سیزدهمین کنفرانس آمار ایران
۴-۲ شهریور ۱۳۹۵

رویکرد رگرسیون ریج در مدل‌های خطی جزئی با خطاهای همبسته

مهدی روزبه[۱]، جلال چاچی[۱]، فاطمه حمیدی[۳]

[۱]استادیار دانشگاه سمنان

[۲]دانشجوی کارشناسی ارشد دانشگاه سمنان

چکیده: در تجزیه و تحلیل مسائل رگرسیونی و بویژه مدل بندی آماری در داده های اقتصادی، روانشناسی، علوم اجتماعی، حیاتی، مهندسی و ... با مشکل همخطی در میان متغیرهای پیشگو و اتوهمبستگی درخطاها مواجه هستیم. در چنین مواقعی برآوردگر کمترین مربعات معمولی منجر به برآوردهای نادقیق از لحاظ مقدار و علامت شده و فواصل اطمینان پهنی را برای پارامترها نتیجه می دهد. هدف از این مقاله، بررسی مشکلات همخطی چندگانه و غلبه بر آن با معرفی برآوردگر رگرسیون ریج تفاضلی محدودشده تعمیم یافته و مقایسه آن با برآوردگر کمترین مربعات تعمیم یافته معمولی است. این برآوردگر، یک برآوردگر تفاضلی محدودشده تعمیم یافته است که از مینیمم کردن مجموع توان های دوم مانده‌ها و با در نظر گرفتن یک محدودیت کروی روی فضای پارامتر تولید شده است. تابع مخاطره برآوردگرهای پیشنهاد شده از تابع زیان متعادل محاسبه می‌شود. سرانجام عملکرد برآوردگر جدید با استفاده از داده‌های شبیه‌سازی شده ارزیابی می‌شود.

واژه‌های کلیدی: برآوردگر تفاضلی، برآوردگر ریج، تابع زیان متعادل، مدل خطی جزئی، محدودیت‌های خطی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62J05، 62G08، 62J07.

۱ مقدمه

یک مدل رگرسیون ناپارامتری را در نظر بگیرید

$$y_i = f(t_i) + \epsilon_i. \quad i = ۱, ۲, ...n \qquad (۱.۱)$$

که در آن y_i ها مشاهدات هستند و $f(.)$ یک تابع نامعلوم است. فرض می‌کنیم ϵ_iها متغیرهای تصادفی مستقل هم‌توزیع با میانگین صفر و واریانس $\sigma^۲$ می‌باشند و tها دارای یک دامنه چگال روی بازه‌ای به طول واحد می‌باشند و به‌ترتیب صعودی مرتب شده‌اند یعنی $t_۱ \leq t_۲ \leq \ldots \leq t_n$ که n تعدادمشاهدات در نمونه می‌باشد. همه آنچه در مورد تابع $f(.)$ می‌دانیم این است که اولین مشتق $f(.)$

[۳]نام ارائه دهنده مقاله :فاطمه حمیدی fhfh1986@gmail.com

توسط یک ثابت که L نامیده می‌شود، کران‌دار شده است. روش مرسوم در برآورد $\sigma^۲$ چنین است که ابتدا تابع $\boldsymbol{f}(.)$ را برآورد نموده و سپس با استفاده از ماتریس هموار سازی یا هت، $\sigma^۲$ را بصورت زیر برآورد می‌کنیم:

$$\sigma^۲_{\boldsymbol{Q}} = \frac{\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}}{tr(\boldsymbol{Q})} \tag{۲.۱}$$

بطوریکه $\boldsymbol{Q} = (\boldsymbol{I}-\boldsymbol{H})'(\boldsymbol{I}-\boldsymbol{H})$ که $\mathbf{H}$ یک ماتریس هموارسازی است. به این نوع برآوردگرها برآوردگرهای براساس مانده میگوئیم. برآوردگرهای براساس مانده، به ماتریس هموارسازی وابسته‌اند و انتخاب پارامتر هموارسازی یک مساله عملی مشکل است. در نوع دیگر برآوردگرها که به برآورگرهای تفاضلی مشهور می‌باشند ابتدا روندی را که توسط تابع $\boldsymbol{f}(.)$ در داده‌ها بوجود می‌آید حذف نموده و سپس به برآورد $\boldsymbol{\sigma}^۲$ می‌پردازیم. در این روش به برآورد تابع $\boldsymbol{f}(.)$ نیاز نداشته و بطور مستقیم $\boldsymbol{\sigma}^۲$ را برآورد می‌کنیم.

هم‌خطی چندگانه به‌عنوان وجود همبستگی خطی نزدیک میان ستون‌های ماتریس طرح $\boldsymbol{X}$ در مدل خطی $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ تعریف می‌شود که $\boldsymbol{y}$ یک بردار $n \times ۱$ پاسخ مشاهدات است، $\boldsymbol{X}$ ماتریس مشاهدات از متغیرهای مستقل با ابعاد $n \times p$ است و فرض می‌شود که $\boldsymbol{X}$ رتبه کامل p است، همچنین $\boldsymbol{\beta}$ یک پارامتر مجهول است و $\boldsymbol{\epsilon}$ یک بردار خطاست. وجود هم خطی چندگانه ممکن است فاصله اطمینان عریضی را برای پارامترهای تکی یا ترکیب خطی از پارامترها به‌وجود آورد و یا ممکن است برآوردهایی با علامت‌های اشتباه تولید کند وغیره. بهترین راه بررسی وجود هم‌خطی چندگانه جستجو در مقادیر ویژه و بردارهای ویژه $\boldsymbol{X}'\boldsymbol{X}$ است. مفهوم رگرسیون ریج در سال ۱۹۷۰ توسط هورل وکنارد برای مقابله با هم‌خطی در مسائل رگرسیون پیشنهاد شده است. از آن زمان بسیاری از نویسندگان و محققان این روش را برای توسعه برآوردگر اصلی مطالعه نمودند.

نیکویی برازش مدل، معیاری برای قضاوت در مورد عملکرد برآوردگرهاست که اغلب به طور کلی نادیده گرفته می‌شود. از جمله مزیت‌های نیکویی برازش به تابع زیان این است که اگر برآوردگر نااریب استفاده شده باشد نیکویی برازش مینیمم می‌شود. به عبارت دیگر، نیکویی برازش یک جریمه روی کم برآورد تحمیل می‌کند. اما زلنر (۱۹۹۴) با در نظر گرفتن ترکیب تؤام نیکویی برازش و دقت برآورد در تابع زیان، تابع زیان متعادل را معرفی کرد. اُهتانی و همکاران (۱۹۹۷) ، شالاب (۲۰۰۱) و تتنبورگ (۲۰۰۵) به مقایسه عملکرد برآوردگرها با استفاده از تابع مخاطره پرداخته‌اند. بیشتر محققان عملکرد برآوردگرهای ریج را براساس تمرکز برآورد اطراف مقدار واقعی پارامتر مورد قضاوت قرار داده‌اند. ما در این مقاله، عملکرد برآوردگرها را با استفاده از تابع مخاطره تحت تابع زیان متعادل بررسی می‌کنیم . در ادامه این مقاله، در بخش ۲، مدل ورویکرد تفاضلی بیان می‌شود. بخش ۳، شامل تعریف برآوردگر ریج تفاضلی تعمیم یافته است و تابع مخاطره آن در بخش ۴ مطرح می‌شود. دربخش ۵، برآوردگرها با یکدیگر مقایسه شده‌اند و سرانجام، برخی تفسیر نتایج بیان شده است.

۲ معرفی مدل

مشاهدات مستقل $(y_۱, \boldsymbol{x}_۱, t_۱)$ ... $(y_n, \boldsymbol{x}_n, t_n)$ که در آن t و y اسکالر و $\boldsymbol{x}$ یک بردار p بعدی است، را در نظر بگیرید. با استفاده از شکل بردار و ماتریس، مدل‌های خطی جزئی به شکل کلی زیر تعریف می‌شوند:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{f}(t) + \boldsymbol{\epsilon} \tag{۱.۲}$$

که در آن $\boldsymbol{y} = (y_۱, y_۲, ..., y_n)'$ بردار $n \times ۱$ از متغیرهای وابسته، $\boldsymbol{X} = (x_۱, x_۲, ..., x_\mathbf{n})'$ یک ماتریس $n \times p$ با p متغیر مستقل و n تکرار برای هریک، $\boldsymbol{\beta} = (\beta_۱, \beta_۲, ...\beta_p)'$ بردار پارامترهای مجهول است، $\boldsymbol{f}(\boldsymbol{t}) = (f(t_۱), ..., f(t_n))'$ بردار $n \times ۱$

اثر غیر خطی و $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)'$ بردار $n \times 1$ از مؤلفه‌های خطای تصادفی می‌باشد. در حالت کلی فرض می‌کنیم $\boldsymbol{\epsilon}$ بردار خطا با $E(\boldsymbol{\epsilon}) = \mathbf{0}$ و $E(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}) = \boldsymbol{\sigma}^2\boldsymbol{V}$ باشد. از آن‌جا که مدل‌های خطی جزئی یک مؤلفه پارامتری و یک مؤلفه ناپارامتری دارند از مدل‌های رگرسیون خطی منعطف‌تر هستند. در واقع استفاده از این مدل زمانی که متغیر وابسته $\boldsymbol{y}$ بطور خطی با متغیر مستقل $\boldsymbol{x}$ و بطور غیر خطی با متغیر مستقل t رابطه داشته باشد. مفید است. ونگ و همکاران (۲۰۰۷) ثابت کردند که با استفاده از روش تفاضلی با مرتبه زیاد، برآوردگر قسمت خطی بطور مجانبی کارا شده و برآوردگر غیر خطی دارای نرخ همگرایی مینیماکس خواهد بود.

فرض کنید m مرتبه تفاضلی کردن و $d_0, ..., d_m$ وزن‌های تفاضلی با شرایط زیر باشند:

$$\sum_{j=0}^{m} d_j = 0 \,, \qquad \sum_{j=0}^{m} d_j^2 = 1 \tag{۲.۲}$$

اکنون با استفاده از این وزن‌ها ماتریس تفاضلی مرتبه m را به‌صورت زیر تعریف می‌کنیم:

$$\boldsymbol{D} = \begin{pmatrix} d_0 & d_1 & \dots & d_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & d_0 & d_1 & \dots & d_m & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & d_0 & d_1 & \dots & d_m \end{pmatrix}$$

استفاده از ماتریس تفاضلی به مدل ۱.۲ اجازه برآورد مستقیم اثر پارامتری را می‌دهد.

با ضرب ماتریس تفاضلی D در طرفین مدل رگرسیونی ۱.۲ خواهیم داشت:

$$\boldsymbol{Dy} = \boldsymbol{DX\beta} + \boldsymbol{Df}(\boldsymbol{t}) + \boldsymbol{D\epsilon} \tag{۳.۲}$$

اکنون می‌توان به نقش شرط ۲.۲ پی برد، نخستین محدودیت در این شرط تضمین می‌کند با نزدیک بودن t_i ها به هم، اثر تابع ناپارامتری پس از تفاضلی کردن از بین می‌رود. بنابراین می‌توانیم معادله ۳.۲ را به‌صورت

$$\boldsymbol{Dy} = \boldsymbol{DX\beta} + \boldsymbol{D\epsilon}$$

یا

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}} \tag{۴.۲}$$

بازنویسی کنیم که در آن $\tilde{\boldsymbol{y}} = \boldsymbol{Dy}$ ، $\tilde{\boldsymbol{X}} = \boldsymbol{DX}$ و $\tilde{\boldsymbol{\epsilon}} = \boldsymbol{D\epsilon}$ هستند. با توجه به تعریف $\boldsymbol{D}$ ، $\tilde{\boldsymbol{\epsilon}}$ یک بردار $(n-m)$ تایی با $E(\tilde{\boldsymbol{\epsilon}}) = \mathbf{0}$ و $E(\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}') = \boldsymbol{\sigma}^2\boldsymbol{V_D}$ می‌باشد، بطوریکه $\boldsymbol{V_D} = \boldsymbol{DVD}' \neq \boldsymbol{I}_{n-m}$. برای برآورد پارامتر $\boldsymbol{\beta}$ با استفاده از ضرایب تفاضلی مرتبه دلخواه در معادله ۲.۲ یاتچیو (۱۹۹۷) برآوردگر تفاضلی $\boldsymbol{\beta}$ را برای مدل نیمه پارامتری زمانی که $\boldsymbol{V} = \boldsymbol{I}_n$ بصورت زیر تعریف کرد:

$$\widehat{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{y}} \tag{۵.۲}$$

بنابراین با روش تفاضلی می‌توان $\boldsymbol{\beta}$ را بدون در نظر گرفتن مؤلفه ناپارامتری $\boldsymbol{f}(\boldsymbol{t})$ در مدل ۱.۲ برآورد کرد.

برای ارزیابی پارامتر $\boldsymbol{\beta}$ در معادله ۱.۲ ما برآوردگر اصلاح شده $\boldsymbol{\sigma}^2$ را به صورت زیر معرفی می‌کنیم:

$$\widehat{\boldsymbol{\sigma}}^2 = \frac{\tilde{\boldsymbol{y}}'\boldsymbol{V_D}^{-\frac{1}{2}}(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{V_D}^{-\frac{1}{2}}\tilde{\boldsymbol{y}}}{tr(\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{D})} \tag{۶.۲}$$

که ماتریس افکنش P به صورت زیر تعریف می‌شود:

$$\boldsymbol{P} = \boldsymbol{V}_{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\boldsymbol{V}_{\boldsymbol{D}}^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{V}_{\boldsymbol{D}}^{-\frac{1}{2}} \quad (۷.۲)$$

۳ برآوردگر تفاضلی ریج محدود شده تعمیم یافته

با توجه به مدل ۴.۲ برآوردگر شناخته شده و نااریب پارامتر $\boldsymbol{\beta}$ که برآوردگر تفاضلی تعمیم یافته نامیده می‌شود، به صورت زیر است:

$$\widehat{\boldsymbol{\beta}}_{GD} = \boldsymbol{C}_{\boldsymbol{D}}^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{V}_{\boldsymbol{D}}^{-1}\tilde{\boldsymbol{y}} \quad (۱.۳)$$

که در آن $\boldsymbol{C}_{\boldsymbol{D}} = \tilde{\boldsymbol{X}}'\boldsymbol{V}_{\boldsymbol{D}}^{-1}\tilde{\boldsymbol{X}}$ می‌باشد. از معادله ۱.۳ مشاهده می‌شود که برآوردگر تفاضلی تعمیم یافته $\boldsymbol{\beta}$ وابستگی شدیدی به ماتریس مشاهدات $\boldsymbol{C}_{\boldsymbol{D}}$ دارد. اگر ماتریس $\boldsymbol{C}_{\boldsymbol{D}}$ دچار بد شرطی شده باشد (وابستگی نزدیک میان ستون‌های ماتریس $\boldsymbol{C}_{\boldsymbol{D}}$)، $\widehat{\boldsymbol{\beta}}_{GD}$ بی دلیل دارای واریانس بزرگ خواهد شد. علاوه براین، ممکن است برخی از ضرایب رگرسیونی با علامت اشتباه برآورد شوند و استنباط آماری برای محقق سخت شود. به عنوان یک چاره، بنابر هورل و کنارد (۱۹۷۰) می‌توان از برآوردگر تفاضلی ریج تعمیم یافته زیر استفاده کرد.

$$\widehat{\boldsymbol{\beta}}_{GD}(k) = \boldsymbol{T}_k\widehat{\boldsymbol{\beta}}_{GD}, \quad \boldsymbol{T}_k = (k\boldsymbol{C}_{\boldsymbol{D}}^{-1} + \boldsymbol{I}_p)^{-1} \quad (۲.۳)$$

که $k \geq ۰$ در آن پارامتر ریج است.

اکنون محدودیت غیر تصادفی و خطی زیر را در نظر می‌گیریم:

$$\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r} \quad (۳.۳)$$

بطوریکه $\boldsymbol{R}$ ماتریس $(q \times p)$ با رتبه $q < p$ و $\boldsymbol{r}$ یک بردار $۱ \times q$ تایی معلوم می‌باشند. فرض رتبه کامل بودن $\boldsymbol{R}$ برای راحتی در محاسبات و باتوجه به این واقعیت که هر معادله خطی سازگار می‌تواند به یک معادله خطی با ماتریس ضرایب دارای رتبه کامل سطری تبدیل شود در نظر گرفته شده است. با در نظر گرفتن فرضیه خطی ۳.۳ ، برآوردگر تفاضلی محدود شده تعمیم یافته برابر است با:

$$\widehat{\boldsymbol{\beta}}_{GRD} = \widehat{\boldsymbol{\beta}}_{GD} - \boldsymbol{C}_{\boldsymbol{D}}^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}_{\boldsymbol{D}}^{-1}\boldsymbol{R}')^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_{GD} - \boldsymbol{r}). \quad (۴.۳)$$

بنابر سوامی (۱۹۷۸) و سوامی و مهتا (۱۹۷۷) می‌توان برآوردگر تفاضلی ریج تعمیم یافته جهت بهبود فرم غیر ریج را با مینیمم کردن مجموع توان‌های دوم مانده‌ها با در نظر گرفتن محدودیت کروی $\boldsymbol{\beta}'\boldsymbol{\beta} \leq \rho^۲$ ،که این محدودیت یک محدودیت اجباری برای غلبه بر مشکل هم‌خطی است، بدست آورد. بنابر ژانگ و یانگ (۲۰۰۷) برای بدست آوردن برآوردگر مورد نظر، باید مساله بهینه سازی زیر را حل نماییم:

$$\begin{aligned} min_{\boldsymbol{\beta}} \quad & [(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta})'\boldsymbol{V}_{\boldsymbol{D}}^{-1}(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta})], \\ s.t. \quad & \boldsymbol{\beta}'\boldsymbol{\beta} \leq \rho^۲ \\ & \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r} \end{aligned} \quad (۵.۳)$$

پس از حل مسئله فوق به روش لاگرانژ، برآوردگر ریج تفاضلی محدود شده تعمیم یافته را بصورت

$$\widehat{\boldsymbol{\beta}}_{GRD}(k) = \widehat{\boldsymbol{\beta}}_{GD}(k) - C_{\boldsymbol{D}}^{-1}(k)\boldsymbol{R}'[(\boldsymbol{R}\boldsymbol{C}_{\boldsymbol{D}}^{-1}(k)\boldsymbol{R}')]^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}}_{GD}(k) - \boldsymbol{r}). \quad (۶.۳)$$

بدست می‌آوریم که $C_D(k) = \tilde{X}'V_D^{-1}\tilde{X} + kI_p$ و $\widehat{\beta}_{GD}(k) = C_D^{-1}(k)\tilde{X}'V_D^{-1}\tilde{Y}$ برای $k \geq 0$ می‌باشند.به سادگی می‌توان نشان داد $\widehat{\beta}_{GRD}(k)$ و $\widehat{\beta}_{GRD}$ با قید $R\beta = r$ مطابقت دارند. همچنین واضح است که برای $k = 0$ ،رابطه $\widehat{\beta}_{GRD}(0) = \widehat{\beta}_{GRD}$ برقرار است.

۴ ارزیابی توابع مخاطره

در این بخش، ما تابع مخاطره را برای برآوردگرهای پیشنهاد شده در بخش قبل محاسبه می‌کنیم. برای محاسبه توابع مخاطره ابتدا باید تابع زیان مورد استفاده را معرفی نماییم. فرض کنید β^* یک برآوردگر دلخواه برای β در مدل ۵.۲ باشد. زلنر (۱۹۹۴) با نظر گرفتن ترکیب توام نیکویی برازش و دقت برآورد در تابع زیان، هردو معیار را برای قضاوت در مورد نحوه عملکرد برآوردگر مورد نظر لحاظ نمود و تابع زیان متعادل (BLF) را که یک ترکیب محدب از دو معیار فوق می‌باشد را به‌صورت زیر معرفی کرد:

$$L(\beta^*, \beta) = \omega(\tilde{X}\beta^* - \tilde{y})'(\tilde{X}\beta^* - \tilde{y}) + (1-\omega)(\tilde{X}\beta^* - \tilde{X}\beta)'(\tilde{X}\beta^* - \tilde{X}\beta) \qquad (۱.۴)$$

که در آن $0 \leq \omega \leq 1$ یک وزن غیر تصادفی است.

حالا،فرض می‌کنیم $S(p)$ فضای ماتریس‌های معین مثبت باشد. در این مقاله، تابع زیان متعادل وزنی را به شکل زیر تعریف می‌کنیم:

$$L(\beta^*\beta) = \omega(\tilde{X}\beta^* - \tilde{y})'V_D^{-1}(\tilde{X}\beta^* - \tilde{y}) + (1-\omega)(\tilde{X}\beta^* - \tilde{X}\beta)'V_D^{-1}(\tilde{X}\beta^* - \tilde{X}\beta) \qquad (۲.۴)$$

که در آن $V_D^{-1}S(p)$ یک ماتریس وزن است. تابع مخاطره مربوط به BLF معادله ۲.۴ بصورت زیر می‌باشد:

$$R(\beta^*, \beta) = E[L(\beta^*, \beta)] \qquad (۳.۴)$$

قبل از محاسبه تابع مخاطره $\widehat{\beta}_{GRD}(k)$ ما فرمول جدیدی برای $\widehat{\beta}_{GRD}(k)$ بدست می‌آوریم که نقش بسزایی در محاسبه تابع مخاطره دارد.

$$\widehat{\beta}_{GRD}(k) = M_D(k)\tilde{X}'V_D^{-1}\tilde{Y} - M_D(k)C_D(k)\beta_0 + \beta_0 \qquad (۴.۴)$$

که $\beta_0 = R'(RR')^{-1}r$ و $M_D(k) = C_D^{-1}(k) - C_D^{-1}(k)R'(RC_D^{-1}(k)R')^{-1}RC_D^{-1}(k)$ می‌باشند.اکنون با استفاده از روابط بدست آمده و این واقعیت که $R\beta = r$ و $R\beta_0 = r$ دلالت بر تساوی $M_D(k)C_D(k)(\beta - \beta_0) = (\beta - \beta_0)$ دارد، براحتی می‌توان اریبی و ماتریس کواریانس $\widehat{\beta}_{GRD}(k)$ را محاسبه کرد.

$$E(\widehat{\beta}_{GRD}(k) - \beta) = -kM_D(k)\beta \qquad (۵.۴)$$

و

$$Cov(\widehat{\beta}_{GRD}(k)) = \sigma^2 M_D(k)C_D M_D(k) \qquad (۶.۴)$$

با جایگذاری $\tilde{Y} \doteq \tilde{X}\beta + \tilde{\epsilon}$ در معادله ۳.۴ ، تابع مخاطره برآوردگر $\widehat{\beta}_{GRD}(k)$ بصورت زیر محاسبه خواهد شد:

$$R(\widehat{\beta}_{GRD}(k), \beta) = \sigma^2(1-2\omega)(p-q) + k\sigma^2(2\omega-1)tr(M_D(k))$$

$$-k\sigma^{۲}tr(\boldsymbol{M}_{\boldsymbol{D}}^{۲}(k)\boldsymbol{C}_{\boldsymbol{D}}) + k^{۲}\boldsymbol{\beta}'(\boldsymbol{M}_{\boldsymbol{D}}(k) - k\boldsymbol{M}_{\boldsymbol{D}}^{۲}(k))\boldsymbol{\beta} + \omega\sigma^{۲}(n-m). \quad (۷.۴)$$

همچنین با مساوی صفر قرار دادن k در عبارت فوق مخاطره $\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}$ بصورت زیر بدست می‌آید:

$$R(\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}, \boldsymbol{\beta}) = \sigma^{۲}(۱ - ۲\omega)(p-q) + +\omega\sigma^{۲}(n-m). \quad (۸.۴)$$

۵ مقایسه نتایج

در این قسمت شرایط لازم و کافی برای اینکه برآوردگر $\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}(k)$ بر $\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}$ برتری داشته باشد را فراهم می‌کنیم.به این معنا که $R(\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}(k), \boldsymbol{\beta}) \leq R(\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}, \boldsymbol{\beta})$ اکنون با استفاده از روابط ۷.۴ و ۸.۴ می‌توان گفت $\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}(k)$ بر $\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}$ تحت تابع زیان متعادل برتری دارد هرگاه به ازای هر $\boldsymbol{\beta} \in R^p$ داشته باشیم:

$$\begin{aligned}\Delta &= \boldsymbol{R}(\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}, \boldsymbol{\beta}) - \boldsymbol{R}(\widehat{\boldsymbol{\beta}}_{GR\boldsymbol{D}}(k), \boldsymbol{\beta}) \\ &= k\sigma^{۲}tr(\boldsymbol{M}_{\boldsymbol{D}}^{۲}\boldsymbol{C}_{\boldsymbol{D}}) + k\sigma^{۲}(۱-۲\omega)tr(M_D(k)) + k^{۲}\boldsymbol{\beta}'(k\boldsymbol{M}_{\boldsymbol{D}}^{۲} - \boldsymbol{M}_{\boldsymbol{D}}(k))\boldsymbol{\beta} \geq \boldsymbol{0} \quad (۱.۵)\end{aligned}$$

همان‌طور که دیده می‌شود Δ دارای یک فرم بسته از k نیست. لذا مقایسه دو برآوردگر را با استفاده از شبیه‌سازی و رسم نمودارهایی به شکل عددی ادامه می‌دهیم.

۱.۵ شبیه سازی نمونه

در این زیر بخش، کارایی برآوردگرهای پیشنهادی را مقایسه می کنیم. این آزمایش ها برای ترکیب مقادیر مختلف $k = \{۰, ۱, ۲, ..., ۸\}$ و $w = \{۰, ۰/۱, ۰/۳, ۰/۵, ۰/۷, ۰/۹\}$ انجام می شود. در این مطالعه، متغیر پاسخ را از مدل زیر به حجم $n = ۱۰۰$ با $۱۰^{۳}$ تکرار تولید می کنیم:

$$Y_i = x_{۱i}\beta_۱ + x_{۲i}\beta_۲ + x_{۳i}\beta_۳ + x_{۴i}\beta_۴ + x_{۵i}\beta_۵ + f(t_i) + \epsilon_i, i = ۱, ..., n$$

که در آن

$$\boldsymbol{\beta} = (۱/۵, ۲, ۳, -۵, ۴)'$$

$$\boldsymbol{x}_i \sim N_۵(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{x}}), \boldsymbol{\mu} = \begin{pmatrix} -۲ \\ ۱ \\ ۲ \\ ۳ \\ ۲ \end{pmatrix}, \boldsymbol{\Sigma}_{\boldsymbol{x}} = \begin{pmatrix} ۰/۸۱ & ۰/۴ & ۰/۳ & -۰/۲ & -۰/۱ \\ ۰/۴ & ۲/۲۵ & ۰/۴ & ۰/۳ & -۰/۲ \\ ۰/۳ & ۰/۴ & ۱ & ۰/۴ & ۰/۳ \\ -۰/۲ & ۰/۳ & ۰/۴ & ۰/۶۴ & ۰/۴ \\ -۰/۱ & -۰/۲ & ۰/۳ & ۰/۴ & ۰/۴۹ \end{pmatrix}$$

$$f(t_i) = \sqrt{t_i(۱-t_i)}\sin\left(\frac{۲/۱\pi}{t_i + ۰/۰۵}\right), t_i = (i - ۰/۵)/n$$

$$\boldsymbol{\epsilon} \sim N(۰, \sigma_\epsilon^{۲}\boldsymbol{V}), \sigma_\epsilon^{۲} = ۰/۰۱, v_{ij} = (\frac{۱}{n})^{|i-j|+۱}$$

در این شبیه سازی ابتدا پارامتر خطی را با حذف تابع ناپارامتری به روش تفاضلی مرتبه چهارم برآورد کرده و سپس با جایگزینی برآورد بدست آمده برای آن، تابع ناپارامتری $f(.)$ را با استفاده از تکنیک هموارسازی کرنل و بکار بردن کرنل نرمال، برآورد می کنیم. ضرایب تفاضلی بهینه از مرتبه چهار عبارتند از $d_0 = ۰/۸۸۷۳$، $d_1 = -۰/۳۰۹۹$، $d_2 = -۰/۲۴۶۴$، $d_3 = -۰/۱۹۰۱$ و $d_4 = -۰/۱۴۰۹$. همچنین، ماتریس محدودیت خطی $\boldsymbol{R}$ و بردار $\boldsymbol{r}$ را بصورت زیر در نظر می گیریم:

$$\boldsymbol{R} = \begin{pmatrix} ۱ & -۲ & ۱ & ۴ & ۵ \\ -۱ & ۲ & -۱ & -۳ & ۰ \\ ۱ & ۲ & -۱ & -۲ & ۳ \\ ۲ & -۱ & ۳ & -۲ & ۰ \end{pmatrix} \Rightarrow \boldsymbol{r} = \begin{pmatrix} ۰/۵ \\ ۱۴ \\ ۲۴ \\ ۲۰ \end{pmatrix}.$$

بحث و نتیجه‌گیری

در این مقاله،دو برآوردگر در مدل جزئی خطی هنگامی‌که خطاها همبسته بوده و برخطی محدودیت‌های خطی روی فضای پارامتر خطی وجود داشته باشد،پیشنهاد شد. سپس با حضور هم‌خطی میان ستون‌های ماتریس طرح، برآوردگر جدید تفاضلی ریج محدودشده تعمیم یافته یعنی $\widehat{\boldsymbol{\beta}}_{\boldsymbol{GRD}}(k)$ در مقابل فرم غیر ریج آن را در مدل جزئی خطی معرفی نمودیم. در ادامه برای مقایسه دو برآوردگر جدید معرفی شده، توابع مخاطره آن‌ها را تحت تابع زیان متعادل وزنی محاسبه کرده وبه دلیل اینکه تفاضل توابع مخاطره دارای صورت بسته‌ای از k نبود،مقایسه را با استراتژی شبیه‌سازی برای مقادیر مختلف پارامتر ریج و ضریب وزنی تابع زیان متعادل وزنی دنبال نمودیم. نتایج بدست آمده از شبیه سازی گویای این مطلبند که برای همه ترکیبات k و ω زمانی که $k < a$ ، $\widehat{\boldsymbol{\beta}}_{\boldsymbol{GRD}}(k)$ از $\widehat{\boldsymbol{\beta}}_{\boldsymbol{GRD}}$ بهتر است،که در آن a یک تابع نزولی از ω است. به‌عنوان مثال،به ازای $\omega = ۰/۱$ ، $\widehat{\boldsymbol{\beta}}_{\boldsymbol{GRD}}(k)$ از $\widehat{\boldsymbol{\beta}}_{\boldsymbol{GRD}}$ بهتر است، هرگاه $a = ۸/۰۳۳$ باشد ولی برای $\omega = ۰/۹$ اگر $a = ۰/۷۸۲$ باشد آنگاه $\widehat{\boldsymbol{\beta}}_{\boldsymbol{GRD}}(k)$ از $\widehat{\boldsymbol{\beta}}_{\boldsymbol{GRD}}$ بهتر است.

مراجع

Hoerl A.E. and Kennard R.W. (1970), *Ridge regression biased estimation for nonorthogonal problems,* Technometrics 12 , pp. 69–82.

Ohtani K. , Giles D.E.A. and Giles J.A. (1997), *The exact risk performance of a pre-test estimator in a heteroskedastic linear regression model under the balanced loss function,* Econ. Rev. 16 (1997), pp. 119–133.

Shalabh (2001), *Least squares estimators in measurement error models under the balanced loss function*, Test 10, pp. 301–308.

Swamy P.A.V.B., Mehta J.S. and Rappoport P.N. (1978), *Two methods of evaluating Hoerl and Kennard's ridge regression,* Commun. Statist. 12, pp. 1133–1155.

Swamy P.A.V.B. and Mehta J.S. (1977), *A note on minimum average risk estimator for coefficient in linear models,* Commun. Statist. 6, pp. 1161–1186. Adv. Appl. Probab., pp. 476–495.

Tothenberg H. and Shalabh (2005), *Estimation of regression coefficients subject to exact linear restrictions when some observations are missing and quadratic error balanced loss function is used,* Test 14, pp. 385–396. 129 (2002), pp. 455–467.

Wang L., Brown L. D., and Cai T. T. (2007), *A difference based approach to semiparametric partial linear model,* Tech. Rep., Department of Statistics, The Wharton School, University of Pennsylvania.

Yatchew A. (1997), *An elementary estimator of the partial linear model,* Econ. Lett. 57, pp. 135–143.

Zellner A. (1986), *Bayesian estimation and prediction using asymmetric loss functions,* J. Am. Stat. Assoc. 81, pp. 446–451.

ZellnerA. (1994), Bayesian and non-Bayesian estimation using balanced loss function, in Statistical Decision Theory and Related Topics, S.S. Gupta and J.O. Berger, eds., Springer Verlag, Berlin, pp. 377–390.

Zhong Z. and Yang H. (2007), *Ridge estimation to the restricted linear model,* Commun. Stat. Theory Methods 36, pp. 2099–2115.