qquad \boldsymbol{b} = (\boldsymbol{b}_1^\top, \ldots, \boldsymbol{b}_n^\top)^\top,$$

$$\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_n^\top)^\top, \qquad \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{Z}_n \end{pmatrix},$$

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_1 & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_2 & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{D}_n \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{\Sigma}_n \end{pmatrix},$$

که $N = \sum_{j=1}^{n} n_i$. در این‌جا بعدهای کمیت‌ها به‌صورت

$$\boldsymbol{y} \sim N \times 1, \quad \boldsymbol{X} \sim N \times p, \quad \boldsymbol{\beta} \sim p \times 1, \quad \boldsymbol{Z} \sim N \times nq,$$
$$\boldsymbol{b} \sim nq \times 1, \quad \boldsymbol{\varepsilon} \sim N \times 1, \quad \boldsymbol{D} \sim nq \times nq, \quad \boldsymbol{\Sigma} \sim N \times N$$

هستند. از آن‌جایی‌که در داده‌های طولی مشاهدات مکرر از هر آزمودنی دارای همبستگی درونی بوده و از مشاهدات سایر آزمودنی‌ها مستقل‌اند، همبستگی بین پاسخ‌ها به‌صورت

$$\mathrm{Cov}(\boldsymbol{y}) = \begin{pmatrix} \mathrm{Cov}(\boldsymbol{y}_1) & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \mathrm{Cov}(\boldsymbol{y}_2) & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \mathrm{Cov}(\boldsymbol{y}_n) \end{pmatrix}_{N \times N}$$

قابل نمایش است به‌طوری‌که

$$\boldsymbol{V}_i = \mathrm{Cov}(\boldsymbol{y}_i) = \boldsymbol{Z}_i^\top \boldsymbol{D}_i \boldsymbol{Z}_i + \boldsymbol{\Sigma}_i.$$

واضح است که همبستگی بین پاسخ‌ها متاثر از تغییرات درون-گروهی ($\boldsymbol{Z}_i^\top \boldsymbol{D}_i \boldsymbol{Z}_i$) و برون-گروهی ($\boldsymbol{\Sigma}_i$) است. در این مدل حتی اگر ε_{ij}ها مستقل از هم فرض شوند؛ یعنی $\boldsymbol{\Sigma}_i = \boldsymbol{I}_i$، با این وجود همبستگی بین پاسخ‌ها به دلیل دخالت ضرایب تصادفی $\boldsymbol{b}_i$، قابل بیان است.

با اندکی تغییر در ساختار مدل و با فرض معلوم بودن مولفه‌های $\boldsymbol{\Sigma}_i$ و $\boldsymbol{D}_i$، می‌توان پارامترهای ثابت مدل را با روش‌های ساده رگرسیونی برآورد کرد. اگر مدل را به‌صورت

$$\boldsymbol{y} = \boldsymbol{X}^\top \boldsymbol{\beta} + \boldsymbol{\varepsilon}^*, \qquad \boldsymbol{\varepsilon}^* = \boldsymbol{Z}^\top \boldsymbol{b} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon}^* \sim \mathcal{N}_N(\boldsymbol{0}, \boldsymbol{V})$$

تغییر دهیم، به‌طوری‌که

$$\boldsymbol{V}=\begin{pmatrix}\boldsymbol{V}_1 & \mathbf{0} & \dots & \mathbf{0}\\ \mathbf{0} & \boldsymbol{V}_2 & \dots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{V}_n\end{pmatrix}_{N\times N}$$

آن‌گاه، برآورد پارامترهای ثابت به روش ML یا GLS به‌صورت

$$\widehat{\boldsymbol{\beta}}=\left(\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{y}$$

حاصل می‌شود. برای برآورد پارامترهای تصادفی مدل، می‌توان از بهترین برآوردگر خطی نااریب[39] ($BLUE$) استفاده کرد. این برآوردگر با محاسبه امید شرطی $\boldsymbol{b}_i$ به شرط $\boldsymbol{y}_i$ و $\widehat{\boldsymbol{\beta}}$ به‌صورت

$$\widehat{\boldsymbol{b}_i}=E[\boldsymbol{b}_i|\boldsymbol{y}_i]=\boldsymbol{D}\boldsymbol{Z}_i\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i^\top\widehat{\boldsymbol{\beta}})$$

حاصل می‌شود. برآوردگرهای حاصل با فرض معلوم بودن مولفه‌های $\boldsymbol{\Sigma}_i$ و $\boldsymbol{D}_i$ به‌دست آمدند؛ لذا این برآوردگرها در عمل ناکارآمد هستند. به منظور غلبه بر این مسئله استفاده از الگوریتم EM پیشنهاد می‌شود که در بخش‌های بعدی به تفصیل بحث خواهد شد.

تاکنون مطالب بیان شده مربوط به متغیرهای پاسخ با توزیع نرمال بود. برای داده‌هایی که توزیع غیرنرمال دارند از مدل اثرات آمیخته تعمیم‌یافته[40] استفاده می‌شود. به عنوان مثال پیش‌بینی تعداد خرابی که کمیتی گسسته است، یا زمان انتظار که کمیتی مثبت است را می‌توان به کمک مدل اثرات آمیخته تعمیم‌یافته انجام داد. این مدل از سه جز زیر تشکیل شده است.

آ) میانگین متغیرهای پاسخ به شرط ضرایب رگرسیونی به‌صورت

$$g[E(y_{ij}|\boldsymbol{b}_i)]=g(\mu_{ij})=\boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\boldsymbol{z}_{ij}^\top\boldsymbol{b}_i$$

است.

ب) متغیرهای پاسخ $y_{i1},y_{i2},\dots,y_{in_i}$ به شرط $\boldsymbol{b}_i$، دو به دو از هم مستقل بوده و دارای توزیعی از خانواده توزیع‌های نمایی به‌صورت

$$f(y_{ij}|\boldsymbol{b}_i)=\exp[\phi^{-1}\big(y_{ij}\theta_{ij}-\psi(\theta_{ij})\big)+c(y_{ij},\phi)],$$

هستند، که در آن ϕ پارامتر مقیاس است.

پ) اگر $g(.)$ تابع پیوند کانونی باشد، آن‌گاه گشتاورهای شرطی $\mu_{ij}=\mathrm{E}(y_{ij}|\boldsymbol{b}_i)=\psi'(\theta_{ij})$ و $\nu_{ij}=\mathrm{Var}(y_{ij}|\boldsymbol{b}_i)=\psi''(\theta_{ij})$، در روابط

$$\theta_{ij}=g(\mu_{ij})=\boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\boldsymbol{z}_{ij}^\top\boldsymbol{b}_i,\quad \nu_{ij}=\nu(\mu_{ij})\phi$$

[39] Best linear unbiased estimator

[40] Generalized mixed effects model

صدق می‌کنند.

بنابراین پاسخ‌های هر آزمودنی به شرط ضرایب تصادفی دوبه‌دو از هم مستقل و دارای توزیع‌های نمایی هستند. معمولاً فرض می‌شود بردار ضرایب تصادفی $\boldsymbol{b}_i$ دارای توزیع نرمال با میانگین صفر و ماتریس واریانس-کوواریانس $\boldsymbol{D}_i$ است. با این شرایط محاسبه توزیع توام پاسخ‌ها و استفاده از روش ML برای برآورد ضرایب ثابت $\boldsymbol{\beta}$ امکان‌پذیر است. بدین ترتیب اگر تابع چگالی y_{ij} به شرط $\boldsymbol{b}_i$ را با نماد $f_{y_{ij}}(y_{ij}|\boldsymbol{b}_i)$ نمایش دهیم، توزیع حاشیه‌ای $\boldsymbol{y}_i$ به‌صورت

$$f_{\boldsymbol{y}_i}(\boldsymbol{y}_i) = \int \prod_{j=1}^{n_i} f_{y_{ij}}(y_{ij}|\boldsymbol{b}_i) f(\boldsymbol{b}_i) d\boldsymbol{b}_i$$

محاسبه می‌شود. با فرض استقلال آزمودنی‌ها، تابع درستنمایی و لگاریتم آن به‌ترتیب، به‌صورت

$$L(\boldsymbol{\beta}, \phi) = \prod_{i=1}^{n} f_{\boldsymbol{y}_i}(\boldsymbol{y}_i) = \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f_{y_{ij}}(y_{ij}|\boldsymbol{b}_i) f(\boldsymbol{b}_i) d\boldsymbol{b}_i$$

$$l(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} log\{f_{\boldsymbol{y}_i}(\boldsymbol{y}_i)\} = \sum_{i=1}^{n} log\Big\{ \int \prod_{j=1}^{n_i} f_{y_{ij}}(y_{ij}|\boldsymbol{b}_i) f(\boldsymbol{b}_i) d\boldsymbol{b}_i \Big\}$$

است. با حل رابطه $\frac{\partial l(\boldsymbol{\beta},\phi)}{\partial \boldsymbol{\beta}} = 0$، می‌توان برآورد ML پارامترهای مدل را به‌دست آورد. صورت بسته برآوردگرها به توزیع $f_{y_{ij}}(y_{ij}|\boldsymbol{b}_i)$ وابسته است. این توزیع متناسب با نوع داده‌ها انتخاب می‌شود، به‌طوری‌که معمولاً در داده‌های پیوسته، دودویی و شمارشی به‌ترتیب توزیع‌های نرمال، دوجمله‌ای و پواسن و همچنین برای داده‌های اسمی و ترتیبی توزیع چندجمله‌ای درنظر گرفته می‌شود. معمولاً برآورد و تحلیل درباره $\boldsymbol{b}_i$ها از طریق روش‌های بیزی انجام می‌گیرد. همچنین به‌جای روش برآورد ML از روش‌های برآورد ماکزیمم درستنمایی مقید، ماکزیمم درستنمایی تاوانیده، شبه‌درستنمایی و بسیاری از روش‌های دیگر استفاده می‌شود. با توجه به این که بیان تمام موارد و شرح جزییات بیش‌تر منجربه دور شدن از بحث این رساله خواهد شد، لذا به مطالب اندک و مختصر شرح داده‌شده بسنده می‌کنیم. برای آشنایی بیش‌تر با این روش‌ها به تحقیقات استیراتلی و همکاران (۱۹۸۴)، هدیکر و گیبونز (۱۹۹۴)، لی و نلدر (۲۰۰۱)، توتز (۲۰۰۵) و لئون و همکاران (۲۰۰۷) رجوع کنید.

۳.۱.۳.۱ مدل‌های انتقالی

سومین مدل حاصل از تعمیم مدل‌های خطی تعمیم‌یافته، مدل‌های انتقالی هستند. مدل‌های انتقالی مدل‌هایی هستند که در آن‌ها متغیر پاسخ در زمان j به متغیرهای پاسخ در زمان‌های q $j-1,\ldots,j-q$ $(q < j-1)$ وابسته است. تحت این مدل، همبستگی بین پاسخ‌ها توسط تاثیر مستقیم مقادیر گذشته $y_{ij-1}, y_{ij-2}, \ldots, y_{i1}$ روی y_{ij} مورد بررسی قرار می‌گیرد و پاسخ‌های گذشته به عنوان متغیرهای تبیینی در مدل درنظر گرفته می‌شوند. این مدل‌ها که به عنوان مدل‌های اتورگرسیو و مارکوف نیز شناخته می‌شوند، ویژگی‌های عمومی به‌صورت زیر دارند:

آ) میانگین پاسخ‌ها به شرط پاسخ‌های گذشته به‌صورت

$$g[E(y_{ij}|\boldsymbol{H}_{ij})] = g[\mu_{ij}] = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \sum_{r=1}^{s} \kappa_r(\boldsymbol{H}_{ij}, \boldsymbol{\alpha})$$

مدل‌بندی می‌شوند، که در آن، $\boldsymbol{H}_{ij} = (y_{i1}, y_{i2}, \ldots, y_{ij-1})^{\top}$ برداری از مقادیر گذشته y_{ij}، $\kappa(.)$ تابعی معلوم و $\boldsymbol{\alpha}$ پارامترهای اضافی است. همچنین، g تابع پیوند است.

ب) متغیرهای پاسخ $y_{i1}, y_{i2}, \ldots, y_{in_i}$ به شرط $\boldsymbol{H}_{ij}$، دوبه‌دو از هم مستقل بوده و دارای توزیع‌های نمایی به‌صورت

$$f(y_{ij}|\boldsymbol{H}_{ij}) = \exp[\phi^{-1}\big(y_{ij}\theta_{ij} - \psi(\theta_{ij})\big) + c(y_{ij}, \phi)]$$

است.

پ) اگر $g(.)$ تابع پیوند کانونی باشد، آن‌گاه گشتاورهای شرطی $\mu_{ij} = \mathrm{E}(y_{ij}|\boldsymbol{H}_{ij}) = \psi'(\theta_{ij})$ و $\nu_{ij} = \mathrm{Var}(y_{ij}|\boldsymbol{H}_{ij}) = \psi''(\theta_{ij})$، در روابط

$$\theta_{ij} = g(\mu_{ij}) = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \sum_{r=1}^{s} \kappa_r(\boldsymbol{H}_{ij}, \boldsymbol{\alpha}), \quad \nu_{ij} = \nu(\mu_{ij})\phi$$

صدق می‌کنند.

یکی از معمول‌ترین مدل‌های انتقالی، مدل اتورگرسیو مرتبه q است. به عبارت دیگر، $(y_{ij}|\boldsymbol{H}_{ij})$ به q مشاهده قبلی خود یعنی $y_{ij-1}, y_{ij-2}, \ldots, y_{ij-q}$ وابسته است. به عنوان مثال برای پاسخ شمارشی، مدل به‌صورت

$$\log[E(y_{ij}|\boldsymbol{H}_{ij})] = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \sum_{r=1}^{q} \alpha_r y_{ij-r}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i$$

بیان می‌گردد. با توجه به این‌که طبق ویژگی دوم مدل‌های انتقالی، $(y_{ij}|\boldsymbol{H}_{ij})$ها دوبه‌دو از هم مستقل‌اند، پس از مشخص نمودن تابع $\kappa(.)$، می‌توان توزیع توام پاسخ‌های مربوط به هر آزمودنی را محاسبه نمود. به عنوان مثال برای مدل اتورگرسیو مرتبه q، توزیع $\boldsymbol{y}_i$ به‌صورت

$$f_{\boldsymbol{y}_i}(\boldsymbol{y}_i) = f(y_{i1}, y_{i2}, \ldots, y_{iq}) \prod_{j=1}^{n_i} f(y_{ij}|y_{ij-1}, y_{ij-2}, \ldots, y_{ij-q})$$

ارائه می‌گردد. سپس با فرض استقلال بین آزمودنی‌ها، تابع درستنمایی به‌صورت

$$\begin{aligned} L(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \prod_{i=1}^{n} f_{\boldsymbol{y}_i}(\boldsymbol{y}_i) \\ &= \prod_{i=1}^{n} \Big[f(y_{i1}, y_{i2}, \ldots, y_{iq}) \prod_{j=1}^{n_i} f(y_{ij}|y_{ij-1}, y_{ij-2}, \ldots, y_{ij-q}) \Big] \end{aligned}$$

است. محاسبه توزیع توام q مشاهده اول، یعنی $f(y_{i۱}, y_{i۲}, \ldots, y_{iq})$ نیازمند اعمال پذیره‌های اضافی و پیچیده در مدل است. در چنین شرایطی استفاده از روش ML شرطی پیشنهاد می‌گردد. در این روش تابع درستنمایی به‌صورت

$$\begin{aligned} L^c(\beta, \alpha) &= \prod_{i=۱}^{n} f(y_{iq+۱}, y_{ig+۲}, \ldots, y_{in_i}) \\ &= \prod_{i=۱}^{n} \prod_{j=q+۱}^{n_i} f(y_{ij}|y_{i۱}, y_{i۲}, \ldots, y_{iq}) \end{aligned}$$

محاسبه می‌شود. با شرطی کردن روی q مقدار اولیه از اطلاعات آن‌ها صرف‌نظر می‌کند. بنابراین کارایی کمتری نسبت به روش درستنمایی عادی خواهد داشت. با این وجود در صورتی که نخواهیم هیچ پذیره اضافی را در مدل اعمال کنیم، این روش مناسب است. علاقه‌مندان جهت آشنایی بیش‌تر با مباحث مدل‌های انتقالی و برآورد پارامترها، می‌توانند به تحقیقات رضایی و گنجعلی (۱۳۸۸)، مولنبرگ و وربک (۲۰۰۵)، چانگ و همکاران (۲۰۰۵)، سنگال و همکاران (۲۰۰۷)، رضایی و گنجعلی (۲۰۰۹a) و رضایی و گنجعلی (۲۰۰۹b) مراجعه کنند.

۲.۳.۱ مدل‌های ناپارامتری

روش‌های رگرسیون ناپارامتری برای داده‌های مستقل در سه دهه گذشته به خوبی توسعه یافته است که در این زمینه می‌توان به کارهای گرین و سیلورمن (۱۹۹۴) و هاردل و همکاران (۲۰۰۰) اشاره نمود. یک مدل ناپارامتری را می‌توان به‌صورت

$$E(y|X = x) = f(x)$$

بیان نمود، که در آن $f(.)$ یک تابع هموار نامعلوم است و باید بر اساس یک نمونه از مشاهدات (x_i, y_i) برآورد شود. در این‌جا منظور از هموار بودن، مشتق‌پذیری تابع $f(.)$ تا یک مرتبه خاص و کوچک است. تقریب این تابع رگرسیونی هموارسازی[۴۱] نامیده می‌شود. بنابراین یک هموارساز تابعی برای خلاصه‌سازی متغیر پاسخ y_i به عنوان تابعی از متغیر تبیینی x_i است. تاکنون روش‌های متعددی برای هموارسازی معرفی شده‌اند که عبارت‌اند از رگرسیون اسپلاین[۴۲]، هموارسازی اسپلاین[۴۳]، اسپلاین تاوانیده[۴۴] و هموارسازی چند جمله‌ای موضعی هسته‌ای[۴۵]. معرفی تمامی این روش‌ها به تنهایی نیازمند رونمایی بحث جدید و گسترده‌ای است که بیان آن‌ها خارج از گنجایش این رساله بوده و منجربه دورشدن از هدف اصلی خواهد شد. لذا در ادامه تنها به مرور اجمالی چند روش متداول خواهیم پرداخت.

[41] Smoothing
[42] Spline regression
[43] Spline smoothing
[44] Penalized spline
[45] Local polynomial kernel smoothing

یک روش رایج در رگرسیون ناپارامتری میانگین وزنی موضعی است. فرض کنید $x_\circ$ مقدار دلخواهی است که x می‌تواند اختیار کند. در این روش برای برآورد تابع $f(.)$ در نقطه $x_\circ$، به مقادیری از y_i که متناضر با مقادیر x_i نزدیکتر به نقطه $x_\circ$ باشند، وزن بیش‌تر و به مقادیر y_i متناظر با نقاط دورتر وزن کمتری نسبت می‌دهیم. وزن‌ها در رگرسیون ناپارامتری می‌توانند بوسیله یک تابع متقارن تک‌مدی حول صفر که مقادیر آن در دو طرف صفر توسط یک پارامتر مقیاس کنترل می‌شود، انتخاب شوند. نخستین نامزدهای آن‌ها برای چنین توابعی که هسته[۴۶] نامیده شدند، توابع چگالی احتمال بودند. سپس با استفاده از توابع کراندار هسته، برآوردگر

$$\widehat{f}(x_\circ) = \frac{1}{n}\sum_{i=1}^{n} W_i y_i, \qquad W_i = \frac{K_h(x_\circ, x_i)}{\sum_{i=1}^{n} K_h(x_\circ, x_i)},$$

را پیشنهاد دادند، که در آن $K_h(.)$ یک تابع هسته بوده و پارامتر h پهنای باند است. در این برآوردگر که به هموارساز هسته‌ای معروف است، شکل وزن‌ها توسط تابع هسته و مقادیر آن‌ها بر اساس فاصله از نقطه $x_\circ$ و پهنای باند h تعیین می‌شود. بزرگ بودن پهنای باند باعث می‌شود که به نقاط نزدیک‌تر وزن‌های بزرگتری داده شود. معمول‌ترین توابع هسته عبارت‌اند از هسته یکنواخت، هسته نرمال و هسته اپانچنیکوف.

یکی دیگر از متداول‌ترین روش‌های هموارسازی، اسپلاین‌ها هستند. اسپلاین توسط شخصی به نام شوئنبرگ در سال ۱۹۴۶ معرفی شد. وی برای نخستین بار قضیه‌ای برای اثبات وجود اسپلاین‌های درون‌یاب بیان کرد و سپس مطالعه چند جمله‌ای‌های قطعه به قطعه برای تابع اسپلاین آغاز شد. فاصله $[a,b]$ به‌صورت اجتماع دو مجموعه $[a,t]\cup[t,b]$ را درنظر بگیرید. یک اسپلاین از مرتبه یک تابعی است که در هر یک از بازه‌های مذکور یک تابع خطی و در نقطه t نیز پیوسته است. این تابع را می‌توان برای تابع پیوسته $f(.)$ به‌صورت

$$f(x) = \begin{cases} \beta_\circ + \beta_1 x & x \leqslant t \\ \beta'_\circ + \beta'_1 x & x > t \end{cases}$$

نوشت. با اعمال شرط پیوستگی تابع f در نقطه $x = t$ داریم

$$\begin{aligned} \beta_\circ + \beta_1 t &= \beta'_\circ + \beta'_1 t \\ \beta'_\circ &= \beta_\circ + \beta_1 t - \beta'_1 t. \end{aligned}$$

با جای‌گذاری $\beta'_\circ$ در ضابطه دوم f داریم

$$\beta'_\circ + \beta'_1 x = \beta_\circ + \beta_1 t - \beta'_1 t + \beta'_1 x = \beta_\circ + \beta_1 t + \beta'_1 (x - t).$$

لذا

$$f(x) = \begin{cases} \beta_\circ + \beta_1 x & x \leqslant t \\ \beta_\circ + \beta_1 t + \beta'_1 (x - t) & x > t \end{cases},$$

[46] Kernel

که می‌توان این تابع دو ضابطه‌ای را به‌صورت

$$f(x) = \beta_۰ + \beta_۱ x + \beta_۲ (x-t)_+$$

نوشت، که در آن $(x-t)_+ = \max(x-t, ۰) = (x-t)I(x>t)$، $I(.)$ تابع نشانگر و $\beta_۲ = \beta'_۱ - \beta_۱$ تفاوت بین شیب خطوط اول و دوم تابع دو ضابطه‌ای را بیان می‌کند. درواقع می‌توان گفت که تابع f، ترکیبی خطی از توابع پایه[۴۷]، ۱، x و $(x-t)_+$ است. اسپلاین مرتبه اول را اسپلاین خطی نیز می‌گویند.

حال فرض کنید که بازه‌ی $[a,b]$ به $K+۱$ زیر بازه به‌صورت

$$[a, t_۱), [t_۱, t_۲), \ldots, [t_K, b]$$

افراز شده باشد، به‌طوری‌که $a < t_۱ < \ldots < t_K < b$. نقاط $\{t_i\}_{i=۱}^K$ را گره[۴۸] می‌نامند. در این صورت اسپلاین مرتبه اول را می‌توان به‌صورت

$$f(x) = \beta_۰ + \beta_۱ x + \sum_{k=۱}^{K} \beta_{۱+k} (x-t_k)_+$$

بیان کرد که ترکیب خطی از توابع پایه ۱، x، $(x-t_۱)_+$، ... و $(x-t_K)_+$ است.

توابع اسپلاین برای مراتب بالاتر به‌طور مشابه تعریف می‌شود. تابع $f(.)$ را یک اسپلاین مرتبه d نامیم اگر در هر یک از زیر بازه‌های بیان‌شده یک چندجمله‌ای از مرتبه d باشد به‌طوری‌که $f(.)$ و مشتق‌های آن از مراتب ۱ تا $d-۱$، در گره‌ها پیوسته است. لذا تابع اسپلاین مرتبه d را می‌توان به‌صورت

$$f(x) = \beta_۰ + \beta_۱ x + \beta_۲ x^۲ + \ldots + \beta_d x^d + \sum_{k=۱}^{K} \beta_{d+k} (x-t_k)_+^d$$

بیان نمود، که در آن $(x-t_k)_+^d = (x-t_k)^d I(x>t_k)$ تابع توانی بریده‌شده[۴۹] از مرتبه d است. یعنی این تابع ترکیب خطی از یک چندجمله‌ای مرتبه d و توابع توانی بریده شده است. به عبارت دیگر مجموعه

$$۱, x, x^۲, \ldots, x^d, (x-t_۱)_+^d, \ldots, (x-t_K)_+^d$$

توابع پایه برای فضای اسپلاین‌های مرتبه d با گره‌های $t_۱$، $t_۲$، ... و t_K است. اگر $d=۳$، آن‌گاه تابع مورد اشاره را یک اسپلاین مکعبی[۵۰] می‌نامند. اسپلاین مکعبی یکی از متداول‌ترین نوع‌های اسپلاین است که به دلیل انعطاف‌پذیری بالا، در بسیاری از علوم مورد استفاده قرار می‌گیرد.

---

[47] Basis fuctions

[48] Knot

[49] Truncated power function

[50] Cubic spline

مدل رگرسیون ناپارامتری

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}$$

را درنظر بگیرید، که در آن

$$\begin{aligned}
\boldsymbol{y} &= (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^\top \\
\boldsymbol{f}(\boldsymbol{x}) &= \big(f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)\big)^\top \\
\boldsymbol{\varepsilon} &= (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n)^\top.
\end{aligned}$$

اگر تابع $f(.)$ را با تابع اسپلاین تعریف‌شده جایگزین کنیم، نتیجه حاصل را رگرسیون اسپلاین نامند، به‌طوری‌که منجر به برازش مدلی به‌صورت

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_d x_i^d + \sum_{k=1}^{K} \beta_{d+k}(x_i - t_k)_+^d + \varepsilon_i, \qquad i = 1, \ldots, n$$

می‌شود که در آن ϵ_iها در پذیره‌های زیربنایی رگرسیون صدق می‌کنند. با استفاده از نماد ماتریسی، رگرسیون اسپلاین به‌صورت

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

نمایش داده می‌شود که در آن $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{d+K})^\top$ و $\boldsymbol{B}$ یک ماتریس طرح به‌صورت

$$\boldsymbol{B} = \begin{pmatrix}
1 & x_1 & x_1^2 & \ldots & x_1^d & (x_1 - t_1)_+^d & (x_1 - t_2)_+^d & \ldots & (x_1 - t_K)_+^d \\
1 & x_2 & x_2^2 & \ldots & x_2^d & (x_2 - t_1)_+^d & (x_2 - t_2)_+^d & \ldots & (x_2 - t_K)_+^d \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_n & x_n^2 & \ldots & x_n^d & (x_n - t_1)_+^d & (x_n - t_2)_+^d & \ldots & (x_n - t_K)_+^d
\end{pmatrix}_{n \times (d+K+1)}$$

است. واضح است که درایه‌های ستون اول آن برابر ۱، درایه‌های d ستون بعدی مربوط به مشاهدات متغیرهای $\boldsymbol{x}$، $\boldsymbol{x}^2$، ... و $\boldsymbol{x}^d$ و درایه‌های ستون‌های $d+2$ تا $d+K+1$ مربوط به مشاهدات حاصل از متغیرهای $(\boldsymbol{x} - t_1)_+^d$، ... و $(\boldsymbol{x} - t_K)_+^d$ به ازای گره‌های از قبل تعیین شده t_1، t_2، ... و t_K هستند. لازم به ذکر است که رگرسیون اسپلاین متعلق به رده رگرسیون خطی است. به عبارت دیگر، تابع $f(.)$ نسبت به پارامترها خطی است و از این رو هزینه‌های محاسباتی برای برآورد منحنی رگرسیون مربوطه ناچیز است. بنابراین برازش یک رگرسیون اسپلاین روی داده‌ها را می‌توان با استفاده از روش‌های مرسوم انجام داد. با به‌کاربردن روش کمترین توان‌های دوم خطا برآوردگر

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \boldsymbol{y}$$

حاصل می‌شود. مشکلی که در این برآوردگر وجود دارد هم‌خطی بین پایه‌های توانی بریده‌شده

$$(\boldsymbol{x} - t_1)_+^d, (\boldsymbol{x} - t_2)_+^d, \ldots, (\boldsymbol{x} - t_K)_+^d$$

است که باعث عدم وارون‌پذیری یا بدشرطی[۵۱] ماتریس $\boldsymbol{B}^\top\boldsymbol{B}$ خواهد شد که ناپایداری ضرایب رگرسیونی را به همراه دارد. یک روش مناسب برای حل مشکل فوق این است که به جای پایه‌های توانی بریده‌شده از پایه‌های B-اسپلاین استفاده شود. بازه $[a,b]$ و تعداد K گره به‌صورت

$$a = t_۰ < t_۱ < \ldots < t_K < t_{K+۱} = b$$

را درنظر بگیرید. متداول‌ترین روش محاسبه پایه‌های B-اسپلاین مرتبه d، به وسیله پایه‌های B-اسپلاین مرتبه $d-۱$ و براساس رابطه بازگشتی

$$\boldsymbol{B}_{i,d}(\boldsymbol{x}) = (\frac{\boldsymbol{x}-t_i}{t_{i+d}-t_i})\boldsymbol{B}_{i,d-۱}(\boldsymbol{x}) + (\frac{t_{i+d+۱}-\boldsymbol{x}}{t_{i+d+۱}-t_{i+۱}})\boldsymbol{B}_{i+۱,d-۱}(\boldsymbol{x}), \quad i = -d, -d+۱, \ldots, K$$

است، که $t_{-d} = t_{-d+۱} = \ldots = t_{-۲} = t_{-۱} = t_۰ = a$ و $t_{K+۱} = b$. بنابراین تعداد پایه‌های B-اسپلاین مرتبه d برابر $d+K+۱$ است. با استفاده از این پایه‌ها می‌توان چندجمله‌ای‌های متعامد به‌دست آورد. توجه کنید که در این رساله به منظور سهولت در نوشتار از نمادگذاری توابع توانی بریده‌شده استفاده می‌شود اما در محاسبات مربوط به مطالعات شبیه‌سازی و تحلیل داده‌های واقعی از پایه‌های B-اسپلاین استفاده شده‌است. برای اطلاعات بیش‌تر درباره این توابع و انواع آن می‌توان به راپرت (۲۰۰۳) مراجعه کرد.

در تحلیل داده‌های طولی به روش ناپارامتری، معمولاً مشاهدات مکرر به‌صورت تابعی نامعلوم و یک ماهیت واحد و نه به‌صورت دنباله‌ای از مشاهدات گسسته ارزیابی می‌شوند. یعنی تغییرات پاسخ‌ها در طول زمان را نامعلوم درنظر می‌گیرند، بنابراین ساده‌ترین مدل رگرسیون ناپارامتری برای این داده‌ها به‌صورت

$$\boldsymbol{y}_i = \boldsymbol{f}(\boldsymbol{t}_i) + \boldsymbol{\varepsilon}_i, \quad i = ۱, \ldots, n \qquad (۱.۱)$$

بیان می‌شود. با وجود پیشرفت چشمگیر روش‌های رگرسیون ناپارامتری برای داده‌های مستقل، تحولات مهم در این زمینه برای داده‌های طولی تنها در سال‌های اخیر رخ داده است. ویژگی‌های خاص داده‌های طولی، چالش‌های عمده‌ای را در توسعه روش‌های ناپارامتری برای این داده‌ها به‌وجود آورده است. بنابراین لازم است روش‌های مرسوم جهت تحلیل داده‌های طولی توسعه و تعمیم یابند.

۱.۲.۳.۱ هموارسازی در داده‌های طولی با استفاده از روش هسته

ایده و پایه اصلی این روش بسط تیلور است، که در آن هر تابع هموار را می‌توان به کمک چند جمله‌ای‌هایی از درجه مشخص تقریب زد. حال با استفاده از بسط تیلور حول نقطه زمانی دلخواه $t_۰$، تابع $f(.)$ در مدل (۱.۱) را می‌توان به‌صورت

$$f(t_{ij}) \approx f(t_۰) + (t_{ij}-t_۰)f^{(۱)}(t_۰) + \ldots + (t_{ij}-t_۰)^d\frac{f^{(d)}(t_۰)}{d!} = \boldsymbol{T}^\top(t_{ij})\boldsymbol{\beta}$$

[51] Ill-condition

تقریب زد، که در آن $\boldsymbol{T}^\top(t_{ij}) = \big(۱, (t_{ij}-t_۰), \ldots, (t_{ij}-t_۰)^d\big)^\top$، $\boldsymbol{\beta} = \big(\beta_۰, \ldots, \beta_d\big)^\top$، $\beta_r = \frac{f^{(r)}(t_۰)}{r!}$ و $f^{(r)}(t_۰)$ مشتق rام تابع f در نقطه $t_۰$ است. جهت برآورد بردار پارامتر $\boldsymbol{\beta}$، معیار کمترین توان‌های دوم موزون[۵۲] (WLS) را به‌صورت

$$\sum_{i=۱}^{n} \boldsymbol{K}_h(\boldsymbol{t}_i - t)\{\boldsymbol{y}_i - \boldsymbol{T}(\boldsymbol{t}_i)\boldsymbol{\beta}\}^۲, \qquad (۲.۱)$$

درنظر بگیرید، به‌طوری‌که $\boldsymbol{K}_h(\boldsymbol{t}_i - t) = \big(K_h(t_{i۱} - t), \ldots, K_h(t_{in_i} - t)\big)$، $K_h(.)$ یک تابع هسته و $\boldsymbol{T}(\boldsymbol{t}_i) = \big(\boldsymbol{T}^\top(t_{i۱}), \ldots, \boldsymbol{T}^\top(t_{in_i})\big)^\top$. با مشتق گیری از رابطه (۲.۱)، برآورد $\boldsymbol{\beta}$ با حل معادلات برآورد هسته‌ای زیر بدست می‌آید.

$$\sum_{i=۱}^{n} \boldsymbol{T}^\top(\boldsymbol{t}_i)\boldsymbol{K}_h(\boldsymbol{t}_i - t)\{\boldsymbol{y}_i - \boldsymbol{T}^\top(\boldsymbol{t}_i)\boldsymbol{\beta}\} = \boldsymbol{0} \qquad (۳.۱)$$

به ازای $d = ۱$، برآوردگر به‌صورت $\widehat{f}(t_{ij}) = \frac{\sum_{i=۱}^{n}\sum_{j=۱}^{n_i} K_h(t_{ij}-t_۰)y_{ij}}{\sum_{i=۱}^{n}\sum_{j=۱}^{n_i} K_h(t_{ij}-t_۰)}$ حاصل می‌شود که همان برآوردگر هسته‌ای یا برآوردگر نادارایا-واتسون است.

به دلیل وجود همبستگی درون گروهی، برآوردهای حاصل از حل معادلات (۳.۱) برای داده‌های طولی ناکارآمد هستند. لین و کارول (۲۰۰۲) با تعمیم نسخه تعمیم‌یافته این روش به داده‌های غیرنرمال و با اعمال ساختار همبستگی، برآوردگر دیگری را به نام GEE چندجمله‌ای موضعی هسته‌ای[۵۳] (LPK) ارائه نمودند. بر این اساس یک مدل رگرسیون ناپارامتری تعمیم‌یافته با سه مشخصه اصلی

$$\text{آ.} \quad g[E(y_{ij}|t_{ij})] = g\big(\mu_{ij}(t_{ij})\big) = f(t_{ij}),$$
$$\text{ب.} \quad \mathrm{Var}(y_{ij}|t_{ij}) = \nu\big(\mu_{ij}(t_{ij})\big)\phi,$$
$$\text{پ.} \quad \mathrm{corr}(y_{ij}, y_{ik}) = \rho\big(\mu_{ij}(t_{ij}), \mu_{ik}(t_{ik}), \boldsymbol{\alpha}\big),$$

را درنظر بگیرید، که در آن $f(t_{ij}) = \boldsymbol{T}^\top(t_{ij})\boldsymbol{\beta}$ و $\boldsymbol{T}^\top(t_{ij}) = \big(۱, (t_{ij} - t_۰), \ldots, (t_{ij} - t_۰)^d\big)^\top$. معادلات برآوردیاب تعمیم‌یافته به‌صورت

$$\sum_{i=۱}^{n} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{t}_i)}{\partial \boldsymbol{\beta}} \boldsymbol{K}_h^{\frac{۱}{۲}}(\boldsymbol{t}_i - t)\boldsymbol{V}_i^{-۱}(\boldsymbol{\alpha})\boldsymbol{K}_h^{\frac{۱}{۲}}(\boldsymbol{t}_i - t)\big(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{t}_i)\big) = ۰ \qquad (۴.۱)$$

ارائه می‌گردد، که در آن

$$\boldsymbol{\mu}_i(\boldsymbol{t}_i) = \big(\mu_{i۱}(t_{i۱}), \ldots, \mu_{in_i}(t_{in_i})\big)^\top,$$
$$\mu_{ij}(t_{ij}) = g^{-۱}(\boldsymbol{T}^\top(t_{ij})\boldsymbol{\beta}),$$
$$\boldsymbol{V}_i(\alpha) = \phi \boldsymbol{A}_i^{\frac{۱}{۲}} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{A}_i^{\frac{۱}{۲}},$$
$$\boldsymbol{A}_i = \mathrm{diag}\,\big(\nu\big(\mu_{i۱}(t_{i۱}), \ldots, \nu\big(\mu_{in_i}(t_{in_i})\big),$$

[52] Weighted least squares

[53] Local polynomial kernel

و

$$\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{t}_i)}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial g^{-1}(\boldsymbol{T}^\top(t_{i1})\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ \frac{\partial g^{-1}(\boldsymbol{T}^\top(t_{i2})}{\partial \boldsymbol{\beta}} \\ \vdots \\ \frac{\partial g^{-1}(\boldsymbol{T}^\top(t_{in_i})}{\partial \boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} 1 & \frac{(t_{i1}-t_\circ)}{g'(\mu_{i1})} & \dots & \frac{(t_{i1}-t_\circ)^d}{g'(\mu_{i1})} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{(t_{in_i}-t_\circ)}{g'(\mu_{in_i})} & \dots & \frac{(t_{in_i}-t_\circ)^d}{g'(\mu_{in_i})} \end{pmatrix}_{n_i \times (d+1)}.$$

$\boldsymbol{R}_i(\boldsymbol{\alpha})$ ماتریس همبستگی عملی و $\boldsymbol{\alpha}$ بردار پارامترهای همبستگی هستند که به تفصیل در بخش ۱.۱.۳.۱ معرفی شدند. جواب حاصل از حل معادلات (۴.۱) یعنی $\widehat{\boldsymbol{\beta}}$ از طریق الگوریتم تکراری بیان‌شده در همان بخش به‌دست آمده و برآوردگر هسته‌ای تابع ناپارامتری به‌صورت $\widehat{f}(t_{ij}) = \boldsymbol{T}_{ij}^\top(t_\circ)\widehat{\boldsymbol{\beta}}$ حاصل می‌شود.

۲.۲.۳.۱ هموارسازی در داده‌های طولی با رگرسیون اسپلاین

همان‌طور که پیش از این نیز بیان شد، رگرسیون اسپلاین با استفاده از تعدادی گره یک رگرسیون پارامتری به کمک توابع پایه تشکیل می‌دهد. رایس و وو (۲۰۰۱) و هوانگ و ژو (۲۰۰۴) این روش را به داده‌های طولی تعمیم دادند. به این صورت که براساس اسپلاین مرتبه d و توابع توانی بریده‌شده، مجموعه توابع پایه در نقاط t_{ij} را به‌صورت $\{1, t_{ij}, t_{ij}^2, \dots, t_{ij}^d, (t_{ij}-t_1)_+^d, \dots, (t_{ij}-t_K)_+^d\}$ درنظر بگیرید که در مجموعه نقاط گره $\{t_1, \dots, t_K\}$ ایجاد شده‌اند. مشابه آن‌چه در ابتدای این بخش شرح داده‌شد، تابع $f(t_{ij})$ را می‌توان به‌صورت

$$f(t_{ij}) \approx \beta_\circ + \beta_1 t_{ij} + \dots + \beta_d t_{ij}^d + \sum_{k=1}^{K} (t_{ij}-t_k)_+^d \beta_{d+k} = \boldsymbol{B}^\top(t_{ij})\boldsymbol{\beta} \qquad (۵.۱)$$

تقریب زد، که $\boldsymbol{B}(t_{ij}) = \left(1, t_{ij}, \dots, t_{ij}^d, (t_{ij}-t_1)_+^d, \dots, (t_{ij}-t_K)_+^d\right)^\top$ و $\boldsymbol{\beta} = \left(\beta_\circ, \beta_1, \dots, \beta_{d+K}\right)^\top$. در این حالت، مدل (۱.۱) را می‌توان به‌صورت

$$\boldsymbol{y}_i = \boldsymbol{B}(\boldsymbol{t}_i)\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

بازنویسی کرد، به‌طوری‌که $\boldsymbol{B}(\boldsymbol{t}_i) = \left(\boldsymbol{B}^\top(t_{i1}), \dots, \boldsymbol{B}^\top(t_{in_i})\right)^\top$ ماتریس طرح $n_i \times (d+K+1)$ بعدی است. با درنظر گرفتن ماتریس کوواریانس $\boldsymbol{V}_i$ و استفاده از روش WLS، برآورد $\boldsymbol{\beta}$ با کمینه کردن رابطه

$$\sum_{i=1}^{n} \{\boldsymbol{y}_i - \boldsymbol{B}(\boldsymbol{t}_i)\boldsymbol{\beta}\}^\top \boldsymbol{V}_i^{-1} \{\boldsymbol{y}_i - \boldsymbol{B}(\boldsymbol{t}_i)\boldsymbol{\beta}\},$$

به‌صورت

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{B}^\top(\boldsymbol{t})\boldsymbol{V}^{-1}\boldsymbol{B}(\boldsymbol{t})\right)^{-1}\left(\boldsymbol{B}^\top(\boldsymbol{t})\boldsymbol{V}^{-1}\boldsymbol{y}\right)$$

حاصل می‌شود، به‌طوری‌که $\boldsymbol{B}(t) = \big(\boldsymbol{B}(\boldsymbol{t}_1), \ldots, \boldsymbol{B}(\boldsymbol{t}_n)\big)^\top$ و $\boldsymbol{V} = \mathrm{diag}(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n)$. به دنبال برآورد پارامتر رگرسیونی، برآورد تابع $f(t)$ به روش رگرسیون اسپلاین به‌صورت $\widehat{f}(t) = \boldsymbol{B}^\top(t)\widehat{\boldsymbol{\beta}}$ بدست می‌آید.

۳.۳.۱ مدل‌های نیمه‌پارامتری

همان‌گونه که پیش از این بیان شد، مدل‌های پارامتری در برخورد با تعدد متغیرهای تبیینی، کارایی خوبی دارند اما تحت تاثیر فرضیات پارامتری هستند. رگرسیون ناپارامتری از انعطاف‌پذیری بالایی برخوردار است، اما افزایش بعد فضای متغیرهای تبیینی استفاده از آن را با مشکل مواجه می‌کند. استون (۱۹۸۲) نشان داد که اگر مشتق تابع $f(.)$ تا مرتبه r موجود باشد، سریع‌ترین نرخ همگرایی قابل حصول برآوردگر $\widehat{f}(.)$ به $f(.)$ برابر $(\log n/n)^{r/(2r+p)}$ است. این نرخ همگرایی نقش بعد متغیرهای تبیینی را در کارایی برآوردگر نشان می‌دهد. به عبارت دیگر، افزایش p منجر به کاهش عملکرد برآوردگر $\widehat{f}(.)$ می‌شود. این مسئله به مشقت بعد معروف است که به دلیل تنک بودن داده‌ها در فضایی با ابعاد بالاتر اتفاق می‌افتد. برای اطلاعات بیش‌تر به گیننس (۲۰۱۱) مراجعه کنید.

برای اجتناب از اریبی زیاد مدل‌های پارامتری و مشقت بعد مدل‌های ناپارامتری، مدل‌های نیمه‌پارامتری به عنوان مدل‌هایی سازشگر و میانه‌رو که ویژگی‌های خوب هر دو مدل را دارد، پیشنهاد می‌شود. این مدل ساختار جمعی رگرسیون پارامتری را با انعطاف‌پذیری رگرسیون ناپارامتری ترکیب می‌کند. تاکنون مدل‌های رگرسیون نیمه‌پارامتری زیادی توسط افراد مختلف معرفی شده است، که رایج‌ترین آن‌ها شامل مدل خطی-جزئی[۵۴]، مدل ضریب متغیر[۵۵] و مدل تک‌شاخص[۵۶] است. تمرکز این رساله روی مدل‌های خطی-جزئی است. این مدل تفسیرپذیری مدل‌های خطی و انعطاف‌پذیری مدل‌های ناپارامتری را با هم ترکیب می‌کند و یکی از رایج‌ترین مدل‌های نیمه‌پارامتری است. انگل و همکاران (۱۹۸۶) از نخستین افرادی بودند که از این مدل برای بررسی رابطه بین دما و مصرف برق استفاده کردند. هاردل و همکاران (۲۰۰۰) بررسی‌های دقیقی از این مدل ارائه داده‌اند.

متغیر پاسخ y_{ij} مربوط به واحد iام در زمان t_{ij} را درنظر بگیرید. فرض کنید میانگین متغیر پاسخ به‌صورت خطی با متغیرهای تبیینی در ارتباط بوده و از طرف دیگر دارای روندی غیرخطی اما با صورت مشخص نسبت به متغیر زمان است. جهت هم‌ساز کردن مدل‌های رگرسیونی با روند غیرخطی بیان‌شده، دیگل و همکاران (۲۰۰۲) مدل نیمه‌پارامتری خطی-جزئی برای داده‌های طولی را به‌صورت

$$y_{ij} = \boldsymbol{x}_{ij}^\top \beta + f(t_{ij}) + \varepsilon_{ij}, \qquad i = 1, \ldots, n; \quad j = 1, \ldots, n_i \tag{۶.۱}$$

ارائه دادند، که در آن $\boldsymbol{x}_{ij} = (x_{1,ij}, x_{2,ij}, \cdots, x_{p,ij})^\top$ و $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^\top$، بردارهایی p بعدی از متغیرهای تبیینی، $f(.)$ تابعی هموار از زمان و ε_{ij} خطای تصادفی نرمال با $E\{\varepsilon_{ij}\} = 0$ است. در

[54] Partially linear models

[55] Varying coefficient model

[56] Single index model

ادامه به مرور کلی و کوتاه بر روش های متداول برآورد پارامترهای $\boldsymbol{\beta}$ و تابع ناپارامتری $f(.)$ خواهیم پرداخت.

۱.۳.۳.۱ برآورد به روش چند جمله‌ای موضعی

در این بخش به معرفی LPK جهت برازش مدل (۶.۱) خواهیم پرداخت. روش اول تعمیمی از اولین روش هسته‌ای اسپیکمن (۱۹۸۸) و بر پایه یک الگوریتم بازگشتی[۵۷] است. روش دوم بر پایه هموارسازی نیمرخ[۵۸] و تعمیمی از دومین روش هسته‌ای اسپیکمن (۱۹۸۸) است. هو و همکاران (۲۰۰۴) و وو و لیانگ (۲۰۰۴) این روش‌ها را در تحلیل داده‌های طولی به‌کار بردند و ما به تبعیت از آن‌ها، روش اول را LPK بازگشتی و روش دوم را LPK نیمرخ می‌نامیم.

مدل (۶.۱) را درنظر بگیرید. مدل ماتریسی که در آن عناصر بر اساس آزمودنی iام مشخص و بر اساس نقاط اندازه گیری ماتریسی شده‌اند و همچنین مدل ماتریسی کلی به‌ترتیب به‌صورت

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}_i) + \boldsymbol{\varepsilon}_i, \\ \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}) + \boldsymbol{\varepsilon}, \end{aligned}$$

نمایش داده می‌شوند، که در آن مولفه‌ها به‌صورت

$$\begin{cases} \boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^\top \\ \boldsymbol{X}_i = (\boldsymbol{x}_{i1}^\top, \ldots, \boldsymbol{x}_{in_i}^\top)^\top \\ \boldsymbol{f}(\boldsymbol{t}_i) = (f(t_{i1}), \ldots, f(t_{in_i}))^\top \\ \boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^\top \end{cases}, \qquad \begin{cases} \boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_n^\top)^\top \\ \boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_n^\top)^\top \\ \boldsymbol{f}(\boldsymbol{t}) = (\boldsymbol{f}^\top(\boldsymbol{t}_1), \ldots, \boldsymbol{f}^\top(\boldsymbol{t}_n))^\top \\ \boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_n^\top)^\top \end{cases}.$$

هستند.

روش LPK بازگشتی

مدل نیمه‌پارامتری معرفی‌شده با فرض معلوم بودن βها، به یک مدل ناپارامتری و با فرض معلوم بودن $f(.)$، به یک مدل پارامتری تبدیل می‌شود. براین اساس، الگوریتم بازگشتی دو مرحله‌ای زیر را که توسط هستی و تیبشیرانی (۱۹۹۰) معرفی شد، درنظر بگیرید.

الف) به ازای مقدار اولیه $\boldsymbol{\beta} = \boldsymbol{\beta}_0$، باقیمانده‌های $r_{ij} = y_{ij} - \boldsymbol{x}_{ij}^\top\boldsymbol{\beta}_0$ را به‌دست آورده و این مقادیر را به عنوان متغیر پاسخ جدید درنظر بگیرید. با درنظر گرفتن مدل ناپارامتری $r_{ij} = f(t_{ij}) + \varepsilon_{ij}$ و با استفاده از روش برآورد LPK، تابع ناپارامتری به‌صورت $\widehat{\boldsymbol{f}} = \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)$ برآورد می‌شود، به‌طوری‌که $\boldsymbol{S}$ ماتریس هموارساز است.

[57] Backfitting algorithm

[58] Profile smoothing

ب) به ازای $\widehat{\boldsymbol{f}}$ به‌دست آمده در مرحله (الف)، باقیمانده‌های $a_{ij} = y_{ij} - \widehat{f}(t_{ij})$ را به‌دست آورید. با درنظرگرفتن این مقادیر به عنوان پاسخ جدید، و درنظرگرفتن مدل پارامتری $a_{ij} = \boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\varepsilon_{ij}$، پارامترهای $\boldsymbol{\beta}$ را به روش ML برآورد کنید.

با تکرار دو مرحله فوق، در نهایت برآوردگرهای LPK بازگشتی برای $\boldsymbol{\beta}$ و $\boldsymbol{f}$ به‌ترتیب به‌صورت

$$\widehat{\boldsymbol{\beta}} = \big[\boldsymbol{X}^\top(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{X}\big]^{-1}\big[\boldsymbol{X}^\top(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{y}\big]$$
$$\widehat{\boldsymbol{f}} = \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}).$$

حاصل می‌شوند. شایان ذکر است که ماتریس هموارساز $\boldsymbol{S}$ بسته به نوع انتخاب توابع هسته، صورت‌های مختلفی دارد. علاقه‌مندان جهت آشنایی بیش‌تر می‌توانند به مراجع ذکر شده در بخش ۲.۳.۱ مراجعه کنند. همچنین در بخش ۴.۳.۱ به ذکر نمونه‌ای از این ماتریس خواهیم پرداخت.

روش LPK نیمرخی

با فرض معلوم بودن $\boldsymbol{\beta}$ و انتقال جمله پارامتری به طرف اول، مدل را به‌صورت $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{f} + \boldsymbol{\varepsilon}$ می‌نویسیم. برآورد تابع ناپارامتری به روش LPK به‌صورت $\widehat{\boldsymbol{f}} = \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ خواهد بود. با جای‌گذاری این مقدار به جای $\boldsymbol{f}$ در مدل اولیه و با اندکی محاسبات خواهیم داشت،

$$(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{y} = (\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

در نتیجه برآودگرهای نیمرخ برای $\boldsymbol{\beta}$ و $\boldsymbol{f}$ به‌ترتیب به‌صورت

$$\widehat{\boldsymbol{\beta}} = \big[\boldsymbol{X}^\top(\boldsymbol{I}_N - \boldsymbol{S})^\top(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{X}\big]^{-1}\big[\boldsymbol{X}^\top(\boldsymbol{I}_N - \boldsymbol{S})^\top(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{y}\big],$$
$$\widehat{\boldsymbol{g}} = \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}^\top\widehat{\boldsymbol{\beta}})$$

حاصل می‌شوند.

۲.۳.۳.۱ برآورد به روش رگرسیون اسپلاین

پایه‌های توانی بریده‌شده از مرتبه d در نقاط گره $t_1, \ldots, t_K$ را به‌صورت

$$\boldsymbol{B}(t_{ij}) = \big(1, t_{ij}, \ldots, t_{ij}^d, (t_{ij} - t_1)_+^d, \ldots, (t_{ij} - t_K)_+^d\big)^\top$$

و تقریب تابع $f(t_{ij})$ بر اساس این پایه‌ها را به‌صورت $f(t_{ij}) = \boldsymbol{B}^\top(t_{ij})\boldsymbol{\alpha}$ درنظر بگیرید، که در آن $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_h)^\top$ بردار ضرایب رگرسیونی و $h = 1 + d + K$ است. مدل نیمه‌پارامتری (۶.۱) را می‌توان به‌صورت

$$y_{ij} = \boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \boldsymbol{B}^\top(t_{ij})\boldsymbol{\alpha} + \varepsilon_{ij} = \tilde{\boldsymbol{x}}_{ij}^\top\boldsymbol{\theta} + \varepsilon_{ij}, \qquad i = 1, \ldots, n; \quad j = 1, \ldots, n_i$$

بیان نمود، به‌طوری‌که $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$، $\tilde{\boldsymbol{x}}_{ij} = \left(\boldsymbol{x}_{ij}^\top, \boldsymbol{B}^\top(t_{ij})\right)^\top$. برآودگر WLS پارامتر $\boldsymbol{\theta}$ به‌صورت

$$\widehat{\boldsymbol{\theta}} = \left(\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1} \tilde{\boldsymbol{X}}_i\right)^{-1} \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{y}_i,$$

حاصل می‌شود، که در آن $\tilde{\boldsymbol{X}}_i = (\tilde{\boldsymbol{x}}_{i1}^\top, \ldots, \tilde{\boldsymbol{x}}_{in_i}^\top)^\top$. بدیهی است که p مولفه اصلی در $\widehat{\boldsymbol{\theta}}$، برآورد مولفه‌های $\boldsymbol{\beta}$ یعنی $\widehat{\boldsymbol{\beta}}$ هستند.

۳.۳.۳.۱ برآورد به روش اسپلاین تاوانیده

مشابه روش رگرسیون اسپلاین، این روش نیز $f(t)$ را به کمک پایه‌ها به‌صورت $f(t) = \boldsymbol{B}^\top(t)\boldsymbol{\alpha}$ تقریب می‌زند. این روش میزان ناهمواری $f(t)$ را کنترل می‌کند، لذا برآوردگرهای $\boldsymbol{\beta}$ و $\boldsymbol{\alpha}$ از طریق کمینه کردن معیار کمترین توان‌های دوم تاوانیده زیر بدست می‌آید،

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left(y_{ij} - \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} - \boldsymbol{B}^\top(t_{ij})\boldsymbol{\alpha}\right)^2 + \lambda \boldsymbol{\alpha}^\top \boldsymbol{G} \boldsymbol{\alpha}$$

که در آن λ پارامتر هموارسازی و $\boldsymbol{G}$ ماتریس ناهمواری[59] است و به‌صورت

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{0}_{(d+1)\times(d+1)} & \boldsymbol{0}_{(d+1)\times K} \\ \boldsymbol{0}_{K\times(d+1)} & \boldsymbol{I}_K \end{bmatrix}_{h\times h}$$

تعریف می‌شود. صورت ماتریسی معیار فوق را می‌توان به‌صورت

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{B}\boldsymbol{\alpha}\|^2 + \lambda \boldsymbol{\alpha}^\top \boldsymbol{G} \boldsymbol{\alpha}$$

بیان نمود، به‌طوری‌که $\boldsymbol{B} = \left(\boldsymbol{B}_1, \ldots, \boldsymbol{B}_n\right)^\top$ و $\boldsymbol{B}_i = \left(\boldsymbol{B}^\top(t_{i1}), \ldots, \boldsymbol{B}^\top(t_{it_i})\right)^\top$. با انجام عملیات جبری، کمینه‌کننده‌ها به‌صورت

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top \boldsymbol{P} \boldsymbol{X})^{-1} (\boldsymbol{X}^\top \boldsymbol{P} \boldsymbol{y}), \\ \widehat{\boldsymbol{\alpha}} &= (\boldsymbol{B}^\top \boldsymbol{Q} B + \lambda \boldsymbol{G})^{-1} (\boldsymbol{B}^\top \boldsymbol{Q} \boldsymbol{y}), \\ \widehat{f}(t) &= \boldsymbol{B}^\top(t) \widehat{\boldsymbol{\beta}} \end{aligned}$$

بدست می‌آیند، که در آن $\boldsymbol{P} = \boldsymbol{I}_N - \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B} + \lambda \boldsymbol{G})^{-1} \boldsymbol{B}^\top$ و $\boldsymbol{Q} = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$.

علاقه‌مندان جهت کسب اطلاعات بیش‌تر درباره روش اسپلان‌های رگرسیونی و تاوانیده می‌توانند به تحقیقات هی و همکاران (۲۰۰۲)، هی و همکاران (۲۰۰۵) و فیتزموریس و همکاران (۲۰۰۹) مراجعه نمایند.

[59] Roughness matrix

۴.۳.۱ مدل نیمه‌پارامتری با اثرات آمیخته

در بخش‌های قبل مدل‌هایی که در تحلیل داده‌های طولی استفاده می‌شوند، اعم از مدل‌های پارامتری، ناپارامتری و نیمه‌پارامتری را معرفی نمودیم. در چند دهه اخیر آماردانان در راستای تحلیل هر چه بهتر این داده‌ها، با خلق ترکیبی از مدل‌های بیان‌شده، رده بسیار منعطف از مدل‌ها را ایجاد نمودند. در این میان، ادغام مدل اثرات آمیخته از مجموعه مدل‌های پارامتری با مدل‌های ناپارامتری بسیار مورد توجه قرار گرفت. بر این اساس ما نیز با درنظر گرفتن ترکیب دو مدل ذکرشده به مطالعه هر چه بیش‌تر مدل نیمه‌پارامتری با اثرات آمیخته پرداختیم. در این مدل اثرات پارامتری و ثابت به مدل‌بندی خطی متغیرهای تبیینی و تابع ناپارامتری به مدل‌بندی غیرخطی نقاط زمانی می‌پردازند. همچنین همبستگی درون موضوعی از طریق اثرات تصادفی، در مدل اعمال شده است. بنابراین یک مدل نیمه‌پارامتری با اثرات آمیخته به‌صورت

$$y_{ij} = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + f(t_{ij}) + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i + \varepsilon_{ij}, \qquad (۷.۱)$$

بیان می‌گردد. در مدل فوق y_{ij}، $(i = ۱, \ldots, n; j = ۱, \ldots, n_i)$، پاسخ مربوط به آزمودنی iام در زمان t_{ij} است، که در آن $\boldsymbol{\beta} = (\beta_۱, \ldots, \beta_p)^{\top}$ بردار $p \times ۱$ از ضرایب ثابت رگرسیونی وابسته به متغیرهای تبیینی $\boldsymbol{x}_{ij}$، $f(.)$ تابعی نامعلوم، هموار از زمان و دارای مشتق دوم، $\boldsymbol{b}_i = (b_{i۱}, \ldots, b_{iq})^{\top}$ بردار $q \times ۱$ بعدی از ضرایب تصادفی، مستقل از هم با میانگین ۰ و ماتریس کوواریانس $\boldsymbol{D}_i$ و وابسته به متغیرهای تبیینی $\boldsymbol{z}_{ij}$، و ε_{ij} خطای تصادفی و مستقل از $\boldsymbol{b}_i$ها با $\mathrm{E}\{\varepsilon_{ij}\} = ۰$ است. بدون کاستن از کلیت مسئله می‌توان فرض کرد که هر t_{ij} اعدادی در فاصله $[۰, ۱]$ می‌باشند.

مشابه کار فن و همکاران (۲۰۰۷)، فرض می‌کنیم که $\mathrm{Var}\{\varepsilon_{ij}\} = \sigma^{۲}(t_{ij})$، یعنی واریانس خطاهای تصادفی تابعی ناپارامتری و هموار نسبت به زمان هستند. از طرف دیگر فرض می‌شود تابع همبستگی بین $\varepsilon_i(t_{ij})$ و $\varepsilon_i(t_{ik})$ دارای صورت تابعی $\mathrm{corr}\{\varepsilon_{ij}, \varepsilon_{ik}\} = \rho(t_{ij}, t_{ik}, \boldsymbol{\theta})$ است، که در آن $\rho(t_{ij}, t_{ik}, \boldsymbol{\theta})$ تابعی معین مثبت از t_{ij} و t_{ik} بوده و $\boldsymbol{\theta}$ برداری از پارامترهای همبستگی است.

مدل ماتریسی که در آن عناصر بر اساس آزمودنی iام مشخص و بر اساس نقاط اندازه‌گیری ماتریسی شده‌اند به‌صورت

$$\boldsymbol{y}_i \;\; = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}_i) + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \qquad (۸.۱)$$

نمایش داده می‌شود، که در آن

$$\begin{cases} \boldsymbol{y}_i = \big(y_i(t_{i۱}), \ldots, y_i(t_{in_i})\big)^{\top} \sim n_i \times ۱ \\ \boldsymbol{X}_i = \big(\boldsymbol{x}_i^{\top}(t_{i۱}), \ldots, \boldsymbol{x}_i^{\top}(t_{in_i})\big)^{\top} \sim n_i \times p \\ \boldsymbol{f}(\boldsymbol{t}_i) = \big(g(t_{i۱}), \ldots, g(t_{in_i})\big)^{\top} \sim n_i \times ۱ \\ \boldsymbol{Z}_i = \big(\boldsymbol{z}_i^{\top}(t_{i۱}), \ldots, \boldsymbol{z}_i^{\top}(t_{in_i})\big)^{\top} \sim n_i \times q \\ \boldsymbol{\varepsilon}_i = \big(\varepsilon_i(t_{i۱}), \ldots, \varepsilon_i(t_{in_i})\big)^{\top} \sim n_i \times ۱. \end{cases}$$

فرض کنید $\boldsymbol{V}_i = \boldsymbol{Z}_i^\top \boldsymbol{D}_i \boldsymbol{Z}_i + \boldsymbol{\Sigma}_i$ ماتریس کوواریانس $n_i \times n_i$ بعدی مربوط به پاسخ $\boldsymbol{y}_i$ بوده که در آن $\boldsymbol{D}_i$ ماتریس کوواریانس $q \times q$ بعدی مربوط به $\boldsymbol{b}_i$ها و $\boldsymbol{\Sigma}_i$ ماتریس کوواریانس $n_i \times n_i$ بعدی مربوط به ε_iها است. به تبعیت از لیانگ و زیگر (۱۹۸۶) ما نیز $\boldsymbol{\Sigma}_i$ را به‌صورت $\boldsymbol{\Sigma}_i = \boldsymbol{A}_i \boldsymbol{R}_i(\boldsymbol{\theta}) \boldsymbol{A}_i$ درنظر گرفتیم، به‌طوری‌که $\boldsymbol{A}_i = \operatorname{diag}\left(\sigma(t_{i1}), \ldots, \sigma(t_{in_i})\right)$ یک ماتریس قطری $n_i \times n_i$ بعدی با عناصر قطری $\sigma(t_{ij})$ و $\boldsymbol{R}_i(\boldsymbol{\theta})$ ماتریس همبستگی بین $\varepsilon_i(t_{ij})$ و $\varepsilon_i(t_{ik})$ است که عضو (j,k)ام آن به‌صورت $\rho(t_{ij}, t_{ik}, \boldsymbol{\theta})$ تعریف می‌شود.

مدل (۸.۱) را می‌توان به‌صورت

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}) + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}$$

بازنویسی کرد، به‌طوری‌که

$$\begin{cases} \boldsymbol{y} = \left(\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_n^\top\right)^\top \sim N \times 1 \\ \boldsymbol{X} = \left(\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_n^\top\right)^\top \sim N \times p \\ \boldsymbol{f}(\boldsymbol{t}) = \left(\boldsymbol{f}^\top(\boldsymbol{t}_1), \ldots, \boldsymbol{f}^\top(\boldsymbol{t}_n)\right)^\top \sim N \times 1 \\ \boldsymbol{Z} = \operatorname{diag}\left(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n\right)^\top \sim N \times nq \\ \boldsymbol{\varepsilon} = \left(\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_n^\top\right)^\top \sim N \times 1 \\ \boldsymbol{b} = \left(\boldsymbol{b}_1^\top, \ldots, \boldsymbol{b}_n^\top\right)^\top \sim nq \times 1 \end{cases}$$

همچنین $\boldsymbol{V} = \operatorname{diag}\left(\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n\right) = \boldsymbol{Z}^\top \boldsymbol{D} \boldsymbol{Z} + \boldsymbol{\Sigma}$ ماتریس کوواریانس $N \times N$ بعدی مربوط به $\boldsymbol{y}$ بوده که در آن $\boldsymbol{D} = \operatorname{diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n)$ ماتریس کوواریانس $nq \times nq$ بعدی مربوط به بردار اثر تصادفی $\boldsymbol{b}$ و $\boldsymbol{\Sigma} = \operatorname{diag}\left(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_n\right)$ ماتریس کوواریانس $N \times N$ بعدی مربوط به ε است.

این مدل تعمیم مدل زیگر و دیگل (۱۹۹۴) است که جهت تحلیل داده‌های خود از یک مدل نیمه‌پارامتری با عرض از مبدأ تصادفی استفاده کردند و تابع ناپارامتری مدل را به روش بازگشتی برآورد نمودند.

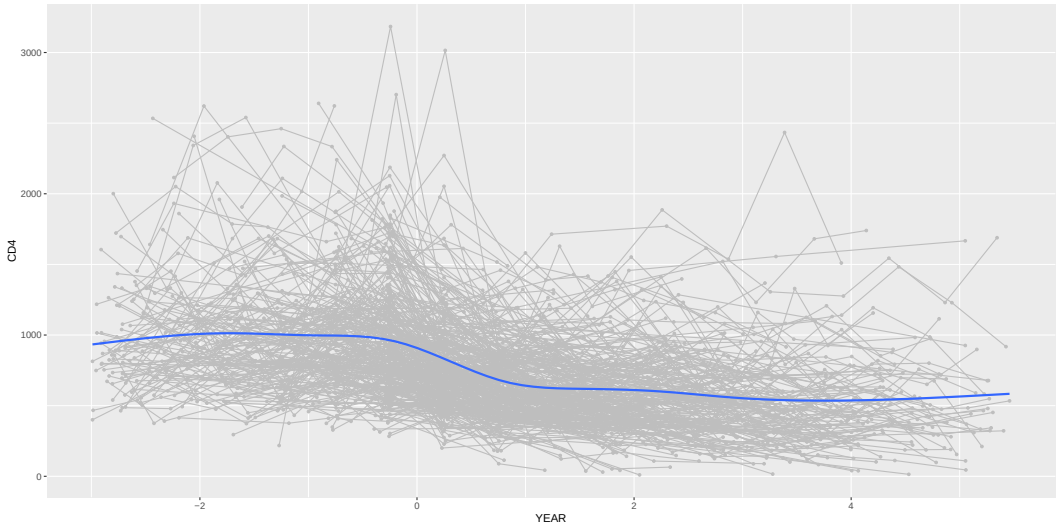
۴.۱ مثال‌های انگیزشی مورد بحث در این رساله

در این بخش مثال‌های واقعی را که در تحلیل مدل‌های ارائه شده در بخش‌های بعدی مورد استفاده قرار می‌گیرند، معرفی می‌کنیم.

۱.۴.۱ داده‌های $CD4$

این داده‌های مربوط به آزمایشات بالینی بیماری ایدز است که تعداد ۳۶۹ مرد مبتلا به HIV را مورد بررسی قرار می‌دهد. در این مثال متغیر پاسخ، تعداد سلول‌های $CD4$ و متغیرهای تبیینی شامل استفاده یا عدم‌استفاده از مواد مخدر $(DRUG)$، تعداد شرکای جنسی $(SEXP)$، تعداد بسته‌های

سیگار مورد استفاده در روز ($SMOKE$)، میزان افسردگی ($CESD$)، سن (AGE) و سال ($YEAR$) هستند. شکل ۳.۱ نمودار پراکنش تعداد سلول‌های $CD۴$ در طول زمان یعنی $YEAR$ را نشان می‌دهد. به منظور اینکه توزیع داده‌ها به توزیع نرمال نزدیک شود، تبدیل ریشه دوم متغیر پاسخ درنظر گرفته می‌شود. این داده‌ها در بسته catdata از نرم‌افزار R موجود است. جهت کسب اطلاعات بیش‌تر در مورد این داده‌ها می‌توان به زیگر و دیگل (۱۹۹۴) و وانگ و همکاران (۲۰۰۵) مراجعه نمود. زیگر و دیگل (۱۹۹۴) دریافتند که داده‌ها دارای همبستگی درونی از نوع ساختار همبستگی تبادل‌پذیر با پارامتر همبستگی $\rho = ۰/۵۰۹$ هستند.



شکل ۳.۱: نمودار پراکنش تعداد سلول‌های $CD۴$ در طول زمان (منحنی‌های خاکستری) به همراه میانگین هموارشده آن‌ها (منحنی آبی).

۲.۴.۱ داده‌های چرخه سلولی مخمر ORF

چرخه سلولی یک فرآیند زندگی کاملاً تنظیم‌شده است که در آن سلول‌ها رشد می‌کنند، DNAهایشان را تکثیر می‌دهند، کروموزوم‌های آن‌ها را جدا می‌کنند و تا جایی‌که محیط اجازه می‌دهد به تعدادی سلول دختر تقسیم می‌شوند. سلول‌های دختر سلول‌هایی هستند که از تقسیم یک سلول تک واحد حاصل می‌شود. روند چرخه سلولی معمولا به مراحل $G۱$، S، $G۲$ و M تقسیم می‌شوند. به‌طوری‌که مرحله $G۱$ مخفف GAP ۱ است. مرحله S مخفف عبارت سنتز (ترکیب یا تجزیه مواد برای تولید ماده یا مواد جدید) است که تکثیر DNA در آن رخ می‌دهد. مرحله $G۲$ مخفف GAP ۲ و مرحله M به معنای مرحله میتوز (تقسیم هسته یک سلول به دو هسته مشابه و روشی برای تکثیر سلولی) است. منظور از بیان‌ژن یعنی استفاده از آن ژن به منظور تولید RNA یا پروتئین است. RNA مخفف اسید ریبونوکلئیک[۶۰]، یک ترکیب پیچیده با وزن مولکولی بالاست که در ساختن پروتئین‌های سلولی نقش دارد. سه نوع اصلی RNA وجود دارند: RNA پیامبر ($mRNA$)، RNA ناقل ($tRNA$) و RNA

[60] Ribonucleic acid

ریبوزومی ($rRNA$). وقتی ژنی مورد استفاده قرار می‌گیرد، گفته می‌شود آن ژن بیان‌شده و روشن است و وقتی که یک ژن مورد استفاده قرار نگیرد، می‌گویند آن ژن خاموش است. این‌که در زمان مشخصی کدام ژن روشن و کدام ژن خاموش است به تنظیم بیان‌ژن معروف است. عوامل رونویسی[۶۱] (TFs) که نقش مهمی در تنظیمات بیان‌ژن دارند پروتئین‌هایی هستند که به دنباله‌های یک DNA خاص متصل می‌شود و از این طریق گردش اطلاعات ژنتیکی از DNA به $mRNA$ را کنترل می‌کند. تعداد این TFها زیاد است و با علائم اختصاری نمادگذاری می‌شوند که Mbp۱، Swi۴، Swi۶، Mcm۱، Fkh۱، Fkh۲، Ndd۱، Swi۵، و Ace۲ نمونه‌هایی از انواع این TFها هستند. برای درک بهتر پدیده اساسی فرآیند چرخه سلولی، مهم است که TFهایی که سطح بیان‌ژن ژن‌های تنظیم‌شده از چرخه سلولی را تنظیم می‌کنند، شناسایی شوند. بنابراین نیازمند استفاده از روش‌های انتخاب متغیر هستیم تا TFهای کلیدی شناسایی شوند. طبق مطالعات انجام‌شده توسط سیمون و همکاران (۲۰۰۱)، پروتئین‌های Swi۶، Swi۴، و Mbp۱ در مرحله G۱، پروتئین‌های McM۱، Ndd۱، و Fkh۱/۲ در مرحله G۲ و پروتئین‌های Ace۲، Swi۵، و Mcm۱ در مرحله M نقش مهمی دارند. داده بیان‌ژن چرخه سلولی مخمر[۶۲] از آزمایشات بیولوژیکی اسپیلمن و همکاران (۱۹۹۸) که روی ژن CDC۱۵ مخمر ORF انجام‌شده، جمع‌آوری شده‌اند. در این داده‌ها سطوح $mRNA$ سراسر ژنوم ۶۱۷۸ مخمر ORF در فواصل ۷ دقیقه و در مجموع به مدت ۱۱۹ دقیقه، شامل دو دوره چرخه سلولی و به تعداد ۱۸ مشاهده، اندازه‌گیری شده است. مجموعه داده‌ای که در این رساله استفاده می‌شود، زیر مجموعه‌ای از بیان‌ژن‌های چرخه سلولی مخمر ORF است که در آن چرخه سلولی ۲۸۳ ژن تنظیم‌شده CDC۱۵ در ۴ نقطه زمانی و در مرحله G۱ مشاهده شده‌اند را دربر می‌گیرد. در این داده‌ها متغیر پاسخ y_{ij} لگاریتم سطح بیان‌ژن مربوط به ژن iام در زمان j است. متغیرهای تبیینی $x_{k,ij}$ و $k = ۱,\ldots,۹۶$ نمره تطابق احتمال اتصال k TFام به ناحیه پیش‌برنده ژن iام است. احتمال اتصال با استفاده از یک روش مدل‌سازی ترکیبی بر اساس داده‌های حاصل از آزمایش اتصال $ChIP$ محاسبه می‌شود. این داده‌ها در بسته lme4 از نرم‌افزار R موجود است. برای آشنایی بیش‌تر با این مجموعه داده‌ها به وانگ و همکاران (۲۰۱۲) مراجعه کنید.

۳.۴.۱ داده‌های بیماری صفراوی

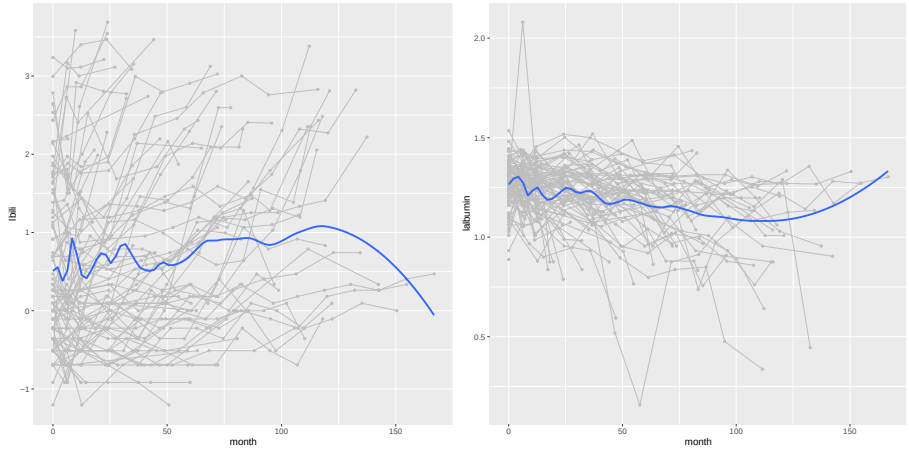
داده‌های بیماری صفراوی مربوط به مطالعه گروهی ۳۱۲ بیمار است که از ژانویه ۱۹۷۴ تا می ۱۹۸۴ به دلیل بیماری سیروز صفراوی اولیه پیاپی[۶۳] به کلینیک مایو[۶۴] که یک مرکز پزشکی-دانشگاهی در کشور آمریکا است، مراجعه کرده‌اند. بیماران شرکت‌کننده تا آوریل ۱۹۸۸ به‌طور تصادفی تحت

[61] Transcription factors
[62] Yeast cell-cycle gene expression
[63] Primary Biliary Cirrhosis sequential
[64] Mayo

سه نوع درمان شامل D-پنیسیلامین[۶۵]، دارونما[۶۶] و آزمایش کور[۶۷] قرار گرفته‌اند. آزمایش کور (نابینا) یا آزمایش چشمی یک آزمایش علمی است که افراد شرکت‌کننده در آن از داشتن اطلاعاتی که منجر شود به‌صورت خودآگاه یا ناخودآگاه به سمتی متمایل شوند، منع می‌شوند. به‌طور مثال وقتی از مصرف‌کننده پرسیده شود که طعم و مزه محصولات مختلف را با هم مقایسه کند، نام سازنده محصول باید پوشانده شود. در غیر این‌صورت معمولاً مصرف‌کننده به سمت محصولی که از قبل نسبت به آن شناخت دارد متمایل می‌شود. به همین شکل وقتی قدرت تأثیر یک دارو مورد آزمایش قرار می‌گیرد، هم بیماران و هم پزشکان شرکت‌کننده در آزمایش از دوز مصرفی بی‌اطلاع هستند. این باعث خواهد شد شانس هرگونه تأثیر تلقینی یا فریب هوشیارانه از بین برود. آزمایش کور ابزار بسیار مهم و کاربردی در شاخه‌های مختلف علوم از جمله علم پزشکی، روانشناسی، عصب‌شناسی و بیولوژی، علوم کاربردی و بازاریابی است. به‌کارگیری روش علمی آزمایش کور در برخی رشته‌ها مانند آزمون دارو ضروری است. مجموعه داده‌های $PBCseq$ شامل شماره شناسایی (ID) بیمار، پنج متغیر وابسته به زمان (مانند سن و تعداد روزهای تحت درمان)، هشت متغیر رسته‌ای (مانند جنسیت، نوع دارو و وضعیت بیماری) و هفت متغیر پیوسته (مانند لگاریتم بیلی‌روبین و آلبومین) است. این داده‌ها در بسته mixAK از نرم‌افزار R و همچنین در آدرس .http : //lib.stat.cmu.edu/datasets/pbcseq قابل دسترسی است. برای توضیحات بیش‌تر درباره داده‌ها به وانگ (۲۰۱۷) مراجعه کنید.

برای بیماران مبتلا به سیروز صفراوی اولیه پیشرفته، پیوند ارتوپدی کبد یک روش جایگزین برای درمان است. بیلی‌روبین و آلبومین دو شاخص اصلی برای ارزیابی وجود یا عدم وجود بیماری‌های کبدی هستند. اگر در صفرا و ادرار سطح بیلی‌روبین بسیار بالاتر از حد استاندارد باشد، نشان‌هنده وجود بیماری‌های خاصی است. سطح آلبومین بسیار زیاد یا خیلی کم نیز ممکن است برای افراد مضر باشد. اعتقاد بر این است که بین سطح بیلی‌روبین و آلبومین رابطه وجود دارد. بنابراین تحلیل مشترک این دو عامل که به‌طور مکرر در طول زمان جمع‌آوری شده‌اند در تشخیص بیماری‌های کبدی بسیار مفید خواهد بود. وانگ (۲۰۱۷) نشان داد که در این داده‌ها، مشاهدات نامتعارف یا پرت وجود دارد، همچنین اشاره کردند که تحلیل جداگانه این دو عامل باعث از دست رفتن بسیاری از اطلاعات مهم می‌شود. بنابراین وی روی مدل‌سازی توام دو متغیر جمع‌آوری شده در طول زمان، لگاریتم سطح بیلی‌روبین ($lbili$) و لگاریتم سطح آلبومین ($lalbumin$)، در حضور سایر متغیرهای مورد علاقه شامل سن (Age)، جنسیت (Sex) و نوع درمان ($Drug$) متمرکز شد. شکل ۴.۱ نمودار پراکنش $lbili$ و $lalbumin$ در طول زمان برای ۱۰۰ بیمار که به‌طور تصادفی انتخاب شده‌اند را نشان می دهد. اجرای چنین تحلیلی نیازمند استفاده از ابزار روش‌های چندمتغیره است و از آن‌جایی‌که دارای مقادیر پرت است باید روش‌هایی اتخاذ شود که نسبت به داده‌های پرت تنومند باشند. در فصل ۳ به تحلیل این داده‌ها پرداخته می‌شود.

[65] D-penicillamine
[66] placebo
[67] Double-blind



شکل ۴.۱: نمودار پراکنش داده‌های $PBCseq$ در طول زمان. سطوح $lbili$ و $lalbumin$ برای ۱۰۰ بیمار (منحنی‌های خاکستری) به همراه میانگین هموارشده آن‌ها (منحنی آبی).

۵.۱ روش‌های برآورد در مدل نیمه‌پارامتری با اثرات آمیخته

چالش اصلی در برازش مدل نیمه‌پارامتری با اثرات آمیخته، نحوه برخورد با مولفه ناپارامتری است که آماردانان بسیاری بسته به نوع مهارت و علاقه خود یکی از روش‌های ناپارامتری معرفی‌شده را به‌کار گرفتند. چن و چو (۲۰۰۷) و چن و چو (۲۰۰۹) از روش هموارسازی اسپلاین، چانگ (۲۰۰۴) و لی و چو (۲۰۱۰) از روش اسپلاین‌های تاوانیده و لیانگ (۲۰۰۹) از روش رگرسیون خطی موضعی استفاده کردند. از آنجایی‌که در بین روش‌های ناپارامتری معرفی‌شده روش‌های هسته‌ای و بازگشتی در این مدل مورد توجه قرار نگرفته‌اند، مطالعات خود را روی این روش‌ها به عنوان مطالعه‌ای جدید متمرکز نموده و نتایج حاصل در ادامه گزارش می‌شوند.

۱.۵.۱ روش برآوردگر هسته‌ای

مدل (۸.۱) را درنظر بگیرید و فرض کنید خطاهای تصادفی دارای توزیع نرمال با میانگین صفر هستند. در این صورت توزیع حاشیه‌ای زیر برای پاسخ $\boldsymbol{y}_i$ نتیجه می‌شود:

$$\boldsymbol{y}_i \sim \mathcal{N}_{n_i}\big(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}_i), \boldsymbol{Z}_i^\top \boldsymbol{D}_i \boldsymbol{Z}_i + \boldsymbol{\Sigma}_i\big). \qquad (۹.۱)$$

اگر امید ریاضی y_{ij} و $\boldsymbol{x}_{ij}$ به شرط نقاط زمانی t_{ij} را به‌ترتیب به‌صورت

$$\mu_Y(t_{ij}) = E\{y_{ij}|t = t_{ij}\}, \qquad \boldsymbol{\mu}_X(t_{ij}) = E\{\boldsymbol{x}_{ij}|t = t_{ij}\},$$

تعریف کنیم، آنگاه رابطه

$$\mu_Y(t_{ij}) = \boldsymbol{\mu}_X^\top(t_{ij})\boldsymbol{\beta} + f(t_{ij})$$

نتیجه می‌شود. با درنظر گرفتن صورت برداری امیدهای شرطی به‌صورت

$$\boldsymbol{\mu}_Y(\boldsymbol{t}_i) = \big(\mu_Y(t_{i1}), \ldots, \mu_Y(t_{in_i})\big)^\top, \qquad \boldsymbol{\mu}_X(\boldsymbol{t}_i) = \big(\boldsymbol{\mu}_X^\top(t_{i1}), \ldots, \boldsymbol{\mu}_X^\top(t_{in_i})\big)^\top,$$

رابطه

$$\boldsymbol{f}(\boldsymbol{t}_i) = \boldsymbol{\mu}_Y(\boldsymbol{t}_i) - \boldsymbol{\mu}_X(\boldsymbol{t}_i)\boldsymbol{\beta}$$

را خواهیم داشت. با جای‌گذاری جمله فوق در رابطه (۹.۱) داریم،

$$\boldsymbol{y}_i - \boldsymbol{\mu}_Y(\boldsymbol{t}_i) \sim \mathcal{N}_{n_i}\big([\boldsymbol{X}_i - \boldsymbol{\mu}_X(\boldsymbol{t}_i)]\boldsymbol{\beta}, \boldsymbol{Z}_i^\top \boldsymbol{D}_i \boldsymbol{Z}_i + \boldsymbol{\Sigma}_i\big).$$

به منظور برآورد ضرایب رگرسیونی $\boldsymbol{\beta}$، معیار WLS را به‌صورت

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} (\breve{\boldsymbol{y}}_i - \breve{\boldsymbol{X}}_i^\top \boldsymbol{\beta})^\top \boldsymbol{V}_i^{-1} (\breve{\boldsymbol{y}}_i - \breve{\boldsymbol{X}}_i^\top \boldsymbol{\beta}),$$

درنظر بگیرید که در آن $\breve{\boldsymbol{y}}_i = \boldsymbol{y}_i - \boldsymbol{\mu}_Y(\boldsymbol{t}_i)$ و $\breve{\boldsymbol{X}}_i = \boldsymbol{X}_i - \boldsymbol{\mu}_X(\boldsymbol{t}_i)$. معیار فوق به دلیل مجهول بودن توابع $\boldsymbol{\mu}_Y(.)$ و $\boldsymbol{\mu}_X(.)$ قابل استفاده نیست. از این رو ما روش هموارسازی هسته‌ای را درنظر گرفته و برآوردگر آن‌ها را به‌صورت

$$\widehat{\boldsymbol{\mu}}_X(t) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} \omega_{ij}\boldsymbol{x}_{ij}, \qquad \widehat{\boldsymbol{\mu}}_Y(t) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} \omega_{ij} y_{ij},$$

ارائه می‌دهیم، که در آن $\omega_{ij} = K_h(t_{ij}-t)/\sum_{k=1}^{n}\sum_{l=1}^{n_k} K_h(t_{kl}-t)$، h پهنای باند، $K_h(.) = K(./h)$ و $K(.)$ یک تابع هسته است. بنابراین با جای‌گذاری $\widehat{\boldsymbol{\mu}}_Y(t)$ و $\widehat{\boldsymbol{\mu}}_X(t)$ به‌ترتیب به جای $\boldsymbol{\mu}_Y(t)$ و $\boldsymbol{\mu}_X(t)$ معیار $l(\boldsymbol{\beta})$ به‌صورت

$$\widehat{l}(\boldsymbol{\beta}) = \sum_{i=1}^{n} (\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta})^\top \boldsymbol{V}_i^{-1} (\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta}) \qquad (۱۰.۱)$$

تقریب زده می‌شود، که در آن $\tilde{\boldsymbol{y}}_i = \boldsymbol{y}_i - \widehat{\boldsymbol{\mu}}_Y(\boldsymbol{t}_i)$، $\widehat{\boldsymbol{\mu}}_Y(\boldsymbol{t}_i) = \big(\widehat{\mu}_Y(t_{i1}), \ldots, \widehat{\mu}_Y(t_{in_i})\big)^\top$؛ $\tilde{\boldsymbol{X}}_i = \boldsymbol{X}_i - \widehat{\boldsymbol{\mu}}_X(\boldsymbol{t}_i)$ و $\widehat{\boldsymbol{\mu}}_X(\boldsymbol{t}_i) = \big(\widehat{\boldsymbol{\mu}}_X^\top(t_{i1}), \ldots, \widehat{\boldsymbol{\mu}}_X^\top(t_{in_i})\big)^\top$.

برآوردگرهای هسته‌ای $\boldsymbol{\beta}$ و $\boldsymbol{f}(\boldsymbol{t}_i)$ از طریق حل مسئله بهینه‌سازی $\arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \widehat{l}(\boldsymbol{\beta})$، به‌ترتیب به‌صورت

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_K &= \Big[\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1} \tilde{\boldsymbol{X}}_i\Big]^{-1}\Big[\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1} \tilde{\boldsymbol{y}}_i\Big], \\
\widehat{\boldsymbol{f}}(\boldsymbol{t}_i) &= \widehat{\boldsymbol{\mu}}_Y(\boldsymbol{t}_i) - \widehat{\boldsymbol{\mu}}_X^\top(\boldsymbol{t}_i)\widehat{\boldsymbol{\beta}}_K.
\end{aligned}$$

حاصل می‌شوند. صورت ماتریسی $\widehat{\boldsymbol{\beta}}_K$ و $\widehat{\boldsymbol{f}}(\boldsymbol{t}_i)$ عبارت‌اند از:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_K &= \Big[\boldsymbol{X}^\top(\boldsymbol{I}_N - \boldsymbol{S})^\top \boldsymbol{V}^{-1}(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{X}\Big]^{-1}\Big[\boldsymbol{X}^\top(\boldsymbol{I}_N - \boldsymbol{S})^\top \boldsymbol{V}^{-1}(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{y}\Big], \\
\widehat{\boldsymbol{f}} &= \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_K).
\end{aligned}$$

در روابط فوق، $\boldsymbol{S}$ یک ماتریس هموارساز است. لازم به ذکر است که برآوردگر معرفی‌شده همان برآوردگر LPK نیمرخی است، با این تفاوت که $\boldsymbol{S}$ به‌صورت

$$\boldsymbol{S} = \text{diag}\left(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_n\right), \qquad \boldsymbol{S}_i = \text{diag}\left(K_h(t_{i1} - t), \ldots, K_h(t_{in_i} - t)\right)$$

تعریف می‌شود. برآورد پارامترهای تصادفی مدل با محاسبه امید شرطی $\boldsymbol{b}_i$ به شرط $\boldsymbol{y}_i$ و $\widehat{\boldsymbol{\beta}}_K$ به‌صورت

$$\widehat{\boldsymbol{b}_i} = E[\boldsymbol{b}_i|\boldsymbol{y}_i] = \boldsymbol{D}_i\boldsymbol{Z}_i\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_K).$$

حاصل می‌شود.

لازم به ذکر است که برآوردگر هسته‌ای نسبت به پهنای باند حساس است، لذا تعیین پهنای باند در این روش از اهمیت بسزایی برخوردار است. روش‌های متعددی جهت انتخاب پهنای باند مناسب وجود داشته که بیان تمامی آن‌ها خارج از بحث این مجموعه است. لذا جهت آشنایی مختصر، به شرح تنها یک روش به نام روش اعتبارسنجی متقابل[۶۸] بسنده می‌کنیم. ایده این روش که توسط رادمو (۱۹۸۲) و بومن (۱۹۸۴) ارائه گردید، به این صورت است که پارامتر بهینه h به وسیله رابطه

$$\widehat{h} = \arg\min_h \, CV(h),$$

تعیین می‌گردد که در آن

$$CV(h) = \int \widehat{\boldsymbol{\mu}}_X^2(t)dt - \frac{2}{n}\sum_i \widehat{\boldsymbol{\mu}}_{(-i)X}(t_{ij}), \quad \widehat{\boldsymbol{\mu}}_{(-i)X}(t_{ij}) = \frac{1}{(n-1)h}\sum_{i'\neq i}\sum_j K\Big(\frac{t_{i'j} - t}{h}\Big).$$

تکنیک دیگری که در تحقیقات جدید از آن برای برآورد $\boldsymbol{\beta}$، $\boldsymbol{f}$ و $\boldsymbol{b}$ استفاده کرده ایم روش بازگشتی است که در بخش بعدی توضیح داده می‌شود.

۲.۵.۱ روش برآوردگر بازگشتی

با درنظر گرفتن رابطه (۹.۱)، روش برآوردگر بازگشتی به این صورت عمل می‌کند که با تکرار دو گام

الف) به ازای مقدار اولیه $\boldsymbol{\beta} = \boldsymbol{\beta}_0$، $\boldsymbol{f}(\boldsymbol{t}_i)$ را به‌صورت $\widehat{\boldsymbol{f}}(\boldsymbol{t}_i) = \widehat{\boldsymbol{\mu}}_Y(\boldsymbol{t}_i) - \widehat{\boldsymbol{\mu}}_X(\boldsymbol{t}_i)$ برآورد کنید.

ب) با اجرای رگرسیون $\boldsymbol{y}_i - \widehat{\boldsymbol{f}}(\boldsymbol{t}_i)$ روی $\boldsymbol{X}_i$، ضرایب رگرسیونی $\boldsymbol{\beta}$ را با کمینه کردن معیار زیر برآورد کنید:

$$\widehat{l}(\boldsymbol{\beta}) = \sum_{i=1}^{n}(\boldsymbol{y}_i - \widehat{\boldsymbol{f}}(\boldsymbol{t}_i) - \boldsymbol{X}_i\boldsymbol{\beta})^\top\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \widehat{\boldsymbol{f}}(\boldsymbol{t}_i) - \boldsymbol{X}_i\boldsymbol{\beta})$$

برآوردگر بازگشتی $\boldsymbol{\beta}$ و $\boldsymbol{f}(\boldsymbol{t}_i)$ به‌ترتیب به‌صورت

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{BF} &= \Big[\sum_{i=1}^{n}\boldsymbol{X}_i^\top\boldsymbol{V}_i^{-1}\tilde{\boldsymbol{X}}_i\Big]^{-1}\Big[\sum_{i=1}^{n}\boldsymbol{X}_i^\top\boldsymbol{V}_i^{-1}\tilde{\boldsymbol{y}}_i\Big],\\ \widehat{\boldsymbol{f}}(\boldsymbol{t}_i) &= \widehat{\boldsymbol{\mu}}_Y(\boldsymbol{t}_i) - \widehat{\boldsymbol{\mu}}_X(\boldsymbol{t}_i)\widehat{\boldsymbol{\beta}}_{BF}\end{aligned}$$

[68] Cross validation

حاصل می‌شوند. همچنین صورت ماتریسی $\widehat{\boldsymbol{\beta}}_{BK}$ و $\widehat{\boldsymbol{f}}(\boldsymbol{t}_i)$ عبارت است از:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{BF} &= \Big[\boldsymbol{X}^\top \boldsymbol{V}^{-1}(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{X}\Big]^{-1}\Big[\boldsymbol{X}^\top \boldsymbol{V}^{-1}(\boldsymbol{I}_N - \boldsymbol{S})\boldsymbol{y}\Big],\\ \widehat{\boldsymbol{f}} &= \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{BF}).\end{aligned}$$

برآورد پارامترهای تصادفی مدل با محاسبه امید شرطی $\boldsymbol{b}_i$ به شرط $\boldsymbol{y}_i$ و $\widehat{\boldsymbol{\beta}}_{BF}$ به‌صورت

$$\widehat{\boldsymbol{b}_i} = E[\boldsymbol{b}_i|\boldsymbol{y}_i] = \boldsymbol{D}_i\boldsymbol{Z}_i\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i^\top\widehat{\boldsymbol{\beta}}_{BF})$$

حاصل می‌شود.

لازم به ذکر است که تقریب مولفه ناپارامتری به روش اسپلاین‌ها بسیار متداول‌تر و محبوب‌تر از روش‌های موضعی قرار گرفته‌اند. زیرا این روش با تعداد کمی از گره‌ها تقریب خوبی را از یک تابع هموار فراهم می‌کند. همچنین با درنظر گرفتن توابع پایه به عنوان شبه‌متغیر، تابعی خطی از مولفه‌های ناپارامتری در اختیار قرار می‌دهد. علیرغم این که در بخش‌های قبل انواع آن را به تفصیل شرح دادیم، به منظور در اختیار داشتن مجموعه‌ای کامل از روش‌های برآورد در مدل نیمه‌پارامتری با اثرات آمیخته، در ادامه به بیان اجمالی این روش می‌پردازیم.

۳.۵.۱ روش برآوردگر اسپلاین

پایه‌های توانی بریده‌شده از مرتبه d در نقاط گره $t_1, \ldots, t_K$ را به‌صورت

$$\boldsymbol{B}(t) = \big(1, t, \ldots, t^d, (t-t_1)_+^d, \ldots, (t-t_K)_+^d\big)^\top$$

و تقریب تابع $f(t)$ بر اساس این پایه‌ها را به‌صورت $f(t) = \boldsymbol{B}^\top(t)\alpha$ درنظر بگیرید، که در آن $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_h)^\top$ بردار ضرایب رگرسیونی و $h = d + K + 1$ است. به این ترتیب مدل (۸.۱) را می‌توان به‌صورت مدل آمیخته

$$\boldsymbol{y}_i = \boldsymbol{X}_i^\top\boldsymbol{\beta} + \boldsymbol{B}_i^\top(\boldsymbol{t}_i)\boldsymbol{\alpha} + \boldsymbol{Z}_i^\top\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \qquad (۱۱.۱)$$

بیان نمود، که در آن $\boldsymbol{B}_i(\boldsymbol{t}_i) = \big(\boldsymbol{B}^\top(t_{i1}), \ldots, \boldsymbol{B}^\top(t_{in_i})\big)^\top$. برای سادگی و کاربردی بودن، مدل (۱۱.۱) را به‌صورت

$$\boldsymbol{y}_i = \tilde{\boldsymbol{X}}_i^\top\boldsymbol{\theta} + \boldsymbol{Z}_i^\top\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i,$$

بازنویسی می‌کنیم، به‌طوری‌که $\tilde{\boldsymbol{X}}_i = (\boldsymbol{X}_i, \boldsymbol{B}_i(\boldsymbol{t}_i))^\top$ ماتریس طرح ادغام‌شده $(p+h)\times n_i$ بعدی از ماتریس اثرات ثابت و اثرات اسپلاین و $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$ بردار پارامتر ادغام شده $(p+h)\times 1$ بعدی می‌باشند. با درنظر گرفتن توزیع حاشیه‌ای پاسخ‌ها به‌صورت

$$\boldsymbol{y}_i \sim N_{n_i}\big(\tilde{\boldsymbol{X}}_i\boldsymbol{\theta}, \boldsymbol{Z}_i^\top\boldsymbol{D}_i\boldsymbol{Z}_i + \boldsymbol{\Sigma}_i\big),$$

برآودگر WLS پارامتر $\boldsymbol{\theta}$ به‌صورت

$$\widehat{\boldsymbol{\theta}}_S = (\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1} \tilde{\boldsymbol{X}}_i)^{-1} \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{y}_i,$$

حاصل می‌شود. مشابه دو برآوردگر قبل برآورد ضرایب تصادفی مدل با محاسبه امید شرطی $\boldsymbol{b}_i$ به شرط $\boldsymbol{y}_i$ و $\widehat{\boldsymbol{\theta}}_S$ به‌صورت

$$\widehat{\boldsymbol{b}_i} = E[\boldsymbol{b}_i|\boldsymbol{y}_i] = \boldsymbol{D}_i \boldsymbol{Z}_i \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i^\top \widehat{\boldsymbol{\theta}}_S)$$

حاصل می‌شود.

۴.۵.۱ خواص مجانبی

در این بخش به بیان توزیع مجانبی برآوردگرهای هسته‌ای و بازگشتی می‌پردازیم. فرض کنید به ازای هر $n_i = m < \infty$، مجموعه سه‌تایی $(\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{t}_i)$، $i = 1, \ldots, n$ متغیرهایی مستقل و هم‌توزیع باشند؛ نقاط زمانی t_{ij} دارای چگالی حاشیه‌ای $f_j(t)$ باشند؛ مشتق rام تابعی مانند $d(.)$ را به‌صورت $d^{(r)}(.)$ تعریف می‌کنیم؛ عضو (j,k)ام ماتریس $\boldsymbol{V}^{-1}$ را به‌صورت ν^{jk} نشان می‌دهیم؛ تابع چگالی هسته $K(.)$ دارای میانگین ۰ و واریانس واحد است؛ یعنی $\int sK(s)ds = 0$ و $\int s^2 K(s)ds = 1$. همچنین فرض کنید $h \propto n^{-\alpha}$، $\frac{1}{5} \leqslant \alpha \leqslant \frac{1}{3}$ و $\tilde{\boldsymbol{X}} = \boldsymbol{X} + \lim_{n\to\infty} \frac{\partial \widehat{\boldsymbol{f}}(t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$. شرایط فوق در لین و کارول (۲۰۰۲) نیز استفاده شده‌اند.

قضیه ۱.۵.۱. تحت شرایط بیان‌شده، وقتی $n \to \infty$، آن‌گاه در رابطه با توزیع مجانبی برآوردگر هسته‌ای داریم

$$\sqrt{n}\Big\{\widehat{\boldsymbol{\beta}}_K - \boldsymbol{\beta} + h^2 \boldsymbol{b}_K(\boldsymbol{\beta}, \boldsymbol{f})/2\Big\} \xrightarrow{d} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{V}_{\mathbf{K}}),$$

که در آن

$$\begin{aligned}
\boldsymbol{b}_K(\boldsymbol{\beta}, \boldsymbol{f}) &= \mathrm{E}\Big\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \tilde{\boldsymbol{X}}\Big\}^{-1} \mathrm{E}\Big\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \boldsymbol{f}^{(2)}(\boldsymbol{t})\Big\}, \\
\boldsymbol{V}_K &= \mathrm{E}\Big\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \tilde{\boldsymbol{X}}\Big\}^{-1} \mathrm{E}\Big\{(\boldsymbol{W}_1 - \boldsymbol{W}_2)^\top \boldsymbol{V}_o (\boldsymbol{W}_1 - \boldsymbol{W}_2)\Big\} \mathrm{E}\Big\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \tilde{\boldsymbol{X}}\Big\}^{-1}, \\
\boldsymbol{V}_o &= \mathrm{Var}(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{t}), \qquad \boldsymbol{W}_1 = \boldsymbol{V}^{-1}\tilde{\boldsymbol{X}}, \qquad \boldsymbol{W}_2 = (\boldsymbol{W}_{21}, \ldots, \boldsymbol{W}_{2m}) \\
\boldsymbol{W}_{2j} &= \frac{\Big\{\sum_{k=1}^{m}\sum_{l=1}^{m} E\big(\tilde{\boldsymbol{X}}_k \nu^{kl} | \boldsymbol{t}_l = \boldsymbol{t}_j\big)\Big\} f_j(\boldsymbol{t}_j)}{\sum_{l=1}^{m} f_l(\boldsymbol{t}_j)}
\end{aligned}$$

اثبات قضیه در پیوست آمده است.

قضیه ۲.۵.۱. تحت شرایط بیان‌شده، وقتی $n \longrightarrow \infty$، آن‌گاه در رابطه با توزیع مجانبی برآوردگر بازگشتی داریم

$$\sqrt{n}\Big\{\widehat{\boldsymbol{\beta}}_{BF} - \boldsymbol{\beta} + h^2 \boldsymbol{b}_{BF}(\boldsymbol{\beta}, \boldsymbol{f})/2\Big\} \xrightarrow{d} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{V}_{BF}),$$

که در آن

$$\begin{aligned}
\boldsymbol{b}_{BF}(\boldsymbol{\beta},\boldsymbol{f}) &= \mathrm{E}\left\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}\tilde{\boldsymbol{X}}\right\}^{-1}\mathrm{E}\left\{\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{f}^{(2)}(\boldsymbol{t})\right\},\\
\boldsymbol{V}_{BF} &= \mathrm{E}\left\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}\tilde{\boldsymbol{X}}\right\}^{-1}\mathrm{E}\left\{(\boldsymbol{W}_1^*-\boldsymbol{W}_2^*)^\top \boldsymbol{V}_o(\boldsymbol{W}_1^*-\boldsymbol{W}_2^*)\right\}\mathrm{E}\left\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}\tilde{\boldsymbol{X}}\right\}^{-1},\\
\boldsymbol{W}_1^* &= \boldsymbol{V}^{-1}\boldsymbol{X}, \qquad \boldsymbol{W}_2^* = (\boldsymbol{W}_{21}^*,\ldots,\boldsymbol{W}_{2m}^*),\\
\boldsymbol{W}_{2j}^* &= \frac{\left\{\sum_{k=1}^m\sum_{l=1}^m E\left(\boldsymbol{X}^k\nu^{kl}|\boldsymbol{t}_l=\boldsymbol{t}_j\right)\right\}f_j(\boldsymbol{t}_j)}{\sum_{l=1}^m f_l(\boldsymbol{t}_i)}.
\end{aligned}$$

اثبات این قضیه مشابه قضیه ۲ در هو و همکاران (۲۰۰۴) است.

قضیه ۳.۵.۱. تحت شرایط نظم بیان‌شده، ماتریس کوواریانس برآوردگر هسته‌ای حداکثر به بزرگی برآوردگر بازگشتی است. به عبارت دیگر ماتریس $\boldsymbol{V}_{BF}-\boldsymbol{V}_K$ یک ماتریس نیمه‌معین مثبت است.

قضیه فوق به مقایسه عملکرد دو برآوردگر هسته‌ای و بازگشتی می‌پردازد و بیان می‌کند که در داده‌های طولی برآوردگر هسته‌ای همواره بر برآوردگر بازگشتی برتری دارد. اثبات قضیه در پیوست آمده است.

۵.۵.۱ الگوریتم تکراری

در تمامی روش‌های معرفی‌شده، کارایی برآورد ضرایب رگرسیونی به برآورد تابع کوواریانس وابسته است. بنابراین برآورد نهایی را از طریق به‌روزکردن $\boldsymbol{\beta}$، $\boldsymbol{b}_i$ و $\boldsymbol{f}(\boldsymbol{t}_i)$ نسبت به برآورد ماتریس کوواریانس و با به‌کارگیری یک الگوریتم تکراری بین آن‌ها به‌دست می‌آوریم. ابتدا برآوردگرهای به‌روزشده $\boldsymbol{\beta}$، $\boldsymbol{b}_i$ و $\boldsymbol{f}(\boldsymbol{t}_i)$ نسبت به برآورد ماتریس کوواریانس برای روش هسته‌ای را به‌صورت

$$\widehat{\boldsymbol{\beta}}_K^{(r)} = \Big(\sum_{i=1}^n \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1(r-1)}\tilde{\boldsymbol{X}}_i\Big)^{-1}\Big(\sum_{i=1}^n \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1(r-1)}\tilde{\boldsymbol{y}}_i\Big), \qquad (۱۲.۱)$$

$$\begin{aligned}
\widehat{\boldsymbol{f}}^{(r)}(\boldsymbol{t}_i) &= \widehat{\boldsymbol{\mu}}_Y(\boldsymbol{t}_i) - \widehat{\boldsymbol{\mu}}_X(\boldsymbol{t}_i)\widehat{\boldsymbol{\beta}}_K^{(r)},\\
\widehat{\boldsymbol{b}}_i^{(r)} &= \boldsymbol{D}_i\boldsymbol{Z}_i\boldsymbol{V}_i^{-1(r-1)}\big(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\widehat{\boldsymbol{\beta}}_K^{(r)}\big),
\end{aligned}$$

درنظر بگیرید. توجه کنید که برای سایر برآوردگرها رابطه فوق به‌طور مشابه تعریف می‌شود و تنها کافی است که به جای $\widehat{\boldsymbol{\beta}}_K^{(r)}$ برآوردگر موردنظر جایگزین شود. الگوریتم پیشنهادی ۱، مراحل را مشخص می‌کند.

الگوریتم ۱ الگوریتم تکراری برای برآوردگر ریج

۱: قرار دهید $r = ۰$. با فرض استقلال، برآورد اولیه برای θ، $\sigma^۲(t)$ و $\boldsymbol{D}_i$، به‌صورت $\widehat{\theta}^{(r)} = ۱$، $\widehat{\sigma}^{۲(r)}(t) = ۱$ و $\widehat{\boldsymbol{D}}_i^{(r)} = I_q$ است.

۲: قرار دهید $r = r + ۱$. بر حسب مقادیر گام قبل، $\widehat{\boldsymbol{f}}(\boldsymbol{t})$، $\widehat{\boldsymbol{\beta}}$ و $\widehat{\boldsymbol{b}}_i$ را به کمک رابطه (۱۲.۱) به‌روز کنید، که در آن

$$\widehat{\boldsymbol{V}}_i^{(r-۱)} = \boldsymbol{Z}_i^\top \boldsymbol{D}_i^{(r-۱)} \boldsymbol{Z}_i + \widehat{\boldsymbol{\Sigma}}_i^{(r-۱)}, \quad \widehat{\boldsymbol{\Sigma}}_i^{(r-۱)} = \widehat{\boldsymbol{A}}_i^{(r-۱)} \boldsymbol{R}_i(\widehat{\theta}^{(r-۱)}) \widehat{\boldsymbol{A}}_i^{(r-۱)},$$
$$\widehat{\boldsymbol{A}}_i^{(r-۱)} = \operatorname{diag}\Big(\widehat{\sigma}^{(r-۱)}(t_{i۱}), \ldots, \widehat{\sigma}^{(r-۱)}(t_{in_i})\Big).$$

۳: مقادیر $\widehat{\boldsymbol{D}}_i^{[r]}$، $\widehat{\sigma}^{۲[r]}(t)$ و $\widehat{\theta}^{[r]}$ را به‌صورت

$$\widehat{\boldsymbol{D}}_i^{(r)} = n^{-۱} \sum_{i=۱}^{n} \Big(\widehat{\boldsymbol{b}}_i^{(r)} \widehat{\boldsymbol{b}}_i^{\top(r)} + [\widehat{\boldsymbol{D}}_i^{(r-۱)} - \widehat{\boldsymbol{D}}_i^{(r-۱)} \boldsymbol{Z}_i \widehat{\boldsymbol{V}}_i^{(r-۱)} \boldsymbol{Z}_i^\top \widehat{\boldsymbol{D}}_i^{(r-۱)}] \Big),$$
$$\widehat{\sigma}^{۲(r)}(t) = \frac{\sum_{i=۱}^{n} \sum_{j=۱}^{n_i} \widehat{r}_{ij}^{۲(r)} K_{h_۱}(t - t_{ij})}{\sum_{i=۱}^{n} \sum_{j=۱}^{n_i} K_{h_۱}(t - t_{ij})},$$
$$\widehat{\theta}^{(r)} = \underset{\theta \in [-۱, ۱]}{\arg\max} \Big(-\frac{۱}{۲} \sum_{i=۱}^{n} \big\{ \log |\boldsymbol{R}_i(\widehat{\theta}^{(r-۱)})| + \widehat{\boldsymbol{r}}_i^{\top(r)} \widehat{\boldsymbol{A}}_i^{-۱} \boldsymbol{R}_i^{-۱}(\widehat{\theta}^{(r-۱)}) \widehat{\boldsymbol{A}}_i^{-۱} \widehat{\boldsymbol{r}}_i^{(r)} \big\} \Big),$$

به‌روز کنید، که در آن $\widehat{r}_{ij}^{(r)} = y_i(t_{ij}) - \boldsymbol{x}_i(t_{ij}) \widehat{\boldsymbol{\beta}}^{(r)} - \boldsymbol{z}_i(t_{ij}) \widehat{\boldsymbol{b}}_i^{(r)} - \widehat{f}^{[r]}(t_{ij})$.

۴: گام‌های دوم و سوم را تا رسیدن به همگرایی ($|\widehat{\beta}_{(r+۱)} - \widehat{\beta}_{(r)}| < \varepsilon$) تکرار کنید.

۶.۵.۱ مطالعات عددی

۱.۶.۵.۱ مطالعه شبیه‌سازی

در این قسمت به منظور مقایسه برآوردگرهای هسته‌ای، بازگشتی و اسپلاین در مدل (۷.۱) برای داده‌های طولی، متغیرهای پاسخ را از مدل

$$y_{ij} = x_{۱,ij}\beta_۱ + x_{۲,ij}\beta_۲ + x_{۳,ij}\beta_۳ + f(t_{ij}) + b_i + \varepsilon_{ij}, \quad i = ۱, \ldots, n. \quad j = ۱, \ldots, ۴$$

شبیه‌سازی می‌کنیم، که در آن مولفه‌های مدل به‌صورت

$$\boldsymbol{\beta}^\top = (\beta_۱, \beta_۲, \beta_۳) = (۰/۵, -۱, ۱), \quad x_{۱,ij}, x_{۲,ij}, x_{۳,ij} \sim \mathcal{N}(۰, ۱),$$
$$f(t_{ij}) = ۲\sin(۲\pi t_{ij}), \quad t_{ij} \sim \mathcal{U}(۰, ۱)$$
$$b_i \sim \mathcal{N}(۰, \sigma_b^۲); \qquad \sigma_b^۲ = ۰/۲۵,$$
$$\varepsilon_{ij} \sim \mathcal{N}(۰, \sigma^۲(t)), \quad \sigma^۲(t) = \exp(\tfrac{t}{۱۲})$$

تعیین شده‌اند. در این شبیه‌سازی، ساختار همبستگی بین مشاهدات مکرر را از طریق مدل همبستگی اتورگرسیو مرتبه اول $AR(۱)$ به‌صورت $\operatorname{corr}(\varepsilon_{is}, \varepsilon_{it}) = \rho^{|t-s|}$ درنظر گرفتیم که در آن $s \neq t$. برای

آزمودن تاثیر شدت همبستگی از نوع ضعیف، متوسط و قوی، مقادیر ۰٫۳، ۰٫۶ و ۰٫۹ را برای ρ اتخاذ کردیم. همچنین از تابع هسته اپانچنیکوف که به‌صورت $K(u) = ۰٫۷۵(۱ - u^۲)_+$ است، استفاده می‌شود. برای بررسی کارایی برآوردگرها با افزایش حجم نمونه مقدار n را یک مقدار نسبتا کوچک، $n = ۵۰$، و یک مقدار نسبتاً بزرگ، $n = ۱۰۰$، درنظر گرفتیم.

ابتدا برای مقایسه عملکرد مدل‌ها، مدل اثرات آمیخته، مدل نیمه‌پارامتری و مدل نیمه‌پارامتری با اثرات آمیخته را به مجموعه داده‌های شبیه‌سازی‌شده برازش دادیم؛ که مولفه ناپارامتری به روش اسپلاین تقریب زده شد. سپس برای مقایسه عملکرد برآوردگرها سه روش اسپلاین، هسته‌ای و بازگشتی را در مدل نیمه‌پارامتری با اثرات آمیخته به‌کار بردیم. به منظور بررسی تاثیر اعمال ساختار همبستگی نادرست و همچنین نادیده گرفتن آن، برآوردها را تحت ساختار همبستگی واقعی (AR)، همبستگی تبادل‌پذیر (EX) که به‌صورت $\text{corr}\left(\varepsilon_i(s), \varepsilon_i(t)\right) = \rho$ بوده، و با فرض استقلال (IND) به‌دست آوردیم. دقت برآورد در هر یک از مدل‌ها و همچنین برآوردگرها را توسط مقادیر اریبی و انحراف استاندارد (SD) پارامترها همچنین توسط معیار کمترین توان‌های دوم خطا (MSE) بر اساس میانگین نتایج ۵۰۰ تکرار از شبیه‌سازی مورد سنجش قرار دادیم. برای آگاهی در مورد این معیارها ضمیمه آ را ببینید.

جدول ۲.۱ نتایج مربوط به مقادیر MSE، اریبی و SD تحت برازش سه مدل است که اعداد داخل پرانتز مقدار SD است. از نظر معیار MSE مدل نیمه‌پارامتری با اثرات آمیخته عملکرد بهتری نسبت به دو مدل دیگر دارد؛ حتی زمانی که ساختار همبستگی نادرست انتخاب شود. در این مدل، حالت همبستگی نادرست EX رفتاری بسیار شبیه به حالت همبستگی درست AR دارد اما هنگامی‌که شدت همبستگی بالاست، AR بر EX پیشی می‌گیرد. حالت IND بدترین عملکرد را داشته است زیرا همبستگی را نادیده گرفته است. همچنین این مدل مقادیر اریبی و SD کمتری نسبت به دو مدل دیگر دارد.

نتایج حاصل از کاربرد برآوردگرهای مختلف مدل نیمه‌پارامتری با اثرات آمیخته در جدول ۳.۱ گزارش شدند. با توجه به نتایج، هر یک از برآوردگرها تحت ساختار همبستگی درست AR عملکرد خوبی نسبت به EX یا IND دارند و نادیده گرفتن همبستگی تاثیر نامطلوبی روی مقادیر اریبی، SD و همچنین MSE گذاشته است. روش اسپلاین در تمام جوانب بهتر از دو برآوردگر دیگر عمل کرده است. همچنین با مقایسه دو روش موضعی می‌توان دریافت که مقادیر اریبی و SD روش هسته‌ای نسبت به روش بازگشتی کمتر است که این مشاهدات با نتایج نظری قضیه ۳.۵.۱ هم‌خوانی دارد.

۲.۶.۵.۱ تحلیل داده‌های $CD۴$

در این بخش به مطالعه رفتار برآوردگرهای پیشنهادی با استفاده از داده‌های $CD۴$ در بخش ۴.۱ می‌پردازیم. داده‌ها با استفاده از سه مدل اثرات آمیخته، نیمه‌پارامتری و نیمه‌پارامتری با اثرات آمیخته تحلیل شدند و برای برآورد جزء ناپارامتری از روش‌های اسپلاین، هسته‌ای و بازگشتی استفاده شد. برای انتخاب پهنای باند در برآوردگرهای هسته‌ای و بازگشتی، از معیار GCV استفاده شد. بدین

ترتیب برآورد پهنای باند h و h_1 به‌ترتیب ۰٫۰۲ و ۰٫۰۱۵ به‌دست آمد. یادآور می‌شویم که h پهنای باند به‌کار رفته در برآورد $\boldsymbol{\mu}_Y(.)$ و $\boldsymbol{\mu}_X(.)$ و h_1 پهنای باند به‌کار رفته در برآورد $\sigma^2(.)$ است.

هدف این است که ضمن انتخاب بهترین مدل، عملکرد سه برآوردگر معرفی شده مقایسه شوند. جهت رسیدن به این هدف معیارهای SD، AIC، BIC و $MSPE$ محاسبه شدند. ضمیمه را ببینید. نتایج حاصل در جدول ۴.۱ ارائه شده است. این نتایج نشان می‌دهد که مدل نیمه‌پارامتری با اثرات آمیخته به‌طور قابل توجهی دارای معیارهای برازش مدل کمتری نسبت به دو مدل رقیب است. اگرچه دو مدل نیمه‌پارامتری با اثرات آمیخته و مدل نیمه‌پارامتری نزدیک به هم عمل کرده‌اند اما این تفاوت در مدل اثرات آمیخته چشمگیرتر است. برآوردگر بازگشتی همواره مقادیر SD بزرگتری از روش هسته‌ای دارد، بنابراین استفاده از این برآوردگر در تحلیل داده‌های طولی توصیه نمی‌شود. این مورد با نتایج خواص مجانبی برآوردگرها مطابقت دارد. همچنین برآوردگر اسپلاین بهتر از دو روش دیگر عمل کرده است.

جدول ۲.۱: نتایج شبیه‌سازی: مقادیر MSE، اریبی (SD) در مدل‌های متفاوت به ازای ρها و nهای مختلف.

همبستگی	پارامتر	$n = ۵۰$			$n = ۱۰۰$		
		$\rho = ۰/۳$	$\rho = ۰/۶$	$\rho = ۰/۹$	$\rho = ۰/۳$	$\rho = ۰/۶$	$\rho = ۰/۹$
مدل نیمه‌پارامتری							
AR	$\beta_۱$	۰/۲۰۸۹(۰/۲۵۶۱)	۰/۱۶۳۰(۰/۱۸۸۴)	۰/۱۲۸۸(۰/۱۴۶۷)	۰/۱۳۰۹(۰/۱۴۵۶)	۰/۰۹۲۲(۰/۱۰۴۲)	۰/۰۶۵۲(۰/۰۵۶۲)
	$\beta_۲$	۰/۱۴۶۱(۰/۱۷۰۴)	۰/۱۱۰۶(۰/۱۲۷۱)	۰/۰۷۳۴(۰/۰۶۴۴)	۰/۱۴۶۶(۰/۱۴۵۹)	۰/۱۲۰۸(۰/۱۳۰۸)	۰/۱۱۹۱(۰/۰۶۸۰)
	$\beta_۳$	۰/۲۵۷۹(۰/۳۰۷۷)	۰/۱۹۸۴(۰/۲۳۴۳)	۰/۱۹۳۰(۰/۲۱۴۷)	۰/۱۱۰۷(۰/۱۳۹۰)	۰/۰۸۴۸(۰/۱۰۴۸)	۰/۰۵۵۶(۰/۰۵۳۵)
	MSE	۰/۲۲۷۳	۰/۱۲۸۶	۰/۰۷۴۹	۰/۰۷۷۶	۰/۰۴۸۱	۰/۰۲۶۴
EX	$\beta_۱$	۰/۲۰۶۹(۰/۲۵۵۷)	۰/۱۶۰۷(۰/۱۸۸۶)	۰/۱۲۵۹(۰/۱۴۵۷)	۰/۱۳۰۹(۰/۱۴۵۶)	۰/۰۹۲۸(۰/۱۰۶۸)	۰/۰۶۵۱(۰/۰۵۶۲)
	$\beta_۲$	۰/۱۴۲۴(۰/۱۶۷۰)	۰/۱۱۵۹(۰/۱۵۲۰)	۰/۰۷۸۲(۰/۱۱۰۵)	۰/۱۴۶۶(۰/۱۴۵۹)	۰/۱۲۰۴(۰/۱۳۰۹)	۰/۱۱۸۹(۰/۰۶۸۲)
	$\beta_۳$	۰/۲۵۸۸(۰/۳۰۷۶)	۰/۲۰۱۶(۰/۲۴۱۲)	۰/۱۹۴۹(۰/۲۱۶۲)	۰/۱۱۰۷(۰/۱۳۹۰)	۰/۰۸۴۴(۰/۱۰۳۰)	۰/۰۵۵۶(۰/۰۵۳۳)
	MSE	۰/۲۲۷۴	۰/۱۲۸۷	۰/۰۷۷۵	۰/۰۷۸۶	۰/۰۵۰۴	۰/۰۲۶۳
IND	$\beta_۱$	۰/۲۱۶۳(۰/۲۷۳۷)	۰/۱۵۶۴(۰/۱۹۳۲)	۰/۱۲۱۹(۰/۱۴۲۲)	۰/۱۲۴۷(۰/۱۴۲۵)	۰/۰۸۹۹(۰/۱۰۴۰)	۰/۰۶۳۶(۰/۰۵۹۶)
	$\beta_۲$	۰/۱۶۶۳(۰/۲۳۴۵)	۰/۱۱۱۲(۰/۱۳۹۰)	۰/۰۶۷۰(۰/۰۶۶۴)	۰/۱۴۸۷(۰/۱۵۸۲)	۰/۱۲۳۶(۰/۱۲۸۷)	۰/۱۱۰۵(۰/۰۷۲۷)
	$\beta_۳$	۰/۲۵۶۸(۰/۳۰۶۱)	۰/۲۱۶۰(۰/۲۶۳۴)	۰/۱۸۸۲(۰/۲۱۰۰)	۰/۱۱۴۵(۰/۱۴۰۸)	۰/۰۸۸۶(۰/۱۰۴۷)	۰/۰۵۶۶(۰/۰۵۳۷)
	MSE	۰/۴۹۵۰	۰/۴۱۲۷	۰/۴۰۹۳	۰/۸۴۳۹	۰/۸۰۴۱	۰/۸۱۱۹
مدل اثرات آمیخته							
AR	$\beta_۱$	۰/۲۰۴۰(۰/۲۵۹۱)	۰/۱۶۵۳(۰/۲۰۷۱)	۰/۱۲۴۵(۰/۱۴۴۲)	۰/۱۲۸۴(۰/۱۴۶۲)	۰/۰۹۱۰(۰/۱۰۵۸)	۰/۰۶۳۸(۰/۰۵۹۱)
	$\beta_۲$	۰/۱۵۶۰(۰/۲۰۲۶)	۰/۱۱۳۲(۰/۱۳۸۶)	۰/۰۶۷۱(۰/۰۶۵۷)	۰/۱۴۹۲(۰/۱۵۵۷)	۰/۱۲۶۸(۰/۱۳۱۳)	۰/۱۱۰۶(۰/۰۷۳۰)
	$\beta_۳$	۰/۲۶۹۹(۰/۳۲۶۶)	۰/۲۰۹۵(۰/۲۵۱۱)	۰/۱۸۹۱(۰/۲۱۱۹)	۰/۱۱۶۹(۰/۱۴۴۵)	۰/۰۹۰۸(۰/۱۰۹۳)	۰/۰۵۶۴(۰/۰۵۳۶)
	MSE	۰/۳۴۰۱	۰/۳۲۱۲	۰/۳۳۰۰	۰/۴۴۶۶	۰/۴۵۰۲	۰/۴۳۳۵
EX	$\beta_۱$	۰/۲۱۶۵(۰/۲۷۳۶)	۰/۱۵۶۲(۰/۱۹۲۸)	۰/۱۲۲۱(۰/۱۴۲۲)	۰/۱۲۵۵(۰/۱۴۳۱)	۰/۰۹۳۹(۰/۱۱۱۱)	۰/۰۶۳۵(۰/۰۵۹۶)
	$\beta_۲$	۰/۱۵۹۰(۰/۲۱۵۶)	۰/۱۱۰۸(۰/۱۳۸۴)	۰/۰۶۶۸(۰/۰۶۶۱)	۰/۱۵۳۳(۰/۱۷۶۴)	۰/۱۲۷۳(۰/۱۴۳۴)	۰/۱۱۰۴(۰/۰۷۲۷)
	$\beta_۳$	۰/۲۵۸۳(۰/۳۰۹۲)	۰/۲۱۵۴(۰/۲۶۳۱)	۰/۱۸۷۷(۰/۲۰۹۵)	۰/۱۱۸۶(۰/۱۵۱۶)	۰/۰۹۵۱(۰/۱۳۱۴)	۰/۰۵۶۶(۰/۰۵۳۸)
	MSE	۰/۳۵۰۶	۰/۳۲۴۰	۰/۳۳۴۷	۰/۴۴۸۰	۰/۴۵۰۵	۰/۴۳۴۹
IND	$\beta_۱$	۰/۲۲۸۱(۰/۱۴۳۵)	۰/۱۹۲۴(۰/۱۰۸۴)	۰/۲۰۳۹(۰/۰۸۴۴)	۰/۲۷۲۰(۰/۰۶۸۹)	۰/۲۶۱۹(۰/۰۵۱۷)	۰/۲۵۸۰(۰/۰۳۳۵)
	$\beta_۲$	۰/۳۹۱۷(۰/۱۵۳۵)	۰/۳۵۹۹(۰/۱۲۵۴)	۰/۳۵۷۷(۰/۱۰۹۹)	۰/۶۵۸۶(۰/۰۸۰۳)	۰/۶۳۴۹(۰/۰۶۶۲)	۰/۶۴۳۴(۰/۰۴۳۹)
	$\beta_۳$	۰/۴۸۱۳(۰/۱۳۸۵)	۰/۴۵۸۶(۰/۱۰۲۳)	۰/۴۶۰۳(۰/۰۹۴۰)	۰/۵۶۳۷(۰/۰۸۵۹)	۰/۵۶۶۷(۰/۰۶۵۵)	۰/۵۷۱۷(۰/۰۳۷۸)
	MSE	۰/۳۵۸۰	۰/۳۳۰۰	۰/۳۳۵۱	۰/۴۴۹۱	۰/۴۴۹۰	۰/۴۳۸۷
مدل نیمه‌پارامتری با اثرات آمیخته							
AR	$\beta_۱$	۰/۰۶۸۵(۰/۰۷۲۸)	۰/۰۷۳۹(۰/۰۷۵۷)	۰/۰۷۲۱(۰/۰۷۲۸)	۰/۰۴۹۲(۰/۰۶۰۳)	۰/۰۴۴۱(۰/۰۵۴)	۰/۰۴۸۹(۰/۰۵۸۶)
	$\beta_۲$	۰/۱۰۹۶(۰/۰۸۱۰)	۰/۱۰۰۹(۰/۰۸۱۴)	۰/۱۰۵۴(۰/۰۷۷۶)	۰/۱۸۲۰(۰/۰۵۱۴)	۰/۱۸۵۴(۰/۰۵۰۲)	۰/۱۸۰۴(۰/۰۴۸۷)
	$\beta_۳$	۰/۰۷۹۶(۰/۰۸۰۷)	۰/۰۷۴۵(۰/۰۸۰۴)	۰/۰۷۷۷(۰/۰۷۸۶)	۰/۰۵۸۶(۰/۰۵۴۵)	۰/۰۵۸۲(۰/۰۵۶۱)	۰/۰۶۱۰(۰/۰۵۳۲)
	MSE	۰/۱۹۷۵	۰/۱۱۳۲	۰/۰۷۲۸	۰/۰۷۵۷	۰/۰۴۸۱	۰/۰۲۶۳
EX	$\beta_۱$	۰/۰۷۱۵(۰/۰۷۴۳)	۰/۰۷۵۱(۰/۰۷۶۲)	۰/۰۷۴۸(۰/۰۷۳۲)	۰/۰۴۸۸(۰/۰۶۰۴)	۰/۰۴۴۳(۰/۰۵۴۶)	۰/۰۴۸۲(۰/۰۵۸۰)
	$\beta_۲$	۰/۱۰۸۹(۰/۰۸۱۹)	۰/۱۰۰۹(۰/۰۸۲۰)	۰/۱۰۵۲(۰/۰۷۸۳)	۰/۱۸۲۳(۰/۰۵۱۱)	۰/۱۸۵۴(۰/۰۴۹۹)	۰/۱۷۹۴(۰/۰۴۸۳)
	$\beta_۳$	۰/۰۸۱۸(۰/۰۸۳۶)	۰/۰۷۵۲(۰/۰۸۰۸)	۰/۰۷۸۱(۰/۰۷۷۷)	۰/۰۵۸۹(۰/۰۵۴۳)	۰/۰۵۸۳(۰/۰۵۵۶)	۰/۰۶۱۷(۰/۰۵۳۹)
	MSE	۰/۱۹۳۷	۰/۱۲۲۷	۰/۰۷۳۰	۰/۰۷۷۶	۰/۰۴۸۱	۰/۰۲۸۶
IND	$\beta_۱$	۰/۰۷۱۶(۰/۰۷۳۶)	۰/۰۷۶۳(۰/۰۷۶۷)	۰/۰۷۴۸(۰/۰۷۲۹)	۰/۰۴۸۸(۰/۰۶۰۵)	۰/۰۴۴۲(۰/۰۵۴۴)	۰/۰۴۸۰(۰/۰۵۷۸)
	$\beta_۲$	۰/۱۰۸۴(۰/۰۸۲۲)	۰/۱۰۱۲(۰/۰۸۲۱)	۰/۱۰۴۸(۰/۰۷۸۳)	۰/۱۸۲۲(۰/۰۵۱۵)	۰/۱۸۴۹(۰/۰۵۰۱)	۰/۱۷۹۲(۰/۰۴۷۸)
	$\beta_۳$	۰/۰۸۳۴(۰/۰۸۶۱)	۰/۰۷۶۲(۰/۰۸۲۱)	۰/۰۷۸۷(۰/۰۷۷۲)	۰/۰۵۹۴(۰/۰۵۴۹)	۰/۰۵۸۷(۰/۰۵۶۲)	۰/۰۶۱۸(۰/۰۵۴۰)
	MSE	۰/۲۳۴۹	۰/۱۲۹۰	۰/۰۸۶۰	۰/۰۸۶۰	۰/۰۶۱۶	۰/۰۲۸۶

جدول ۳.۱: نتایج شبیه‌سازی: مقادیر MSE، اریبی (SD) برای روش‌های متفاوت برآورد به ازای ρها و nهای مختلف در مدل اثرات آمیخته نیمه‌پارامتری.

همبستگی	پارامتر	$n = ۵۰$			$n = ۱۰۰$		
		$\rho = ۰/۳$	$\rho = ۰/۶$	$\rho = ۰/۹$	$\rho = ۰/۳$	$\rho = ۰/۶$	$\rho = ۰/۹$
				روش اسپلاین			
AR	$\beta_۱$	۰/۰۶۸۵(۰/۰۷۲۸)	۰/۰۷۳۹(۰/۰۷۵۷)	۰/۰۷۲۱(۰/۰۷۲۸)	۰/۰۴۹۲(۰/۰۶۰۳)	۰/۰۴۴۱(۰/۰۵۴)	۰/۰۴۸۹(۰/۰۵۸۶)
	$\beta_۲$	۰/۱۰۹۶(۰/۰۸۱۰)	۰/۱۰۰۹(۰/۰۸۱۴)	۰/۱۰۵۴(۰/۰۷۷۶)	۰/۱۸۲۰(۰/۰۵۱۴)	۰/۱۸۵۴(۰/۰۵۰۲)	۰/۱۸۰۴(۰/۰۴۸۷)
	$\beta_۳$	۰/۰۷۹۶(۰/۰۸۰۷)	۰/۰۷۴۵(۰/۰۸۰۴)	۰/۰۷۷۷(۰/۰۷۸۶)	۰/۰۵۸۶(۰/۰۵۴۵)	۰/۰۵۸۲(۰/۰۵۶۱)	۰/۰۶۱۰(۰/۰۵۳۲)
	MSE	۰/۱۹۷۵	۰/۱۱۳۲	۰/۰۷۲۸	۰/۰۷۵۷	۰/۰۴۸۱	۰/۰۲۶۳
EX	$\beta_۱$	۰/۰۷۱۵(۰/۰۷۴۳)	۰/۰۷۵۱(۰/۰۷۶۲)	۰/۰۷۴۸(۰/۰۷۳۲)	۰/۰۴۸۸(۰/۰۶۰۴)	۰/۰۴۴۳(۰/۰۵۴۶)	۰/۰۴۸۲(۰/۰۵۸۰)
	$\beta_۲$	۰/۱۰۸۹(۰/۰۸۱۹)	۰/۱۰۰۹(۰/۰۸۲۰)	۰/۱۰۵۲(۰/۰۷۸۳)	۰/۱۸۲۳(۰/۰۵۱۱)	۰/۱۸۵۴(۰/۰۴۹۹)	۰/۱۷۹۴(۰/۰۴۸۳)
	$\beta_۳$	۰/۰۸۱۸(۰/۰۸۳۶)	۰/۰۷۵۲(۰/۰۸۰۸)	۰/۰۷۸۱(۰/۰۷۷۷)	۰/۰۵۸۹(۰/۰۵۴۳)	۰/۰۵۸۳(۰/۰۵۵۶)	۰/۰۶۱۷(۰/۰۵۳۹)
	MSE	۰/۱۹۳۷	۰/۱۲۲۷	۰/۰۷۳۰	۰/۰۷۷۶	۰/۰۴۸۱	۰/۰۲۸۶
IND	$\beta_۱$	۰/۰۷۱۶(۰/۰۷۳۶)	۰/۰۷۶۳(۰/۰۷۶۷)	۰/۰۷۴۸(۰/۰۷۲۹)	۰/۰۴۸۸(۰/۰۶۰۵)	۰/۰۴۴۲(۰/۰۵۴۴)	۰/۰۴۸۰(۰/۰۵۷۸)
	$\beta_۲$	۰/۱۰۸۴(۰/۰۸۲۲)	۰/۱۰۱۲(۰/۰۸۲۱)	۰/۱۰۴۸(۰/۰۷۸۳)	۰/۱۸۲۲(۰/۰۵۱۵)	۰/۱۸۴۹(۰/۰۵۰۱)	۰/۱۷۹۲(۰/۰۴۷۸)
	$\beta_۳$	۰/۰۸۳۴(۰/۰۸۶۱)	۰/۰۷۶۲(۰/۰۸۲۱)	۰/۰۷۸۷(۰/۰۷۷۲)	۰/۰۵۹۴(۰/۰۵۴۹)	۰/۰۵۸۷(۰/۰۵۶۲)	۰/۰۶۱۸(۰/۰۵۴۰)
	MSE	۰/۲۳۴۹	۰/۱۲۹۰	۰/۰۸۶۰	۰/۰۸۶۰	۰/۰۶۱۶	۰/۰۲۸۶
				روش هسته‌ای			
AR	$\beta_۱$	۰/۲۰۴۴(۰/۲۵۰۶)	۰/۱۶۰۰(۰/۱۸۶۹)	۰/۱۲۵۱(۰/۱۴۴۳)	۰/۱۳۰۰(۰/۱۴۲۲)	۰/۰۹۲۸(۰/۱۰۶۳)	۰/۰۶۵۱(۰/۰۵۶۲)
	$\beta_۲$	۰/۱۴۳۴(۰/۱۶۵۷)	۰/۱۱۱۱(۰/۱۲۸۵)	۰/۰۷۱۱(۰/۰۶۲۳)	۰/۱۴۴۴(۰/۱۴۵۴)	۰/۱۱۹۱(۰/۱۲۹۰)	۰/۱۱۷۸(۰/۰۶۸۷)
	$\beta_۳$	۰/۲۵۴۸(۰/۳۰۵۰)	۰/۱۹۴۰(۰/۲۳۰۲)	۰/۱۹۰۷(۰/۲۱۱۸)	۰/۱۰۷۷(۰/۱۳۲۳)	۰/۰۸۳۴(۰/۱۰۲۳)	۰/۰۵۵۴(۰/۰۵۴۲)
	MSE	۰/۱۹۰۱	۰/۱۱۰۳	۰/۰۷۴۸	۰/۰۷۴۸	۰/۰۴۷۳	۰/۰۲۸۳
EX	$\beta_۱$	۰/۲۰۸۹(۰/۲۵۷۳)	۰/۱۶۳۴(۰/۱۹۴۱)	۰/۱۲۵۸(۰/۱۴۵۶)	۰/۱۲۹۱(۰/۱۴۲۲)	۰/۰۹۲۷(۰/۱۰۶۶)	۰/۰۶۴۹(۰/۰۵۵۸)
	$\beta_۲$	۰/۱۴۶۷(۰/۱۸۵۱)	۰/۱۱۱۳(۰/۱۲۸۸)	۰/۰۷۲۱(۰/۰۶۳۱)	۰/۱۴۶۱(۰/۱۴۶۶)	۰/۱۲۰۳(۰/۱۳۰۶)	۰/۱۱۹۵(۰/۰۶۷۸)
	$\beta_۳$	۰/۲۶۶۳(۰/۳۲۲۷)	۰/۱۹۸۷(۰/۲۳۶۰)	۰/۱۹۵۸(۰/۲۱۷۲)	۰/۱۰۷۱(۰/۱۳۳۳)	۰/۰۸۴۴(۰/۱۰۳۱)	۰/۰۵۵۴(۰/۰۵۳۳)
	MSE	۰/۲۱۱۵	۰/۱۱۵۲	۰/۰۷۷۵	۰/۰۷۵۱	۰/۰۴۸۱	۰/۰۲۸۶
IND	$\beta_۱$	۰/۲۰۵۸(۰/۲۵۶۲)	۰/۱۸۲۴(۰/۲۰۳۹)	۰/۱۶۳۷(۰/۱۵۵۰)	۰/۴۱۳۱(۰/۵۰۲۴)	۰/۴۳۳۰(۰/۵۷۲۰)	۰/۴۲۳۷(۰/۵۳۶۲)
	$\beta_۲$	۰/۲۸۹۵(۰/۳۷۸۲)	۰/۲۴۴۵(۰/۳۲۸۶)	۰/۱۳۰۵(۰/۱۶۶۳)	۰/۳۶۳۷(۰/۵۳۴۴)	۰/۴۱۱۷(۱/۰۶۱۸)	۰/۲۰۳۰(۰/۲۸۳۹)
	$\beta_۳$	۰/۳۴۵۷(۰/۴۰۸۷)	۰/۲۹۷۷(۰/۳۵۴۰)	۰/۲۲۵۰(۰/۲۵۸۶)	۰/۴۲۷۱(۰/۵۵۷۸)	۰/۵۲۶۵(۰/۹۲۹۲)	۰/۴۹۸۱(۱/۰۵۴۶)
	MSE	۰/۴۶۴۶	۰/۳۶۶۰	۰/۱۵۴۷	۰/۹۳۰۴	۲/۴۵۸۸	۱/۵۹۷۴
				روش بازگشتی			
AR	$\beta_۱$	۰/۲۳۲۴(۰/۳۰۷۲)	۰/۱۹۹۸(۰/۲۴۸۲)	۰/۱۴۹۰(۰/۱۵۶۵)	۰/۲۵۴۵(۰/۳۳۶۲)	۰/۲۰۶۳(۰/۲۸۵۸)	۰/۱۵۴۴(۰/۲۱۹۸)
	$\beta_۲$	۰/۲۴۴۲(۰/۳۳۰۸)	۰/۱۸۵۷(۰/۲۷۴۴)	۰/۱۲۴۹(۰/۱۹۴۳)	۰/۲۸۹۱(۰/۴۰۵۲)	۰/۲۱۰۳(۰/۳۱۰۸)	۰/۱۳۷۰(۰/۱۵۴۲)
	$\beta_۳$	۰/۳۰۵۳(۰/۳۷۱۹)	۰/۲۴۵۴(۰/۳۱۵۹)	۰/۲۳۸۸(۰/۲۸۳۸)	۰/۳۰۵۶(۰/۴۴۰۵)	۰/۲۶۳۵(۰/۴۱۴۶)	۰/۱۷۶۳(۰/۳۳۵۹)
	MSE	۰/۳۹۱۴	۰/۲۷۱۳	۰/۱۶۷۰	۰/۵۰۶۹	۰/۳۷۰۸	۰/۲۰۱۶
EX	$\beta_۱$	۰/۲۰۶۰(۰/۲۵۵۹)	۰/۱۹۱۰(۰/۲۴۲۷)	۰/۱۶۵۷(۰/۱۵۸۷)	۰/۴۰۴۶(۰/۴۹۶۰)	۰/۴۲۶۴(۰/۵۰۷۶)	۰/۴۰۹۳(۰/۴۱۰۲)
	$\beta_۲$	۰/۲۸۳۶(۰/۳۷۱۵)	۰/۲۳۴۷(۰/۳۱۵۲)	۰/۱۲۶۲(۰/۱۵۲۹)	۰/۳۴۰۹(۰/۴۸۳۶)	۰/۳۳۵۱(۰/۵۰۷۰)	۰/۲۲۶۸(۰/۴۴۴۵)
	$\beta_۳$	۰/۳۴۲۳(۰/۴۰۳۳)	۰/۲۹۱۷(۰/۳۴۴۶)	۰/۲۲۴۳(۰/۲۵۶۵)	۰/۴۴۰۶(۰/۵۸۷۳)	۰/۴۹۲۵(۰/۷۲۵۴)	۰/۴۶۳۸(۰/۷۳۰۱)
	MSE	۰/۴۵۲۰	۰/۳۶۵۶	۰/۱۴۷۷	۰/۹۰۲۷	۱/۱۳۴۹	۱/۰۲۶۳
IND	$\beta_۱$	۰/۲۳۶۲(۰/۱۴۷۰)	۰/۱۹۲۹(۰/۱۰۶۷)	۰/۲۰۳۲(۰/۰۸۱۲)	۰/۲۷۳۴(۰/۰۶۷۶)	۰/۲۶۶۱(۰/۰۵۲۲)	۰/۲۵۸۶(۰/۰۲۸۲)
	$\beta_۲$	۰/۴۰۲۹(۰/۱۶۹۱)	۰/۳۵۹۴(۰/۱۱۸۱)	۰/۳۵۲۷(۰/۰۸۵۵)	۰/۶۶۴۱(۰/۰۷۸۹)	۰/۶۴۳۲(۰/۰۶۵۱)	۰/۶۴۵۰(۰/۰۳۴۳)
	$\beta_۳$	۰/۴۹۲۵(۰/۱۵۳۷)	۰/۴۵۸۶(۰/۰۹۶۴)	۰/۴۵۷۴(۰/۰۸۳۵)	۰/۵۶۹۸(۰/۰۹۰۷)	۰/۵۷۳۹(۰/۰۶۷۵)	۰/۵۷۳۱(۰/۰۳۱۰)
	MSE	۰/۵۲۹۰	۰/۴۰۹۱	۰/۳۹۵۶	۰/۸۵۹۴	۰/۸۲۵۳	۰/۸۱۴۲

جدول ۴.۱: نتایج تحلیل داده‌های $CD4$: برآورد پارامترهای مدل رگرسیونی (SD) تحت برازش مدل‌های مختلف.

	مدل				
		نیمه‌پارامتری با اثرات آمیخته		نیمه‌پارامتری	اثرات آمیخته
	اسپلاین	هسته‌ای	بازگشتی	اسپلاین	
AGE	۰٫۰۸۵۱(۰٫۰۰۰۲)	۰٫۰۲۱۲(۰٫۰۰۲۱)	۰٫۰۳۲۹(۰٫۰۰۲۱)	۰٫۱۵۵۱(۰٫۰۵۷۸)	۰٫۴۹۱۹(۰٫۰۲۳۲)
$SMOKE$	۰٫۷۲۹۰(۰٫۰۰۶۰)	۰٫۵۹۴۱(۰٫۰۰۸۲)	۰٫۶۱۱(۰٫۰۰۵۱)	۰٫۱۵۷۰(۰٫۱۷۲۴)	۲٫۹۱۴۱(۰٫۱۰۳۹)
$DRUG$	۰٫۸۰۱۱(۰٫۰۱۴۳)	۰٫۵۳۵۰(۰٫۰۱۹۴)	۰٫۰۵۴۴(۰٫۰۵۲۱)	۰٫۰۱۱۰(۰٫۰۵۹۰)	۱٫۴۴۵۱(۰٫۳۸۴۴)
$SEXP$	۰٫۰۷۱۱(۰٫۰۰۱۴)	۰٫۰۵۸۵(۰٫۰۰۲۴)	۰٫۰۶۴۰(۰٫۰۰۴۶)	۰٫۰۱۸۹(۰٫۰۰۴۴)	۰٫۹۷۱۲(۰٫۰۲۰۰)
$CESD$	−۰٫۰۵۴۲(۰٫۰۰۰۳)	−۰٫۰۵۳۱(۰٫۰۰۰۹)	−۰٫۰۶۲۲(۰٫۰۰۲۲)	−۰٫۰۱۷۴(۰٫۰۰۳۵)	−۰٫۰۳۸۴(۰٫۰۰۶۹)
AIC	۳۱۳۷٫۴۶	۳۲۹۱٫۱۴۱	۳۳۱۹٫۳۰۵	۳۴۹۱٫۱۶۴	۱۰۱۸۷٫۰۶
BIC	۳۱۹۶٫۱۸	۳۳۱۰٫۶۲۷	۳۳۳۸٫۷۹۱	۳۵۱۰٫۶۴۹	۱۰۲۰۶٫۵۵
$MSPE$	۹۷٫۵۲۱۲	۹۸٫۸۱۹۴۵	۹۸٫۳۶۵۷۲	۱۳۶٫۰۲۳۹	۲۰۴٫۸۱۱

فصل ۲

انتخاب متغیر برای داده‌های طولی نرمال

۱.۲ انتخاب متغیر در مدل‌های رگرسیونی

مدل رگرسیونی چندگانه

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon = \boldsymbol{x}^\top \boldsymbol{\beta} + \varepsilon \qquad (۱.۲)$$

را درنظر بگیرید، که در آن $\boldsymbol{x} = (1, x_1, \ldots, x_p)^\top$ بردار متغیرهای تبیینی، $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ بردار ضرایب رگرسیونی و ε خطای تصادفی با میانگین صفر و واریانس ثابت است. در این رساله بدون از دست دادن کلیت، فرض می‌شود عرض از مبدا برابر با صفر است. روش‌های متعددی برای برآورد پارامترهای مدل وجود دارند، که برآورد ماکزیمم درستنمایی (MLE) و کمترین توان‌های دوم معمولی (OLS) از جمله معمول‌ترین این روش‌ها هستند. اگر داده‌ها دارای توزیع نرمال باشند این دو روش معادل‌اند. هنگامی‌که بین متغیرهای تبیینی هم‌خطی وجود دارد، استنباط بر اساس این روش می‌تواند گمراه‌کننده باشد. زیرا وجود هم‌خطی، برآوردهایی با واریانس بزرگ را نتیجه می‌دهد که منجر به کاهش دقت پیشگویی می‌شود. تاکنون روش‌های متعددی برای مقابله با مشکلات هم‌خطی بین متغیرهای تبیینی معرفی شده است. برای اطلاعات بیش‌تر در این‌باره به مونت‌گومری و همکاران (۲۰۱۲) و چاترجی و هادی (۲۰۱۲) مراجعه کنید. یک روش متداول برای مقابله با هم‌خطی، استفاده از برآوردگرهای اریب مانند ریج[1] است که توسط هورل و کنارد

[1] Ridge

(۱۹۷۰) معرفی شده است. یکی از عوامل ایجاد هم‌خطی، زیاد بودن تعداد متغیرهای تبیینی در مدل است. در مدل‌های با بعد بالا مسئله هم‌خطی بسیار شدید است، یعنی هر یک از متغیرهای تبیینی را می‌توان به‌صورت ترکیب خطی از سایر متغیرها نوشت (جیمز و همکاران ، ۲۰۱۳). در این حالت روش پیشنهادی، انتخاب زیرمجموعه‌ای از متغیرهای تبیینی برای ورود به مدل است. در این رهیافت علاقه‌مندیم زیرمجموعه‌ای از متغیرها را طوری انتخاب کنیم که رابطه بین متغیر پاسخ و متغیرهای تبیینی را به خوبی توصیف کند. ساختن یک مدل رگرسیونی که تنها شامل زیرمجموعه‌ای از متغیرهای تبیینی است، انتخاب متغیر[۲] نامیده می‌شود. مسئله انتخاب متغیر دو واقعیت ناسازگار را دربر دارد. اولاً علاقه‌مندیم که مدل تا حد امکان متغیرهای تبیینی بیش‌تری را شامل شود، به نحوی که اطلاعات موجود در این متغیرها بتواند در پیشگویی متغیر پاسخ مؤثر باشد. دوماً می‌خواهیم مدل حتی‌الامکان دارای متغیرهای رگرسیونی کمتری باشد، چون با افزایش تعداد متغیرهای تبیینی در مدل، هزینه محاسبات افزایش می‌یابد و تفسیر مدل دشوارتر می‌شود. علاوه بر این، وجود تعداد زیادی متغیر تبیینی ممکن است باعث بیش‌برازش[۳] شود، یعنی مدل برازش‌شده برای داده‌های آموزشی[۴] که برای ساخت مدل از آن‌ها استفاده شده است، مناسب باشد اما برای پیشگویی مشاهدات آینده بسیار ضعیف باشد. مسئله انتخاب متغیر امری بسیار مهم در تحلیل داده‌هاست. فرآیند به‌دست آوردن یک مدل که سازگاری و توافق بین این دو واقعیت است، انتخاب بهترین معادله رگرسیونی نامیده می‌شود. بدیهی است برای تعیین مدل پیشگو، لازم است پارامترهای مدل برآورد شوند. لذا فرآیندی که علاوه بر انتخاب متغیر و بهترین معادله رگرسیونی، برآورد ضرایب را نیز انجام دهد، موردنظر است. یک روش کارا برای برآورد پارامترها که هم‌زمان انتخاب متغیر را نیز انجام می‌دهد، روش مبتنی بر تابع تاوان است. این روش با اضافه کردن جمله‌ای منظم‌ساز به معادلات برآورد، رویکرد نوینی تحت عنوان رگرسیون تاوانیده را گسترش داد که تاکنون توسط محققین زیادی مورد مطالعه قرار گرفته است. در این میان می‌توان روش‌های رگرسیون بریج[۵] (فرانک و فریدمن ، ۱۹۹۳)، لاسو[۶] (تیبشیرانی ، ۱۹۹۶)، لاسوی سازوار[۷](زو ، ۲۰۰۶)، الاستیک نت [۸] (زو و هستی ، ۲۰۰۵)، اسکد [۹](فن و لی ، ۲۰۰۱) و MCP (ژانگ ، ۲۰۱۰) را نام برد.

امروزه توسعه و گسترش سریع علوم در زمینه‌های مختلف از جمله ژنومیک، علوم اعصاب، علوم زمین، مالی و غیره، افزایش سرعت ثبت و تحلیل اطلاعات توسط ابررایانه‌ها، آماردانان را با حجم عظیمی از اطلاعات پیچیده در تحلیل مسائل مواجه نموده است. اگرچه محققین در زمینه‌های مختلف اهداف متفاوتی را دنبال می‌کنند، اما با این حال یک موضوع مشترک دارند؛ تحقیق آن‌ها به شدت به استخراج اطلاعات مفید از داده‌های حجیم وابسته است و تعداد متغیرهای تبیینی یعنی بعد متغیرها

[2]Variable selection
[3]Overfitting
[4]Training data set
[5]Bridge
[6]Lasso
[7]Adaptive lasso
[8]Elastic-net
[9]SCAD

در مقایسه با حجم نمونه می‌تواند بزرگ باشد. به‌طور کلی داده‌ها را از نظر بعد می‌توان به دو گروه بعد پایین[۱۰] و بعد بالا[۱۱] تقسیم نمود. در داده‌های با بعد پایین، تعداد متغیرها، p، بسیار کوچک‌تر از تعداد مشاهدات، n، است. به عنوان مثال مجموعه داده با حجم $n = ۱۰۰$ و بعد $p = ۱۰$ مجموعه داده با بعد پایین است. به‌علاوه در این داده‌ها برای نظریه مجانبی فرض می‌شود که با افزایش n، مقدار p ثابت می‌ماند. در داده‌های با بعد بالا هم‌چنان p کوچک‌تر از n است ولی مقدار آن بزرگ است یا p از n بزرگ‌تر است. به عنوان مثال $n = ۱۰۰$ و $p = ۲۰۰$. در این نوع داده‌ها با افزایش n، مقدار p می‌تواند افزایش یابد. مسئله "p بزرگ و n کوچک " چالش‌ها و موهبت‌های فراوانی را پیش‌روی آماردانان قرار داده است و تحلیل این نوع داده‌ها مستلزم به‌کارگیری روش‌های نوین آماری است. یک پذیره اساسی که استنباط پیرامون این نوع داده‌ها را امکان‌پذیر می‌سازد، پذیره تنکی[۱۲] است که براساس آن، فقط مجموعه کوچکی از متغیرهای تبیینی بر متغیر پاسخ تاثیرگذارند و بقیه متغیرهای تبیینی ارتباطی با متغیر پاسخ ندارند. بنابراین کاهش بعد و انتخاب متغیر نقش مهمی را در تحلیل این نوع داده‌ها ایفا می‌کند. تاکنون روش‌های آماری مختلفی برای انتخاب متغیر در داده‌های با بعد بالا معرفی شده است که از جمله می‌توان لاسو، لاسوی سازوار، اسکد، الاستیک نت، گاروت نامنفی و انتخابگر دنزیک[۱۳] را نام برد. در حالت خاص، ممکن است p با نرخ نمایی رشد کند، به عبارت دیگر رابطه $\log(p) = O(n^a)$ برای برخی مقادیر مثبت a برقرار باشد. داده‌های بیان‌ژنی، ریزآرایه‌ها، داده‌های تابعی، داده‌های مالی، داده‌های پردازش تصویر، داده‌های تصویربرداری پزشکی و طبقه‌بندی تومور نمونه‌هایی از این نوع داده‌ها هستند. به عنوان مثال در مطالعه رابطه بین ویژگی‌های ظاهری[۱۴] و ویژگی‌های ژنتیکی[۱۵] از میلیون‌ها SNP[۱۶] استفاده می‌کنند، یا در طبقه‌بندی نوعی تومور از داده‌های ریزآرایه استفاده می‌کنند که شامل هزاران عامل ژنتیکی است و هدف شناسایی ژن‌های موثر بر بیماری است. در این مثال‌ها تعداد متغیرهای تبیینی بسیار بزرگ‌تر از تعداد مشاهدات است. برای تمایز این نوع داده‌ها از داده‌های با بعد بالای با نرخ رشد چندجمله‌ای، آن‌ها را داده‌های با بعد بسیار بالا[۱۷] نامند. لذا عبارت بعد بالا برای نرخ رشد چندجمله‌ای ($p = O(n^b)$) و عبارت بعد بسیار بالا برای نرخ رشد نمایی ($p = \exp(O(n^a))$) استفاده می‌شود. برای درک بهتر مطلب دو حالت $p = n^b$ و $p = \exp(n^a)$ را به ترتیب برای بعد بالا و بعد بسیار بالا در نظر گرفته و نسبت $\frac{p}{n}$ را برای مقادیر مختلف (n, a, b) در جدول ۱.۲ گزارش کرده‌ایم. با توجه به نتایج جدول مذکور ملاحظه می‌شود که در بعد بسیار بالا، با افزایش n مقدار p به سرعت رشد می‌کند و به سمت بینهایت می‌رود. لازم به ذکر است که تمرکز این رساله روی داده‌ها با بعد بالا است نه بعد بسیار بالا. برای آگاهی بیش‌تر درباره داده‌های با بعد بالا و چالش‌های تحلیل این نوع داده‌ها فن و لی (۲۰۰۶)، فن و لیو

[10] Low dimension
[11] High dimension
[12] Sparsity
[13] Dantzig selector
[14] Phenotypes
[15] Genotypes
[16] Single-nucleotide polymorphism
[17] Ultrahigh-dimensional data

(۲۰۰۸)، فن و همکاران (۲۰۰۹)، فن و همکاران (۲۰۱۴)، جان‌استون و تیترینگتون (۲۰۰۹)، بولمن و فن‌دگیر (۲۰۱۱) و گیراد (۲۰۱۴) مراجعه کنید.

جدول ۱.۲: مقادیر تقریبی $\frac{p}{n}$ برای داده‌های با ابعاد بالا و بسیار بالا به ازای مقادیر مختلف n، a و b.

	بعد بسیار بالا مقادیر مختلف a				بعد بالا مقادیر مختلف b			
n	۱	$\frac{۷}{۸}$	$\frac{۵}{۸}$	$\frac{۳}{۸}$	۲/۵	۲	۱/۵	۱
۱۰	$۲/۲ \times ۱۰^{۴}$	$۱/۸ \times ۱۰^{۳}$	$۶/۸ \times ۱۰^{۱}$	$۱/۱ \times ۱۰^{۱}$	$۳/۲ \times ۱۰^{۲}$	$۱/۰ \times ۱۰^{۲}$	$۳/۲ \times ۱۰^{۱}$	$۱/۰ \times ۱۰^{۱}$
۳۰	$۱/۱ \times ۱۰^{۱۳}$	$۳/۳ \times ۱۰^{۸}$	$۴/۴ \times ۱۰^{۳}$	$۳/۶ \times ۱۰^{۱}$	$۴/۹ \times ۱۰^{۳}$	$۹/۰ \times ۱۰^{۲}$	$۱/۶ \times ۱۰^{۲}$	$۳/۰ \times ۱۰^{۱}$
۱۰۰	$۲/۷ \times ۱۰^{۴۳}$	$۲/۶ \times ۱۰^{۲۴}$	$۵/۳ \times ۱۰^{۷}$	$۲/۸ \times ۱۰^{۲}$	$۱/۰ \times ۱۰^{۵}$	$۱/۰ \times ۱۰^{۴}$	$۱/۰ \times ۱۰^{۳}$	$۱/۰ \times ۱۰^{۲}$
۲۰۰	$۷/۲ \times ۱۰^{۸۶}$	$۶/۲ \times ۱۰^{۴۴}$	$۸/۱ \times ۱۰^{۱۱}$	$۱/۵ \times ۱۰^{۳}$	$۵/۶ \times ۱۰^{۵}$	$۴/۰ \times ۱۰^{۴}$	$۲/۸ \times ۱۰^{۳}$	$۲/۰ \times ۱۰^{۲}$
۵۰۰	$۱/۴ \times ۱۰^{۲۱۷}$	$۷/۲ \times ۱۰^{۹۹}$	$۱/۳ \times ۱۰^{۲۱}$	$۲/۹ \times ۱۰^{۴}$	$۵/۶ \times ۱۰^{۶}$	$۲/۵ \times ۱۰^{۵}$	$۱/۱ \times ۱۰^{۴}$	$۵/۰ \times ۱۰^{۲}$

در بخش بعد، ابتدا تاثیر وجود هم‌خطی بر عملکرد برآوردگرها و برآوردگر ریج، به عنوان ابزاری برای مقابله با مشکل هم‌خطی بررسی شده و سپس برآوردگر ریج برای داده‌های طولی در مدل نیمه‌پارامتری با اثرات آمیخته معرفی می‌شود. همچنین با معرفی رایج‌ترین روش‌های انتخاب متغیر به مسئله برآورد و انتخاب متغیر همزمان در مدل مذکور پرداخته است.

۱.۱.۲ مسئله هم‌خطی و برآوردگر ریج

فرض کنید $\{y_i, \boldsymbol{x}_i\}_{i=۱}^{n}$ نمونه‌ای تصادفی از مدل (۱.۲) باشد. بنابراین مدل رگرسیون خطی چندگانه را می‌توان با نماد ماتریسی $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ نوشت، که در آن $\boldsymbol{y} = (y_۱, \ldots, y_n)^\top$ بردار متغیر پاسخ، $\boldsymbol{X} = (\boldsymbol{x}_۱^\top, \ldots, \boldsymbol{x}_n^\top)^\top$ ماتریس طرح $n \times p$ بعدی با رتبه کامل ستونی، $\boldsymbol{\beta}$ بردار ضرایب رگرسیونی و $\boldsymbol{\varepsilon} = (\varepsilon_۱, \ldots, \varepsilon_n)^\top$ بردار خطاست. برآوردگر OLS بردار ضرایب به‌صورت

$$\widehat{\boldsymbol{\beta}}_{OLS} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^۲ = (\boldsymbol{X}^\top \boldsymbol{X})^{-۱} \boldsymbol{X}^\top \boldsymbol{y}$$

است. یک مسئله جدی که کاربرد برآوردگرهای معمول را با مشکل مواجه می‌کند وجود هم‌خطی چندگانه[18] بین متغیرهای تبیینی است، بدین معنی که متغیرهای تبیینی به‌صورت کامل یا تقریبا کامل وابسته خطی باشند. اگر یکی از متغیرهای تبیینی یک تابع دقیق خطی از یک یا چند متغیر تبیینی باشد، گوییم رگرسیون دارای هم‌خطی کامل است. هم‌خطی ناقص هنگامی اتفاق می‌افتد که یکی از متغیرهای تبیینی به‌طور تقریبی یک تابع خطی از یک یا چند متغیر تبیینی دیگر باشد. به عبارت دیگر هم‌خطی کامل هنگامی رخ می‌دهد که حداقل به ازای یک k $(k = ۱, \ldots, p)$ داشته باشیم $R_k^۲ = ۱$ و هم‌خطی ناقص زمانی پیش می‌آید که حداقل به ازای یک k داشته باشیم $R_k^۲ \approx ۱$، به‌طوری‌که $R_k^۲$ ضریب تعیین متغیر $\boldsymbol{x}_k$ بر سایر متغیرهای تبیینی در رگرسیون خطی است.

[18] Multicollinearity

وجود هم‌خطی تاثیرات جدی بر برآوردها دارد. مثلاً در برآوردگر OLS منجر به برآوردهایی با واریانس بزرگ می‌شود که دقت پیشگویی را کاهش می‌دهد. برای درک بهتر موضوع، مدل رگرسیون خطی با دو متغیر تبیینی را درنظر بگیرید.

$$y = \beta_۱ x_۱ + \beta_۲ x_۲ + \varepsilon$$

فرض کنید $x_۱$، $x_۲$ و y استاندارد شده‌اند، به‌طوری‌که $\bar{x}_k = ۰$، $\sum_{i=۱}^{n}(x_{ik} - \bar{x}_k)^۲ = ۱$، $\bar{y} = ۰$ و $\sum_{i=۱}^{n}(y_i - \bar{y})^۲ = ۱$. در این صورت معادلات نرمال روش OLS به‌صورت

$$(\boldsymbol{X}^\top \boldsymbol{X})\widehat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{X}^\top \boldsymbol{y}$$

یا به‌طور معادل به‌صورت

$$\begin{bmatrix} ۱ & r_{۱۲} \\ r_{۱۲} & ۱ \end{bmatrix} \begin{bmatrix} \widehat{\beta}_۱ \\ \widehat{\beta}_۲ \end{bmatrix} = \begin{bmatrix} r_{۱y} \\ r_{۲y} \end{bmatrix}$$

است که $r_{۱۲}$ همبستگی ساده بین $x_۱$ و $x_۲$؛ و r_{ky} همبستگی ساده بین y و x_k به ازای $k = ۱, ۲$ هستند. معکوس $\boldsymbol{X}^\top \boldsymbol{X}$ عبارت است از

$$\boldsymbol{C} = (\boldsymbol{X}^\top \boldsymbol{X})^{-۱} = \begin{bmatrix} \frac{۱}{۱-r_{۱۲}^۲} & \frac{-r_{۱۲}}{۱-r_{۱۲}^۲} \\ \frac{-r_{۱۲}}{۱-r_{۱۲}^۲} & \frac{۱}{۱-r_{۱۲}^۲} \end{bmatrix} \qquad (۲.۲)$$

و برآوردهای ضرایب رگرسیونی به‌صورت

$$\widehat{\beta}_۱ = \frac{r_{۱y} - r_{۱۲}r_{۲y}}{۱ - r_{۱۲}^۲}, \qquad \widehat{\beta}_۲ = \frac{r_{۲y} - r_{۱۲}r_{۱y}}{۱ - r_{۱۲}^۲}$$

است. اگر بین $x_۱$ و $x_۲$ هم‌خطی ناقص وجود داشته باشد، آن‌گاه ضریب همبستگی $r_{۱۲}$ بزرگ خواهد بود. از رابطه (۲.۲) می‌توان دریافت که اگر $|r_{۱۲}| \to ۱$، آنگاه به ازای $k = ۱, ۲$ خواهیم داشت $\mathrm{Var}(\widehat{\beta}_k) = C_{kk}\sigma^۲ \to \infty$ و $\mathrm{Cov}(\widehat{\beta}_۱, \widehat{\beta}_۲) = C_{۱۲}\sigma^۲ \to \infty$ که C_{kk} مولفه kام روی قطر اصلی ماتریس $\boldsymbol{C}$ است. بنابراین هم‌خطی زیاد بین $x_۱$ و $x_۲$ منجر به واریانس‌ها و کوواریانس‌های بزرگ برای برآوردگرهای OLS ضرایب رگرسیونی خواهد شد. برای بیش از دو متغیر تبیینی نیز هم‌خطی اثرات مشابهی ایجاد می‌کند. می‌توان نشان داد که اعضای قطر اصلی ماتریس عبارتند از

$$C_{kk} = \frac{۱}{۱ - R_k^۲}, \quad k = ۱, \ldots, p,$$

که در آن $R_k^۲$ ضریب تعیین حاصل از رگرسیون x_j بر $p - ۱$ متغیر تبیینی دیگر است. اگر هم‌خطی شدید بین x_k و هر زیرمجموعه‌ای از $p - ۱$ متغیر دیگر وجود داشته باشد، در این صورت مقدار $r_k^۲$ نزدیک به یک خواهد بود. چون $\mathrm{Var}(\widehat{\beta}_k) = C_{kk}\sigma^۲ = \sigma^۲(۱ - R_k^۲)^{-۱}$، هم‌خطی شدید موجب می‌شود که واریانس برآوردگر ضرایب رگرسیون بسیار بزرگ شود. در این صورت $\widehat{\boldsymbol{\beta}}_{OLS}$ برآورد دقیقی از $\boldsymbol{\beta}$

نیست. به عبارت دیگر، دقت برآورد ضرایب رگرسیونی بسیار کم می‌شود. همچنین هم‌خطی باعث می‌شود که بردار $\widehat{\boldsymbol{\beta}}_{OLS}$ دارای اعضایی باشد که از لحاظ قدر مطلق بسیار بزرگ هستند. علاوه بر این، $\widehat{\boldsymbol{\beta}}_{OLS}$ ناپایدار است، یعنی با حذف یا اضافه کردن یک متغیر تبیینی یا با یک تغییر کوچک در مقادیر متغیرهای تبیینی، مقادیر $\widehat{\boldsymbol{\beta}}_{OLS}$ به‌طور قابل توجهی تغییر می‌کند. در صورت وجود هم‌خطی کامل، ماتریس وارون‌پذیر نیست، زیرا در این حالت پررتبه نیست. بنابراین جواب یکتا برای برآورد ضرایب وجود ندارد. لازم به ذکر است که هم‌خطی کامل معمولا در عمل رخ نمی‌دهد، زیرا همواره با خطای اندازه گیری مواجه هستیم. در عمل هم‌خطی کامل هنگامی رخ می‌دهد که کاربر به اشتباه در مدل رگرسیونی یکی از متغیرها را تابعی از متغیر دیگر قرار دهد، یا وقتی تعداد متغیرهای تبیینی بیش‌تر از تعداد مشاهدات باشد. برای اطلاعات بیش‌تر درباره منابع پیدایش هم‌خطی به روزبه (۱۳۹۰) مراجعه کنید. بنابراین لازم است بدانیم که چگونه وجود هم‌خطی را کشف کنیم تا مواظب پیامدهای ممکن باشیم. روش‌های مختلفی برای تشخیص هم‌خطی وجود دارند. در این‌جا به برخی از این روش‌ها اشاره می‌کنیم.

۱. عامل تورم واریانس (VIF): یکی از روش‌های معروف برای تشخیص هم‌خطی استفاده از کمیت VIF است که به‌صورت

$$VIF_k = \frac{1}{1 - R_k^2}, \qquad k = 1, \ldots, p$$

محاسبه می‌شود. واضح است که اگر x_k رابطه خطی قوی با سایر متغیرهای تبیینی داشته باشد، آنگاه R_k^2 نزدیک به یک بوده و VIF_k بزرگ می‌شود. اگر x_k بر $p-1$ متغیر تبیینی دیگر عمود باشد، آنگاه R_k^2 صفر بوده و VIF_k برابر یک خواهد شد. بنابراین انحراف VIF_k از مقدار یک، انحراف از متعامد بودن و گرایش به هم‌خطی را نشان می‌دهد. اگر هر یک از مقادیر VIF بیش‌تر از ۱۰ باشد، هم‌خطی بین متغیرهای تبیینی وجود دارد و اگر هر یک از مقادیر آن بیش‌تر از ۳۰ باشد، هم‌خطی بسیار شدید است و ممکن است موجب بروز مشکلات در برآورد شود (گجراتی ، ۲۰۰۴). چون $\mathrm{Var}(\widehat{\beta}_k) = VIF_k \sigma^2$، می‌توان VIF_k را به عنوان عاملی دانست که به وسیله آن واریانس $\widehat{\beta}_k$ به دلیل وابستگی خطی بین متغیرهای تبیینی افزایش یافته است. به عبارت دیگر، VIF_k میزان افزایش در واریانس $\widehat{\beta}_k$ به واسطه ارتباط خطی x_k با بقیه متغیرهای تبیینی را نسبت به واریانس آن در صورت ناهمبسته بودن متغیر x_k با سایر متغیرهای تبیینی اندازه می‌گیرد. این مطلب دلیل نام‌گذاری این کمیت خاص را نشان می‌دهد.

۲. عدد شرطی[19]: مقادیر ویژه ماتریس $\boldsymbol{X}^\top \boldsymbol{X}$ می‌توانند برای اندازه‌گیری میزان هم‌خطی مورد استفاده قرار گیرند. یک روش متداول برای کشف وجود هم‌خطی محاسبه عدد شرطی به‌صورت

$$\kappa = \sqrt{\frac{\max_{1 \leqslant k \leqslant p} \lambda_k}{\min_{1 \leqslant k \leqslant p} \lambda_k}}$$

[19] Condition number

است که در آن $\lambda_1, \ldots, \lambda_p$ مقادیر ویژه ماتریس $\boldsymbol{X}^\top \boldsymbol{X}$ هستند. بلسلی و همکاران (۱۹۸۰) پیشنهاد کردند که اگر عدد شرطی برای ماتریس $\boldsymbol{X}^\top \boldsymbol{X}$ بزرگ‌تر از ۱۰ باشد، آنگاه در ماتریس طرح هم‌خطی وجود دارد. اگر $۳۰ < \kappa < ۱۰۰$ باشد، هم‌خطی شدید و چنانچه $\kappa > ۱۰۰$ باشد هم‌خطی بسیار جدی است. برای اطلاعات بیش‌تر در باره عدد شرطی به آرست (۱۳۹۵) مراجعه کنید.

یکی از روش‌های مقابله با اثرات نامطلوب هم‌خطی، استفاده از برآوردگرهای اریب است که رگرسیون ریج یکی از مهم‌ترین و کاراترین روش‌هاست. هورل و کنارد (۱۹۷۰) بر اساس روش منظم‌سازی تیکونوف (۱۹۴۳)، برآوردگر رگرسیونی ریج را معرفی کردند. این روش با اضافه کردن جمله منظم‌ساز، $\lambda\boldsymbol{\beta}^\top\boldsymbol{\beta}$، به معادلات برآورد، به نوعی یک برآوردگر رگرسیونی تاوانیده به‌شمار می‌آید. اگر چه برآوردگر حاصل اریب است، با این وجود برآوردهایی حاصل می‌شود که واریانس آن در مقایسه با برآوردگر OLS کمتر است. این کاهش واریانس، کاهش MSE را به دنبال خواهد داشت. در نتیجه برآوردگر ریج در مقایسه با برآوردگر OLS پایدارتر بوده و دقت پیش بینی را بهبود می‌بخشد. توضیحات بیش‌تر درباره این روش در زیربخش ۱.۳.۱.۲ ارائه می‌شود.

هنگامی‌که تعداد متغیرهای تبیینی بزرگ است، هم‌خطی می‌تواند جدی باشد. ورود تعداد متغیرهای فراوان به مدل اگرچه انعطاف‌پذیری مدل را افزایش می‌دهد، اما علاوه بر ایجاد هم‌خطی شدید در ماتریس طرح، می‌تواند باعث کاهش توان پیشگویی و مشکل شدن تفسیر مدل شود. برآوردگر ریج اگرچه توان پیشگویی بالایی دارد، اما مدل پیچیده‌ای را نتیجه می‌دهد، یعنی تعداد پارامترهای آن بزرگ است. تفسیر یک مدل پیچیده بسیار مشکل است. در مدل‌های با بعد بالا، معمولا تنها تعداد اندکی از متغیرهای تبیینی با متغیر پاسخ مرتبط هستند. بهترین روش در این حالت، حذف متغیرهای زائد و مدل‌بندی با زیرمجموعه‌ای از موثرترین متغیرهای تبیینی است. تاکنون روش‌های مختلف برای انتخاب متغیر معرفی شده‌اند. در ادامه به مهم‌ترین روش‌های انتخاب متغیر می‌پردازیم. این روش‌ها شامل روش‌های انتخاب متغیر کلاسیک و روش‌های مبتنی بر تابع تاوان هستند.

۲.۱.۲ انتخاب متغیر با معیارهای کلاسیک

تاکنون الگوریتم‌های متعددی برای انتخاب متغیر در مدل‌های خطی معرفی شده است. این الگوریتم‌ها براساس معیارهایی مانند معیار اطلاع آکائیک (AIC)، معیار اطلاع بیزی[۲۰] (BIC)، اعتبارسنجی متقابل تعمیم‌یافته[۲۱] (GCV)، ضریب تعیین تعدیل‌شده (R^2_{adj})، مجموع توان‌های دوم باقیمانده[۲۲] (RSS) و C_p–مالو[۲۳]، زیرمجموعه‌های متفاوتی از متغیرهای تبیینی را به عنوان بهترین زیرمجموعه

[20] Bayesian information criterion
[21] Generalized Cross-Validation
[22] Residual sum of squares
[23] Mallows's Cp

انتخاب می‌کنند. از جمله این روش‌ها می‌توان انتخاب بهترین زیرمجموعه[۲۴]، انتخاب پیش‌رو[۲۵]، روش حذفی پس‌رو[۲۶]، و روش گام به گام[۲۷] را نام برد.

در روش انتخاب بهترین زیرمجموعه، تمام زیرمجموعه‌های ممکن را درنظر گرفته و سپس براساس معیارهای فوق، بهترین زیرمجموعه از مدل کامل انتخاب می‌شود. در این حالت، اگر p تعداد کل متغیرهای تبیینی باشد، تعداد 2^p زیرمجموعه باید مورد بررسی قرار گیرند. به عنوان مثال، اگر $p = 10$، آنگاه باید ۱۰۲۴ مدل ممکن درنظر گرفته‌شوند و برای $p = 20$ باید بیش از یک میلیون مدل ممکن بررسی شوند. بنابراین با افزایش تعداد p، زیرمجموعه‌های ممکن به سرعت افزایش می‌یابد. لذا روش انتخاب بهترین زیر مجموعه از نظر محاسباتی، حتی در صورت استفاده از کامپیوترهای با سرعت بالا، برای $p > 40$ قابل استفاده نیست. در این راستا، دراپر و اسمیت (۱۹۸۱) روش‌های انتخاب پیش‌رو، حذفی پس‌رو و گام به گام را معرفی کردند. روش انتخاب پیش‌رو با مدلی که تنها شامل عرض از مبدا است، شروع می‌شود و سپس در هر گام متغیری که بیش‌ترین تاثیر را در بهبود مدل داشته باشد، انتخاب می‌شود. این رویه تا زمانی ادامه می‌یابد که تاثیر هیچ یک از متغیرهای باقیمانده در بهبود مدل معنی‌دار نباشد. برای اندازه‌گیری میزان تأثیر هر یک از متغیرها در بهبود مدل، از آماره‌هایی مانند R^2، RSS، F و t استفاده می‌شود. در روش حذفی پس‌رو، فرآیند با مدل کامل شروع می‌شود و هر بار یک متغیر از مدل حذف می‌شود. متغیر حذف‌شده همان متغیری است که کمترین تاثیر را در بهبود مدل دارد. به عبارت دیگر در هر گام، متغیر تبیینی با کوچک‌ترین آماره F جزئی برای حذف از مدل نامزد می‌شود. اگر آماره F جزئی متغیر منتخب کمتر از مقدار F^* باشد، متغیر مورد نظر از مدل حذف می‌شود. فرآیند حذف متغیرها زمانی متوقف می‌شود که کوچک‌ترین مقدار آماره F جزئی کوچک‌تر از F^* نباشد، یا همه متغیرهای مدل حذف شده باشند. در روش‌های انتخاب پیش‌رو و حذفی پس‌رو، هنگامی‌که یک متغیر حذف یا اضافه می‌شود، در مراحل بعدی معنی‌داری این متغیرها مجدداً بررسی نمی‌شود. روش گام به گام تعدیل روش انتخاب پیش‌رو است که در هر گام همه متغیرهای واردشده به مدل در گام‌های قبل، مجدداً ارزیابی می‌شوند. بنابراین یک متغیر تبیینی اضافه شده در گام قبلی ممکن است در گام بعدی از مدل حذف شود. برای اطلاعات بیش‌تر درباره سایر روش‌های انتخاب متغیر می‌توان به دزیاک و همکاران (۲۰۰۵)، دسبولتس (۲۰۱۸) و دینگ (۲۰۱۸) مراجعه کرد.

اگرچه روش‌های انتخاب زیرمجموعه با معیارهای کلاسیک از ویژگی‌های نمونه‌ای مطلوبی برخوردارند، اما این روش‌ها پایدار نیستند، بدین معنی که یک تغییر بسیار کوچک در داده‌ها ممکن است مدل‌های کاملاً متفاوت را نتیجه دهد. لذا ممکن است این روش‌ها پیشگویی بدی را ایجاد کنند (بریمن ، ۱۹۹۶). علاوه بر این، زمانی که p بزرگ باشد، هزینه محاسباتی این روش‌ها بسیار بالاست. برای رفع این چالش‌ها، محققین رگرسیون تاوانیده را معرفی کردند که با انتخاب یک تابع

[24] Best subset selection
[25] Forward selection
[26] Backward elimination
[27] Stepwise

تاوان مناسب، دو عمل انتخاب متغیر و برآورد ضرایب را به‌طور همزمان انجام می‌دهد.

۳.۱.۲ انتخاب متغیر مبتنی بر تابع تاوان

در این بخش، به توصیف روش رگرسیون تاوانیده می‌پردازیم. بدیهی است وقتی تعداد متغیرهای تبیینی p زیاد است، اندازه مجموع قدر مطلق ضرایب β_k، $k = ۱, \ldots, p$، افزایش می‌یابد.

از آنجایی‌که در روش کمترین توان‌های دوم به دنبال کمینه کردن $\|\varepsilon\|^۲ = \varepsilon^\top \varepsilon$ نسبت به $\boldsymbol{\beta}$ هستیم، پس چنان‌چه بخواهیم علاوه بر برآورد ضرایب رگرسیونی، تعداد متغیرهای مدل را نیز کاهش دهیم، کافی است بزرگی قدرمطلق ضرایب یا تابعی از آن را نیز به‌طور همزمان کمینه کنیم. بنابراین در رگرسیون تاوانیده، محک $\|\varepsilon\|^۲$ را تحت این قید که نباید مجموع تابعی از قدرمطلق ضرایب رگرسیونی مانند $\sum_{k=۱}^{p} P(|\beta_k|)$ بزرگ باشد، کمینه می‌کنند. این یک مسئله بهینه‌سازی مقید[۲۸] (CO) است. به‌طور دقیق‌تر، علاقه‌مندیم مسئله CO به‌صورت

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\varepsilon}\|^۲ \qquad s.t. \qquad \sum_{k=۱}^{p} P_\lambda(|\beta_k|) \leqslant s, \tag{۳.۲}$$

را حل کنیم که در آن $P_\lambda(|.|)$ تابع تاوان نامیده می‌شود و λ پارامتر تاوان است که میزان تاوان را کنترل می‌کند. یکی از راه‌های حل مسئله CO استفاده از روش ضریب لاگرانژ[۲۹] است. بنابراین مسئله CO معادل با مسئله

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \Big\{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^۲ + \lambda \sum_{k=۱}^{p} P(|\beta_k|) \Big\} \tag{۴.۲}$$

است. با دقت در ساختار مسئله (۴.۲) دیده می‌شود که به تابع مجموع توان‌های دوم خطا در رگرسیون معمولی، عبارت تاوان $\Big\{\lambda \sum_{k=۱}^{p} P(|\beta_k|)\Big\}$ اضافه شده است. در اینجا، $P_\lambda(|.|) = \lambda P(|.|)$ تابع تاوان و $\lambda > ۰$ پارامتر تنظیم‌کننده[۳۰] است که میزان انقباض را برحسب بزرگی ضرایب رگرسیونی کنترل می‌کند. در این حالت بین λ در (۴.۲) و s در (۳.۲) رابطه‌ای معکوس وجود دارد. به‌طور دقیق‌تر، اگر بخواهیم تعداد متغیرهای بیش‌تری از مدل حذف شوند، کافی است s را کوچک یا λ را بزرگ انتخاب کنیم و برعکس. پارامتر تنظیم‌کننده λ موازنه‌ای بین خطای پیشگویی و پیچیدگی مدل برقرار می‌کند. جواب مسئله بهینه‌سازی (۴.۲) به شدت به مقدار λ وابسته است. تعیین مقدار مناسب برای λ یک مسئله چالش برانگیز است. روش‌های مختلف انتخاب مقدار بهینه λ در زیربخش ۳.۳.۱.۲ داده شده است. بنابراین با انتخاب یک مقدار بهینه برای λ، مقادیر برخی از پارامترها ممکن است صفر شوند که در این صورت متغیر تبیینی متناظر با آن پارامتر، به عنوان متغیر بی‌تاثیر بر متغیر پاسخ شناخته شده و بدین ترتیب مجموعه متغیرهای مهم و بی‌اهمیت از یکدیگر تشخیص داده می‌شوند و در اصطلاح ''انتخاب متغیر'' صورت می‌گیرد. در ادامه چند تابع تاوان مهم را معرفی می‌کنیم.

[28] Constraint optimization

[29] Lagrange multiplier

[30] Tuning parameter

۱.۳.۱.۲ انواع تابع تاوان

تاکنون توابع تاوان زیادی توسط افراد مختلف معرفی شده‌اند. در این بخش به معرفی برخی از توابع تاوان معروف می‌پردازیم.

۱. **تاوان L_q:** با فرض $P(|\beta_k|) = |\beta_k|^q$ در رابطه (۳.۲)، تابع تاوان بریج[۳۱] حاصل می‌شود. این تابع رگرسیون بریج را نتیجه می‌دهد که توسط فرانک و فریدمن (۱۹۹۳) معرفی شد. برآوردگر بریج به‌صورت

$$\widehat{\boldsymbol{\beta}}^{Bridge} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\Big\{\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\sum_{k=1}^{p}|\beta_k|^q\Big\}, \qquad 0 < q \leqslant 2.$$

تعریف می‌شود. این تابع تاوان، برخی از توابع تاوان مهم را به عنوان یک حالت خاص شامل می‌شود.

رگرسیون ریج: برای حالت خاص $q = 2$، مدل رگرسیونی حاصل را رگرسیون ریج[۳۲] گویند و برآورد پارامترها به‌صورت

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{Ridge} &= \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\Big\{\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\sum_{k=1}^{p}|\beta_k|^2\Big\}\\ &= (\boldsymbol{X}^\top\boldsymbol{X} + K\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top\boldsymbol{y},\end{aligned}$$

به‌دست می‌آید که $K = 2\lambda$ را پارامتر تنظیم‌کننده یا پارامتر ریج گویند. یک مزیت رگرسیون ریج پیاده‌سازی آسان و صورت صریح و بسته جواب آن است. اگر ماتریس طرح $\boldsymbol{X}$ دارای رتبه کامل ستونی نباشد، ماتریس $\boldsymbol{X}^\top\boldsymbol{X}$ وارون‌پذیر نخواهد بود و لذا جواب یکتایی برای برآوردگر وجود نخواهد داشت. اما برای هر ماتریس دلخواه $\boldsymbol{X}$ و $K > 0$، ماتریس $\boldsymbol{X}^\top\boldsymbol{X} + K\boldsymbol{I}_p$ همواره وارون‌پذیر است و در نتیجه پاسخ یکتا برای برآوردگر ریج وجود دارد. برآوردگر ریج اریب است، اما در مقایسه با برآوردگر OLS دارای واریانس کوچکتری است. همچنین همواره λای وجود دارد به‌طوری‌که $MSE(\widehat{\boldsymbol{\beta}}^{Ridge}) < MSE(\widehat{\boldsymbol{\beta}}^{OLS})$ (صالح و همکاران ، ۲۰۱۹). بنابراین رگرسیون ریج پیشگویی بهتری از رگرسیون کمترین توان‌های دوم معمولی ارائه می‌دهد. با این وجود برآوردگر ریج اگرچه ضرایب رگرسیون را منقبض می‌کند اما هیچ یک از ضرایب را صفر برآورد نمی‌کند، یعنی این روش توانایی انتخاب متغیر ندارد. این موضوع خللی در پیشگویی ایجاد نمی‌کند، اما با توجه به حضور همه متغیرها در مدل، تفسیر مدل برازش‌شده وقتی p بزرگ است به سادگی امکان‌پذیر نیست. در حالتی که $\boldsymbol{X}^\top\boldsymbol{X} = \boldsymbol{I}_p$، ماتریس طرح متعامد خواهد بود. در این حالت برآوردگر ریج به‌صورت

$$\widehat{\boldsymbol{\beta}}^{Ridge} = \frac{1}{1+K}\widehat{\boldsymbol{\beta}}^{OLS}$$

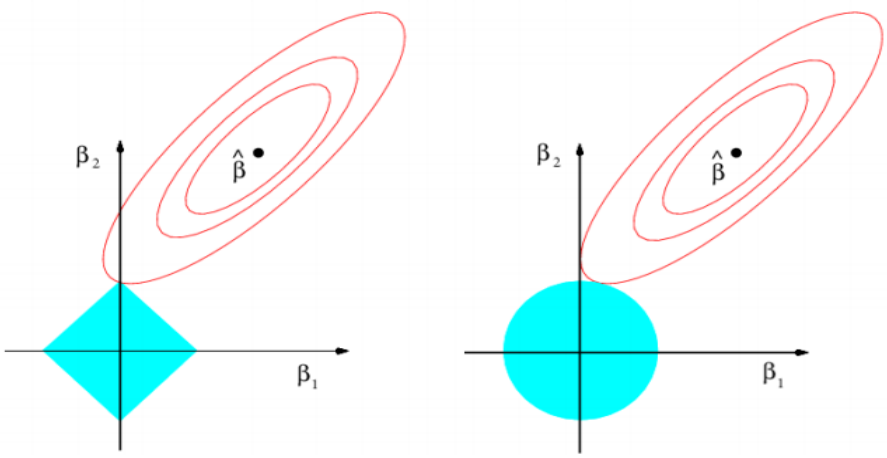
است. این رابطه ماهیت انقباضی رگرسیون تاوانیده را نشان می‌دهد، چون با افزایش K ضرایب به سمت صفر منقبض می‌شوند. هنگامی که $K = 0$، برآوردگرهای OLS و ریج معادل‌اند.

[31] Bridge
[32] Ridge

رگرسیون لاسو: در این حالت مدل رگرسیونی حاصل را که توسط تیبشیرانی (۱۹۹۶) معرفی شد، رگرسیون لاسو گویند. از دیدگاه آماری، لاسو مخفف عملگر انتخاب‌کننده، منقبض‌کننده و کمینه‌کننده از جنس قدرمطلق است. در رگرسیون لاسو برآورد ضرایب رگرسیونی به‌صورت

$$\widehat{\boldsymbol{\beta}}^{Lasso} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\Big\{\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\sum_{j=1}^{p}|\beta_j|\Big\}$$

به‌دست می‌آید که در آن λ پارامتر تنظیم‌کننده است و سطح تنکی (تعداد ضرایب صفر) را کنترل می‌کند. وقتی $\lambda \to 0$، آنگاه $\widehat{\boldsymbol{\beta}}^{Lasso} \to \widehat{\boldsymbol{\beta}}^{OLS}$ و وقتی $\lambda \to \infty$، آنگاه $\widehat{\boldsymbol{\beta}}^{Lasso} \to 0$. به عبارت دیگر، هر چه تاوان بزرگ‌تری به کار رود، تعداد بیش‌تری از ضرایب به سمت صفر منقبض می‌شوند. برآورد بدست‌آمده با رگرسیون لاسو تنک است، یعنی برخی از ضرایب صفر برآورد می‌شوند. به عبارت دیگر این روش انتخاب متغیر و برآورد پارامتر را به‌طور همزمان انجام می‌دهد. تابع تاوان قدرمطلق باعث می‌شود که جواب صریحی برای برآوردگر لاسو وجود نداشته باشد. تیبشیرانی این برآوردگر را از طریق برنامه‌ریزی درجه دوم[۳۳] به‌دست آورده بود. سپس افرون و همکاران (۲۰۰۴) رگرسیون کمترین زاویه[۳۴] را معرفی نمودند که برآوردهای لاسو را نتیجه می‌دهد و هزینه محاسباتی آن با برآوردهای کمترین توان‌های دوم برابر است.



شکل ۱.۲: منحنی‌های تراز و نواحی محدودیت رگرسیون لاسو (نمودار چپ) و رگرسیون ریج (نمودار راست). ناحیه‌های محدودیت لاسو و ریج به‌ترتیب به‌صورت $|\beta_1| + |\beta_2| \leqslant s$ و $\beta_1^2 + \beta_2^2 \leqslant s$ هستند. بیضی‌ها، منحنی‌های تراز تابع خطای برآورد کمترین توان های دوم هستند.

برای روشن‌تر شدن مفهوم انتخاب متغیر توسط تابع تاوان لاسو، حالت خاصی از مدل رگرسیونی را به‌صورت

$$\boldsymbol{y} = \beta_1\boldsymbol{x}_1 + \beta_2\boldsymbol{x}_2 + \boldsymbol{\varepsilon},$$

درنظر بگیرید. بنابراین مجموع توان‌های دوم خطا $Q(\boldsymbol{\beta})$، تابعی از دو پارامتر β_1 و β_2 است که در

[33] Quadaratic programming
[34] Least angle regression

آن

$$Q(\boldsymbol{\beta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_1 x_{i1} - \beta_2 x_{i2})^2.$$

نمودار تابع $Q(.)$ یک رویه درجه دوم به‌صورت

$$Q(\boldsymbol{\beta}) = a_0 + a_1\beta_1 + a_2\beta_2 + a_3\beta_1^2 + a_4\beta_2^2 + a_5\beta_1\beta_2$$

است که منحنی‌های تراز آن در شکل ۱.۲ نشان داده شده است. اکنون شرط‌های کمینه‌سازی ریج و لاسو را به‌صورت

$$\beta_1^2 + \beta_2^2 = s \qquad \text{و} \qquad |\beta_1| + |\beta_2| = s$$

درنظر بگیرید که شکل هندسی آن‌ها به‌صورت دایره برای رگرسیون ریج و لوزی برای رگرسیون لاسو نشان داده شده است. با توجه به نمودار سمت چپ شکل ۱.۲، در رگرسیون لاسو می‌تواند حالتی رخ دهد که یکی از ضرایب صفر شده و دیگری همواره برابر s باشد، در صورتی‌که در نمودار سمت راست این امکان وجود ندارد، به عبارت دقیق‌تر، در قاب سمت راست، اگر یکی از ضرایب صفر شود، آن‌گاه دیگری می‌تواند $\sqrt{s}$، و نه s، باشد. بنابراین برای لاسو امکان صفر برآورد شدن ضرایب وجود دارد، اما برای ریج این اتفاق هرگز رخ نمی‌دهد. در صورت صفر شدن هر یک از ضرایب در رگرسیون لاسو، متغیر متناظر با آن ضریب به عنوان متغیر بی‌اهمیت شناخته می‌شود. لازم به ذکر است که در هر دو روش لاسو و ریج، با انتخاب مقدار کوچک s، هر دو پارامتر β_1 و β_2 (در حالت دو بعدی) تا حد ممکن می‌توانند کوچک شوند. بنابراین هر دو روش ریج و لاسو انقباضی هستند.

اگر ماتریس طرح متعامد باشد، آن‌گاه می‌توان برآوردگر لاسو را به عنوان تابعی از برآوردگر کمترین توان‌های دوم معمولی β_k به‌صورت

$$\widehat{\beta}_k^{Lasso} = \operatorname{sgn}(z_k)(|z_k| - \lambda)_+, \qquad j = 1, \ldots, p, \tag{۵.۲}$$

به‌دست آورد که در آن $\operatorname{sgn}(x) = I(x \geqslant 0) - I(x < 0)$، $I(A)$ تابع نشانگر، $x_+ = \max\{0, x\}$ و $z_k = \boldsymbol{X}_k^\top \boldsymbol{y}$ برآورد OLS ضریب β_k بوده که $\boldsymbol{X}_k$، ستون kام ماتریس طرح $\boldsymbol{X}$ است. برای اطلاعات بیش‌تر به فن و لی (۲۰۰۱) مراجعه کنید. برآوردگر (۵.۲) را به‌صورت

$$\widehat{\beta}_k^{Lasso} = S(z_k|\lambda), \tag{۶.۲}$$

نشان می‌دهیم که $S(.|\lambda)$ عملگر آستانه‌ای نرم نامیده می‌شود. رابطه (۶.۲) در بررسی ویژگی‌های مطلوب رگرسیون لاسو مفید است که در پایان این بخش به این موضوع خواهیم پرداخت.

رگرسیون لاسو با اینکه در بسیاری از مسائل انتخاب متغیر عملکرد خوبی از خود نشان داده، اما دارای برخی محدودیت‌ها است. به عنوان مثال، لاسو برای یک مدل رگرسیونی خطی با p متغیر پیشگو و n مشاهده، حداکثر n متغیر را انتخاب می‌کند. بنابراین اگر تعداد متغیرهای تبیینی معنی‌دار در مدل بیش‌تر از n باشد، توسط لاسو انتخاب نمی‌شوند. لاسو در بین یک گروه از متغیرهای به شدت همبسته فقط یک متغیر را انتخاب می‌کند (زو و هستی ، ۲۰۰۵). برای حالت معمولی $n > p$، اگر

همبستگی بالایی بین متغیرهای پیشگو وجود داشته باشد، کارآیی پیشگویی لاسو به خوبی برآوردگر ریج نیست (تیبشیرانی ، ۱۹۹۶). رگرسیون لاسو برآوردگری را نتیجه می‌دهد که اریبی آن زیاد است. در حالتی که ماتریس طرح متعامد است، می‌توان نشان داد که برای $k = ۱, \ldots, p$ داریم

$$\begin{cases} \mathrm{E}\,|\widehat{\beta}_k^{Lasso} - \beta_k| = ۰, & \beta_k = ۰ \\ \mathrm{E}\,|\widehat{\beta}_k^{Lasso} - \beta_k| \approx \beta_k, & \beta_k \in [۰, \lambda] \\ \mathrm{E}\,|\widehat{\beta}_k^{Lasso} - \beta_k| \approx \lambda & |\beta_k| > \lambda \end{cases} . \tag{۷.۲}$$

بنابراین مقدار اریبی برآورد لاسو برای ضرایب بزرگ، نزدیک به λ است (برهنی و هووانگ ، ۲۰۱۱). برای از بین بردن این محدودیت‌ها، توابع تاوان متعددی معرفی شدند که در ادامه به برخی از آن‌ها اشاره می‌کنیم. قبل از معرفی این توابع، ابتدا ویژگی پیشگویی را تعریف می‌کنیم.

ویژگی پیشگویی: فرض کنید $\mathcal{A} = \{k : \beta_k \neq ۰\}$ مجموعه اندیس ضرایب غیرصفر باشد، به‌طوری که $|\mathcal{A}| = p_۰ < p$. در این صورت، مدل درست تنها به زیرمجموعه‌ای از متغیرها ($p_۰$ تا متغیر) وابسته است. منظور از مدل درست مدلی است که تنها شامل ضرایب غیرصفر باشد. براساس فن و لی (۲۰۰۱)، یک روش برآورد ضرایب دارای ویژگی پیشگویی[۳۵] است، اگر به‌طور مجانبی دارای خواص زیر باشد:

۱. بتواند ضرایب غیرصفر را به درستی شناسایی کند، یعنی اگر $\widehat{\mathcal{A}} = \{k : \widehat{\beta}_k \neq ۰\}$، آن‌گاه $\lim_{n\to\infty} \mathbb{P}(\widehat{\mathcal{A}} = \mathcal{A}) = ۱$.

۲. دارای نرخ برآورد بهینه باشد، به عبارت دیگر

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\mathcal{A}} - \boldsymbol{\beta}_{\mathcal{A}}) \xrightarrow{d} \mathcal{N}(۰, \boldsymbol{\Sigma}^*)$$

که $\boldsymbol{\Sigma}^*$ ماتریس کوواریانس زیر مدل درست یعنی مدلی برپایه $\mathcal{A}$، است.

بنابراین مطلوب است که یک برآوردگر دارای ویژگی پیشگویی باشد. زو (۲۰۰۶) نشان داد که برآوردگر لاسو دارای ویژگی پیشگویی نیست. با توجه به رابطه (۷.۲) ملاحظه می‌شود که مقدار اریبی برآوردگر لاسو توسط λ تعیین می‌شود. بنابراین یک روش کاهش اریبی لاسو، استفاده از تاوان موزون به‌صورت $\lambda_k = \omega_k \lambda$ است. اگر بتوانیم وزن‌ها را طوری انتخاب کنیم که به ضرایب بزرگ‌تر تاوان کوچتری تخصیص داده شود، آن‌گاه با حفظ ویژگی تنکی، اریبی برآورد لاسو کاهش می‌یابد. بنابراین زو (۲۰۰۶) لاسوی سازوار را معرفی کرد و نشان داد که این روش از ویژگی پیشگویی برخوردار است.

۲. تاوان لاسوی سازوار: صورت کلی تابع تاوان لاسوی سازوار مانند لاسو است، اما برخلاف لاسو برای ضرایب مختلف، تاوان‌های مختلفی درنظر می‌گیرد. برآوردگر لاسوی سازوار به‌صورت

$$\widehat{\boldsymbol{\beta}}^{aLasso} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\Big\{ \frac{۱}{۲}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^۲ + \lambda \sum_{k=۱}^{p} \omega_k |\beta_k| \Big\}$$

[35] Oracle property

بیان می‌شود، که ω_k وزن β_k است تا ضرایب کوچک را بیش‌تر از ضرایب بزرگ منقبض کند. معمولا در رگرسیون لاسوی تطبیقی، وزن $\omega_k = \frac{1}{|\widehat{\beta}_k^\circ|^\gamma}$ را انتخاب می کنند که $\gamma > 0$ و $\widehat{\beta}_k^\circ$ یک برآوردگر $\sqrt{n}$-سازگار از β_k مانند برآوردگر OLS، ریج یا لاسو است.

۳. تاوان اسکد: فن و لی (۲۰۰۱) یک تابع تاوان مقعر به نام اسکد[۳۶] ($SCAD$) را معرفی کردند و نشان دادند که این تابع تاوان، از همه ویژگی‌های مطلوب یک تابع تاوان برخوردار است (زیر بخش ۲.۳.۱.۲ را ببینید). آن‌ها همچنین ویژگی پیشگویی را نیز برای این روش اثبات کردند. برآوردگر اسکد به‌صورت

$$\widehat{\boldsymbol{\beta}}^{SCAD} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\Big\{\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\sum_{k=1}^{p} P(|\beta_k|;a,\lambda)\Big\}$$

است که $P(.;a,\lambda)$ تابع تاوان اسکد است و به‌صورت

$$P(\beta;a,\lambda) = \begin{cases} \lambda\beta, & 0 \leqslant |\beta| \leqslant \lambda \\ \frac{2a\lambda\beta-\beta^2-\lambda^2}{2(a-1)}, & \lambda < |\beta| < a\lambda \\ \frac{\lambda^2(a+1)}{2}, & |\beta| \geqslant a\lambda \end{cases} \qquad (۸.۲)$$

تعریف می‌شود که $\lambda > 0$ و $a > 2$ پارامترهای تاوان می‌باشند و اندازه انقباض ضرایب را کنترل می‌کنند. فن و لی (۲۰۰۱) مقدار $a = 3.7$ را پیشنهاد دادند. شکل ۲.۲ نمودار تابع اسکد را به ازای $\lambda = 1$ نشان می‌دهد. با توجه به شکل ۲.۲ و رابطه (۸.۲) ملاحظه می‌شود که در فاصله $|\beta| \leqslant \lambda$ تابع تاوان اسکد بر لاسو منطبق است و پس از آن تا زمانی که $|\beta| \leqslant a\lambda$، تاوان اسکد یک تابع درجه دوم است. سپس به ازای $|\beta| > a\lambda$ به یک تابع ثابت تبدیل می‌شود. هم‌چنین، با توجه به (۸.۲)، وقتی $a \to \infty$، برآوردگر اسکد با برآوردگر لاسو معادل است، یعنی تابع تاوان اسکد در حالت خاص $a = \infty$ تابع تاوان لاسو را نتیجه می‌دهد.

با فرض متعامد بودن ماتریس طرح، برآوردگرهای اسکد ضرایب را می‌توان به‌صورت

$$\widehat{\beta}_j^{SCAD} = \begin{cases} S(z_k|\lambda), & |z_k| \leqslant 2\lambda \\ \frac{a-1}{(a-2)}S(z_k|\frac{a\lambda}{a-1}), & 2\lambda < |z_k| \leqslant a\lambda \\ |z_k|, & |z_k| > a\lambda \end{cases}$$

نشان داد که z_k برآورد OLS ضریب β_k و $S(.|\lambda)$ عملگر آستانه‌ای نرم در (۵.۲) است.

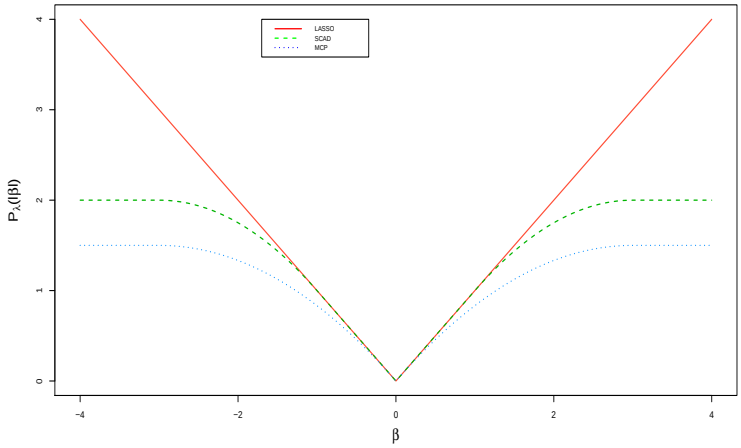
۴. تاوان MCP: ژانگ (۲۰۱۰) تابع تاوان دیگری به نام MCP به‌صورت

$$P(\beta;\lambda,\gamma) = \lambda\int_0^\beta (1-\frac{x}{\gamma\lambda})_+ dx; \qquad \beta \geqslant 0,$$

معرفی کرد که در آن λ پارامتر تاوان و γ میزان مقعر بودن تابع $P(.)$ را کنترل می‌کند. در این حالت، برآوردگر MCP به‌صورت

$$\widehat{\boldsymbol{\beta}}^{MCP} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min}\Big\{\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\sum_{j=1}^{p} P(|\beta_j|;\lambda,\gamma)\Big\},$$

[36] Smoothly clipped absolute deviation



شکل ۲.۲: مقایسه نمودارهای توابع تاوان لاسو، اسکد و MCP.

تعریف می‌شود. تابع تاوان MCP به ازای $\gamma = \infty$ تابع تاوان لاسو را نتیجه می‌دهد. تحت فرض متعامد بودن ماتریس طرح، مشابه برآوردگرهای لاسو و اسکد، برآوردگرهای MCP ضرایب را می‌توان به‌صورت

$$\hat{\beta}_k^{MCP}(\gamma,\lambda) = \begin{cases} \frac{\gamma}{\gamma-1} S(z_k|\lambda), & |z_k| \leqslant \gamma\lambda \\ z_k, & |z_k| > \gamma\lambda \end{cases}$$

نوشت. شکل ۲.۲ نمودار تابع MCP را به ازای $\lambda = 1$ و $\gamma = 3$ نشان می‌دهد. تابع تاوان MCP در فاصله $|\beta| \leqslant \gamma\lambda$ یک تابع درجه دوم و پس از آن یک تابع ثابت است.

۲.۳.۱.۲ ویژگی‌های یک تابع تاوان مناسب

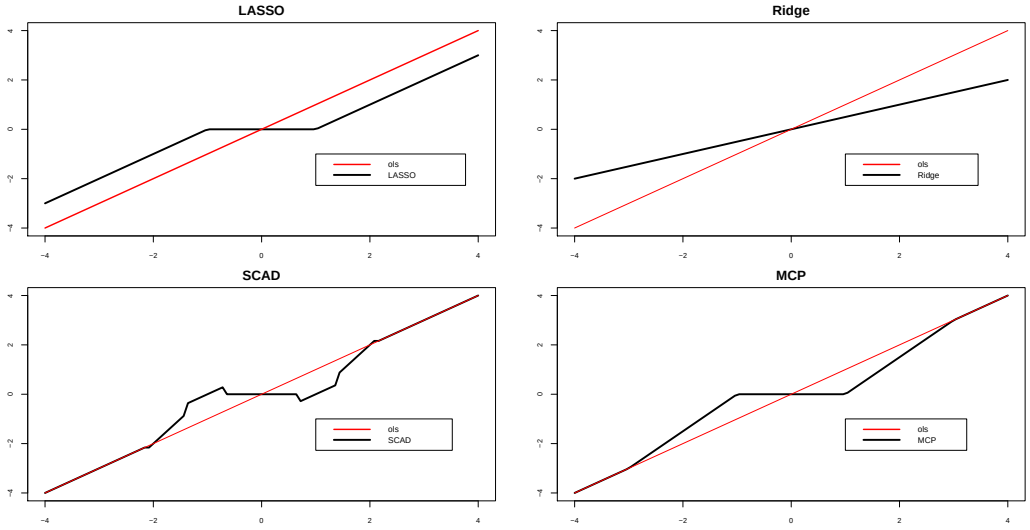
برای استفاده از رگرسیون تاوانیده، سوالی که پیش می‌آید این است که از چه نوع تابع تاوان باید استفاده کرد. فن و لی (۲۰۰۱) نشان دادند که تابع تاوان خوب باید برآوردگری با سه ویژگی مطلوب زیر را نتیجه دهد:

۱. تنکی: برآوردگر نتیجه شده باید به‌طور خودکار ضرایب برآوردشده‌ای که مقدار کوچک دارند را برابر صفر قرار دهد تا متغیرهای مناسب را انتخاب کند. این کار پیچیدگی مدل را کاهش می‌دهد.

۲. نااریبی: برآوردگر به‌دست آمده برای ضرایب رگرسیونی که مقادیر واقعی آن‌ها بزرگ است، تقریبا نااریب باشد. این ویژگی اریبی مدل را کاهش می‌دهد.

۳. پیوستگی: برآوردگر نتیجه‌شده پیوسته باشد تا بتواند ناپایداری در پیشگویی مدل را کاهش دهد.

شکل ۳.۲ ارتباط برآوردگر OLS را با برآوردگرهای لاسو، ریج، اسکد و MCP با فرض متعامد بودن ماتریس طرح $\boldsymbol{X}$ نشان می‌دهد. محور افقی نشان‌دهنده برآوردگر OLS و محور عمودی



شکل ۳.۲: مقایسه نمودارهای برآوردهای کمترین توان‌های دوم تاوانیده (PLS) در مقابل برآورد کمترین توان‌های دوم معمولی (OLS) وقتی که ماتریس طرح متعامد است. محور افقی OLS و محور عمودی PLS است.

برآوردهای کمترین توان‌های دوم تاوانیده است. چون برآوردگر OLS ناااریب است، انتظار داریم که برای ضرایب بزرگ، برآورد تاوانیده برابر یا نزدیک به برآورد OLS باشد تا شرط ناااریبی برقرار شود. از طرفی برای برقراری شرط تنکی، لازم است روش کمترین توان‌های دوم تاوانیده، ضرایب رگرسیونی کوچک را صفر برآورد کند. همچنین لازم است برآوردگر موردنظر پیوسته باشد.

با توجه به شکل ۳.۲، پیوستگی برآوردگر لاسو بدیهی است. همچنین برای مقادیر کوچک ضرایب، برآورد لاسو برابر صفر است. در نتیجه شرط تنکی برقرار است. اما با بزرگ شدن مقادیر ضرایب، برآورد لاسو از برآورد OLS فاصله می‌گیرد، یعنی شرط ناااریبی برقرار نیست. در نتیجه عیب روش لاسو این است که برای مقادیر بزرگ ضرایب، برآوردهای اریب را نتیجه می‌دهد. برآورد ریج در مقایسه با لاسو، علی‌رغم پیوسته بودن، ضرایب کوچک را صفر برآورد نمی‌کند. علاوه بر این، با افزایش قدرمطلق ضرایب، مقدار اریبی به شدت افزایش می‌یابد. بنابراین، تابع تاوان ریج یک تابع مناسب برای انتخاب متغیر نمی‌باشد. با توجه به نمودار برآورد اسکد در مقابل برآورد OLS، برقراری ویژگی‌های تنکی و پیوستگی واضح است. علاوه بر این، در ابتدا میزان اریبی برآوردگر تاوانیده با اریبی برآوردگر لاسو معادل است، اما با افزایش قدر مطلق ضرایب، اریبی به سمت صفر تنزل می‌یابد. لذا ویژگی ناااریبی نیز برقرار است. همچنین، با توجه به شکل، تابع تاوان MCP نیز برآوردگری با ویژگی‌های مطلوب را نتیجه می‌دهد. رفتار تابع تاوان MCP بسیار شبیه اسکد است، با این تفاوت که با فاصله گرفتن ضرایب از صفر، اریبی برآوردگر MCP فوراً به سمت صفر می‌رود، اما اریبی برآوردگر اسکد در یک فاصله معین برابر اریبی لاسو است و سپس به سرعت کاهش می‌یابد. بنابراین از توابع تاوان MCP و اسکد می‌توان به عنوان توابع تاوان مناسب نام برد.

همچنین فن و لی (۲۰۰۱) نشان دادند که تحت شرایط کافی زیر برآوردگر کمترین توان‌های دوم تاوانیده دارای ویژگی‌های مطلوب فوق است:

۱. تنکی: برآوردگر دارای ویژگی تنکی است اگر $\inf_{\beta \geqslant ۰}\{\beta + P'_\lambda(\beta)\} > ۰$.

۲. نا اریبی: برآوردگر به‌دست‌آمده نااریب است اگر، برای مقادیر بزرگ β، داشته باشیم $P'_\lambda(\beta) = ۰$.

۳. پیوستگی: برآوردگر پیوسته است اگر و فقط اگر $\arg\inf_{\beta \geqslant ۰}\{\beta + P'_\lambda(\beta)\} = ۰$.

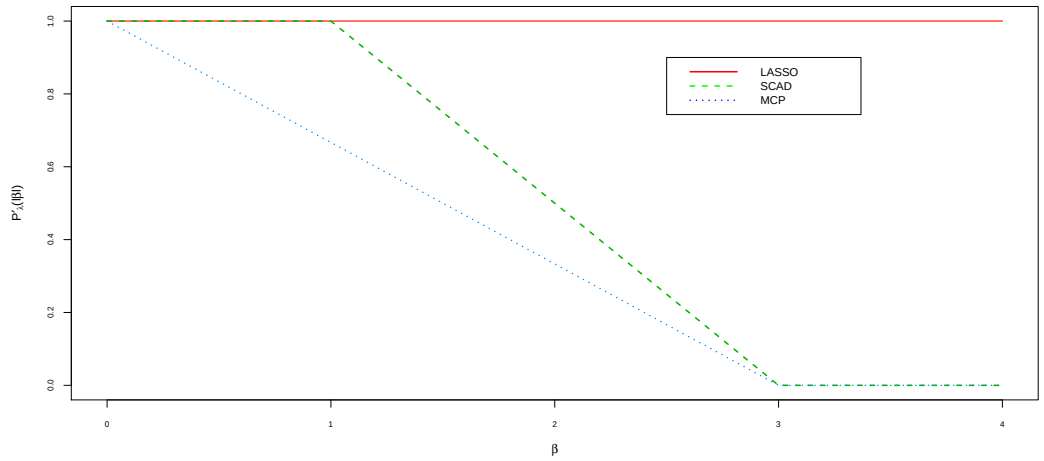
همان‌طور که ملاحظه می‌شود، شرایط کافی برای داشتن ویژگی‌های مطلوب بیان شده توسط فن و لی (۲۰۰۱) وابسته به مشتق تابع تاوان هست. علاوه بر این، شکل مشتق تابع تاوان در حل مسئله بهینه‌سازی تاوانیده تاثیر بسزایی دارد. مشتق تابع تاوان لاسو به‌صورت $P'_\lambda(\beta) = \lambda$ و مشتق توابع تاوان MCP و اسکد به‌ترتیب، به‌صورت

$$P'_\lambda(\beta; a, \lambda) = \lambda\Big\{I(\beta \leqslant \lambda) + \frac{(a\lambda - \beta)_+}{(a - ۱)\lambda} I(\beta > \lambda)\Big\}, \qquad \beta \geqslant ۰$$

و

$$P'_\lambda(\beta; \lambda, \gamma) = \lambda\Big(۱ - \frac{\beta}{\gamma\lambda}\Big), \qquad \beta \geqslant ۰$$

به‌دست می‌آید. بررسی شرایط کافی ویژگی‌های مطلوب توابع تاوان بسیار ساده است. به عنوان مثال، برای تابع تاوان لاسو، $\min_{\beta \geqslant ۰}\{\beta + P'_\lambda(\beta)\} = \lambda > ۰$ و $\arg\min_{\beta \geqslant ۰}\{\beta + P'_\lambda(\beta)\} = ۰$، اما $P'_\lambda(\beta) = \lambda > ۰$. بنابراین، تابع تاوان لاسو دارای ویژگی نااریبی نیست.



شکل ۴.۲: نمودار مشتقات توابع تاوان لاسو، اسکد و MCP.

نمودار مشتق توابع تاوان لاسو، اسکد و MCP به ازای $\lambda = ۱$ و $a, \gamma = ۳$ در شکل ۴.۲ داده شده است. با توجه به شکل می‌بینیم که نرخ تاوان توابع MCP و اسکد در مبدا با نرخ تاوان تابع لاسو برابر است، اما با افزایش قدر مطلق ضرایب، نرخ تاوان توابع MCP و اسکد به سمت صفر تنزل می‌کند. لازم به ذکر است که نرخ تاوان تابع اسکد در ابتدا با فاصله گرفتن ضرایب از صفر معادل نرخ تاوان لاسو است، سپس با افزایش قدر مطلق ضرایب به سرعت کاهش می‌یابد، اما تابع

تاوان MCP به محض خارج شدن از مبدا، نرخ تاوان آن فوراً به سمت صفر تنزل می‌یابد. علاوه بر این، شکل نشان می‌دهد که روش MCP مقعر مینیماکس است، یعنی در بین همه توابع تاوان پیوسته مشتق‌پذیر بر بازه $(۰,\infty)$ که برای هر $\beta \geqslant \gamma\lambda$ در دو شرط $P'_\lambda(\beta;\lambda) = ۰$ و $P'_\lambda(۰_+;\lambda) = \lambda$ صدق می‌کند، MCP بیشینه تقعر زیر را کمینه می‌کند

$$\mathcal{K} = \sup_{۰<\beta_۱<\beta_۲} \frac{P'_\lambda(\beta_۱;\lambda) - P'_\lambda(\beta_۲;\lambda)}{\beta_۲ - \beta_۱}.$$

مقعر مینماکس بودن تابع MCP بدین معنی است که در بین همه توابعی که دارای سه ویژگی مطلوب فوق هستند، MCP اندازه بیشینه تقعر را کمینه می‌کند. در شکل مشتق‌های توابع MCP و اسکد در نقاط ۰ و $\gamma\lambda = ۳$ برابر هستند، اما MCP در این ناحیه دارای تقعر $\mathcal{K} = \frac{۱}{\gamma} = \frac{۱}{۳}$ است، در حالی که اسکد دارای ماکزیمم تقعر $\mathcal{K} = \frac{۱}{\gamma-۱} = \frac{۱}{۲}$ است. لازم به ذکر است که اغلب توابع تاوانی که شرایط مطلوب تنکی، نااریبی و پیوستگی را دارا می‌باشند، توابع غیرمحدب هستند. این موضوع، مسئله بهینه‌سازی را با چالش مواجه می‌کند. برای اطلاعات بیش‌تر به برهنی و هووانگ (۲۰۱۹) مراجعه شود.

۳.۳.۱.۲ انتخاب پارامتر تاوان

انتخاب پارامتر تاوان در رگرسیون تاوانیده نقش بسیار مهمی در کنترل پیچیدگی مدل بازی می‌کند. به عنوان مثال، به ازای $\lambda = ۰$ برآوردگر تاوانیده به برآوردگر OLS تبدیل می‌شود و همه متغیرهای تبیینی وارد مدل می‌شوند. اگر $\lambda \to \infty$، جزء دوم تابع (۴.۲) به ∞ همگرا می‌شود، و در نتیجه تمام ضرایب صفر شده و هیچ متغیری وارد مدل نمی‌شود. برای انتخاب بهترین مقدار بین این دو حالت فرین، از معیارهای کلاسیک AIC، BIC و GCV استفاده می‌شود. به‌طور مشخص، برای λ داده‌شده، برآورد کمترین توان‌های دوم تاوانیده، $\widehat{\beta}_\lambda$، را به‌دست آورده و سپس مقدار معیار انتخاب کلاسیک را براساس $\widehat{\beta}_\lambda$ محاسبه می‌کنیم. این معیارهای کلاسیک به‌صورت

$$\begin{aligned}
AIC_\lambda &= \|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda\|^۲ + ۲df_\lambda\widehat{\sigma}^۲,\\
BIC_\lambda &= \|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda\|^۲ + \log(n)\widehat{\sigma}^۲,\\
GCV_\lambda &= \frac{\frac{۱}{n}\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_\lambda\|^۲}{(۱ - df_\lambda/n)^۲},
\end{aligned}$$

محاسبه می‌شوند، که $\widehat{\sigma}^۲ = \frac{RSS_p}{n-p}$ و RSS_p مجموع توان‌های دوم باقیمانده‌ها تحت مدل کامل است. منظور از مدل کامل مدلی است که همه متغیرهای تبیینی در مدل حضور دارند. در معیارهای فوق، درجه آزادی مدل برآورد شده df_λ معمولا با استفاده از تعداد ضرایب غیرصفر به‌صورت

$$df_\lambda = \sum_{k=۱}^{p} I(\widehat{\beta}_{k,\lambda} \neq ۰),$$

به‌دست می‌آید که در آن $I(.)$ تابع نشانگر است. راه دیگر محاسبه df_λ استفاده از ماتریس طرح به‌صورت

$$df_\lambda = \mathrm{tr}\left(\boldsymbol{X}_\lambda(\boldsymbol{X}_\lambda^\top \boldsymbol{X}_\lambda + n\boldsymbol{\Sigma}_\lambda)^{-1}\boldsymbol{X}_\lambda^\top\right),$$

است که $\boldsymbol{X}_\lambda$ ماتریس طرح مدل متناظر با λ داده‌شده و $\boldsymbol{\Sigma}_\lambda$ ماتریس بلوکی به‌صورت

$$\boldsymbol{\Sigma}_\lambda = \underset{\widehat{\beta}_{k,\lambda}\neq 0}{\mathrm{diag}}\left\{P'_\lambda(|\widehat{\beta}_{k,\lambda}|)/|\widehat{\beta}_{k,\lambda}|\right\}$$

بوده و $\widehat{\beta}_{k,\lambda}$، kامین مولفه از برآورد کمترین توان‌های دوم تاوانیده $\widehat{\boldsymbol{\beta}}_\lambda$ است. در پایان، برای یک مجموعه از نقاط داده شده $\lambda_1,\ldots,\lambda_M$، آن مقداری از λ را به عنوان مقدار بهینه انتخاب می‌کنیم که دارای کمترین مقدار معیار کلاسیک باشد.

یک روش دیگر برای انتخاب پارامتر تاوان استفاده از روش اعتبارسنجی متقابل k تایی[37] است. در این روش، برای یک λ داده شده، مجموعه مشاهدات را به‌طور تصادفی به k گروه با اندازه یکسان افراز می‌کنیم. سپس گروه اول را به عنوان داده‌های آزمون[38] و سایر گروه‌ها را به عنوان داده‌های آموزشی درنظر می‌گیریم. پس از برازش مدل با استفاده از داده‌های آموزشی، مقدار متغیر پاسخ به ازای داده‌های آزمون پیشگویی می‌شود. سپس خطای پیشگویی را به‌صورت $PE_1 = \sum_{i\in I_1}(y_i-\widehat{y}_i)^2$ محاسبه می‌کنیم. که I_1 اندیس مشاهدات گروه اول است. این روند k بار تکرار می‌شود و در هر تکرار یک گروه مختلف به عنوان داده‌های آزمون و بقیه گروه‌ها به عنوان داده‌های آموزشی درنظر گرفته می‌شود. در نتیجه k برآورد خطای آزمون $PE_1,\ldots,PE_k$ به‌دست می‌آید. سپس اعتبارسنجی متقابل با استفاده از رابطه

$$CV(\lambda) = \frac{1}{k}\sum_{i=1}^{k} PE_i$$

محاسبه می‌شود. آن λای انتخاب می‌شود که مقدار $CV(\lambda)$ را کمینه کند. در عمل معمولاً مقدار k برابر ۵ یا ۱۰ درنظر گرفته می‌شود.

۴.۳.۱.۲ الگوریتم محاسباتی روش مبتنی بر تابع تاوان

رگرسیون تاوانیده در بخش‌های قبل را به‌صورت

$$\ell(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{k=1}^{p} P_\lambda(|\beta_k|) \qquad (۹.۲)$$

درنظر بگیرید. با یک تابع تاوان محدب، مانند لاسو، تابع هدف (۹.۲) یک تابع محدب است و لذا برآوردهای رگرسیون تاوانیده را می‌توان با بهینه‌سازی محدب به‌دست آورد. اما اغلب توابع تاوانی که دارای سه ویژگی مطلوب تنکی، ناااریبی و پیوستگی می‌باشند، توابع غیرمحدب بوده و این موضوع باعث می‌شود که تابع هدف (۹.۲) محدب نباشد. می‌توان توابع تاوان غیرمحدب را با برخی توابع

[37] k-fold cross validation

[38] Test dataset

محدب تقریب زد و مسئله بهینه‌سازی غیرمحدب را با استفاده از الگوریتم‌های بهینه‌سازی محدب حل کرد. در این قسمت دو تقریب زیر را برای حل مشکل بیان‌شده، معرفی می‌کنیم.

تقریب درجه دوم موضعی: چون عبارت اول در رابطه (۹.۲) یک تابع محدب است، لذا کافی است تابع تاوان غیر محدب $P_\lambda(|\beta_k|)$ را با یک تابع محدب تقریب کنیم. در این راستا، فن و لی (۲۰۰۱) الگوریتم درجه دوم موضعی[39] (LQA) را برای تقریب توابع تاوان غیرمحدب ارائه دادند. فرض کنید $\boldsymbol{\beta}^{(0)} = (\beta_1^{(0)}, \ldots, \beta_p^{(0)})^\top$ مقدار اولیه‌ای برای $\boldsymbol{\beta}$ باشد. اگر $\beta_k^{(0)}$ نزدیک به صفر باشد، قرار می‌دهیم $\widehat{\beta}_k = 0$، در غیر این‌صورت با استفاده از تقریب موضعی داریم

$$[P_\lambda(|\beta_k|)]' = P'_\lambda(|\beta_k|)\frac{\beta_k}{|\beta_k|} \approx P'_\lambda(|\beta_k^{(0)}|)\frac{\beta_k}{|\beta_k^{(0)}|}, \qquad \beta_k \approx \beta_k^{(0)}. \tag{۱۰.۲}$$

با انتگرال‌گیری از طرفین نسبت به β_k از $\beta_k^{(0)}$ تا β_k داریم

$$P_\lambda(|\beta_k|) - P_\lambda(|\beta_k^{(0)}|) \approx \frac{1}{2}\frac{P'_\lambda(|\beta_k^{(0)}|)}{|\beta_k^{(0)}|}\{\beta_k^2 - \beta_k^{(0)2}\}$$

سپس با انتقال جمله $P_\lambda(|\beta_k^{(0)}|)$ به طرف دوم معادله، عبارت زیر

$$P_\lambda(|\beta_k|) \approx P_\lambda(|\beta_k^{(0)}|) + \frac{1}{2}\frac{P'_\lambda(|\beta_k^{(0)}|)}{|\beta_k^{(0)}|}\{\beta_k^2 - \beta_k^{(0)2}\}.$$

به‌دست می‌آید. بنابراین،

$$\begin{aligned}
\ell(\boldsymbol{\beta}) &= \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{k=1}^{p} P_\lambda(|\beta_k|) \\
&\approx \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{k=1}^{p}\left\{P_\lambda(|\beta_k^{(0)}|) + \frac{1}{2}\frac{P'_\lambda(|\beta_k^{(0)}|)}{|\beta_k^{(0)}|}\{\beta_k^2 - \beta_k^{(0)2}\}\right\} \\
&= \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{k=1}^{p}\frac{1}{2}\frac{P'_\lambda(|\beta_k^{(0)}|)}{|\beta_k^{(0)}|}\beta_k^2 + c,
\end{aligned}$$

که ثابت c تابعی از $\beta_k^{(0)}$ است. با استفاده از تقریب فوق و حذف جملات ثابت، تابع هدف (۹.۲) به یک تابع درجه دو تبدیل می‌شود. در نتیجه، برآورد کمترین توان‌های دوم تاوانیده را می‌توان به‌صورت

$$\boldsymbol{\beta}^{(1)} = \left\{\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{\Sigma}_\lambda(\boldsymbol{\beta}^{(0)})\right\}^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$

به‌دست آورد که در آن

$$\boldsymbol{\Sigma}_\lambda(\boldsymbol{\beta}^{(0)}) = \operatorname{diag}\left\{\frac{P'_\lambda(|\beta_1^{(0)}|)}{|\beta_1^{(0)}|}, \ldots, \frac{P'_\lambda(|\beta_p^{(0)}|)}{|\beta_p^{(0)}|}\right\}.$$

[39] Local quadratic approximation

ساختار فوق شبیه برآوردگر ریج است و با توجه به این واقعیت که رگرسیون ریج نمی‌تواند متغیرهای معنی‌دار را انتخاب کند، ضرایب رگرسیونی برآوردشده را که بسیار نزدیک به صفر می‌باشند، برابر صفر قرار می‌دهیم تا متغیرهای بی‌اهمیت از مدل حذف شوند.

یکی از مزایای تقریب LQA تعمیم‌پذیری آن به هر تابع زیان هموار دیگری غیر از تابع زیان کمترین توان‌های دوم است. با درنظر گرفتن تابع زیان $\ell(\boldsymbol{\beta})$ تابع هدف تاوانیده به‌صورت

$$\ell(\boldsymbol{\beta}) + \sum_{k=1}^{p} P_\lambda(|\beta_k|) \qquad (۱۱.۲)$$

تعریف می‌شود. در این حالت، ابتدا تابع زیان $\ell(\boldsymbol{\beta})$ را با یک تابع درجه دو به‌صورت

$$\ell(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}^{(0)}) + \nabla\ell^\top(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})^\top \nabla^2\ell(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}),$$

تقریب می‌کنیم که

$$\nabla\ell^\top(\boldsymbol{\beta}^{(0)}) = \frac{\partial\ell^\top(\boldsymbol{\beta}^{(0)})}{\partial\boldsymbol{\beta}}, \qquad \nabla^2\ell^\top(\boldsymbol{\beta}^{(0)}) = \frac{\partial^2\ell^\top(\boldsymbol{\beta}^{(0)})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}.$$

با حذف جمله ثابت، تابع هدف (۱۱.۲) را می‌توان به‌صورت

$$\nabla\ell^\top(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})^\top \nabla^2\ell(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}) + \frac{1}{2}\boldsymbol{\beta}^\top\boldsymbol{\Sigma}_\lambda(\boldsymbol{\beta}^{(0)})\boldsymbol{\beta} \qquad (۱۲.۲)$$

نوشت و سپس با استفاده از الگوریتم نیوتن–رافسون جواب بهینه تابع (۱۱.۲) را به‌دست آورد. به‌طور دقیق‌تر، با اجرای مراحل زیر، الگوریتم LQA را می‌توان برای تقریب تابع (۱۱.۲) به کار برد.

- مرحله اول: مقدار اولیه $\boldsymbol{\beta}^{(0)}$ را برای $\boldsymbol{\beta}$ انتخاب کنید. این مقدار اولیه می‌تواند برآورد OLS باشد.
- مرحله دوم: تابع هدف (۱۱.۲) را براساس (۱۲.۲) تقریب کنید.
- مرحله سوم: الگوریتم نیوتن–رافسون را برای محاسبه مقدار کمینه کننده تابع درجه دو بدست آمده در مرحله دوم به کار ببرید تا مقدار $\beta_k^{(0)}$ به‌روزرسانی شود. سپس اگر $|\beta_k^{(0)}| < \epsilon$، متغیر X_k را حذف کنید.
- مرحله چهارم: مراحل دوم و سوم را تا رسیدن به همگرایی تکرار کنید.

فن و لی (۲۰۰۱) نشان دادند که در صورت انتخاب یک مقدار مناسب برای $\boldsymbol{\beta}^{(0)}$، الگوریتم LQA پس از تعداد کمی تکرار به همگرایی می‌رسد. علاوه بر این، با یک مقدار اولیه مناسب، کارایی الگوریتم LQA با یک تکرار معادل با تکرار کامل الگوریتم تا رسیدن به همگرایی است. اما ایراد وارد بر این الگوریتم این است که اگر یک متغیر در هر تکرار حذف شود، در تکرار بعدی وارد مدل نمی‌شود و در نتیجه از مدل نهایی حذف می‌شود. بنابراین، این روش مشابه روش انتخاب پیشرو عمل می‌کند.

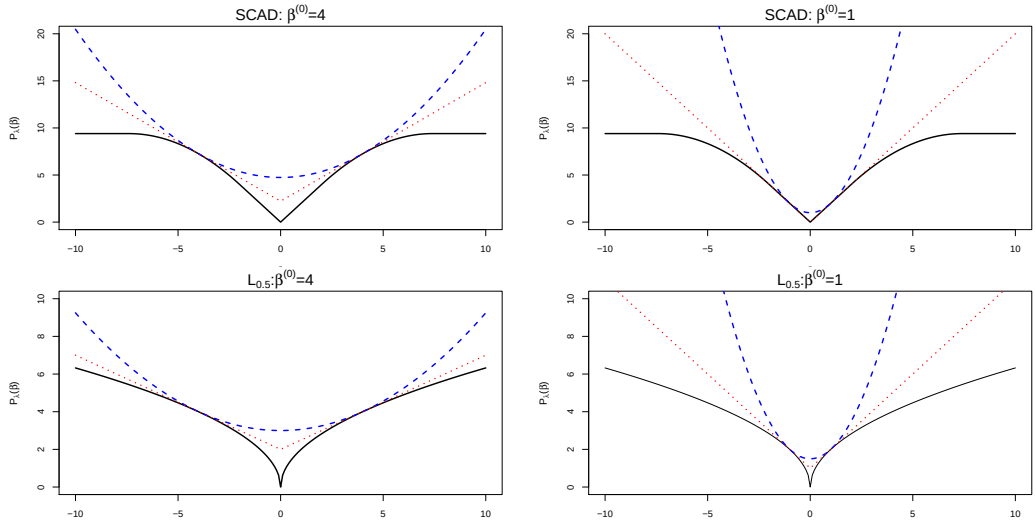
تقریب خطی موضعی: برای غلبه بر مشکل بیان‌شده در تقریب LQA، ژو و لی (۲۰۰۸) الگوریتم تقریب خطی موضعی[۴۰] (LLA) را برای تقریب توابع تاوان غیرمحدب پیشنهاد دادند. با استفاده از بسط تیلور $P_\lambda(.)$ در نقطه $\beta_k^{(0)}$ داریم

$$P_\lambda(|\beta_k|) \approx P_\lambda(|\beta_k^{(0)}|) + P'_\lambda(|\beta_k^{(0)}|)\{|\beta_k| - |\beta_k^{(0)}|\}; \qquad \beta_k \approx \beta_k^{(0)}.$$

بنابراین می‌توان تابع هدف (۹.۲) را به‌صورت

$$\begin{aligned}
\ell(\boldsymbol{\beta}) &= \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{k=1}^{p} P_\lambda(|\beta_k|) \\
&\approx \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{k=1}^{p} \left\{P_\lambda(|\beta_k^{(0)}|) + P'_\lambda(|\beta_k^{(0)}|)\{|\beta_k| - |\beta_k^{(0)}|\}\right\} \\
&= \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{k=1}^{p} \frac{1}{2} P'_\lambda(|\beta_k^{(0)}|)|\beta_k| + c,
\end{aligned}$$

تقریب نمود که ثابت c تابعی از $\boldsymbol{\beta}^{(0)}$ است. تقریب LLA تابع هدف تاوانیده غیرمحدب را به تابع هدف لاسوی وزنی تبدیل می‌کند. این تابع هدف تبدیل‌یافته را می‌توان با استفاده از برنامه‌ریزی درجه دو یا الگوریتم کمترین زاویه حل کرد. شکل ۵.۲ تقریب LQA و LLA را برای توابع تاوان $L_{0.5}$ و اسکد نشان می‌دهد. در این شکل برای هر دو تابع $\lambda = 2$ و دو مقدار اولیه مختلف درنظر گرفته شده است.



شکل ۵.۲: تقریب درجه دو و خطی برای توابع جریمه $L_{0.5}$ و $SCAD$. منحنی آبی: تقریب درجه دو؛ منحنی قرمز: تقریب خطی؛ منحنی مشکی: تابع جریمه.

تاکنون، علاوه بر دو تقریب فوق، الگوریتم‌های دیگری نیز برای حل مسئله توابع هدف تاوانیده معرفی شده است. الگوریتم پیش‌برنده[۴۱] (فو ، ۱۹۹۸)، الگوریتم مختصات نزولی[۴۲] برای لاسوی

[40] Local linear approximation
[41] Shooting
[42] Coordinate descent

گروهی (وو و لانگ ، ۲۰۰۸)، الگوریتم ماکزیمم‌سازی شرطی تکراری[۴۳] (ژانگ و لی ، ۲۰۱۱)، الگوریتم مختصات غیرنزولی برای توابع تاوان غیرمحدب (برهنی و هووانگ ، ۲۰۱۱) و الگوریتم مختصات نزولی گروهی (برهنی و هووانگ ، ۲۰۱۵) نمونه‌هایی از این الگوریتم‌ها می‌باشند.

۲.۲ برآوردگر ریج برای داده‌های طولی

همان‌گونه که در برخورد با داده‌های معمولی و بسیاری از کاربردهای رگرسیون کلاسیک با پدیده هم‌خطی روبرو هستیم، داده‌های طولی نیز از این امر مستثنی نیستند. استفاده از برآوردگر ریج در تحلیل داده‌های طولی پیشینه چندانی نداشته و تنها می‌توان کارهای ژانگ و هورواس (۲۰۰۶)، مالو و همکاران (۲۰۰۸)، الیوت و همکاران (۲۰۱۱) و رحمانی و همکاران (۲۰۱۸) را نام برد. لازم به ذکر است که در هیچ کدام از این موارد، مدل نیمه‌پارامتری با اثرات آمیخته بررسی نشده است و تحقیقات ما را می‌توان به عنوان تعمیمی از نتایج آن‌ها درنظر گرفت. بنابراین ضمن معرفی برآوردگر ریج، رفتار مجانبی برآوردگر معرفی شده ارائه شده است. در پایان نتایج عددی حاصل از مطالعات شبیه‌سازی و تحلیل داده‌های واقعی ارائه می‌گردد.

شرایطی را درنظر بگیرید که بین ستون‌های ماتریس طرح اثرات ثابت داده‌های طولی، هم‌خطی چندگانه وجود داشته باشد. در این حالت استفاده از برآوردگر رگرسیونی ریج برای $\boldsymbol{\beta}$ پیشنهاد می‌شود.

با فرض برآوردگر برای تابع ناپارامتری در رابطه (۸.۱)، صورت تاوانیده رابطه (۱۰.۱) عبارت است از

$$l^*(\boldsymbol{\beta}) = \widehat{l}(\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^\top \boldsymbol{\beta}.$$

در این حالت برآوردگر ریج هسته‌ای موزون به‌صورت

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{RK} &= \arg\min_{\boldsymbol{\beta}} \{(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1}(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^\top \boldsymbol{\beta}\} \\
&= (\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \tilde{\boldsymbol{X}} + \lambda \boldsymbol{I}_{n_i})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \tilde{\boldsymbol{y}} \\
&= \boldsymbol{R}(\lambda) \widehat{\boldsymbol{\beta}}_K,
\end{aligned}$$

حاصل می‌شود، که در آن $\boldsymbol{R}(\lambda) = \left((\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1} \tilde{\boldsymbol{X}})^{-1}\lambda + \boldsymbol{I}_{n_i}\right)^{-1}$. واضح است که $\boldsymbol{f}(\boldsymbol{t}_i)$ و $\boldsymbol{b}_i$ مشابه حالت‌های قبل به‌صورت

$$\begin{aligned}
\widehat{\boldsymbol{f}}(\boldsymbol{t}_i) &= \widehat{\boldsymbol{\mu}}_Y(\boldsymbol{t}_i) - \widehat{\boldsymbol{\mu}}_X^\top(\boldsymbol{t}_i)\widehat{\boldsymbol{\beta}}_{RK}, \\
\widehat{\boldsymbol{b}}_i &= E[\boldsymbol{b}_i|\boldsymbol{y}_i] = \boldsymbol{D}_i \boldsymbol{Z}_i \boldsymbol{V}_i^{-1}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i \widehat{\boldsymbol{\beta}}_{RK}) = \boldsymbol{D}_i \boldsymbol{Z}_i \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}_{RK} - \widehat{\boldsymbol{f}}(\boldsymbol{t}_i))
\end{aligned}$$

برآورد می‌شوند. به منظور انتخاب مقداری مناسب برای پارامتر تنظیم λ، فن و لی (۲۰۰۱) معیار

[43] Iterative conditional maximization

اعتبار سنجی متقابل تعمیم‌یافته[۴۴] (GCV) را به‌صورت

$$GCV(\lambda) = \frac{RSS(\lambda)/n}{(۱ - tr\big(d(\lambda)\big)/n)^۲}, \quad \widehat{\lambda} = \underset{\lambda \in \mathbb{R}^+}{\arg\min}\ GCV(\lambda)$$

معرفی کردند، که در آن

$$RSS(\lambda) = (\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}_R)^\top \boldsymbol{V}^{-۱}(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}_R),$$

$$d(\lambda) = \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-۱}\tilde{\boldsymbol{X}} + \lambda \boldsymbol{I})^{-۱}\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-۱} + \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^\top(\boldsymbol{I} - \tilde{\boldsymbol{X}}[(\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-۱}\tilde{\boldsymbol{X}} + \lambda \boldsymbol{I})^{-۱}\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-۱}]).$$

از آنجایی‌که برآوردگر ریج مضربی از برآوردگر هسته‌ای $(\widehat{\beta}_K)$ است، به راحتی می‌توان قضیه زیر را از قضیه ۱.۵.۱ نتیجه گرفت.

قضیه ۱.۲.۲. تحت مفروضات این بخش وقتی $n \longrightarrow \infty$، آنگاه برآوردگر ریج دارای توزیع مجانبی زیر است.

$$\sqrt{n}\Big\{\widehat{\boldsymbol{\beta}}_R - \boldsymbol{R}(\lambda)\big(\boldsymbol{\beta} + h^۲ b_K(\boldsymbol{\beta}, \boldsymbol{f})/۲\big)\Big\} \xrightarrow{d} \mathcal{N}_p\big(\mathbf{0}, \mathbf{R}(\lambda)\boldsymbol{V}_{\mathbf{K}}\mathbf{R}^\top(\lambda)\big).$$

که در آن مولفه‌های $b_K(\boldsymbol{\beta}, \boldsymbol{f})$ و $\boldsymbol{V}_K$ در قضیه ۱.۵.۱ آمده است.

از آنجایی‌که برآوردگر معرفی‌شده در تعامل با ماتریس کوواریانس پاسخ‌ها می‌باشد، به منظور بهبود کارایی برآوردگرها از یک الگوریتم تکراری مشابه الگوریتم ۱ بین برآورد ضرایب رگرسیونی و برآورد ماتریس کوواریانس استفاده می‌کنیم.

۱.۲.۲ مطالعات عددی

۱.۱.۲.۲ مطالعات شبیه‌سازی

در این قسمت یک مطالعه شبیه‌سازی جهت مقایسه برآوردگرهای هسته‌ای و هسته‌ای ریج در شرایطی که متغیرهای تبیینی دارای هم‌خطی باشند، انجام می‌گیرد. داده‌ها از مدل

$$\begin{aligned} y_i(t_{ij}) = \ & x_{۱,i}(t_{ij})\beta_۱ + x_{۲,i}(t_{ij})\beta_۲ + x_{۳,i}(t_{ij})\beta_۳ + x_{۴,i}(t_{ij})\beta_۴ + x_{۵,i}(t_{ij})\beta_۵ \\ & + f(t_{ij}) + b_i + \varepsilon_i(t_{ij}), \end{aligned}$$

شبیه‌سازی شده‌اند، که در آن $\boldsymbol{\beta}^T = (۰/۵, ۱, ۱/۵, ۲, ۰/۱, ۰/۲)$. سایر مولفه‌ها به‌صورت

$$\begin{aligned} f(t_{ij}) &= ۲\sin(۲\pi t_{ij}); \qquad t_{ij} \sim \mathcal{U}(۰, ۱), \\ b_i &\sim \mathcal{N}(۰, \sigma_b^۲); \qquad \sigma_b^۲ = ۰/۲۵, \\ \varepsilon_{ij} &\sim \mathcal{N}(۰, \sigma^۲(t)); \qquad \sigma^۲(t) = \exp(\tfrac{t}{۱۲}). \end{aligned}$$

[44] Generalized cross validation

جهت ایجاد شرایط هم‌خطی، متغیرهای تبیینی به کمک رابطه

$$x_{ik} = (۱ - \gamma^۲)^{\frac{۱}{۲}}\omega_{ik} + \gamma^۲\omega_{ip}, \quad i = ۱, \ldots, n = ۵۰, \qquad k = ۱, \ldots, p = ۵,$$

تولید می‌شوند، به‌طوری‌که ω_{ik}ها مستقل از هم بوده و دارای توزیع نرمال استاندارد هستند. از آنجایی‌که میزان همبستگی بین هر دو متغیر تبیینی $\gamma^۲$ می‌باشد به منظور رسیدن به درجات مختلف همبستگی مقادیر ۰/۷، ۰/۸، ۰/۹ و ۰/۹۹ را برای γ درنظر گرفتیم. به منظور بررسی تاثیر اعمال ساختار همبستگی نادرست بر عملکرد برآوردگرها، ابتدا $\widehat{\beta}_{RK}$ و $\widehat{\beta}_{K}$ را تحت ساختار همبستگی واقعی $AR(۱)$ برآورد کرده و به‌ترتیب با نماد $RK - C$ و $K - C$ نشان دادیم. سپس برآوردگرها را تحت ساختار همبستگی نادرست تبادل‌پذیر برآورد کرده و به‌ترتیب با نماد $RK - I$ و $K - I$ نشان دادیم.

جدول ۲.۲ نتایج مربوط به مقادیر MSE، اریبی و SD برآوردگرهای مذکور بر اساس ۵۰۰ تکرار از شبیه‌سازی بیان‌شده را گزارش می‌کند. از نتایج جداول می‌توان نتیجه گرفت که برآوردگر ریج هسته‌ای، MSE کمتری نسبت به برآوردگر هسته‌ای دارد حتی هنگامی که ساختار همبستگی به درستی تعیین نشود. برآوردگر ریج هسته‌ای از نظر معیارهای اریبی و SD عملکرد بهتری نسبت به برآوردگر هسته‌ای دارد. علاوه بر این، هنگامی که میزان هم‌خطی بین متغیرهای تبیینی افزایش می‌یابد، کارایی برآوردگر هسته‌ای در مقایسه با برآوردگر ریج هسته‌ای کاهش می‌یابد.

۲.۱.۲.۲ تحلیل داده‌های $CD۴$

در این بخش به مطالعه رفتار برآوردگرهای هسته‌ای و ریج هسته‌ای معرفی‌شده با استفاده از داده‌های $CD۴$ می‌پردازیم.

برآورد پارامترهای رگرسیونی به همراه SD آن‌ها در جدول ۳.۲ ارائه شده است. برای محاسبه مقادیر SD از روش اعتبارسنجی یک‌طرفه[۴۵] استفاده شد. این نتایج نشان می‌دهد که روش ریج همواره برآوردهایی با مقادیر SD کمتر را نتیجه می‌دهد.

۳.۲ انتخاب متغیر در مدل نیمه‌پارامتری با اثرات آمیخته برای داده‌های طولی نرمال با بعد بالا

در بسیاری از مطالعات طولی با تعداد فراوانی از متغیرها روبرو هستیم که عموما وارد کردن همه آن‌ها در تحلیل موجب انحراف مدل‌های معمول از الگوی واقعی داده‌ها می‌شود. بنابراین مهم است که روش‌های انتخاب متغیر و برآورد برای داده‌های طولی نیز توسعه یابد. پس از آن که فو (۲۰۰۳) و فن و لی (۲۰۰۴) به تعمیم روش مبتنی بر تابع تاوان، به‌ترتیب در مدل‌های خطی تعمیم‌یافته و مدل‌های نیمه‌پارامتری با داده‌های طولی پرداختند، مسئله انتخاب متغیر در زمینه داده‌های طولی با پیشرفت چمشگیری روبرو شد. در این میان می‌توان به کارهای کانتونی و همکاران (۲۰۰۵)، زیاک

[45] Leave-one-out cross-validation

جدول ۲.۲: نتایج شبیه‌سازی: مقادیر MSE و اریبی (SD) برآوردگرهای $\widehat{\beta}_{RK}$ و $\widehat{\beta}_K$ به ازای γهای مختلف.

روش‌ها	پارامترها	$\gamma = ۰/۷۰$	$\gamma = ۰/۸۰$	$\gamma = ۰/۹۰$	$\gamma = ۰/۹۹$
$RK-C$	$\beta_۱$	۰/۰۳۱۶(۰/۰۴۰۰)	۰/۰۳۷۶(۰/۰۴۷۵)	۰/۰۵۱۴(۰/۰۶۴۹)	۰/۱۴۰۱(۰/۱۷۴۸)
	$\beta_۲$	۰/۰۳۱۰(۰/۰۳۹۰)	۰/۰۳۶۸(۰/۰۴۶۴)	۰/۰۵۰۶(۰/۰۶۳۵)	۰/۱۴۴۶(۰/۱۷۰۶)
	$\beta_۳$	۰/۰۲۸۲(۰/۰۳۵۵)	۰/۰۳۳۳(۰/۰۴۲۲)	۰/۰۴۵۲(۰/۰۵۷۷)	۰/۱۳۵۹(۰/۱۵۷۲)
	$\beta_۴$	۰/۰۳۰۶(۰/۰۳۸۸)	۰/۰۳۶۴(۰/۰۴۶۱)	۰/۰۴۹۸(۰/۰۶۳۰)	۰/۱۳۹۷(۰/۱۷۴۵)
	$\beta_۵$	۰/۰۳۵۱(۰/۰۴۵۴)	۰/۰۴۵۴(۰/۰۵۹۲)	۰/۰۶۹۰(۰/۰۹۱۰)	۰/۲۲۰۸(۰/۲۶۴۴)
	MSE	۰/۰۰۷۹	۰/۰۱۱۸	۰/۰۲۳۸	۰/۲۱۰۳
$RK-I$	$\beta_۱$	۰/۰۳۳۷(۰/۰۴۱۴)	۰/۰۴۰۲(۰/۰۴۹۲)	۰/۰۵۵۳(۰/۰۶۷۷)	۰/۱۶۸۶(۰/۲۰۵۹)
	$\beta_۲$	۰/۰۳۳۰(۰/۰۴۰۸)	۰/۰۳۹۳(۰/۰۴۸۴)	۰/۰۵۴۰(۰/۰۶۶۵)	۰/۱۶۲۴(۰/۲۰۱۱)
	$\beta_۳$	۰/۰۲۹۵(۰/۰۳۷۵)	۰/۰۳۵۱(۰/۰۴۴۶)	۰/۰۴۸۲(۰/۰۶۱۳)	۰/۱۴۵۳(۰/۱۸۴۷)
	$\beta_۴$	۰/۰۳۶۲(۰/۰۴۵۲)	۰/۰۴۳۱(۰/۰۵۳۷)	۰/۰۵۹۳(۰/۰۷۳۹)	۰/۱۸۰۰(۰/۲۲۳۶)
	$\beta_۵$	۰/۰۳۴۲(۰/۰۴۲۷)	۰/۰۴۴۴(۰/۰۵۵۶)	۰/۰۶۸۷(۰/۰۸۶۵)	۰/۲۵۱۸(۰/۳۲۱۹)
	MSE	۰/۰۰۸۷	۰/۰۱۲۸	۰/۰۲۵۸	۰/۲۷۰۴
$K-C$	$\beta_۱$	۰/۰۳۱۵(۰/۰۴۰۰)	۰/۰۳۷۵(۰/۰۴۷۶)	۰/۰۵۱۷(۰/۰۶۵۶)	۰/۱۵۹۶(۰/۲۰۲۷)
	$\beta_۲$	۰/۰۳۱۳(۰/۰۳۹۰)	۰/۰۳۷۳(۰/۰۴۶۴)	۰/۰۵۱۳(۰/۰۶۳۹)	۰/۱۵۸۵(۰/۱۹۷۵)
	$\beta_۳$	۰/۰۲۹۶(۰/۰۳۵۸)	۰/۰۳۵۲(۰/۰۴۲۶)	۰/۰۴۸۵(۰/۰۵۸۶)	۰/۱۴۹۹(۰/۱۸۱۱)
	$\beta_۴$	۰/۰۳۰۸(۰/۰۳۸۹)	۰/۰۳۶۶(۰/۰۴۶۳)	۰/۰۵۰۴(۰/۰۶۳۸)	۰/۱۵۵۸(۰/۱۹۷۱)
	$\beta_۵$	۰/۰۳۵۶(۰/۰۴۵۶)	۰/۰۴۶۳(۰/۰۵۹۶)	۰/۰۷۲۰(۰/۰۹۳۰)	۰/۲۷۸۲(۰/۳۶۰۸)
	MSE	۰/۰۰۸۱	۰/۰۱۲۱	۰/۰۲۴۹	۰/۲۸۷۱
$K-I$	$\beta_۱$	۰/۰۳۳۷(۰/۰۴۱۴)	۰/۰۴۰۲(۰/۰۴۹۳)	۰/۰۵۵۳(۰/۰۶۷۹)	۰/۱۷۰۸(۰/۲۰۹۷)
	$\beta_۲$	۰/۰۳۳۲(۰/۰۴۰۷)	۰/۰۳۹۵(۰/۰۴۸۵)	۰/۰۵۴۳(۰/۰۶۶۷)	۰/۱۶۷۹(۰/۲۰۶۲)
	$\beta_۳$	۰/۰۲۹۶(۰/۰۳۷۶)	۰/۰۳۵۳(۰/۰۴۴۷)	۰/۰۴۸۵(۰/۰۶۱۵)	۰/۱۵۰۰(۰/۱۹۰۲)
	$\beta_۴$	۰/۰۳۶۲(۰/۰۴۵۲)	۰/۰۴۳۱(۰/۰۵۳۸)	۰/۰۵۹۴(۰/۰۷۴۰)	۰/۱۸۳۵(۰/۲۲۸۸)
	$\beta_۵$	۰/۰۳۴۵(۰/۰۴۲۸)	۰/۰۴۴۴(۰/۰۵۵۷)	۰/۰۶۸۸(۰/۰۸۶۸)	۰/۲۶۳۳(۰/۳۳۸۰)
	MSE	۰/۰۰۸۷	۰/۰۱۲۹	۰/۰۲۶۱	۰/۲۹۱۹

(۲۰۰۶)، وانگ و کوای (۲۰۰۹) و سوای و همکاران (۲۰۱۰) در مدل‌های خطی تعمیم‌یافته، باندل و همکاران (۲۰۱۰) در مدل اثرات آمیخته، نی و همکاران (۲۰۱۰) و ما و همکاران (۲۰۱۳) در مدل‌های نیمه‌پارامتری اشاره کرد. رخداد مسئله بعد بالا در داده‌های طولی نیز غیر قابل اجتناب است. با این وجود، مسئله انتخاب متغیر برای داده‌های طولی با بعد بالا به واسطه چالش‌های تحمیل‌شده از جمله وجود همبستگی درون گروهی پیشینه چندانی نداشته و تنها می توان کارهای سو و همکاران (۲۰۱۳) و وانگ و همکاران (۲۰۱۲) را نام برد. اخیرا کاظمی و همکاران (۱۳۹۷)، در مدل‌های خطی-جزئی با بعد بسیار بالا مسئله انتخاب متغیر را بررسی کرده است. در عین حال، تحقیقات در زمینه انتخاب متغیر داده‌های طولی با بعد بالا در مدل‌های رگرسیونی نیمه‌پارامتری با اثرات آمیخته تا حد زیادی ناشناخته مانده است. با توجه به اهمیت مسئله بیان‌شده در داده‌های طولی با بعد بالا، در این بخش، برآورد و انتخاب متغیر در مدل نیمه‌پارامتری با اثرات آمیخته را مبتنی

جدول ۳.۲: نتایج تحلیل داده‌های $CD4$: برآورد پارامترهای رگرسیونی (SD) تحت برآوردگرهای مختلف.

	روش			
	ریج اسپلاین	ریج هسته‌ای	اسپلاین	هسته‌ای
AGE	۰٫۰۸۱۱(۰٫۰۰۰۱)	۰٫۰۲۰۸(۰٫۰۰۲۱)	۰٫۰۸۵۱(۰٫۰۰۰۲)	۰٫۰۲۱۲(۰٫۰۰۲۱)
$SMOKE$	۰٫۷۸۲۳(۰٫۰۰۳۱)	۰٫۵۸۰۸(۰٫۰۰۸۰)	۰٫۷۲۹۰(۰٫۰۰۶۰)	۰٫۵۹۴۱(۰٫۰۰۸۲)
$DRUG$	۰٫۸۴۰۱(۰٫۰۱۴۲)	۰٫۴۵۸۲(۰٫۰۱۶۶)	۰٫۸۰۱۱(۰٫۰۱۴۳)	۰٫۵۳۵۰(۰٫۰۱۹۴)
$SEXP$	۰٫۰۷۹۰(۰٫۰۰۱۳)	۰٫۰۵۹۷(۰٫۰۰۲۴)	۰٫۰۷۱۱(۰٫۰۰۱۴)	۰٫۰۵۸۵(۰٫۰۰۲۴)
$CESD$	−۰٫۰۵۴۲(۰٫۰۰۰۳)	−۰٫۰۵۳۱(۰٫۰۰۰۹)	−۰٫۰۶۸۶(۰٫۰۰۰۳)	−۰٫۰۵۳۲(۰٫۰۰۰۹)

بر رهیافت توابع تاوان، درنظر گرفتیم. روش پیشنهادی به علت وجود توام اثرات تصادفی و مولفه ناپارامتری در مدل، نسبتا با کارهای انجام‌شده متفاوت است. از آنجایی‌که برآوردگر معرفی‌شده در تعامل با ماتریس کوواریانس پاسخ‌ها و اثرات تصادفی می‌باشند، به منظور بهبود کارآیی برآوردگر بیشینه درستنمایی تاوانیده (PML) از الگوریتم تکراری EM استفاده می‌کنیم. علاوه بر این، خواص مجانبی برآوردگرها در چارچوب داده‌های بعد بالا که در آن تعداد پارامترها متناسب با افزایش حجم نمونه افزایش می‌یابد، و نحوه انتخاب پارامترهای تابع تاوان مورد بررسی قرار می‌گیرد. به منظور بررسی عملکرد روش معرفی‌شده، مطالعه‌ای شبیه‌سازی و همچنین تحلیل یک مجموعه داده واقعی فراهم آمده است.

فرض کنید یک نمونه تصادفی n تایی از داده‌های طولی با بعد بالا در اختیار داشته باشیم. مدل رگرسیونی نیمه‌پارامتری با اثرات آمیخته برای داده‌های طولی که در بخش ۴.۳.۱ شرح داده شد را به یاد آورید. همچنین فرض کنید مولفه ناپارامتری مدل به روش اسپلاین‌های رگرسیونی شرح داده شده در ۳.۵.۱ تقریب زده شود. بدین ترتیب یک مدل رگرسیون خطی به‌صورت

$$\boldsymbol{y}_i = \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \qquad (۱۳.۲)$$

خواهیم داشت، به‌طوری‌که ماتریس طرح ادغام شده $\tilde{\boldsymbol{X}}_i$ بعدی از ماتریس اثرات ثابت و اثرات اسپلاین و $\boldsymbol{\theta}$ بردار پارامتر ادغام شده $(p_n + h_n) \times ۱$ بعدی می‌باشند. به‌طوری‌که $h_n = d + K_n + ۱$، d درجه تقریب اسپلاین و K_n تعداد گره‌ها در تقریب اسپلاین است. حال از طریق بهینه‌سازی تابع هدف مبتنی بر تابع تاوان مانند بیشینه درستنمایی تاوانیده PML به برآورد و انتخاب متغیر می‌پردازیم.

از آنجایی‌که در مبحث داده‌های با بعد بالا فرض می‌شود که تعداد پارامترها (p) متناسب با افزایش حجم نمونه افزایش می‌یابد، معمولاً p را با نماد p_n نمایش می‌دهند. متعاقباً نمادهای h، K و λ به صورت h_n، K_n و λ_n تغییر می‌کنند.

۱.۳.۲ الگوریتم EM برای بیشینه درستنمایی تاوانیده

۱.۱.۳.۲ برآوردگر بیشینه درستنمایی

در روش برآورد ML، اگر بردار مشاهدات غیرکامل باشد، یعنی برخی از داده‌ها به دلیل سانسور شدن یا مقادیر گمشده در دسترس نباشند، می‌توان از روش‌های تکرارشونده همچون روش نیوتن–رافسون و الگوریتم امیدریاضی بیشینه‌سازی (دمپستر و همکاران، ۱۹۷۷) یا به اختصار الگوریتم EM استفاده کرد. در هر تکرار از الگوریتم دو گام وجود دارد. گام E که امید ریاضی شرطی براساس توزیع شرطی لگاریتم تابع درستنمایی داده‌های کامل به شرط داده‌های مشاهده‌شده را محاسبه می‌کند و گام M که برآورد بیشینه درستنمایی پارامترهای مدل را پیدا می‌کند. ما نیز برای برآورد پارامترهای مدل (۱۳.۲) از روش مذکور استفاده می‌کنیم.

فرض کنید $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{D}, \boldsymbol{\Sigma},)$ مجموعه ای از تمام پارامترهای مدل باشد. اگر اثرات تصادفی $\boldsymbol{b}_i$ را به عنوان مقادیر گمشده و مجموعه $\{(\boldsymbol{y}_i, \boldsymbol{b}_i), i = ۱, \ldots, n\}$ را داده‌های کامل درنظر بگیریم، آنگاه لگاریتم تابع درستنمایی بدون درنظر گرفتن جمله ثابت به‌صورت

$$\begin{aligned} \ell_c(\boldsymbol{\Theta}) &= \sum_{i=۱}^{n} \ell_c^{[i]}(\boldsymbol{\Theta}) \qquad (۱۴.۲) \\ &= \sum_{i=۱}^{n} \frac{۱}{۲}\Big\{ \log|\boldsymbol{\Sigma}_i^{-۱}| + \log|\boldsymbol{D}_i^{-۱}| - [\boldsymbol{\varepsilon}_i^\top \boldsymbol{\Sigma}_i^{-۱}\boldsymbol{\varepsilon}_i + \boldsymbol{b}_i^\top \boldsymbol{D}_i^{-۱}\boldsymbol{b}_i]\Big\} \end{aligned}$$

است. برای ارزیابی امید شرطی (۱۴.۲) به شرط داده‌های مشاهده‌شده $\boldsymbol{y} = \{\boldsymbol{y}_i, i = ۱, \ldots, n\}$ و برآوردهای جاری $\widehat{\boldsymbol{\Theta}}^{(r)} = (\widehat{\boldsymbol{\theta}}^{(r)}, \widehat{\boldsymbol{D}}^{(r)}, \widehat{\boldsymbol{\Sigma}}^{(r)})$، ما از نتیجه زیر استفاده می‌کنیم.

نتیجه ۱.۳.۲.

$$\begin{aligned} \begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{bmatrix} &\sim \mathcal{N}_{n_i+q}\left(\begin{bmatrix} \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{V}_i & \boldsymbol{Z}_i\boldsymbol{D}_i \\ \boldsymbol{D}_i\boldsymbol{Z}_i^\top & \boldsymbol{D}_i \end{bmatrix} \right), \\ \boldsymbol{b}_i|\boldsymbol{y}_i &\sim \mathcal{N}_q\Big(\boldsymbol{D}_i\boldsymbol{Z}_i^\top\boldsymbol{V}_i^{-۱}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta}), (\boldsymbol{D}_i^{-۱} + \boldsymbol{Z}_i^\top\boldsymbol{\Sigma}_i^{-۱}\boldsymbol{Z}_i)^{-۱}\Big). \end{aligned}$$

مراحل اصلی الگوریتم EM به شرح زیر انجام می‌شود:

گام E: محاسبه تابع $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \mathrm{E}\left(\ell_c(\boldsymbol{\Theta})|\boldsymbol{y}, \widehat{\boldsymbol{\Theta}}^{(r)}\right)$ که مجموع $Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \mathrm{E}\left(\ell_c^{[i]}(\boldsymbol{\Theta})|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}\right)$ برای $i = ۱, \ldots, n$ است و $Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$ به‌صورت

$$Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \frac{۱}{۲}\Big\{ \log|\boldsymbol{\Sigma}_i^{-۱}| + \log|\boldsymbol{D}_i^{-۱}| - \mathrm{tr}(\boldsymbol{D}_i^{-۱}\widehat{\boldsymbol{B}}_i^{(r)}) - \mathrm{tr}(\boldsymbol{\Sigma}_i^{-۱}\widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\theta}))\Big\},$$

تعریف می‌شوند، به‌طوری‌که بر اساس نتیجه ۱.۳.۲ داریم:

$$\begin{aligned} \widehat{\boldsymbol{B}}_i^{(r)} &= \mathrm{E}(\boldsymbol{b}_i\boldsymbol{b}_i^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\boldsymbol{b}}_i^{(r)}\widehat{\boldsymbol{b}}_i^{(r)^\top} + \widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)}, \\ \widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\beta}) &= \mathrm{E}(\boldsymbol{e}_i\boldsymbol{e}_i^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) \\ &= (\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)})(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)})^\top + \boldsymbol{Z}_i\widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)}\boldsymbol{Z}_i^\top, \end{aligned}$$

و

$$\begin{aligned}\widehat{\boldsymbol{b}}_i^{(r)} &= E(\boldsymbol{b}_i|\boldsymbol{y}_i,\widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\boldsymbol{D}}_i^{(r)}\boldsymbol{Z}_i^\top \widehat{\boldsymbol{V}}_i^{(r)^{-1}}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\widehat{\boldsymbol{\theta}}^{(r)}),\\ \widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)} &= \mathrm{Cov}(\boldsymbol{b}_i|\boldsymbol{y}_i,\widehat{\boldsymbol{\Theta}}^{(r)}) = \big(\widehat{\boldsymbol{D}}_i^{(r)^{-1}} + \boldsymbol{Z}_i^\top\widehat{\boldsymbol{\Sigma}}_i^{(r)^{-1}}\boldsymbol{Z}_i\big)^{-1},\end{aligned}$$

و $\boldsymbol{e}_i = (\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta}^{(r)})$.

گام M: به‌روزکردن $\widehat{\boldsymbol{\Theta}}^{(r)}$ توسط $\widehat{\boldsymbol{\Theta}}^{(r+1)} = \max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$.

۲.۱.۳.۲ برآوردگر PML و الگوریتم EM

به منظور انتخاب متغیرهای مهم و برآورد ضرایب آن‌ها، به لگاریتم تابع درستنمایی (۱۴.۲)، جمله تاوان $\sum_{k=1}^{p} P_{\lambda_n}(|\beta_k|)$ را اضافه می‌کنیم، آنگاه لگاریتم تابع درستنمایی تاوانیده به صورت

$$\ell_c^{\mathrm{Pen}}(\boldsymbol{\Theta}) = \ell_c(\boldsymbol{\Theta}) - n\sum_{k=1}^{p} P_{\lambda_n}(|\beta_k|). \qquad (۱۵.۲)$$

حاصل می‌شود. به‌طور مشابه تابع Q تاوانیده در گام E را به صورت

$$Q_i^{\mathrm{Pen}}(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) - n\sum_{k=1}^{p_n} P_{\lambda_n}(|\beta_k|).$$

درنظر می‌گیریم. در ادامه فرض می‌کنیم تابع تاوان مورد استفاده در این‌جا و کل رساله، تابع تاوان اسکد، رابطه (۸.۲) است و برای حل مسئله بهینه‌سازی در گام M، تقریب درجه دو موضعی (LQA) را به کار می‌بریم. در مجموع گام M الگوریتم EM به صورت الگوریتم ۲ پیشنهاد می‌شود.

۳.۱.۳.۲ انتخاب پارامترهای تاوان

برای پیاده‌سازی روش پیشنهادشده، باید درباره چندین پارامتر تصمیم‌گیری مناسبی اتخاذ شود. این پارامترها شامل درجه چند جمله‌ای d، تعداد نقاط گره K_n و تعداد توابع پایه h_n در تقریب B-اسپلاین و پارامترهای a و λ_n در تابع تاوان اسکد هستند.

در تقریب اسپلاین، اگر $d = ۳$، آنگاه تابع مورد اشاره را اسپلاین مکعبی می‌نامند. اسپلاین مکعبی یکی از متداول‌ترین نوع اسپلاین‌ها است و به دلیل انعطاف‌پذیری بالایی که دارد، در بسیاری از علوم مورد استفاده قرار می‌گیرد، لذا در قسمت مطالعات عددی از این نوع اسپلاین استفاده می‌شود. این نکته که تعداد گره‌ها با افزایش حجم نمونه افزایش یابند اهمیت داشته و از سوی دیگر تعداد زیاد گره‌ها ممکن است واریانس برآوردگرها را افزایش دهد. بنابراین، تعداد گره‌ها باید به درستی انتخاب شود تا بین اریبی و واریانس برآوردگرها تعادل ایجاد گردد. برای سهولت در محاسبات معمولا گره‌هایی با فواصل برابر و تعداد گره‌ها به صورت $K_n \approx n^{1/(2r+1)}$ انتخاب می‌شود. این استراتژی در بسیاری از متون آماری از جمله هی و همکاران (۲۰۰۲) ، چن و چو (۲۰۰۷) و سینها و ستار (۲۰۱۵)

الگوریتم ۲ برآوردگر PML و الگوریتم EM

۱: **گام** M: به‌روز کردن $\widehat{\boldsymbol{\beta}}^{(r)}$، $\widehat{\boldsymbol{D}}^{(r)}$ و $\widehat{\boldsymbol{\Sigma}}^{(r)}$ از طریق بیشینه کردن $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$ که به‌صورت زیر حاصل می شوند:

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^{(r+1)} &= \Big(\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^{\top} \widehat{\boldsymbol{\Sigma}}_i^{(r)^{-1}} \tilde{\boldsymbol{X}}_i + n\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}^{(r)})\Big)^{-1} \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^{\top} \widehat{\boldsymbol{\Sigma}}_i^{(r)^{-1}} (\boldsymbol{y}_i - \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i^{(r)}),\\
\widehat{\boldsymbol{D}}^{(r+1)} &= n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{B}}_i^{(r)},\\
\widehat{\boldsymbol{\Sigma}}^{(r+1)} &= n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\beta}),
\end{aligned}$$

که در آن

$$\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}^{(r)}) = \text{diag}\Big\{\frac{q_{\lambda_n}(|\beta_1|)}{\epsilon + |\beta_1|}, \ldots, \frac{q_{\lambda_n}(|\beta_p|)}{\epsilon + |\beta_p|}, \mathbf{0}_{h_n}\Big\}, \qquad \epsilon = 10^{-6},$$

و $\mathbf{0}_{h_n}$ برداری h_n بعدی با عناصر صفر است.

۲: **گام همگرایی:** گام های E و M را تا زمان همگرایی تکرار می‌کنیم تا برآوردگر PML، $\widehat{\boldsymbol{\Theta}} = (\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{D}}, \widehat{\boldsymbol{\Sigma}})$ ، به‌دست آید.

اجرا شده است. واضح است که $r > 0$ زیرا $r = 0$ تعداد n گره را درنظر می‌گیرد که انتخاب درستی نمی‌باشد. معمولاً r را برابر ۱ درنظر می‌گیرند؛ به‌طور مثال برای نمونه‌هایی به حجم ۵۰، ۱۰۰ و ۲۰۰، تعداد گره‌ها به‌ترتیب برابر ۴، ۵ و ۶ خواهد بود. بنابراین اگر اسپلاین را مکعبی و تعداد گره‌ها را به‌صورت $K_n \approx n^{1/(2r+1)}$ درنظر بگیریم، تعداد توابع پایه به‌صورت $h_n \approx n^{1/(2r+1)} + 4$ است؛ که برای حجم نمونه‌های ۵۰، ۱۰۰ و ۲۰۰ به‌ترتیب ۸، ۹ و ۱۰ تابع پایه خواهیم داشت.

در بخش ۳.۳.۱.۲ به منظور انتخاب مقداری مناسب برای پارامتر تنظیم λ_n، تکنیک‌های متعدد پیشنهادی توسط افراد مختلف، معرفی شده‌اند. همچنین فن و لی (۲۰۰۱) از دیدگاه بیزی مقدار $a = 3.7$ را به عنوان یک مقدار مناسب برای مسائل مختلف پیشنهاد دادند. در اینجا از معیار اعتبار سنجی متقابل تعمیم‌یافته (GCV) استفاده می‌کنیم که به‌صورت

$$GCV(\lambda_n) = \frac{RSS(\lambda_n)/n}{(1 - tr\big(d(\lambda_n)\big)/n)^2}, \quad \widehat{\lambda}_n = \underset{\lambda_n \in \text{Re}^+}{\arg\min}\, GCV(\lambda_n),$$

است و در آن

$$\begin{aligned}
RSS(\lambda_n) &= (\boldsymbol{y} - \tilde{\boldsymbol{X}}^{\top}\widehat{\boldsymbol{\theta}} - \boldsymbol{Z}^{\top}\widehat{\boldsymbol{b}})^{\top} \widehat{\boldsymbol{V}}^{-1} (\boldsymbol{y} - \tilde{\boldsymbol{X}}^{\top}\widehat{\boldsymbol{\theta}} - \boldsymbol{Z}^{\top}\widehat{\boldsymbol{b}}),\\
d(\lambda_n) &= \boldsymbol{X}(\boldsymbol{X}^{\top}\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X} + \lambda_n \boldsymbol{I})^{-1}\boldsymbol{X}^{\top}\widehat{\boldsymbol{V}}^{-1}\\
&\quad + \boldsymbol{Z}\widehat{\boldsymbol{D}}\boldsymbol{Z}^{\top}(\boldsymbol{I} - \boldsymbol{X}[(\boldsymbol{X}^{\top}\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}^{\top}\widehat{\boldsymbol{V}}^{-1}]).
\end{aligned}$$

۲.۳.۲ خواص مجانبی

در این بخش به بیان توزیع مجانبی برآوردهای $\widehat{\beta}_n$ و $\widehat{f}$ می‌پردازیم. در اکثر متون آماری در حیطه انتخاب متغیر به منظور سادگی در محاسبات، پارامترهای درست $\boldsymbol{\beta}_{n\circ}$ به‌صورت $\boldsymbol{\beta}_{n\circ} = (\boldsymbol{\beta}_{n\circ 1}^\top, \boldsymbol{\beta}_{n\circ 2}^\top)^\top$ و ماتریس طرح مدل به‌صورت $\boldsymbol{X}_i = (\boldsymbol{X}_{i(1)}, \boldsymbol{X}_{i(2)})$ تفکیک می‌شوند. بدون کاستن از کلیت مسئله فرض می‌شود $\boldsymbol{\beta}_{n\circ 2} = \mathbf{0}$ و تمام اعضای $\boldsymbol{\beta}_{n\circ 1}$ که برداری s_n^* بعدی است غیرصفر باشد. در اینجا، ضرایب رگرسیونی درست به‌صورت $\boldsymbol{\theta}_{n\circ} = (\boldsymbol{\beta}_{n\circ 1}^\top, \boldsymbol{\beta}_{n\circ 2}^\top, \boldsymbol{\alpha}_{n\circ}^\top)^\top$ هستند که در آن $\boldsymbol{\alpha}_{n\circ}$ برداری h_n بعدی و وابسته به تابع ناپارامتری $f_\circ$ است. به منظور جداسازی فضای پارامتر به فضای صفر و غیرصفر، ما $\boldsymbol{\theta}_{n\circ}$ را به‌صورت $(\boldsymbol{\theta}_{n\circ 1}^\top, \boldsymbol{\theta}_{n\circ 2}^\top)^\top$ بازتفکیک می‌کنیم، به‌طوری‌که $\boldsymbol{\theta}_{n\circ 1} = (\boldsymbol{\beta}_{n\circ 1}^\top, \boldsymbol{\alpha}_{n\circ}^\top)^\top$ ، $\boldsymbol{\theta}_{n\circ 2} = \boldsymbol{\beta}_{n\circ 2} = \mathbf{0}$ یک بردار $(s_n = s_n^* + h_n)$ بعدی است که تمام اعضای آن غیرصفراند. متعاقباً، مقادیر برآورد و ماتریس طرح به‌ترتیب به‌صورت $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\theta}}_{n1}^\top, \widehat{\boldsymbol{\theta}}_{n2}^\top)^\top$ و $\tilde{\boldsymbol{X}}_i =$ $\big(\tilde{\boldsymbol{X}}_{i(1)}^\top, \tilde{\boldsymbol{X}}_{i(2)}^\top\big)^\top$ بازتفکیک می‌شوند، که در آن $\widehat{\boldsymbol{\theta}}_{n1} = (\widehat{\boldsymbol{\beta}}_{n1}^\top, \widehat{\boldsymbol{\alpha}}_n^\top)^\top$، $\tilde{\boldsymbol{X}}_{i(1)} = \big(\boldsymbol{X}_{i(1)}^\top, \boldsymbol{B}_i(\boldsymbol{t}_i)^\top\big)^\top$، $\widehat{\boldsymbol{\theta}}_{n2} = \widehat{\boldsymbol{\beta}}_{n2}$ و $\tilde{\boldsymbol{X}}_{i(2)} = \boldsymbol{X}_{i(2)}$. فرض کنید $a_n = \max_{1\leqslant k\leqslant p_n}\{P'_{\lambda_n}(|\beta_{n\circ,k}|),\ \beta_{n\circ,k}\neq 0\}$ و $b_n = \max_{1\leqslant k\leqslant p_n}\{P''_{\lambda_n}(|\beta_{n\circ,k}|),\ \beta_{n\circ,k}\neq 0\}$ به‌طوری‌که، $\beta_{n\circ,k}$ مولفه kام $\boldsymbol{\beta}_{n\circ}$ می‌باشد.

درنظر داشته باشید که اگر گام M الگوریتم EM دارای چندین جواب باشد، تنها دنباله‌ای از برآوردگر سازگار $\widehat{\boldsymbol{\theta}}_n$ را درنظر می‌گیریم. دنباله $\widehat{\boldsymbol{\theta}}_n$ را سازگار گوییم اگر $n\to\infty$ آنگاه $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_{n\circ}\to\mathbf{0}$ و $\sup_t|\boldsymbol{B}^\top(t)\widehat{\boldsymbol{\alpha}}_n - f_\circ(t)| \xrightarrow{p} \mathbf{0}$.

علاوه بر این، فرض کنید به ازای هر i، $n_i = m < \infty$ باشد؛ مجموعه سه تایی $(\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{t}_i)$، $i = 1,\ldots,n$ متغیرهایی مستقل و هم توزیع باشند و مشتق kام تابعی مانند $d(.)$ را به‌صورت $d^{(k)}(.)$ تعریف می‌کنیم. برای مطالعه رفتار حدی برآوردگرها، شرایط نظم زیر لازم می‌باشد:

(۱.A) به ازای هر $k\geqslant 2$، مشتق kام $\boldsymbol{f}_\circ(\boldsymbol{t}_i)$ کراندار است.

(۲.A) ماتریس کوواریانس برآورد شده $\widehat{\boldsymbol{\Sigma}}_i$ در شرط $\|\widehat{\boldsymbol{\Sigma}}_i - \boldsymbol{\Sigma}_i\| = O_p(n^{-1/2})$ صدق می کند.

(۳.A)

$$b_1 \leqslant \lambda_{\min}(n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{X}}_i^\top\boldsymbol{\Sigma}_i^{-1}\tilde{\boldsymbol{X}}_i) \leqslant \lambda_{\max}(n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{X}}_i^\top\boldsymbol{\Sigma}_i^{-1}\tilde{\boldsymbol{X}}_i) \leqslant b_2,$$

به‌طوری‌که b_1 و b_2 مقادیری ثابت و مثبت و $\lambda_{\min}$ و $\lambda_{\max}$ به‌ترتیب کوچکترین و بزرگترین مقدار ویژه ماتریس هستند.

(۴.A) وقتی $n\longrightarrow\infty$، آنگاه $\min_{1\leqslant k\leqslant s_n}|\theta_{n\circ k}|/\lambda_n\to\infty$، $\lambda_n\to 0$، $s_n^3 n^{-1} = o(1)$، $s_n^2(\log n)^4 =$ $o(n\lambda_n^2)^2$، $\log(p_n) = o(n\lambda_n^2/(\log n)$ و $p_n s_n^4(\log n)^6 = o(n^2\lambda_n^2)$، $p_n s_n^3(\log n)^8 = o(n^2\lambda_n^4)$.

(۵.A) تابع تاوان در شرایط $\liminf_{n\to\infty}\liminf_{\theta\to 0_+} p'_{\lambda_n}(\theta)/\lambda_n > 0$، $a_n = O(n^{-\frac{1}{2}})$، $b_n\to\infty$، صدق می‌کند و همچنین مقادیر ثابت C و D وجود دارند به‌طوری‌که به ازای هر $\theta_1,\theta_2 > \lambda_n C$ خواهیم داشت $|P''_{\lambda_n}(\theta_1) - P''_{\lambda_n}(\theta_2)| \leqslant D|\theta_1-\theta_2|$.

قضیه ۱.۳.۲. با درنظر گرفتن شرایط بیان‌شده، وقتی $n \longrightarrow \infty$ و $n^{-1}p_n^2 = o(1)$، آنگاه $\widehat{\boldsymbol{\theta}}_n$ یک جواب الگوریتم EM است، به‌طوری‌که

$$(i) \qquad \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{n0}\| = O_p\big(\sqrt{p_n}(n^{-1/2} + a_n)\big),$$
$$(ii) \qquad \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\big(\widehat{g}(t_{ij}) - g_0(t_{ij})\big)^2 = O_p(n^{-2m/(2m+1)}).$$

قضیه ۲.۳.۲. با درنظر گرفتن شرایط بیان‌شده، وقتی $n \longrightarrow \infty$، اگر $K_n = O_p(n^{1/(2r+1)})$ و $n^{-1}p_n^3 = o(1)$، آن‌گاه

$$(i) \qquad \mathbb{P}(\widehat{\boldsymbol{\beta}}_{n2} = \mathbf{0}) \to \mathbf{1},$$
$$(ii) \qquad \widehat{\boldsymbol{\Theta}} \xrightarrow{p} \boldsymbol{\Theta}, \quad \sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{\boldsymbol{\eta\eta}}^{-1}),$$
$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n01}) \xrightarrow{d} \mathcal{N}_{s_n}(\mathbf{0}, \boldsymbol{I}^{-1}_{\boldsymbol{\beta}_{\mathbf{n01}}\boldsymbol{\beta}_{\mathbf{n01}}}),$$

که در آن، $\boldsymbol{\eta} = (vech(D)^\top, vech(\boldsymbol{\Sigma})^\top)^\top$، $n^{-1}\boldsymbol{J}_{\boldsymbol{\eta\eta}} \to \boldsymbol{I}_{\boldsymbol{\eta\eta}}$، $n^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_{n01}\boldsymbol{\beta}_{n01}} \to \boldsymbol{I}_{\boldsymbol{\beta}_{n01}\boldsymbol{\beta}_{n01}}$، $\boldsymbol{J}_{\boldsymbol{\eta\eta}}$ و $\boldsymbol{J}_{\boldsymbol{\beta}_{n01}\boldsymbol{\beta}_{n01}}$ ماتریس‌های اطلاع فیشر هستند.

اثبات قضایای ۱.۳.۲ و ۲.۳.۲ و همچنین ماتریس‌های اطلاع فیشر در قسمت ضمیمه آمده است.

۳.۳.۲ مطالعات عددی

در این بخش، فرایند برازش مدل، برآورد پارامترهای مدل و انتخاب متغیر را به‌صورت عددی بررسی می‌کنیم. ابتدا مجموعه‌ای از داده‌های شبیه‌سازی‌شده و یک مورد مطالعاتی با داده‌های واقعی را توسط روش پیشنهادی مقاله یعنی برآوردگر بیشینه درستنمایی تاوانیده در مدل نیمه‌پارامتری با اثرات آمیخته ($P-SMM$) مورد تحلیل قرار می‌دهیم. سپس به منظور بررسی عملکرد روش $P-SMM$ در مسئله برآورد و انتخاب متغیر همزمان، نتایج حاصل از تحلیل داده‌ها بدون رهیافت تابع تاوان یعنی روش برآوردگر بیشینه درستنمایی در مدل خطی-جزئی با اثرات آمیخته (SMM) نیز گزارش می‌شود. درنظر داشته باشید که اگر داده‌ها ذاتاً دارای جزء ناپارامتری باشند، مدل نیمه‌پارامتری بهتر از مدل خطی عمل می‌کند. برای نشان دادن این موضوع روش $P-SMM$ را با روش برآوردگر بیشینه درستنمایی تاوانیده در مدل خطی با اثرات آمیخته ($P-LMM$) مقایسه می‌کنیم.

۱.۳.۳.۲ مطالعات شبیه‌سازی

در این قسمت متغیرهای پاسخ را از مدل

$$y_{ij} = \sum_{k=1}^{p_n} x_{k,ij}\beta_k + f(t_{ij}) + b_i + \varepsilon_{ij},$$

شبیه‌سازی می‌کنیم، که در آن مولفه‌های مدل به‌صورت

$$i = ۱,\ldots,n, \quad j = ۱,\ldots,n_i, \quad n = ۵۰, ۱۰۰, ۲۰۰, \quad n_i = ۵$$
$$\boldsymbol{\beta}^\top = (۲, ۳, ۱/۵, ۲, ۰, \ldots, ۰), \quad b_i \sim \mathcal{N}(۰, \sigma_b^۲); \qquad \sigma_b^۲ = ۰/۲۵,$$
$$\boldsymbol{X}_{ij}^\top = (x_{۱,ij}, \ldots, x_{p_n,ij}), \quad x_{k,ij} \sim \mathcal{U}(-۱, ۱)$$
$$f(t_{ij}) = ۲\sin(۲\pi t_{ij}), \quad t_{ij} \sim \mathcal{U}(۰, ۱).$$

تعیین شده‌اند. به منظور انتخاب تعداد مولفه‌های پارامتری مدل، نویسندگان پیشنهادهای متفاوتی از جمله $p_n = [\frac{n}{۲}]$، $p_n = [۴/۵n^{۱/۴}]$ و $p_n = [\frac{n}{b\log(n)}]$ داشته‌اند که در آن‌ها $b > ۱$ و $[a]$ جزء صحیح مقدار a است. موارد بیان‌شده برای حالت $p_n < n$ است. برای زمانی که $p_n > n$، نرخ رشد چند جمله‌ای ($p_n = O(n^b)$) و هنگامی که $p_n >> n$ (p_n بسیار بزرگتر از n)، نرخ رشد نمایی ($p_n = \exp\{O(n^b)\}$) پیشنهاد شده است که در آن $۰ < b < ۱$. در حالت p_n بسیار بزرگتر از n اگر چه احتمال انتخاب متغیرهای موثر در مدل افزایش می‌یابد در عین حال متغیرهای بی‌تاثیر بیش‌تری در مدل وارد خواهند شد که این موضوع بر عملکرد روش‌های برآورد و انتخاب متغیر تاثیر منفی خواهد گذاشت. تحلیل داده‌های با بعد بسیار بالا با چالش‌هایی روبرو است که کاربرد مستقیم روش‌های انتخاب متغیر اشاره‌شده را با مشکل مواجه می‌کند. یکی از راه حل‌های رفع این مشکل استفاده از روش‌های غربالگری است. از آنجایی‌که در این رساله به داده‌های با بعد بسیار بالا پرداخته نمی‌شود از توضیحات بیش‌تر پیرامون این مبحث صرف‌نظر می‌کنیم و علاقه‌مندان جهت آشنایی بیش‌تر می‌توانند به کاظمی (۱۳۹۷) مراجعه کنند. به منظور همپوشانی تمامی نرخ‌های پیشنهادی، موارد $(۵۰, ۱۱), (۱۰۰, ۱۴), (۲۰۰, ۱۶)$ برای حالت $p_n < n$ و موارد $(۲۰۰, ۲۰۰۰), (۱۰۰, ۵۰۰), (۳۰, ۱۰۰)$ برای حالت $p_n > n$ درنظر گرفته می‌شود.

به منظور مقایسه عملکرد روش پیشنهادی، هر مجموعه از داده‌های شبیه‌سازی شده را توسط سه روش شامل برآوردگر بیشینه درستنمایی تاوانیده در مدل خطی-جزئی با اثرات آمیخته ($P-SMM$)، برآوردگر بیشینه درستنمایی در مدل خطی-جزئی با اثرات آمیخته (SMM) و برآوردگر بیشینه درستنمایی تاوانیده در مدل خطی با اثرات آمیخته ($P-LMM$)، مورد تحلیل قرار دادیم. دقت برآورد هر یک از روش‌ها توسط معیار میانگین کمترین توان‌های دوم خطا (MSE) براساس ۱۰۰ تکرار شبیه‌سازی، مورد سنجش قرار گرفت. ارزیابی عملکرد روش‌ها در مسئله انتخاب متغیر توسط معیارهای (C, I) سنجیده می‌شود به‌طوری‌که C میانگین ضرایب صفر که به درستی صفر تخمین زده شده‌اند و I میانگین ضرایب غیرصفر است که به اشتباه صفر شده‌اند. برای ارائه توصیف جامع‌تر از عملکرد مسئله انتخاب متغیر روش‌ها از معیارهای دیگری نیز استفاده کردیم. UF بیانگر نسبت تکرارهایی است که ضرایب تاثیرگذار را به اشتباه صفر درنظر گرفته است، CF بیانگر نسبت تکرارهایی است که مدل را به درستی تعیین کرده‌اند و نسبت تکرارهایی که علاوه بر متغیرهای مهم، متغیرهای بی تاثیر را نیز در مدل وارد کرده‌اند با OF نشان دادیم.

جدول ۴.۲، خلاصه‌ای از نتایج مربوط به عملکرد روش‌های $P-PSM$، $P-LMM$ و SMM در برآورد و انتخاب مدل برای مقادیر مختلف (n, p_n) است. از نظر دقت برآورد روش SMM،

جدول ۴.۲: نتایج شبیه‌سازی: مقایسه نتایج انتخاب متغیر تحت روش‌های $P-SMM$، SMM و $P-LMM$ برای حالت‌های $p_n < n$ و $p_n > n$.

method	حالت $p_n < n$						حالت $p_n > n$					
	$(n,p)=(۵۰,۱۱)$						$(n,p)=(۵۰,۱۰۰)$					
	MSE	$C(۸)$	$I(۰)$	UF	CF	OF	MSE	$C(۹۷)$	$I(۰)$	UF	CF	OF
SMM	۰/۳۹۶	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰	۸/۰۵۰	۰/۱۰	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰
$P-LMM$	۰/۱۸۳	۶/۲۸	۰/۰۰	۰/۰۰	۰/۵۰	۰/۵۰	۰/۴۵۵	۹۴/۱۹	۰/۰۰	۰/۰۰	۰/۲۳	۰/۷۷
$P-SMM$	۰/۱۹۱	۶/۴۸	۰/۰۰	۰/۰۰	۰/۵۴	۰/۴۶	۰/۳۸۹	۹۴/۶۵	۰/۰۰	۰/۰۰	۰/۲۷	۰/۷۳
	$(n,p)=(۱۰۰,۱۴)$						$(n,p)=(۱۰۰,۵۰۰)$					
	MSE	$C(۱۱)$	$I(۰)$	UF	CF	OF	MSE	$C(۴۹۷)$	$I(۰)$	UF	CF	OF
SMM	۰/۲۶۵	۰/۰۲	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰	۱۴۹۹/۱۳۶	۴۷/۰۲	۰/۰۳	۰/۰۳	۰/۰۰	۱/۰۰
$P-LMM$	۰/۰۹۳	۹/۴۷	۰/۰۰	۰/۰۰	۰/۶۲	۰/۳۸	۰/۰۶۲	۴۹۵/۷۲۰	۰/۰۰	۰/۰۰	۰/۳۰	۰/۷۰
$P-SMM$	۰/۰۹۶	۹/۵۶	۰/۰۰	۰/۰۰	۰/۶۹	۰/۳۱	۰/۰۳۸	۴۹۶/۲۵۰	۰/۰۰	۰/۰۰	۰/۵۴	۰/۴۶
	$(n,p)=(۱۵۰,۱۵)$						$(n,p)=(۲۰۰,۲۰۰۰)$					
	MSE	$C(۱۲)$	$I(۰)$	UF	CF	OF	MSE	$C(۱۹۹۷)$	$I(۰)$	UF	CF	OF
SMM	۰/۱۳۹	۰/۰۹	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰	۱۲۵/۴۰۶	۱۱۳۷/۶۲	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰
$P-LMM$	۰/۰۴۳	۱۱/۹۳	۰/۰۰	۰/۰۰	۰/۹۳	۰/۰۷	۰/۰۱۸	۱۹۹۶/۸۹	۰/۰۰	۰/۰۰	۰/۸۹	۰/۱۱
$P-SMM$	۰/۰۴۳	۱۱/۹۶	۰/۰۰	۰/۰۰	۰/۹۶	۰/۰۴	۰/۰۱۸	۱۹۹۶/۹۳	۰/۰۰	۰/۰۰	۰/۹۲	۰/۰۸

جدول ۵.۲: نتایج شبیه‌سازی: کارآیی روش $P-SMM$ برای حالت‌های $p_n < n$ و $p_n > n$.

حالت $p_n < n$						حالت $p_n > n$					
(n,p_n)		$\beta_۱$	$\beta_۲$	$\beta_۳$	$\beta_۴$	(n,p_n)		$\beta_۱$	$\beta_۲$	$\beta_۳$	$\beta_۴$
$(۵۰,۱۱)$	$Bias$	۰/۰۳۹	۰/۰۰۸	۰/۰۰۳	۰/۰۰۵	$(۵۰,۱۰۰)$	$Bias$	۰/۰۸۰	۰/۱۰۸	۰/۰۴۶	۰/۰۵۸
	$SD۱$	۰/۱۰۵	۰/۱۰۷	۰/۱۰۶	۰/۱۰۳		$SD۱$	۰/۰۹۳	۰/۰۸۷	۰/۰۸۰	۰/۰۸۱
	$SD۲$	۰/۰۷۷	۰/۰۸۲	۰/۰۸۰	۰/۰۸۰		$SD۲$	۰/۰۸۱	۰/۰۷۹	۰/۰۷۷	۰/۰۷۷
	CP	۰/۹۵	۰/۹۷	۰/۹۵	۰/۹۱		CP	۰/۹۶	۰/۹۶	۰/۹۹	۰/۹۳
$(۱۰۰,۱۴)$	$Bias$	۰/۰۱۳	۰/۰۴۹	۰/۰۱۱	۰/۰۱۷	$(۱۰۰,۵۰۰)$	$Bias$	۰/۰۴۱	۰/۰۳۶	۰/۰۱۹	۰/۰۲۹
	$SD۱$	۰/۰۷۵	۰/۰۷۶	۰/۰۷۳	۰/۰۷۷		$SD۱$	۰/۰۷۸	۰/۰۷۹	۰/۰۷۷	۰/۰۷۸
	$SD۲$	۰/۰۵۵	۰/۰۵۵	۰/۰۵۵	۰/۰۵۵		$SD۲$	۰/۰۵۵	۰/۰۵۴	۰/۰۵۵	۰/۰۵۴
	CP	۰/۹۱	۰/۹۴	۰/۹۵	۰/۹۴		CP	۰/۹۳	۰/۹۳	۰/۹۵	۰/۹۴
$(۲۰۰,۱۶)$	$Bias$	۰/۰۳۷	۰/۰۰۳	۰/۰۱۳	۰/۰۱۴	$(۲۰۰,۲۰۰۰)$	$Bias$	۰/۰۱۶	۰/۰۴۲	۰/۰۱۳	۰/۰۰۵
	$SD۱$	۰/۰۵۷	۰/۰۵۷	۰/۰۵۸	۰/۰۵۹		$SD۱$	۰/۰۷۶	۰/۰۷۷	۰/۰۷۵	۰/۰۷۹
	$SD۲$	۰/۰۳۹	۰/۰۳۹	۰/۰۴۰	۰/۰۴۰		$SD۲$	۰/۰۵۵	۰/۰۵۵	۰/۰۵۵	۰/۰۵۶
	CP	۰/۹۶	۰/۹۵	۰/۹۵	۰/۹۵		CP	۰/۹۳	۰/۹۶	۰/۹۲	۰/۹۴

MSE بیش‌تری نسبت به دو روش دیگر دارد، در حالی‌که دو روش $P-LMM$ و $P-SMM$ بسیار نزدیک به هم عمل می‌کنند. با این وجود MSE روش پیشنهادی همواره کمتر از $P-LMM$ است. از نظر انتخاب مدل مشاهده می‌کنیم که SMM انتخاب مدل انجام نمی‌دهد اما روش‌های $P-LMM$ و $P-SMM$ تمامی متغیرهای مهم را انتخاب کرده است، زیرا معیار I آن‌ها صفر است. همچنین روش پیشنهادی به دلیل داشتن مقادیر بالا در نرخ‌های C و CF و مقادیر کمتر در OF، نسبت به رقیب خود یعنی $P-LMM$ عملکرد بهتری داشته است. با توجه به معیارهای CF و OF، هنگامی که $p_n > n$ نسبت به حالت $p_n < n$ روش‌ها در انتخاب مدل عملکرد ضعیفی دارند و ضرایب صفر تمایل بیش‌تری دارند که در مدل گنجانده شوند، اما با افزایش حجم نمونه هر دو

معیار بهبود می‌یابند. برای بررسی بیش‌تر عملکرد روش پیشنهادی، نتایج مربوط به اریبی ($Bias$)، انحراف استاندارد مجانبی(SD۱)، انحراف استاندارد نمونه‌ای (SD۲) و احتمال پوشش نمونه‌ای[۴۶] (CP) برای فاصله اطمینان ۹۵٪ در جدول ۵.۲ گزارش شده است. نتایج نشان می‌دهد که خطاهای استاندارد نمونه‌ای و مجانبی نزدیک به هم برآورد شده‌اند و احتمال پوشش در اغلب موارد نزدیک به ۹۵٪ است که این‌ها بیانگر عملکرد خوب روش است. به عبارت دیگر برآوردگرها رفتار مجانبی خوبی را نشان داده‌اند که این امر با نتایج قضایای حدی بیان‌شده مطابقت دارد.

۲.۳.۳.۲ تحلیل داده‌های CD۴

برای بهره‌مندی از انعطاف بالای مدل‌های نیمه‌پارامتری ما متغیر $YEAR$ را به عنوان جزء ناپارامتری و سایر متغیرها را به‌صورت خطی وارد مدل کردیم. از آنجایی‌که اثرات متقابل نیز ممکن است در مدل اهمیت داشته باشند، ما تمامی اثرات متقابل متغیرها را نیز در مدل وارد کرده و از روش پیشنهادی برای انتخاب متغیرهای مهم استفاده کردیم. به منظور مقایسه عملکرد روش $P-SMM$ با دو روش دیگر یعنی $P-LMM$ و SMM از معیار انحراف استاندارد مجانبی SD استفاده کردیم. همچنین برای شناسایی بهترین مدل پشتیبانی‌شده توسط داده‌ها، معیار اطلاع آکائیک (AIC) و معیار اطلاع بیزی (BIC) را محاسبه کردیم. این معیارها در قسمت ضمیمه معرفی شده‌اند. نتایج حاصل در جدول ۶.۲ ارائه شده است. این نتایج نشان می‌دهد که برآوردگر $P-SMM$ در مقایسه با دو روش دیگر دارای مقادیر SD کمتری می‌باشند. همچنین معیار های AIC و BIC نیز عملکرد بهتر روش پیشنهادی را تایید می‌کنند. تحت مدل $P-SMM$، متغیرهای $SMOKE$، $SEXP$، $CESD$ و $SMOKE*SEXP$ به عنوان متغیرهای مهم شناسایی شدند.

نتایج نشان می‌دهند که مدل در مقایسه با سایر مدل‌های رقیب، در برآورد ضرایب غیر صفر و انتخاب مدل عملکرد بهتری دارد. به عبارت دیگر وقتی داده‌ها دارای روند غیرخطی هستند، مدل‌های خطی-جزئی نسبت به مدل‌های خطی کارآمدتر هستند. نتایج در هر دو حالت $p_n < n$ و $p_n > n$ برقرار است.

[46] Coverage probability

جدول ۶.۲: نتایج تحلیل داده‌های $CD4$: برآورد پارامترهای رگرسیونی (SD) تحت برازش سه مدل $P-SMM$، SMM و $P-LMM$.

متغیرها	SMM $\widehat{\beta}(SD)$	$P-LMM$ $\widehat{\beta}(SD)$	$P-SMM$ $\widehat{\beta}(SD)$
AGE	۰٫۰۸۴ (۰٫۰۰۴)	−۰٫۶۶۱ (۰٫۰۰۵)	۰ (۰)
$SMOKE$	۰٫۷۳۰ (۰٫۰۲۴)	۲٫۴۹۸ (۰٫۰۳۰)	۰٫۴۱۷(۰٫۰۱۴)
$DRUG$	۰٫۸۱۱ (۰٫۰۴۷)	۶٫۶۰۸ (۰٫۰۳۸)	۰ (۰)
$SEXP$	۰٫۱۷۰ (۰٫۰۰۹)	۱٫۱۴۵ (۰٫۰۰۸)	۰٫۰۸۴(۰٫۰۰۶)
$CESD$	−۰٫۰۶۹ (۰٫۰۰۴)	۰ (۰)	−۰٫۰۵۴(۰٫۰۰۳)
$AGE*SMOKE$	۰٫۰۱۲ (۰٫۰۰۱)	۰٫۰۱۶ (۰٫۰۰۱)	۰ (۰)
$AGE*DRUG$	−۰٫۰۲۰ (۰٫۰۰۴)	۰٫۰۷۰ (۰٫۰۰۵)	۰ (۰)
$AGE*SEXP$	−۰٫۰۱۲ (۰٫۰۰۱)	۰ (۰)	۰ (۰)
$AGE*CESD$	−۰٫۰۰۵ (۰٫۰۰۰۲)	۰ (۰)	۰ (۰)
$SMOKE*DRUG$	−۰٫۲۷۸ (۰٫۰۲۲)	−۱٫۱۶۶ (۰٫۰۳۱)	۰ (۰)
$SMOKE*SEXP$	۰٫۰۳۸ (۰٫۰۰۳)	۰ (۰)	۰٫۰۴۲(۰٫۰۰۳)
$SMOKE*CESD$	−۰٫۰۰۵ (۰٫۰۰۱)	۰ (۰)	۰ (۰)
$DRUG*SEXP$	−۰٫۰۷۹ (۰٫۰۰۹)	−۰٫۵۵۸ (۰٫۰۱۰)	۰ (۰)
$DRUG*CESD$	۰٫۰۲۰ (۰٫۰۰۳)	۰ (۰)	۰ (۰)
$SEXP*CESD$	−۰٫۰۰۱ (۰٫۰۰۰۳)	۰ (۰)	۰ (۰)
$\ell_{\max}$	−۱۶۵۶۴٫۶۲	−۱۷۳۰۶٫۱۹	−۱۶۵۵۳٫۷۲
AIC	۳۳۱۵۹٫۲۴	۳۴۶۴۲٫۳۷	۳۳۱۳۷٫۴۵
BIC	۳۳۲۱۷٫۹۰	۳۴۷۰۱٫۰۳	۳۳۱۹۶٫۱۱

فصل ۳

انتخاب متغیر برای داده‌های طولی در خانواده توزیع‌های نمایی

در بسیاری از تحقیقات علمی، با داده‌های رسته‌ای[1] یا داده‌های غیرنرمال، مانند پاسخ‌های دودویی و شمارشی مواجه هستیم. نلدر و ودربرن (۱۹۷۲) اولین کسانی بودند که مدل‌های خطی تعمیم یافته[2] (GLM) را معرفی کردند و مک‌کلاچ و نلدر (۱۹۸۹) مدل‌های GLM را برای مدل‌بندی متغیرهای پاسخ غیرنرمال پیشنهاد دادند. در این مدل‌ها با پذیرفتن استقلال مشاهدات، با استفاده از یک تابع پیوند بین میانگین مشاهدات و متغیرهای کمکی ارتباط خطی برقرار می‌شود. هنگامی که بین مشاهدات همبستگی وجود دارد، تعمیمی از مدل‌های مذکور به نام مدل‌های آمیخته خطی تعمیم‌یافته ($GLMM$) استفاده می‌شود که توسط بریسلو و کلیتون (۱۹۹۳) ارائه شد؛ که در آن پیشگوهای مدل علاوه بر اثرات ثابت شامل اثرات تصادفی نیز هستند. در $GLMM$ها، تابع درستنمایی برخلاف مدل‌های خطی به دلیل غیرنرمال بودن متغیرهای پاسخ و وجود اثرات تصادفی، صورت بسته‌ای نداشته و برآورد پارامترها به راحتی امکان‌پذیر نیست. لذا در اکثر مقالات با فرض نرمال بودن اثرات تصادفی به ارائه راه‌حلی برای برآورد پارامترهای مدل و اثرات تصادفی با بیشینه کردن توابع درستنمایی، شبه درستنمایی[3] و درستنمایی سلسله مراتبی[4] به روش‌های عددی پرداخته

[1] Categorical
[2] Generalized linear models
[3] Quasi likelihood
[4] Hierarchical likelihood

شده است. از جمله مک‌کلاچ (۱۹۹۷)، با بکار بردن روش‌های عددی مانند بیشینه‌سازی امیدریاضی مونت‌کارلویی[۵] ($MCEM$)، نیوتون رافسون مونت‌کارلویی[۶] ($MCNR$) و بیشینه درستنمایی شبیه‌سازی شده[۷] (SML) برآورد بیشینه درستنمایی پارامترها را در مدل $GLMM$ به‌دست آورد. همچنین در مدل‌های GLM، پن و تامپسون (۲۰۰۷)، برآورد شبه مونت کارلو[۸] و لین (۲۰۰۷) برآورد پارامترهای GLM پواسنی را با ترکیب درستنمایی نما[۹] و شبه درستنمایی تاوانیده مورد مطالعه قرار دادند. ناتاراجان و کّس (۲۰۰۰) با بسط روش‌های بیزی در تحلیل مدل‌های GLM، نحوه انتخاب پیشین در این مدل‌ها را مطالعه کردند.

مطالعات طولی که متغیر پاسخ آن‌ها غیرنرمال باشد نیز در علوم مختلف جایگاه ویژه‌ای یافته‌اند. استفاده از $GLMM$ در تحلیل داده‌های طولی (زیگر و کریم ، ۱۹۹۱) بسیار مفید واقع شده‌اند، زیرا تفاوت بین گروه‌ها را می‌توان با استفاده از اثرات تصادفی مدل‌بندی کرد. با این حال این مدل‌ها به دلیل اعمال فرضیات پارامتری اثرات ثابت، در تحلیل داده‌های طولی که متغیر پاسخ آن‌ها دارای رفتارهای پیچیده در طول زمان است، دارای محدودیت‌هایی هستند. برای از بین بردن محدودیت $GLMM$ برای مدل‌سازی روند غیرخطی زمان، مدل آمیخته نیمه‌پارامتری تعمیم‌یافته[۱۰] ($GSMM$)، برای تحلیل داده‌های طولی استفاده می‌شود که در آن همبستگی درون آزمودنی‌ها با استفاده از اثرات تصادفی و مدل‌سازی اثر زمان توسط یک تابع هموار درنظر گرفته می‌شود. فن و همکاران (۲۰۰۷)، چن و چو (۲۰۰۷)، چن و چو (۲۰۰۹)، لیانگ (۲۰۰۹) و کروم و همکارارن (۲۰۱۶) از مدل $GSMM$ در تحلیل داده‌های طولی استفاده کرده‌اند.

اگرچه انتخاب مدل برای داده‌های طولی نرمال تا اندازه‌ای گسترش یافته است، تحقیقات در مورد انتخاب مدل برای داده‌های طولی غیرنرمال در چارچوب GLM تا حد زیادی ناشناخته مانده است. در این میان پن (۲۰۰۱) معیار اطلاع شبه‌درستنمایی[۱۱] (QIC)، کانتونی و همکاران (۲۰۰۵) معیار C_p-مالو و وانگ و کوای (۲۰۰۹) معیار BIC را برای انتخاب متغیر داده‌های طولی در مدل‌های GLM تعمیم دادند. این روش‌ها انتخاب متغیر را بر اساس معیارهای کلاسیک انجام می‌دهند. همان‌طور که پیش از این نیز بیان شد هنگامی که بعد داده‌ها بالاست این روش‌ها از نظر محاسباتی کارایی ندارند. در زمینه انتخاب متغیر داده‌های طولی براساس روش مبتنی بر تابع تاوان، فو (۲۰۰۳) رگرسیون بریج و لاسو، سوای و همکاران (۲۰۱۰) روش غربالگری مستقل و زیاک (۲۰۰۶) روش اسکد را در مدل‌های خطی تعمیم یافته پیشنهاد دادند. در تمامی تحقیقات فوق که تعمیم GLM تاوانیده به داده‌های طولی است، بعد مدل ثابت است. در زمینه داده‌های طولی با بعد بالا، وانگ و همکاران (۲۰۱۲) و سو و همکاران (۲۰۱۳) روش GEE تاوانیده با تابع تاوان

[5] Monte Carlo Expectation Maximization
[6] Monte carlo Newton Raphson
[7] Simulated Maximum Likelihood
[8] Quasi Monte Carlo
[9] Pseudo Likelihood
[10] Generalized semiparametric mixed models
[11] Quasi-likelihood information criterion

اسکد را پیشنهاد دادند.

در این فصل با تمرکز روی مدل‌های $GSMM$، به معرفی اجمالی مدل‌های آمیخته نیمه‌پارامتری تعمیم‌یافته پرداخته می‌شود. سپس برآورد و انتخاب متغیر مبتنی بر رهیافت تابع تاوان برای داده‌های طولی خانواده توزیع‌های نمایی با بعد بالا و بعد پایین درنظر گرفته می‌شود. برآوردگرهای معرفی‌شده صورت بسته ندارند؛ لذا با بکارگیری روش سو و همکاران (۲۰۱۳)، برآوردگر GEE تاوانیده را معرفی کرده و با استفاده از یک الگوریتم تکراری به نام الگوریتم نیوتن رافسون مونت کارلویی تاوانیده، به طور همزمان برآورد و انتخاب متغیر انجام می‌شود. علاوه بر این، خواص مجانبی برآوردگرها در چارچوب داده‌های بعد بالا که در آن تعداد پارامترها متناسب با افزایش حجم نمونه افزایش می‌یابد، مورد بررسی قرار می‌گیرد. به منظور بررسی عملکرد روش معرفی‌شده، مطالعه‌ای شبیه‌سازی و همچنین تحلیل مجموعه داده‌های واقعی فراهم آمده است.

۱.۳ انتخاب متغیر برای داده‌ها با بعد بالا

۱.۱.۳ مدل آمیخته نیمه‌پارامتری تعمیم یافته

یک نمونه تصادفی n تایی از داده‌های طولی با بعد بالا را درنظر بگیرید. فرض کنید متغیرهای پاسخ به شرط اثرات تصادفی از هم مستقل و هر $y_{ij}|\boldsymbol{u}_i$ دارای خانواده توزیع‌های نمایی باشد، به طوری که تابع چگالی احتمال به صورت

$$p(y_{ij}|\boldsymbol{u}_i,\boldsymbol{\beta}_n,\phi)=\exp\left[\phi^{-1}\{y_{ij}\theta_{ij}-b(\theta_{ij})\}+c(y_{ij},\phi)\right], \qquad (۱.۳)$$

است، که در آن θ_{ij} پارامتر کانونی، ϕ پارامتر پراکندگی، $a(.)$ و $b(.)$ توابع معلوم هستند. همچنین امید و واریانس شرطی به ترتیب به صورت $\mu_{ij}=E(y_{ij}|\boldsymbol{u}_i)=b'(\theta_{ij})$ و $\nu_{ij}=var(y_{ij}|\boldsymbol{u}_i)=\phi b''(\theta_{ij})$ هستند، به طوری که $b'(\theta)=\frac{\partial b(\theta)}{\partial\theta}$ و $b''(\theta)=\frac{\partial^2 b(\theta)}{\partial\theta^2}$. اکنون با فرض استقلال شرطی مشاهدات، با استفاده از تابع پیوند معلوم g، داریم

$$g(\mu_{ij})\triangleq\eta_{ij}=\boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta}_n+\boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i+f(t_{ij}),\quad i=1,\ldots,n;\ j=1,\ldots,n_i. \qquad (۲.۳)$$

در مدل فوق y_{ij} پاسخ مربوط به آزمودنی iام در زمان t_{ij} است، که در آن $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_{p_n})^{\top}$ برداری $p_n\times 1$ بعدی از ضرایب ثابت رگرسیونی وابسته به متغیرهای تبیینی $\boldsymbol{x}_{ij}$؛ $f(.)$ تابعی نامعلوم، هموار و داری مشتق دوم؛ $\boldsymbol{u}_i=(u_{i1},\ldots,u_{iq})^{\top}$ برداری $q\times 1$ بعدی از ضرایب تصادفی، مستقل از هم و وابسته به متغیرهای تبیینی z_{ij} و دارای توزیعی به صورت

$$\boldsymbol{u}_i\sim f_u(\boldsymbol{u}_i|\boldsymbol{\Sigma}), \qquad (۳.۳)$$

هستند. مجموع روابط (۱.۳)–(۳.۳) به عنوان مدل آمیخته نیمه‌پارامتری تعمیم‌یافته ($GSMM$) شناخته می‌شود.

حال فرض کنید مولفه ناپارامتری مدل به روش اسپلاین‌های رگرسیونی شرح داده‌شده در ۳.۵.۱ تقریب زده شود. بدین ترتیب رابطه (۲.۳) یک مدل رگرسیونی به صورت

$$g(\mu_{ij}) \triangleq \eta_{ij} = \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta}_n + \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i + \boldsymbol{B}(t_{ij})\boldsymbol{\alpha}_n, \quad i = ۱,\ldots,n; j = ۱,\ldots,n_i. \qquad (۴.۳)$$

را نتیجه می‌دهد. برای سادگی در نوشتار، مدل را می‌توان به صورت

$$g(\mu_{ij}) \triangleq \eta_{ij} = \boldsymbol{D}_{ij}^{\top}\boldsymbol{\theta}_n + \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i, \quad i = ۱,\ldots,n; j = ۱,\ldots,n_i,$$

بازنویسی کرد، به طوری که $\boldsymbol{D}_{ij} = (\boldsymbol{X}_{ij}^{\top}, \boldsymbol{B}_j(\boldsymbol{t}_i)^{\top})^{\top}$ ماتریس طرح ادغام‌شده $(p_n + h_n) \times n_i$ بعدی از ماتریس اثرات ثابت و اثرات اسپلاین و $\boldsymbol{\theta}_n = (\boldsymbol{\beta}_n^{\top}, \boldsymbol{\alpha}_n^{\top})^{\top}$ بردار پارامتر ادغام‌شده $(p_n + h_n) \times ۱$ بعدی هستند. همچنین $h_n = d + K_n + ۱$، d درجه تقریب اسپلاین و K_n تعداد گره‌ها در تقریب اسپلاین است.

لذا به طور خلاصه یک $GSMM$ را به صورت

$$\begin{aligned} p(y_{ij}|\boldsymbol{u}_i, \boldsymbol{\theta}_n, \phi) &= \exp\left[\phi^{-۱}\{y_{ij}\theta_{ij} - b(\theta_{ij})\} + c(y_{ij}, \phi)\right], \\ \boldsymbol{u}_i &\sim f_u(\boldsymbol{u}_i|\boldsymbol{\Sigma}), \\ \mu_{ij} &= \mathrm{E}(y_{ij}|\boldsymbol{u}_i), \\ g(\mu_{ij}) \triangleq \eta_{ij} &= \boldsymbol{D}_{ij}^{\top}\boldsymbol{\theta}_n + \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i, \quad i = ۱,\ldots,n; j = ۱,\ldots,n_i, \end{aligned} \qquad (۵.۳)$$

خواهیم داشت. اکنون با توجه به مدل (۵.۳) که یک $GLMM$ به شمار می‌آید، می‌توان تمام تکنیک‌های برآورد در $GLMM$ را برای این مدل نیز بکار برد.

۲.۱.۳ انتخاب متغیر در $GSMM$

۱.۲.۱.۳ الگوریتم $MCNR$

یک روش معمول برای برآورد پارامترها در $GLMM$ها، روش بیشینه درستنمایی است. برای مدل (۵.۳)، تابع درستنمایی به صورت

$$L(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}, \phi) = \prod_{i=۱}^{n} \int_{\mathbb{R}^q} p_{\boldsymbol{y}_i|\boldsymbol{u}_i}(\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{\theta}_n, \phi) p_{\boldsymbol{u}}(\boldsymbol{u}_i|\boldsymbol{\Sigma}) d\boldsymbol{u}_i \qquad (۶.۳)$$

است که در آن $\boldsymbol{y}_i = (\boldsymbol{y}_{i۱}, \ldots, \boldsymbol{y}_{in_i})^{\top}$، $\boldsymbol{u} = (\boldsymbol{u}_۱, \ldots, \boldsymbol{u}_n)$ و

$$p_{\boldsymbol{y}_i|\boldsymbol{u}_i}(\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{\theta}_n, \phi) = \prod_{j=۱}^{n_i} p(\boldsymbol{y}_{ij}|\boldsymbol{u}_i, \boldsymbol{\theta}_n, \phi).$$

تابع درستنمایی (۵.۳) دارای صورت بسته نیست و بعد انتگرال‌گیری برابر با تعداد اثرات تصادفی مدل است. لذا برای به‌دست آوردن برآورد ML از الگوریتم بیشینه‌سازی امیدریاضی مونت‌کارلویی (مک‌کلاچ ، ۱۹۹۷) استفاده می‌کنیم. بدین ترتیب که اثرات تصادفی $\boldsymbol{u}_i$ را به عنوان مقادیر گمشده

و $(\boldsymbol{y}_i, \boldsymbol{u}_i)$ را داده‌های کامل در ظر می‌گیرد. سپس لگاریتم تابع درستنمایی برای داده‌های کامل به صورت

$$\ell(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}, \phi) = \sum_{i=1}^{n} \ln p_{\boldsymbol{y}_i|\boldsymbol{u}_i}(\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{\theta}_n, \phi) + \sum_{i=1}^{n} \ln p_{\boldsymbol{u}_i}(\boldsymbol{u}_i|\boldsymbol{\Sigma}). \qquad (۷.۳)$$

ارائه می‌شود. درنظر گرفتن $\boldsymbol{u}_i$ به عنوان مقادیر گمشده این مزیت را دارد که در گام-M، برای بیشینه‌سازی نسبت به پارامترهای $\boldsymbol{\theta}_n$ و ϕ کافی است که تنها امید عبارت اول در رابطه (۷.۳) یعنی $\mathrm{E}\left[\ln p_{y_{ij}|\boldsymbol{u}_i}(y_{ij}|\boldsymbol{u}_i, \boldsymbol{\theta}_n)|y_{ij}\right]$ را بیشینه کنیم و برای بیشینه‌سازی نسبت به پارامترهای $\boldsymbol{\Sigma}$ تنها از امید عبارت دوم رابطه (۷.۳) یعنی $\mathrm{E}\left[\ln p_{\boldsymbol{u}_i}(\boldsymbol{u}_i|\boldsymbol{\Sigma})|y_{ij}\right]$ استفاده می‌شود. بنابراین معادلات درستنمایی برای $\boldsymbol{\theta}_n$ و $\boldsymbol{\Sigma}$، به ترتیب به صورت

$$\mathrm{E}\Big[\frac{\partial \ln p_{y_{ij}|\boldsymbol{u}_i}(y_{ij}|\boldsymbol{u}_i, \boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n}|y_{ij}\Big] = \mathbf{0},$$
$$\mathrm{E}\Big[\frac{\partial \ln p_{\boldsymbol{u}_i}(\boldsymbol{u}_i|\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}}|y_{ij}\Big] = \mathbf{0},$$

حاصل می‌شود. ملاحظه می‌شود که عبارت اول در رابطه فوق مشابه بیشینه‌سازی در مدل خطی تعمیم‌یافته است. بنابراین در این مرحله منطقی است که برای متناسب‌سازی با مدل خطی تعمیم‌یافته، رویکردی نظیر روش نیوتن رافسون را به کار ببریم. در این شرایط مک‌کلاچ (۱۹۹۷) الگوریتم نیوتن رافسون مونت کارلویی[12] $(MCNR)$ را پیشنهاد داد. با بسط دادن این الگوریتم به داده‌های طولی، معادلات برآورد به صورت

$$\mathrm{E}_{\boldsymbol{u}|\boldsymbol{y}}\Big[n^{-1}\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)}{\partial \boldsymbol{\theta}_n^\top} \boldsymbol{V}_i^{-1}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big)\Big] = \mathbf{0}, \qquad (۸.۳)$$

ارائه می‌گردد، که در آن $\boldsymbol{\mu}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i) = (\mu_{i1}, \ldots, \mu_{in_i})^\top$ و $\boldsymbol{V}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)$ ماتریس کوواریانس $\boldsymbol{y}_i|\boldsymbol{u}_i$ است. الگوریتم $MCNR$ معرفی‌شده را می‌توان تعمیمی از روش GEE از بخش ۱.۱.۳.۱ در نظر گرفت. لذا مشابه روش GEE، ماتریس کوواریانس $\boldsymbol{V}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)$ را براساس ماتریس همبستگی $\boldsymbol{R}(\boldsymbol{\rho}) : \boldsymbol{V}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i) = \boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)$، که $\boldsymbol{A}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i) = \mathrm{diag}(\nu_{i1}, \ldots, \nu_{in_i})$ و $\boldsymbol{\rho}$ برداری از پارامترهای همبستگی است، مشخص می‌کنیم. در فصل ۱ برای $\boldsymbol{R}(\boldsymbol{\rho})$ ساختارهای متعددی معرفی شد که به ازای یک ساختار خاص می‌توان پارامترهای همبستگی را به روش‌های متفاوت برآورد کرد. با درنظر گرفتن برآورد ماتریس کوواریانس به صورت $\widehat{\boldsymbol{R}} \equiv \boldsymbol{R}(\widehat{\boldsymbol{\rho}})$، رابطه (۸.۳) به صورت

$$\mathrm{E}_{\boldsymbol{u}|\boldsymbol{y}}\Big[n^{-1}\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\widehat{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{-\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big)\Big] = \mathbf{0}, \qquad (۹.۳)$$

درمی‌آید، که در آن $\boldsymbol{D}_i = (\boldsymbol{D}_{i1}^\top, \ldots, \boldsymbol{D}_{in_i}^\top)^\top$. برآوردگر $\widehat{\boldsymbol{\theta}}_n$ با حل معادلات (۹.۳) به دست می‌آید. در ادامه برای سادگی در نوشتار و تفسیر فرض می‌کنیم $\phi = 1$ و $n_i = m < \infty$.

حال از طریق بهینه‌سازی تابع هدف مبتنی بر تابع تاوان به برآورد و انتخاب متغیر می‌پردازیم.

[12] Monte carlo newton raphson

۲.۲.۱.۳ الگوریتم $MCNR$ تاوانیده

به منظور انتخاب متغیرهای مهم و برآورد ضرایب آن‌ها، به لگاریتم تابع درستنمایی (۷.۳)، جمله تاوان $\sum_{k=1}^{p} P_{\lambda_n}(|\beta_k|)$ را اضافه می‌کنیم، آنگاه لگاریتم تابع درستنمایی تاوانیده به‌صورت

$$\ell^p(\boldsymbol{\beta}_n, \boldsymbol{\alpha}_n, \boldsymbol{D}, \phi) = \sum_{i=1}^{n} \ln p_{\boldsymbol{y}_i|\boldsymbol{u}_i}(\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{\theta}_n) + \sum_{i=1}^{n} p_{\boldsymbol{u}_i}(\boldsymbol{u}_i|\boldsymbol{\Sigma}) - n\sum_{k=1}^{p_n} p_{\lambda_n}(|\beta_{nk}|), \quad (۱۰.۳)$$

به‌دست می‌آید. از آن‌جایی که ضرایب $\boldsymbol{\theta}_n$ با جملات اول و سوم عبارت فوق وابسته است، معادلات برآورد تاوانیده به صورت

$$\boldsymbol{U}_n(\boldsymbol{\theta}_n) = \boldsymbol{S}_n(\boldsymbol{\theta}_n) - q_{\lambda_n}(|\boldsymbol{\beta}_n|)\mathrm{sgn}(\boldsymbol{\beta}_n),$$

پیشنهاد می‌شود، به‌طوری‌که

$$\boldsymbol{S}_n(\boldsymbol{\theta}_n) = \mathrm{E}_{u|y}\Big[\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\widehat{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{-\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big)\Big],$$

$q_{\lambda_n}(|\boldsymbol{\beta}_n|) = \big(q_{\lambda_n}(|\beta_{n1}|), \ldots, q_{\lambda_n}(|\beta_{np_n}|)\big)^\top$ برداری $p_n \times 1$ بعدی از توابع تاوان، $\mathrm{sgn}(\boldsymbol{\beta}_n) = \big(\mathrm{sgn}(\beta_{n1}), \ldots, \mathrm{sgn}(\beta_{np_n})\big)^\top$ و $q_{\lambda_n}(|\beta_{nk}|) = p'_{\lambda_n}(|\beta_{nk}|)$ هستند.

درنظر داشته باشید که برای سادگی، فرض می‌کنیم قسمت ناپارامتری سهم قابل توجهی در مدل داشته باشد و معادلات برآورد تاوانیده تنها مولفه‌های کوچک $\boldsymbol{\beta}_n$ را منقبض می‌کند و نه $\boldsymbol{\alpha}_n$ها را. در اینجا تابع تاوان اسکد با تقریب درجه دو موضعی (LQA) را به کار می‌بریم. برآوردگر $\boldsymbol{\theta}_n$ باید جواب معادله $\boldsymbol{U}_n(\boldsymbol{\theta}_n) = \mathbf{0}$ باشد، اما از آن‌جایی که $\boldsymbol{U}_n(\boldsymbol{\theta}_n)$ دارای نقاط ناپیوسته است، ممکن است جواب دقیق برای $\boldsymbol{U}_n(\boldsymbol{\theta}_n) = \mathbf{0}$ وجود نداشته باشد. لذا برآوردگر $\widehat{\boldsymbol{\theta}}_n$ را به عنوان جواب تقریبی معادلات برآورد می‌شناسیم، به عبارت دیگر به ازای دنباله $a_n \to 0$ داشته باشیم $\boldsymbol{U}_n(\widehat{\boldsymbol{\theta}}_n) = o(a_n)$. اکنون برای حل مسئله بهینه سازی تاوانیده $\boldsymbol{U}_n(\widehat{\boldsymbol{\theta}}_n) = o(a_n)$ از الگوریتم تکراری نیوتن رافسون استفاده شده و رابطه بازگشتی به‌صورت

$$\widehat{\boldsymbol{\theta}}_n^{(m+1)} = \widehat{\boldsymbol{\theta}}_n^{(m)} + \Big\{\boldsymbol{H}_n(\widehat{\boldsymbol{\theta}}_n^{(m)}) + n\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}_n^{(m)})\Big\}^{-1} \times \Big\{\boldsymbol{S}_n(\widehat{\boldsymbol{\theta}}_n^{(m)}) + n\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}_n^{(m)})\widehat{\boldsymbol{\theta}}_n^{(m)}\Big\}, \quad (۱۱.۳)$$

حاصل می‌شود، به‌طوری‌که

$$\boldsymbol{H}_n(\widehat{\boldsymbol{\theta}}_n^{(m)}) = \mathbb{E}_{u|y}\Big[\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\widehat{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{D}_i\Big],$$

$$\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}_n^{(m)}) = \mathrm{diag}\Big\{\frac{q_{\lambda_n}(|\beta_{n1}|)}{\epsilon + |\beta_{n1}|}, \ldots, \frac{q_{\lambda_n}(|\beta_{np_n}|)}{\epsilon + |\beta_{np_n}|}, \mathbf{0}_{h_n}\Big\}, \quad \epsilon = 10^{-6}$$

و $\mathbf{0}_{h_n}$ بردار h_n بعدی با عناصر 0 است.

عبارت‌های امید ریاضی در رابطه (۱۱.۳) صورت بسته ندارند، زیرا توزیع شرطی $\boldsymbol{u}_i|\boldsymbol{y}_i$ به توزیع حاشیه‌ای $\boldsymbol{y}_i$ وابسته است که محاسبه صریح آن کار ساده‌ای نیست. مشابه [۷۸] از یک روش جایگزین

استفاده می‌کنیم بدین ترتیب که با استفاده از الگوریتم متروپولیس تانر (۱۹۹۹۳) نمونه‌هایی تصادفی از توزیع شرطی $\boldsymbol{u}_i|\boldsymbol{y}_i$ تولید می‌کنیم که دیگر به تعیین توزیع حاشیه‌ای $\boldsymbol{y}_i$ نیازی نیست. الگوریتم متروپولیس، یک روش زنجیر مارکوف مونت کارلویی[۱۳] برای به‌دست آوردن نمونه‌های تصادفی از یک توزیع احتمالی است که نمونه‌برداری مستقیم از آن دشوار می‌باشد. این نمونه‌ها را می‌توان برای برآورد یک توزیع یا برای محاسبه برخی انتگرال‌ها (به‌طور مثال یک امید ریاضی) استفاده کرد. در این الگوریتم یک توزیع پیشنهادی متناسب با مسئله مورد نظر انتخاب می‌شود به طوری که مشاهدات جدید از آن ساخته می‌شوند. توزیع پیشنهادی باید به گونه‌ای ساخته شود که نمونه‌گیری از آن ساده باشد. سپس با تعیین یک تابع پذیرش، احتمال پذیرش مشاهدات جدید در مقابل مشاهدات گذشته محاسبه می‌شود. حال فرض کنید $\boldsymbol{U}$ مشاهدات قبلی از توزیع شرطی $\boldsymbol{U}|\boldsymbol{y}$ باشد و مقادیر جدید u_k^* را برای مولفه kام $(u_1, \ldots, u_{k-1}, u_k^*, u_{k+1}, \ldots, u_{nq})$ $\boldsymbol{U}^*$ تولید می‌کنیم. با در نظر گرفتن $p_{\boldsymbol{u}}$ به عنوان توزیع پیشنهادی، $\boldsymbol{U}^*$ را با احتمال

$$\alpha_k(\boldsymbol{U}, \boldsymbol{U}_*) = \min\Big\{1, \frac{p_{u|y}(\boldsymbol{U}^*|\boldsymbol{y}, \boldsymbol{\theta}_n, \boldsymbol{D}) p_u(\boldsymbol{U}|\boldsymbol{D})}{p_{u|y}(\boldsymbol{U}|\boldsymbol{y}, \boldsymbol{\theta}_n, \boldsymbol{D}) p_u(\boldsymbol{U}^*|\boldsymbol{D})}\Big\}. \qquad (۱۲.۳)$$

می‌پذیریم. در صورت رد مقادیر جدید $\boldsymbol{U}^*$ ، مقادیر قبلی $\boldsymbol{U}$ باقی می‌مانند. عبارت دوم در رابطه (۱۲.۳)، را می‌توان به صورت

$$\begin{aligned}\frac{p_{\boldsymbol{u}|\boldsymbol{y}}(\boldsymbol{U}^*|\boldsymbol{y}, \boldsymbol{\theta}_n, \boldsymbol{D}) p_u(\boldsymbol{U}|\boldsymbol{D})}{p_{\boldsymbol{u}|\boldsymbol{y}}(\boldsymbol{U}|\boldsymbol{y}, \boldsymbol{\theta}_n, \boldsymbol{D}) p_u(\boldsymbol{U}^*|\boldsymbol{D})} &= \frac{p_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{U}^*, \boldsymbol{\theta}_n)}{p_{\boldsymbol{y}|\boldsymbol{u}}(\boldsymbol{y}|\boldsymbol{U}, \boldsymbol{\theta}_n)} \\ &= \frac{\prod_{i=1}^n p_{\boldsymbol{y}_i|\boldsymbol{u}}(\boldsymbol{y}_i|\boldsymbol{U}^*, \boldsymbol{\theta}_n)}{\prod_{i=1}^n p_{\boldsymbol{y}_i|\boldsymbol{u}}(\boldsymbol{y}_i|\boldsymbol{U}, \boldsymbol{\theta}_n)},\end{aligned}$$

نوشت. محاسبه تابع پذیرش $\alpha_k(\boldsymbol{U}, \boldsymbol{U}^*)$ تنها شامل توزیع شرطی $\boldsymbol{y}|\boldsymbol{u}$ است که صورت بسته آن قابل محاسبه است. در پایان الگوریتم $MCNR$ به صورت الگوریتم ۳ پیشنهاد می‌شود.

این روش برآوردگرهای سازگاری برای ضرایب مدل ارائه می‌دهد که مستقل از انتخاب ساختار همبستگی بوده و $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ دارای توزیع نرمال مجانبی چندمتغیره با میانگین $\mathbf{0}$ و ماتریس واریانس-کوواریانس به‌صورت

$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_n) \approx [\boldsymbol{H}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i) + n\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}_n)]^{-1} \boldsymbol{M}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i) [\boldsymbol{H}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i) + n\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i)]^{-1},$$

است، به‌طوری‌که

$$\boldsymbol{M}_n(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i) = \sum_{i=1}^n \boldsymbol{D}_i^\top \boldsymbol{A}_i^{1/2}(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i) \widehat{\boldsymbol{R}}^{-1} \big[\boldsymbol{\epsilon}_i(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i) \boldsymbol{\epsilon}_i^\top(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i)\big] \widehat{\boldsymbol{R}}^{-1} \boldsymbol{A}_i^{1/2}(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{u}_i) \boldsymbol{D}_i^\top,$$

و

$$\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i) = \big(\epsilon_{i1}(\boldsymbol{\theta}_n, \boldsymbol{u}_i), \ldots, \epsilon_{in_i}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big)^\top = \boldsymbol{A}_i^{-1/2}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big).$$

[13] Markov chain Monte Carlo

الگوریتم ۳ الگوریتم $MCNR$ تاوانیده

گام ۱. فرض کنید $m_k = ۰$. برای $\boldsymbol{\theta}_n$ و $\boldsymbol{\Sigma}$ مقادیر اولیه به صورت $\boldsymbol{\theta}_n^{(۰)}$ و $\boldsymbol{\Sigma}^{(۰)}$ انتخاب کنید.

گام ۲. با استفاده از الگوریتم متروپولیس، N مشاهده به صورت $\boldsymbol{U}^{(۱)}, \ldots, \boldsymbol{U}^{(N)}$ از توزیع $p_{\boldsymbol{u}|\boldsymbol{y}}(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\theta}_n^{(m_k)}, \boldsymbol{\Sigma}^{(m_k)})$ تولید کنید. از این مشاهدات برای برآورد مونت کارلویی امید ریاضی‌ها استفاده کنید. یعنی

a) $\boldsymbol{\theta}_n^{(m_k+۱)}$ را به صورت

$$\begin{aligned}\boldsymbol{\theta}_n^{(m_k+۱)} &= \boldsymbol{\theta}_n^{(m_k)} + \Big\{\frac{۱}{N}\sum_{k=۱}^{N}\Big[\boldsymbol{H}_n\big(\widehat{\boldsymbol{\theta}}_n^{(m_k)}, \boldsymbol{U}^{(k)}\big)\Big] + n\boldsymbol{E}_n(\widehat{\boldsymbol{\beta}}_n^{(m_k)})\Big\}^{-۱} \\ &\quad \times\Big\{\frac{۱}{N}\sum_{k=۱}^{N}\Big[\boldsymbol{S}_n(\widehat{\boldsymbol{\theta}}_n^{(m_k)}, \boldsymbol{U}^{(k)})\Big] - n\boldsymbol{E}_n(\widehat{\boldsymbol{\beta}}_n^{(m_k)})\widehat{\boldsymbol{\beta}}_n^{(m_k)}\Big\},\end{aligned}$$

محاسبه کنید، به طوری که

$$\begin{aligned}\boldsymbol{H}_n\big(\widehat{\boldsymbol{\theta}}_n^{(m_k)}, \boldsymbol{U}^{(k)}\big) &= \sum_{i=۱}^{n}\boldsymbol{D}_i^{\top}\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n^{(m_k)}, U_i^{(k)})\widehat{\boldsymbol{R}}^{-۱}\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n^{(m_k)}, U_i^{(k)})\boldsymbol{D}_i, \\ \boldsymbol{S}_n\big(\widehat{\boldsymbol{\theta}}_n^{(m_k)}, \boldsymbol{U}^{(k)}\big) &= \sum_{i=۱}^{n}\boldsymbol{D}_i^{\top}\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n^{(m_k)}, U_i^{(k)})\widehat{\boldsymbol{R}}^{-۱}\boldsymbol{A}_i^{-\frac{۱}{۲}}(\boldsymbol{\theta}_n^{(m_k)}, U_i^{(k)})\big(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}_n^{(m_k)}, U_i^{(k)})\big).\end{aligned}$$

b) $\boldsymbol{\Sigma}^{(m_k+۱)}$ را از طریق بیشینه‌سازی عبارت $\frac{۱}{N}\sum_{k=۱}^{N}\ln f_u(\boldsymbol{U}^{(k)}|\boldsymbol{\Sigma})$ به دست آورید.

c) قرار دهید $m_k = m_{k+۱}$.

گام ۳. گام ۲ را تا رسیدن به همگرایی تکرار کنید. $\boldsymbol{\theta}_n^{(m_k+۱)}$ و $\boldsymbol{\Sigma}^{(m_k+۱)}$ برآوردگر $MCNR$ تاوانیده برای $\boldsymbol{\theta}_n$ و $\boldsymbol{\Sigma}$ هستند.

۳.۲.۱.۳ انتخاب پارامترهای تاوان

برای پیاده‌سازی روش پیشنهادشده، باید درباره چندین پارامتر تصمیم‌گیری مناسبی اتخاذ شود. این پارامترها شامل درجه چند جمله‌ای d، تعداد نقاط گره L_n و تعداد توابع پایه h_n در تقریب B-اسپلاین و پارامترهای a و λ_n در تابع تاوان اسکد هستند. تصمیم‌گیری درباره d، K_n، h_n و a مشابه روند بیان‌شده در بخش ۳.۱.۳.۲ است. همچنین برای پارامتر تاوان λ_n از معیار GCV معرفی‌شده در همان بخش که به‌صورت

$$GCV_{\lambda_n} = \frac{RSS(\lambda_n)/n}{(۱ - d(\lambda_n)/n)^۲},$$

است، استفاده می‌کنیم با این تفات که مولفه‌های آن به‌صورت

$$RSS(\lambda_n) = \frac{۱}{N}\sum_{k=۱}^{N}\Big[\sum_{i=۱}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\theta}}_n, U_i^{(k)}))^\top \boldsymbol{W}_i^{-۱}(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\widehat{\boldsymbol{\theta}}_n, U_i^{(k)}))\Big] \qquad (۱۳.۳)$$

و

$$d(\lambda_n) = \mathrm{tr}\Big[\Big\{\frac{۱}{N}\sum_{k=۱}^{N}\Big[\boldsymbol{H}_n\big(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{U}^{(k)}\big)\Big] + n\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}_n)\Big\}^{-۱} \times \Big\{\frac{۱}{N}\sum_{k=۱}^{N}\Big[\boldsymbol{H}_n\big(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{U}^{(k)}\big)\Big]\Big)\Big\}\Big]$$

تعیین می‌شوند. $\boldsymbol{W}_i$ در (۱۳.۳) ماتریس کوواریانس $n_i \times n_i$ برای پاسخ‌های $\boldsymbol{y}_i$ است و به صورت $\boldsymbol{W}_i = \mathrm{E}_{u|y}\Big(\mathrm{Var}(\boldsymbol{y}_i|\boldsymbol{u}_i)\Big) + \mathrm{Var}_{u|y}\Big(\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{u}_i)\Big)$، محاسبه می‌شود، به طوری که

$$\begin{aligned}
\mathrm{E}_{u|y}\Big(\mathrm{var}(\boldsymbol{y}_i|\boldsymbol{u}_i)\Big) &= \frac{۱}{N}\sum_{k=۱}^{N}\Big[\boldsymbol{V}_i(\widehat{\boldsymbol{\theta}}_n, U_i^{(k)})\Big], \\
\mathrm{Var}_{u|y}\Big(\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{u}_i)\Big) &= \frac{۱}{N}\sum_{k=۱}^{N}\Big[\boldsymbol{\mu}_i(\widehat{\boldsymbol{\theta}}_n, U_i^{(k)})\Big]^۲ - \Big[\frac{۱}{N}\sum_{k=۱}^{N}\Big[\boldsymbol{\mu}_i(\widehat{\boldsymbol{\theta}}_n, U_i^{(k)})\Big]\Big]^۲.
\end{aligned}$$

۳.۱.۳ خواص مجانبی

در این بخش رفتار مجانبی برآوردگرهای $\widehat{\boldsymbol{\beta}}_n$ و $\widehat{f}$ تحت الگوریتم $MCNR$ پیشنهادشده، مطالعه می‌شوند. توجه داشته باشید که در این بخش نیز، تعاریف، شرایط و تفکیک‌های معرفی‌شده در بخش ۲.۳.۲ برقرار است. همچنین شرایط نظم زیر را درنظر بگیرید:

(۱.A) به ازای هر $k \geqslant ۲$، مشتق kام $\boldsymbol{f}_۰(\boldsymbol{t}_i)$ کراندار است.

(۲.A) $\boldsymbol{X}_{ij}$، $۱ \leqslant i \leqslant n$، $۱ \leqslant j \leqslant m$ به‌طور یکنواخت کراندار است.

(۳.A) این شرط مشابه شرط (۳.A) در بخش ۲.۳.۲ است.

(۴.A) ماتریس همبستگی واقعی $\boldsymbol{R}_۰$ در دارای مقادیر ویژه کراندار است؛ ماتریس همبستگی برآوردشده $\overline{\boldsymbol{R}}$ در شرط $\|\widehat{\boldsymbol{R}}^{-۱} - \overline{\boldsymbol{R}}^{-۱}\| = O_p(n^{-۱/۲})$ صدق می‌کند به‌طوری‌که $\overline{\boldsymbol{R}}$ یک ماتریس معین مثبت و دارای مقادیر ویژه کراندار است و لازم نیست که $\overline{\boldsymbol{R}}$ برابر با $\boldsymbol{R}_۰$ باشد.

(۵.A) فرض کنید

$$\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i) = \big(\epsilon_{i1}(\boldsymbol{\theta}_n, \boldsymbol{u}_i), \ldots, \epsilon_{in_i}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big)^\top = \boldsymbol{A}_i^{-1/2}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big).$$

مقدار ثابت و متناهی $M_1 > 0$ وجود دارد که به ازای تمام iها و بعضی از $\delta > 0$ها داریم

$$\mathrm{E}(\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n0}, \boldsymbol{u}_i)\|^{2+\delta}) \leqslant M_1;$$

مقادیر ثابت و مثبت M_2 و M_3 وجود دارند که

$$\mathrm{E}\Big[\exp\big(M_2|\epsilon_{ij}(\boldsymbol{\theta}_{n0}, \boldsymbol{u}_i)|\big)\big|\boldsymbol{X}_i\Big] \leqslant M_3.$$

(۶.A) فرض کنید $B_n = \{\boldsymbol{\theta}_n : \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}\| \leqslant \Delta\sqrt{p_n/n}\}$. آن‌گاه $\boldsymbol{\mu}'(\boldsymbol{D}_{ij}^\top\boldsymbol{\theta}_n)$، $\boldsymbol{\mu}''(\boldsymbol{D}_{ij}^\top\boldsymbol{\theta}_n)$ و $\boldsymbol{\mu}^{(3)}(\boldsymbol{D}_{ij}^\top\boldsymbol{\theta}_n)$ به‌طور یکنواخت کراندار هستند.

(۷.A) وقتی $n \longrightarrow \infty$، آن‌گاه $\min_{1\leqslant k\leqslant s_n}|\theta_{n0k}|/\lambda_n \to \infty$، $\lambda_n \to 0$، $s_n^3 n^{-1} = o(1)$، $s_n(\log n)^2 = o(n\lambda_n^2)$، $\log(p_n) = o(n\lambda_n^2/(\log n))$، $p_n s_n^4(\log n)^6 = o(n^2\lambda_n^2)$ و $p_n s_n^3(\log n)^8 = o(n^2\lambda_n^4)$.

(۸.A) تابع تاوان در شرایط $\liminf_{n\to\infty}\liminf_{\theta\to 0+} p'_{\lambda_n}(\theta)/\lambda_n > 0$، $a_n = O(n^{-\frac{1}{2}})$، $b_n \to \infty$، صدق می‌کند و همچنین مقادیر ثابت C و D وجود دارند به‌طوری‌که به ازای هر $\theta_1, \theta_2 > \lambda_n C$ خواهیم داشت $|p''_{\lambda_n}(\theta_1) - p''_{\lambda_n}(\theta_2)| \leqslant D|\theta_1 - \theta_2|$.

حال معادلات برآوردیاب زیر را درنظر بگیرید،

$$\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}) = \mathrm{E}_{u|y}\Big[\sum_{i=1}^{n}\boldsymbol{D}_i^\top\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}, \boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{-\frac{1}{2}}(\boldsymbol{\theta}, \boldsymbol{u}_i)\big(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\theta}, \boldsymbol{u}_i)\big)\Big].$$

فرض کنید $\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_n)$ ماتریس کوواریانس $\overline{\boldsymbol{S}}_n(\boldsymbol{\theta})$ است، پس

$$\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}) = \mathrm{E}_{u|y}\Big[\sum_{i=1}^{n}\boldsymbol{D}_i^\top\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}, \boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{R}_0\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}, \boldsymbol{u}_i)\boldsymbol{D}_i\Big].$$

برای مطالعه رفتار حدی برآوردگرها به بیان نتیجه‌های دیگری نیاز داریم که در قالب لم در زیر بخش ب.۱.۳ از پیوست ارائه شده است. با استفاده از قضیه ۱۲.۷ در اسچوماکر (۱۹۸۱)، $f_0(t)$ توسط $\boldsymbol{B}(t)\boldsymbol{\alpha}_0$ تقریب زده می‌شود، پس داریم

$$\eta_{ij}(\boldsymbol{\theta}_0) = g\big(\mu_{ij}(\boldsymbol{\theta}_0)\big) = \boldsymbol{X}_{ij}^\top\boldsymbol{\beta}_0 + \boldsymbol{B}(t_{ij})\boldsymbol{\alpha}_0 + \boldsymbol{Z}_{ij}^\top\boldsymbol{u}_i, \quad \boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0^\top)^\top_{(p_n+h_n)\times 1}.$$

در ادامه قضایای ۱.۱.۳-۳.۱.۳ بیان می‌شوند که به ترتیب موجود بودن، سازگاری و نرمال بودن برآوردگرهای تاوانیده را هنگامی که $p_n \to \infty$ نشان می‌دهند.

قضیه ۱.۱.۳. تحت شرایط بیان‌شده، برآوردگر $MCNR$ تاوانیده $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1^\top, \widehat{\boldsymbol{\theta}}_2^\top)^\top$ وجود دارد به‌طوری‌که

(i) $\quad \mathbb{P}_n\Big(|U_{nk}(\widehat{\boldsymbol{\theta}}_n)| = 0,\ k = 1, \ldots, s_n^*, (s_n^*+1), \ldots, (s_n = s_n^* + h_n)\Big) \to 1,$

$$\text{(ii)} \qquad \mathbb{P}_n\Big(|U_{nk}(\widehat{\boldsymbol{\theta}}_n)| \leqslant \frac{\lambda_n}{\log n},\ k=(s_n^*+h_n+۱),\ldots,p_n\Big) \to ۱,$$

که در آن

$$U_{nk}(\widehat{\boldsymbol{\theta}}_n) = \begin{cases} S_{nk}(\widehat{\boldsymbol{\theta}}_n) - n\frac{q_{\lambda_n}(|\widehat{\beta}_{nk}|)}{\epsilon+|\widehat{\beta}_{nk}|}\widehat{\beta}_{nk} & k=۱,\ldots,s_n, \\ S_{nk}(\widehat{\boldsymbol{\theta}}_n) & k=(s_n+۱),\ldots,p_n \end{cases},$$

و $S_{nk}(\widehat{\boldsymbol{\theta}}_n)$ مولفه kام $\boldsymbol{S}_n(\widehat{\boldsymbol{\theta}}_n)$ است.

قضیه ۲.۱.۳. تحت شرایط بیان‌شده، وقتی $n^{-۱}p_n^۲ = o(۱)$ آنگاه $\boldsymbol{U}_n(\boldsymbol{\theta}_n) = o(۱)$ دارای جواب $\widehat{\boldsymbol{\theta}}_n$ است به‌طوری‌که

$$\begin{aligned} &(i) \qquad \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{n\circ}\| = O_p(\sqrt{p_n/n}), \\ &(ii) \qquad \frac{۱}{n}\sum_{i=۱}^{n}\sum_{j=۱}^{n_i}\big(\widehat{f}(t_{ij}) - f_\circ(t_{ij})\big)^۲ = O_p(n^{-۲r/(۲r+۱)}). \end{aligned}$$

قضیه ۳.۱.۳. تحت شرایط بیان‌شده، وقتی $L_n = O_p(n^{۱/(۲r+۱)})$ و $n^{-۱}p_n^۳ = o(۱)$ آنگاه به ازای هر $\boldsymbol{\xi}_n \in \mathcal{R}^{p_n}$ که $\|\boldsymbol{\xi}_n\| = ۱$، داریم

$$\begin{aligned} &(i) \qquad \mathbb{P}_n(\widehat{\boldsymbol{\beta}}_{n۲} = \mathbf{0}) \to ۱, \\ &(ii) \qquad \boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{*^{-۱/۲}}(\boldsymbol{\beta}_{n\circ})\overline{\boldsymbol{H}}_n^*(\boldsymbol{\beta}_{n\circ})(\widehat{\boldsymbol{\beta}}_{n۱} - \boldsymbol{\beta}_{n\circ ۱}) \xrightarrow{d} \mathcal{N}_{p_n}(\mathbf{0}, ۱), \end{aligned}$$

به‌طوری‌که

$$\begin{aligned} \overline{\boldsymbol{M}}_n^* &= \mathrm{E}_{u|y}\Big[\sum_{i=۱}^{n}\boldsymbol{X}_i^{*\top}\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{R}_\circ\overline{\boldsymbol{R}}^{-۱}\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\boldsymbol{X}_i^*\Big], \\ \overline{\boldsymbol{H}}_n^* &= \mathrm{E}_{u|y}\Big[\sum_{i=۱}^{n}\boldsymbol{X}_i^{*\top}\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\boldsymbol{X}_i^*\Big], \\ \boldsymbol{X}_i^* &= (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{X}_i \\ \boldsymbol{P} &= \boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{\Omega}\boldsymbol{B})^{-۱}\boldsymbol{B}^\top\boldsymbol{\Omega} \\ \boldsymbol{\Omega} &= \mathrm{diag}\{\boldsymbol{\Omega}_i\} \\ \boldsymbol{\Omega}_i &= \mathrm{E}_{u|y}\Big[\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{A}_i^{\frac{۱}{۲}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\Big] \end{aligned}$$

قضیه ۳.۱.۳ نشان می‌دهد که تحت شرایط نظم بیان‌شده، برآوردگر پیشنهادی می‌تواند ضریب متغیرهای بی‌تاثیر در مدل را با احتمال نزدیک به ۱ به سمت صفر منقبض کند. همچنین ضرایب متغیرهای مهم در مدل، دارای توزیع نرمال مجانبی هستند. اثبات قضایای ۱.۱.۳-۳.۱.۳ در قسمت پیوست ارائه شده است.

۴.۱.۳ مطالعات عددی

در این بخش، فرآیند برازش مدل شامل برآورد پارامترهای مدل و انتخاب متغیر را بر اساس مطالعات شبیه‌سازی و تحلیل داده‌های $CD۴$ بخش ۴.۱ بررسی می‌کنیم.

ابتدا مجموعه‌ای از داده‌های شبیه‌سازی‌شده و یک مورد مطالعاتی با داده‌های واقعی را توسط روش پیشنهادی مقاله یعنی برآوردگر $MCNR$ تاوانیده در مدل نیمه‌پارامتری با اثرات آمیخته تعمیم‌یافته ($P-GSMM$) مورد تحلیل قرار می‌دهیم. سپس به منظور بررسی عملکرد روش $P-GSMM$ در مسئله برآورد و انتخاب متغیر همزمان، نتایج حاصل از تحلیل داده‌ها بدون رهیافت تابع تاوان یعنی روش برآوردگر $MCNR$ در مدل آمیخته خطی تعمیم‌یافته گزارش می‌شود. درنظر داشته باشید که اگر داده‌ها ذاتاً دارای جزء ناپارامتری باشند، مدل نیمه‌پارامتری بهتر از مدل خطی عمل می‌کند. برای نشان دادن این موضوع روش $P-GSMM$ را با روش تاوانیده در مدل آمیخته خطی تعمیم‌یافته ($P-GLMM$) مقایسه می‌کنیم.

۱.۴.۱.۳ مطالعه شبیه‌سازی

در این قسمت متغیرهای پاسخ را از مدل پواسن با عرض از مبدا تصادفی و به صورت

$$y_{ij}|b_i \sim \text{Pois}(\mu_{ij}), \qquad i = ۱,\ldots,n, \quad j = ۱,\ldots,n_i,$$
$$\eta_{ij} = \log(\mu_{ij}) = \textstyle\sum_{k=۱}^{p_n} x_{k,ij}\beta_k + f(t_{ij}) + b_i$$

تولید می‌کنیم که در آن مولفه‌های مدل به صورت

$$\boldsymbol{\beta} = (-۱,-۱,۲,۰,\ldots,۰), \quad b_i \sim \mathcal{N}(۰,\sigma_b^۲); \qquad \sigma_b^۲ = ۰/۲۵,$$
$$\boldsymbol{X}_{ij}^\top = (x_{۱,ij},\ldots,x_{p_n ij}) \sim \mathcal{U}(-۱,۱)$$
$$f(t_{ij}) = ۲sin(۲\pi t_{ij}), \qquad t_{ij} \sim \mathcal{U}(۰,۱),$$

تعیین شده‌اند. تعداد متغیرهای مدل، p_n، با افزایش حجم نمونه افزایش می‌یابد اما بعد متغیرهای تاثیرگذار در مدل ثابت و برابر ۳ است. با توجه به مطالب بیان‌شده در بخش ۱.۳.۳.۲، برای (n,p_n)، موارد $(۲۰۰,۱۶),(۱۰۰,۱۴),(۵۰,۱۱)$ را برای حالت $p_n < n$ و برای حالت $p_n > n$، موارد $(۲۰۰,۲۰۰۰),(۱۰۰,۵۰۰),(۳۰,۱۰۰)$ درنظر گرفته شده است.

برای مقایسه عملکرد روش‌های برآورد، از معیارهای مشابه بخش ۱.۳.۳.۲ از فصل ۲ استفاده شد. نتایج مربوط به این معیارها برای سه روش برآورد بیان‌شده و به ازای مقادیر مختلف (n,p_n) در جداول ۱.۳ و ۲.۳ گزارش شده است. مشابه استدلالی که در بخش ۱.۳.۳.۲ آمده است، روش $P-GSMM$، از نظر دقت برآورد و همچنین انتخاب متغیر بهتر از دو مدل دیگر عمل می‌کند. نتایج جدول ۲.۳ نیز گواهی بر صحت قضایای حدی بیان‌شده است.

همچنین رفتار تابع $\boldsymbol{f(t)}$ از طریق رسم نمودارهایی بررسی شده است. این نمودارها شامل منحنی‌های $\widehat{\boldsymbol{f}}(\boldsymbol{t})$، اریبی، انحراف استاندارد و احتمال پوشش هستند. شکل‌های ۱.۳ و ۲.۳ به

جدول ۱.۳: نتایج شبیه‌سازی: مقایسه نتایج انتخاب متغیر روش‌های $P-GSMM$، $GSMM$ و $P-GLMM$ برای حالت‌های $p_n < n$ و $p_n > n$.

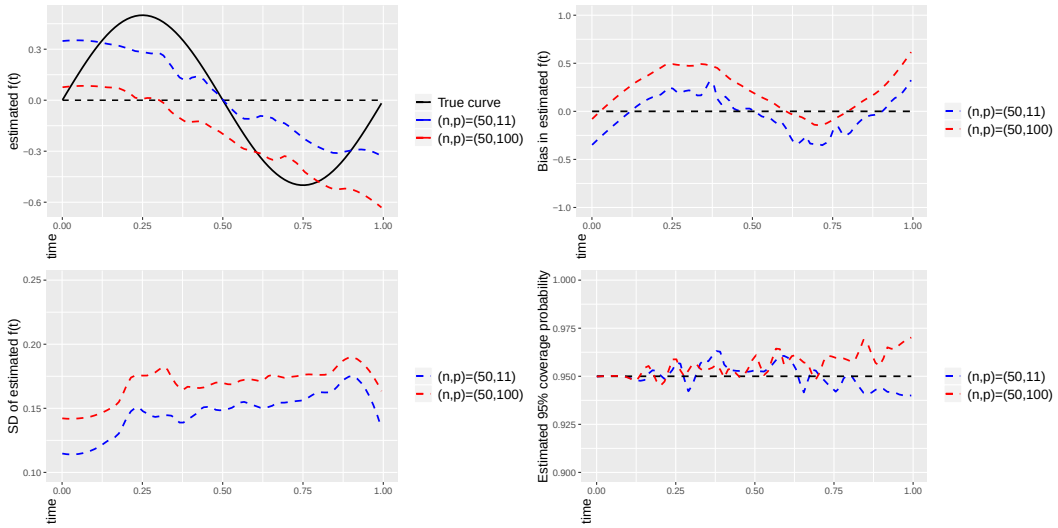
method	حالت $p_n < n$						حالت $p_n > n$					
	$(n,p)=(۵۰,۱۱)$						$(n,p)=(۵۰,۱۰۰)$					
	MSE	$C(۸)$	$I(۰)$	UF	CF	OF	MSE	$C(۹۷)$	$I(۰)$	UF	CF	OF
$GSMM$	۰٫۱۱۶	۰٫۰۹	۰٫۰۰	۰٫۰۰	۰٫۰۰	۱٫۰۰	۶۸٫۰۲۸	۰٫۰۷۴	۰٫۰۰	۰٫۰۰	۰٫۰۰	۱٫۰۰
$P-GLMM$	۰٫۰۶۰	۶٫۵۴	۰٫۰۰	۰٫۰۰	۰٫۱۳	۰٫۸۷	۰٫۴۳۵	۹۶٫۴۸	۰٫۰۰	۰٫۰۰	۰٫۵۵	۰٫۴۵
$P-GSMM$	۰٫۰۵۲	۷٫۵۹	۰٫۰۰	۰٫۰۰	۰٫۶۴	۰٫۳۶	۰٫۳۹۱	۹۶٫۴۱	۰٫۰۰	۰٫۰۰	۰٫۶۰	۰٫۴۰
	$(n,p)=(۱۰۰,۱۴)$						$(n,p)=(۱۰۰,۵۰۰)$					
	MSE	$C(۱۱)$	$I(۰)$	UF	CF	OF	MSE	$C(۴۹۷)$	$I(۰)$	UF	CF	OF
$GSMM$	۰٫۰۷۲	۰٫۱۶	۰٫۰۰	۰٫۰۰	۰٫۰۰	۱٫۰۰	۱۴۹۹٫۱۳۶	۴۷٫۰۲	۰٫۰۳	۰٫۰۳	۰٫۰۰	۱٫۰۰
$P-GLMM$	۰٫۰۴۱	۱۰٫۵۲	۰٫۰۰	۰٫۰۰	۰٫۷۷	۰٫۲۳	۰٫۰۶۲	۴۹۵٫۷۲۰	۰٫۰۰	۰٫۰۰	۰٫۸۹	۰٫۱۱
$P-GSMM$	۰٫۰۳۶	۱۰٫۷۰	۰٫۰۰	۰٫۰۰	۰٫۹۳	۰٫۰۷	۰٫۰۳۸	۴۹۶٫۲۵۰	۰٫۰۰	۰٫۰۰	۰٫۹۲	۰٫۰۸
	$(n,p)=(۱۵۰,۱۵)$						$(n,p)=(۲۰۰,۲۰۰۰)$					
	MSE	$C(۱۲)$	$I(۰)$	UF	CF	OF	MSE	$C(۱۹۹۷)$	$I(۰)$	UF	CF	OF
$GSMM$	۰٫۰۶۰	۰٫۲۶	۰٫۰۰	۰٫۰۰	۰٫۰۰	۱٫۰۰	۱۲۵٫۴۰۶	۱۱۳۷٫۶۲	۰٫۰۰	۰٫۰۰	۰٫۰۰	۱٫۰۰
$P-GLMM$	۰٫۰۴۴	۱۱٫۲۵	۰٫۰۰	۰٫۰۰	۰٫۹۲	۰٫۰۸	۰٫۰۱۸	۱۹۹۶٫۸۹	۰٫۰۰	۰٫۰۰	۰٫۳۰	۰٫۷۰
$P-GSMM$	۰٫۰۴۵	۱۱٫۸۷	۰٫۰۰	۰٫۰۰	۰٫۹۶	۰٫۰۴	۰٫۰۱۸	۱۹۹۶٫۹۳	۰٫۰۰	۰٫۰۰	۰٫۵۴	۰٫۴۶

جدول ۲.۳: نتایج شبیه‌سازی: کارآیی روش $P-GSMM$ برای حالت‌های $p_n < n$ و $p_n > n$.

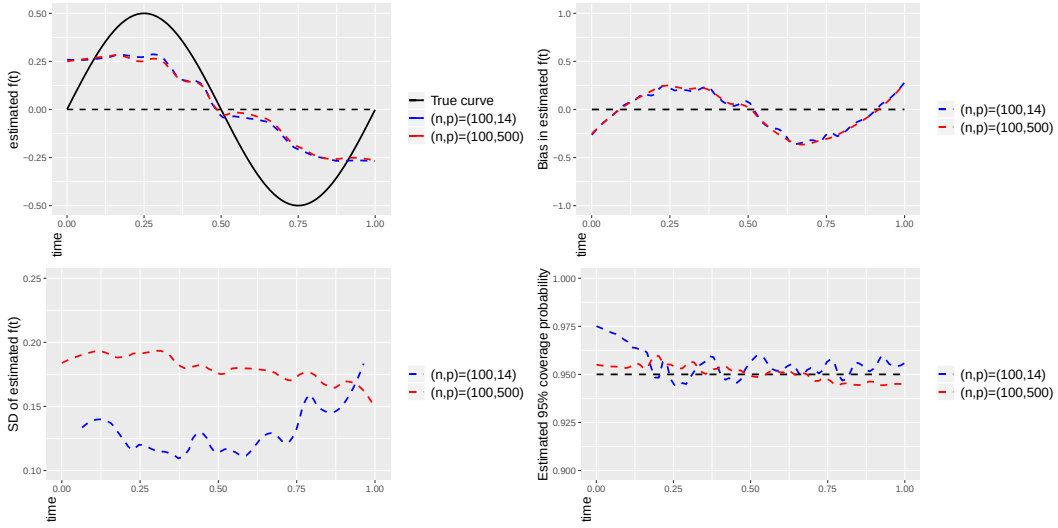
حالت $p_n < n$					حالت $p_n > n$				
(n,p_n)		$\beta_۱$	$\beta_۲$	$\beta_۳$	(n,p_n)		$\beta_۱$	$\beta_۲$	$\beta_۳$
$(۵۰,۱۱)$	$Bias$	۰٫۰۴۷	۰٫۰۹۶	۰٫۰۶۹	$(۵۰,۱۰۰)$	$Bias$	۰٫۱۴۷	۰٫۳۴۴	۰٫۳۶۵
	$SD۱$	۰٫۰۹۲	۰٫۰۸۵	۰٫۱۱۳		$SD۱$	۰٫۰۸۴	۰٫۰۷۱	۰٫۰۹۴
	$SD۲$	۰٫۰۹۷	۰٫۰۹۴	۰٫۰۹۷		$SD۲$	۰٫۱۶۲	۰٫۱۷۸	۰٫۱۴۸
	CP	۰٫۹۶	۰٫۹۵	۰٫۹۲		CP	۰٫۹۵	۰٫۹۶	۰٫۹۷
$(۱۰۰,۱۴)$	$Bias$	۰٫۰۷۶	۰٫۱۰۳	۰٫۰۵۳	$(۱۰۰,۵۰۰)$	$Bias$	۰٫۰۷۸	۰٫۰۶۱	۰٫۰۱۷
	$SD۱$	۰٫۰۷۱	۰٫۰۶۷	۰٫۰۸۴		$SD۱$	۰٫۰۴۹	۰٫۰۴۴	۰٫۰۶۶
	$SD۲$	۰٫۰۷۲	۰٫۰۶۷	۰٫۰۷۸		$SD۲$	۰٫۰۷۰	۰٫۰۷۲	۰٫۰۸۶
	CP	۰٫۹۶	۰٫۹۵	۰٫۹۶		CP	۰٫۹۴	۰٫۹۵	۰٫۹۶
$(۱۵۰,۱۵)$	$Bias$	۰٫۱۰۱	۰٫۱۲۴	۰٫۰۹۹	$(۲۰۰,۲۰۰۰)$	$Bias$	۰٫۰۴۹	۰٫۰۷۲	۰٫۰۲۸
	$SD۱$	۰٫۰۶۰	۰٫۰۵۹	۰٫۰۷۰		$SD۱$	۰٫۰۳۹	۰٫۰۴۰	۰٫۰۵۴
	$SD۲$	۰٫۰۵۲	۰٫۰۵۹	۰٫۰۵۹		$SD۲$	۰٫۰۵۱	۰٫۰۵۲	۰٫۰۵۸
	CP	۰٫۹۷	۰٫۹۶	۰٫۹۵		CP	۰٫۹۴	۰٫۹۲	۰٫۹۴

ترتیب نتایج مربوط به حجم نمونه ۵۰ و ۱۰۰ هستند. در هر دو تصویر نتایج برای حالت‌های $p_n < n$ و $p_n > n$ رسم شده‌اند. مشاهده می‌شود که منحنی‌های برآوردشده تقریبا دارای تابع سینوسی هستند و این امر در حجم نمونه بیشتر یعنی شکل ۲.۳ مشهودتر است. منحنی اریبی برآوردگرها حول خط صفر در نوسان هستند و برای حجم نمونه ۱۰۰ میزان اریبی کمتر از حجم نمونه ۵۰ است. همچنین برای حجم نمونه ۵۰ میزان اریبی حالت $p_n < n$ کمتر از $p_n > n$ است در حالی که در حجم نمونه بالا

میزان اریبی در دو حالت رفتار یکسانی را نشان می‌دهد. در هر دو شکل مقادیر انحراف استاندارد حالت $p_n < n$ کمتر از $p_n > n$ است . همچنین منحنی احتمال پوشش در هر دو شکل حول خط ۹۵٪ در نوسان است.



شکل ۱.۳: نتایج شبیه‌سازی: نتایج مربوط به $\boldsymbol{f(t)}$ برآوردشده برای حالت‌های $p_n < n$ و $p_n > n$ با حجم نمونه $n = ۵۰$. تصویر بالا-چپ: منحنی $\boldsymbol{f(t)}$ برآوردشده، تصویر بالا-راست: منحنی اریبی، تصویر پایین-چپ: منحنی انحراف استاندارد، تصویر پایین-راست: احتمال پوشش.



شکل ۲.۳: نتایج شبیه‌سازی: نتایج مربوط به $\boldsymbol{f(t)}$ برآوردشده برای حالت‌های $p_n < n$ و $p_n > n$ در مطالعه شبیه‌سازی با حجم نمونه $n = ۱۰۰$. تصویر بالا-چپ: منحنی $\boldsymbol{f(t)}$ برآوردشده، تصویر بالا-راست: منحنی اریبی، تصویر پایین-چپ: منحنی انحراف استاندارد، تصویر پایین-راست: احتمال پوشش.

۲.۴.۱.۳ مثال واقعی

تحلیل داده‌های $CD۴$

در این بخش به مطالعه رفتار برآوردگرهای پیشنهادی با استفاده از داده‌های $CD۴$ می‌پردازیم. مشابه بخش ۳.۳.۲ علاوه بر اثرات اصلی، اثرات متقابل را نیز وارد مدل کردیم. با این تفاوت که در آن بخش، تبدیل ریشه دوم متغیر پاسخ (تعداد سلول‌های $CD۴$) درنظر گرفته شد تا توزیع داده‌ها به توزیع نرمال نزدیک باشد؛ اما در اینجا بدون تبدیل متغیرهای پاسخ، مدل رگرسیون پواسن را درنظر می‌گیریم. متغیر $YEAR$ به عنوان جزء ناپارامتری و سایر متغیرها به صورت خطی وارد مدل می‌شوند. بدین ترتیب یک مدل آمیخته نیمه‌پارامتری تعمیم‌یافته به داده‌ها برازش داده می‌شود. به منظور مقایسه عملکرد روش $P-GSMM$ با دو روش دیگر یعنی $P-GLMM$ و $GSMM$ از معیار انحراف استاندارد مجانبی (SD) استفاده و برای شناسایی بهترین مدل پشتیبانی‌شده توسط داده‌ها، معیار اطلاع آکائیک (AIC) و معیار اطلاع بیزی (BIC) محاسبه می‌شود. نتایج حاصل در جدول ۳.۳ گزارش شده است. روش $P-GSMM$ مقادیر SD کمتری در مقایسه با دو روش دیگر یعنی $P-GLMM$ و $GSMM$ دارد. همچنین معیارهای AIC و BIC مدل پیشنهادی در مقایسه با دو مدل رقیب دیگر مقادیر کمتری دارند که نشان‌دهنده عملکرد بهتر آن است. تحت مدل $P-GSMM$، متغیرهای $SMOKE$، $DRUGS$، $SEXP$، $SOMKE*DRUG$ و $DRUG*SEXP$ به عنوان متغیرهای مهم شناسایی شده‌اند. انتخاب متغیر تحت مدل $P-GLMM$ کمی متفاوت بوده است و علاوه بر متغیرهای ذکرشده، متغیرهای $SMOKE*AGE$، $DRUG*AGE$، و $SMOKE*SEXP$ نیز وارد مدل شده‌اند. در این تحلیل برخی از اثرات متقابل نیز به عنوان عوامل مهم وارد مدل شده‌اند که در تحقیقات قبلی وانگ و همکاران (۲۰۰۵) و هووانگ و همکاران (۲۰۰۷) نادیده گرفته شده بودند.

نتایج مربوط به برآورد قسمت ناپارامتری مدل توسط نمودارهایی در شکل ۳.۳ ارائه شده است. این شکل شامل نمودار تابع ناپارامتری برآوردشده $\boldsymbol{f(t)}$ به همراه نمودارهای فواصل اطمینان ٪۹۵، منحنی‌های انحراف استاندارد مجانبی و نمونه‌ای، و منحنی‌های احتمال پوشش مجانبی و نمونه‌ای می‌باشند. مشاهده می‌شود که با گذشت زمان تابع ناپارامتری روند نزولی دارد. همچنین در نقاط مرزی بین انحراف استاندارد مجانبی و انحراف استاندارد نمونه‌ای اختلاف چشمگیری وجود دارد. در این نقاط انحراف استاندارد مجانبی به شدت کاهش می‌یابد، این امر باعث می‌شود که احتمالات پوشش از ٪۹۵ به ٪۹۲ تنزل یابند.

تحلیل داده‌های چرخه سلولی مخمر بیان‌ژن ORF

در این بخش به داده‌های چرخه سلولی مخمر بیان‌ژن ORF از بخش ۲.۴.۱، مدل آمیخته نیمه‌پارامتری به صورت

$$y_{ij} = \sum_{k=۱}^{۹۶} x_{ij}^{(k)} + f(t_{ij}) + b_i; \qquad i = ۱,\ldots,۲۸۳; j = ۱,\ldots,۴$$

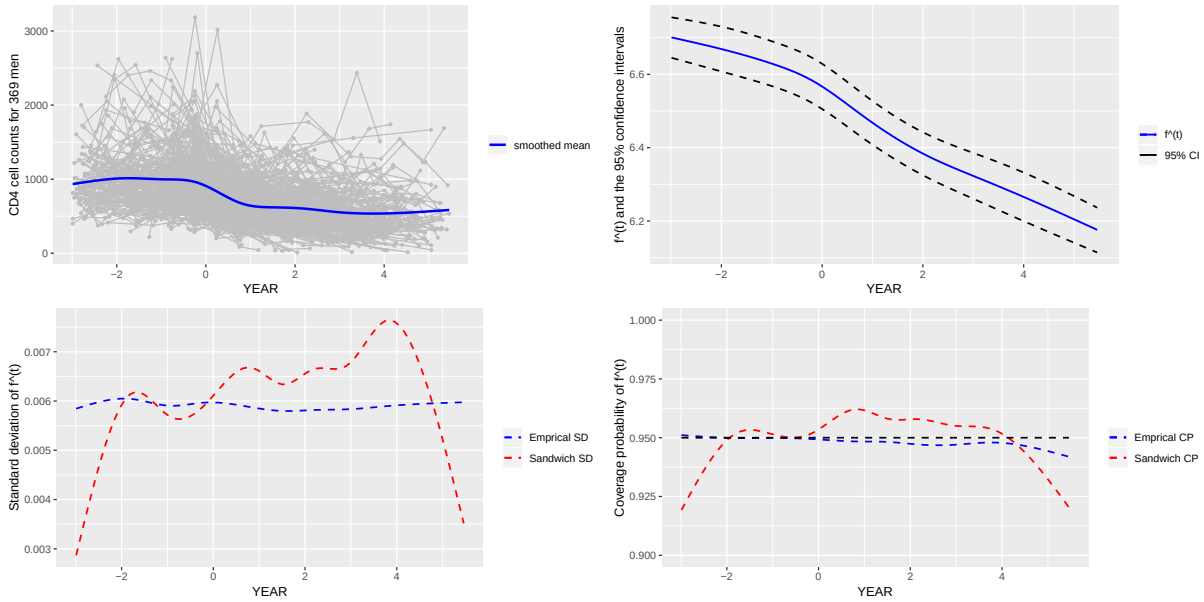
را برازش می‌دهیم. هدف شناسایی TFهایی است که ممکن است با الگوهای بیان‌ژن ۲۸۳ ژن

جدول ۳.۳: نتایج تحلیل داده‌های $CD4$: برآورد پارامترهای رگرسیونی (SD) تحت برازش سه مدل $P-GSMM$، $GSMM$ و $P-GLMM$.

متغیرها	$GSMM$ $\widehat{\beta}(SE)$	$P-GLMM$ $\widehat{\beta}(SE)$	$P-GSMM$ $\widehat{\beta}(SE)$
AGE	۰٫۰۷۳ (۰٫۰۳۹)	−۰٫۰۹۲ (۰٫۰۵۱)	۰ (۰)
$SMOKE$	۰٫۱۸۸ (۰٫۱۷۹)	۰٫۸۸۸ (۰٫۱۹۲)	۰٫۰۷۹ (۰٫۰۴۵)
$DRUG$	۰٫۱۳۰ (۰٫۱۴۳)	۶٫۰۶۸(۰٫۱۲۵)	۰٫۱۴۲ (۰٫۰۷۴)
$SEXP$	−۰٫۰۴۹ (۰٫۰۳۱)	۰٫۶۷۲ (۰٫۰۳۰)	۰٫۰۱۷ (۰٫۰۱۲)
$CESD$	−۰٫۰۰۱ (۰٫۰۱۱)	۰ (۰)	۰ (۰)
$AGE*SMOKE$	۰٫۰۰۲ (۰٫۰۱۴)	۰٫۰۱۴ (۰٫۰۰۴)	۰ (۰)
$AGE*DRUG$	−۰٫۰۳۴ (۰٫۰۲۴)	۰٫۰۳۲ (۰٫۰۳۵)	۰ (۰)
$AGE*SEXP$	−۰٫۰۰۹ (۰٫۰۰۳)	۰ (۰)	۰ (۰)
$AGE*CESD$	۰٫۰۰۱ (۰٫۰۰۲)	۰ (۰)	۰ (۰)
$SMOKE*DRUG$	۰٫۰۰۹ (۰٫۰۵۴)	−۰٫۵۸۴ (۰٫۱۵۰)	−۰٫۰۱۴ (۰٫۰۳۸)
$SMOKE*SEXP$	−۰٫۰۱۰ (۰٫۰۱۲)	−۰٫۰۳۴ (۰٫۰۱۰)	۰ (۰)
$SMOKE*CESD$	−۰٫۰۰۶ (۰٫۰۰۹)	۰ (۰)	۰ (۰)
$DRUG*SEXP$	−۰٫۰۲۵ (۰٫۰۱۹)	−۰٫۵۹۸ (۰٫۰۴۱)	−۰٫۰۲۲ (۰٫۰۱۲)
$DRUG*CESD$	۰٫۰۰۶ (۰٫۰۰۶)	۰ (۰)	۰ (۰)
$SEXP*CESD$	۰٫۰۰۱ (۰٫۰۰۳)	۰ (۰)	۰ (۰)
$\ell_{\max}$	۸۴۶۳۰۰۷	۷۵۲۹۱۵۸	۸۶۲۴۴۲۹
AIC	−۱۶۹۲۵۹۸۳	−۱۵۰۵۸۲۸۶	−۱۷۲۴۸۸۲۷
BIC	−۱۶۹۲۵۹۲۴	−۱۵۰۵۸۲۲۸	−۱۷۲۴۸۷۶۹

تنظیم‌شده از چرخه سلولی مرتبط باشد. بنابراین یک روش تاوانیده را بر اساس مدل $P-GSMM$ به کار می‌بریم. همچنین با نادیده گرفتن مولفه ناپارامتری $f(t_{ij})$، مدل $P-GLMM$ را نیز استفاده می‌کنیم.

جدول ۴.۳، خلاصه‌ای از نتایج مربوط به TFهای برگزیده‌شده بر اساس مدل‌های $P-GSMM$ و $P-GLMM$ است که نشان می‌دهد در مجموع به تعداد ۱۳ و ۱۶ تا از TFها مربوط به فرآیند چرخه سلولی مخمر به ترتیب توسط $P-GSMM$ و $P-GLMM$ شناسایی می‌شوند. تعداد TFهای انتخاب‌شده که در دو مدل مشترک باشند کم بوده و شامل $GAT3$، $MBP1$، $MSN4$، $PHD1$،



شکل ۳.۳: نتایج تحلیل داده‌های $CD۴$: نتایج مربوط به $\boldsymbol{f}(t)$ برآوردشده تحت مدل $P-GSMM$. تصویر بالا-چپ: نمودار پراکنش متغیر پاسخ، تصویر بالا-راست: $\hat{f}(t)$ (آبی) و فواصل اطمینان (مشکی)، تصویر پایین-چپ: منحنی‌های انحراف استاندارد مجانبی (آبی) و نمونه‌ای (قرمز)، تصویر پایین-راست: منحنی احتمالات پوشش مجانبی (آبی) و نمونه‌ای (قرمز).

جدول ۴.۳: نتایج تحلیل داده‌های چرخه سلولی مخمر ORF: برآورد پارامترهای رگرسیونی (SD) تحت برازش سه مدل $P-GSMM$، $GSMM$ و $P-GLMM$.

متغیرها	$P-GLMM$ $\hat{\beta}(SE)$	$P-GSMM$ $\hat{\beta}(SE)$	متغیرها ادامه	$P-GLMM$ $\hat{\beta}(SE)$	$P-GSMM$ $\hat{\beta}(SE)$
$ARG۸۱$	۰٫۰۲۲ (۰٫۰۱۹)	۰(۰)	$PHD۱$	۰٫۰۶۵ (۰٫۰۲۷)	−۰٫۰۱۹ (۰٫۰۰۶)
$DOT۶$	۰٫۰۱۸ (۰٫۰۱۷)	۰(۰)	$RAP۱$	۰٫۰۵۳ (۰٫۰۲۷)	۰ (۰)
$FKH۱$	۰ (۰)	۰٫۰۰۳ (۰٫۰۰۵)	$RGM۱$	۰ (۰)	−۰٫۰۲۲ (۰٫۰۱۳)
$FKH۲$	۰ (۰)	۰٫۱۶۶ (۰٫۰۰۸)	$RLM۱$	۰ (۰)	−۰٫۰۰۲ (۰٫۰۰۴)
$GAT۱$	−۰٫۰۰۳ (۰٫۰۰۷)	۰ (۰)	$RME۱$	۰٫۰۷۲ (۰٫۰۲۸)	۰ (۰)
$GAT۳$	۰٫۰۱۲ (۰٫۰۱۴)	−۰٫۰۲۲۳ (۰٫۰۱۲)	$SMP۱$	۰٫۰۴۵ (۰٫۰۲۴)	−۰٫۰۱۵ (۰٫۰۰۶)
$MBP۱$	۰٫۱۴۷ (۰٫۰۳۵)	−۰٫۱۴۷۷ (۰٫۰۰۷)	$STB۱$	۰ (۰)	−۰٫۰۰۸ (۰٫۰۰۵)
$MIG۱$	−۰٫۰۰۳ (۰٫۰۰۷)	۰ (۰)	$STP۱$	۰٫۰۰۲ (۰٫۰۰۵)	۰ (۰)
$MSN۴$	۰٫۰۶۰ (۰٫۰۲۷)	−۰٫۰۰۸ (۰٫۰۰۶)	$SWI۴$	۰٫۰۷۶ (۰٫۰۳۰)	−۰٫۰۰۷ (۰٫۰۰۶)
$NDD۱$	۰ (۰)	۰٫۰۸۴ (۰٫۰۰۸)	$SWI۶$	۰٫۱۱۵۱ (۰٫۰۳۴)	−۰٫۰۲۰ (۰٫۰۰۷)
$PDR۱$	۰٫۰۲۲۸ (۰٫۰۱۷)	۰ (۰)	$YAP۵$	۰٫۰۰۷ (۰٫۰۱۱)	۰ (۰)
$\ell_{\max}$	$-۵٫۷۱ \times ۱۰^{۱۴}$	$-۱٫۵۸ \times ۱۰^{۱۴}$			
AIC	$۱٫۱۴ \times ۱۰^{۱۵}$	$۳٫۱۷ \times ۱۰^{۱۴}$			
BIC	$۱٫۱۴ \times ۱۰^{۱۵}$	$۳٫۱۷ \times ۱۰^{۱۴}$			

$SMP۱$، $SWI۴$، $SWI۶$ و $SWI۶$ است. در آزمایشات بیولوژیکی پروتئین‌های $MBP۱$، $SWI۴$ و $SWI۶$ به عنوان سه TF مهم در مرحله $G۱$ از چرخه سلولی شناخته می‌شوند که این TFها در هر دو مدل وارد شده‌اند. با این حال، معیارهای انتخاب مدل، از جمله مقادیر خطاهای استاندارد و معیارهای AIC و BIC عملکرد بهتر مدل پیشنهادی ما را تأیید می کنند.

۲.۳ انتخاب متغیر برای داده‌های با بعد پایین

فرآیند انتخاب متغیر در مدل آمیخته نیمه‌پارامتری تعمیم‌یافته برای داده‌های طولی غیرنرمال با بعد پایین حالت خاصی از بعد بالا است، با این تفاوت که بعد مدل p_n ثابت است و با افزایش حجم نمونه تغییر نمی‌کند. لذا در این بخش تنها به بیان قضایای حدی مربوط به آن می‌پردازیم. تمام شرایط نظم بیان‌شده در بخش ۳.۱.۳ در این‌جا نیز برقرار است با این تفاوت که تعداد کل متغیرها p_n و تعداد متغیرهای مهم s_n ثابت بوده و به p و s تغییر می‌یابند.

قضیه ۱.۲.۳. تحت شرایط بیان‌شده در بخش ۳.۱.۳، اگر تعداد گره‌ها به صورت $L_n = O_p(n^{۱/(۲r+۱)})$ باشد، آن‌گاه به ازای هر $\boldsymbol{\xi}_n \in \mathcal{R}^p$ که $\|\boldsymbol{\xi}_n\| = ۱$، داریم

$$\mathbb{P}(\widehat{\boldsymbol{\beta}}_{n۲} = ۰) \to ۱,$$

$$\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{*^{-۱/۲}}(\boldsymbol{\beta}_{n۰})\overline{\boldsymbol{H}}_n^*(\boldsymbol{\beta}_{n۰})(\widehat{\boldsymbol{\beta}}_{n۱} - \boldsymbol{\beta}_{n۰۱}) \xrightarrow{d} \mathcal{N}(۰, ۱),$$

که در آن‌ها

$$\begin{aligned}
\overline{\boldsymbol{M}}_n^* &= \mathrm{E}_{\boldsymbol{u}|\boldsymbol{y}}\Big[\sum_{i=۱}^{n} \boldsymbol{X}_i^{*\top}\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{R}_۰\overline{\boldsymbol{R}}^{-۱}\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{X}_i^*\Big],\\
\overline{\boldsymbol{H}}_n^* &= \mathrm{E}_{\boldsymbol{u}|\boldsymbol{y}}\Big[\sum_{i=۱}^{n} \boldsymbol{X}_i^{*\top}\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{X}_i^*\Big],\\
\boldsymbol{X}_i^* &= (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{X}_i\\
\boldsymbol{P} &= \boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{\Omega}\boldsymbol{B})^{-۱}\boldsymbol{B}^\top\boldsymbol{\Omega}\\
\boldsymbol{\Omega} &= \mathrm{diag}\{\boldsymbol{\Omega}_i\}\\
\boldsymbol{\Omega}_i &= \mathrm{E}_{\boldsymbol{u}|\boldsymbol{y}}\Big[\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\Big].
\end{aligned}$$

فصل ۴

انتخاب متغیر برای داده‌های طولی چندمتغیره

در تحقیقات میدانی معمولاً چند متغیر پاسخ به طور همزمان اندازه‌گیری می‌شوند که این متغیرها با یکدیگر همبستگی داشته و داده‌های حاصل را داده‌های چندمتغیره می‌نامند. در مطالعات طولی نیز داشتن داده‌هایی با بیش از یک متغیر پاسخ که به طور مکرر در یک موضوع خاص و برای یک دوره زمانی خاص اندازه‌گیری شوند، منجر به داده‌های طولی چندمتغیره می‌شود. روش‌های مدل‌سازی ویژه‌ای برای رویارویی با داده‌های طولی چندمتغیره توسعه یافته‌اند. در ابتدا شاه و همکاران (۱۹۹۷) مدل آمیخته خطی چندمتغیره[1] ($MLMM$) را معرفی نمود و برای برآورد پارامترها از الگوریتم EM استفاده کرد. سپس ساموئل و همکاران (۱۹۹۹)، سونگ وهمکاران (۲۰۰۲)، روی (۲۰۰۶) و وانگ و فن (۲۰۱۰) به تحقیقات بیشتر پیرامون این مدل پرداختند. در تحقیقات ذکرشده با پذیرش نرمال بودن اثرات آمیخته و خطاهای مدل به ارائه راه‌حلی برای برآورد پارامترهای مدل و اثرات آمیخته با بیشینه کردن تابع درستنمایی پرداخته شده است. در عمل به دلیل غیر قابل مشاهده بودن اثرات آمیخته، بررسی نرمال بودن آن‌ها مقدور نیست و پذیرش ناصحیح این فرض می‌تواند بر روی دقت برآورد پارامترها و پیشگوها تاثیر سوء داشته باشد. اگرچه بررسی هر توزیع دیگری نیز امکان‌پذیر نیست اما اگر مطالعاتی داشته باشیم که برای اثرات آمیخته توزیع نرمال درنظر گرفته شده باشد، می‌توان از برخی توزیع‌های رقیب برای بهبود نتایج مدل نرمال، استفاده کرد. اما مهم‌تر هنگامی

[1] Multivariate linear mixed model

که داده‌ها دارای مقادیر پرت باشند، پذیرش توزیع نرمال برای خطاهای مدل می‌تواند گمراه‌کننده باشد. برای مدل‌سازی تنومند داده‌های طولی در حضور مقادیر پرت یا مشاهدات غیرمعمول، اتخاذ یک رده انعطاف‌پذیرتر از توزیع‌ها به عنوان جایگزینی برای توزیع نرمال پیشنهاد شده است. در این راستا پینهیرو و همکاران (۲۰۰۱) با درنظر گرفتن توزیع t برای اثرات تصادفی و خطاهای مدل، به تعمیم LMM پرداخته و مدلی تحت عنوان مدل آمیخته خطی t ($tMLMM$) را پیشنهاد دادند. توسعه بیشتر این روش توسط روزا و همکاران (۲۰۰۴)، لین و لی (۲۰۰۶)، لین و لی (۲۰۰۷) و سونگ و همکاران (۲۰۰۷) انجام گرفته است. با این حال روش آن‌ها به داده‌هایی با یک متغیر پاسخ محدود می‌شود. برای استنباط تنومند در برابر مقادیر پرت برای داده‌های طولی چند متغیره، وانگ و فن (۲۰۱۱) نسخه چندمتغیره مدل $tLMM$ را تحت عنوان مدل آمیخته خطی چندمتغیره ($MtLMM$) گسترش دادند. مقالات وانگ (۲۰۱۳)، وانگ و فن (۲۰۱۲) و وانگ و لین (۲۰۱۵) از جمله تحقیقات جامعی هستند که استراتژی‌های محاسباتی و برنامه‌های کاربردی $MtLMM$ را پوشش می‌دهند.

علیرغم سازگاری $MtLMM$ها با داده‌های چندمتغیره و تنومند بودن در مقابل مقادیر پرت، این مدل‌ها تنها محدود به رابطه خطی بین متغیرهای پاسخ و متغیرهای تبیینی هستند. اخیرا وانگ و لین (۲۰۱۴) و وانگ و کاسترو (۲۰۱۸) به مطالعه مدل آمیخته غیرخطی t چندمتغیره ($MtNLMM$) پرداخته‌اند که تعمیم مدل آمیخته غیرخطی چندمتغیره مارشال و همکاران (۲۰۰۶) به شمار می‌آید. اگرچه $MtNLMM$ مانند $MNLMM$ می‌تواند طیف گسترده‌تری از کاربردهای عملی را نسبت به مدل‌های خطی فراهم کند، اما به طور قابل توجهی پیچیده‌تر و از نظر محاسباتی سنگین‌تر هستند.

تا کنون استفاده از مدل‌های نیمه‌پارامتری برای داده‌های طولی چندمتغیره و مسئله انتخاب متغیر در $MtLMM$ها نسبتاً نادر بوده است. بر این اساس، در این فصل ابتدا به مسئله انتخاب متغیر در $MtLMM$ها پرداخته می‌شود. سپس نسخه نیمه‌پارامتری $MtLMM$ تحت عنوان مدل آمیخته نیمه‌پارامتری چندمتغیره ($MtSMM$) معرفی می‌شود.

۱.۴ انتخاب متغیر در مدل آمیخته با توزیع t چندمتغیره

۱.۱.۴ معرفی $MtLMM$

n نمونه تصادفی از داده‌های طولی را در نظر بگیرید، به‌طوری‌که نمونه iام دارای r متغیر پاسخ و s_i مشاهده مکرر در طول زمان باشد. فرض کنید

$$\boldsymbol{Y}_i = \begin{pmatrix} y_{i1,1} & y_{i2,1} & \cdots & y_{ir,1} \\ y_{i1,2} & y_{i2,2} & \cdots & y_{ir,2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{i1,s_1} & y_{i2,s_2} & \cdots & y_{ir,s_r} \end{pmatrix} = [\boldsymbol{y}_{i1} : \ldots : \boldsymbol{y}_{ir}] = [\boldsymbol{y}_{i,1}^\top : \ldots : \boldsymbol{y}_{i,s_i}^\top]^\top$$

ماتریس متغیرهای پاسخ $s_i \times r$ بعدی برای نمونه iام به ازای $i = ۱,\ldots,n$ باشد، به‌طوری‌که $\boldsymbol{y}_{ij} = (y_{ij,۱},\ldots,y_{ij,s_i})^\top$ برداری ستونی از مشاهدات مکرر متغیر پاسخ jام، $j = ۱,\ldots,r$، در طول زمان است. $\boldsymbol{y}_{i,k} = (y_{i۱,k},\ldots,y_{ir,k})$ بردار سطری از متغیرهای پاسخ مشاهده‌شده در موقعیت زمانی kام، $k = ۱,\ldots,s_i$، است. همچنین فرض کنید $\boldsymbol{E}_i = [\boldsymbol{e}_{i۱} : \ldots : \boldsymbol{e}_{ir}] = [\boldsymbol{e}_{i,۱}^\top : \ldots : \boldsymbol{e}_{i,s_i}^\top]^\top$ ماتریس $s_i \times r$ بعدی از خطاهای درون-گروهی، $\boldsymbol{e}_{ij} = (e_{ij,۱},\ldots,e_{ij,n_i})^\top$ بردار ستونی خطاهای هم‌ارز با $\boldsymbol{y}_{ij}$ و $\boldsymbol{e}_{i,k} = (e_{i۱,k},\ldots,e_{ir,k})$ بردار سطری خطاهای هم‌ازر با $\boldsymbol{y}_{i,k}$ هستند. برای سهولت در نمادگذاری از عملگر بردارساز[۲] (vec) استفاده می‌شود و بردارهای پاسخ و خطای مدل به ترتیب به‌صورت $\boldsymbol{y}_i = \mathrm{vec}(\boldsymbol{Y}_i)$ و $\boldsymbol{\varepsilon}_i = \mathrm{vec}(\boldsymbol{E}_i)$ با بعد $n_i = s_i r$ تعریف می‌شوند.

برای نمونه iام $MtLMM$ به‌صورت

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \qquad (۱.۴)$$

بیان می‌شود و فرض می‌کنیم $\boldsymbol{\varepsilon}_i$ و $\boldsymbol{b}_i$ دارای توزیع توام t چندمتغیره به‌صورت زیر است:

$$\begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} \sim \mathcal{T}_{q+n_i}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_i \end{bmatrix}, \nu \right). \qquad (۲.۴)$$

در مدل (۱.۴)، $\boldsymbol{X}_i$ و $\boldsymbol{Z}_i$ به‌صورت $\boldsymbol{X}_i = \mathrm{diag}\{\boldsymbol{X}_{i۱},\ldots,\boldsymbol{X}_{ir}\}$ و $\boldsymbol{Z}_i = \mathrm{diag}\{\boldsymbol{Z}_{i۱},\ldots,\boldsymbol{Z}_{ir}\}$ تعریف می‌شوند، به‌طوری‌که $\boldsymbol{X}_{ij}$ یک ماتریس از اثرات ثابت با بعد $s_i \times p_j$ و هم‌ارز با پاسخ jام از نمونه i است و $\boldsymbol{Z}_{ij}$ که معمولاً توسط زیرمجموعه‌ای از $\boldsymbol{X}_{ij}$ها ساخته می‌شود، یک ماتریس طرح $s_i \times q_j$ بعدی مربوط به اثرات تصادفی است. ساختار قطری بلوکی $\boldsymbol{X}_i$ و $\boldsymbol{Z}_i$ به واسطه تعیین ماتریس طرح مجزا برای هر متغیر پاسخ، به تحلیل‌گران اجازه می‌دهد که به راحتی روابط بین متغیرهای پاسخ و تبیینی را که ممکن است در فواصل زمانی نابرابر جمع‌آوری شده باشند، به یکدیگر پیوند دهند. اگر بعد کلی اثرات ثابت با $p = \sum_{j=۱}^{r} p_j$ و بعد کلی اثرات تصادفی با $q = \sum_{j=۱}^{r} q_j$ مشخص شوند، آنگاه $\boldsymbol{\beta} = (\boldsymbol{\beta}_۱^\top,\ldots,\boldsymbol{\beta}_r^\top)^\top$ یک بردار $p \times ۱$ بعدی از اثرات ثابت است و $\boldsymbol{\beta}_j$ زیربرداری $p_j \times ۱$ بعدی است که برای توصیف مشخصات پاسخ jام مورد استفاده قرار می‌گیرد. $\boldsymbol{b}_i = (\boldsymbol{b}_{i۱}^\top,\ldots,\boldsymbol{b}_{ir}^\top)^\top$ یک بردار $q \times ۱$ بعدی از اثرات تصادفی مشاهده نشده است و $\boldsymbol{b}_{ij}$ زیربرداری $q_j \times ۱$ بعدی هم‌ارز با $\boldsymbol{y}_{ij}$ است. در فرضیات توزیعی رابطه (۲.۴)، $\boldsymbol{D}$ و $\boldsymbol{R}_i$ به ترتیب ماتریس‌های کوواریانس مربوط به اثرات تصادفی و خطاهای مدل هستند. درجه آزادی ν یک پارامتر تنظیم‌کننده است که میزان سنگینی دم توزیع را تعیین می‌کند. علاوه بر این، $\boldsymbol{D} = [D_{jj'}]$ یک ماتریس $q \times q$ بعدی متقارن و معین مثبت با عناصر $D_{jj'}$ است. به ازای $j = j'$، D_{jj} ساختار کوواریانس اثرات تصادفی برای پاسخ jام است و به ازای $j \neq j'$، $D_{jj'}$ ساختار کوواریانس اثرات تصادفی برای جفت متغیر پاسخ j و j' است. به منظور شناسایی بهتر ماتریس کوواریانس $\boldsymbol{R}_i$، فرض کنید به ازای $j = ۱,\ldots,r$، $\boldsymbol{e}_{ij} \sim \mathcal{T}_{s_i}(۰, \sigma_{jj}\boldsymbol{\Omega}_i, \nu)$ و به ازای $k = ۱,\ldots,s_i$، داشته باشیم $\boldsymbol{e}_{i,k} \sim \mathcal{T}_r(۰, \boldsymbol{\Sigma}, \nu)$، داشته باشیم به‌طوری‌که $\boldsymbol{\Sigma} = [\sigma_{jj'}]$ واریانس‌ها و کوواریانس‌های بین r متغیر پاسخ را توصیف می‌کند و $\boldsymbol{\Omega}_i$

[۲] Vectorial operation

ماتریس همبستگی وابسته به موقعیت‌های زمانی است که همبستگی سریالی بین s_i موقعیت زمانی را نشان می‌دهد. بر این اساس، ماتریس خطای درون-گروهی $\boldsymbol{E}_i$، دارای توزیع t ماتریسی کیبریا (۲۰۰۶) بوده و بردار خطای ε_i دارای توزیع t چندمتغیره با پارامتر مکان ۰ و پارامتر مقیاس R_i است که توسط ساختار ضرب کرونکر[3] (KP) به‌صورت $\boldsymbol{R}_i = \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}_i$ تعریف می‌شود. ساختار KP کمک می‌کند تا R_i را دقیق‌تر برآورد کنیم. برای غلبه بر مسئله عدم شناساپذیری ناشی از جواب‌های غیریکتا برای $\boldsymbol{\Sigma}$ و $\boldsymbol{\Omega}_i$ در برآورد $\boldsymbol{R}_i$ با ساختار KP، لازم است که $\boldsymbol{\Omega}_i$ را یک ماتریس همبستگی بدانیم نه یک ماتریس کوواریانس. برای این که $\boldsymbol{\Omega}_i$ دقیق‌تر برآورد شود، می‌توان بر اساس ویژگی‌های داده‌های موجود یک ساختار پارامتری برای آن در نظر گرفت که تابعی از پارامترهای همبستگی ϕ و نقاط زمانی $\boldsymbol{t}_i$ به‌صورت $\boldsymbol{\Omega}_i = \boldsymbol{\Omega}_i(\phi, \boldsymbol{t}_i)$ باشد. تحت مدل (۱.۴) به راحتی می‌توان نتیجه گرفت که $\boldsymbol{y}_i \sim \mathcal{T}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Lambda}_i, \nu)$ و $\boldsymbol{\Lambda}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^\top + \boldsymbol{R}_i$.

به تبعیت از وانگ و فن (۲۰۱۱)، مجموعه‌ای از پارامترهای مقیاس τ_i را بکار گرفته و به کمک آن‌ها مدل $MtLMM$ در رابطه (۱.۴)را توسط یک مدل سلسله مراتبی به‌صورت

$$\begin{aligned} \boldsymbol{y}_i|(\boldsymbol{b}_i, \tau_i) &\sim \mathcal{N}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \tau_i^{-1}\boldsymbol{R}_i), \\ \boldsymbol{b}_i|\tau_i &\sim \mathcal{N}_q(0, \tau_i^{-1}\boldsymbol{D}), \\ \tau_i &\sim \mathcal{G}(\nu/2, \nu/2). \end{aligned} \tag{۳.۴}$$

بیان می‌کنیم. بر اساس فرضیات رابطه (۲.۴)، $\boldsymbol{b}_i|\tau_i$ و $\varepsilon_i|\tau_i$ از هم مستقل‌اند. با انتگرال‌گیری مدل (۳.۴) نسبت به $\boldsymbol{b}_i$، مدل را می‌توان به‌صورت

$$\boldsymbol{y}_i|\tau_i \sim \mathcal{N}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \tau_i^{-1}\boldsymbol{\Lambda}_i), \qquad \tau_i \sim \mathcal{G}(\nu/2, \nu/2). \tag{۴.۴}$$

نوشت. حال از طریق بهینه‌سازی تابع هدف مبتنی بر تابع تاوان مانند PML به برآورد و انتخاب متغیر می‌پردازیم.

۲.۱.۴ الگوریتم ECM برای PML

در بخش ۱.۳.۲ از فصل ۲، توضیحات مختصری پیرامون برآورد ML پارامترها و الگوریتم EM ارائه شد. در این بخش نیز برای برآورد پارامترهای مدل از روش مذکور استفاده می‌شود. اما در مدل $MtLMM$، گام M الگوریتم EM از نظر محاسباتی پیچیده است و صورت بسته‌ای برای برآوردگرها نمی‌توان یافت. برای حل این مسئله، الگوریتم دیگری به نام الگوریتم بیشینه‌سازی امید ریاضی شرطی[4] (ECM) استفاده می‌شود. الگوریتم ECM که توسط مینگ و رابین (۱۹۹۳) معرفی شد مانند الگوریتم EM است با این تفاوت که گام M آن توسط گام بیشینه‌سازی شرطی (CM) که از نظر محاسباتی ساده‌تر است، جایگزین می‌شود.

[3] Kronecker product

[4] Expectation conditional maximization

۱.۲.۱.۴ برآوردگر بیشینه درستنمایی

فرض کنید $\boldsymbol{\Theta} = \Big(\boldsymbol{\beta}, \text{vech}(\boldsymbol{D}), \text{vech}(\boldsymbol{\Sigma}), \phi, \nu\Big)$ مجموعه‌ای از تمام پارامترهای مدل باشد. اگر اثرات تصادفی $\boldsymbol{b}_i$ را به عنوان مقادیر گمشده و مجموعه $\{(\boldsymbol{y}_i, \boldsymbol{b}_i), i = ۱, \ldots, n\}$ را داده‌های کامل در نظر بگیریم، آنگاه لگاریتم تابع درستنمایی کامل بدون در نظر گرفتن جمله ثابت به‌صورت

$$\begin{aligned}\ell_c(\boldsymbol{\Theta}) = \sum_{i=۱}^{n} \ell_c^{[i]}(\boldsymbol{\Theta}) = \sum_{i=۱}^{n} \frac{۱}{۲}\Big\{ &\log|\boldsymbol{R}_i^{-۱}| + \log|\boldsymbol{D}^{-۱}| - \tau_i[\boldsymbol{\varepsilon}_i^\top \boldsymbol{R}_i^{-۱}\boldsymbol{\varepsilon}_i \\ &+ \boldsymbol{b}_i^\top \boldsymbol{D}^{-۱}\boldsymbol{b}_i] + \nu\log(\tfrac{\nu}{۲}) - ۲\log\tfrac{\nu}{۲} + \nu(\log\tau_i - \tau_i)\Big\}.\end{aligned} \qquad (۵.۴)$$

است. برای ارزیابی امید شرطی (۵.۴) به شرط داده‌های مشاهده شده $\boldsymbol{y} = \{\boldsymbol{y}_i, i = ۱, \ldots, n\}$ و برآوردهای جاری $\widehat{\boldsymbol{\Theta}}^{(r)} = (\widehat{\boldsymbol{\beta}}^{(r)}, \text{vech}(\widehat{\boldsymbol{D}}^{(r)}), \text{vech}(\widehat{\boldsymbol{\Sigma}}^{(r)}), \widehat{\phi}^{(r)}, \widehat{\nu}^{(r)})$، از قضیه زیر استفاده می‌کنیم.

قضیه ۱.۱.۴. (وانگ و فن ، ۲۰۱۱) تحت مدل سلسله مراتبی (۳.۴)، داریم

$$\begin{aligned}\begin{bmatrix}\boldsymbol{y}_i \\ \boldsymbol{b}_i\end{bmatrix}|\tau_i &\sim \mathcal{N}_{n_i+q}\left(\begin{bmatrix}\boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{0}\end{bmatrix}, \tau_i^{-۱}\begin{bmatrix}\boldsymbol{\Lambda}_i & \boldsymbol{Z}_i\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{Z}_i^\top & \boldsymbol{D}\end{bmatrix}\right), \\ \boldsymbol{b}_i|(\boldsymbol{y}_i, \tau_i) &\sim \mathcal{N}_q\Big(\boldsymbol{D}\boldsymbol{Z}_i^\top\boldsymbol{\Lambda}_i^{-۱}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}), \tau_i^{-۱}(\boldsymbol{D}^{-۱} + \boldsymbol{Z}^\top\boldsymbol{R}_i^{-۱}\boldsymbol{Z}_i)^{-۱}\Big), \\ \tau_i|\boldsymbol{y}_i &\sim \mathcal{G}\Big((\nu+n_i)/۲, (\nu+\Delta_i)/۲\Big),\end{aligned}$$

که در آن $\Delta_i = \Delta_i(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{\Sigma}, \phi, \nu) = \boldsymbol{\epsilon}_i^\top\boldsymbol{\Lambda}_i^{-۱}\boldsymbol{\epsilon}_i$ فاصله ماهالانوبیس[5] بین $\boldsymbol{y}_i$ و $\boldsymbol{X}_i\boldsymbol{\beta}$ است و $\boldsymbol{\epsilon}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$.

مراحل اصلی الگوریتم EM به شرح زیر انجام می‌شود:

گام E: محاسبه تابع $Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \text{E}\left(\ell_c^{[i]}(\boldsymbol{\Theta})|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}\right)$ که مجموع $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \text{E}\left(\ell_c(\boldsymbol{\Theta})|\boldsymbol{y}, \widehat{\boldsymbol{\Theta}}^{(r)}\right)$ برای $i = ۱, \ldots, n$ است و $Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$ به‌صورت

$$\begin{aligned}Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \frac{۱}{۲}\Big\{ &\log|\boldsymbol{R}_i^{-۱}| + \log|\boldsymbol{D}^{-۱}| - \text{tr}(\boldsymbol{D}^{-۱}\widehat{\boldsymbol{B}}_i^{(r)}) - \text{tr}(\boldsymbol{R}_i^{-۱}\widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\theta})) \\ &+ \nu\big(\log(\nu/۲) + \widehat{\kappa}_i^{(r)} - \widehat{\tau}_i^{(r)}\big)\Big\} - \log(\nu/۲),\end{aligned}$$

تعریف می‌شود، به‌طوری‌که

$$\begin{aligned}\widehat{\tau}_i^{(r)} &= \text{E}(\tau_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = (\widehat{\nu}^{(r)} + n_i)/(\widehat{\nu}^{(r)} + \widehat{\Delta}_i^{(r)}), \\ \widehat{\kappa}_i^{(r)} &= \text{E}(\log\tau_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \mathcal{D}_g(\frac{\widehat{\nu}^{(r)} + n_i}{۲}) - \log(\frac{\widehat{\nu}^{(r)} + \widehat{\Delta}_i^{(r)}}{۲}), \\ \widehat{\boldsymbol{B}}_i^{(r)} &= \text{E}(\tau_i\boldsymbol{b}_i\boldsymbol{b}_i^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\tau}_i^{(r)}\widehat{\boldsymbol{b}}_i^{(r)}\widehat{\boldsymbol{b}}_i^{(r)^\top} + \widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)}, \\ \widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\beta}) &= \text{E}(\tau_i\boldsymbol{e}_i\boldsymbol{e}_i^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) \\ &= \widehat{\tau}_i^{(r)}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)})^\top + \boldsymbol{Z}_i\widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)}\boldsymbol{Z}_i^\top,\end{aligned}$$

[5] Mahalanobis distance

و

$$\begin{aligned}
\widehat{\Delta}_i^{(r)} &= (\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}^{(r)})^\top \widehat{\boldsymbol{\Lambda}}_i^{(r)^{-1}} (\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}^{(r)}),\\
\widehat{\boldsymbol{\Lambda}}_i^{(r)} &= \boldsymbol{Z}_i\widehat{\boldsymbol{D}}^{(r)}\boldsymbol{Z}_i^\top + \widehat{\boldsymbol{R}}_i^{(r)},\\
\widehat{\boldsymbol{R}}_i^{(r)} &= \widehat{\boldsymbol{\Sigma}}^{(r)} \otimes \boldsymbol{\Omega}_i(\widehat{\phi}^{(r)}),\\
\widehat{\boldsymbol{b}}_i^{(r)} &= E(\boldsymbol{b}_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\boldsymbol{D}}^{(r)}\boldsymbol{Z}_i^\top \widehat{\boldsymbol{\Lambda}}_i^{(r)^{-1}} (\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}^{(r)}),\\
\widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)} &= \widehat{\tau}_i^{(r)}\mathrm{cov}(\boldsymbol{b}_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \big(\widehat{\boldsymbol{D}}^{(r)^{-1}} + \boldsymbol{Z}_i^\top \widehat{\boldsymbol{R}}_i^{(r)^{-1}} \boldsymbol{Z}_i\big)^{-1}.
\end{aligned}$$

گام M: به روزکردن $\widehat{\boldsymbol{\Theta}}^{(r)}$ توسط $\widehat{\boldsymbol{\Theta}}^{(r+1)} = \max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$.

۲.۲.۱.۴ برآوردگر PML و الگوریتم ECM

به منظور انتخاب متغیرهای مهم و برآورد ضرایب آن‌ها، به لگاریتم تابع درستنمایی (۵.۴)، جمله تاوان $\sum_{k=1}^{p} P_{\lambda_n}(|\beta_k|)$ را اضافه می‌کنیم، آنگاه لگاریتم تابع درستنمایی تاوانیده به‌صورت

$$\ell_c^{\mathrm{Pen}}(\boldsymbol{\Theta}) = \ell_c(\boldsymbol{\Theta}) - n\sum_{k=1}^{p} P_{\lambda_n}(|\beta_k|). \qquad (۶.۴)$$

حاصل می‌شود. به طور مشابه تابع Q تاوانیده در گام E را به‌صورت

$$Q_i^{\mathrm{Pen}}(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) - n\sum_{k=1}^{p_n} P_{\lambda_n}(|\beta_k|)$$

در نظر می‌گیریم. در این‌جا از تابع تاوان اسکد استفاده می‌کنیم و برای حل مسئله بهینه سازی در گام M، تقریب درجه دو موضعی را به کار می‌بریم. بیشینه‌سازی گام M از نظر محاسباتی پیچیده است، لذا از گام CM استفاده می‌کنیم. در این گام فرض می‌شود $\widehat{\boldsymbol{e}}_{il}^{(r)} = \boldsymbol{y}_{il} - \boldsymbol{X}_{il}\boldsymbol{\theta}_l - \boldsymbol{Z}_{il}\widehat{\boldsymbol{b}}_{il}^{(r)}$ و $\widehat{\boldsymbol{e}}_{is}^{(r)} = \boldsymbol{y}_{is} - \boldsymbol{X}_{is}\boldsymbol{\theta}_s - \boldsymbol{Z}_{is}\widehat{\boldsymbol{b}}_{is}^{(r)}$، به‌طوری‌که $\widehat{\boldsymbol{b}}_{il}^{(r)}$ یک زیربردار ۱ × q_l بعدی و شامل عناصر $(\sum_{j=1}^{l-1} q_j + 1)$ام تا $(\sum_{j=1}^{l} q_j)$ام از بردار $\widehat{\boldsymbol{b}}_i^{(r)}$ است. در این حالت داریم $\widehat{\boldsymbol{\Psi}}^{(r)}(\boldsymbol{\theta}) = [\widehat{\psi}_{ils}^{(r)}(\boldsymbol{\theta})]$، که در آن $\widehat{\psi}_{ils}^{(r)}(\boldsymbol{\theta}) = E(\tau_i \boldsymbol{e}_{il}\boldsymbol{e}_{is}^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\tau}_i^{(r)}\widehat{\boldsymbol{e}}_{il}^{(r)}\widehat{\boldsymbol{e}}_{is}^{(r)^\top} - \boldsymbol{Z}_{il}\widehat{\boldsymbol{V}}_{\boldsymbol{b}_{ils}}^{(r)}\boldsymbol{Z}_{is}^\top$ یک ماتریس مربعی با بعد s_i است که در آن $\widehat{\boldsymbol{V}}_{\boldsymbol{b}_{ils}}^{(r)}$ یک زیرماتریس $q_l \times q_s$ بعدی شامل سطرهای $(\sum_{j=1}^{l-1} q_j + 1)$ام تا $(\sum_{j=1}^{l} q_j)$ام و ستون‌های $(\sum_{j=1}^{s-1} q_j + 1)$ام تا $(\sum_{j=1}^{s} q_j)$ام ماتریس $\widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)}$ به ازای $l, s = 1, \ldots, r$ است. در مجموع گام CM الگوریتم ECM به‌صورت الگوریتم ۴ پیشنهاد می‌شود.

۳.۱.۴ خواص مجانبی

در این بخش خواص مجانبی برآوردگر تاوانیده معرفی‌شده، مطالعه می‌شود. موقعیتی را در نظر بگیرید که مقادیر درست $\boldsymbol{\beta}$ به‌صورت $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ و ماتریس طرح هم‌ارز با آن به‌صورت $\boldsymbol{X}_i = (\boldsymbol{X}_{i(1)}, \boldsymbol{X}_{i(2)})$ تفکیک شوند. $\boldsymbol{\beta}_1$ بردار s بعدی است که همه مقادیر آن غیر صفراند و $\boldsymbol{\beta}_2 = \mathbf{0}$. در

الگوریتم ۴ برآوردگر PML و الگوریتم ECM

۱: **گام CM-۱:** ثابت نگه‌داشتن $\widehat{\phi}=\widehat{\phi}^{(r)}$ و $\widehat{\nu}=\widehat{\nu}^{(r)}$، به روزکردن $\widehat{\boldsymbol{\beta}}^{(r)}$، $\widehat{\boldsymbol{D}}^{(r)}$ و $\widehat{\boldsymbol{\Sigma}}^{(r)}$ از طریق بیشینه کردن $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$ که به‌صورت زیر حاصل می شوند:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(r+1)} &= \Big(\sum_{i=1}^{n}\widehat{\tau}_i^{(r)}\boldsymbol{X}_i^{\top}\widehat{\boldsymbol{R}}_i^{(r)^{-1}}\boldsymbol{X}_i+n\boldsymbol{E}_n(\widehat{\boldsymbol{\beta}}^{(r)})\Big)^{-1}\sum_{i=1}^{n}\widehat{\tau}_i^{(r)}\boldsymbol{X}_i^{\top}\widehat{\boldsymbol{R}}_i^{(r)^{-1}}(\boldsymbol{y}_i-\boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)}),\\
\widehat{\boldsymbol{D}}^{(r+1)} &= n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{B}}_i^{(r)},\\
\widehat{\sigma}_{lm}^{(r+1)} &= \begin{cases}(\sum_{i=1}^{n}s_i)^{-1}\sum_{i=1}^{n}\mathrm{tr}\Big(\boldsymbol{\Omega}_i^{-1}(\widehat{\phi}^{(r)})\widehat{\boldsymbol{\psi}}_{ils}^{(r)}(\widehat{\boldsymbol{\beta}}^{(r+1)})\Big) & \text{for } l=s,\\ (2\sum_{i=1}^{n}s_i)^{-1}\sum_{i=1}^{n}\mathrm{tr}\Big(\boldsymbol{\Omega}_i^{-1}(\widehat{\phi}^{(r)})\big(\widehat{\boldsymbol{\psi}}_{ils}^{(r)}(\widehat{\boldsymbol{\beta}}^{(r+1)})+\widehat{\boldsymbol{\psi}}_{isl}^{(r)}(\widehat{\boldsymbol{\beta}}^{(r+1)})\big)\Big) & \text{for } l\neq s,\end{cases}
\end{aligned}$$

که در آن

$$\boldsymbol{E}_n(\widehat{\boldsymbol{\theta}}^{(r)})=\mathrm{diag}\Big\{\frac{q_{\lambda_n}(|\beta_1|)}{\epsilon+|\beta_1|},\ldots,\frac{q_{\lambda_n}(|\beta_p|)}{\epsilon+|\beta_p|}\Big\},\qquad \epsilon=10^{-6}.$$

۲: **گام CM-۲:** بر اساس $\widehat{\boldsymbol{\beta}}^{(r+1)}$، $\widehat{\boldsymbol{D}}^{(r+1)}$، و $\widehat{\boldsymbol{\Sigma}}^{(r+1)}$، محاسبه شده در گام قبل، $(\widehat{\phi}^{(r+1)},\widehat{\nu}^{(r+1)})$ از طریق بیشینه کردن تابع Q به دست می‌آید. به عبارت دیگر

$$\begin{aligned}
(\widehat{\phi}^{(r+1)},\widehat{\nu}^{(r+1)})=\arg\max_{(\phi,\nu)}\Big\{\sum_{i=1}^{n}\Big(r\log|\boldsymbol{\Omega}_i^{-1}(\phi)|-\mathrm{tr}\big((\widehat{\boldsymbol{\Sigma}}^{(r+1)^{-1}}\otimes\boldsymbol{\Omega}_i^{-1}(\phi))\times\widehat{\boldsymbol{\Psi}}_i^{(r+1/2)}\big)\Big)\\
+\nu\log(\tfrac{\nu}{2})-2\log\Gamma(\tfrac{\nu}{2})+\nu(\widehat{\kappa}_i^{(r)}-\widehat{\tau}_i^{(r)})\Big\},
\end{aligned}$$

به‌طوری‌که $\widehat{\boldsymbol{\Psi}}_i^{(r+1/2)}=\widehat{\boldsymbol{\Psi}}_i^{(r)}(\widehat{\boldsymbol{\beta}}^{(r+1)})$. از آنجایی که با برابر صفر قرار دادن مشتق اول عبارت Q نسبت به ϕ و ν، صورت بسته‌ای نتیجه نمی‌دهد؛ برای به روزکردن $(\widehat{\phi}^{(r+1)},\widehat{\nu}^{(r+1)})$، باید از روش های عددی استفاده کرد که در این رساله از پکیج nlminb در نرم‌افزار R استفاده شده است.

۳: **گام همگرایی:** گام های E و CM را تا زمان همگرایی به مقداری ثابت تکرار می‌کنیم تا برآوردگر PML بدست آید.

نتیجه مقادیر برآوردگرها نیز به‌صورت $\widehat{\boldsymbol{\beta}}=(\widehat{\boldsymbol{\beta}}_1^{\top},\widehat{\boldsymbol{\beta}}_2^{\top})^{\top}$ تفکیک می‌شوند. فرض کنید $\boldsymbol{\eta}=(\boldsymbol{\omega}^{\top},\nu)^{\top}$، که در آن $\boldsymbol{\omega}=\big(vec(\boldsymbol{D})^{\top},vec(\boldsymbol{\Sigma})^{\top},\boldsymbol{\phi}^{\top}\big)^{\top}$. همچنین $\boldsymbol{J}_{\beta\beta}$، $\boldsymbol{J}_{\beta_1\beta_1}$، $\boldsymbol{J}_{\beta_2\beta_2}$ و $\boldsymbol{J}_{\eta\eta}\to\boldsymbol{I}_{\eta\eta}$ به ترتیب ماتریس‌های اطلاع فیشر مربوط به $\boldsymbol{\beta}$، $\boldsymbol{\beta}_1$، $\boldsymbol{\beta}_2$ و $\boldsymbol{\eta}$ هستند. برای مطالعه رفتار حدی برآوردگرها، شرایط نظم زیر را درنظر می‌گیریم:

(۱.A) اگر $n \to \infty$، آن‌گاه

$$n^{-1}\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} \to \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}, \quad n^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} \to \boldsymbol{I}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1}$$
$$n^{-1}\boldsymbol{J}_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2} \to \boldsymbol{I}_{\boldsymbol{\beta}_2\boldsymbol{\beta}_2}, \quad n^{-1}\boldsymbol{J}_{\boldsymbol{\eta}\boldsymbol{\eta}} \to \boldsymbol{I}_{\boldsymbol{\eta}\boldsymbol{\eta}}$$

(۲.A) به ازای بردار $\boldsymbol{a} \neq 0$ که $pr \times 1$ بعدی است داریم

$$\frac{\max_i\{\boldsymbol{a}^\top \boldsymbol{X}_i^\top \boldsymbol{\Lambda}_i^{-1}\boldsymbol{X}_i\boldsymbol{a}\}}{\boldsymbol{a}^\top\big(\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{\Lambda}_i^{-1}\boldsymbol{X}_i\big)\boldsymbol{a}} \to 0. \qquad (۷.۴)$$

(۳.A) مقادیر ثابت و مثبت C_1 و C_1 وجود دارند، به‌طوری‌که

$$C_1 < \lambda_{\min}\Big(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{\Lambda}_i^{-1}\boldsymbol{X}_i\Big) < \lambda_{\max}\Big(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{\Lambda}_i^{-1}\boldsymbol{X}_i\Big) < C_2,$$

که در آن $\lambda_{\min}(\boldsymbol{A})$ و $\lambda_{\max}(\boldsymbol{A})$ به ترتیب کوچکترین و بزرگترین مقدار ویژه ماتریس $\boldsymbol{A}$ هستند.

قضیه ۲.۱.۴. تحت مفروضات این بخش و شرایط نظم (۱.A)-(۳.A) اگر

$$\liminf_{n\to\infty} \liminf_{\lambda_n \to 0^+} \sqrt{n} P'_{\lambda_n} \to \infty,$$

آن‌گاه $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1^\top, \widehat{\boldsymbol{\beta}}_2^\top)^\top$ یک جواب الگوریتم ECM است، به‌طوری‌که

$$\mathbb{P}_n(\widehat{\boldsymbol{\beta}}_2 = 0) \to 1,$$

$$\widehat{\boldsymbol{\Theta}} \xrightarrow{p} \boldsymbol{\Theta}, \qquad \sqrt{n}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \xrightarrow{d} \mathcal{N}_s(\boldsymbol{0}, \boldsymbol{I}^{-1}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1}),$$
$$\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}^{-1}_{\boldsymbol{\eta}\boldsymbol{\eta}}).$$

قضیه ۲.۱.۴ بیان می‌کند که با احتمال نزدیک به یک، ضرایب صفر واقعی یعنی متغیرهای بی‌تاثیر شناسایی می‌شوند و ضرایب غیر صفر کارآیی بالایی دارند. علاوه بر این با مطالعه رفتار مجانبی برآوردگرهای $\widehat{\beta}$ و $\widehat{\eta}$، ابزار اولیه برای اجرای سایر استنباط‌های آماری مانند فاصله اطمینان و آزمون‌های فرض را فراهم می‌آورد. برای این منظور لازم است ماتریس کوواریانس مجموعه برآوردگرهای $\widehat{\Theta}$ شناسایی شوند. از آنجا که $\widehat{\Theta}$ سازگار و دارای توزیع نرمال مجانبی است، می‌توان معکوس ماتریس اطلاع فیشر در نقطه $\Theta = \widehat{\Theta}$ را به عنوان تقریبی از ماتریس کوواریانس مجانبی برای $\widehat{\Theta}$ درنظر گرفت. ماتریس‌های اطلاع فیشر و اثبات قضیه ۲.۱.۴ در قسمت ضمیمه آمده است.

۴.۱.۴ مطالعات عددی

در این بخش با طراحی یک مطالعه شبیه‌سازی شده و تحلیل داده‌های واقعی نشان داده می‌شود که روش معرفی‌شده علاوه بر انتخاب متغیرهای مهم به دلیل استفاده از توزیع t به جای توزیع نرمال، در مقابل داده‌های پرت تنومند است. بدین منظور عملکرد برآوردگرها تحت رویکردهای $MtLMM$ تاوانیده، $MtLMM$، $MNMM$ تاوانیده و $MNMM$ را با هم مقایسه خواهیم کرد.

۱.۴.۱.۴ مطالعه شبیه‌سازی

در این بخش با طراحی یک مطالعه شبیه‌سازی شده، نشان داده می‌شود که روش معرفی‌شده در انتخاب متغیرهای مهم از توزیع t از کارایی خوبی برخوردار است.

متغیرهای پاسخ $\boldsymbol{y}_i$ را به ازای $r = ۲$ متغیر پاسخ از مدل (۱.۴) با فرضیات (۲.۴) تولید می‌کنیم. تعداد تکرارها برای هر نمونه را ثابت و برابر $s_i = ۵$ در نظر می‌گیریم. به‌طوری‌که مولفه‌های مدل به‌صورت

$$\boldsymbol{X}_{i۱}, \boldsymbol{X}_{i۲} \sim \mathcal{U}(-۰/۵, ۰/۵)$$

$$\boldsymbol{D} = \begin{bmatrix} ۱ & ۰/۲۵ \\ ۰/۲۵ & ۱ \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} ۲ & ۰/۷۵ \\ ۰/۷۵ & ۱ \end{bmatrix}, \quad \text{و} \quad \boldsymbol{\Omega} = \boldsymbol{I}_{s_i}$$

تعیین شده‌اند. مقادیر فرضی برای پارامترهای مدل به‌صورت $\boldsymbol{\beta} = (\beta_{۱۱}, \ldots, \beta_{۱p}, \beta_{۲۱}, \ldots, \beta_{۲p})^\top$ بوده که در آن $p = ۱۰$، $(\beta_{۱۱}, \beta_{۱۲}, \beta_{۱۳}, \beta_{۲۱}, \beta_{۲۲}, \beta_{۲۳}) = (-۱, -۱, ۲, -۱, -۱, ۲)$ و سایر اعضای بردار پارامترها برابر صفر هستند. برای درنظر گرفتن داده‌هایی با دم سنگین و همچنین داده‌هایی نزدیک به توزیع نرمال، مقادیر درجه آزادی را به ترتیب $\nu = ۵$ و $\nu = ۵۰$ در نظر گرفتیم. برای بررسی خواص نمونه‌ای برآورد ML، اندازه حجم نمونه به‌صورت $n = ۲۵$ برای نمونه نسبتاً کوچک و $n = ۱۰۰$ برای نمونه نسبتاً بزرگ تعیین شده است. هر مجموعه از داده‌های شبیه‌سازی شده توسط رهیافت‌های $MtLMM$ تاوانیده، $MtLMM$، $MNMM$ تاوانیده و $MNMM$ و به ازای ترکیب‌های مختلف ν و n مورد تحلیل قرار گرفتند. به منظور سنجش دقت برآوردگرها از معیارهای AIC ، BIC و MSE بر اساس میانگین‌گیری ۱۰۰ تکرار از شبیه‌سازی، استفاده می‌شود. برای ارزیابی عملکرد روش‌ها در مساله انتخاب متغیر مشابه بخش ۱.۳.۳.۲ از معیارهای C، I، UF، CF و OF استفاده می‌کنیم. جدول ۱.۴، خلاصه‌ای از نتایج مربوط به عملکرد چهار روش ذکر شده در برآورد و انتخاب مدل برای مقادیر مختلف (n, ν) است. برای بررسی بیشتر عملکرد روش‌های رقیب $MtLMM$ تاوانیده و $MNMM$ تاوانیده، نتایج مربوط به برآورد (ESt)، SD و MSE برای هر یک از پارامترها در جدول ۲.۴ گزارش شده است.

از نظر انتخاب مدل، همان‌طور که انتظار می‌رود مشاهده می‌کنیم که $MtLMM$ و $MNMM$ انتخاب متغیر انجام نمی‌دهند. علاوه بر این، $MtLMM$ تاوانیده و $MNMM$ تاوانیده با موفقیت همه متغیرها با ضرایب غیر صفر را انتخاب می‌کنند زیرا معیار I آن‌ها صفر است. اما بدیهی است که رویکرد پیشنهادی با تفاوت بسیار کم بهتر از $MNMM$ تاوانیده است زیرا C بزرگ‌تر است. در $MtLMM$ تاوانیده، احتمال شناسایی مدل درست حداقل ۷۰٪ است و این نرخ با افزایش حجم نمونه بهبود می‌یابد. بنابراین رفتار مجانبی برآوردگرها به خوبی تایید می‌شود. از نظر دقت برآورد، $MtLMM$ تاوانیده نسبت به سایر روش‌ها به طور قابل توجهی منجر به SD و Mse کمتری شده است. این موضوع در مورد معیارهای AIC، BIC و MSE نیز صدق می‌کند. این برتری در حالت $\nu = ۵$ که دم توزیع کلفت‌تر است، یا به عبارتی دارای مقادیر پرت بیشتری است، مشهودتر است. هنگامی که توزیع داده‌ها به نرمال نزدیک‌تر است، $\nu = ۵۰$، یا حجم نمونه بزرگ است، $n = ۱۰۰$،

هر دو روش تقریباً عملکرد یکسانی داشته‌اند. همچنین دریافتیم که مقادیر SD و Mse با افزایش حجم نمونه کاهش می‌یابند، لذا رفتار مجانبی خوب برآوردگرهای ML را تائید می‌کند.

جدول ۱.۴: نتایج شبیه‌سازی: مقایسه نتایج انتخاب متغیر روش‌های $MtLMM$ تاوانیده، $MtLMM$، $MNMM$ تاوانیده و $MNMM$ به ازای nها و νهای مختلف.

$n = ۲۵$

	ν	MSE	ℓ	AIC	BIC	C	I	UF	CF	OF
$MtLMM$ تاوانیده	۵	۰/۴۵۲	−۸۹/۸۵	۲۳۳/۷۰	۲۶۶/۶۱	۱۳/۷۵	۰/۰۰	۰/۰۰	۰/۸۳	۰/۱۷
	۵۰	۰/۴۸۰	−۹۰/۱۵	۲۳۴/۳۰	۲۶۷/۲۱	۱۳/۶۶	۰/۰۰	۰/۰۰	۰/۷۳	۰/۲۷
$MtLMM$	۵	۰/۸۶۷	−۷۸/۶۹	۲۱۱/۳۹	۲۴۴/۲۹	۲/۷۵۰	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰
	۵۰	۰/۸۳۵	−۸۴/۷۷	۲۲۳/۵۵	۲۵۶/۴۶	۲/۸۹	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰
$MNMM$ تاوانیده	۵	۰/۷۱۱	−۲۴۶/۰۹	۵۴۶/۱۹	۵۷۹/۱۰	۱۳/۶۰	۰/۰۰	۰/۰۰	۰/۷۲	۰/۲۸
	۵۰	۰/۵۵۰	−۱۸۱/۵۳	۴۱۷/۰۷	۴۴۹/۹۸	۱۳/۵۷	۰/۰۰	۰/۰۰	۰/۶۹	۰/۳۱
$MNMM$	۵	۲/۵۱۰	−۲۴۰/۸۹	۵۳۵/۷۷	۵۶۸/۶۸	۱/۹۷	۰/۰۰	۰/۰۱	۰/۰۰	۱/۰۰
	۵۰	۱/۶۶۱	−۱۷۷/۱۳	۴۰۸/۲۸	۴۴۱/۱۸	۲/۱۵	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰

$n = ۱۰۰$

$MtLMM$ تاوانیده	۵	۰/۱۰۱	−۴۷۰/۷۳	۹۹۵/۴۶	۱۰۶۵/۸۰	۱۳/۹۷	۰/۰۰	۰/۰۰	۰/۹۷	۰/۰۳
	۵۰	۰/۰۸۳	−۵۲۴/۵۲	۱۱۰۳/۰۴	۱۱۷۳/۳۸	۱۳/۹۴	۰/۰۰	۰/۰۰	۰/۹۴	۰/۰۶
$MtLMM$	۵	۰/۱۸۹	−۴۹۰/۹۲	۱۰۳۵/۸۴	۱۱۰۶/۱۸	۵/۵۹	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰
	۵۰	۰/۱۸۳	−۵۴۳/۷۹۶	۱۱۴۱/۵۹۳	۱۲۱۱/۹۳۲	۵/۶۷	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰
$MNMM$ تاوانیده	۵	۰/۲۱۹	−۱۰۵۱/۱۴	۲۱۵۶/۲۹	۲۲۲۶/۶۳	۱۳/۹۱	۰/۰۱	۰/۰۱	۰/۹۰	۰/۰۹
	۵۰	۰/۰۷۹	−۷۷۵/۸۳	۱۶۰۵/۶۵	۱۶۷۵/۹۹	۱۳/۸۸	۰/۰۰	۰/۰۰	۰/۸۹	۰/۱۱
$MNMM$	۵	۰/۵۴۹	−۱۰۴۳/۰۲۸	۲۱۴۰/۰۵	۲۲۱۰/۳۹	۳/۹۴	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰
	۵۰	۰/۳۴۶	−۷۶۸/۶۰	۱۵۹۱/۲۰	۱۶۶۱/۵۴	۴/۵۹	۰/۰۰	۰/۰۰	۰/۰۰	۱/۰۰

۲.۴.۱.۴ تحلیل داده‌های بیماری صفراوی

در این بخش داده‌های بیماری صفراوی را توسط روش‌های $MtLMM$ تاوانیده، $MtLMM$، $MNMM$ تاوانیده و $MNMM$ تحلیل کرده و نتایج با هم مقایسه می‌شوند. $\boldsymbol{y}_i = (\boldsymbol{y}_{i۱}^\top, \boldsymbol{y}_{i۲}^\top)^\top$ متغیرهای پاسخ برای بیمار iام است، به‌طوری‌که $\boldsymbol{y}_{i۱}$ و $\boldsymbol{y}_{i۲}$ به ترتیب بیانگر سطوح $lbili$ و $lalbumin$ است. هدف از تحلیل این داده‌ها تعیین ارتباط بین این دو متغیر پاسخ و متغیرهای تبیینی Age، Sex و $Drug$ است. بنابراین ماتریس طرح برای اثرات ثابت $\boldsymbol{\beta} = (\beta_{۱۰}, \beta_{۱۱}, \beta_{۱۲}, \beta_{۱۳}, \beta_{۱۴}, \beta_{۲۰}, \beta_{۲۱}, \beta_{۲۲}, \beta_{۲۳}, \beta_{۲۴})$ به‌صورت

$$\boldsymbol{X}_i = \boldsymbol{I}_۲ \otimes [\boldsymbol{I}_{s_i} : \boldsymbol{t}_i : sex_i \boldsymbol{I}_{s_i} : drug_i \boldsymbol{I}_{s_i} : age_i \boldsymbol{I}_{s_i}],$$

تعیین می‌شوند، به‌طوری‌که $\boldsymbol{I}_{s_i}$ یک بردار ۱ × s_i بعدی با عناصر ۱، $\boldsymbol{t}_i = (t_{i۱}, \ldots, t_{is_i})^\top$ با $t_{i,k} = month_{i,k}/۱۲$ $(Year)$، sex_i شاخص جنسیت (۰ برای مرد ۰ و ۱ برای زن)، $drug_i$ شاخص دارو (۰ برای دارونما و ۱ برای D-پنیسیلامین) ، age_i سن بیمار iام هستند. ماتریس طرح برای عرض از مبدا و شیب تصادفی به‌صورت $\boldsymbol{Z}_i = \boldsymbol{I}_۲ \otimes [\boldsymbol{I}_{s_i} : \boldsymbol{t}_i]$ است. برای سادگی در محاسبات

فرض می‌کنیم $\boldsymbol{\Omega}_i = \boldsymbol{I}$. برای مقایسه عملکرد روش $MLMM$ تاوانیده با سه روش دیگر، خطای استاندار برآوردگرها از طریق روش اعتبار سنجی متقابل یک طرفه محاسبه می‌شود. همچنین برای شناسایی بهترین مدل پشتیبانی شده توسط داده‌ها، از معیارهای AIC، BIC و $MSPE$ (میانگین توان‌های دوم خطای پیشگویی) استفاده می‌شود. جدول ۳.۴، نتایج شامل برآورد پارامترها، خطاهای استاندارد و معیارهای AIC، BIC و $MSPE$ را تحت برازش روش‌های مذکور ارائه می‌دهد. نتایج نشان می‌دهند که پارامترها تحت روش‌های $MtLMM$ تاوانیده و $MNMM$ تاوانیده، خطاهای استاندارد کمتری نسبت به $MtLMM$ و $MNMM$ دارند. اگر چه از نظر معیار $MSPE$ تفاوت عملکرد روش‌ها چشمگیر نیست با این وجود معیارهای AIC، BIC مدل پیشنهادی در مقایسه با سه مدل دیگر مقدار کمتری دارند. براساس مدل $MtLMM$ تاوانیده، نتیجه می‌شود که $Drug$ اثر معنی‌داری روی $lbili$ ندارد و $lalbumin$ تحت تاثیر Sex و $Drug$ نیست.

۲.۴ مدل نیمه پارامتری با اثرات آمیخته با توزیع t چندمتغیره

در این بخش ابتدا به معرفی مدل آمیخته نیمه‌پارامتری t چندمتغیره ($MtSMM$) پرداخته می‌شود. مولفه ناپارامتری مدل به روش اسپلاین‌های رگرسیونی شرح داده شده در بخش ۳.۵.۱ تقریب زده می‌شود. سپس از الگوریتم تکراری ECM برای برآورد پارامترها استفاده می‌کنیم. برای بررسی کارایی مدل پیشنهادی، نتایج حاصل از تحلیل مجموعه داده شبیه‌سازی شده و یک مجموعه داده واقعی گزارش می‌شود.

۱.۲.۴ مدل $MtSMM$ و تقریب اسپلاین

n نمونه تصادفی از داده‌های طولی با مقادیر پرت را در نظر بگیرید، به‌طوری‌که نمونه iام دارای r متغیر پاسخ و s_i مشاهده مکرر در طول زمان باشد. مدل $MtSMM$ برای نمونه iام به‌صورت

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}_i) + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \qquad (۸.۴)$$

بیان می‌شود، که در آن $\boldsymbol{f}(\boldsymbol{t}_i) = \left(\boldsymbol{f}_1(\boldsymbol{t}_i)^\top, \ldots, \boldsymbol{f}_r(\boldsymbol{t}_i)^\top\right)^\top$، $\boldsymbol{g}_j(\boldsymbol{t}_i) = \left(g_j(t_{i,1}), \ldots, g_j(t_{i,s_i})\right)^\top$ و $g_j(t_{i,k})$ یک تابع هموار و مشتق‌پذیر برای پاسخ jام اندازه‌گیری شده در زمان $t_{i,k}$ است. فرضیات توزیعی مدل به‌صورت رابطه (۲.۴) است. سایر مولفه‌های مدل مشابه بخش ۱.۴ تعیین می‌شوند.

برای تقریب تابع ناپارامتری $g_j(t_{i,k})$ به روش اسپلاین‌های رگرسیونی فرض کنید $t_{i,1} = t_{ij}^{(1)} < \ldots < t_{ij}^{(L_j)} = t_{i,s_i}$ افرازی برای بازه $[t_{i,1}, t_{i,s_i}]$ باشد که هم‌ارز با متغیر پاسخ j از نمونه iام است. $t_{ij}^{(l)}$، lامین گره به ازای $l = 1, \ldots, L_j$ است. اگر برای تقریب $g_j(t_{i,k})$ از تابع اسپلاین مرتبه $d_j(\geqslant 1)$ استفاده کنیم، آن‌گاه خواهیم داشت

$$g_j(t_{i,k}) = \alpha_{j,0} + \alpha_{j,1}t_{i,k} + \ldots + \alpha_{j,d_j}t_{i,k}^{d_j} + \sum_{l=1}^{L_j}\alpha_{(d_j+1)+l}(t_{i,k} - t_{ij}^{(l)})_+^{d_j} = \boldsymbol{B}_j(t_{i,k})^\top\boldsymbol{\alpha}_j,$$

که در آن $\boldsymbol{B}_j(t_{i,k}) = \Big(۱, t_{i,k}, \ldots, t_{i,k}^{d_j}, (t_{i,k} - t_{ij}^{(۱)})_+^{d_j}, \ldots, (t_{i,k} - t_{ij}^{(L_j)})_+^{d_j}\Big)$ برداری $h_j \times ۱$ بعدی از توابع پایه بوده و $\boldsymbol{\alpha}_j = (\alpha_{j,۰}, \ldots, \alpha_{j,d_j}, \alpha_{j,d+۱}, \ldots, \alpha_{j,h_j},)^\top$ برداری از ضرایب اسپلاین با بعد $h_j = d_j + ۱ + L_j$ است. بنابراین مدل نیمه‌پارامتری (۸.۴) به یک مدل خطی به‌صورت

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{B}(\boldsymbol{t}_i)\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \qquad (۹.۴)$$

تبدیل می‌شود. که $\boldsymbol{B}_j(\boldsymbol{t}_i) = \big(\boldsymbol{B}_j(t_{i,۱})^\top, \ldots, \boldsymbol{B}_j(t_{i,s_i})^\top\big)^\top$ و $\boldsymbol{B}(\boldsymbol{t}_i) = \operatorname{diag}\big(\boldsymbol{B}_۱(\boldsymbol{t}_i), \ldots, \boldsymbol{B}_r(\boldsymbol{t}_i)\big)$ یک ماتریس طرح $s_i \times h_i$ بعدی هم‌ارز با بخش اسپلاینی مدل است. اگر بعد کلی اثرات اسپلاین را با $h = \sum_{j=۱}^r h_j$ نشان دهیم، آنگاه $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_۱^\top, \ldots, \boldsymbol{\alpha}_r^\top)^\top$ یک بردار $h \times ۱$ بعدی از اثرات اسپلاین است. برای سادگی، مدل (۹.۴) را می‌توان به‌صورت

$$\boldsymbol{y}_i = \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \varepsilon_i, \qquad (۱۰.۴)$$

نوشت، به‌طوری‌که $\tilde{\boldsymbol{X}}_i = \operatorname{diag}\{\tilde{\boldsymbol{X}}_{i۱}, \ldots, \tilde{\boldsymbol{X}}_{ir}\}$ و $\tilde{\boldsymbol{X}}_{ij} = \big(\boldsymbol{X}_{ij}, \boldsymbol{B}_j(\boldsymbol{t}_i)\big)$ ماتریس طرح $s_i \times (p_j + h_j)$ بعدی برای پاسخ j از نمونه iام است. $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$ بردار پارامترهای رگرسیونی ترکیب شده با بعد $(p_j + h_j) \times ۱$ است.

طبق راهکار ارائه شده در بخش ۱.۴، مدل (۱۰.۴) را می‌توان به‌صورت مدل سلسله مراتبی زیر

$$\begin{aligned} \boldsymbol{y}_i|(\boldsymbol{b}_i, \tau_i) &\sim \mathcal{N}_{n_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \tau_i^{-۱}\boldsymbol{R}_i), \qquad (۱۱.۴) \\ \boldsymbol{b}_i|\tau_i &\sim \mathcal{N}_q(۰, \tau_i^{-۱}\boldsymbol{D}), \\ \tau_i &\sim \mathcal{G}(\nu/۲, \nu/۲). \end{aligned}$$

نوشت. حال به برآورد ML پارامترها بر اساس الگوریتم ECM می‌پردازیم. این روش کاملاً مشابه الگوریتم بکار رفته در بخش است با این تفاوت که

۲.۲.۴ برآورد ML و الگوریتم ECM

فرض کنید $\boldsymbol{\Theta} = \Big(\boldsymbol{\theta}, \operatorname{vech}(\boldsymbol{D}), \operatorname{vech}(\boldsymbol{\Sigma}), \phi, \nu\Big)$ مجموعه‌ای از تمام پارامترهای مدل باشد. اگر اثرات تصادفی $\boldsymbol{b}_i$ را به عنوان مقادیر گمشده و مجموعه $\{(\boldsymbol{y}_i, \boldsymbol{b}_i), i = ۱, \ldots, n\}$ را داده‌های کامل در نظر بگیریم، آنگاه لگاریتم تابع درستنمایی شرطی بدون در نظر گرفتن جمله ثابت به‌صورت

$$\begin{aligned} \ell_c(\boldsymbol{\Theta}) = \sum_{i=۱}^n \ell_c^{[i]}(\boldsymbol{\Theta}) = \sum_{i=۱}^n \tfrac{۱}{۲}\Big\{ &\log|\boldsymbol{R}_i^{-۱}| + \log|\boldsymbol{D}^{-۱}| - \tau_i[\boldsymbol{\varepsilon}_i^\top\boldsymbol{R}_i^{-۱}\boldsymbol{\varepsilon}_i \qquad (۱۲.۴) \\ &+ \boldsymbol{b}_i^\top\boldsymbol{D}^{-۱}\boldsymbol{b}_i] + \nu\log\big(\tfrac{\nu}{۲}\big) - ۲\log\tfrac{\nu}{۲} + \nu(\log\tau_i - \tau_i)\Big\}. \end{aligned}$$

است. برای ارزیابی امید شرطی (۱۲.۴) به شرط داده‌های مشاهده شده $\boldsymbol{y} = \{\boldsymbol{y}_i, i = ۱, \ldots, n\}$ و برآوردهای جاری $\widehat{\boldsymbol{\Theta}}^{(r)} = \Big(\widehat{\boldsymbol{\theta}}^{(r)}, \operatorname{vech}(\widehat{\boldsymbol{D}}^{(r)}), \operatorname{vech}(\widehat{\boldsymbol{\Sigma}}^{(r)}), \widehat{\phi}^{(r)}, \widehat{\nu}^{(r)}\Big)$، از قضیه زیر استفاده می‌کنیم.

نتیجه ۱.۲.۴. (وانگ و فن، ۲۰۱۱). تحت مدل سلسله مراتبی (۳.۴)، داریم

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{bmatrix} | \tau_i &\sim \mathcal{N}_{n_i+q}\left(\begin{bmatrix} \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} \\ \boldsymbol{0} \end{bmatrix}, \ \tau_i^{-1} \begin{bmatrix} \boldsymbol{\Lambda}_i & \boldsymbol{Z}_i\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{Z}_i^\top & \boldsymbol{D} \end{bmatrix} \right), \\
\boldsymbol{b}_i|(\boldsymbol{y}_i, \tau_i) &\sim \mathcal{N}_q\Big(\boldsymbol{D}\boldsymbol{Z}_i^\top\boldsymbol{\Lambda}_i^{-1}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta}), \tau_i^{-1}(\boldsymbol{D}^{-1} + \boldsymbol{Z}^\top\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i)^{-1}\Big), \\
\tau_i|\boldsymbol{y}_i &\sim \mathcal{G}\Big((\nu + n_i)/2, (\nu + \Delta_i)/2\Big),
\end{aligned}$$

به‌طوری‌که $\Delta_i = \Delta_i(\boldsymbol{\theta}, \boldsymbol{D}, \boldsymbol{\Sigma}, \phi, \nu) = \tilde{\boldsymbol{\epsilon}}_i^\top\boldsymbol{\Lambda}_i^{-1}\tilde{\boldsymbol{\epsilon}}_i$ فاصله ماهالانوبیس بین $\boldsymbol{y}_i$ و $\tilde{\boldsymbol{X}}_i\boldsymbol{\theta}$ است و $\tilde{\boldsymbol{\epsilon}}_i = \boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta}$.

مراحل اصلی الگوریتم EM به شرح زیر انجام می‌شود:

گام E: محاسبه تابع $Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \mathrm{E}\left(\ell_c^{[i]}(\boldsymbol{\Theta})|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}\right)$ که مجموع $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) = \mathrm{E}\left(\ell_c(\boldsymbol{\Theta})|\boldsymbol{y}, \widehat{\boldsymbol{\Theta}}^{(r)}\right)$ برای $i = 1, \ldots, n$ است و $Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$ به‌صورت

$$\begin{aligned}
Q_i(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)}) &= \frac{1}{2}\Big\{ \log|\boldsymbol{R}_i^{-1}| + \log|\boldsymbol{D}^{-1}| - \mathrm{tr}(\boldsymbol{D}^{-1}\widehat{\boldsymbol{B}}_i^{(r)}) - \mathrm{tr}(\boldsymbol{R}_i^{-1}\widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\theta})) \\
&\quad + \nu\big(\log(\nu/2) + \widehat{\kappa}_i^{(r)} - \widehat{\tau}_i^{(r)}\big)\Big\} - \log(\nu/2),
\end{aligned}$$

تعریف می‌شوند، به‌طوری‌که

$$\begin{aligned}
\widehat{\tau}_i^{(r)} &= \mathrm{E}(\tau_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = (\widehat{\nu}^{(r)} + n_i)/(\widehat{\nu}^{(r)} + \widehat{\Delta}_i^{(r)}), \\
\widehat{\kappa}_i^{(r)} &= \mathrm{E}(\log\tau_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \mathcal{D}_g(\frac{\widehat{\nu}^{(r)} + n_i}{2}) - \log(\frac{\widehat{\nu}^{(r)} + \widehat{\Delta}_i^{(r)}}{2}), \\
\widehat{\boldsymbol{B}}_i^{(r)} &= \mathrm{E}(\tau_i\boldsymbol{b}_i\boldsymbol{b}_i^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\tau}_i^{(r)}\widehat{\boldsymbol{b}}_i^{(r)}\widehat{\boldsymbol{b}}_i^{(r)^\top} + \widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)}, \\
\widehat{\boldsymbol{\Psi}}_i^{(r)}(\boldsymbol{\theta}) &= \mathrm{E}(\tau_i\boldsymbol{e}_i\boldsymbol{e}_i^\top|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) \\
&= \widehat{\tau}_i^{(r)}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)})(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i^{(r)})^\top + \boldsymbol{Z}_i\widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)}\boldsymbol{Z}_i^\top,
\end{aligned}$$

و

$$\begin{aligned}
\widehat{\Delta}_i^{(r)} &= (\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\widehat{\boldsymbol{\theta}}^{(r)})^\top\widehat{\boldsymbol{\Lambda}}_i^{(r)^{-1}}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\widehat{\boldsymbol{\theta}}^{(r)}), \\
\widehat{\boldsymbol{\Lambda}}_i^{(r)} &= \boldsymbol{Z}_i\widehat{\boldsymbol{D}}^{(r)}\boldsymbol{Z}_i^\top + \widehat{\boldsymbol{R}}_i^{(r)}, \\
\widehat{\boldsymbol{R}}_i^{(r)} &= \widehat{\boldsymbol{\Sigma}}^{(r)} \otimes \boldsymbol{\Omega}_i(\widehat{\phi}^{(r)}), \\
\widehat{\boldsymbol{b}}_i^{(r)} &= E(\boldsymbol{b}_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \widehat{\boldsymbol{D}}^{(r)}\boldsymbol{Z}_i^\top\widehat{\boldsymbol{\Lambda}}_i^{(r)^{-1}}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\widehat{\boldsymbol{\theta}}^{(r)}), \\
\widehat{\boldsymbol{V}}_{\boldsymbol{b}_i}^{(r)} &= \widehat{\tau}_i^{(r)}\mathrm{cov}(\boldsymbol{b}_i|\boldsymbol{y}_i, \widehat{\boldsymbol{\Theta}}^{(r)}) = \big(\widehat{\boldsymbol{D}}^{(r)^{-1}} + \boldsymbol{Z}_i^\top\widehat{\boldsymbol{R}}_i^{(r)^{-1}}\boldsymbol{Z}_i\big)^{-1}.
\end{aligned}$$

گام M: به روزکردن $\widehat{\boldsymbol{\Theta}}^{(r)}$ توسط $\widehat{\boldsymbol{\Theta}}^{(r+1)} = \max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$.

در مجموع گام CM الگوریتم ECM به‌صورت الگوریتم ۵ پیشنهاد می‌شود.

الگوریتم ۵ برآوردگر PML و الگوریتم ECM

۱: **گام CM-۱:** ثابت نگه‌داشتن $\phi = \widehat{\phi}^{(r)}$ و $\nu = \widehat{\nu}^{(r)}$، به روزکردن $\widehat{\boldsymbol{\theta}}^{(r)}$، $\widehat{\boldsymbol{D}}^{(r)}$ و $\widehat{\boldsymbol{\Sigma}}^{(r)}$ از طریق بیشینه کردن $Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(r)})$ که به‌صورت زیر حاصل می شوند:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(r+1)} &= \Big(\sum_{i=1}^{n} \widehat{\tau}_i^{(r)} \tilde{\boldsymbol{X}}_i^\top \widehat{\boldsymbol{R}}_i^{(r)^{-1}} \tilde{\boldsymbol{X}}_i\Big)^{-1} \sum_{i=1}^{n} \widehat{\tau}_i^{(r)} \tilde{\boldsymbol{X}}_i^\top \widehat{\boldsymbol{R}}_i^{(r)^{-1}} (\boldsymbol{y}_i - \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i^{(r)}),\\
\widehat{\boldsymbol{D}}^{(r+1)} &= n^{-1} \sum_{i=1}^{n} \widehat{\boldsymbol{B}}_i^{(r)},\\
\widehat{\sigma}_{lm}^{(r+1)} &= \begin{cases} (\sum_{i=1}^{n} s_i)^{-1} \sum_{i=1}^{n} \operatorname{tr}\Big(\boldsymbol{\Omega}_i^{-1}(\widehat{\phi}^{(r)}) \widehat{\boldsymbol{\psi}}_{ils}^{(r)}(\widehat{\boldsymbol{\theta}}^{(r+1)})\Big) & \text{for } l = s,\\ (2\sum_{i=1}^{n} s_i)^{-1} \sum_{i=1}^{n} \operatorname{tr}\Big(\boldsymbol{\Omega}_i^{-1}(\widehat{\phi}^{(r)})\big(\widehat{\boldsymbol{\psi}}_{ils}^{(r)}(\widehat{\boldsymbol{\theta}}^{(r+1)}) + \widehat{\boldsymbol{\psi}}_{isl}^{(r)}(\widehat{\boldsymbol{\theta}}^{(r+1)})\big)\Big) & \text{for } l \neq s, \end{cases}
\end{aligned}$$

۲: **گام CM-۲:** بر اساس $\widehat{\boldsymbol{\theta}}^{(r+1)}$، $\widehat{\boldsymbol{D}}^{(r+1)}$ و $\widehat{\boldsymbol{\Sigma}}^{(r+1)}$، محاسبه شده در گام قبل، $(\widehat{\phi}^{(r+1)}, \widehat{\nu}^{(r+1)})$ از طریق بیشینه کردن تابع Q به دست می‌آید. به عبارت دیگر

$$\begin{aligned}
(\widehat{\phi}^{(r+1)}, \widehat{\nu}^{(r+1)}) = \arg\max_{(\phi,\nu)} \Big\{ \sum_{i=1}^{n} \Big(& r \log|\boldsymbol{\Omega}_i^{-1}(\phi)| - \operatorname{tr}\big((\widehat{\boldsymbol{\Sigma}}^{(r+1)^{-1}} \otimes \boldsymbol{\Omega}_i^{-1}(\phi)) \times \widehat{\boldsymbol{\Psi}}_i^{(r+1/2)}\big)\Big)\\
& + \nu \log(\tfrac{\nu}{2}) - 2\log\Gamma(\tfrac{\nu}{2}) + \nu(\widehat{\kappa}_i^{(r)} - \widehat{\tau}_i^{(r)})\Big\},
\end{aligned}$$

به‌طوری‌که $\widehat{\boldsymbol{\Psi}}_i^{(r+1/2)} = \widehat{\boldsymbol{\Psi}}_i^{(r)}(\widehat{\boldsymbol{\beta}}^{(r+1)})$.

۳: **گام همگرایی:** گام های E و CM را تا زمان همگرایی به مقداری ثابت تکرار می‌کنیم تا برآوردگر ML بدست آید.

۳.۲.۴ خواص مجانبی

برای مطالعه رفتار حدی برآوردگرهای $\widehat{\boldsymbol{\Xi}} = (\widehat{\boldsymbol{\beta}}^\top, \boldsymbol{\omega}^\top, \widehat{\nu})$ و $\widehat{\boldsymbol{g}}(\boldsymbol{t}_i) = \boldsymbol{B}(\boldsymbol{t}_i)\widehat{\boldsymbol{\alpha}}$، مجموعه‌ای از شرایط نظم مشابه بخش ۱.۴ مورد نیاز است. در این جا $\boldsymbol{\omega} = \big(\operatorname{vech}(\boldsymbol{D})^\top, \operatorname{vech}(\boldsymbol{\Sigma})^\top, \boldsymbol{\phi}^\top\big)^\top$. اگر $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{p} 0$ و هنگامی که $n \to \infty$، داشته باشیم $\sup_t |\boldsymbol{B}(\boldsymbol{t}_i)\widehat{\boldsymbol{\alpha}} - \boldsymbol{g}(\boldsymbol{t}_i)| \xrightarrow{p} 0$، آن‌گاه می‌گوییم برآوردگرهای $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{g}}(\boldsymbol{t}_i))$ سازگاراند. تحت شرایط نظم، قضیه ۱.۲.۴ را بیان می‌کنیم که رفتار حدی برآوردگرهای $\widehat{\boldsymbol{\Xi}}$ و $\widehat{\boldsymbol{g}}(\boldsymbol{t}_i)$ را نشان می‌دهد.

قضیه ۱.۲.۴. اگر $L_j \approx N^{1/(2m_j+1)}$ آن‌گاه

$$\begin{aligned}
(i) \quad &: \quad \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{s_i} \big(\widehat{g}_j(t_{i,k}) - g_j(t_{i,k})\big)^2 = O_p(N^{2m_j/(2m_j+1)}), \quad j = 1, \ldots, r,\\
(ii) \quad &: \quad \widehat{\boldsymbol{\Xi}} \xrightarrow{p} \boldsymbol{\Xi}, \quad \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{\boldsymbol{\beta\beta}}^{-1}), \quad \sqrt{\mathbf{n}}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{\mathbf{d}} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{\boldsymbol{\eta\eta}}^{-1}).
\end{aligned}$$

که در آن $N = \sum_{i=1}^{n} s_i$ تعداد کل مشاهدات بوده و ماتریس‌های $\boldsymbol{I}_{\boldsymbol{\beta\beta}}$ و $\boldsymbol{I}_{\boldsymbol{\eta\eta}}$ در شرط (۱.A) معرفی شده‌اند.

۴.۲.۴ مطالعات عددی

در این بخش، فرایند برازش مدل و برآورد پارامترها را به‌صورت عددی بررسی می‌کنیم. ابتدا مجموعه‌ای از داده‌های شبیه‌سازی شده و یک مورد مطالعاتی با داده‌های واقعی را توسط روش پیشنهادی یعنی برآوردگر ML در مدل نیمه‌پارامتری با اثرات آمیخته t چندمتغیره ($MtSMM$) مورد تحلیل قرار می‌دهیم.

۱.۴.۲.۴ مطالعه شبیه‌سازی

متغیرهای پاسخ $\boldsymbol{y}_i$ را به ازای $r = ۲$ متغیر پاسخ از مدل (۸.۴) و با فرضیات (۲.۴) تولید می‌کنیم. به‌طوری‌که

$$\begin{aligned} g_۱(t_{i,k}) &= ۲/۵ - ۱/۵t_{i,k} + ۰/۴t_{i,k}^۲, \\ g_۲(t_{i,k}) &= ۲\sin(۲\pi t_{i,k}) + ۲\cos(۲\pi t_{i,k}), \end{aligned}$$

و پارامترهای مدل به‌صورت $\boldsymbol{\beta} = (\beta_{۱۱}, \beta_{۱۲}, \beta_{۲۱}, \beta_{۲۲})^\top = (۱, ۲, -۲, ۴,)^\top$ تعیین می‌شوند. سایر مولفه‌های مدل مشابه مطالعه شبیه سازی در بخش ۱.۴ است. برای $g_۱(t_{i,k})$ تقریب اسپلاین درجه دو، $d_۱ = ۲$، و برای $g_۲(t_{i,k})$ تقریب اسپلاین مکعبی، $d_۲ = ۳$، را درنظر می‌گیریم. مطابق دلایل بیان شده در بخش (۲.۴) حجم نمونه و مقادیر درجه آزادی به ترتیب با مقادیر $n = ۵۰, ۱۰۰$ و $\nu = ۵, ۵۰$ تعیین می‌شوند. دقت برآورد با مقادیر اریبی، SD و MSE پارامترهای برآورد شده تحت برازش مدل‌های مذکور اندازه‌گیری می‌شود. علاوه بر این، انتخاب مدل با محاسبه معیارهای AIC و BIC انجام می‌شود. نتایج برای حالت‌های $n = ۲۵$ و $n = ۱۰۰$ به ترتیب در جداول ۴.۴ و ۵.۴ گزارش شده است. همان‌طور که مشاهده می‌شود، $MtSMM$ نسبت به $MNSMM$ مقادیر SD، MSE کمتری دارد. این قضیه در مورد معیارهای AIC وBIC نیز صادق است. این رفتار در حالت $(n, \nu) = (۲۵, ۵)$ مشهودتر است. هنگامی که داده‌ها به توزیع نرمال نزدیکترند ($\nu = ۵۰$)، هر دو مدل برآوردها و انحراف‌های استاندارد تقریباً مشابهی را نتیجه می‌دهند. همچنین دریافتیم که مقادیر SD و MSE با افزایش اندازه نمونه کاهش می‌یابند که تأیید کننده ویژگی‌های مجانبی خوب برآوردگرهای ML، هستند.

۲.۴.۲.۴ تحلیل داده‌های بیماری صفراوی

در این بخش داده‌های $PBCseq$ تحلیل می‌شوند. شکل ۴.۱ نشان می‌دهد رفتار دو متغیر در طول زمان متفاوت از هم است. بنابراین به نظر می‌رسد که برای مطالعه تغییر رفتار آن‌ها در طول زمان، استفاده از طرح ناپارامتری مجزا برای هر یک از متغیرها، مناسب باشد. طرح باید به گونه‌ای باشد که تاثیر مقادیر پرت موجود در داده‌ها را نیز کنترل کند. لذا استفاده از مدل $MtSMM$ پیشنهاد می‌شود. استفاده از این مدل این اجازه را می‌دهد که علاوه بر مطالعه ارتباط بین بیلیروبین و آلبومین، نحوه

تغییرات و ارتباط این دو متغیر در طول زمان را بررسی کنیم. برای بررسی تنومند بودن این روش در حضور مشاهدات پرت، علاوه بر مدل پیشنهادی، مدل $MNSMM$ را نیز به داده‌ها برازش داده و نتایج با هم مقایسه می‌شوند. تنضیمات مدل مشابه بخش ۱.۴ است. در اینجا برای ماتریس $\mathbf{\Omega}_i$، سه نوع اختیار همبستگی شامل همبستگی استقلال (UNC)، همبستگی تبادل‌پذیر (EX) و همبستگی اتورگرسیو مرتبه یک ($AR(۱)$) را درنظر می‌گیریم. در مجموع با ۶ مدل نامزد روبرو هستیم.

جدول ۶.۴ خلاصه‌ای از برازش ۶ مدل نامزد را نشان می‌دهد که شامل مقادیر AIC، BIC و $MSPE$ است. با توجه به تمام معیارها، استفاده از ساختار همبستگی $AR(۱)$، برای هر دو مدل $MtSMM$ و $MtSMM$ مناسب بوده است. جدول ۷.۴ برآورد پارامترهای رگرسیونی به همراه انحراف‌های استاندارد اثرات ثابت برای $MtSMM$ و $MNSMM$ با همبستگی $AR(۱)$ را ارائه می‌دهد. نتایج بیان می‌کنند که برآورد اثرات ثابت در دو مدل $MtSMM$ و $MNSMM$ تقریباً یکسان است اما $MtSMM$ در مقایسه با $MNSMM$، انحراف استاندارد بسیار کمتری را نتیجه می‌دهد، که بیانگر عملکرد مناسب $MtSMM$ است.

۳.۴ پیشنهادات برای آینده تحقیق

مدل اثرات آمیخته نیمه‌پارامتری با جمع‌آوری ویژگی‌های مثبت هر سه دسته مدل‌های رگرسیونی پارامتری، ناپارامتری و نیمه‌پارامتری در یک مدل واحد، از انعطاف بسیار بالایی در تحلیل داده‌های طولی برخوردار است. همان‌گونه که نتایج گزارش شده در مطالعات شبیه‌سازی و تحلیل داده‌های واقعی، برتری این مدل نسبت به مدل‌های معمول و کارایی بالای برآوردگرهای معرفی‌شده را نشان می‌دهد، بر این اساس، تحلیل داده‌های طولی در چارچوب مدل معرفی‌شده زمینه‌ای نوپا و ابزاری کارآمد برای فعالیت‌های آتی آماردانان فراهم خواهد آورد. ما نیز با تکیه بر این مهم تصمیم داریم، مطالعات و فعالیت‌های خود را در این راستا قرار داده و این مدل را در بسیاری از مسائل و استنباط‌های آماری دیگر وارد کنیم. در ادامه پیشنهادات خود را به طور مختصر ارائه خواهیم نمود.

۱.۳.۴ مدل اثرات آمیخته تعمیم‌یافته نیمه‌پارامتری برای داده‌های چندجمله‌ای

با بررسی و مطالعه اجمالی مباحث این رساله، می‌توان پی برد که یکی از اولین موضوعات کاربردی که می‌توان به عنوان مطالعه بعدی درنظر گرفت، ادغام بخش‌های ۱.۴ و ۲.۴ و پرداختن به مسئله انتخاب متغیر در مدل نیمه‌پارامتری با اثرات آمیخته با توزیع t چندمتغیره است. همچنین فرض کنید که داده‌های چندجمله‌ای مورد بحث در این فصل، ترکیبی از انواع داده‌های رسته‌ای باشند که هر کدام نیازمند به‌کارگیری تابع ربط منحصر به خود است. در این حالت می‌توان یک مدل اثرات

آمیخته تعمیم‌یافته نیمه‌پارامتری را به‌صورت

$$g_j(\boldsymbol{\mu}_{ij}) = \boldsymbol{\eta}_{ij} = \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i + \boldsymbol{f}(\boldsymbol{t}_{ij})$$

در نظر گرفت. اگرچه این مدل می‌تواند طیف گسترده‌تری از کاربردهای عملی را دربر بگیرد، اما به‌طور قابل توجهی پیچیده‌تر و از نظر محاسباتی سنگین‌تر خواهد بود. لذا نیازمند تحقیقات جامعی پیرامون استراتژی‌های محاسباتی و برنامه‌های کاربردی است.

۲.۳.۴ استفاده از توابع تاوان تجمعی

در این رساله تمرکز اصلی روی برآورد و انتخاب متغیر جزء پارامتری مدل‌ها بود. با این حال می‌توان مشابه کار سوای (۲۰۰۹)، تکنیک پیشنهادی را برای جزء ناپارامتری مدل‌ها نیز به‌کار برد. به عنوان مثال در فصل ۲ با به‌کارگیری جملات تاوان تجمعی، مسئله برآورد و انتخاب متغیر هم‌زمان در رابطه (۱۵.۲) به حل مسئله بیشینه‌سازی لگاریتم تابع درستنمایی تاوانیده، به‌صورت

$$\ell_c^{\mathrm{Pen}}(\boldsymbol{\Theta}) = \ell_c(\boldsymbol{\Theta}) - n\sum_{k=1}^{p} P_{\lambda_n}^{(1)}(|\beta_k|) - n\sum_{k=1}^{p} P_{\lambda_n}^{(2)}(|\alpha_k|)$$

تبدیل می‌شود، که $P_{\lambda_n}^{(1)}(.)$ ضرایب پارامتری مدل و $P_{\lambda_n}^{(2)}(.)$ ضرایب ناپارامتری مدل را کنترل می‌کنند. جزئیات نظری مسئله بیان‌شده نیازمند توجه ویژه‌ای است و می‌توان به عنوان موضوع جالبی برای تحقیقات آینده درنظر گرفت.

۳.۳.۴ مدل ضرایب متغیر نیمه‌پارامتری با اثرات آمیخته

همان‌گونه که در فصل اول این رساله بیان شد، در تحلیل داده‌های طولی با سه هدف اصلی زیر روبرو هستیم:

۱. مطالعه تغییرات آزمودنی‌ها در طول زمان.
۲. مطالعه رابطه بین متغیرها (اثر متغیرهای تبیینی بر روی متغیر پاسخ).
۳. مطالعه تاثیر زمان بر رابطه بین متغیرها.

مدل اثرات آمیخته، تنها قادر به برآورد نقاطی است که مشاهده واقعی هم ارز با آن موجود باشد؛ بنابراین اثر متغیرهای تبیینی بر روی متغیر پاسخ را می‌توان دریافت. از آنجایی‌که، برآورد متغیر پاسخ در مدل اثرات آمیخته نیمه‌پارامتری به دلیل وجود تابع ناپارامتری نسبت به زمان، به‌صورت یک منحنی هموار می‌باشد، هدف مطالعه تغییرات آزمودنی‌ها در طول زمان محقق می‌گردد. حال اگر متغیرهایی تبیینی داشته باشیم که در هر نقطه زمانی تاثیر متفاوتی روی پاسخ‌ها داشته باشند و علاقه‌مند به مطالعه تاثیر زمان بر رابطه بین متغیرها باشیم، مدل اثرات آمیخته نیمه‌پارامتری با ضرایب متغیر را می‌توان به‌صورت

$$\boldsymbol{y}(t) = \boldsymbol{X}(t)^{\top}\boldsymbol{\beta}(\boldsymbol{t}) + \boldsymbol{W}(\boldsymbol{t})^{\top}\boldsymbol{\alpha} + \boldsymbol{Z}(\boldsymbol{t})^{\top}\boldsymbol{b} + \boldsymbol{\varepsilon}(\boldsymbol{t}),$$

در نظر گرفت، به‌طوری‌که پارامتر $\boldsymbol{\beta}(\boldsymbol{t}) = \left(\boldsymbol{\beta}_1(\boldsymbol{t}), \ldots, \boldsymbol{\beta}_p(\boldsymbol{t})\right)^\top \in \mathbb{R}^p$ تابعی نامعلوم از زمان است. این ویژگی باعث می‌شود که اطلاعات حجیمی از تغییر رفتار پدیده‌ها و نحوه تاثیر متغیرها روی یکدیگر در زمان های مختلف به‌دست آیند. با درنظر گرفتن هر یک از روش های ناپارامتری معرفی‌شده برای $\boldsymbol{\beta}(\boldsymbol{t})$، می‌توان مساله برآورد در این نوع مدل‌های بسیار منعطف را حل نمود.

اگرچه تاکنون تحقیقات بسیار متعددی در زمینه مدل‌بندی داده‌های طولی با ضرایب متغیر از جمله هوور و همکاران (۱۹۹۸)، وو و یانگ (۲۰۰۰)، زو و ژو (۲۰۰۷)، تانگ و چنگ (۲۰۰۸)، نوه و پارک (۲۰۱۰)، تانگ و همکاران (۲۰۱۳)، یانگ و همکاران (۲۰۱۴) و چنگ و همکاران (۲۰۱۴) انجام گرفته‌است؛ با این وجود در هیچ‌کدام از این تحقیقات مدل پیشنهادی این مجموعه مورد بررسی قرار نگرفته‌است. از این‌رو پیشنهاد می‌شود ضمن بررسی ویژگی‌ها و اهمیت کاربردی مدل مذکور به مباحث و موضوعات زیر نیز پرداخته شود.

۱. مسئله برآورد و انتخاب متغیر همزمان.

۲. بررسی نظریه‌های مرتبط با داده‌های بعد بالا و بعد بسیار بالا.

۳. مطالعه نسخه تعمیم‌یافته مدل مذکور جهت انطباق با داده‌های رسته‌ای.

جدول ۲.۴: نتایج شبیه‌سازی: کارآیی روش‌های $MtLMM$ تاوانیده و $MNMM$ تاوانیده به ازای nها و νهای مختلف.

		$n = ۲۵$				$n = ۱۰۰$			
		$\nu = ۵$		$\nu = ۵۰$		$\nu = ۵$		$\nu = ۵۰$	
		$MtLMM$	$MNMM$	$MtLMM$	$MNMM$	$MtLMM$	$MNMM$	$MtLMM$	$MNMM$
$\beta_{۱۱}$	Est	$-۱/۰۲۰$	$-۱/۰۳۵$	$-۱/۰۶۷$	$-۱/۰۲۷$	$-۰/۹۷۷$	$-۰/۹۵۶$	$-۱/۰۰۰$	$-۱/۰۰۱$
	SE	$۰/۲۸۹$	$۰/۲۸۶$	$۰/۳۲۹$	$۰/۳۳۱$	$۰/۱۸۰$	$۰/۲۷۰$	$۰/۱۴۵$	$۰/۱۴۵$
	MSE	$۰/۰۸۳$	$۰/۰۸۲$	$۰/۱۰۸$	$۰/۱۱۲$	$۰/۰۳۳$	$۰/۰۷۴$	$۰/۰۲۱$	$۰/۰۲۱$
$\beta_{۱۲}$	Est	$-۱/۰۱۰$	$-۱/۰۲۷$	$-۱/۰۰۷$	$-۱/۰۱۳$	$-۰/۹۸۶$	$-۰/۹۵۸$	$-۰/۹۹۵$	$-۱/۰۰۰$
	SE	$۰/۲۸۲$	$۰/۳۴۲$	$۰/۲۶۰$	$۰/۲۵۸$	$۰/۱۴۰$	$۰/۲۵۴$	$۰/۱۳۸$	$۰/۱۳۵$
	MSE	$۰/۰۷۹$	$۰/۱۱۶$	$۰/۰۶۷$	$۰/۰۶۶$	$۰/۰۲۰$	$۰/۰۶۶$	$۰/۰۱۹$	$۰/۰۱۸$
$\beta_{۱۳}$	Est	$۲/۰۶۰$	$۲/۰۴۹$	$۲/۰۳۳$	$۲/۰۳۶$	$۱/۹۸۲$	$۱/۹۳۷$	$۲/۰۲۰$	$۲/۰۲۲$
	SE	$۰/۲۴۷$	$۰/۳۲۵$	$۰/۲۶۷$	$۰/۲۹۱$	$۰/۱۵۲$	$۰/۱۸$	$۰/۱۲۵$	$۰/۱۲۰$
	MSE	$۰/۰۶۴$	$۰/۱۰۶$	$۰/۰۷۲$	$۰/۰۸۵$	$۰/۰۲۳$	$۰/۰۳۶$	$۰/۰۱۶$	$۰/۰۱۶$
$\beta_{۲۱}$	Est	$-۰/۹۷۹$	$-۱/۰۲۲$	$-۰/۹۸۷$	$-۰/۹۸۰$	$-۰/۹۹۱$	$-۰/۹۸۰$	$-۰/۹۷۹$	$-۰/۹۸۱$
	SE	$۰/۲۰۰$	$۰/۲۱۸$	$۰/۲۰۵$	$۰/۲۲۰$	$۰/۱۰۴$	$۰/۱۳۳$	$۰/۱۰۰$	$۰/۰۹۲$
	MSE	$۰/۰۴۰$	$۰/۰۴۷$	$۰/۰۴۲$	$۰/۰۴۸$	$۰/۰۱۱$	$۰/۰۱۸$	$۰/۰۱۰$	$۰/۰۰۹$
$\beta_{۲۲}$	Est	$-۱/۰۱۲$	$-۰/۹۸۹$	$-۰/۹۹۰$	$-۱/۰۰۰$	$-۰/۹۹۹$	$-۱/۰۰۶$	$-۱/۰۰۳$	$-۱/۰۰۲$
	SE	$۰/۱۹۸$	$۰/۲۵۷$	$۰/۲۰۷$	$۰/۱۹۷$	$۰/۰۸۷$	$۰/۱۱۸$	$۰/۰۹۱$	$۰/۰۸۳$
	MSE	$۰/۰۳۹$	$۰/۰۶۵$	$۰/۰۳۸$	$۰/۰۴۲$	$۰/۰۰۸$	$۰/۰۱۴$	$۰/۰۰۸$	$۰/۰۰۷$
$\beta_{۲۳}$	Est	$۲/۰۰۳$	$۲/۰۱۸$	$۱/۹۶۶$	$۱/۹۵۸$	$۲/۰۰۹$	$۲/۰۰۴$	$۱/۹۹۱$	$۱/۹۹۱$
	SE	$۰/۱۷۶$	$۰/۱۹۴$	$۰/۱۸۰$	$۰/۲۰۳$	$۰/۰۸۵$	$۰/۱۰۰$	$۰/۰۸۵$	$۰/۰۸۵$
	MSE	$۰/۰۳۱$	$۰/۰۳۷$	$۰/۰۳۴$	$۰/۰۴۲$	$۰/۰۰۷$	$۰/۰۱۰$	$۰/۰۰۷$	$۰/۰۰۷$
$d_{۱۱}$	Est	$۱/۱۷۸$	$۳/۴۴۷$	$۱/۱۸۵$	$۲/۲۰۵$	$۱/۲۸۹$	$۳/۵۵۱$	$۱/۳۹۳$	$۲/۲۲۵$
	SE	$۰/۴۰۲$	$۱/۳۵۰$	$۰/۷۲۲$	$۰/۴۴۹$	$۰/۲۶۹$	$۰/۷۹۴$	$۰/۲۶۸$	$۰/۳۴۶$
	MSE	$۰/۸۳۵$	$۳/۸۸۶$	$۰/۵۵۷$	$۰/۸۶۴$	$۰/۵۷۷$	$۳/۰۳۰$	$۰/۴۴۰$	$۰/۱۶۹$
$d_{۱۲}$	Est	$۰/۱۴۲$	$۰/۳۳۰$	$۰/۱۶۱$	$۰/۳۱۸$	$۰/۱۷۳$	$۰/۴۴۶$	$۰/۲۰۳$	$۰/۳۱۱$
	SE	$۰/۲۴۷$	$۰/۹۶۰$	$۰/۵۰۱$	$۰/۲۸۲$	$۰/۱۶۱$	$۰/۵۹۹$	$۰/۱۵۶$	$۰/۲۴۹$
	MSE	$۰/۰۷۲$	$۰/۹۱۳$	$۰/۲۵۳$	$۰/۰۸۷$	$۰/۰۳۲$	$۰/۳۹۳$	$۰/۰۲۶$	$۰/۰۶۵$
$d_{۲۲}$	Est	$۱/۱۵۰$	$۳/۴۳۴$	$۱/۲۳۵$	$۲/۲۴۱$	$۱/۲۱۰$	$۳/۴۲۲$	$۱/۳۵۶$	$۲/۱۵۸$
	SE	$۰/۳۷۰$	$۱/۳۶۰$	$۰/۶۶۰$	$۰/۵۴۱$	$۰/۲۷۴$	$۰/۸۵۳$	$۰/۲۹۱$	$۰/۳۵۲$
	MSE	$۰/۸۵۷$	$۳/۸۷۴$	$۰/۴۸۸$	$۰/۸۷۵$	$۰/۶۹۹$	$۲/۷۴۲$	$۰/۴۹۹$	$۰/۱۴۷$
$\sigma_{۱۱}$	Est	$۰/۶۴۶$	$۲/۱۷۲$	$۰/۶۵۹$	$۱/۳۳۷$	$۰/۷۳۷$	$۲/۳۱۲$	$۰/۸۴۴$	$۱/۴۴۷$
	SE	$۰/۱۵۹$	$۰/۵۵۲$	$۰/۲۱۰$	$۰/۲۱۲$	$۰/۱۴۳$	$۰/۳۹۲$	$۰/۱۳۹$	$۰/۱۲۲$
	MSE	$۱/۸۵۹$	$۰/۳۲۹$	$۰/۴۸۴$	$۱/۸۴۳$	$۱/۶۱۶$	$۰/۲۴۹$	$۱/۳۵۵$	$۰/۳۲$
$\sigma_{۱۲}$	Est	$۰/۲۵۲$	$۰/۸۶۰$	$۰/۲۶۲$	$۰/۵۴۰$	$۰/۲۷۷$	$۰/۸۸۴$	$۰/۳۲۴$	$۰/۵۵۲$
	SE	$۰/۰۶۶$	$۰/۲۵۹$	$۰/۱۲۵$	$۰/۰۹۵$	$۰/۰۵۸$	$۰/۱۹۳$	$۰/۰۵۹$	$۰/۰۶۸$
	MSE	$۰/۲۵۲$	$۰/۰۷۸$	$۰/۲۴۷$	$۰/۰۶۰$	$۰/۲۲۷$	$۰/۰۵۵$	$۰/۱۸۵$	$۰/۰۴۴$
$\sigma_{۲۲}$	Est	$۰/۳۰۸$	$۱/۰۴۹$	$۰/۳۲۶$	$۰/۶۷۴$	$۰/۳۶۳$	$۱/۱۴۵$	$۰/۴۱۴$	$۰/۷۱۲$
	SE	$۰/۰۷۹$	$۰/۲۷۸$	$۰/۱۱۰$	$۰/۰۹۴$	$۰/۰۶۷$	$۰/۱۷۸$	$۰/۰۶۶$	$۰/۰۶۰$
	MSE	$۰/۴۸۶$	$۰/۰۷۸$	$۰/۴۶۳$	$۰/۱۱۸$	$۰/۴۱۱$	$۰/۰۵۲$	$۰/۳۴۸$	$۰/۰۸۷$
ν	Est	$۱/۹۴۶$	-	$۳/۱۷ \times ۱۰^{۶}$	-	$۲/۱۵۰$	-	$۳/۹۷۱$	-
	SE	$۰/۳۵۲$	-	$۲/۰۲ \times ۱۰^{۷}$	-	$۰/۴۲۵$	-	$۰/۹۰۲$	-
	MSE	$۹/۴۵۱$	-	$۴/۱۴ \times ۱۰^{۱۴}$	-	$۸/۳۰۱$	-	$۲۱۱۹/۴۵$	-

جدول ۳.۴: نتایج تحلیل داده‌های $PBCseq$: برآورد پارامترهای رگرسیونی (SD) تحت برازش سه مدل $MtLMM$ تاوانیده، $MtLMM$، $MNMM$ تاوانیده و $MNMM$.

پارامترها		$MtLMM$ تاوانیده	$MtLMM$	$MNMM$ تاوانیده	$MNMM$
اثرات ثابت	$\beta_{1\circ}$ $(Intercept)$	۱٫۲۷۹ (۰٫۰۲۰)	۱٫۶۴۵ (۰٫۱۲۱)	۱٫۲۳۶ (۰٫۰۲۰)	۱٫۴۹۷ (۰٫۱۸۰)
	β_{11} $(Time)$	۰٫۰۲۴ ($۸٫۸ \times ۱۰^{-۴}$)	۰٫۰۰۴ (۰٫۰۱۱)	۰٫۰۱۷۱ ($۸٫۹ \times ۱۰^{-۴}$)	۰٫۰۱۷ (۰٫۰۲۰)
	β_{12} (Sex)	−۰٫۳۷۲ (۰٫۰۱۰)	−۰٫۵۵۸ (۰٫۰۴۰)	−۰٫۳۳۱ (۰٫۰۱۱)	−۰٫۵۳۰ (۰٫۰۶۴)
	β_{13} $(Drug)$	۰٫۰۰۰ (۰٫۰۰۰)	−۰٫۰۰۸ (۰٫۰۲۰)	۰٫۰۰۰ (۰٫۰۰۰)	−۰٫۰۲۳ (۰٫۰۱۸)
	β_{14} (Age)	−۰٫۰۱۰ (۰٫۰۰۶)	−۰٫۰۱۱ (۰٫۰۰۱)	−۰٫۰۰۸ (۰٫۰۰۷)	−۰٫۰۱۰ (۰٫۰۰۲)
	$\beta_{2\circ}$ $(Intercept)$	۱٫۰۱۳ (۰٫۰۰۲)	۱٫۴۰۲ (۰٫۰۸۶)	۱٫۳۴۳ (۰٫۰۰۲)	۱٫۳۱۲ (۰٫۰۱۷)
	β_{21} $(Time)$	−۰٫۰۱۵ ($۱٫۱ \times ۱۰^{-۴}$)	−۰٫۰۲۰ (۰٫۰۰۶)	−۰٫۰۱۳ ($۱٫۲ \times ۱۰^{-۴}$)	−۰٫۰۱۴ (۰٫۰۰۱)
	β_{22} (Sex)	۰٫۰۰۰ ($۱٫۴ \times ۱۰^{-۴}$)	−۰٫۰۲۴ (۰٫۰۳۵)	۰٫۰۰۰ ($۱٫۵ \times ۱۰^{-۴}$)	−۰٫۰۰۱ (۰٫۰۰۵)
	β_{23} $(Drug)$	۰٫۰۰۰ ($۸٫۴ \times ۱۰^{-۴}$)	۰٫۰۲۸ (۰٫۰۱۶)	۰٫۰۰۰ ($۸٫۸ \times ۱۰^{-۴}$)	۰٫۰۲۵ (۰٫۰۰۳)
	β_{24} (Age)	۰٫۰۰۶ ($۳٫۶ \times ۱۰^{-۵}$)	−۰٫۰۰۲ (۰٫۰۰۱)	−۰٫۰۰۲ ($۳٫۶ \times ۱۰^{-۵}$)	−۰٫۰۰۱ (۲×۱۰^{-۴})
اثرات تصادفی	d_{11}	۰٫۸۷۹	۲٫۱۲۵	۱٫۰۴۰	۱٫۱۱۴
	d_{21}	۰٫۰۲۵	۰٫۰۰۰	۰٫۰۲۸	−۰٫۳۴۳
	d_{22}	۰٫۰۷۴	۰٫۲۱۱	۰٫۰۸۲	۲٫۶۳۱
	d_{31}	−۰٫۰۳۸	−۰٫۱۰۰	−۰٫۰۵۹	−۰٫۰۶۵
	d_{32}	−۰٫۰۰۳	−۰٫۰۰۱	−۰٫۰۰۱	۰٫۰۱۸
	d_{33}	۰٫۰۱۳	۰٫۰۲۲	۰٫۰۱۲	۰٫۰۱۳
	d_{41}	−۰٫۰۱۱	−۰٫۰۲۳	−۰٫۰۰۷	۰٫۰۱۶
	d_{42}	−۰٫۰۰۹	−۰٫۰۲۴	−۰٫۰۰۹	−۰٫۱۵۷
	d_{43}	۰٫۰۰۱	−۰٫۰۰۲	−۰٫۰۰۱	−۰٫۰۱۵
	d_{44}	۰٫۰۰۳	۰٫۰۱۰	۰٫۰۰۲	۰٫۰۳۷
خطاها	σ_{11}	۰٫۰۴۶	۰٫۱۲۴	۰٫۰۶۶	۰٫۱۰۲
	σ_{21}	−۰٫۰۰۱	−۰٫۰۰۱	−۰٫۰۰۰۵	−۰٫۰۰۲
	σ_{22}	۰٫۰۰۳	۰٫۰۱۲	۰٫۰۰۶۱	۰٫۰۱۹
	ν	۱٫۰۸۲	۱٫۰۰۰	-	-
معیارها	ℓ_{max}	۸۶۷۵٫۴۰	۵۴۳۳٫۴۴	۶۵۰۱٫۰۵	۵۳۴۶٫۲۷
	AIC	−۱۷۳۰۲٫۸۱	−۱۰۸۱۸٫۸۸	−۱۲۹۵۶٫۱۱	−۱۰۶۴۶٫۵۴
	BIC	−۱۷۲۱۲٫۹۸	−۱۰۷۲۹٫۰۵	−۱۲۸۷۰٫۰۲	−۱۰۵۶۰٫۴۵
	$MSPE$	۷٫۷۴	۶٫۸۶	۷٫۷۵	۷٫۸۴

جدول ۴.۴: نتایج شبیه‌سازی: کارآیی روش‌های $MtSMM$، $MtLMM$ و $MNSMM$ به ازای $n = ۲۵$ و νهای مختلف.

پارامترها		$n = ۲۵$					
		$\nu = ۵$			$\nu = ۵۰$		
		$MtSMM$	$MtLMM$	$MNSMM$	$MtSMM$	$MtLMM$	$MNSMM$
$\beta_{۱۱}$	$BIAS$	۰٫۶۴۱	۷٫۹۶۶	۰٫۶۴۵	۰٫۷۴۶	۷٫۹۰۶	۰٫۷۴۶
	SD	۰٫۳۶۶	۰٫۴۸۹	۰٫۴۱۸	۰٫۳۶۸	۰٫۳۹۳	۰٫۳۶۸
	MSE	۰٫۵۴۳	۶۳٫۷۰۱	۰٫۵۸۹	۰٫۶۹۰	۶۲٫۶۵۸	۰٫۶۹۰
$\beta_{۱۲}$	$BIAS$	۰٫۰۳۰	۰٫۵۵۵	۰٫۰۶۴	۰٫۰۵۸	۰٫۶۴۸	۰٫۰۶۴
	SD	۰٫۲۲۴	۰٫۷۷۸	۰٫۲۶۵	۰٫۲۷۱	۰٫۷۳۰	۰٫۲۷۰
	MSE	۰٫۰۵۰	۰٫۹۰۸	۰٫۰۷۴	۰٫۰۷۶	۰٫۹۴۷	۰٫۰۷۶
$\beta_{۲۱}$	$BIAS$	۰٫۹۸۷	۱٫۰۷۴	۰٫۹۵۵	۰٫۸۹۲	۱٫۰۶۷	۰٫۸۷۹
	SD	۰٫۳۷۷	۰٫۴۱۷	۰٫۳۷۷	۰٫۳۷۵	۰٫۳۹۱	۰٫۳۷۳
	MSE	۱٫۱۱۶	۱٫۳۲۴	۱٫۰۵۳	۰٫۹۳۵	۱٫۲۸۹	۰٫۹۱۰
$\beta_{۲۲}$	$BIAS$	۰٫۳۱۲	۱٫۲۹۲	۰٫۲۱۵	۰٫۱۷۳	۱٫۲۸۷	۰٫۱۵۱
	SD	۰٫۴۸۴	۰٫۵۱۲	۰٫۴۱۴	۰٫۳۱۴	۰٫۴۳۵	۰٫۳۰۰
	MSE	۰٫۳۲۹	۱٫۹۳۰	۰٫۲۱۶	۰٫۱۲۸	۱٫۸۴۴	۰٫۱۱۲
$d_{۱۱}$	$BIAS$	۰٫۸۴۹	۰٫۹۵۴	۰٫۹۶۵	۰٫۱۳۱	۰٫۹۵۷	۰٫۱۱۰
	SD	۰٫۳۷۹	۰٫۲۶۱	۱٫۳۲۵	۰٫۶۴۴	۰٫۲۵۹	۰٫۷۳۷
	MSE	۰٫۸۶۲	۰٫۹۷۷	۲٫۶۷۰	۰٫۴۲۷	۰٫۹۸۳	۰٫۵۵۰
$d_{۱۲}$	$BIAS$	۰٫۰۳۶	۰٫۱۶۰	۰٫۳۰۳	۰٫۰۵۰	۰٫۱۹۴	۰٫۰۸۶
	SD	۰٫۳۳۸	۰٫۱۶۵	۰٫۹۹۰	۰٫۵۰۲	۰٫۱۶۶	۰٫۵۶۲
	MSE	۰٫۱۱۴	۰٫۰۵۳	۱٫۰۶۲	۰٫۲۵۱	۰٫۰۶۵	۰٫۳۲۰
$d_{۲۲}$	$BIAS$	۰٫۵۵۳	۱٫۶۵۰	۱٫۷۰۳	۰٫۰۴۳	۱٫۷۳۴	۰٫۳۰۲
	SD	۰٫۵۴۶	۰٫۱۵۲	۲٫۰۷۰	۰٫۸۲۸	۰٫۱۰۰	۰٫۹۳۹
	MSE	۰٫۶۰۱	۲٫۷۴۴	۷٫۱۴۳	۰٫۶۸۰	۳٫۰۱۷	۰٫۹۶۴
$\sigma_{۱۱}$	$BIAS$	۱٫۲۲۶	۳٫۱۳۵	۰٫۱۸۸	۰٫۷۷۰	۳٫۲۴۷	۰٫۵۹۵
	SD	۰٫۱۷۷	۰٫۵۱۱	۰٫۷۳۴	۰٫۱۹۰	۰٫۶۱۴	۰٫۲۲۳
	MSE	۱٫۵۳۳	۱۰٫۰۸۵	۰٫۵۶۹	۰٫۶۲۷	۱۰٫۹۱۳	۰٫۴۰۳
$\sigma_{۱۲}$	$BIAS$	۰٫۴۵۳	۰٫۷۶۸	۰٫۰۸۳	۰٫۲۲۷	۰٫۷۷۰	۰٫۱۵۶
	SD	۰٫۱۵۷	۰٫۱۲۷	۰٫۴۳۲	۰٫۲۴۰	۰٫۰۹۶	۰٫۲۷۰
	MSE	۰٫۲۳۰	۰٫۶۰۶	۰٫۱۹۲	۰٫۱۰۸	۰٫۶۰۲	۰٫۱۰۰
$\sigma_{۲۲}$	$BIAS$	۰٫۳۹۸	۰٫۶۳۹	۲٫۱۶۸	۱٫۵۵۶	۰٫۶۴۶	۱٫۸۸۳
	SD	۰٫۲۲۲	۰٫۰۵۷	۰٫۵۰۷	۰٫۳۱۰	۰٫۰۶۰	۴۰٫۳۴۴
	MSE	۰٫۲۰۷	۰٫۴۱۱	۴٫۹۵۳	۲٫۵۱۵	۰٫۴۲۰	۳٫۶۶۱
ν	$BIAS$	۰٫۰۲۳	۳٫۲۶۲	-	۱٫۱۵۷	۴۸٫۹۹۹	-
	SD	۰٫۱۳۲	۰٫۴۳۹	-	۰٫۹۷۹	۰٫۰۰۸	-
	MSE	۰٫۰۱۸	۱۰۸	-	۲٫۲۸۸	۲۴۰۰٫۸۸۴	-
	ℓ_{max}	−۲۱۵٫۳۲۹	−۲۳۲٫۳۷۴	−۳۲۷٫۲۵۳	−۲۶۵٫۹۲۳	−۲۲۸٫۷۵۷	−۲۸۷٫۰۸۹
	AIC	۴۵۲٫۶۵۸	۴۸۶٫۷۴۸	۶۷۴٫۵۰۶	۵۵۳٫۸۴۶	۴۷۹٫۵۱۵	۵۹۴٫۱۷۷
	BIC	۴۶۶٫۰۶۵	۵۰۰٫۱۵۶	۶۸۶٫۶۹۵	۵۶۷٫۲۵۴	۴۹۲٫۹۲۲	۶۰۶٫۳۶۶

جدول ۵.۴: نتایج شبیه‌سازی: کارآیی روش‌های $MtSMM$، $MtLMM$ و $MNSMM$ به ازای $n = ۱۰۰$ و νهای مختلف.

Parameters		$n = ۱۰۰$					
		$\nu = ۵$			$\nu = ۵۰$		
		$MtSMM$	$MtLMM$	$MNSMM$	$MtSMM$	$MtLMM$	$MNSMM$
$\beta_{۱۱}$	$BIAS$	۰٫۷۲۴	۸٫۱۸۷	۰٫۷۲۷	۰٫۷۴۰	۸٫۱۹۴	۰٫۷۴۰
	SD	۰٫۱۹۰	۰٫۲۵۵	۰٫۲۱۹	۰٫۱۶۲	۰٫۱۷۶	۰٫۱۶۳
	MSE	۰٫۵۶۰	۶۷٫۰۸۵	۰٫۵۷۶	۰٫۵۷۳	۶۷٫۱۷۳	۰٫۵۷۳۴
$\beta_{۱۲}$	$BIAS$	۰٫۰۳۳	۰٫۸۶۹	۰٫۰۲۳	۰٫۰۰۶	۰٫۸۲۵	۰٫۰۰۶
	SD	۰٫۱۲۰	۰٫۲۷۲	۰٫۱۳۸	۰٫۱۰۰	۰٫۲۴۷	۰٫۱۰۳
	MSE	۰٫۰۱۵	۰٫۸۲۸	۰٫۰۱۹	۰٫۰۱۰	۰٫۷۴۲	۰٫۰۱۱
$\beta_{۲۱}$	$BIAS$	۰٫۰۱۸	۰٫۱۶۲	۰٫۰۱۴	۰٫۰۶۱	۰٫۱۷۹	۰٫۰۵۶
	SD	۰٫۲۲۳	۰٫۲۳۹	۰٫۲۳۹	۰٫۱۴۸	۰٫۲۱۴	۰٫۱۴۷
	MSE	۰٫۰۵۰	۰٫۰۸۳	۰٫۰۵۷	۰٫۰۲۵	۰٫۰۷۷	۰٫۰۲۵
$\beta_{۲۲}$	$BIAS$	۰٫۷۰۸	۰٫۵۶۸	۰٫۶۲۲	۰٫۶۴۳	۰٫۵۰۹	۰٫۶۳۲
	SD	۰٫۲۳۰	۰٫۲۸۶	۰٫۲۱۸	۰٫۱۵۵	۰٫۲۱۱	۰٫۱۵۷
	MSE	۰٫۵۵۲	۰٫۴۰۴	۰٫۴۳۴	۰٫۴۳۷	۰٫۳۰۳	۰٫۴۲۴
$d_{۱۱}$	$BIAS$	۰٫۵۶۲	۱٫۰۲۸	۱٫۵۷۰	۰٫۱۱۴	۱٫۰۴۴	۰٫۲۱۲
	SD	۰٫۲۷۷	۰٫۱۶۷	۰٫۹۷۵	۰٫۳۳۹	۰٫۱۵۸	۰٫۳۹۴
	MSE	۰٫۳۹۲	۱٫۰۸۴	۳٫۴۰۴	۰٫۱۲۷	۱٫۱۱۴	۰٫۱۹۹
$d_{۱۲}$	$BIAS$	۰٫۰۴۴	۰٫۲۴۴	۰٫۲۵۳	۰٫۰۲۱	۰٫۲۳۷	۰٫۰۲۱
	SD	۰٫۲۰۸	۰٫۰۸۷	۰٫۷۴۷	۰٫۲۷۱	۰٫۰۷۹	۰٫۳۱۱
	MSE	۰٫۰۴۵	۰٫۰۶۷	۰٫۶۱۶	۰٫۰۷۳	۰٫۰۶۲	۰٫۰۹۶
$d_{۲۲}$	$BIAS$	۰٫۳۳۲	۱٫۶۳۱	۱٫۶۹۷	۰٫۱۰۲	۱٫۶۸۳	۰٫۴۵۸
	SD	۰٫۲۹۳	۰٫۰۸۰	۰٫۸۶۱	۰٫۳۵۱	۰٫۰۵۸	۰٫۴۰۳
	MSE	۰٫۱۹۵	۲٫۶۶۸	۳٫۶۱۲	۰٫۱۳۳	۲٫۸۳۶	۰٫۳۷۰
$\sigma_{۱۱}$	$BIAS$	۱٫۱۳۷	۴٫۰۰۹	۴۰٫۲۹۸	۰٫۷۶۸	۴٫۳۲۰	۰٫۵۴۲
	SD	۰٫۱۲۳	۰٫۶۲۳	۰٫۴۰۳	۰٫۱۰۶	۰٫۵۸۰	۰٫۱۲۳
	MSE	۱٫۳۰۸	۱۶٫۴۵۲	۰٫۲۵۰	۰٫۶۰۱	۱۸٫۹۹۸	۰٫۳۰۹
$\sigma_{۱۲}$	$BIAS$	۰٫۴۲۳	۰٫۷۴۱	۰٫۱۲۱	۰٫۲۴۷	۰٫۷۸۰	۰٫۱۵۶
	SD	۰٫۰۷۹	۰٫۰۵۸	۰٫۲۲۳	۰٫۱۱۴	۰٫۰۵۶	۰٫۱۳۰
	MSE	۰٫۱۸۵	۰٫۵۵۲	۰٫۰۶۴	۰٫۰۷۴	۰٫۶۱۱	۰٫۰۴۱
$\sigma_{۲۲}$	$BIAS$	۰٫۸۰۷۸	۰٫۵۶۹	۲٫۶۹۵	۱٫۹۲۱	۰٫۵۶۷	۲٫۳۹۱
	SD	۰٫۲۰۷	۰٫۰۵۴	۰٫۲۶۰	۰٫۱۸۳	۰٫۰۵۲	۰٫۱۷۵
	MSE	۰٫۶۹۵	۰٫۳۲۶	۷٫۳۲۷	۳٫۷۲۳	۰٫۳۲۵	۵٫۷۴۵
ν	$BIAS$	۰٫۰۱۴	۳٫۲۹۵	-	۰٫۷۶۰	۴۸٫۹۹۸	-
	SD	۰٫۱۴۳	۰٫۰۸۰	-	۰٫۹۷۶	۰٫۰۰۶	-
	MSE	۰٫۰۲۰	۱۰٫۸۶۴	-	۱٫۵۲۰	۲۴۰۰٫۸۴۶	-
	ℓ_{max}	−۹۹۴٫۸۰۸	−۱۰۳۴٫۴۹۱	−۱۴۰۰٫۳۷۷	−۱۱۳۵٫۹۴۸	−۱۰۳۸٫۵۰۰	−۱۲۲۶٫۶۳۱
	AIC	۲۰۱۱٫۶۱۷	۲۰۹۰٫۹۸۲	۲۸۲۰٫۷۵۳	۲۲۹۳٫۸۹۶	۲۰۹۸٫۹۹۹	۲۴۷۳٫۲۶۲
	BIC	۲۰۴۰٫۲۷۴	۲۱۱۹٫۶۳۹	۲۸۴۶٫۸۰۵	۲۳۲۲٫۵۵۲	۲۱۲۷٫۶۵۶	۲۴۹۹٫۳۱۴

جدول ۶.۴: نتایج تحلیل داده‌های $PBCseq$: نتایج معیارهای انتخاب مدل تحت برازش مدل‌های $MtSMM$ و $MNSMM$ و ساختارهای همبستگی مختلف.

		Criteria			
مدل	$\mathbf{\Omega}_i$	ℓ_{max}	AIC	BIC	$MSPE$
$MtSMM$	UNC	۵۰۹۰٫۴۹	−۱۰۱۴۲٫۹۸	−۱۰۰۵۴٫۶۳	۶٫۴۰
	EX	۵۲۵۱٫۲۴	−۱۰۴۶۲٫۴۸	−۱۰۳۷۶٫۱۳	۶٫۳۸
	$AR(۱)$	۵۳۵۸٫۴۱	−۱۰۶۷۲٫۸۲	−۱۰۵۹۰٫۴۷	۶٫۱۳
$MNSMM$	UNC	۳۸۱۷٫۹۴	−۷۵۹۳٫۸۸	−۷۵۱۵٫۲۸	۷٫۲۴
	EX	۳۹۳۸٫۵۰	−۷۸۳۵٫۰۰	−۷۷۵۶٫۴۰	۷٫۱۸
	$AR(۱)$	۴۰۱۸٫۸۸	−۷۹۹۵٫۷۷	−۷۹۱۷٫۱۶	۷٫۱۰

جدول ۷.۴: نتایج تحلیل داده‌های $PBCseq$: برآورد پارامترهای رگرسیونی (SD) تحت برازش مدل‌های $MtSMM$ و $MNSMM$.

پارامترها		$MtSMM$	$MNSMM$
اثرات ثابت	$\beta_{10}(Intercept)$	۱/۴۳۲(۰/۰۱۸)	۱/۴۲۸(۰/۱۱۲)
	$\beta_{11}(Sex)$	−۰/۵۴۲(۰/۰۱۰)	−۰/۵۳۸(۰/۰۴۲)
	$\beta_{12}(Drug)$	−۰/۰۱۵(۰/۰۰۷)	−۰/۰۱۹(۰/۰۱۳)
	$\beta_{13}(Age)$	$-۰/۰۱۰(۳/۳\times۱۰^{-۴})$	−۰/۰۱۰۲(۰/۰۰۲)
	$\beta_{20}(Intercept)$	۱/۱۷۱(۰/۰۰۲)	۱/۱۷۲(۰/۰۰۷)
	$\beta_{21}(Sex)$	−۰/۰۰۴(۰/۰۰۱)	−۰/۰۰۵(۰/۰۰۳)
	$\beta_{22}(Drug)$	$۰/۰۱۰(۸/۹\times۱۰^{-۴})$	۰/۰۰۹(۰/۰۰۱)
	$\beta_{23}(Age)$	$-۰/۰۰۲(۴/۱\times۱۰^{-۵})$	$-۰/۰۰۱(۱/۱\times۱۰^{-۴})$
اثرات تصادفی	d_{11}	۰/۸۰۷	۰/۷۳۸
	d_{21}	−۰/۰۱۲	−۰/۰۲۱
	d_{22}	۰/۳۱۷	۰/۳۰۹
	d_{31}	−۰/۰۲۲	−۰/۰۱۹
	d_{32}	−۰/۰۰۳	−۰/۰۰۲
	d_{33}	۰/۳۱۰	۰/۳۰۹
	d_{41}	−۰/۰۰۷	−۰/۰۰۶
	d_{42}	−۰/۰۰۵	−۰/۰۰۴
	d_{43}	−۰/۰۹۵	−۰/۰۹۵
	d_{44}	۰/۲۵۶	۰/۲۵۶
خطاها	σ_{11}	−۰/۰۸۵	−۰/۱۱۴
	σ_{21}	−۰/۰۰۳	−۰/۰۰۲
	σ_{22}	−۰/۲۲۲	−۰/۲۲۴
	ν	۳/۱۲۰	-

مراجع

[۱] آرست، م. (۱۳۹۵)، مقایسه رفتار برخی برآوردگرهای انقباضی بریج در مدل رگرسیون چندگانه، پایان‌نامه کارشناسی ارشد، دانشگاه صنعتی شاهرود.

[۲] تعاونی، م. (۱۳۹۲)، استفاده از مدل‌های مختلف رگرسیونی در تحلیل داده‌های طولی، پایان‌نامه کارشناسی ارشد، دانشگاه گلستان.

[۳] رضایی‌قهرودی، م.، گنجعلی، م. و هرندی، ف. (۱۳۸۸)، مدل انتقالی برای تحلیل داده‌های طولی با پاسخ‌های دو متغیره ترتیبی و اسمی، پژوهش‌های آماری ایران.

[۴] روزبه، م. (۱۳۹۰)، برآورد در مدل‌های خطی جزئی، رساله دکتری، دانشگاه فردوسی مشهد.

[۵] کاظمی، م.، شاهسونی، د. و آرشی، م. (۱۳۹۷)، انتخاب متغیر و تشخیص ساختار در بعد بالا برای مدل‌های جمعی خطی-جزیی، مجله علوم آماری، ۱۲، ۴۸۵-۵۱۲.

[۶] گنجعلی، م. و رضایی‌قهرودی، ز. (۱۳۸۹)، تحلیل چند متغیره‌ی گسسته در مطالعات مقطعی و طولی، پژوهشکده آمار: تهران.

[7] Airy, G.B. (1861), *On the Algebraical and Numerical Theory of Errors of Observation and the Combination of Observations*, Macmillan, London.

[8] Akaike, H. (1973), *Information Theory and an Extension of the Maximum Likelihood Principle* In 2nd Int. Symp. on Information Theory. (Edited by B. N. Petrov and F. Csaki), Akademiai Kiado, Budapest, 267–281.

[9] Belsley, D.A., Kuh, E. and Welsch, R.E. (1980), *Regression Diagnostics*, John Wiley, New York.

[10] Bondell, H.D., Krishna, A. and Ghosh, S.K. (2010), Joint Variable Selection for Fixed and Random Effects in Linear Mixed-effects Models, *Biometrics*, 66, 1069–1077.

[11] Bowman, A. (1984), An Alternative Method of Cross-Validation for the Smoothing of Density Estimates, *Biometrika*, 71, 353–360.

[12] Breiman, L. (1996), Heuristics of Instability and Stabilization in Model Selection, *Annal. Statist.*, 24(6), 2350–-2383.

[13] Breheny, P., and Huang, J.(2011), Coordinate Descent Algorithms for Nonconvex Penalized Regression, with Applications to Biological Feature Selection, *Annal. Appl. Statist.*, 5, 232–253.

[14] Breheny, P., and Huang, J. (2015), Group Descent Algorithms for Nonconvex Penalized Linear and Logistic Regression Models with Grouped Predictors, *Statist. Comp.*, 25(2), 173–187.

[15] Breheny, P., and Huang, J. (2019), *High-Dimensional Regression Modeling: Methodology, Applications, and Software*, Chapman and Hall.

[16] Breslow, N.E. and Clayton, D.G. (1993), Approximate Inference in Generalized Linear Mixed Models, *J. Amer. Statist. Assoc.*, 88(421), 9–25.

[17] Buhlmann, P., and Van De Geer, S. (2011), *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Science and Business Media.

[18] Cantoni, E., Mills Flemming, J. and Ronchetti, E. (2005), Variable Selection for Marginal Longitudinal Generalized Linear Models, *Biometrics*, 61(2), 507–514.

[19] Chatterjee, S., and Hadi, A.S. (2012), *Regression Analysis by Example*, Wiley, New York.

[20] Cheng, M.Y., Honda, T., Li, J. and Peng, H. (2014), Nonparametric Independence Screening and Structural Identification for Ultra-high Dimensional Longitudinal Data, *Annal. Statist.*, 42, 1819–1849.

[21] Chung, H., Park, Y. and Lanza, S.T. (2005), Latent Transition Analysis with Covariates: Pubertal Timing and Substance Use Behaviours in Adolescent Females, *Statist. Med.*, 24, 2895–2910.

[22] Dempster AP, Laird NM, Rubin DB. (1977), Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J. Roy. Statist. Soc.* B, 39, 1–38.

[23] Diggle, P.J., Heagerty, P., Liang, K.Y. and Zeger, S.L. (2002), *Analysis of Longitudinal Data*, Oxford University Press, NewYork.

[24] Dziak, J.J. (2006), *Penalized Quadratic Inference Functions for Variable Selection in Longitudinal Research*, Ph.D Thesis, the Pennsylvania State University.

[25] Drapper, N.R., and Smith, H. (1981), *Applied Regresion Analysis*, Wiley.

[26] Dziak, J., Li, R., and Collins, L. (2005), *Critical Review and Comparison of Variable Selection Procedures for Linear Regression*, State College, Pennsylvania State University.

[27] Desboulets, L. (2018), A Review on Variable Selection in Regression Analysis, *Econometrics*, 6(4), ; doi: 10.3390/econometrics6040045.

[28] Ding, J., Tarokh, V., and Yang, Y. (2018), Model Selection Techniques: An Overview. *IEEE Sig. Proc. Mag.*, 35(6), 16–34.

[29] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004), Least Angle Regression, *Annal. Statisti.*, 32, 407–451.

[30] Eliot, M., Ferguson, J., Reilly M.P. and Foulkes, A.S. (2011), Ridge Regression for Longitudinal Biomarker Data, *Int. J. Biostat.*, 7, Article 37.

[31] Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric Estimates of the Relation Between Weather and Electricity Sales, *J. Amer. Statist. Assoc.*, 81, 310–320.

[32] Fan, J.Q., Huangm, T. and Li, R. (2007), Analysis of Longitudinal Data with Semiparametric Estimation of Covariance Function, *J. Amer. Statist. Assoc.*, 102, 632–641.

[33] Fan, J.Q. and Li, R. (2001), Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties, *J. Amer. Statist. Assoc.*, 96, 1348–1360.

[34] Fan, J.Q. and Li, R. (2004), New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis, *J. Amer. Statist. Assoc.*, 99, 710–723.

[35] Fan, J.Q, and Li, R. (2006), Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery, arXiv preprint math/0602133.

[36] Fan, J.Q, and Lv, J. (2008), Sure Independence Screening for Ultrahigh Dimensional Feature Space, *J. Roy. Statist. Soc.*, B, 70(5), 849–911.

[37] Fan, J.Q, Ma, Y., and Dai, W. (2014), Nonparametric Independence Screening in Sparse Ultra-high-dimensional Varying Coefficient Models, *J. Amer. Statist. Assoc.*, 109(507), 1270–1284.

[38] Fan, J.Q, Samworth, R., and Wu, Y. (2009), Ultrahigh Dimensional Feature Selection: Beyond the Linear Model, *J. Mach. Learn. Res.*, 10, 2013–2038.

[39] Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2009), *Longitudinal Data Analysis*, Chapman and Hall, London.

[40] Fitzmaurice, G.M. and Laird, N.M. (1993), A Likelihood-Based Method for Analysing Longitudinal Binary Responses, *Biometrika*, 80, 141- 151.

[41] Fitzmaurice, G.M., Laird, N.M. and Ware, J.H. (2011), *Applied Longitudinal Analysis*, Wiley, NewYork.

[42] Frank, I.E. and Friedman, J.H. (1993), A statistical View of Some Chemometrics Regression Tools (with discussion), *Technometrics*, 35, 109–148.

[43] Fu, W.J. (1998), Penalized Regressions: the Bridge Versus the Lasso, *J. Comp. Grap. Gtatist.*, 7(3), 397–416.

[44] Fu, W.J. (2003), Penalized Estimating Equations, *Biometrics*, 59, 126–132.

[45] Geenens, G. (2011). Curse of Dimensionality and Related Issues in Nonparametric Functional Regression, *Statist. Surveys*, 5, 30–43.

[46] Giraud, C. (2014), *Introduction to High-dimensional Statistics*, Chapman and Hall.

[47] Green, P.J. and Silverman, B.W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, London.

[48] Gujrati, D.N. (2004), *Basic Econometrics*, Tata McGrow- Hill.

[49] Hardin, J.W. and Hilbe, J.M. (2003), *Generalized Estimating Equations*, Chapman and Hall, New York.

[50] Hardle, W., Liang, H. and Gao, J.T. (2000), *Partially Linear Models*, Springer, Heidelberg.

[51] Harvile, D.A. (1977), Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems, *J. Amer. Statist. Assoc.*, 72, 320–338.

[52] Hastie, T. and Tibshirani, R. (1990), *Generalized Additive Models*, Chapman and Hall, London.

[53] Hedeker, D. and Gibbons, R.D. (1994), A Random-Effects Ordinal Regression Model for Multilevel Analysis, *Biometrics*, 50, 933–944.

[54] Hedeker, D. and Gibbons, R,D. (2006), *Longitudinal Data Analysis*, Wiley, New York.

[55] He, X, Fung, W.K., Zhu, Z.Y. (2005), Robust Estimation in Generalized Partial Linear Models for Clustered Data, *J. Amer. Statist. Assoc.*, 100, 1176–1184

[56] He, X.M., Zhu, Z.Y. and Fung, W.K. (2002), Estimation in a Semiparametric Model for Longitudinal Data with Unspecified Dependence Structure, *Biometrika*, 89, 579–590.

[57] Hoerl, A. and Kennard, R. (1970b), Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*, 12, 55–67.

[58] Hoover, D.R., Rice, J.A., Wu, C.O. and Yang, L.P. (1998), Nonparametric Smoothing Estimates of Timevarying Coefficient Models with Longitudinal Data, *Biometrika*, 85, 809–822.

[59] Huang, J., Wu, C., and Zhou, L. (2004), Polynomial Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data, *Statist. Sinica*, 14, 763–788.

[60] Huang, J.Z., Zhang, L., and Zhou, L. (2007), Efficient Estimation in Marginal Partially Linear Models for Longitudinal/Clustered Data using Splines, *Scand. J. Stat.*, 34, 451–477.

[61] Hu, Z., Wang, N. and Carroll, J. R. (2004), Profile-Kernel Versus Backfitting in the Partially Linear Models for Longitudinal/Clustered Data, *Biometrika*, 91, 251–262.

[62] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013), An Introduction to Statistical Learning, springer, New York.

[63] Johnstone, I.M., and Titterington, D.M. (2009), Statistical Challenges of Highdimensional Data, *Philos. Trans. R. Soc. B* A, 367, 4237–4253.

[64] Kibria, B.M.G. (2006), The Matrix-t Distribution and its Applications in Predictive Inference, *J. Multivariate Anal.* 97, 785–795.

[65] Kurum, E., Li, R., Shiffman, S. and Yao, W. (2016), Time- varying Coefficient Models for Joint Modeling Binary and Continues Outcome in Longitudinal Data, *Stat. Sinica*, 29, 979–1000

[66] Lee, Y. and Nelder, J.A. (2001), Hierarchical Generalized Linear Models: A Synthesis of Generalized Linear Models, Random-Effect Models and Structured Dispersions, *Biometrika*, 88, 987–1006.

[67] Leon, A.C., Hedeker, D. and Teres, J.J. (2007), Bias Reduction in Effectiveness Analyses of Longitudinal Ordinal Doses with a Mixed-Effects Propensity Adjustment, *Statist. Med.*, 26, 110-123.

[68] Liang, H. (2009), Generalized Partially Linear Mixed-Effects Models Incorporating Mismeasured Covariates, *Ann. Inst. Statist. Math.*, 61, 27–46.

[69] Liang, K.Y. and Zeger, S.L. (1986), Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, 73, 13–22.

[70] Li, Z. and Zhu, L. (2010), On Variance Components in Semiparametric Mixed Models for Longitudinal Data, *Scan. J. Statist.*, 37, 442–457.

[71] Lin, X. (2007), Estimation using Penalized Quasilikelihood and Quasi-pseudo-likelihood in Poisson Mixed Models, *Lifetime Data Analysis*, 13, 533–544.

[72] Lin, X. and Carroll, R.J. (2000), Nonparametric Function Estimation for Clustered Data when the Predictor is Measured without/with Error, *J. Amer. Statist. Assoc.*, 95, 520–534.

[73] Lin, T.I. and Lee, J.C. (2006), A Robust Approach to t Linear Mixed Models Applied to Multiple Sclerosis Data, *Statist. Med.* 25, 1397–1412.

[74] Lin, T.I. and Lee, J.C. (2007), Bayesian Analysis of Hierarchical Linear Mixed Modeling using the Multivariate t Distribution, *J. Statist. Plann. Inference.* 137, 484–495.

[75] Ma, S., Song, Q. and Wang, L. (2013), Simultaneous Variable Selection and Estimation in Semiparametric Modeling of Longitudinal/Clustered Data, *Bernoulli*, 19(1), 252–274.

[76] Malo, N., Libiger, O. and Schork, N. (2008), Accommodating Linkage Disequilibrium in Genetic-Association Analyses via Ridge Regression, *Amer. J. Hum. Genet.*, 82, 375–85.

[77] Marshall, G., De la Cruz-Mesia, R., Baron, A.E., Rutledge, J.H. and Zerbe, G.O. (2006), Non-linear Random Effects Model for Multivariate Responses with Missing Data, *Statist. Med.*, 25, 2817–2830.

[78] Mcculloch, C.E. (1997), Maximum Likelihood Algorithms for Generalized Linear Mixed Models, *J. Amer. Statist. Assoc.*, 92, 162–170.

[79] McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, Chapman and Hall, London, second edition.

[80] Meng, X.L. and Rubin, D.B. (1993), Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika*, 80, 267–278.

[81] Molenberghs, G. and Verbeke, G. (2005), *Models for Discrete Longitudinal Data*, Springer, New York.

[82] Montgomery, D.C., Peck, E.A. and Vining, G.G. (2012), *Introduction to Linear Regression Analysis*, Wiley, New York.

[83] Natarajan, R. and Kass, R.E. (2000), Reference Bayesian methods for generalized linear mixed models, *Amer. Statist.*, 95, 227–237.

[84] Nelder, J.A. and Wedderburn, R.W.M. (1972), Generalized Linear Models, *J. Roy. Statist. Soc.*, 135, 370–384.

[85] Ni, X., Zhang, D. and Zhang, H.H. (2010), Variable Selection for Semiparametric Mixed Models in Longitudinal Studies, *Biometrics*, 66, 79–88.

[86] Noh, H. and Park, B., (2010), Sparse Varying Coefficient Models for Longitudinal Data, *Statist. Sinica*, 20, 1183–1202.

[87] Pan, W. (2001), Akaike's Information Criterion in Generalized Estimating Equations, *Biometrics*, 57, 120–125.

[88] Pan, W. (2002), Goodness-of-Fit Tests for GEE with Correlated Binary Data, *Scand. J. Statist.*, 29, 101–110.

[89] Pan, J. and Thompson, R. (2007), Quasi-Monte Carlo Estimation in Generalized Linear Mixed Models, *Comput. Statist. Data Anal.*, 51, 5765–5775.

[90] Pinheiro, J.C., Liu, C.H. and Wu, Y.N. (2001), Efficient Algorithms for Robust Estimation in Linear Mixed Effects Model using the Multivariate t Distribution, *J. Comput. Graph. Statist.*, 10,249–276.

[91] Prentice, R.L. (1998), Correlated Binary Regression with Covariates Specific to Each Binary Observation, *Biometrics*, 44, 1033–1048.

[92] Qin, G.Y. and Zhu, Z.Y. (2007), Robust Estimation in Generalized Semiparametric Mixed Models for Longitudinal Data, *J. Mult. Anal.*, 98, 1658–1683.

[93] Qin, G.Y. and Zhu, Z.Y., (2009), Robustified Maximum Likelihood Estimation in Generalized Partial Linear Mixed Model for Longitudinal Data, *Biometrics*, 65, 52–59.

[94] Rahmani, M., Arashi, M., Mamode Khan, N. and Sunecher, Y. (2018), Improved Mixed Model for Longitudinal Data Analysis Using Shrinkage m Method, *Math. Sciences*, 12, 305–312.

[95] Rezaee, Z., and Ganjali, M. (2009), Testing Homogeneity in Markov Models for Analyzing Longitudinal Ordinal Response Data with Random Dropout, *J. Statist. Theo. App.* , 8, 125–139.

[96] Rezaee, Z., Ganjali, M., and Berridge, D. (2009), A Transition Model for Ordinal Response Data with Random Dropout: An Application to the Fluvoxamine Data, *J. Biopharm. Statist.*, 19, 658–671.

[97] Rice, J.A. and Wu, C.O. (2001), Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves, *Biometrics*, 57, 253–259.

[98] Rosa, G.J.M., Gianola و D. and Padovani و C.R. (2004), Bayesian Longitudinal Data Analysis with Mixed Models and Thick-tailed Distributions using MCMC, *J. Appl. Stat.* 31, 855–873.

[99] Roy, A. (2006), Estimating Correlation Coefficient Between two Variables with Repeated Observations using Mixed Effects Model, *Biom J.*, 48, 286–301.

[100] Rudemo, M. (1982), Empirical Choice of Histograms and Kernel Density Estimators, *Scand. J. Statist.*, 9, 65–78.

[101] Ruppert, D., Wand, M.P. and Carroll, R.J. (2003), *Semiparametric Regression*, Cambridge University Press, Cambridge.

[102] Saleh, A.K.Md.E., Arashi, M. and Kibria, B.M.G. (2019), *Theory of Ridge Regression Estimation with Applications*. Wiley, New York.

[103] Sammel, M., Lin, X. and Ryan, L. (1999), Multivariate Linear Mixed Models for Multiple Outcomes, *Statist. Med.*, 18, 2479–2492.

[104] Schumaker, L.L. (1981), *Spline Functions*, Wiley, New York.

[105] Schwarz, G. (1978), Estimating the Dimension of a Model, *Ann. Statist.*, 6, 461–464.

[106] Sengul, T.K., Stoffer, D.S. and Day, N.L. (2007), A Residuals-Based Transition Model for Longitudinal Analysis with Estimation in the Presence of Missing Data, *Statist. Med.*, 26, 3330–3341.

[107] Shah, A., Laird, N. and Schoenfeld, D. (1997), A Random-effects Model for Multiple Characteristics with Possibly Missing Data, *J. Roy. Statist. Soc.*, 92, 775–779.

[108] Simon, I., Barnett, J., Hannett, N., Harbison, C. T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., Jaakola, T. S., and Young, R.A. (2001), Serial Regulation of Transcriptional Regulators in the Yeast Cell Cycle, *Cell*, 106, 697–708.

[109] Sinha, S.K. and Sattar, A. (2015), Inference in Semi-Parametric Spline Mixed Models for Longitudinal Data, *METRON*, 73, 377–395.

[110] Song, X., Davidian, M. and Tsiatis, A.A. (2002), An Estimator for the Proportional Hazards Model with Multiple Mongitudinal Covariates Measured with Error, *Biostatistics*, 3, 511–528.

[111] Song, P.X.K., Zhang, P. and Qu, A. (2007), Maximum Likelihood Inference in Robust Linear Mixed-Effects Models using Multivariate t Distributions,*Statist. Sinica*, 17,929–943.

[112] Speckman, P. (1988), Kernel Smoothing in Partial Linear Models, *J. Roy. Statist. Soc.*, B, 50, 413–436.

[113] Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998), Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast Saccharomyces Cerevisiae by Microarray Hybridization, *Molecular Biology of Cell*, 9, 3273–3297.

[114] Stiratelli, R., Laird, N.M. and Ware, J.H. (1984), Random Effects Models for Serial Observations with Binary Response, *Biometrics*, 40, 961–971.

[115] Stone, C. (1982), Optimal Global Rates of Convergence for Nonparametric Regression, Ann. Statist., 10, 1040-1053.

[116] Sutradhar, B.C. and Das, K. (1999), On the Efficiency of Regression Estimators in Generalized Linear Models for Longitudinal Data, *Biometrika*, 86, 459-465.

[117] Sutradhar, B.C. and Farrell, P.J. (2004), Analyzing Multivariate Longitudinal Binary Data: A Generalized Estimating Equations Approach, *Can. J. Statist.*, 32, 39–55.

[118] Tanner, M.A. (1993), *Tools for Statistical Inference: Methods for the Exploration*, Springer.

[119] Tang, Q.G. and Cheng, L.S. (2008), M-Estimation and B-spline Approximation for Varying Coefficient Models with Longitudinal Data, *J. Nonp. Statist.*, 20, 611–625.

[120] Tang, Y., Wang, H.J. and Zhu, Z. (2013), Variable Selection in Quantile Varying Coefficient Models with Longitudinal Data, *Comp. Statist. Data Anal.*, 57, 435–449.

[121] Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso. *J. Roy. Statist. Soc*, B, 58, 267–288.

[122] Tikhonov, A. (1943), On the Stability of Inverse Problems, *Proceedings of the USSR Academy of Sciences*, 39, 267–88.

[123] Tutz, G. (2005), Modelling of Repeated Ordered Measurements by Isotonic Sequential Regression, *Statist. Model.*, 5, 269- -87.

[124] Verbeke, G. and Molenberghs, G. (2009), *Linear Mixed Models for Longitudinal Data*, Springer, New York.

[125] Wang, W.L. (2013), Multivariate t Linear Mixed Models for Irregularly Observed Multiple Repeated Measures with Missing Outcomes, *Biom. J.*, 55, 554–571.

[126] Wang N., Carroll R. and Lin X.H. (2005), Efficient Semiparametric Marginal Estimation for Longitudinal/Clustered Data, *J. Amer. Statist. Assoc.*, 100, 147–157.

[127] Wang, W.L. and Castro, L.M. (2018), Bayesian Inference on Multivariate-t Nonlinear Mixed-effects Models for Multiple Longitudinal Data with Missing Values, *Stat. Interface.*, 11, 251–264.

[128] Wang, W.L. and Fan, T.H. (2010), ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors, *Computat. Stat. Data Anal.* 54, 1328–1341.

[129] Wang, W.L. (2017), Mixture of Multivariate t Linear Mixed Models for Multi-Outcome Longitudinal Data with Heterogeneity, *Statist. Sinica*, 27, 733–760.

[130] Wang, W.L. and Fan, (2011), T.H. Estimation in Multivariate t Linear Mixed Models for Multiple Longitudinal Data,*Statist. Sinica*, 21, 1857–1880.

[131] Wang, W.L. and Fan, T.H. (2012), Bayesian Analysis of Multivariate t Linear Mixed Models using a Combination of IBF and Gibbs Samplers, *J Multivariate Anal.*, 105, 300–310.

[132] Wang, W.L. and Lin, T.I. (2015), Bayesian Analysis of Multivariate t Linear Mixed Models with Missing Responses at Random, *J. Stat. Comput. Simul.*, 85, 3594–3612.

[133] Wang, W.L. and Lin, T.I. (2014), Multivariate t Nonlinear Mixed-effects Models for Multi-outcome Longitudinal Data with Missing Values, *Statist. Med.*, 33, 3029–3046.

[134] Wang, L. and Qu, A. (2009), Consistent Model Selection and Data-Driven Smooth Tests for Longitudinal Data in the Estimating Equations Approach, *J. Roy. Statist. Soc.*, B, 71, 177–190.

[135] Wang, L., Zhou, J. and Qu, A. (2012), Penalized Generalized Estimating Equations for High-Dimensional Longitudinal Data Analysis, *Biometrics*, 68(2), 353–360.

[136] Wedderburn, R.W.M. (1974), Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss- Newton Method, *Biometrika*, 61, 439–47.

[137] Wu, C.O. and Chiang, C.T. (2000), Kernel Smoothing on Varying Coefficient Models with Longitudinal Dependent Variable, *Statist. Sinica*, 10, 433–456.

[138] Wu, T.T., and Lange, K. (2008), Coordinate Descent Algorithms for Lasso Penalized Regression, *Annal. Appl. Statist.*, 2(1), 224–244.

[139] Wu, H. and Liang, H. (2004), Backfitting Random Varying-Coefficient Models with Time-Dependent Smoothing Covariates, *Scand. J. Statist.*, 31, 3–19.

[140] Wu, H. and Zhang, T. (2006), *Nonparametric Regression Methods for Longitudinal Data Analysis*, Wily, New York.

[141] Xu, P.R., Fu, W. and Zhu, L.X. (2013), Shrinkage Estimation Analysis of Correlated Binary Data with a Diverging Number of Parameters, *Science China Math.*, 56, 359–377.

[142] Xue, L. (2009), Consistent Variable Selection in Additive Models. *Statist. Sinica*, 19, 1281–1296.

[143] Xue, L., Qu, A. and Zhou, J. (2010), Consistent Model Selection for Marginal Generalized Additive Model for Correlated Data. *J. Amer. Statist. Assoc.*, 105, 1518–1530.

[144] Xue, L.G. and Zhu, L.X. (2007), Empirical Likelihood for a Varying Coefficient Model with Longitudinal Data, *J. Amer. Statist. Assoc.*, 102, 642–654.

[145] Xue, L., Qu, A., and Zhou, J. (2010), Consistent model selection for marginal generalized additive model for correlated data, *J. Amer. Statist. Assoc.*, 105, 1518–1530.

[146] Yang, Y., Li, G. and Peng, H. (2014), Empirical Likelihood of Varying Coefficient Errors-in-Variables Models with Longitudinal Data, *J. Mult. Anal.*, 127, 1–18.

[147] Zeger, S.L. and Diggle, P.J. (1994), Semi-Parametric Models for Longitudinal Data with Application to CD4 Cell Numbers in HIV Seroconverters. *Biometrics*, 50, 689–99.

[148] Zeger, S.L. and Karim, M.R. (1991), Generalized Linear Models with Random Effects: A Gibbs sampling approach, *J. Amer. Statist. Assoc.*, 86, 79–86.

[149] Zhang, B. and Horvath, S. (2006), Ridge Regression Based Hybrid Genetic Algorithms for Multi-Locus Quantitative Trait Mapping, *Int. J. Bioinformatics Res. App.*, 1, 261–272.

[150] Zhang, C.H. (2010). Nearly Unbiased Variable Selection Under Minimax Concave Penalty, *Annal. statist.*, 38(2), 894–942.

[151] Zhang, D. (2004), Generalized Linear Mixed Models with Varying Coefficients for Longitudinal Data, *Biometrics*, 60, 8–15.

[152] Zhang, Y., and Li, R. (2011), Iterative Conditional Maximization Algorithm for Nonconcave Penalized Likelihood, *Nonpar. Statist. Mixture Models*, 336–351.

[153] Zhao, L.P. and Prentice, R.L. (1990), Correlated Binary Regression Using a Quadratic Exponential Model, *Biometrika*, 77, 642–648.

[154] Ziegler, A. and Vens, M. (2010). Generalized Estimating Equations: Notes on the Choice of the Working Correlation Matrix, *Methods Inf. Med.*, 49, 421–425.

[155] Zou, H. (2006), The Adaptive Lasso and its Oracle Properties, *J. Amer. Statist. Assoc.*, 101, 1418–1429.

[156] Zou, H. and Hastie, T. (2005), Regularization and Variable Selection via the Elasticnet, *J. Roy. Statist. Soc.*, B, 67, 301–320.

[157] Zou, H., and Li, R. (2008), One-step Sparse Estimates in Nonconcave Penalized Likelihood Models. *Annal. Statist.*, 36(4), 1509–1533.

پیوست آ

تعاریف و ملزومات

در این پیوست در بخش‌های مختلف ملزومات و تعاریف استفاده‌شده در فصل‌های این رساله، گردآوری شده‌اند.

۱.آ معیارهای نیکویی برازش

در رساله حاضر، سنجش نیکویی برازش مدل‌های مختلف توسط معیارهای متعدد انجام می‌شود؛ از جمله معیار اطلاع آکائیک (AIC)، معیار اطلاع بیزی (BIC) و میانگین توان‌های دوم خطای پیشگویی[1] ($MSPE$). همچنین برای تشخیص کارایی برآوردگرهای مختلف از معیارهایی مانند اریبی برآوردگرها ($Bias$)، انحراف استاندارد برآوردگرها (SD)، میانگین توان دوم خطای (Mse) برآوردگرها و میانگین توان دوم خطا (MSE) برای بردار برآوردگرها استفاده می‌شود. در ادامه هر یک از این معیارها را تعریف می‌کنیم.

تعریف آ.۱.۱. AIC: این معیار که توسط آکائیک (۱۹۷۳) معرفی شد به صورت

$$AIC = ۲m - ۲\ell_{\max}$$

تعریف می‌شود که در آن $\ell_{\max}$ مقدار بیشینه تابع درستنمایی است و m تعداد پارامترهای مدل است.

[1] Mean squared prediction error

تعریف آ.۱.۲. BIC: این معیار که توسط شوارتز (۱۹۷۸) معرفی شد به صورت

$$BIC = m \log n - ۲\ell_{\max}$$

تعریف می‌شود که در آن $\ell_{\max}$ و m مشابه تعریف AIC بوده و n حجم نمونه است.

تعریف آ.۱.۳. $MSPE$: این معیار مدل را بر اساس کارایی آن در پیش‌بینی مقدارهای جدید (مشاهده‌نشده) ارزیابی می‌کند. به عبارت دیگر بر داده‌هایی تکیه دارد که مشاهده شده‌اند ولی در هنگام ساختن مدل به‌کار گرفته نمی‌شوند. این داده‌ها به منظور بررسی و سنجش کارایی مدل برای پیش‌بینی داده‌های جدید به‌کار می‌روند. بدین منظور داده‌ها به دو زیرمجموعه مکمل افراز می‌شوند. بر روی یکی از آن زیرمجموعه‌ها (داده‌های آموزشی) تحلیل انجام می‌شود و بر اساس نتایج حاصل، داده‌های مجموعه دیگر (داده‌های آزمایش) پیشگویی می‌شوند. برای افزایش دقت، عمل پیشگویی چندین بار با افرازهای مختلف انجام و از خطاهای پیشگویی میانگین گرفته می‌شود. اگر فرض کنیم داده‌ها به S زیرمجموعه افراز شده‌اند. از این S زیرمجموعه، هر بار یکی به عنوان داده‌های آزمایش و $S - ۱$ تای دیگر برای آموزش بکار می‌روند. این روال S بار تکرار می‌شود و همه داده‌ها یک بار برای آموزش و یک بار برای آزمایش به‌کار می‌روند. $MSPE$ به صورت

$$MSPE = \frac{۱}{S} \sum_{s=۱}^{S} (\boldsymbol{Y}_{test}^{(-s)} - \widehat{\boldsymbol{Y}}_{test}^{(-s)})^{\top} (\boldsymbol{Y}_{test}^{(-s)} - \widehat{\boldsymbol{Y}}_{test}^{(-s)}),$$

تعریف می‌شود که در آن $\boldsymbol{Y}_{test}^{(-s)}$ داده‌های آزمایشی در افراز sام و $\widehat{\boldsymbol{Y}}_{test}^{(-i)}$ مقادیر پیشگویی بر اساس نتایج کنار گذاشتن داده‌های افراز sام هستند.

تعریف آ.۱.۴. $Bias$: اریبی یک برآوردگر همان اختلاف بین امید ریاضی آن برآوردگر و مقدار واقعی پارامتر برآوردشده است. اریبی برآوردگر $\widehat{\beta}$ به صورت

$$Bias(\widehat{\beta}) = \mathrm{E}(\widehat{\beta}) - \beta$$

تعریف می‌شود. در مطالعات شبیه‌سازی اگر S مجموعه تولیدشده داشته باشیم، $\mathrm{E}(\widehat{\beta})$ به صورت

$$\widehat{\mathrm{E}}(\widehat{\beta}) = \frac{۱}{S} \sum_{s=۱}^{S} \widehat{\beta}^{(s)}$$

تقریب می‌شود که $\widehat{\beta}^{(s)}$ برآورد β بر اساس مجموعه تولیدشده sام است.

تعریف آ.۱.۵. SD: در مطالعات شبیه‌سازی مقاله حاضر، SD برآوردگر $\widehat{\beta}$ بر اساس S تکرار از شبیه‌سازی به صورت

$$SD = \sqrt{\frac{\sum_{s=۱}^{S} \left(\widehat{\beta}^{s} - \mathrm{E}(\widehat{\beta})\right)^{۲}}{S - ۱}}$$

تعریف می‌شود که در آن $\widehat{\mathrm{E}}(\widehat{\beta}) = \frac{1}{S}\sum_{s=1}^{S}\widehat{\beta}^{(s)}$ و $\widehat{\beta}^s$ برآورد β در s-امین مجموعه داده تولیدشده است.

در مجموعه داده‌های واقعی، مشابه روش بیان شده در تعریف $MSPE$، داده‌ها به S زیرمجموعه افراز می‌شوند. هر بار با کنار گذاشتن داده‌های آزمایش، برآوردگرها را بر اساس داده‌های آموزشی به‌دست می‌آوریم. بدین ترتیب تعداد S برآورد برای پارامتر مدل خواهیم داشت. با محاسبه انحراف استاندارد مجموعه برآوردگرهای حاصل، معیار SD موردنظر به‌دست می‌آید که معمولاً با عنوان انحراف استاندارد نمونه‌ای از آن نام می‌بریم.

تعریف آ.۱.۶. Mse: در مطالعات شبیه‌سازی مقاله حاضر، Mse برآوردگر $\widehat{\beta}$ بر اساس S تکرار از شبیه‌سازی به صورت

$$Mse = \frac{1}{S}\sum_{s=1}^{S}(\widehat{\beta}^s - \beta)^2$$

تعریف می‌شود که در آن $\widehat{\beta}^s$ برآورد β در s-امین مجموعه داده تولیدشده است.

تعریف آ.۱.۷. MSE: در مطالعات شبیه‌سازی مقاله حاضر، MSE بر اساس S تکرار از شبیه‌سازی به صورت

$$MSE = \frac{1}{S}\sum_{s=1}^{S}\|\widehat{\boldsymbol{\beta}}^s - \boldsymbol{\beta}\| = \frac{1}{S}\sum_{s=1}^{S}(\widehat{\boldsymbol{\beta}}^s - \boldsymbol{\beta})^\top(\widehat{\boldsymbol{\beta}}^s - \boldsymbol{\beta})$$

تعریف می‌شود که در آن $\widehat{\boldsymbol{\beta}}^s$ برآورد بردار پارامتر $\boldsymbol{\beta}$ در s-امین مجموعه داده تولیدشده است.

آ.۲ تعاریف و نامساوی‌ها

تعریف آ.۱.۲. نرم اقلیدسی ($\|.\|$): یک تابع حقیقی است که بر فضای برداری $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ به صورت

$$\|\boldsymbol{x}\| = \Big(\sum_{i=1}^{n}|x_i|^2\Big)^{1/2}$$

تعریف می‌شود.

تعریف آ.۲.۲. پایه: در جبر خطی منظور از یک پایه برای یک فضای برداری، مجموعه‌ای از بردارهای موجود در آن فضا بوده که دارای دو خاصیت زیر است:

۱. هیچ کدام از بردارهای آن مجموعه را نتوان به صورت ترکیب خطی از بقیه بردارها نوشت، به عبارت دیگر، بردارها به طور خطی از هم مستقل باشند.

۲. هر بردار دیگر در آن فضای برداری را بتوان از ترکیب خطی این بردارها به‌دست آورد.

تعریف آ.۳.۲. همگرایی در توزیع: دنباله متغیرهای تصادفی $\{X_n; n \geqslant ۱\}$ در توزیع به متغیر تصادفی X همگراست اگر برای هر $x \in \mathbb{R}$ که تابع F در آن پیوسته است، داشته باشیم

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

که در آن F_n تابع توزیع تجمعی $\{X_n; n \geqslant ۱\}$ و F تابع توزیع تجمعی X است. این نوع همگرایی را به صورت $X_n \xrightarrow{d} X$ نشان می‌دهند.

تعریف آ.۴.۲. همگرایی در احتمال: دنباله متغیرهای تصادفی $\{X_n\}$ در احتمال به متغیر تصادفی X همگراست اگر برای هر $\varepsilon > ۰$، داشته باشیم

$$\lim_{n\to\infty} P(|X_n - X| > ۰) = ۰.$$

این نوع همگرایی را به صورت $X_n \xrightarrow{P} X$ نشان می‌دهند.

تعریف آ.۵.۲. سازگاری: دنباله برآوردگرهای $\{T_n; n \geqslant ۱\}$ برای $a(\theta)$ سازگار است، اگر $\{T_n\}$ در احتمال به $a(\theta)$ همگرا باشد.

تعریف آ.۶.۲. گوییم برآوردگر $\widehat{\theta}_n$ یک برآوردگر $\sqrt{n}$-سازگار برای θ است، اگر رابطه

$$\widehat{\theta}_n - \theta = O_p(n^{-۱/۲}),$$

برقرار باشد.

تعریف آ.۷.۲. گوییم $f(n)$ از مرتبه $g(n)$ است یا $f(n) = O\big(g(n)\big)$، اگر وقتی $n \to \infty$، آنگاه $\frac{f(n)}{g(n)}$ متناهی باشد. به عبارت دیگر، عدد حقیقی مثبت M و عدد حقیقی $n_۰$ موجود باشند به‌طوری‌که

$$\forall n > n_۰ \qquad |f(n)| \leqslant M|g(n)|.$$

یعنی برای مقادیر به اندازه کافی بزرگ n، قدرمطلق $f(n)$ حداکثر M برابر قدرمطلق $g(n)$ باشد.

تعریف آ.۸.۲. گوییم $f(n)$ از مرتبه کوچک‌تر $g(n)$ است یا $f(n) = o\big(g(n)\big)$، اگر وقتی $n \to \infty$، آنگاه داشته باشیم

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = ۰.$$

به عبارت دیگر $g(n)$ بسیار سریع‌تر از $f(n)$ رشد می‌کند. بنابراین o حالت قوی‌تر O است. هر عبارتی که $o(g)$ باشد، $O(g)$ نیز هست، اما عکس آن برقرار نیست.

تعریف آ.۹.۲. فرض کنید $\{X_n\}$ یک دنباله از متغیرهای تصادفی و $\{a_n\}$ یک دنباله از ثابت‌های متناضر باشد. نماد $X_n = O_p(a_n)$ بدین معنی است که مقادیر $\frac{X_n}{a_n}$ در احتمال کراندار است، یعنی برای هر $\varepsilon > ۰$، اعداد متناهی $M, N > ۰$ وجود دارند به‌طوری‌که

$$\forall n > N, \qquad \mathbb{P}(|\frac{X_n}{a_n}| > M) > \varepsilon.$$

تعریف آ.۱۰.۲. بسط تیلور: فرض کنید که تابع $f(x)$ بر بازه $[a,b]$ تعریف شده و $n+1$ بار مشتق‌پذیر باشد. نقطه $x_0 \in (a,b)$ را درنظر بگیرید. برای هر $x \in (a,b)$ که $x = x_0 + \Delta x$، عددی مانند ξ بین دو نقطه x و x_0 وجود دارد، به‌طوری‌که تساوی

$$f(x_0+\Delta x) = f(x_0) + \Delta x f'(x_0) + \frac{\Delta x^2}{2!} f''(x_0) + \ldots + \frac{\Delta x^n}{n!} f^{(n)}(x_0) + R_n(x_0+\Delta x),$$

برقرار است. که در آن

$$R_n(x_0+\Delta x) = \frac{\Delta x^{n+1}}{(n+1)!} f^{(n+1)}(\xi), \qquad x_0 \leqslant \xi \leqslant x.$$

در این روابط منظور از $f^{(n)}$ مشتق مرتبه nام است. در عمل معمولاً تعداد اندکی از جملات اول سری برای تقریب استفاده می‌شود. در این حالت اگر مقدار Δx به قدر کافی کوچک باشد، تقریب مناسبی از تابع در همسایگی x_0 حاصل می‌شود. در صورتی که تقریب تابع با سری تیلور برای n جمله اول انجام شود، مانده سری یا خطای تقریب متناسب با Δx^n خواهد بود و آن را با $O(\Delta x^n)$ نشان می‌دهند. یعنی

$$f(x_0+\Delta x) = f(x_0) + \Delta x f'(x_0) + \frac{\Delta x^2}{2!} f''(x_0) + \ldots + \frac{\Delta x^n}{n!} f^{(n)}(x_0) + O(\Delta x^n).$$

تعریف آ.۱۱.۲. عملگر بردارساز[2]: در ریاضیات، به ویژه در جبر خطی و نظریه ماتریس، بردارسازی یک ماتریس یک تبدیل خطی است که ماتریس را به یک بردار ستونی تبدیل می‌کند. به عبارت دیگر بردارسازی ماتریس A با بعد $m \times n$ ، که با نماد $\mathrm{vec}(A)$ نمایش داده می‌شود، یک بردار ستونی $mn \times 1$ بعدی است که با انباشتن ستون‌های ماتریس A بر روی یک‌دیگر به‌دست می‌آید. به $\mathrm{vec}(.)$ عملگر بردارساز گویند و به صورت

$$\mathrm{vec}(A) = (a_{1,1}, \ldots, a_{m,1}, a_{1,2}, \ldots, a_{m,2}, \ldots, a_{1,n}, \ldots, a_{m,n})^\top$$

تعریف می‌شود، به‌طوری‌که $a_{i,j}$ عضو مربوط به سطر iام و ستون jام ماتریس A است.

تعریف آ.۱۲.۲. عملگر نیمه‌بردارساز[3]: برای ماتریس متقارن A، بردار $\mathrm{vec}(A)$ حاوی اطلاعات غیرضروری است؛ زیرا به ازای $i \neq j$ عضو a_{ij} برابر عضو a_{ji} است. اگر A یک ماتریس متقارن $n \times n$ بعدی باشد، عملگر نیمه‌بردارساز، $\mathrm{vech}(A)$، با بردارسازی مثلث پایین ماتریس A یک بردار $\frac{n(n+1)}{2} \times 1$ بعدی را نتیجه می‌دهد. به عبارت دیگر

$$\mathrm{vech}(A) = (a_{1,1}, \ldots, a_{n,1}, a_{2,2}, \ldots, a_{n,2}, \ldots, a_{n-1,n-1}, a_{n,n-1}, a_{n,n})^\top.$$

[2] Vectorization

[3] Half-vectorization

تعریف آ.۱۳.۲. ضرب کرونکر: ضرب کرونکر ماتریس‌های $A \in \mathbb{R}^{m_A \times n_A}$ و $B \in \mathbb{R}^{m_B \times n_B}$ به صورت $A \otimes B$ نمایش داده می‌شود و برابر است با:

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n_A}B \\ \vdots & \ddots & \vdots \\ a_{m_A1}B & \dots & a_{m_An_A}B \end{pmatrix}$$

هر $a_{ij}B$ یک بلوک در سایز $m_B \times n_B$ است، بنابراین $A \otimes B$ یک ماتریس $m_Am_B \times n_An_B$ بعدی است.

تعریف آ.۱۴.۲. قضیه لیندبرگ[۴]: متغیرهای تصادفی X_{ij} را درنظر بگیرید که یک آرایه مثلثی به صورت

$$\begin{array}{ccccccc} X_{11} & X_{12} & \dots & X_{1n_1} & & & \\ X_{21} & X_{22} & \dots & \dots & X_{2n_2} & & \\ X_{31} & X_{32} & \dots & \dots & \dots & X_{3n_3} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}$$

را تشکیل می‌دهند. فرض می‌شود که X_{ij}ها در شروط زیر که شروط آرایه مثلثی نامیده می‌شود، صدق می‌کنند.

۱. برای هر i، n_i متغیر تصادفی $X_{i1}, X_{i2}, \dots, X_{in_i}$ در سطر iام، دو به دو از هم مستقل‌اند.

۲. به ازای تمام iها و jها، $\mathrm{E}(X_{ij}) = 0$.

۳. به ازای تمام iها، $\sum_j \mathrm{E}(X_{ij}^2) = 1$.

دنباله تصادفی که از مجموع عناصر سطری آرایه مثلثی به دست می‌آید را به صورت $S_i = \sum_j X_{ij}$ تعریف می‌کنیم. فرض می‌کنیم $L(s_i)$ تابع توزیع این دنباله باشد. به‌علاوه فرض می‌کنیم X_{ij}ها در شرط لیندبرگ

$$\forall \epsilon > 0, \quad \lim_{i \to \infty} \sum_{j=1}^{n_i} \mathrm{E}\left(X_{ij}^2 I(|X_{ij}| > \epsilon)\right) = 0$$

صدق می‌کنند. آن‌گاه داریم

$$\lim_{i \to \infty} L(S_i) \to \mathcal{N}(0, 1).$$

تعریف آ.۱۵.۲. نامساوی مارکف: فرض کنید X_n یک متغیر تصادفی با مقادیر مثبت باشد، یعنی $\mathbb{P}(X_n \geqslant 0) = 1$. آن‌گاه این نامساوی بیان می‌کند که

$$\mathbb{P}(X_n \geqslant c) \leqslant \frac{\mathrm{E}(X_n)}{c}.$$

این نامساوی یک کران بالا برای مقدار احتمال ارائه می‌دهد.

[۴] Lindeberg

تعریف آ.۱۶.۲. نامساوی جنسن[۵]: فرض کنید g یک تابع محدب در بازه (a,b) باشد و امید ریاضی متغیر تصادفی X_n با ویژگی $\mathbb{P}\Big(X_n \in (a,b)\Big) = ۱$، وجود داشته‌باشد. آن‌گاه نامساوی جنسن به صورت زیر برقرار است.

$$g\Big(\mathrm{E}(X_n)\Big) \leqslant \mathrm{E}\Big(g(X_n)\Big).$$

تعریف آ.۱۷.۲. نامساوی کوشی‌شوارتز[۶]: این نامساوی بیان می‌کند که برای هر دو متغیر تصادفی X_n و Y_n، داریم

$$\mathrm{E}^{۲}(X_nY_n) \leqslant \mathrm{E}(X_n^{۲})\,\mathrm{E}(Y_n^{۲}).$$

هم‌چنین با گرفتن ریشه دوم طرفین نامساوی فوق داریم

$$\Big|\mathrm{E}(X_nY_n)\Big| \leqslant \mathrm{E}^{۱/۲}(X_n^{۲})\,\mathrm{E}^{۱/۲}(Y_n^{۲}).$$

تعریف آ.۱۸.۲. الگوریتم نیوتن–رافسون: در آنالیز عددی این روش یک الگوریتم ریشه‌یابی است که تقریب‌های خوبی در نزدیکی ریشه یک تابع می‌زند. در پایه‌ای‌ترین حالت، الگوریتم نیوتن برای یک تابعی چون f با متغیر x و با مشتق f' به همراه حدس اولیه $x_۰$ به‌کار می‌رود. شکل عمومی الگوریتم به شرح زیر می‌باشد:

$$x_۱ = x_۰ - \frac{f(x_۰)}{f'(x_۰)}$$

اگر تابع حدس کافی و دقیقی را برآورد سازد و همچنین حدس اولیه نزدیک به ریشه تابع مفروض باشد آن‌گاه $x_۱$ تقریب بهتری نسبت به $x_۰$ به حساب می‌آید. چرا که با احتساب همگرایی جواب‌ها، هر تقریب نسبت به تقریب قبل از خودش از دقت بالاتری برخوردار بوده و به ریشه تابع نزدیک‌تر است.

[5] Jensen

[6] Cauchy–Schwarz

پیوست ب
اثبات قضایا

در این پیوست اثبات قضایایی که در متن رساله (فصل‌های گذشته) آمده، به تفصیل آورده شده است.

ب.۱ اثبات قضایای فصل ۱

ب.۱.۱ اثبات قضیه ۱.۵.۱

برای برآوردگر هسته‌ای، $\widehat{\boldsymbol{\beta}}_K$، داریم

$$\sqrt{n}\Big(\widehat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}\Big) = D_n^{-1}\{\sqrt{n}C_n\} + o_p(1),$$

به‌طوری‌که

$$D_n = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1}\tilde{\boldsymbol{X}}_i, \quad \text{و} \quad C_n = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1}[\boldsymbol{Y}_i - (\boldsymbol{X}_i^\top\boldsymbol{\beta} + \widehat{\boldsymbol{f}}(\boldsymbol{t}_i))].$$

فرض کنید $D = \lim_{n\longrightarrow\infty} D_n = E\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}\tilde{\boldsymbol{X}}\}$. با اندکی محاسبات ساده می‌توان نشان داد $C_n = C_{1n} - C_{2n} + o_p(1)$ که در آن C_{1n} و C_{2n} به ترتیب به صورت

$$C_{1n} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1}[\boldsymbol{Y}_i - \boldsymbol{\mu}_i] = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{W}_{1i}^\top[\boldsymbol{Y}_i - \boldsymbol{\mu}_i]$$

و

$$C_{۲n} = \frac{۱}{n}\sum_{i=۱}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-۱}[\hat{\boldsymbol{f}}(\boldsymbol{t}_i,\boldsymbol{\beta}) - \boldsymbol{f}(\boldsymbol{t}_i)]$$

تعریف می‌شوند، به‌طوری‌که $\boldsymbol{\mu}_i = \boldsymbol{X}_i^\top\boldsymbol{\beta} + \boldsymbol{f}(\boldsymbol{t}_i)$ و $\boldsymbol{W}_{۱i} = \boldsymbol{V}_i^{-۱}\tilde{\boldsymbol{X}}_i$. مشابه اثبات صورت‌گرفته در لین و کارول (۲۰۰۲)، می‌توان نشان داد که

$$C_{۲n} = \frac{۱}{n}\sum_{i=۱}^{n} \boldsymbol{W}_{۲i}^\top[\boldsymbol{Y}_i - \boldsymbol{\mu}_i] + \frac{h^۲}{۲}E\{\boldsymbol{X}^\top\boldsymbol{V}^{-۱}\boldsymbol{f}^{۲}(\boldsymbol{t})\} + o_p(۱),$$

به‌طوری‌که $\boldsymbol{W}_{۲i} = \{\boldsymbol{W}_{۲i۱},\ldots,\boldsymbol{W}_{۲im}\}^\top$ و

$$\boldsymbol{W}_{۲ij} = \frac{\{\sum_{k=۱}^{m}\sum_{l=۱}^{m} E(\tilde{\boldsymbol{X}}_k\boldsymbol{V}^{kl}|\boldsymbol{t}_l = t_{ij})\}f_j(t_{ij}))}{\sum_{l=۱}^{m} f_l(t_{ij}))}.$$

بنابراین

$$\sqrt{n}\Big(\widehat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}\Big) = D^{-۱}\frac{۱}{\sqrt{n}}\sum_{i=۱}^{n}(\boldsymbol{W}_{۱i} - \boldsymbol{W}_{۲i})(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) + \sqrt{nh^{۴}}\boldsymbol{b}_K(\boldsymbol{\beta},\boldsymbol{f})/۲ + o_p(۱),$$

که $\boldsymbol{b}_K(\boldsymbol{\beta},\boldsymbol{g}) = D^{-۱}E\{\tilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-۱}\boldsymbol{g}^{(۲)}(\boldsymbol{t})\}$ با این فرض که $\boldsymbol{V}_o = \mathrm{Var}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{t})$ و همچنین $\boldsymbol{V}_K = D^{-۱}\,\mathrm{E}\,\{(\boldsymbol{W}_۱ - \boldsymbol{W}_۲)^\top\boldsymbol{V}_o(\boldsymbol{W}_۱ - \boldsymbol{W}_۲)\}D^{-۱}$، نتیجه موردنظر حاصل می‌شود.

ب.۲.۱ اثبات قضیه ۳.۵.۱

می‌خواهیم ثابت کنیم $\boldsymbol{V}_{BF} - \boldsymbol{V}_K$ یک ماتریس نیمه معین مثبت است. جزء اصلی ماتریس $\boldsymbol{V}_{BF}$ به صورت $\boldsymbol{X}^\top\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\varepsilon^*$ است که $\varepsilon^* = \boldsymbol{Z}^\top\boldsymbol{b} + \varepsilon$. همچنین جزء اصلی ماتریس $\boldsymbol{V}_K$ به صورت $\tilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\varepsilon^*$ است. برای مقایسه $\boldsymbol{X}^\top(\boldsymbol{I}-\boldsymbol{S})^\top\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\varepsilon^*$ و یا به عبارتی به صورت دو ماتریس $\boldsymbol{V}_{BF}$ و $\boldsymbol{V}_K$ کافی است که اجزای اصلی آن‌ها را مقایسه کنیم. اکنون نشان می‌دهیم که

$$\mathrm{Cov}\{\boldsymbol{X}^\top\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\varepsilon^*\} \geqslant \mathrm{Cov}\{\tilde{\boldsymbol{X}}^\top\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\varepsilon^*\}.$$

برای برآوردگر بازگشتی داریم

$$\begin{aligned}\mathrm{Cov}\{\boldsymbol{X}^\top\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\varepsilon^*\} &= \mathrm{E}\Big\{\boldsymbol{X}^\top\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{V}_o(\boldsymbol{I}-\boldsymbol{S})^\top\boldsymbol{V}^{-۱}\boldsymbol{X}|\boldsymbol{t}\Big\} \qquad (ب.۱)\\ &= \boldsymbol{\mu}_{\boldsymbol{X}}^\top(\boldsymbol{t})\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{V}_o(\boldsymbol{I}-\boldsymbol{S})^\top\boldsymbol{V}^{-۱}\boldsymbol{\mu}_{\boldsymbol{X}}(\boldsymbol{t})\\ &\quad + \mathrm{tr}\left[\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{V}_o(\boldsymbol{I}-\boldsymbol{S})^\top\boldsymbol{V}^{-۱}\,\mathrm{Cov}(\boldsymbol{X}|\boldsymbol{t})\right].\end{aligned}$$

جمله اول عبارت فوق نیمه‌معین مثبت است زیرا $\boldsymbol{V}^{-۱}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{V}_o(\boldsymbol{I}-\boldsymbol{S})^\top\boldsymbol{V}^{-۱}$ یک ماتریس نیمه‌معین مثبت است. همچنین عبارت $\boldsymbol{\mu}_X(\boldsymbol{t})$ مقداری غیرصفر است. از طرفی برای برآوردگر

هسته‌ای داریم

$$\begin{aligned}\mathrm{Cov}\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{\varepsilon}^*\} &= \mathrm{E}\left\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{V}_o(\boldsymbol{I}-\boldsymbol{S})^\top \boldsymbol{V}^{-1}\tilde{\boldsymbol{X}}|\boldsymbol{t}\right\} \qquad (\text{ب}.۲)\\ &= \mathrm{E}(\tilde{\boldsymbol{X}}|\boldsymbol{t})^\top \boldsymbol{V}^{-1}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{V}_o(\boldsymbol{I}-\boldsymbol{S})^\top \boldsymbol{V}^{-1}\,\mathrm{E}(\tilde{\boldsymbol{X}}|\boldsymbol{t})\\ &+ \mathrm{tr}\left[\boldsymbol{V}^{-1}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{V}_o(\boldsymbol{I}-\boldsymbol{S})^\top \boldsymbol{V}^{-1}\,\mathrm{tr}(\tilde{\boldsymbol{X}}|\boldsymbol{t})\right].\end{aligned}$$

با توجه به این که

$$\mathrm{E}(\tilde{\boldsymbol{X}}_i|\boldsymbol{t}_i) = \mathrm{E}\left\{\boldsymbol{X}_i - \boldsymbol{\mu}_X(t_i)|\boldsymbol{t}_i\right\} = 0, \quad \mathrm{Cov}(\tilde{\boldsymbol{X}}_i|\boldsymbol{t}_i) = \mathrm{Cov}\left\{\boldsymbol{X}_i - \boldsymbol{\mu}_X(\boldsymbol{t}_i)|\boldsymbol{t}_i\right\} = \mathrm{Cov}(\boldsymbol{X}_i|\boldsymbol{t}_i).$$

بنابراین جمله اول در عبارت (ب.۲) برابر ۰ بوده و جملات دوم در عبارات (ب.۱) و (ب.۲) برابر هستند. لذا نتیجه می‌گیریم که $\mathrm{Cov}\{\boldsymbol{X}^\top \boldsymbol{V}^{-1}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{\varepsilon}^*\} \geqslant \mathrm{Cov}\{\tilde{\boldsymbol{X}}^\top \boldsymbol{V}^{-1}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{\varepsilon}^*\}$، و نتیجه مورد نظر حاصل می‌شود.

ب.۲ اثبات قضایای فصل ۲

ب.۲.۱ بردار رتبه و ماتریس اطلاع فیشر

مشتق اول جزئی از لگاریتم تابع درستنمایی تاوانیده (۱۵.۲) نسبت به هر یک از اعضای پارامتر $\boldsymbol{\Theta}$ را بردار رتبه ($\boldsymbol{S}_\Theta$) می‌گویند که اعضای آن شامل

$$\begin{aligned}\boldsymbol{S}_{\boldsymbol{\theta}} &= \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^\top \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta}) - n\boldsymbol{E}_n(\boldsymbol{\theta}),\\ [\boldsymbol{S}_{\boldsymbol{\eta}}]_l &= -\frac{1}{2}\sum_{i=1}^{n}\left\{\mathrm{tr}(\boldsymbol{V}_i^{-1}\overset{\circ}{\boldsymbol{V}}_{il}) - (\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta})^\top \boldsymbol{V}_i^{-1}\overset{\circ}{\boldsymbol{V}}_{il}\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta})\right\},\end{aligned}$$

است، که در آن $l = 1,\ldots,g$، $g = \frac{q(q+1)}{2} + \frac{m(m+1)}{2}$ ،

$$\overset{\circ}{\boldsymbol{V}}_{il} = \frac{\partial \nabla_i(\boldsymbol{D},\boldsymbol{\Sigma})}{\partial \boldsymbol{\eta}_l} = \begin{cases} \boldsymbol{Z}_i \frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\eta}_l}\boldsymbol{Z}_i^\top & \text{اگر } \boldsymbol{\eta}_l = \mathrm{vec}(\boldsymbol{D}) \\ \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\eta}_l} & \text{اگر } \boldsymbol{\eta}_l = \mathrm{vec}(\boldsymbol{\Sigma}) \end{cases}$$

و

$$\Delta_i(\boldsymbol{D},\boldsymbol{\Sigma}) = (\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} - \boldsymbol{Z}_i\boldsymbol{b}_i)^\top \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\theta} - \boldsymbol{Z}_i\boldsymbol{b}_i).$$

همچنین امید مشتق دوم جزئی از لگاریتم تابع درستنمایی تاوانیده (۱۵.۲) نسبت به هر یک از اعضای پارامتر $\boldsymbol{\Theta}$ را ماتریس اطلاع فیشر ($\boldsymbol{J}_\Theta$) می‌گویند که اعضای آن به صورت

$$\boldsymbol{J}_{\Theta\Theta} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \boldsymbol{J}_{\boldsymbol{\theta}\boldsymbol{\eta}} \\ \boldsymbol{J}_{\boldsymbol{\eta}\boldsymbol{\theta}}^\top & \boldsymbol{J}_{\boldsymbol{\eta}\boldsymbol{\eta}} \end{bmatrix},$$

تعیین می‌شود، به طوری که

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\theta\eta}} &= \mathbf{0}, \\
\boldsymbol{J}_{\boldsymbol{\theta\theta}} &= \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^{\top} \boldsymbol{V}_i^{-1} \tilde{\boldsymbol{X}}_i - n \frac{\partial \boldsymbol{E}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\
[\boldsymbol{J}_{\boldsymbol{\eta\eta}}]_{ls} &= \frac{1}{2} \sum_{i=1}^{n} \Big(\operatorname{tr}\left(\boldsymbol{V}_i^{-1} \overset{\circ}{\boldsymbol{V}}_{il} \boldsymbol{V}_i^{-1} \overset{\circ}{\boldsymbol{V}}_{is}\right) - \operatorname{tr}\left(\boldsymbol{V}_i^{-1} \overset{\circ}{\boldsymbol{V}}_{il}\right) \operatorname{tr}\left(\boldsymbol{V}_i^{-1} \overset{\circ}{\boldsymbol{V}}_{is}\right) \Big),
\end{aligned}$$

و $l, s = 1, \ldots, g$. به‌علاوه، سایر ماتریس‌های اطلاع مورد نیاز در اثبات قضایا به صورت‌های

$$\boldsymbol{J}_{\boldsymbol{\theta\theta}} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} & \boldsymbol{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2} \\ \boldsymbol{J}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_1}^{\top} & \boldsymbol{J}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} \end{bmatrix}, \qquad \boldsymbol{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} & \boldsymbol{J}_{\boldsymbol{\beta}_1\boldsymbol{\alpha}} \\ \boldsymbol{J}_{\boldsymbol{\alpha}\boldsymbol{\beta}_1}^{\top} & \boldsymbol{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \end{bmatrix}$$

قابل تفکیک هستند و اعضای آن‌ها به صورت

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} &= \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_{i1}^{\top} \boldsymbol{V}_{i11}^{-1} \tilde{\boldsymbol{X}_{i1}} - n \frac{\partial \boldsymbol{E}_n(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}, \\
\boldsymbol{J}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2} &= \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_{i2}^{\top} \boldsymbol{V}_{i22}^{-1} \tilde{\boldsymbol{X}_{i2}} - n \frac{\partial \boldsymbol{E}_n(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2}, \\
\boldsymbol{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2} &= \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_{i1}^{\top} \boldsymbol{V}_{i12}^{-1} \tilde{\boldsymbol{X}_{i2}}, \\
\boldsymbol{J}_{\boldsymbol{\beta}_1\boldsymbol{\beta}_1} &= \sum_{i=1}^{n} \boldsymbol{X}_{i1}^{\top} \boldsymbol{V}_{i11}^{-1} \boldsymbol{X}_{i1} - n \frac{\partial \boldsymbol{E}_n(\boldsymbol{\beta}_1)}{\partial \boldsymbol{\beta}_1}, \\
\boldsymbol{J}_{\boldsymbol{\beta}_1\boldsymbol{\alpha}} &= \sum_{i=1}^{n} \boldsymbol{X}_{i1}^{\top} \boldsymbol{V}_{i11}^{-1} \boldsymbol{B}_i, \\
\boldsymbol{J}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} &= \sum_{i=1}^{n} \boldsymbol{B}_i^{\top} \boldsymbol{V}_{i11}^{-1} \boldsymbol{B}_i;
\end{aligned}$$

تعیین می‌شوند، که در آن‌ها $\boldsymbol{V}_{i11}$، $\boldsymbol{V}_{i22}$ و $\boldsymbol{V}_{i12}$ اعضای ماتریس

$$\boldsymbol{V}_i = \begin{bmatrix} \boldsymbol{V}_{i11} & \boldsymbol{V}_{i12} \\ \boldsymbol{V}_{i12} & \boldsymbol{V}_{i22} \end{bmatrix}$$

هستند.

ب.۲.۲ اثبات قضیه ۱.۳.۲

فرض کنید $\alpha_n = \sqrt{p_n}(n^{-1/2} + a_n)$ و $\|U\| = C$، به طوری که C یک مقدار ثابت و به اندازه کافی بزرگ است. می‌خواهیم نشان دهیم که به ازای هر $\xi > 0$ یک مقدار ثابت C وجود دارد که

$$\mathbb{P}\Big\{ \inf_{\|U\|=C} \ell_c^{Pen}(\boldsymbol{\theta}_{n0} + \alpha_n U) \geqslant \ell_c^{Pen}(\boldsymbol{\theta}_{n0}) \Big\} \geqslant 1 - \xi.$$

یعنی با احتمال نزدیک به ۱، یک جواب محلی برای $\ell_c^{Pen}(\boldsymbol{\theta}_{n\circ})$ در مجموعه $\{\boldsymbol{\theta}_{n\circ}+\alpha_n U:\ \|U\|\leqslant C\}$ وجود دارد. با استفاده از $p_{\lambda_n}(\circ)=\circ$، داریم

$$\begin{aligned} D_n(U) &= \ell_c^{Pen}(\boldsymbol{\theta}_{n\circ}+\alpha_n U)-\ell_c^{Pen}(\boldsymbol{\theta}_{n\circ}) \\ &\geqslant \ell_c(\boldsymbol{\theta}_{n\circ}+\alpha_n U)-\ell_c(\boldsymbol{\theta}_{n\circ})+n\sum_{i=1}^{n}\{p_{\lambda_n}(|\theta_{n\circ,k}+\alpha_n U_k|)-p_{\lambda_n}(|\theta_{n\circ,k}|)\} \\ &\equiv I+II. \end{aligned}$$

با اندکی محاسبات ساده داریم

$$\begin{aligned} I &= \frac{n\alpha_n^2}{2}U^\top\hat{\Gamma}_n U-\alpha_n U^\top\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_i-\tilde{\boldsymbol{X}}_i\boldsymbol{\theta}-\boldsymbol{Z}_i\boldsymbol{b}_i) \\ &\equiv I_1-I_2, \end{aligned}$$

که در آن $\Gamma_n=n^{-1}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i^\top\boldsymbol{\Sigma}_i^{-1}\tilde{\boldsymbol{X}}_i$، $\hat{\Gamma}_n=n^{-1}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i^\top\hat{\boldsymbol{\Sigma}}_i^{-1}\tilde{\boldsymbol{X}}_i$ و

$$\begin{aligned} I_1 &= \frac{n\alpha_n^2}{2}U^\top(\Gamma_n-\hat{\Gamma}_n+\hat{\Gamma}_n)U \\ &= \frac{n\alpha_n^2}{2}U^\top\hat{\Gamma}_n U+O_p(n^{-1/2})n\alpha_n^2\|U\|^2, \end{aligned}$$

$$I_2 = \alpha_n U^\top\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i.$$

با استفاده از شرط (۳.A) می‌توان نشان داد که

$$\begin{aligned} |\alpha_n U^\top\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\epsilon}_i| &\leqslant \alpha_n\|\alpha_n\|\|U\| \\ &= O_p(\alpha_n\sqrt{np_n})\|U\|=O_p(\alpha_n^2 n)\|U\|. \end{aligned}$$

بنابراین $I_2=O_p(\alpha_n^2 n)\|U\|$. مشابه اثبات فن و لی (۲۰۰۱) می‌توان نشان داد که

$$II=O_p(\sqrt{s_n}n\alpha_n a_n)\|U\|+O_p(n\alpha_n^2 b_n)\|U\|.$$

بر اساس نتایج بدست آمده، شرط (۳.A) و اینکه $\|U\|$ به اندازه کافی بزرگ باشد، می‌توان دریافت که عبارت‌های I_2 و II تحت الشعاع عبارت I_1 می‌باشند که این عبارت مقداری مثبت است. بنابراین قسمت (i) قضیه ۱.۳.۲ اثبات می‌شود.

برای اثبات قسمت (ii) قضیه ۱.۳.۲، با استفاده از قضیه ۱۲.۷ در اسچوماکر (۱۹۸۱) تحت شرط (۱.A) مقدار ثابتی مانند C وجود دارد به طوری که

$$\sup_{t\in[\circ,1]}|f_\circ(\boldsymbol{t})-\boldsymbol{B}(\boldsymbol{t})\boldsymbol{\alpha}_\circ|\leqslant CK_n^{-m}.$$

حال با انتخاب $K_n \approx n^{۱/(۲r+۱)}$ داریم

$$\begin{aligned}\big(\widehat{f}(t)-f_{\circ}(t)\big)^{۲} &= \big(\boldsymbol{B}(t)\widehat{\boldsymbol{\alpha}}_n - f_{\circ}(t)\big)^{۲} \simeq \big(\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ} - f_{\circ}(\boldsymbol{t})\big)^{۲}\\ &\leqslant |\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ} - f_{\circ}(t)||\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ} - f_{\circ}(t)|\\ &\leqslant \sup_{t\in[\circ,۱]}|\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ} - f_{\circ}(t)| \sup_{t\in[\circ,۱]}|\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ} - f_{\circ}(t)|\\ &\leqslant CL_n^{-r}CL_n^{-r} = C^{۲}n^{-۲r/(۲r+۱)} = O_p(n^{-۲r/(۲r+۱)}).\end{aligned}$$

بنابراین قسمت (ii) قضیه ۱.۳.۲ اثبات می‌شود.

ب.۳.۲ اثبات قضیه ۲.۳.۲

برای اثبات قسمت (i)، باید نشان داد که به ازای هر $\boldsymbol{\theta}_{n۱}$ ای که $\|\boldsymbol{\theta}_{n۱} - \boldsymbol{\theta}_{n\circ ۱}\| = O_p(\sqrt{p_n/n})$ و هر مقدار ثابت C، $(\boldsymbol{\theta}_{n۱}^{\top}, \mathbf{0})^{\top}$ جواب بهینه تابع درستنمایی تاوانیده است. به عبارت دیگر

$$\ell_c^{Pen}\{(\boldsymbol{\theta}_{n۱}^{\top}, \mathbf{0})^{\top}\} = \max_{\|\boldsymbol{\theta}_{n۲}\|\leqslant C\sqrt{p_n/n}} \ell_c^{Pen}\{(\boldsymbol{\theta}_{n۱}^{\top}, \boldsymbol{\theta}_{n۲}^{\top})^{\top}\}.$$

برای اثبات عبارت فوق کافی است ثابت کنیم که

$$\begin{aligned}\frac{\partial \ell_c^{Pen}\{(\boldsymbol{\theta}_n)}{\partial \theta_{n,k}}, &< \circ \qquad \forall\ \circ < \theta_{n,k} < \xi_n;\\ \frac{\partial \ell_c^{Pen}\{(\boldsymbol{\theta}_n)}{\partial \theta_{n,k}}, &> \circ \qquad \forall\ -\xi_n < \theta_{n,k} < \circ.\end{aligned}$$

به ازای هر $\theta_{n,k} \neq \circ$ و $k = s_n + ۱, \ldots, p_n$ برقرار است، به طوری که $\xi_n = C\sqrt{p_n/n}$. واضح است که

$$\frac{\partial \ell_c^{Pen}\{(\boldsymbol{\theta}_n)}{\partial \theta_{n,k}} = \frac{\partial \ell_c\{(\boldsymbol{\theta}_n)}{\partial \theta_{n,k}} - np'_{\lambda_n}(|\theta_{n,k}|)\,\mathrm{sgn}(\theta_{n,k}),$$

و مشابه فن و لی (۲۰۰۱) می‌توان نشان که

$$\frac{\partial \ell_c\{(\boldsymbol{\theta}_n)}{\partial \theta_{n,k}} = O_p(\sqrt{np_n}).$$

بنابراین می‌توان عبارت

$$\frac{\partial \ell_c^{Pen}\{(\boldsymbol{\theta}_n)}{\partial \theta_{n,k}} = n\lambda_n\Big\{ -\frac{p'_{\lambda_n}(|\theta_{n,k}|)}{\lambda_n}\,\mathrm{sgn}(\theta_{n,k}) + O_p\big(\frac{\sqrt{p_n/n}}{\lambda_n}\big)\Big\}$$

را نتیجه گرفت. از آنجایی که $\frac{\sqrt{p_n/n}}{\lambda_n} \to \circ$ و $\liminf_{n\to\infty}\liminf_{\theta\to\circ_+} p'_{\lambda_n}(\theta)/\lambda_n > \circ$، علامت $\theta_{n,k}$ تعیین‌کننده علامت $\frac{\partial \ell_c^{Pen}\{(\boldsymbol{\theta}_n)}{\partial \theta_{n,k}}$ است. بنابراین اثبات کامل می‌شود.

بخش همگرایی در احتمال و همگرایی در توزیع مستقیما از طریق قانون ضعیف اعداد بزرگ، قضیه حد مرکزی و قضیه اسلاتسکی اثبات می‌شود.

ب.۳ اثبات قضایای فصل ۳

فرض کنید C یک مقدار ثابت و مثبت باشد، که در جاهای مختلف این بخش ممکن است مقادیر مختلفی داشته باشد. در این بخش $\overline{\boldsymbol{S}}_n(\boldsymbol{\theta})$ را به صورت $\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}) = \big(\overline{S}_{n1}(\boldsymbol{\theta}), \ldots, \overline{S}_{np_n}(\boldsymbol{\theta}_n)\big)^\top$ تعریف می‌شود به طوری که $\overline{S}_{nk}(\boldsymbol{\theta}_n) = \boldsymbol{e}_k^\top \overline{\boldsymbol{S}}_n(\boldsymbol{\theta})$ و $\boldsymbol{e}_k$، $1 \leqslant k \leqslant p_n$ یک بردار p_n بعدی است که عضو kام آن برابر ۱ بوده و سایر اعضا برابر ۰ هستند. همچنین مشتق $\overline{S}_{nk}(\boldsymbol{\theta}_n)$ نسبت به $\boldsymbol{\theta}_n$ به صورت $\overline{D}_{nk}(\boldsymbol{\theta}_n) = \frac{\partial}{\partial \boldsymbol{\theta}_n^\top} \overline{S}_{nk}(\boldsymbol{\theta}_n)$ تعریف می‌شود و داریم

$$\overline{D}_{nk}(\boldsymbol{\theta}_n) = \overline{H}_{nk}(\boldsymbol{\theta}_n) + \overline{E}_{nk}(\boldsymbol{\theta}_n) + \overline{G}_{nk}(\boldsymbol{\theta}_n),$$

که در آن $\overline{H}_{nk}$، $\overline{E}_{nk}$ و $\overline{G}_{nk}$ عضو kام $\overline{\boldsymbol{H}}_n$، $\overline{\boldsymbol{E}}_n$ و $\overline{\boldsymbol{G}}_n$ هستند که به طور مشابه از مشتق $\overline{\boldsymbol{S}}_n(\boldsymbol{\theta})$ نسبت به $\boldsymbol{\theta}_n$ استخراج شده‌اند. در ادامه لم‌هایی را که در اثبات قضایا استفاده می‌شوند، بیان می‌کنیم.

ب.۱.۳ لم‌ها

ایده اصلی برای اثبات سازگاری برآوردگرها این است که $\boldsymbol{S}_n(\boldsymbol{\theta})$ را توسط $\overline{\boldsymbol{S}}_n(\boldsymbol{\theta})$ تقریب بزنیم؛ زیرا ارزیابی گشتاورهای آن آسان‌تر است. لم ب.۱.۳ صحت این تقریب را تعیین می‌کند، همچنین این لم نقش مهمی در اثبات نرمال مجانبی بودن توزیع برآوردگرها در قضیه ۳.۱.۳ دارد.

لم ب.۱.۳. تحت شرایط نظم، اگر $n^{-1}p_n^2 = o(1)$، آن‌گاه

$$\|\boldsymbol{U}_n(\boldsymbol{\theta}_{n0}) - \overline{\boldsymbol{U}}_n(\boldsymbol{\theta}_{n0})\| = \|\boldsymbol{S}_n(\boldsymbol{\theta}_{n0}) - \overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n0})\| = O_p(p_n).$$

برای سادگی در یافتن بسط تیلور تابع برآوردیابی $\boldsymbol{S}_n(\boldsymbol{\theta}_n)$، از $\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n) = \frac{\partial}{\partial \boldsymbol{\theta}_n^\top} \overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_n)$ برای تقریب تابع گرادیانت، $\boldsymbol{D}_n(\boldsymbol{\theta}_n) = \frac{\partial}{\partial \boldsymbol{\theta}_n^\top} \boldsymbol{S}_n(\boldsymbol{\theta}_n)$، استفاده می‌شود. لم ب.۲.۳ نمایش کاربردی برای $\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n)$ ارائه می‌دهد.

لم ب.۲.۳.

$$\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n) = \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n) + \overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n) + \overline{\boldsymbol{G}}_n(\boldsymbol{\theta}_n), \qquad (ب.۳)$$

به طوری که

$$\begin{aligned}
\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n) &= -n^{-1}\,\mathrm{E}_{u|y}\Big[\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{D}_i\Big],\\
\overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n) &= -(2n)^{-1}\,\mathrm{E}_{u|y}\Big[\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{\frac{-3}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{C}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{F}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{D}_i\Big],\\
\overline{\boldsymbol{G}}_n(\boldsymbol{\theta}_n) &= (2n)^{-1}\,\mathrm{E}_{u|y}\Big[\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{F}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{J}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{D}_i\Big],
\end{aligned}$$

و

$$\begin{aligned}
\boldsymbol{C}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i) &= \operatorname{diag}\Big(\boldsymbol{y}_{i1}-\boldsymbol{\mu}_{i1}(\boldsymbol{\theta}_n,\boldsymbol{u}_i),\ldots,\boldsymbol{y}_{im}-\boldsymbol{\mu}_{im}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\Big),\\
\boldsymbol{F}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i) &= \operatorname{diag}\Big(\boldsymbol{\mu}^{\cdot\cdot}(\boldsymbol{D}_{i1}^{\top}\boldsymbol{\theta}_n+\boldsymbol{Z}_{i1}^{\top}\boldsymbol{u}_i),\ldots,\boldsymbol{\mu}^{\cdot\cdot}(\boldsymbol{D}_{im}^{\top}\boldsymbol{\theta}_n+\boldsymbol{Z}_{im}^{\top}\boldsymbol{u}_i)\Big),\\
\boldsymbol{J}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i) &= \operatorname{diag}\Big(\overline{\boldsymbol{R}}_i^{-1}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big)\Big),
\end{aligned}$$

در عبارت‌های فوق، هر سه نماد $a=(a_1,\ldots,a_{n_i})^{\top}$، $\operatorname{diag}(a_1,\ldots,a_m)$ و $\operatorname{diag}(\boldsymbol{a})$، یک ماتریس قطری $m\times m$ بعدی است که عناصر قطر اصلی آن $(a_1,\ldots,a_m)$ هستند.

لم ب.۳.۳. تحت شرایط نظم، اگر $n^{-1}p_n^2=o(1)$، به ازای هر $\Delta>0$ و $\boldsymbol{b}_n\in\mathcal{R}^{p_n}$، داریم

$$\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n0}\|\leqslant\Delta\sqrt{p_n/n}}\ \sup_{\|\boldsymbol{b}_n\|=1}\Big|\boldsymbol{b}_n^{\top}\big[\boldsymbol{D}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n)\big]\boldsymbol{b}_n\Big|=O_p(\sqrt{np_n}).$$

ماتریس $\boldsymbol{D}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n)$ یک ماتریس متقارن است. لم فوق گزاره زیر را نتیجه می‌دهد.

نتیجه ب.۱.۳.

$$\begin{aligned}
&\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n0}\|\leqslant\Delta\sqrt{p_n/n}}\Big|\lambda_{\min}\big[\boldsymbol{D}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n)\big]\Big|=O_p(\sqrt{np_n}),\\
&\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n0}\|\leqslant\Delta\sqrt{p_n/n}}\Big|\lambda_{\max}\big[\boldsymbol{D}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n)\big]\Big|=O_p(\sqrt{np_n}).
\end{aligned}$$

لم ب.۴.۳. تحت شرایط نظم، اگر $n^{-1}p_n^2=o(1)$، آن‌گاه به ازای هر $\Delta>0$ و $\boldsymbol{b}_n\in\mathcal{R}^{p_n}$، داریم

$$\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n0}\|\leqslant\Delta\sqrt{p_n/n}}\ \sup_{\|\boldsymbol{b}_n\|=1}\Big|\boldsymbol{b}_n^{\top}\big[\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n)\big]\boldsymbol{b}_n\Big|=O_p(\sqrt{n}p_n).$$

لم ب.۵.۳. تحت شرایط نظم، اگر $n^{-1}p_n^2=o(1)$، آن‌گاه به ازای هر $\Delta>0$ و $\boldsymbol{b}_n\in\mathcal{R}^{p_n}$، داریم

$$\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n0}\|\leqslant\Delta\sqrt{p_n/n}}\ \sup_{\|\boldsymbol{b}_n\|=1}\Big|\boldsymbol{b}_n^{\top}\big[\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n0})\big]\boldsymbol{b}_n\Big|=O_p(\sqrt{n}p_n).$$

لم ب.۶.۳. تحت شرایط نظم، اگر $n^{-1}p_n^3=o(1)$، آن‌گاه به ازای هر $\boldsymbol{\xi}_n\in\mathcal{R}^{p_n}$ که $\|\boldsymbol{\xi}_n\|=1$، داریم

$$\boldsymbol{\xi}_n^{\top}\overline{\boldsymbol{M}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}_{n0})\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n0})\xrightarrow{d}\mathcal{N}_{p_n}(0,1).$$

ب.۲.۳ اثبات لم‌ها

اثبات لم ب.۱.۳

فرض کنید $\boldsymbol{Q}=\{q_{j_1,j_2}\}_{1\leqslant j_1,j_2\leqslant m}$ نمایانگر ماتریس $\widehat{\boldsymbol{R}}^{-1}-\overline{\boldsymbol{R}}^{-1}$ باشد. آن‌گاه

$$\begin{aligned}
\boldsymbol{S}_n(\boldsymbol{\theta}_{n0})-\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n0}) &= \sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}q_{j_1,j_2}\boldsymbol{A}_{ij_1}^{1/2}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\boldsymbol{A}_{ij_2}^{-1/2}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\\
&\quad\times\big(y_{ij_1}-\mu_{ij_2}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\big)\boldsymbol{D}_{ij_1}\\
&= \sum_{j_1=1}^{m}\sum_{j_2=1}^{m}q_{j_1,j_2}\Big[\sum_{i=1}^{n}\boldsymbol{A}_{ij_1}^{1/2}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\epsilon_{ij_2}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\boldsymbol{D}_{ij_1}\Big],
\end{aligned}$$

به طوری که $\epsilon_{ij_۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i) = \boldsymbol{A}_{ij_۲}^{-۱/۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i)\big(y_{ij_۱} - \mu_{ij_۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i)\big)$. در نظر داشته باشید که

$$\begin{aligned}\mathrm{E}\Big[\Big\|\sum_{i=۱}^{n} \boldsymbol{A}_{ij_۱}^{۱/۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i)\epsilon_{ij_۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i)\boldsymbol{D}_{ij_۱}\Big\|^{۲}\Big] &= \sum_{i=۱}^{n} \boldsymbol{A}_{ij_۱}^{۱/۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i)\,\mathrm{E}[\epsilon_{ij_۲}^{۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i)]\boldsymbol{D}_{ij_۱}^{\top}\boldsymbol{D}_{ij_۱} \\ &\leqslant \sum_{i=۱}^{n} \boldsymbol{D}_{ij_۱}^{\top}\boldsymbol{D}_{ij_۱} = O(np_n).\end{aligned}$$

بنابراین $\big\|\sum_{i=۱}^{n} \boldsymbol{A}_{ij_۱}^{۱/۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i)\epsilon_{ij_۲}(\boldsymbol{\theta}_{n۰}, \boldsymbol{u}_i)\big\| = O_p(\sqrt{np_n})$. بر اساس شرط (۴.A)، به ازای هر $۱ \leqslant j_۱$ و $j_۲ \leqslant m$، داریم $q_{j_۱,j_۲} = O_p(\sqrt{p_n/n})$، لذا اثبات کامل می‌شود.

اثبات لم ب.۲.۳

اثبات مشابه پن (۲۰۰۲) است.

اثبات لم ب.۳.۳

فرض کنید $\boldsymbol{H}_n(\boldsymbol{\theta}_n)$، $\boldsymbol{E}_n(\boldsymbol{\theta}_n)$ و $\boldsymbol{G}_n(\boldsymbol{\theta}_n)$ به ترتیب مشابه $\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n)$، $\overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n)$ و $\overline{\boldsymbol{G}}_n(\boldsymbol{\theta}_n)$، تعریف شوند با این تفاوت که $\overline{\boldsymbol{R}}$ توسط $\widehat{\boldsymbol{R}}$ جایگزین می‌شود. بر اساس لم ب.۲.۳، کافی است که نتایج زیر را اثبات کنیم.

$$\sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰}\| \leqslant \Delta\sqrt{p_n/n}} \sup_{\|\boldsymbol{b}_n\|=۱} \Big|\boldsymbol{b}_n^{\top}\big[\boldsymbol{H}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n)\big]\boldsymbol{b}_n\Big| = O_p(\sqrt{np_n}), \qquad (۴.ب)$$

$$\sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰}\| \leqslant \Delta\sqrt{p_n/n}} \sup_{\|\boldsymbol{b}_n\|=۱} \Big|\boldsymbol{b}_n^{\top}\big[\boldsymbol{E}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n)\big]\boldsymbol{b}_n\Big| = O_p(\sqrt{np_n}), \qquad (۵.ب)$$

$$\sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰}\| \leqslant \Delta\sqrt{p_n/n}} \sup_{\|\boldsymbol{b}_n\|=۱} \Big|\boldsymbol{b}_n^{\top}\big[\boldsymbol{G}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{G}}_n(\boldsymbol{\theta}_n)\big]\boldsymbol{b}_n\Big| = O_p(\sqrt{np_n}). \qquad (۶.ب)$$

حال داریم

$$\begin{aligned}\Big|\boldsymbol{b}_n^{\top}\big[\boldsymbol{H}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n)\big]\boldsymbol{b}_n\Big| &= \Big|\sum_{i=۱}^{n} \boldsymbol{b}_n^{\top}\boldsymbol{D}_i^{\top}\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)[\widehat{\boldsymbol{R}}^{-۱} - \overline{\boldsymbol{R}}^{-۱}]\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{D}_i\boldsymbol{b}_n\Big| \\ &\leqslant \|\widehat{\boldsymbol{R}}^{-۱} - \overline{\boldsymbol{R}}^{-۱}\|\lambda_{\max}\big(\boldsymbol{A}_i(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big)\lambda_{\max}\Big(\sum_{i=۱}^{n} \boldsymbol{D}_i^{\top}\boldsymbol{D}_i\Big)\|\boldsymbol{b}_n\|^{۲}.\end{aligned}$$

تحت شرط (۴.A)، عبارت (ب.۴) اثبات می‌شود. سپس در نظر داشته باشید که

$$\begin{aligned}\Big|\boldsymbol{b}_n^{\top}\big[\boldsymbol{E}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n)\big]\boldsymbol{b}_n\Big| &= \frac{۱}{۲}\Big|\sum_{i=۱}^{n}\sum_{j=۱}^{m} \big(۱ - ۲\mu_{ij}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big)\epsilon_{ij}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\boldsymbol{b}_n^{\top}\boldsymbol{D}_i^{\top}\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i) \\ &\qquad \times[\widehat{\boldsymbol{R}}^{-۱} - \overline{\boldsymbol{R}}^{-۱}]\boldsymbol{e}_j\boldsymbol{e}_j^{\top}\boldsymbol{D}_i\boldsymbol{b}_n\Big| \\ &\leqslant \sum_{i=۱}^{n}\sum_{j=۱}^{m} \boldsymbol{A}_i^{-۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\big|\boldsymbol{b}_n^{\top}\boldsymbol{D}_i^{\top}\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)[\widehat{\boldsymbol{R}}^{-۱} - \overline{\boldsymbol{R}}^{-۱}]\boldsymbol{e}_j\big|.\big|\boldsymbol{e}_j^{\top}\boldsymbol{D}_i\boldsymbol{b}_n\big| \\ &\leqslant \sum_{i=۱}^{n}\sum_{j=۱}^{m} \boldsymbol{A}_i^{-۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\|\widehat{\boldsymbol{R}}^{-۱} - \overline{\boldsymbol{R}}^{-۱}\|.\|\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n, \boldsymbol{u}_i)\|.\|\boldsymbol{D}_i\boldsymbol{b}_n\|^{۲}.\end{aligned}$$

تحت برقراری شرط(۴.A)، داریم

$$\begin{aligned}\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}} & \sup_{\|\boldsymbol{b}_n\|=۱}\Big|\boldsymbol{b}_n^\top\big[\boldsymbol{E}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n)\big]\boldsymbol{b}_n\Big| \\ &\leqslant C\|\widehat{\boldsymbol{R}}^{-۱}-\overline{\boldsymbol{R}}^{-۱}\|.\sum_{i=۱}^{n}\sum_{j=۱}^{m}\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}}A_{ij}^{-۱/۲}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\sup_{\|\boldsymbol{b}_n\|=۱}\|\boldsymbol{D}_i\boldsymbol{b}_n\|^۲ \\ &= O_p(\sqrt{p_n/n})O(n)=O_p(\sqrt{np_n}).\end{aligned}$$

بنابراین عبارت (ب.۵) نیز اثبات می‌شود. عبارت (ب.۶) مشابه دو عبارت قبل برقرار است.

اثبات لم ب.۴.۳

بر اساس لم ب.۲.۳ تنها کافی است نشان دهیم که

$$\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}}\sup_{\|\boldsymbol{b}_n\|=۱}\Big|\boldsymbol{b}_n^\top\overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n)\boldsymbol{b}_n\Big|=O_p(\sqrt{np_n}), \qquad (ب.۷)$$

و

$$\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}}\sup_{\|\boldsymbol{b}_n\|=۱}\Big|\boldsymbol{b}_n^\top\overline{\boldsymbol{G}}_n(\boldsymbol{\theta}_n)\boldsymbol{b}_n\Big|=O_p(\sqrt{np_n}). \qquad (ب.۸)$$

در نظر داشته باشید که می‌توان $\overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n)$ را به صورت

$$\begin{aligned}\overline{\boldsymbol{E}}_n(\boldsymbol{\theta}_n) &= \frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{m}\big(۱-۲\mu_{ij}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\big)\epsilon_{ij}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\boldsymbol{D}_i^\top\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{e}_j\boldsymbol{e}_j^\top\boldsymbol{D}_i \\ &+\frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{m}\big(۱-۲\mu_{ij}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\big)\epsilon_{ij}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\boldsymbol{D}_i^\top\Big[\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n,\boldsymbol{u}_i) \\ &-\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\Big]\overline{\boldsymbol{R}}^{-۱}\boldsymbol{e}_j\boldsymbol{e}_j^\top\boldsymbol{D}_i \\ &+\frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{m}\Big[\big(۱-۲\mu_{ij}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big)\boldsymbol{A}_{ij}^{-۱/۲}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)-\big(۱-۲\mu_{ij}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\big) \\ &\times\boldsymbol{A}_{ij}^{-۱/۲}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\Big]\big(y_{ij}-\mu_{ij}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\big)\boldsymbol{D}_i^\top\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{e}_j\boldsymbol{e}_j^\top\boldsymbol{D}_i \\ &+\frac{۱}{۲}\sum_{i=۱}^{n}\sum_{j=۱}^{m}\big(۱-۲\mu_{ij}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big)\boldsymbol{A}_{ij}^{-۱/۲}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big(\mu_{ij}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)-\mu_{ij}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big) \\ &\times\boldsymbol{D}_i^\top\boldsymbol{A}_i^{۱/۲}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-۱}\boldsymbol{e}_j\boldsymbol{e}_j^\top\boldsymbol{D}_i \\ &= \overline{\boldsymbol{E}}_{۱n}(\boldsymbol{\theta}_{n\circ})+\sum_{k=۲}^{۴}\overline{\boldsymbol{E}}_{kn}(\boldsymbol{\theta}_n)\end{aligned}$$

تجزیه کرد. بنابراین برای عبارت (ب.۷) کافی است نشان دهیم که

$$\sup_{\|\boldsymbol{b}_n\|=۱}\Big|\boldsymbol{b}_n^\top\overline{\boldsymbol{E}}_{۱n}(\boldsymbol{\theta}_{n\circ})\boldsymbol{b}_n\Big|=O_p(\sqrt{n}p_n)$$

و

$$\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}}\ \sup_{\|\boldsymbol{b}_n\|=1}\left|\boldsymbol{b}_n^\top\overline{\boldsymbol{E}}_{kn}(\boldsymbol{\theta}_n)\boldsymbol{b}_n\right|=O_p(\sqrt{n}p_n).$$

ابتدا نشان می‌دهیم که $\sup_{\|\boldsymbol{b}_n\|=1}\left|\boldsymbol{b}_n^\top\overline{\boldsymbol{E}}_{1n}(\boldsymbol{\theta}_{n\circ})\boldsymbol{b}_n\right|=O_p(\sqrt{n}p_n)$ برقرار است. اگر نشان دهیم که $\|\overline{\boldsymbol{E}}_{1n}(\boldsymbol{\theta}_{n\circ})\|=O_p(\sqrt{n}p_n)$، به طوری که $\|\overline{\boldsymbol{E}}_{1n}(\boldsymbol{\theta}_{n\circ})\|=\sqrt{\operatorname{tr}\left(\overline{\boldsymbol{E}}_{1n}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{E}}_{1n}^\top(\boldsymbol{\theta}_{n\circ})\right)}$، آنگاه نتیجه حاصل می‌شود. حال داریم

$$\begin{aligned}
&\mathrm{E}[\|\overline{\boldsymbol{E}}_{1n}(\boldsymbol{\theta}_{n\circ})\|^2]\\
=\ &\frac{1}{4}\sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\big(1-2\mu_{ij_1}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\big)\big(1-2\mu_{ij_2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\big)\\
&\quad\times\mathrm{E}[\epsilon_{ij_1}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\epsilon_{ij_2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)]\\
&\quad\quad\times\operatorname{tr}\left[\boldsymbol{D}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{e}_{j_1}\boldsymbol{e}_{j_1}^\top\boldsymbol{D}_i\boldsymbol{D}_i^\top\boldsymbol{e}_{j_2}\boldsymbol{e}_{j_2}^\top\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\boldsymbol{D}_i\right]\\
\leqslant\ &C\sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\left|\boldsymbol{e}_{j_1}^\top\boldsymbol{D}_i\boldsymbol{D}_i^\top\boldsymbol{e}_{j_2}\boldsymbol{e}_{j_2}^\top\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\boldsymbol{D}_i\boldsymbol{D}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{e}_{j_1}\right|\\
\leqslant\ &C\sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\|\boldsymbol{e}_{j_1}^\top\boldsymbol{D}_i\|.\|\boldsymbol{D}_i^\top\boldsymbol{e}_{j_2}\|.\|\boldsymbol{e}_{j_2}^\top\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\boldsymbol{D}_i\|\\
&\quad\times\|\boldsymbol{D}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{e}_{j_1}\|.
\end{aligned}$$

در نظر داشته باشید که $\|\boldsymbol{e}_{j_1}^\top\boldsymbol{D}_i\|=\|\boldsymbol{D}_{ij_1}\|$، $\|\boldsymbol{D}_i^\top\boldsymbol{e}_{j_2}\|=\|\boldsymbol{D}_{ij_2}\|$، $\|\boldsymbol{e}_{j_2}^\top\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\boldsymbol{D}_i\|\leqslant C\big(\operatorname{tr}(\boldsymbol{D}_i\boldsymbol{D}_i^\top)\big)^{1/2}$ و $\|\boldsymbol{D}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{e}_{j_1}\|\leqslant C\big(\operatorname{tr}(\boldsymbol{D}_i\boldsymbol{D}_i^\top)\big)^{1/2}$. بنابراین براساس شرایط (۲.A) و (۳.A)، داریم

$$\begin{aligned}
\mathrm{E}[\|\overline{\boldsymbol{E}}_{1n}(\boldsymbol{\theta}_{n\circ})\|^2]\ &\leqslant\ C\sum_{i=1}^{n}\sum_{j_1=1}^{m}\sum_{j_2=1}^{m}\|\boldsymbol{D}_{ij_1}\|.\|\boldsymbol{D}_{ij_2}\|\operatorname{tr}(\boldsymbol{D}_i\boldsymbol{D}_i^\top)\\
&\leqslant\ C\max_{i,j}\|\boldsymbol{D}_{ij}\|^2\operatorname{tr}\Big(\sum_{i=1}^{n}\boldsymbol{D}_i\boldsymbol{D}_i^\top\Big)=O(np_n^2),
\end{aligned}$$

که $\sup_{\|\boldsymbol{b}_n\|=1}\left|\boldsymbol{b}_n^\top\overline{\boldsymbol{E}}_{1n}(\boldsymbol{\theta}_{n\circ})\boldsymbol{b}_n\right|=O_p(\sqrt{n}p_n)$. سپس داریم

$$\begin{aligned}
\left|\boldsymbol{b}_n^\top\overline{\boldsymbol{E}}_{2n}(\boldsymbol{\theta}_{n\circ})\boldsymbol{b}_n\right|\ &=\ \Big|\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\big(1-2\mu_{ij}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\big)\epsilon_{ij}^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\\
&\quad-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)]\overline{\boldsymbol{R}}^{-1}\boldsymbol{e}_j\boldsymbol{e}_j^\top\boldsymbol{D}_i\boldsymbol{b}_n\Big|\\
&\leqslant\ C\sum_{i=1}^{n}\sum_{j=1}^{m}\left|\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)]\overline{\boldsymbol{R}}^{-1}\boldsymbol{e}_j\right|.\left|\boldsymbol{e}_j^\top\boldsymbol{D}_i\boldsymbol{b}_n\right|\\
&\leqslant\ C\sum_{i=1}^{n}\sum_{j=1}^{m}\|\boldsymbol{D}_i\boldsymbol{b}_n\|^2\lambda_{\max}(\overline{\boldsymbol{R}}^{-1})\max_j|\boldsymbol{A}_{ij}^{1/2}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)-\boldsymbol{A}_{ij}^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)|.
\end{aligned}$$

بین $\boldsymbol{\theta}_n$ و $\boldsymbol{\theta}_{n\circ}$، $\boldsymbol{\theta}_n^*$هایی را می‌توان یافت که

$$\begin{aligned}
\boldsymbol{A}_{ij}^{1/2}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)-\boldsymbol{A}_{ij}^{1/2}(\boldsymbol{\theta}_{n\circ},\boldsymbol{u}_i)\ &=\ \frac{1}{2}\boldsymbol{A}_{ij}^{1/2}(\boldsymbol{\theta}_n^*,\boldsymbol{u}_i)\big(1-2\mu_{ij}(\boldsymbol{\theta}_n^*,\boldsymbol{u}_i)\big)\boldsymbol{D}_{ij}^\top(\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ})\\
&\leqslant\ C\|\boldsymbol{D}_{ij}\|.\|(\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ})\|.
\end{aligned}$$

بنابراین

$$\begin{aligned}
&\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}} \quad \sup_{\|\boldsymbol{b}_n\|=1}\Big|\boldsymbol{b}_n^\top\overline{\boldsymbol{E}}_{2n}(\boldsymbol{\theta}_n)\boldsymbol{b}_n\Big| \\
&\qquad\leqslant\ C\max_{ij}\|\boldsymbol{D}_{ij}\|\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}}\|(\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ})\|.\lambda_{\max}\Big(\sum_{i=1}^n\boldsymbol{D}_i\boldsymbol{D}_i^\top\Big) \\
&\qquad=\ O(\sqrt{p_n})O(\sqrt{p_n/n})O(n)=O(\sqrt{n}p_n).
\end{aligned}$$

به طور مشابه می‌توان نشان داد که $\sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}}\ \sup_{\|\boldsymbol{b}_n\|=1}\Big|\boldsymbol{b}_n^\top\overline{\boldsymbol{E}}_{kn}(\boldsymbol{\theta}_n)\boldsymbol{b}_n\Big|=O_p(\sqrt{n}p_n),\ k=$ ۳,۴. لذا عبارت (ب۷.) اثبات می‌شود. برقراری عبارت (ب۸.) از طریق تکنیک‌های مشابه نتیجه می‌شود.

اثبات لم ب.۵.۳

$$\begin{aligned}
&\boldsymbol{b}_n^\top\big[\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n\circ})\big]\boldsymbol{b}_n \\
=\ &\Big|\sum_{i=1}^n\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top\big[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\big]\boldsymbol{D}_i\boldsymbol{b}_n\Big| \\
\leqslant\ &\Big|\sum_{i=1}^n\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top\big[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\big]\boldsymbol{D}_i\boldsymbol{b}_n\Big| \\
&+\Big|\sum_{i=1}^n\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top\big[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\big]\boldsymbol{D}_i\boldsymbol{b}_n\Big| \\
=\ &I_{n1}+I_{n2}.
\end{aligned}$$

بر اساس نامساوی کوشی-شوارتز می توان

$$\begin{aligned}
I_{n1}\ &\leqslant\ \sum_{i=1}^n\Big|\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)\overline{\boldsymbol{R}}^{-1}\big[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\big]\boldsymbol{D}_i\boldsymbol{b}_n\Big| \\
&\leqslant\ \sum_{i=1}^n\|\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)\overline{\boldsymbol{R}}^{-1/2}\|.\|\overline{\boldsymbol{R}}^{-1/2}\big[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\big]\boldsymbol{D}_i\boldsymbol{b}_n\|,
\end{aligned}$$

را نتیجه گرفت. سپس از عبارت‌های

$$\begin{aligned}
\|\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1/2}\|\ &=\ \Big[\boldsymbol{b}_n^\top\boldsymbol{D}_i^\top\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\boldsymbol{D}_i\boldsymbol{b}_n\Big] \\
&\leqslant\ \lambda_{\max}^{1/2}\big(\overline{\boldsymbol{R}}^{-1}\big)\lambda_{\max}^{1/2}\big(\boldsymbol{A}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big)\|\boldsymbol{D}_i\boldsymbol{b}_n\|,
\end{aligned}$$

و

$$\begin{aligned}
&\|\overline{\boldsymbol{R}}^{-1/2}\big[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\big]\boldsymbol{D}_i\boldsymbol{b}_n\| \\
\leqslant\ &\lambda_{\max}^{1/2}\big(\overline{\boldsymbol{R}}^{-1}\big)\lambda_{\max}^{1/2}\big(\big[\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_n)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ})\big]^2\big)\|\boldsymbol{D}_i\boldsymbol{b}_n\|,
\end{aligned}$$

می‌توان نشان داد که

$$\begin{aligned} I_{n1} &\leqslant C \max_{i,j} \left| \boldsymbol{A}_{ij}^{1/2}(\boldsymbol{\theta}_n) - \boldsymbol{A}_{ij}^{1/2}(\boldsymbol{\theta}_{n\circ}) \right| \sum_{i=1}^{n} \|\boldsymbol{D}_i \boldsymbol{b}_n\|^2 \\ &\leqslant C \max_{i,j} \|\boldsymbol{D}_{ij}\| . \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n\circ}\| . \lambda_{\max}\Big(\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{D}_i \Big) \|\boldsymbol{b}_n\|^2 . \end{aligned}$$

بنابراین خواهیم داشت

$$\sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n\circ}\| \leqslant \Delta\sqrt{p_n/n}} \quad \sup_{\|\boldsymbol{b}_n\|=1} I_{n1} = O_p(\sqrt{n} p_n).$$

به طور مشابه می‌توان نشان داد که $\sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n\circ}\| \leqslant \Delta\sqrt{p_n/n}} \sup_{\|\boldsymbol{b}_n\|=1} I_{n2} = O_p(\sqrt{n} p_n)$ و لم اثبات می‌شود.

اثبات لم ب.۶.۳

عبارت مورد نظر که باید توزیع مجانبی آن را یافت به صورت $\boldsymbol{\alpha}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ}) \overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) = \sum_{i=1}^{n} Z_{ni}$ قابل بیان است، به طوری که $Z_{ni} = \boldsymbol{\alpha}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ}) \boldsymbol{D}_i^\top \boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ}, \boldsymbol{u}_i) \overline{\boldsymbol{R}}^{-1} \boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n\circ}, \boldsymbol{u}_i)$. از آنجایی که $\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ}) = \operatorname{Cov}\big(\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ})\big)$، داریم $\operatorname{Var}\big(\boldsymbol{\alpha}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ}) \overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ})\big) = 1$. برای نشان دادن توزیع نرمال مجانبی کافی است که شرط لیندبرگ را بررسی کنیم، به این صورت که به ازای هر $\epsilon > 0$ داشته باشیم $\sum_{i=1}^{n} \mathrm{E}[Z_{ni}^2 I(|Z_{ni}| > \epsilon)] \to 0$. بر اساس نامساوی کوشی-شوارتز داریم

$$\begin{aligned} Z_{ni}^2 &\leqslant \|\boldsymbol{\alpha}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ}) \boldsymbol{D}_i^\top \boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_{n\circ}, \boldsymbol{u}_i) \overline{\boldsymbol{R}}^{-1}\|^2 . \|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n\circ}, \boldsymbol{u}_i)\|^2 \\ &\leqslant \lambda_{\max}\big(\overline{\boldsymbol{R}}^{-2}\big) \lambda_{\max}\big(\boldsymbol{A}_i(\boldsymbol{\theta}_{n\circ}, \boldsymbol{u}_i)\big) \big(\boldsymbol{\alpha}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ}) \boldsymbol{D}_i^\top \boldsymbol{D}_i \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ}) \boldsymbol{\alpha}_n\big) \|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n\circ}, \boldsymbol{u}_i)\|^2 \\ &\leqslant C \gamma_{ni} \|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n\circ}, \boldsymbol{u}_i)\|^2, \end{aligned}$$

به طوری که $\gamma_{ni} = \boldsymbol{\alpha}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ}) \boldsymbol{D}_i^\top \boldsymbol{D}_i \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ}) \boldsymbol{\alpha}_n$. سپس نشان خواهیم داد هنگامی که $n \to \infty$ آنگاه $\max_{1 \leqslant i \leqslant n} \gamma_{ni} \to 0$. در نظر داشته باشید که $\gamma_{ni} \leqslant \lambda_{\max}\big(\boldsymbol{D}_i^\top \boldsymbol{D}_i\big) \lambda_{\min}^{-1}\big(\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ})\big)$. از آنجایی که $\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ})$ متقارن است، برای ارزیابی $\lambda_{\min}\big(\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ})\big)$، به ازای هر $\boldsymbol{b}_n \in \mathcal{R}^{p_n}$، داریم

$$\begin{aligned} \boldsymbol{b}_n^\top \overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ}) \boldsymbol{b}_n &\geqslant \lambda_{\min}\big(\boldsymbol{R}_\circ\big) \lambda_{\min}\big(\overline{\boldsymbol{R}}^{-2}\big) \sum_{i=1}^{n} \lambda_{\min}\big(\boldsymbol{A}_i(\boldsymbol{\theta}_{n\circ}, \boldsymbol{u}_i)\big) \boldsymbol{b}_n^\top \boldsymbol{D}_i^\top \boldsymbol{D}_i \boldsymbol{b}_n \\ &\geqslant C \boldsymbol{b}_n^\top \Big(\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{D}_i \Big) \boldsymbol{b}_n \geqslant C \lambda_{\min}\Big(\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{D}_i \Big) \|\boldsymbol{b}_n\|^2 . \end{aligned}$$

بنابراین نامساوی $\inf_{\|\boldsymbol{b}_n\|=1} |\boldsymbol{b}_n^\top \overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ}) \boldsymbol{b}_n| \geqslant C \lambda_{\min}\Big(\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{D}_i \Big)$ را نتیجه می‌گیریم و این نشان می‌دهد که $\lambda_{\min}\big(\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ}) \geqslant C \lambda_{\min}\Big(\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{D}_i \Big)$. حال می‌توان نامساوی

$$\gamma_{ni} \leqslant \frac{\lambda_{\max}\Big(\boldsymbol{D}_i^\top \boldsymbol{D}_i\Big)}{C \lambda_{\min}\Big(\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{D}_i \Big)} \leqslant \frac{\operatorname{tr}\Big(\boldsymbol{D}_i^\top \boldsymbol{D}_i\Big)}{C \lambda_{\min}\Big(\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{D}_i \Big)} = \frac{\sum_{j=1}^{m} \boldsymbol{D}_{ij}^\top \boldsymbol{D}_{ij}}{C \lambda_{\min}\Big(\sum_{i=1}^{n} \boldsymbol{D}_i^\top \boldsymbol{D}_i \Big)},$$

و به دنبال آن $\max_{۱\leqslant i\leqslant n}\gamma_{ni}\leqslant O(n^{-۱}p_n)=o(۱)$ را نتیجه گرفت. بنابراین داریم

$$\sum_{i=۱}^{n}\mathrm{E}[Z_{ni}^{۲}I(|Z_{ni}|>\epsilon)]\leqslant\sum_{i=۱}^{n}C\gamma_{ni}\,\mathrm{E}\left[\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^{۲}\mathrm{I}\Big\{\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^{۲}>\frac{\epsilon^{۲}}{C\gamma_{ni}}\Big\}\right].$$

بر اساس شرط (۳.A)، عبارت $\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^{۲}$ به طور یکنواخت کراندار است، لذا به ازای هر $\epsilon>۰$ و $\delta>۰$، یک عدد ثابت و مثبت مانند C وجود دارد که $\mathrm{I}\Big\{\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^{۲}>\frac{\epsilon^{۲}}{C\gamma_{ni}}\Big\}=۰$. این تضمین می‌کند که

$$\sum_{i=۱}^{n}C\gamma_{ni}\,\mathrm{E}\left[\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^{۲}\mathrm{I}\Big\{\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^{۲}>\frac{\epsilon^{۲}}{C\gamma_{ni}}\Big\}\right]\to۰.$$

لذا شرط لیندبرگ برقرار است.

ب.۳.۳ اثبات قضیه ۱.۱.۳

قسمت (i): بر اساس تعریف $\widehat{\boldsymbol{\theta}}_n$ بدیهی است که به ازای $k=۱,\ldots,s_n$، $S_{nk}(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{u}_i)=۰$. کافی است ثابت کنیم که $\mathbb{P}(|\widehat{\theta}_{nk}|\geqslant a\lambda_n,\ k=۱,\ldots,s_n)\to۱$، یعنی با احتمال نزدیک به ۱، تابع تاوان برابر ۰ است. می‌دانیم که

$$\min_{۱\leqslant k\leqslant s_n}|\widehat{\theta}_{nk}|\geqslant\min_{۱\leqslant k\leqslant s_n}|\widehat{\theta}_{n۰k}|-\max_{۱\leqslant k\leqslant s_n}|\widehat{\theta}_{n۰k}-\widehat{\theta}_{nk}|\geqslant\min_{۱\leqslant k\leqslant s_n}|\widehat{\theta}_{n۰k}|-\|\boldsymbol{\theta}_{n۱۰}-\widehat{\boldsymbol{\theta}}_{n۱۰}\|.$$

همچنین فرض کنید

$$\|\boldsymbol{\theta}_{n۱۰}-\widehat{\boldsymbol{\theta}}_{n۱۰}\|=\sqrt{s_n/n}. \qquad (ب.۹)$$

از آنجایی که $\min_{۱\leqslant k\leqslant s_n}|\widehat{\theta}_{n۰k}|/\lambda\to\infty$ و $\|\boldsymbol{\theta}_{n۱۰}-\widehat{\boldsymbol{\theta}}_{n۱۰}\|=o(\lambda_n)$ می‌توان نتیجه گرفت که

$$\mathbb{P}(\min_{۱\leqslant k\leqslant s_n}|\widehat{\theta}_{n۰k}|-\|\boldsymbol{\theta}_{n۱۰}-\widehat{\boldsymbol{\theta}}_{n۱۰}\|\geqslant a\lambda_n)=\mathbb{P}(\|\boldsymbol{\theta}_{n۱۰}-\widehat{\boldsymbol{\theta}}_{n۱۰}\|\leqslant\min_{۱\leqslant k\leqslant s_n}|\widehat{\theta}_{n۰k}|-a\lambda_n)\to۱.$$

بنابراین هنگامی که $n\to\infty$، داریم $\mathbb{P}(\min_{۱\leqslant k\leqslant s_n}|\widehat{\theta}_{nk}|\geqslant a\lambda_n)\to۱$.

قسمت (ii): به ازای هر $k=s_n+۱,\ldots,p_n$، داریم $\widehat{\theta}_{nk}=۰$، لذا $q_{\lambda_n}(|\theta_{nk}|)\,\mathrm{sgn}(\theta_{nk})=۰$. کافی است نشان دهیم که

$$\mathbb{P}\left(\max_{s_n+۱\leqslant k\leqslant p_n}|S_{nk}(\widehat{\boldsymbol{\theta}})|\leqslant\frac{\lambda_n}{\log n}\right)\to۱. \qquad (ب.۱۰)$$

به عبارت دیگر کافی است نشان دهیم که

$$\mathbb{P}\left(\max_{s_n+۱\leqslant k\leqslant p_n}|S_{nk}(\widehat{\boldsymbol{\theta}}_n)-\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_n)|>\frac{\lambda_n}{۲\log n}\right)\to۰, \qquad (ب.۱۱)$$

و

$$\mathbb{P}\left(\max_{s_n+۱\leqslant k\leqslant p_n}|\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_n)|>\frac{\lambda_n}{۲\log n}\right)\to۰. \qquad (ب.۱۲)$$

عبارت سمت چپ در (ب.۱۱)، توسط عبارت

$$\begin{aligned}
&\mathbb{P}\ \Big(\max_{s_n+1\leqslant k\leqslant p_n} n^{-1}\,\mathrm{E}_{u|y}\Big[\sum_{i=1}^{n}\Big|\boldsymbol{e}_k\boldsymbol{D}_i^{\top}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)[\widehat{\boldsymbol{R}}^{-1}-\overline{\boldsymbol{R}}^{-1}]\boldsymbol{A}_i^{-\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\\
&\qquad\times\big(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big)\Big|\Big]>\frac{\lambda_n}{2\log n}\Big)\\
\leqslant\ &\mathbb{P}\ \Big(\max_{s_n+1\leqslant k\leqslant p_n} n^{-1}\,\mathrm{E}_{u|y}\Big[\sum_{i=1}^{n}\|\boldsymbol{e}_k\boldsymbol{D}_i^{\top}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\|.\|\widehat{\boldsymbol{R}}^{-1}-\overline{\boldsymbol{R}}^{-1}\|.\|\boldsymbol{A}_i^{-\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\\
&\qquad\times\big(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big)\|\Big]>\frac{\lambda_n}{2\log n}\Big)\\
\leqslant\ &\mathbb{P}\ \Big(\|\widehat{\boldsymbol{R}}^{-1}-\overline{\boldsymbol{R}}^{-1}\|n^{-1}E_{u|y}\Big[\sum_{i=1}^{n}(\max_{s_n+1\leqslant k\leqslant p_n}\|\boldsymbol{e}_k\boldsymbol{D}_i^{\top}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\|)\|\boldsymbol{A}_i^{-\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\\
&\qquad\times\big(\boldsymbol{y}_i-\boldsymbol{\mu}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\big)\|\Big]>\frac{\lambda_n}{2\log n}\Big)\\
\leqslant\ &\mathbb{P}\ \Big(n^{-1}E_{u|y}\Big[\sum_{i=1}^{n}\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\|\Big]>\frac{\lambda_n\sqrt{n}}{2\sqrt{s_n}\log n}\Big)\\
\leqslant\ &C\ \ \frac{n^{-1}\,\mathrm{E}_{u|y}\Big[\sum_{i=1}^{n}\mathrm{E}(\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\|)\Big]\sqrt{s_n}\log n}{\lambda_n\sqrt{n}}=O(\frac{\sqrt{s_n}\log n}{\lambda_n\sqrt{n}})=o(1),
\end{aligned}$$

از بالاکراندار است، به طوری‌که نامساوی سوم از شرط‌های (۲.A) و (۴.A) نتیجه می‌شود. همچنین شرط (۴.A) بیان می‌کند که $\sqrt{s_n}\log n/\lambda_n\sqrt{n}\to 0$. برای اثبات عبارت (ب.۱۲)، بسط تیلور عبارت $\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{u}_i)$ را به صورت

$$\begin{aligned}
\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{u}_i)=\overline{S}_{nk}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\ &+\ \frac{\partial\overline{S}_{nk}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)}{\partial\boldsymbol{\theta}_n^{\top}}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{n0})\\
&+\ (\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{n0})^{\top}\frac{\partial^2\overline{S}_{nk}(\boldsymbol{\theta}_n^*,\boldsymbol{u}_i)}{\partial\boldsymbol{\theta}_n\partial\boldsymbol{\theta}_n^{\top}}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{n0}),
\end{aligned}\qquad(\text{ب.۱۳})$$

در نظر می‌گیریم، که $\boldsymbol{\theta}_n^*$ بین $\boldsymbol{\theta}_{n0}$ و $\widehat{\boldsymbol{\theta}}_n$ است. فرض کنید $\boldsymbol{D}_k(\boldsymbol{\theta}_n,\boldsymbol{u}_i)=\frac{\partial\overline{S}_{nk}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)}{\partial\boldsymbol{\theta}_n^{\top}}$ و $\boldsymbol{\nabla}_k(\boldsymbol{\theta}_n,\boldsymbol{u}_i)=\frac{\partial^2\overline{S}_{nk}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)}{\partial\boldsymbol{\theta}_n\partial\boldsymbol{\theta}_n^{\top}}$. $\boldsymbol{D}_{k1}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)$ زیربرداری است که شامل s_n عضو اول بردار $\boldsymbol{D}_k(\boldsymbol{\theta}_n,\boldsymbol{u}_i)$ است. همچنین $\boldsymbol{\nabla}_{k1}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)$ یک ماتریس $s_n\times s_n$ بعدی است که همه اعضای s_n سطر و ستون اول ماتریس $\boldsymbol{\nabla}_k(\boldsymbol{\theta}_n,\boldsymbol{u}_i)$ را شامل می‌شود. از آنجایی که $\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{n0}=((\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n01})^{\top},\boldsymbol{0}^{\top})^{\top}$، عبارت (ب.۱۳) را می‌توان به صورت

$$\begin{aligned}
\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{u}_i)=\overline{S}_{nk}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\ &+\ \boldsymbol{D}_{k1}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n10})\\
&+\ (\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n10})^{\top}\boldsymbol{\nabla}_{k1}(\boldsymbol{\theta}_n^*,\boldsymbol{u}_i)(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n10}),
\end{aligned}$$

بیان نمود. حال داریم

$$\begin{aligned}
&\mathbb{P}\ \big(\max_{s_n+1\leqslant k\leqslant p_n}|\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_n)|>\frac{\lambda_n}{2\log n}\big)\\
\leqslant\ &\mathbb{P}\ \big(\max_{s_n+1\leqslant k\leqslant p_n}|\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_n)|>\frac{\lambda_n}{6\log n}\big)\\
&\qquad+\mathbb{P}\big(\max_{s_n+1\leqslant k\leqslant p_n}|\boldsymbol{D}_{k1}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n10})|>\frac{\lambda_n}{6\log n}\big)\\
&\qquad+\mathbb{P}\big(\max_{s_n+1\leqslant k\leqslant p_n}|(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n10})^\top\boldsymbol{\nabla}_{k1}(\boldsymbol{\theta}_n^*,\boldsymbol{u}_i)(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n10})|>\frac{\lambda_n}{6\log n}\big)\\
=\ &I_{n1}+I_{n2}+I_{n3}.
\end{aligned}$$

اگر نشان دهیم $I_{ni}=o(1)$ که $i=1,2,3$، آن‌گاه رابطه (ب.۱۲) برقرار خواهد بود. ابتدا نشان می‌دهیم که $I_{n1}=o(1)$. بدیهی است که $I_{n1}\leqslant\sum_{k=s_n+1}^{p_n}\mathbb{P}\Big(|\overline{S}_{nk}(\widehat{\boldsymbol{\theta}})|>\frac{\lambda_n}{6\log n}\Big)$. حال می‌توان نوشت $\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_{n0},\boldsymbol{u}_i)=n^{-1}\sum_{i=1}^n Z_i$، به طوری که $Z_i=\boldsymbol{e}_k\boldsymbol{D}_i^\top\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)$ متغیرهای تصادفی مستقل با میانگین ۰ هستند. به ازای هر $l\geqslant 2$ و ثابت‌های $M>0$ و $\delta>0$، داریم

$$\begin{aligned}
\mathrm{E}\,|Z_i|^l\ &\leqslant\ \mathrm{E}\Big[\|\boldsymbol{e}_k\boldsymbol{D}_i^\top\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\overline{\boldsymbol{R}}_i^{-1}\|^l\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\|^l\Big]\\
&\leqslant\ C^lE\Big[\|\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\|^l\Big]=C^l\,\mathrm{E}\Big[\big(\sum_{j=1}^m\epsilon_{ij}^2(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\big)^{l/2}\Big]\\
&\leqslant\ C^lm^{l/2-1}\sum_{j=1}^m\mathrm{E}\,|\epsilon_{ij}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)|^l\leqslant C^lm^{l/2-1}\sum_{j=1}^m l!M_2^{-l}\,\mathrm{E}\,\big(\exp(M_2|\epsilon_{ij}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i))\big)\\
&\leqslant\ C^lm^{l/2-1}ml!M_3\leqslant l!M^{l-2}\delta/2.
\end{aligned}$$

در عبارت فوق نامساوی دوم از شرایط (A.۲) و (A.۳) نتیجه می‌شود. نامساوی سوم را می‌توان از نتیجه نامساوی جنسن که بیان می‌کند برای $m\geqslant 1$ و $p\geqslant 1$ داریم $|\sum_{j=1}^m a_i|^p\leqslant m^{p-1}\sum_{j=1}^m|a_i|^p$، نتیجه گرفت. نامساوی چهارم از بسط تیلور و نامساوی آخر از شرط (A.۵) نتیجه می‌شوند. بنابراین Z_iها در شرط نامساوی برنشتاین صدق می‌کنند و داریم

$$\begin{aligned}
\mathbb{P}\big(|\overline{S}_{nk}(\widehat{\boldsymbol{\theta}}_{n0})|>\frac{\lambda_n}{6\log n}\big)\ &\leqslant\ 2\exp\Big[-\frac{1}{2}\frac{n^2\lambda_n^2/\big(36(\log n)\big)^2}{n\delta+M^*n\lambda_n/(6\log n)}\Big]\\
&\leqslant\ 2\exp\Big[-C\frac{n\lambda_n^2}{(\log n)^2}\Big].
\end{aligned}$$

بر اساس شرط (A.۷) $\log p_n=o\big(n\lambda_n^2/(\log n)^2\big)$ و $n\lambda_n^2/(\log n)^2\to\infty$ برقرار هستند، بنابراین داریم

$$I_{n1}\leqslant 2\exp\Big[\log p_n-C\frac{n\lambda_n^2}{(\log n)^2}\Big]=o(1),$$

و $I_{n1}=o(1)$ اثبات می‌شود.

سپس نشان خواهیم داد که $I_{n۲}=o(۱)$. فرض کنید $\boldsymbol{H}_{nk۱}=(H_{nk۱},\ldots,H_{nks_n})^\top$ زیربرداری باشد که s_n عضو اولیه بردار $\boldsymbol{H}_{nk}$ را شامل می‌شود. $\boldsymbol{E}_{nk۱}$ و $\boldsymbol{G}_{nk۱}$ به طور مشابه تعریف می‌شوند. حال داریم

$$\begin{aligned}
I_{n۲} &= \mathbb{P}\big(\max_{s_n+۱\leqslant k\leqslant p_n}|\boldsymbol{D}_{k۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)(\widehat{\boldsymbol{\theta}}_{n۱}-\boldsymbol{\theta}_{n۱۰})|>\frac{\lambda_n}{۶\log n}\big)\\
&= \mathbb{P}\big(\max_{s_n+۱\leqslant k\leqslant p_n}|\boldsymbol{D}_{k۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)(\widehat{\boldsymbol{\theta}}_{n۱}-\boldsymbol{\theta}_{n۱۰})|>\frac{\lambda_n}{۶\log n},\ \|\widehat{\boldsymbol{\theta}}_{n۱}-\boldsymbol{\theta}_{n۱۰}\|\leqslant\sqrt{s_n/n}\log n\big)\\
&\quad +\mathbb{P}\big(\|\widehat{\boldsymbol{\theta}}_{n۱}-\boldsymbol{\theta}_{n۱۰}\|>\sqrt{s_n/n}\log n\big)\\
&\leqslant \mathbb{P}\big(\max_{s_n+۱\leqslant k\leqslant p_n}\|\boldsymbol{D}_{k۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|>\frac{\lambda_n\sqrt{n}}{۶\sqrt{s_n}(\log n)^۲}\big)+o(۱)\\
&\leqslant \mathbb{P}\big(\max_{s_n+۱\leqslant k\leqslant p_n}\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|>\frac{\lambda_n\sqrt{n}}{۶\sqrt{s_n}(\log n)^۲}\big)\\
&\quad +\mathbb{P}\big(\max_{s_n+۱\leqslant k\leqslant p_n}\|\boldsymbol{E}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|>\frac{\lambda_n\sqrt{n}}{۶\sqrt{s_n}(\log n)^۲}\big)\\
&\quad +\mathbb{P}\big(\max_{s_n+۱\leqslant k\leqslant p_n}\|\boldsymbol{G}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|>\frac{\lambda_n\sqrt{n}}{۶\sqrt{s_n}(\log n)^۲}\big)+o(۱)\\
&= I_{n۲۱}+I_{n۲۲}+I_{n۲۳}+o(۱).
\end{aligned}$$

برای ارزیابی $I_{n۲۱}$، تحت شرایط (۵.A)، (۴.A) و (۶.A) و اینکه $|H_{nkk'}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)|$ به طور یکنواخت کراندار است، می‌توان نشان داد که

$$\begin{aligned}
I_{n۲۱} &\leqslant \mathbb{P}\big(\max_{s_n+۱\leqslant k\leqslant p_n}\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲>C\frac{n\lambda_n^۲}{s_n(\log n)^۴}\big)\\
&\leqslant \mathbb{P}\Big(\max_{s_n+۱\leqslant k\leqslant p_n}\Big|\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲-\mathrm{E}\,\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲\Big|\\
&\quad +\max_{s_n+۱\leqslant k\leqslant p_n}\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲>C\frac{n\lambda_n^۲}{s_n(\log n)^۴}\Big).
\end{aligned}$$

بنابراین داریم $\max_{s_n+۱\leqslant k\leqslant p_n}\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲=\max_{s_n+۱\leqslant k\leqslant p_n}\mathrm{E}\left(\sum_{k'=۱}^{s_n}H_{nkk'}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\right)\leqslant Cs_n$. از آنجایی که تحت شرط (۷.A)، $s_n^۲(\log n)^۴=o(n\lambda_n^۲)$ و $p_ns_n^۳(\log n)^۸/(n^۲\lambda_n^۴)=o(۱)$ برقرار است، داریم

$$\begin{aligned}
I_{n۲۱} &\leqslant \mathbb{P}\Big(\max_{s_n+۱\leqslant k\leqslant p_n}\Big|\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲-\mathrm{E}\,\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲\Big|>\frac{C}{۲}\frac{n\lambda_n^۲}{s_n(\log n)^۴}\Big)\\
&\leqslant \sum_{k=s_n+۱}^{p_n}\mathbb{P}\Big(\Big|\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲-\mathrm{E}\,\|\boldsymbol{H}_{nk۱}(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\|^۲\Big|>\frac{C}{۲}\frac{n\lambda_n^۲}{s_n(\log n)^۴}\Big)\\
&\leqslant C\sum_{k=s_n+۱}^{p_n}\frac{\mathrm{E}\left[\sum_{k'=۱}^{s_n}\left(H_{nkk'}^۲(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)-\mathrm{E}\left(H_{nkk'}^۲(\boldsymbol{\theta}_{n۰},\boldsymbol{u}_i)\right)\right)\right]^۲s_n^۲(\log n)^۸}{n^۲\lambda_n^۴}\\
&= O(p_ns_n^۳(\log n)^۸/(n^۲\lambda_n^۴))=o(۱),
\end{aligned}$$

به طوری که نامساوی سوم از نامساوی مارکوف نتیجه می‌شود. به طور مشابه می‌توان نشان داد که $I_{n۲۲}=o(۱)$ و $I_{n۲۳}=o(۱)$. لذا عبارت $I_{n۲}=o(۱)$ اثبات می‌شود.

در پایان اثبات می‌کنیم که $I_{n۳} = o(۱)$. نا مساوی‌های زیر را در نظر بگیرید

$$\begin{aligned} I_{n۳} &\leqslant \mathbb{P}\Big(\max_{s_n+۱\leqslant k\leqslant p_n} |(\widehat{\boldsymbol{\theta}}_{n۱} - \boldsymbol{\theta}_{n۱۰})^\top \boldsymbol{\nabla}_{k۱}(\boldsymbol{\theta}_n^*, \boldsymbol{u}_i)(\widehat{\boldsymbol{\theta}}_{n۱} - \boldsymbol{\theta}_{n۱۰})| > \frac{\lambda_n}{۶\log n}\Big) \\ &\leqslant \mathbb{P}\Big(\max_{s_n+۱\leqslant k\leqslant p_n} |(\widehat{\boldsymbol{\theta}}_{n۱} - \boldsymbol{\theta}_{n۱۰})^\top \boldsymbol{\nabla}_{k۱}(\boldsymbol{\theta}_n^*, \boldsymbol{u}_i)(\widehat{\boldsymbol{\theta}}_{n۱} - \boldsymbol{\theta}_{n۱۰})| > \frac{\lambda_n}{۶\log n}, \\ &\quad \|\widehat{\boldsymbol{\theta}}_{n۱} - \boldsymbol{\theta}_{n۱۰}\| \leqslant \sqrt{s_n/n}\log n\Big) + \mathbb{P}\Big(\|\widehat{\boldsymbol{\theta}}_{n۱} - \boldsymbol{\theta}_{n۱۰}\| > \sqrt{s_n/n}\log n\Big) \\ &\leqslant \sum_{k=s_n+۱}^{p_n} \mathbb{P}\Big(\operatorname{tr}(\boldsymbol{\nabla}_{k۱}(\boldsymbol{\theta}_n^*, \boldsymbol{u}_i)) > \frac{n\lambda_n}{s_n(\log n)^۳}\Big) + o(۱) \\ &\leqslant C \sum_{k=s_n+۱}^{p_n} \frac{\mathrm{E}\big[\operatorname{tr}(\boldsymbol{\nabla}_{k۱}(\boldsymbol{\theta}_n^*, \boldsymbol{u}_i))^۲\big] s_n^۲ (\log n)^۶}{n^۲\lambda_n^۲} + o(۱), \end{aligned}$$

به طوری که نامساوی سوم از عبارت (ب.۹) و نامساوی آخر از نامساوی مارکوف نتیجه می شوند. تحت شرایط (۲.A)، (۴.A)، (۵.A) و (۶.A) داریم

$$\mathrm{E}\Big[\operatorname{tr}(\boldsymbol{\nabla}_{k۱}(\boldsymbol{\theta}_n^*, \boldsymbol{u}_i))^۲\Big] = \mathrm{E}\Big[\sum_{k'=۱}^{s_n} \frac{\partial^۲ \overline{S}_{nk}}{\partial \beta_{nk'}^۲}(\boldsymbol{\theta}_n^*, \boldsymbol{u}_i)\Big]^۲ \leqslant C s_n^۲.$$

از آنجایی که تحت شرط (۷.A) $p_n s_n^۴ (\log n)^۶ = (n^۲\lambda_n^۲) = o(۱)$ برقرار است لذا $I_{n۳} = O(p_n s_n^۴ (\log n)^۶ = (n^۲\lambda_n^۲)) + o(۱) = o(۱)$.

با جمع‌آوری همه شرایط مورد نیاز، قضیه اثبات می‌شود.

ب.۴.۳ اثبات قضیه ۲.۱.۳

قسمت (i): کافی است نشان دهیم، به ازای هر $\epsilon > ۰$، یک ثابت $\Delta > ۰$ وجود دارد که برای n به اندازه کافی بزرگ، داریم

$$\mathbb{P}\left(\sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰}\| = \Delta\sqrt{p_n/n}} (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰})^\top \boldsymbol{U}_n(\boldsymbol{\theta}_n) < ۰ \right) \geqslant ۱ - \epsilon.$$

لذا لازم است که علامت $(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰})^\top \boldsymbol{U}_n(\boldsymbol{\theta}_n)$ را روی $\{\boldsymbol{\theta}_n : \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰}\| = \Delta\sqrt{p_n/n}\}$ بررسی کنیم. به ازای $\boldsymbol{\theta}_n^*$ بین $\boldsymbol{\theta}_{n۰}$ and $\boldsymbol{\theta}_n$، یعنی به ازای $۰ < t < ۱$ داشته باشیم $\boldsymbol{\theta}_n^* = t\boldsymbol{\theta}_n + (۱-t)\boldsymbol{\theta}_{n۰}$، داریم

$$\begin{aligned} (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰})^\top \boldsymbol{U}_n(\boldsymbol{\theta}_n) &= (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰})^\top \boldsymbol{U}_n(\boldsymbol{\theta}_{n۰}) - (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰})^\top \boldsymbol{D}_n(\boldsymbol{\theta}_n^*)(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰}) \\ &= J_{n۱} + J_{n۲}. \end{aligned}$$

برای $J_{n۱}$ می‌توان نوشت

$$J_{n۱} = (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰})^\top \overline{\boldsymbol{U}}_n(\boldsymbol{\theta}_{n۰}) + (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n۰})^\top [\boldsymbol{U}_n(\boldsymbol{\theta}_{n۰}) - \overline{\boldsymbol{U}}_n(\boldsymbol{\theta}_{n۰})] = J_{n۱۱} + J_{n۱۲}.$$

بر اساس نامساوی کوشی شوارتز داریم $|J_{n11}| \leqslant \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}\|.\|\overline{\boldsymbol{U}}_n(\boldsymbol{\theta}_{n0})\| = \Delta\sqrt{p_n/n}\|\overline{\boldsymbol{U}}_n(\boldsymbol{\theta}_{n0})\|$. بنابراین

$$\begin{aligned}
&\mathrm{E}\Big[\|\overline{\boldsymbol{U}}_n(\boldsymbol{\theta}_{n0})\|^2\Big] \\
&\quad= \mathrm{E}\Big[\sum_{i=1}^{n} \boldsymbol{\epsilon}_i^\top(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\boldsymbol{D}_i\boldsymbol{D}_i^\top\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{\epsilon}_i^\top(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\Big] \\
&\quad\leqslant \sum_{i=1}^{n}\lambda_{\max}(\boldsymbol{D}_i\boldsymbol{D}_i^\top)\lambda_{\max}(\boldsymbol{A}_i(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i))\lambda_{\max}(\overline{\boldsymbol{R}}^{-2})\,\mathrm{E}\Big[\boldsymbol{\epsilon}_i^\top(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\boldsymbol{\epsilon}_i(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\Big] \\
&\quad\leqslant C\,\mathrm{tr}\,\Big(\sum_{i=1}^{n}\boldsymbol{D}_i\boldsymbol{D}_i^\top\Big) = C\sum_{i=1}^{n}\sum_{j=1}^{n_i}\boldsymbol{D}_{ij}^\top\boldsymbol{D}_{ij} = O(np_n).
\end{aligned}$$

از عبارات فوق $\|\overline{\boldsymbol{U}}_n(\boldsymbol{\theta}_{n0})\| = O_p(\sqrt{np_n})$ و به دنبال آن $|J_{n11}| = \Delta O_p(p_n)$ را می‌توان نتیجه گرفت. برای عبارت J_{n12} بر اساس لم ب.۱.۳ می‌توان نشان داد که

$$|J_{n12}| \leqslant \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}\|.\|\boldsymbol{U}_n(\boldsymbol{\theta}_{n0}) - \overline{\boldsymbol{U}}_n(\boldsymbol{\theta}_{n0})\| = \Delta\sqrt{p_n/n}O_p(p_n) = \Delta o_p(p_n).$$

بنابراین تساوی $|J_{n1}| = \Delta O_p(p_n)$ برقرار است. در ادامه J_{n2} را می‌توان به صورت

$$\begin{aligned}
J_{n2} &= -(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0})^\top\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n^*)(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}) - (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0})^\top[\boldsymbol{D}_n(\boldsymbol{\theta}_n^*) - \overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n^*)](\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}) \\
&= J_{n21} + J_{n22},
\end{aligned}$$

نشان داد. ابتدا بر اساس لم ب.۳.۳ می‌توان نشان داد که

$$\begin{aligned}
|J_{n22}| &\leqslant \max\Big(|\lambda_{\max}\big(\boldsymbol{D}_n(\boldsymbol{\theta}_n^*) - \overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n^*)\big)|, |\lambda_{min}\big(\boldsymbol{D}_n(\boldsymbol{\theta}_n^*) - \overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n^*)\big)|\Big) \times \|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}\|^2 \\
&= O_p(\sqrt{np_n})\Delta^2\frac{p_n}{n} = \Delta^2 o_p(p_n).
\end{aligned}$$

به عبارت دیگر

$$\begin{aligned}
J_{n21} = &- (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0})^\top\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n0})(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}) \\
&- (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0})^\top[\boldsymbol{H}_n(\boldsymbol{\theta}_n^*) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n0})](\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}) \\
&- (\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0})^\top[\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n^*) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n^*)]((\boldsymbol{\theta}_n^*)(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}) \\
&= J_{n21}^a + J_{n21}^b + J_{n21}^c.
\end{aligned}$$

بر اساس لم ب.۵.۳، عبارت $J_{n21}^b = \Delta^2 o_p(p_n)$؛ و براساس لم ب.۴.۳، عبارت $J_{n21}^c = \Delta^2 o_p(p_n)$ نتیجه می‌شود. لذا کافی است که عبارت J_{n21}^a را بررسی کنیم. تحت شرط (۳.A) داریم

$$\begin{aligned}
J_{n21}^a &= -(\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0})^\top\Big[\sum_{i=1}^{n}\boldsymbol{D}_i^\top\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i)\boldsymbol{D}_i\Big](\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}) \\
&\leqslant \lambda_{\min}(\overline{\boldsymbol{R}}^{-1})\min_i\lambda_{\min}(\boldsymbol{A}_i(\boldsymbol{\theta}_{n0},\boldsymbol{u}_i))\lambda_{\min}(\sum_{i=1}^{n}\boldsymbol{D}_i^\top\boldsymbol{D}_i)\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n0}\|^2 \leqslant -C\Delta^2 p_n.
\end{aligned}$$

بنابراین در فضای پارامتر $(\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ})^\top \boldsymbol{U}_n(\boldsymbol{\theta}_n)$، عبارت $(\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ})^\top \boldsymbol{U}_n(\boldsymbol{\theta}_n)$ در احتمال به عبارت $J_{n11}+J^a_{n21}$ همگراست که مقداری منفی است. لذا نتیجه اثبات می‌شود.
قسمت (ii): با استفاده از قضیه ۱۲.۷ در اسچوماکر (۱۹۸۱) تحت شرط (۱.A) مقدار ثابتی مانند C وجود دارد به طوری که

$$\sup_{t\in[\circ,1]}|f_\circ(\boldsymbol{t})-\boldsymbol{B}(\boldsymbol{t})\boldsymbol{\alpha}_\circ|\leqslant CK_n^{-m}.$$

حال با انتخاب $K_n\approx n^{1/(2r+1)}$، داریم

$$\begin{aligned}\big(\widehat{f}(t)-f_\circ(t)\big)^2 &= \big(\boldsymbol{B}(t)\widehat{\boldsymbol{\alpha}}_n-f_\circ(t)\big)^2\simeq\big(\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ}-f_\circ(t)\big)^2\\ &\leqslant |\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ}-f_\circ(t)||\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ}-f_\circ(t)|\\ &\leqslant \sup_{t\in[\circ,1]}|\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ}-f_\circ(t)|\sup_{t\in[\circ,1]}|\boldsymbol{B}(t)\boldsymbol{\alpha}_{n\circ}-f_\circ(t)|\\ &\leqslant CL_n^{-r}CL_n^{-r}=C^2n^{-2r/(2r+1)}=O_p(n^{-2r/(2r+1)}),\end{aligned}$$

و اثبات کامل می‌شود.

ب.۵.۳ اثبات قضیه ۳.۱.۳

قسمت (i): ابتدا نشان می‌دهیم که $\widehat{\boldsymbol{\theta}}_n$ در شرط تنکی، $\widehat{\boldsymbol{\theta}}_{n2}=\circ$، صدق می‌کند. برای این که نشان دهیم به ازای $\widehat{\boldsymbol{\theta}}_{n2}=\circ$، جواب بهینه بدست می‌آید، کافی است نشان دهیم هنگامی که $n\to\infty$، برای هر $\widehat{\boldsymbol{\theta}}_{n1}$ و $\widehat{\boldsymbol{\theta}}_{n2}$ که به ترتیب در شرایط $\|\widehat{\boldsymbol{\theta}}_{n1}-\widehat{\boldsymbol{\theta}}_{n1\circ}\|=O_p(\sqrt{p_n/n})$ و $\|\widehat{\boldsymbol{\theta}}_{n2}\|\leqslant C(\sqrt{p_n/n})$ صدق می‌کنند، عبارت $U_{nk}(\boldsymbol{\theta})$ برای $\beta_k\in\big(-C(\sqrt{p_n/n}),C(\sqrt{p_n/n})\big)$ که $k=s_n+1,\ldots,p_n$، با احتمال همگرا به ۱ دارای علامت‌های متفاوت هستند. عبارت $U_{nk}(\boldsymbol{\theta})$ را به صورت

$$S_{nk}(\widehat{\boldsymbol{\theta}})=O_p(\sqrt{p_n/n}),$$

در نظر بگیرید، بنابراین می‌توان نشان داد که

$$U_{nk}(\boldsymbol{\theta})=n\lambda_n\Big\{\frac{q_{\lambda_n}(|\beta_k|)}{\lambda_n}\operatorname{sgn}(\beta_k)+O_p(\sqrt{p_n/n})\Big\}.$$

از آنجایی که $\sqrt{p_n/n}/\lambda_n\to\circ$ و $\liminf_{n\to\infty}\liminf_{\beta_k\to\circ+}\frac{q_{\lambda_n}(|\beta_k|)}{\lambda_n}>\circ$، می‌توان نشان داد که علامت β_k تعیین کننده علامت $U_{nk}(\boldsymbol{\theta})$ است. لذا نتیجه مورد نظر حاصل می‌شود.
قسمت (ii): ابتدا نشان می‌دهیم که $\boldsymbol{\xi}_n^\top\overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n\circ})(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n\circ1})\xrightarrow{d}\mathcal{N}_{s_n}(\circ,1)$. می‌توان نشان داد

$$\begin{aligned}&\boldsymbol{\xi}_n^\top\overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ})\\ =\;&\boldsymbol{\xi}_n^\top\overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})+\boldsymbol{\xi}_n^\top\overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ})-\boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})]\\ =\;&\boldsymbol{\xi}_n^\top\overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\boldsymbol{D}_n(\boldsymbol{\theta}_n^*)(\widehat{\boldsymbol{\theta}_n}-\boldsymbol{\theta}_{n\circ})+\boldsymbol{\xi}_n^\top\overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ})-\boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})]\end{aligned}$$

$$
\begin{aligned}
= \ & \boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n\circ})(\widehat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_{n\circ}) \\
& + \boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\boldsymbol{D}_n(\boldsymbol{\theta}_n^*) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n\circ})](\widehat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_{n\circ}) \\
& + \boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})],
\end{aligned}
$$

به طوری که برای رسیدن به تساوی دوم فرض می‌شود $\boldsymbol{S}_n(\hat{\boldsymbol{\theta}}_n) = 0$ و برای $\boldsymbol{\theta}_n^*$ بین $\widehat{\boldsymbol{\theta}_n}$ و $\boldsymbol{\theta}_{n\circ}$ و بر اساس بسط تیلور داریم $\boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ}) = \boldsymbol{D}_n(\boldsymbol{\theta}_n^*)(\widehat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_{n\circ})$. بر اساس لم ب.۶.۳، داریم $\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) \xrightarrow{\mathcal{D}} N_{s_n}(0, 1)$. بنابراین برای اثبات قضیه کافی است نشان دهیم که به ازای هر $\Delta > 0$، عبارت‌های

$$
\sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n\circ}\| \leqslant \Delta\sqrt{p_n/n}} |\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\boldsymbol{D}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n\circ})](\widehat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_{n\circ})| = o_p(1) \qquad (ب.۱۴)
$$

و

$$
|\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})]| = o_p(1) \qquad (ب.۱۵)
$$

برقرار است. ابتدا عبارت (ب.۱۵) را اثبات می‌کنیم. در نظر بگیرید که

$$
\begin{aligned}
& \big[\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})]\big]^2 \\
= \ & \boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})][\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})]^\top \overline{\boldsymbol{M}}_n^{-1/2}\boldsymbol{\xi}_n \\
\leqslant \ & \lambda_{\max}\big(\overline{\boldsymbol{M}}_n^{-1}(\boldsymbol{\theta}_{n\circ})\big)\lambda_{\max}\big([\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})][\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})]^\top\big) \\
\leqslant \ & \frac{\|\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})\|^2}{\lambda_{\min}\big(\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ})\big)} \\
\leqslant \ & \frac{\|\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}_{n\circ}) - \boldsymbol{S}_n(\boldsymbol{\theta}_{n\circ})\|^2}{C\lambda_{\min}\big(\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{X}_i\big)} = O_p(p_n^2/n) = o_p(1).
\end{aligned}
$$

بر اساس لم ب.۱.۳ و این که

$$
\lambda_{\min}\big(\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ})\big) \geqslant C\lambda_{\min}\big(\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{X}_i\big), \qquad (ب.۱۶)
$$

می‌توان عبارت (ب.۱۵) را مشابه لم ب.۶.۳ اثبات کرد. حال به بررسی عبارت (ب.۱۴) می‌پردازیم. تساوی‌های و نامساوی‌های

$$
\begin{aligned}
& \sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n\circ}\| \leqslant \Delta\sqrt{p_n/n}} |\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\boldsymbol{D}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n\circ})](\widehat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_{n\circ})| \\
\leqslant \ & \sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n\circ}\| \leqslant \Delta\sqrt{p_n/n}} |\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\boldsymbol{D}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n)](\widehat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_{n\circ})| \\
+ \ & \sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n\circ}\| \leqslant \Delta\sqrt{p_n/n}} |\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n)](\widehat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_{n\circ})| \\
+ \ & \sup_{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n\circ}\| \leqslant \Delta\sqrt{p_n/n}} |\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})[\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_n) - \overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n\circ})](\widehat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_{n\circ})| \\
= \ & K_{n1} + K_{n2} + K_{n3},
\end{aligned}
$$

را در نظر بگیرید. براساس نامساوی کوشی شوارتز و نتیجه ب.۱.۳، داریم

$$\begin{aligned} K_{n1} &\leqslant \sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}} [\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\big(\boldsymbol{D}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n)\big)^2 \\ &\qquad \times \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\boldsymbol{\xi}_n]^{1/2}\|\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{n\circ}\| \\ &\leqslant \sup_{\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_{n\circ}\|\leqslant\Delta\sqrt{p_n/n}} \max\big(|\lambda_{\min}(\boldsymbol{D}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n))|, |\lambda_{\max}(\boldsymbol{D}_n(\boldsymbol{\theta}_n)-\overline{\boldsymbol{D}}_n(\boldsymbol{\theta}_n))|\big) \\ &\qquad \times \lambda_{\min}^{-1/2}\big(\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ})\big) O_p(p_n^{1/2}n^{-1/2}) \\ &= O_p(\sqrt{n}p_n)O(n^{-1/2})O_p(p_n^{1/2}n^{-1/2}) = O_p(n^{-1/2}p_n^{3/2}) = o_p(1). \end{aligned}$$

بنابراین $K_{n1}=o_p(1)$. با استدلال مشابه لم‌های ب.۴.۳ و ب.۵.۳، می‌توان نشان داد که $K_{n2}=o_p(1)$ و $K_{n3}=o_p(1)$. لذا عبارت (ب.۱۴) اثبات می‌شود. تا این مرحله نشان دادیم که

$$\boldsymbol{\xi}_n^\top \overline{\boldsymbol{M}}_n^{-1/2}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{H}}_n(\boldsymbol{\theta}_{n\circ})(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n\circ 1}) \xrightarrow{\mathcal{D}} \mathcal{N}_{s_n}(0,1),$$

بنابراین $(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n\circ 1}) \xrightarrow{\mathcal{D}} \mathcal{N}_{s_n}\big(0, \overline{\boldsymbol{H}}_n^{-1^\top}(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{M}}_n(\boldsymbol{\theta}_{n\circ})\overline{\boldsymbol{H}}_n^{-1}(\boldsymbol{\theta}_{n\circ})\big)$. بدیهی است که

$$(\widehat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_{n\circ 1}) = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{n1}-\boldsymbol{\beta}_{n\circ 1} \\ \widehat{\boldsymbol{\alpha}}_{n1}-\boldsymbol{\alpha}_{n\circ 1} \end{pmatrix} \xrightarrow{\mathcal{D}} N_2\left(\begin{pmatrix} \mathbf{0}_{s_n} \\ \mathbf{0}_{h_n} \end{pmatrix}, \begin{pmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \\ \boldsymbol{Q}_3 & \boldsymbol{Q}_4 \end{pmatrix} \right).$$

در نتیجه برای اثبات قضیه کافی است نشان دهیم که $\boldsymbol{Q}_1 = \overline{\boldsymbol{H}}_n^{*-1^\top}(\boldsymbol{\beta}_{n\circ})\overline{\boldsymbol{M}}_n^{*}(\boldsymbol{\beta}_{n\circ})\overline{\boldsymbol{H}}_n^{*-1}(\boldsymbol{\beta}_{n\circ})$. $\overline{\boldsymbol{H}}_n$ را به صورت

$$\begin{aligned} \overline{\boldsymbol{H}}_n &= \begin{pmatrix} \boldsymbol{X}^\top \\ \boldsymbol{B}^\top \end{pmatrix} \boldsymbol{\Omega} \begin{pmatrix} \boldsymbol{X} & \boldsymbol{B} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X} & \boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{B} \\ \boldsymbol{B}^\top\boldsymbol{\Omega}\boldsymbol{X} & \boldsymbol{B}^\top\boldsymbol{\Omega}\boldsymbol{B} \end{pmatrix}, \\ \overline{\boldsymbol{M}}_n &= \begin{pmatrix} \boldsymbol{X}^\top \\ \boldsymbol{B}^\top \end{pmatrix} \boldsymbol{\Omega}_\circ \begin{pmatrix} \boldsymbol{X} & \boldsymbol{B} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^\top\boldsymbol{\Omega}_\circ\boldsymbol{X} & \boldsymbol{X}^\top\boldsymbol{\Omega}_\circ\boldsymbol{B} \\ \boldsymbol{B}^\top\boldsymbol{\Omega}_\circ\boldsymbol{X} & \boldsymbol{B}^\top\boldsymbol{\Omega}_\circ\boldsymbol{B} \end{pmatrix}, \end{aligned}$$

بازتعریف می‌کنیم، به طوری که $\boldsymbol{\Omega}_\circ = \mathrm{diag}\{\boldsymbol{\Omega}_{\circ i}\}$ و $\boldsymbol{\Omega}_{\circ i} = \mathrm{E}_{u|y}\left[\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\overline{\boldsymbol{R}}^{-1}\boldsymbol{R}_\circ\overline{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{\frac{1}{2}}(\boldsymbol{\theta}_n,\boldsymbol{u}_i)\right]$. این نشان می‌دهد که

$$\overline{\boldsymbol{H}}_n^{-1} = \begin{pmatrix} (\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X})^{-1} + (\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{B}^\top\boldsymbol{\Omega}\boldsymbol{X}Q\boldsymbol{B}^\top\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X}), & \boldsymbol{Q}^* \\ \boldsymbol{Q}^{*\top}, & \boldsymbol{Q} \end{pmatrix},$$

به طوری که $\boldsymbol{Q} = \left(\boldsymbol{B}^\top\boldsymbol{\Omega}\boldsymbol{B} - \boldsymbol{B}^\top\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{B}\right)^{-1}$ و $\boldsymbol{Q}^* = -(\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\boldsymbol{B}\boldsymbol{Q}$.

با کمی محاسبات جبری می‌توان Q_1 را به صورت

$$\begin{aligned} Q_1 &= (I - Q^* B^\top \Omega X)(X^\top \Omega X)^{-1} \\ &\times \Big[X^\top \Omega_\circ X + X^\top \Omega_\circ B Q^{*\top} (X^\top \Omega X)(I - Q^* B^\top \Omega X)^{-1} \\ &+ (X^\top \Omega X)(I - Q^* B^\top \Omega X)^{-1} Q^* B^\top \Omega_\circ X \\ &+ (X^\top \Omega X)(I - Q^* B^\top \Omega X)^{-1} Q^* B^\top \Omega_\circ B Q^{*\top} (X^\top \Omega X)(I - Q^* B^\top \Omega X)^{-1} \Big] \\ &\times (I - Q^* B^\top \Omega X)(X^\top \Omega X)^{-1}, \end{aligned}$$

بیان کرد. بنابراین کافی است نشان دهیم که

$$(I - Q^* B^\top \Omega X)(X^\top \Omega X)^{-1} = \Big(X^\top \Omega X (I - Q^* B^\top \Omega X)^{-1} \Big)^{-1} = \overline{H}_n^{*-1} \quad (۱۷.ب)$$

و

$$\begin{aligned} X^\top \Omega_\circ X &+ X^\top \Omega_\circ B Q^{*\top} (X^\top \Omega X)(I - Q^* B^\top \Omega X)^{-1} \\ &+ (X^\top \Omega X)(I - Q^* B^\top \Omega X)^{-1} Q^* B^\top \Omega_\circ X \\ &+ (X^\top \Omega X)(I - Q^* B^\top \Omega X)^{-1} Q^* B^\top \Omega_\circ B Q^{*\top} (X^\top \Omega X)(I - Q^* B^\top \Omega X)^{-1} \\ &= \overline{M}_n^*. \end{aligned} \quad (۱۸.ب)$$

هر دو عبارت (ب.۱۷) و (ب.۱۸) شامل جمله $(I - Q^* B^\top \Omega X)^{-1}$ هستند که

$$(I - Q^* B^\top \Omega X)^{-1} = I - (X^\top \Omega X)^{-1} X^\top \Omega B (B^\top \Omega B)^{-1} B^\top \Omega X.$$

با استفاده از

$$X^\top \Omega X (I - Q^* B^\top \Omega X)^{-1} = X^\top \Omega X - X^\top \Omega B (B^\top \Omega B)^{-1} B^\top \Omega X = \overline{H}_n^*,$$

عبارت (ب.۱۷) اثبات می‌شود. همچنین با استفاده از

$$X^\top \Omega X (I - Q^* B^\top \Omega X)^{-1} Q^* = -X^\top \Omega B (B^\top \Omega B)^{-1},$$

عبارت (ب.۱۸) را می‌توان به صورت

$$\begin{aligned} X^\top \Omega_\circ X &- X^\top \Omega_\circ B (B^\top \Omega B)^{-1} B^\top \Omega X \\ &- X^\top \Omega B (B^\top \Omega B)^{-1} B^\top \Omega_\circ X \\ &+ X^\top \Omega B (B^\top \Omega B)^{-1} B^\top \Omega_\circ B (B^\top \Omega B)^{-1} B^\top \Omega X = \overline{M}_n^*, \end{aligned}$$

بیان کرد. لذا (ب.۱۸) نیز برقرار بوده و اثبات قضیه کامل می‌شود.

ب.۴ اثبات قضایای فصل ۴

ب.۱.۴ بردار رتبه و ماتریس اطلاع فیشر

مشتق اول جزئی از لگاریتم تابع درستنمایی تاوانیده نسبت به هر یک از اعضای پارامتر $\boldsymbol{\Theta}$ را بردار رتبه ($\boldsymbol{S}_{\boldsymbol{\Theta}}$) می‌گویند که اعضای آن شامل

$$\begin{aligned}
\boldsymbol{S}_{\boldsymbol{\beta}} &= \sum_{i=۱}^{n} \Big(\frac{\nu + n_i}{\nu + \Delta_i}\Big) \boldsymbol{X}_i^{\top} \boldsymbol{\Lambda}_i^{-۱} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}), \\
[\boldsymbol{S}_{\boldsymbol{\omega}}]_l &= -\frac{۱}{۲} \sum_{i=۱}^{n} \Big\{ \operatorname{tr}(\boldsymbol{\Lambda}_i^{-۱} \overset{\circ}{\boldsymbol{\Lambda}}_{il}) - \Big(\frac{\nu + n_i}{\nu + \Delta_i}\Big) (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^{\top} \boldsymbol{\Lambda}_i^{-۱} \overset{\circ}{\boldsymbol{\Lambda}}_{il} \boldsymbol{\Lambda}_i^{-۱} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \Big\}, \\
s_{\nu} &= \frac{۱}{۲} \sum_{i=۱}^{n} \Big\{ \mathcal{D}_g\Big(\frac{\nu + n_i}{۲}\Big) - \mathcal{D}_g\Big(\frac{\nu}{۲}\Big) - \frac{n_i}{\nu} - \log\Big(۱ + \frac{\Delta_i}{\nu}\Big) + \frac{(\nu + n_i)\Delta_i}{(\nu + \Delta_i)\nu} \Big\},
\end{aligned}$$

است، که در آن $l = ۱, \ldots, g$، $g = q(q+۱)/۲ + r(r+۱)/۲ + ۲$، $\mathcal{D}_g(x) = \partial \Gamma(x)/\partial x$ تابع دیاگاما بوده و

$$\overset{\circ}{\boldsymbol{\Lambda}}_{il} = \frac{\partial \Delta_i(\boldsymbol{D}, \boldsymbol{\Sigma}, \boldsymbol{\phi})}{\partial \boldsymbol{\omega}_l} = \begin{cases} \boldsymbol{Z}_i \frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\omega}_l} \boldsymbol{Z}_i^{\top} & \text{if } \boldsymbol{\omega}_l = \operatorname{vec}(\boldsymbol{D}), \\ \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\omega}_l} \otimes \boldsymbol{\Omega}_i & \text{if } \boldsymbol{\omega}_l = \operatorname{vec}(\boldsymbol{\Sigma}), \\ \boldsymbol{\Sigma} \otimes \frac{\partial \boldsymbol{\Omega}_i}{\partial \boldsymbol{\omega}_l} & \text{if } \boldsymbol{\omega}_l = \boldsymbol{\phi}. \end{cases}$$

همچنین امید مشتق دوم جزئی از لگاریتم تابع درستنمایی تاوانیده نسبت به هر یک از اعضای پارامتر $\boldsymbol{\Theta}$ را ماتریس اطلاع فیشر ($\boldsymbol{J}_{\boldsymbol{\Theta}}$) می‌گویند که اعضای آن به صورت

$$\boldsymbol{J}_{\boldsymbol{\Theta}\boldsymbol{\Theta}} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\eta}} \\ \boldsymbol{J}_{\boldsymbol{\eta}\boldsymbol{\beta}}^{\top} & \boldsymbol{J}_{\boldsymbol{\eta}\boldsymbol{\eta}} \end{bmatrix},$$

تعیین می‌شود، به طوری که

$$\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \sum_{i=۱}^{n} \frac{(\nu + n_i)}{(\nu + n_i + ۲)} \boldsymbol{X}_i^{\top} \boldsymbol{\Lambda}_i^{-۱} \boldsymbol{X}_i, \; \boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\eta}} = \boldsymbol{m}\circ, \quad \text{and} \quad \boldsymbol{J}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\omega}\boldsymbol{\omega}} & \boldsymbol{J}_{\boldsymbol{\omega}\nu} \\ \boldsymbol{J}_{\nu\boldsymbol{\omega}}^{\top} & \boldsymbol{J}_{\nu\nu} \end{bmatrix}$$

و

$$\begin{aligned}
[\boldsymbol{J}_{\boldsymbol{\omega}\boldsymbol{\omega}}]_{ls} &= \frac{۱}{۲} \sum_{i=۱}^{n} \frac{۱}{\nu + n_i + ۲} \Big((\nu + n_i) \operatorname{tr}\big(\boldsymbol{\Lambda}_i^{-۱} \overset{\circ}{\boldsymbol{\Lambda}}_{il} \boldsymbol{\Lambda}_i^{-۱} \overset{\circ}{\boldsymbol{\Lambda}}_{is}\big) - \operatorname{tr}\big(\boldsymbol{\Lambda}_i^{-۱} \overset{\circ}{\boldsymbol{\Lambda}}_{il}\big) \operatorname{tr}\big(\boldsymbol{\Lambda}_i^{-۱} \overset{\circ}{\boldsymbol{\Lambda}}_{is}\big) \Big), \\
[\boldsymbol{J}_{\boldsymbol{\omega}\nu}]_l &= -\sum_{i=۱}^{n} \frac{۱}{(\nu + n_i)(\nu + n_i + ۲)} \operatorname{tr}\big(\boldsymbol{\Lambda}_i^{-۱} \overset{\circ}{\boldsymbol{\Lambda}}_{il}\big), \\
\boldsymbol{J}_{\nu\nu} &= \frac{۱}{۴} \sum_{i=۱}^{n} \Big(\mathcal{T}_g\big(\frac{\nu}{۲}\big) - \mathcal{T}_g\big(\frac{\nu + n_i}{۲}\big) - \frac{۲ n_i (\nu + n_i + ۴)}{\nu(\nu + n_i)(\nu + n_i + ۲)} \Big),
\end{aligned}$$

که در آن $l, s = ۱, \ldots, g$ و $\mathcal{T}_g(x) = \partial^۲ \log\Gamma(x)/\partial x^۲$ تابع تری‌گاما است. به علاوه $\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ را می‌توان به صورت

$$\boldsymbol{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\beta}_۱\boldsymbol{\beta}_۱} & \boldsymbol{J}_{\boldsymbol{\beta}_۱\boldsymbol{\beta}_۲} \\ \boldsymbol{J}_{\boldsymbol{\beta}_۲\boldsymbol{\beta}_۱}^\top & \boldsymbol{J}_{\boldsymbol{\beta}_۲\boldsymbol{\beta}_۲} \end{bmatrix},$$

تفکیک کرد، به طوری که

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\beta}_۱\boldsymbol{\beta}_۱} &= \textstyle\sum_{i=۱}^n \frac{(\nu+n_i)}{(\nu+n_i+۲)} \boldsymbol{X}_{i۱}^\top \boldsymbol{\Lambda}_{i۱۱}^{-۱} \boldsymbol{X}_{i۱}, \\
\boldsymbol{J}_{\boldsymbol{\beta}_۲\boldsymbol{\beta}_۲} &= \textstyle\sum_{i=۱}^n \frac{(\nu+n_i)}{(\nu+n_i+۲)} \boldsymbol{X}_{i۲}^\top \boldsymbol{\Lambda}_{i۲۲}^{-۱} \boldsymbol{X}_{i۲}, \\
\boldsymbol{J}_{\boldsymbol{\beta}_۱\boldsymbol{\beta}_۲} &= \textstyle\sum_{i=۱}^n \frac{(\nu+n_i)}{(\nu+n_i+۲)} \boldsymbol{X}_{i۱}^\top \boldsymbol{\Lambda}_{i۱۲}^{-۱} \boldsymbol{X}_{i۲};
\end{aligned}$$

و $\boldsymbol{\Lambda}_{i۱۱}$، $\boldsymbol{\Lambda}_{i۲۲}$، $\boldsymbol{\Lambda}_{i۱۲}$ عضوهای

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} \boldsymbol{\Lambda}_{i۱۱} & \boldsymbol{\Lambda}_{i۱۲} \\ \boldsymbol{\Lambda}_{i۱۲} & \boldsymbol{\Lambda}_{i۲۲} \end{bmatrix}.$$

هستند.

ب.۲.۴ اثبات قضیه ۲.۱.۴

مشابه فن و لی (۲۰۰۱)، می‌توان نشان داد که با احتمال همگرا به ۱، $\widehat{\boldsymbol{\beta}}$ در شرط

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| = O_p(n^{-۱/۲} + a_n).$$

صدق می‌کند. این نشان می‌دهد که نرخ همگرایی $\widehat{\boldsymbol{\beta}}$ به پارامتر تاوان λ_n وابسته است. برای اینکه نشان دهیم $\widehat{\boldsymbol{\beta}}$ یک برآوردگر $\sqrt{n}$-سازگار است، فرض می‌کنم $a_n = O_p(n^{-۱/۲})$ و پارامتر تاوان λ_n را مقداری بسیار کوچک در نظر می‌گیریم. ابتدا نشان می‌دهیم که $\widehat{\boldsymbol{\beta}}$ در شرط تنکی، $\widehat{\boldsymbol{\beta}}_۲ = ۰$، صدق می‌کند. فرض کنید که U_k و S_k، به ترتیب مشتق اول ℓ_c^{Pen} و ℓ_c نسبت به β_k هستند. برای این که نشان دهیم به ازای $\widehat{\boldsymbol{\beta}}_۲ = ۰$، جواب بهینه بدست می‌آید، کافی است نشان دهیم هنگامی که $n \to \infty$، برای هر $\boldsymbol{\beta}_۱$ که در شرط $\|\widehat{\boldsymbol{\beta}}_۱ - \boldsymbol{\beta}_۱\| = O_p(n^{-۱/۲} + a_n)$ صدق می‌کند، عبارت $U_k(\boldsymbol{\beta})$ برای $k = s+۱, \ldots, p$ که $\beta_k \in \big(-C(n^{-۱/۲} + a_n), C(n^{-۱/۲} + a_n)\big)$ با احتمال همگرا به ۱ دارای علامت‌های متفاوت هستند. عبارت $U_k(\boldsymbol{\beta})$ را به صورت

$$U_k(\boldsymbol{\beta}) = S_k(\widehat{\boldsymbol{\beta}}) + n q_{\lambda_n}(|\beta_k|)\,\text{sgn}(\beta_k),$$

در نظر بگیرید. می‌توان نشان داد $S_k(\widehat{\boldsymbol{\beta}}) = O_p(n^{-۱/۲} + a_n)$. بنابراین داریم

$$U_k(\boldsymbol{\beta}) = n\lambda_n \Big\{ \frac{q_{\lambda_n}(|\beta_k|)}{\lambda_n}\,\text{sgn}(\beta_k) + O_p(n^{-۱/۲} + a_n) \Big\}.$$

از آنجایی که $O_p(n^{-۱/۲} + a_n) \to ۰$ و $\liminf_{n\to\infty} \liminf_{\beta_k \to ۰^+} \frac{q_{\lambda_n}(|\beta_k|)}{\lambda_n} > ۰$، می‌توان نشان داد که علامت β_k تعیین کننده علامت $U_{nk}(\boldsymbol{\theta})$ است. لذا نتیجه مورد نظر حاصل می‌شود.

ب.۳.۴ اثبات قضیه ۱.۲.۴

با استفاده از قضیه ۱۲.۷ در اسچوماکر (۱۹۸۱) می‌توان $\boldsymbol{g}_j(\boldsymbol{t}_i)$ را توسط $\boldsymbol{B}_j(\boldsymbol{t}_i)\boldsymbol{\alpha}_j$ تقریب زد. سپس با انتخاب $K_j \approx N^{1/(2r_j+1)}$، داریم

$$\begin{aligned}\big(\widehat{g}_j(t) - g_j(t)\big)^2 &= |\widehat{g}_j(t) - g_j(t)||\widehat{g}_j(t) - g_j(t)| \\ &\leqslant \sup_{t\in[0,1]}|\widehat{g}_j(t) - g_j(t)| \sup_{t\in[0,1]}|\widehat{g}_j(t) - g_j(t)| \\ &\leqslant C_j L_j^{-m_j} C_j L_j^{-m_j} = C_j^2 N^{2m_j/(2m_j+1)} = O_p(N^{2m_j/(2m_j+1)}),\end{aligned}$$

که بخش (i) قضیه اثبات می‌شود. قسمت (ii) قضیه مستقیما از طریق قانون ضعیف اعداد بزرگ، قضیه حد مرکزی و قضیه اسلاتسکی اثبات می‌شود.

Abstract

This thesis considers the estimation methods in semiparametric mixed effects regression models for longitudinal data analysis. The model is a combination of mixed effects and semiparametric models. The nonparametric function of the model is estimated using kernel, backfitting and spline methods, then consider the multicollinearity problem and introduce a new estimator called the ridge estimator. Also considers the problem of simultaneous variable selection and estimation in the generalized semiparametric mixed effects model for gaussian and non-gaussian high dimensional data. A penalization type of generalized estimating equation method is proposed while using regression spline to approximate the nonparametric component. In addition, for analyzing multivariate longitudinal data in the presence of atypical observations or outliers, a robust method based on multivariate t-distribution is offered. In each study framework, the asymptotic behavior of proposed estimators is studied and for practical implementation, an appropriate iterative algorithm is developed. The performance of the proposed methods are compared through both simulation examples and real data sets.

Keywords: Smoothing Spline, Variable selection, Ridge estimator, High dimention, Penalized, Multivariate t-distribution, Asymptotic distribution, Outliers, Longitudinal data, Heavy-tailed distribution, Backfitting, Kernel, Generalized estimating equations, Semiparametric mixed effects model.

Shahrood University of Technology

Faculty of Mathematical Sciences

PhD Thesis in: Statistics

Estimation in Semiparametric Modeling for Longitudinal Data Analysis

By: Mozhgan Taavoni

Supervisor

Mohammad Arashi

August 2020