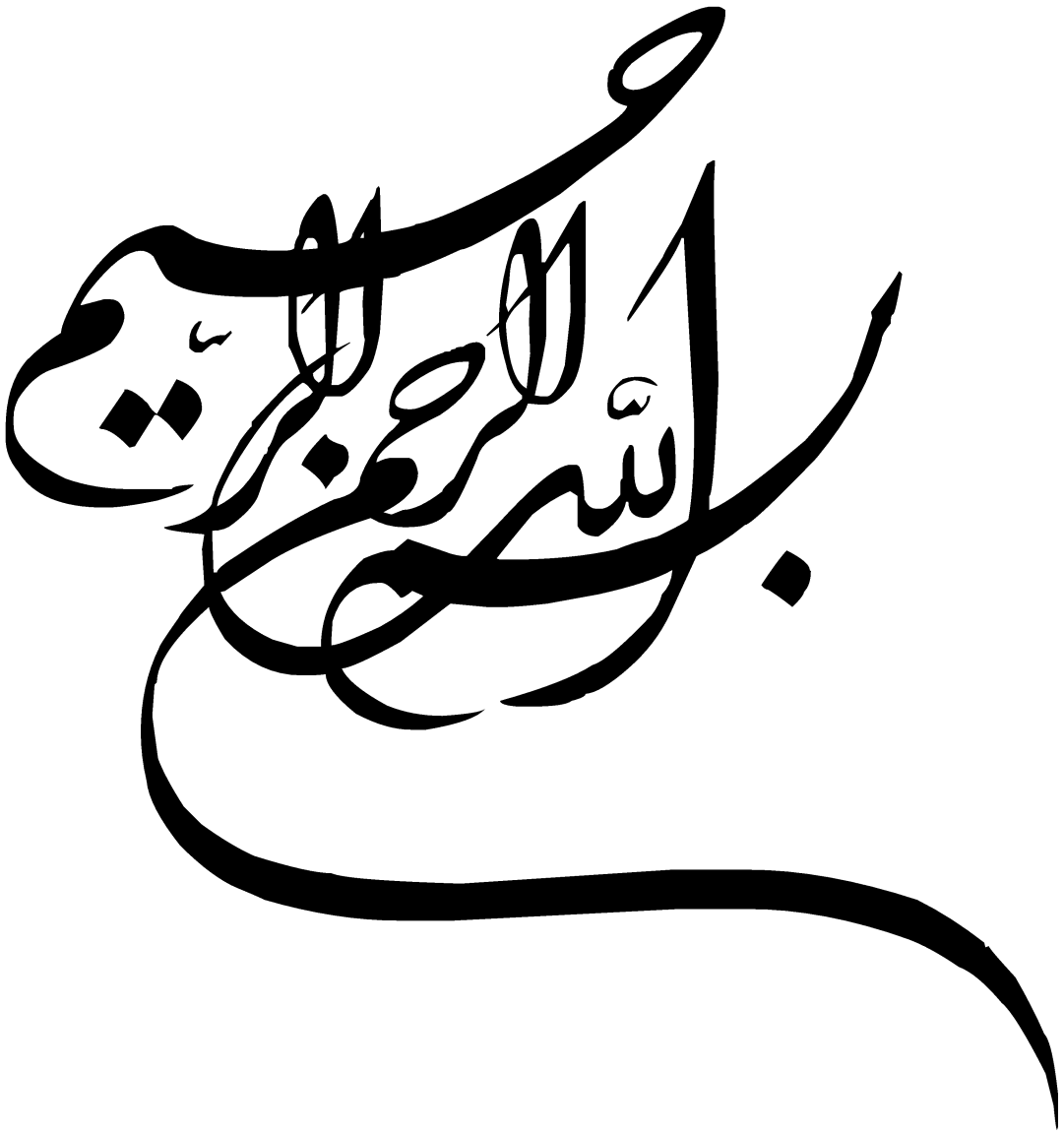
بسم الله الرحمن الرحيم

دانشکده علوم

پایان نامه کارشناسی ارشد در رشته آماررياضی

بررسی کارایی برآوردگرهای تقریبا ناااریب در مدل رگرسیون خطی، در حضور هم‌خطی چندگانه

توسط:

مرضیه توانگر

استاد راهنما:

دکتر مینا توحیدی

دکترعبدالرسول برهانی حقیقی

مهر ماه۱۳۹۱

بـــه نـــام خـــدا

اظـــهـــارنـــامـــه

اینجانب مرضیه توانگر(۸۹۰۸۹۲) دانشجوی کارشناسی ارشد رشته‌ی آمار ریاضی اظهار می‌کنم که این پایان نامه حاصل پژوهش خودم بوده و در جاهایی که از منابع دیگران استفاده کرده‌ام، نشانی دقیق و مشخصات کامل آن را نوشته‌ام. همچنین اظهار می‌کنم که تحقیق و موضوع پایان نامه‌ام تکراری نیست و تعهد می‌نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده و یا در اختیار غیر قرار ندهم. کلیه حقوق این اثر مطابق با آئین نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است.

نام و نام خانوادگی : مرضیه توانگر

تاریخ و امضا : ۱۳۹۱/۰۷/۱۲

به نام خدا

بررسی کارایی برآوردگرهای تقریبا نااریب در مدل رگرسیون خطی، در حضور هم‌خطی چندگانه

به وسیله :

مرضیه توانگر

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه به عنوان بخشی از فعالیت های لازم برای اخذ درجه کارشناسی ارشد

در رشته :

آمار ریاضی

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی کمیته ی پایان نامه ،با درجه ی : عالی

دکتر مینا توحیدی، دانشیار بخش آمار(رئیس کمیته).....

دکترعبدالرسول برهانی حقیقی، استادیار بخش آمار(رئیس کمیته).....

دکتر سلطان محمد صدوقی الوندی، استاد بخش آمار.....

دکتر عبدالرضا بازرگان لاری، استادیار بخش آمار.....

مهر ماه ۱۳۹۱

تقدیم به

او که مرا آفرید

و آنها که بودن تا گشتیم پدید

و ایشان که دادند از علم بر من نوید

و مایی که داریم عشق و امید

و شما اکنون .....

تقدیم به پدر و مادر عزیزم که

وجودشان برایم همه عشق

و وجودم برایشان همه رنج

توانشان رفت تا به توانایی رسم

و مویشان گرد سپیدی گرفت تا رویم سپید ماند

سپاسگزاری

با سپاس بیکران به درگاه یگانه معبود هستی که این توفیق را نصیب من کرد تا در مرحله ای دیگر، گامی هرچند کوچک در تحقق اهداف خویش بردارم. اکنون که به یاری پروردگار، مراحل پژوهش و تدوین این پایان نامه به اتمام رسیده است، بر خود لازم می دانم که از تلاش و کمک تمام کسانی که مرا در این امر یاری نموده اند، صمیمانه سپاسگزاری نمایم. از مادر و پدر عزیزم به پاس محبت بی دریغشان که همواره مشوق من بوده اند و در سختی ها و دشواری های زندگی، صمیمانه و با خلوص قلب مرا یاری داده اند، خاضعانه تشکر می کنم .

مراتب سپاس و قدردانی بیکران خود را به محضر استاد بزرگوارم سرکار خانم دکتر مینا توحیدی که این پژوهش مرهون رهنمودهای ایشان می باشد و همواره با سعه ی صدر و ظرافت طبع در دوران تحصیل مرا یاری داده اند، تقدیم می دارم. همچنین از جناب آقای دکتر دکترعبدالرسول برهانی حقیقی که در پایان نامه مرا یاری کردند نهایت سپاس را دارم. از اساتید محترم جناب آقای دکتر محمد صدوقی و جناب آقای دکتر عبدالرضا بازرگان لاری که زحمت مشاوره ی این تحقیق را تقبل نمودند و مرا مدیون محبت ها و راهنمایی های ارزنده ی خود کردند، سپاسگزارم. امید آن دارم که با یاری خداوند متعال در جهت پیشرفت ایران عزیز قدم بردارم.

چکیده

بررسی کارایی برآوردگرهای تقریبا ناریب در مدل رگرسیون خطی، در حضور هم‌خطی چندگانه

به وسیله ی :

مرضیه توانگر

برآوردگر حداقل مربعات معمولی $\hat{\beta} = (X'X)^{-1}X'Y$ (OLS) اغلب برای برآورد ضرایب رگرسیونی در مدل رگرسیون خطی $Y = X\beta + \varepsilon$ استفاده می‌شود. اما این برآوردگر به شدت به خصوصیات ماتریس $X'X$ بستگی دارد. هم خطی چندگانه بین متغیرهای توضیحی در مدل رگرسیون خطی، یک مسأله مهم در بکارگیری این مدل می‌باشد. در این حالت برآوردکننده کمترین مربعات $(\hat{\beta})$ دارای واریانس بزرگی است. در این پایان‌نامه، ابتدا با الهام از کلاس برآوردکننده‌های ریج، $\hat{\beta}_R = (\mathrm{X'X + kI})^{-1}\mathrm{X'Y}$، $\mathrm{k} > 0$ ، کلاس جدیدی از برآوردگرهای اریب، $\hat{\beta}_d = (\mathrm{X'X + I})^{-1}(\mathrm{X'Y + d\hat{\beta}})$، $0 < d < 1$، معرفی می‌شود. هر عضو این کلاس جدید یک برآوردکننده لیو می‌باشد. برآوردگر بهینه، با استفاده از محک مینیمم توان دوم خطا، MSE ، انتخاب می‌شود. با ادغام اعضای دو کلاس از برآوردکننده مذکور، کلاسی دیگر از برآوردکننده‌ها به نام کلاس برآوردکننده‌های دوپارامتری (TP) ساخته می‌شود. با استفاده از فرآیند جک-نایف، برآوردگر تقریبا ناریب لیو $(AULE)$ و برآوردگر تقریبا ناریب دوپارامتری $(AUTP)$ به دست آورده می‌شوند. برآوردگرهای ساخته شده در این پایان نامه، با برآوردگر LS با معیار MSE مقایسه می‌گردند و در آخر، یک مطالعه شبیه‌سازی، برای مشاهده کارایی این برآوردگرهای ساخته شده انجام گرفته است.

فهرست مطالب

عنوان **صفحه**

فصل چهارم : کارایی برآوردگر تقریبا نااریب دوپارامتری

فهرست جداول

عنوان و شماره **صفحه**

فهرست شکل‌ها

عنوان **صفحه**

فصل اول

مقدمه

۱-۱- مقدمه

مدل رگرسیون خطی زیر را در نظر بگیرید:

$$Y = X\beta + \varepsilon\,, \qquad \varepsilon \sim (0, \sigma^2 I_n) \qquad (1.1.1)$$

که Y بردار پاسخ n×1، X ماتریس n×p از متغیرهای توضیحی، β بردار p×1 از ضرایب رگرسیونی نامعلوم و ε بردار n×1 از خطاهای تصادفی مستقل و هم توزیع هستند که دارای میانگین صفر و واریانس $\sigma^2 I_n$ می‌باشند.

حال علاقمند هستیم تا پارامتر β را جهت استنباط‌های پیش‌بینی کننده برآورد کنیم.

با توجه به روش حداقل‌مربعات بردار پارامتر β، به صورت زیر برآورد می‌شود:

$$S = (Y - X\beta)'(Y - X\beta) = \varepsilon'\varepsilon$$
$$S = Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$$
$$\frac{\partial S}{\partial \beta} = -2X'Y + 2X'X\beta = 0$$
$$X'X\beta = X'Y$$
$$\hat{\beta} = (X'X)^{-1}X'Y$$

برآوردگر $\hat{\beta}$ را، برآوردگرحداقل‌مربعات معمولی (OLS)[1] می‌نامند، که برآوردگر نااریب می‌باشد و با توجه به قضیه گوس مارکوف[2] در بین برآوردگرهای خطی نااریب، دارای کوچکترین واریانس می‌باشد و به این ویژگی $BLUE$[3]، به‌معنای بهترین برآوردگر نااریب خطی، گویند.

[1] Ordinary least square
[2] Gauss Markov Theorem
[3] Best Linear Unbiased Estimator

۱-۲- برآوردگر انقباضی[1]

حال می‌خواهیم برآوردگرهای انقباضی، که برآوردگرهای اریبی می‌باشند را معرفی کنیم. با توجه به اینکه برآوردگر (OLS) در بین برآوردگرهای خطی نااریب کوچکترین واریانس را دارد، اما نمی‌توان گفت که در بین تمام برآوردگرها، دارای کوچکترین MSE می‌باشد. بنابراین اگر فرض نااریبی را نادیده بگیریم، می‌توان برآوردگرهای اریبی را یافت که در بعضی موقعیت‌ها به دلیل MSE کوچکتر عملکرد بهتری داشته باشند.

در آمار برآوردگر انقباضی، برآوردگری است که با بهبود بخشیدن برآوردگر اولیه با استفاده از سایر اطلاعات درباره پارامتر، به دست می‌آید. در واقع برآوردگر بهبود یافته نسبت به برآوردگر اولیه به مقدار واقعی نزدیک‌تر می‌باشد و دارای ریسک کمتری می‌باشد.

فرض کنید برآوردگر اولیه صفر نباشد. با ضرب برآوردگر اولیه در یک پارامتر مشخص می‌توان برآوردگر دیگری به دست آورد و سپس مقدار پارامتر مشخص را طوری تعیین کرد که MSE برآوردگر جدید مینیمم شود. به ازای مقدار به دست آمده برای پارامتر، برآوردگر جدید MSE کوچکتری نسبت به برآوردگر اولیه خواهد داشت و در نتیجه برآوردگر جدید بهبود یافته است. یک مثال قابل ذکر در این رابطه، برآورد واریانس جامعه بر اساس یک نمونه ساده می‌باشد. برای نمونه‌ای به حجم n ، واریانس نمونه به صورت زیر می‌باشد:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

و می‌دانیم که s^2، برآوردگری نااریب برای واریانس جامعه می‌باشد اما اگر به جای کسر $\frac{1}{n-1}$ در s^2، کسر $\frac{1}{n+1}$ را جایگزین کنیم، برآورد اریبی با MSE مینیمم حاصل می‌شود. مبحث انقباض در استنباط بیزی ضمنی و استنباط درستنمایی تاوانیده[2] و استنباط از نوع $James - Stein$ مطرح می‌شود. در مقابل فرآیند برآوردگر ماکسیمم درستنمایی و حداقل مربعات شامل تاثیرات انقباضی نمی‌باشند اگرچه در طرح‌های برآورد انقباضی مورد استفاده قرار می‌گیرند.

[1] Shrinkage Estimator
[2] Penalized

در استفاده از برآوردگرهای انقباضی در زمینه تحلیل رگرسیونی که ممکن است تعداد زیادی متغیرهای توضیحی وجود داشته باشند، توسط copas (۱۹۸۳) توصیف شده است. در این حالت مقدار ضرایب رگرسیونی برآورد شده به سمت صفر منقبض می‌شوند و تاثیرات آن در کاهش MSE مقادیر پیش‌بینی از مدل مشخص شده است.

یک نوع دیگر از برآوردگرهای اریب که برآوردگرهای انقباضی می‌باشند و مورد توجه بسیاری از پژوهشگران قرار گرفته است، برآوردگرریج[1] و برآوردگراشتاین[2] می‌باشندکه به ترتیب به صورت زیر می‌باشند:

$$\hat{\beta}_R = (X'X + kI)^{-1}X'Y$$
$$\hat{\beta}_s = c\hat{\beta}$$

که $0 < c < 1$ و $k > 0$ پارامتر می‌باشند و در بخش‌های بعدی به‌طور کامل معرفی می‌شوند.

۱-۳- هم‌خطی چندگانه[3]

بکارگیری و تعبیر یک مدل رگرسیون خطی چندگانه به طور ضمنی یا به طور صریح به برآوردهای ضرایب منفرد رگرسیون بستگی دارد. استنباط‌هایی که به طور معمول انجام می‌شود، شامل موارد زیر است:

۱. مشخص کردن اثرات نسبی متغیرهای رگرسیونی.
۲. پیش‌بینی و یا برآورد
۳. انتخاب یک مجموعه مناسب از متغیرها برای مدل

اگر هیچ رابطه خطی بین متغیرهای رگرسیونی نباشد، گفته می‌شود که متعامدند، اگر متغیرهای رگرسیونی بر هم عمود باشند، استنباط‌هایی که در بالا ذکر شد نسبتا ساده صورت می‌پذیرد. متاسفانه در بیشتر کاربردهای رگرسیون متغیرهای رگرسیونی متعامد نیستند. گاهی اوقات متعامد نبودن چیز جدی نیست. در حالی که در بعضی وضعیت‌ها متغیرهای رگرسیونی تقریبا به طور کامل ارتباط خطی دارند و در چنین مواردی استنباط‌های بر اساس این

[1] Ridge
[2] Stein
[3] multicollinearity

رگرسیون می‌تواند گمراه کننده یا غلط باشد. هنگامی‌که ارتباط خطی نزدیکی بین متغیرهای رگرسیونی وجود دارد، گفته می‌شود مسئله هم‌خطی چندگانه وجود دارد(به کتاب مقدمه‌ای بر تحلیل رگرسیون خطی از Montgomery, Douglas C ترجمه سید ابراهیم رضوی پاریزی مراجعه شود).

دراین بخش علل وقوع هم‌خطی چندگانه و بعضی از اثرات آن روی استنباط، بیان می‌شود و روش‌هایی برای تشخیص وجود هم‌خطی و بعضی فنون برخورد با این مسئله، مورد بحث و بررسی قرار خواهد گرفت.

۱-۳-۱- نتایج وجود هم‌خطی چندگانه

وجود هم‌خطی چندگانه بالقوه آثار جدی متعددی بر برآوردهای حداقل مربعات ضرایب رگرسیون دارد. بعضی از این اثرها را به سادگی می‌توان نمایش داد. فرض کنیم فقط دو متغیر رگرسیونی X_1 و X_2 وجود داشته باشد و با این فرض که X_1 و X_2 و Y به طول واحد مقیاس‌سازی شده‌اند. مدل را به صورت زیر در نظر می‌گیریم :

$$Y = \beta_1 X_1 + \beta_1 X_2 + \varepsilon$$

و معادلات حداقل مربعات نرمال عبارتند از :

$$(X'X)\hat{\beta} = X'Y$$

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

که در آن r_{12} همبستگی ساده بین X_1 و X_2 و r_{jy} همبستگی ساده بین X_j و Y می‌باشد و $j = 1,2$. معکوس $(X'X)$ عبارتست از :

$$C = (X'X)^{-1} = \begin{pmatrix} \frac{1}{(1-r_{12}^2)} & \frac{-r_{12}}{(1-r_{12}^2)} \\ \frac{-r_{12}}{(1-r_{12}^2)} & \frac{1}{(1-r_{12}^2)} \end{pmatrix} \qquad (1.3.1)$$

و برآوردهای ضرایب رگرسیون عبارتست از :

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12}r_{2y}}{(1-r_{12}^2)} \quad , \quad \hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{(1-r_{12}^2)} \qquad (1.3.2)$$

اگر بین X_1 و X_2 هم‌خطی چندگانه شدید موجود باشد در این‌صورت ضریب همبستگی r_{12} بزرگ خواهد بود. با توجه به اینکه اگر $|r_{12}| \to 1$ آنگاه $V(\hat{\beta}_j) = C_{jj}\sigma^2 \to \infty$ و $Cov(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\sigma^2 \to \pm\infty$ بسته به اینکه $r_{12} \to +1$ یا $r_{12} \to -1$. بنابراین هم‌خطی زیاد بین X_1 و X_2 منجر به واریانس ها و کواریانس های بزرگ برای برآوردگرهای حداقل مربعات ضرایب رگرسیون خواهد شد.

در نتیجه اگر از سطوح یکسان X نمونه‌های مختلف گرفته شود، به برآوردگرهای با تفاوت زیاد از هر پارامتر منجر خواهد شد.

تعریف ضریب تعیین[1] چندگانه:

ضریب تعیین چندگانه یک محک مناسب برای ارزیابی مناسبت مدل می‌باشد و با R_p^2 نشان داده می‌شود. R_p^2 ضریب تعیین چندگانه برای یک مدل زیر مجموعه با p جمله می‌باشد و به‌صورت زیر تعریف می‌شود:

$$R_p^2 = \frac{SS_R(p)}{S_{yy}} = 1 - \frac{SS_E(p)}{S_{yy}}$$

که $SS_R(p)$ و $SS_E(p)$ ، به ترتیب نشان‌دهندهٔ مجموع مربعات رگرسیون و مجموع مربعات باقیمانده برای یک مدل زیر مجموعه با p جمله می‌باشد.

وقتی که بیش از دو متغیر رگرسیونی وجود دارد، هم‌خطی اثرات مشابهی ایجاد می‌کند، می‌توان نشان داد که اعضای قطر ماتریس $C = (X'X)^{-1}$ عبارتند از

$$C_{jj} = \frac{1}{(1-R_j^2)} \qquad j = 1, \dots, p \qquad (1.3.3)$$

که در آن R_j^2 ضریب تعیین چندگانه از رگرسیون X_j نسبت به p-1 متغیر رگرسیونی باقیمانده است. اگر هم‌خطی شدید بین X_j و هر زیرمجموعه‌ای از p-1 متغیر رگرسیونی دیگر وجود داشته باشد، در این‌صورت مقدار R_j^2 نزدیک به واحد خواهد بود. چون واریانس $\hat{\beta}_j$ برابر است با $V(\hat{\beta}_j) = C_{jj}\sigma^2 = (1 - R_j^2)^{-1}\sigma^2$، هم‌خطی چندگانه شدید موجب می‌شود که واریانس برآورد حداقل مربعات ضریب رگرسیونی β_j بسیار زیاد شود. بطور کلی اگر X_i و X_j در یک رابطه هم‌خطی چندگانه درگیر باشند کواریانس $\hat{\beta}_i$ و $\hat{\beta}_j$ نیز بزرگ خواهد بود.

[1] Coefficient of Determination

هم‌خطی چندگانه همچنین به ایجاد برآوردهای حداقل مربعات $\hat{\beta}_j$ که از نظر قدرمطلق خیلی بزرگ می‌باشند منجر خواهد شد. برای ملاحظه این مطلب، مربع فاصله $\hat{\beta}$ تا بردار واقعی β را مورد توجه قرار می‌دهیم، یعنی فاصله β از $\hat{\beta}$ را به صورت زیر در نظر می‌گیریم:

$$L_1^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta) \qquad (1.3.4)$$

امید مربع فاصله، $E(L_1^2)$، برابر است با :

$$\begin{aligned} E(L_1^2) &= E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) \\ &= \sum_{j=1}^{p} E(\hat{\beta}_j - \beta_j)^2 \\ &= \sum_{j=1}^{p} V(\hat{\beta}_j) \qquad (1.3.5) \\ &= \sigma^2 Tr(X'X)^{-1} \end{aligned}$$

که اثر[1] یک ماتریس (با Tr نمادگذاری می‌شود) با مجموع اعضای قطراصلی برابر است. وقتی که هم‌خطی چندگانه وجود داشته باشد، بعضی از مقادیر ویژه $X'X$ کوچک خواهند بود. چون اثر یک ماتریس نیز با مجموع مقادیر ویژه آن برابر است(1.3.5) به صورت زیر در می‌آید:

$$E(L_1^2) = \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j} \qquad (1.3.6)$$

که درآن $\lambda_j > 0\ ,\ j = 1,2,\dots,p$ مقادیر ویژه $X'X$ می‌باشند. بنابراین اگر ماتریس $X'X$ به لحاظ هم‌خطی شرایط بیمارگونه داشته باشد حداقل یکی از λ_jها کوچک خواهد بود و(1.3.6) نتیجه می‌دهد که فاصله برآورد حداقل مربعات $\hat{\beta}$ تا مقدار واقعی پارامتر β می‌تواند بزرگ باشد. به‌طور معادل می‌توان نشان داد که

$$\begin{aligned} E(L_1^2) &= E(\hat{\beta} - \beta)'(\hat{\beta} - \beta) \\ &= E(\hat{\beta}'\hat{\beta} - 2\hat{\beta}'\beta + \beta'\beta) \qquad (1.3.7) \end{aligned}$$

یا

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2 Tr(X'X)^{-1} \qquad (1.3.8)$$

یعنی به‌طورکلی طول بردار $\hat{\beta}$ بزرگتر از بردار β می‌باشد. این بدین معنی است که روش حداقل مربعات، ضرایب رگرسیونی را ایجاد می‌کند که از نظر قدرمطلق خیلی بزرگ می‌باشند.

[1] Trace

به‌طورکلی هرگاه هم‌خطی چندگانه شدید وجود داشته باشد، روش حداقل مربعات برآوردهای ضعیفی از هر یک از پارامترهای منفرد مدل بدست می‌دهد. اما این لزوما نتیجه نمی‌دهد که مدل برازش شده یک پیش بینی ضعیفی می‌باشد. اگر پیش بینی‌ها محدود به ناحیه‌هایی از فضای X باشد که هم‌خطی چندگانه تقریبا برقرار است، مدل برازش شده اغلب پیش بینی قابل قبولی به دست می‌دهد. این بدین علت می‌تواند اتفاق بیفتد که ترکیب خطی $\sum_{j=1}^{p}\beta_j x_{ij}$ کاملا خوب برآورد می‌شود، هرچند پارامترهای منفرد β_j به‌طور ضعیفی برآورد می‌شوند. بدین معنی که اگر داده‌های اصلی در امتداد ابر صفحه تعریف شده در رگرسیون قرار گیرند در این صورت مشاهدات آینده که نزدیک این ابر صفحه قرار می‌گیرند علیرغم برآوردهای نامناسب تک پارامترها می‌توانند بصورت دقیق پیش بینی شوند.

به طور خلاصه می‌توان پیامدهای وجود هم‌خطی را به صورت زیر فهرست کرد.

۱- چون در حالت هم‌خطی اطلاعات مستقل در مورد هر یک از متغیرهای مستقل وجود ندارد؛ لذا نمی‌توان اثرات جزئی متغیرهای مذکور روی متغیر وابسته را برآورد کرد.

۲- هنگامی‌که همبستگی شدید بین متغیرهای مستقل وجود داشته باشد کواریانس و واریانس ضرایب بزرگتر برآورد خواهند شد.

۳- برآورد ضرایب رگرسیونی نااریب همچنان خاصیت $Blue$ را دارند. گرچه کمیت به‌دست آمده برای آنها غیر قابل اعتماد است.

۴- در حالتی که با هم‌خطی شدید در مدل مواجه هستیم، پیش‌بینی‌های صورت گرفته در آن غیر قابل اعتماد خواهد بود. در این حالت پیش‌بینی‌ها بر اساس مدلی که دارای زیر مجموعه‌ای از متغیرهای مستقل مدل اصلی است، بهتر صورت می‌گیرد.

۱-۳-۲- علل وقوع هم‌خطی چندگانه

مدل رگرسیون چندمتغیره(1.1.1) را درنظر بگیرید. فرض کنید متغیرهای رگرسیونی و پاسخ، مرکزی شده و به طول واحد مقیاس‌سازی شده‌اند. در نتیجه $X'X$ یک ماتریس $p \times p$ از همبستگی‌های بین متغیرهای رگرسیونی و $X'y$ یک بردار $p \times 1$ از همبستگی‌های بین متغیرهای رگرسیونی و پاسخ است. اگر j امین ستون ماتریس X با X_j نمایش داده شود،