سیزدهمین کنفرانس
آمار ایران
۲-۴ شهریور ۱۳۹۵

رگرسیون ریج با رویکرد امکانی برای متغیرهای ورودی-خروجی فازی

محمدرضا ربیعی [۱]، منصوره مقدس [۲]

[۱] استادیار گروه آمار، دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود

[۲] کارشناس ارشد آمار اقتصادی و اجتماعی، دانشگاه فردوسی مشهد

چکیده: پارامترهای مدل‌های رگرسیون فازی به طوری معمول به کمک مسائل بهینه‌سازی برآورد می‌گردند. در این مسائل عموماً فاصله بین پاسخ مشاهده‌شده و برآورد متناظر آن‌ها با شرایط خاصی مینیمم می‌گردد. در رگرسیون فازی نیز همانند رگرسیون معمولی وجود همخطی در بین متغیرهای پیشگو می‌تواند از کارایی مدل به دست آمده کم کند. از این رو لازم است به روش‌هایی مانند رگرسیون ریج روی آورد. در این مقاله یک مدل رگرسیون ریج فازی با متغیرهای پیشگو و پاسخ فازی و ضرایب دقیق پیشنهاد می‌کنیم.

واژه‌های کلیدی: رگرسیون چندمتغیره فازی، رگرسیون ریج، همخطی ، مسئله بهینه‌سازی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62A86، 62J07.

۱ مقدمه

بررسی رابطه بین یك متغیر وابسته و یك یا چند متغیر مستقل مورد توجه بسیاری از محققین می‌باشد تا بتوانند با استفاده از این رابطه به توصیف و پیش بینی متغیر وابسته بپردازند، از این رابطه در آمار با عنوان رگرسیون نام برده می‌شود. حال اگر متغیرها و یا رابطه بین آن‌ها اعداد مبهمی باشند محیطی که در آن به دنبال رابطه رگرسیونی بین آن‌ها هستیم یك محیط فازی بوده و در نتیجه یك تحلیل رگرسیون فازی خواهیم داشت. در حقیقت رگرسیون فازی یك روش قدرتمند برای تحلیل پدیده ها در محیط فازی است. رگرسیون فازی اولین بار توسط تاناکا و همکاران (۱۹۸۲) معرفی شد و پس از آن توسط بسیاری از محققان اصلاح و گسترش یافت.

[۱] محمدرضا ربیعی : Rabie1354@yahoo.com

A-10-434-3

مدل‌های رگرسیونی در حوزه‌های کاربردی وسیعی مورد استفاده قرار می‌گیرد. یک مسئله جدی که می‌تواند استفاده از مدل رگرسیونی را با اشکال مواجه کند، همخطی یا وابستگی خطی نزدیک بین متغیرهای رگرسیونی است. وجود همخطی توانایی برآورد ضرایب مدل رگرسیون را با مشکل روبرو می‌کند. بنابراین لازم است روش‌هایی معرفی گردد که با وجود همخطی در بین متغیرهای پیشگو، بتواند همچنان مدل رگرسیون کارایی لازم را داشته باشد. برای این منظور، روش‌هایی مانند رگرسیون ریج معرفی شده است.

وجود همخطی در رگرسیون فازی نیز می‌تواند دردسر ساز باشد از این رو محققان به معرفی روش‌های رگرسیون ریج فازی روی آوردند. دونوسو و همکاران (۲۰۰۷) یک روش رگرسیون فازی بر پایه رگرسیون ریج پیشنهاد کردند که به مقابله با مشکل همخطی در محیط فازی پرداخت. در مدل رگرسیون آن‌ها متغیرهای پیشگو(ورودی) دقیق و متغیرهای پاسخ (خروجی) و پارامترهای مدل فازی در نظر گرفته شده است، اما باید توجه داشت، گاهی متغیرهای پیشگو مبهم بوده و لازم است از مدل رگرسیون با متغیر پیشگوی فازی استفاده شود.

ما در این مقاله، یک روش رگرسیون ریج فازی برای حالت متغیرهای پیشگو و پاسخ فازی متقارن و پارامترهای دقیق ارائه می‌کنیم. این مقاله متشکل از بخش‌های زیر است: در بخش بعد به بیان یک سری تعاریف و قضایای به کاربرده در مقاله می‌پردازیم. بخش ۳ مقدماتی درباره مدل رگرسیون ریج بیان می‌گردد. بخش ۴ مدل رگرسیون ریج فازی با متغیرهای پیشگو و پاسخ فازی متقارن و ضرایب دقیق معرفی شده و یک مثال عددی ارائه می‌شود. بخش آخر نیز به نتیجه‌گیری اختصاص دارد.

۲ تعاریف و قضایا

در این بخش به ارائه تعاریف و قضایای مورد نیاز می‌پردازیم که از مرجع دوبوا و پراد (۱۹۸۰) گرفته شده است.

تعریف ۱.۲. *تابع عضویت عدد فازی مثلثی $\tilde{A}=(a,c_L,c_R)$ به صورت زیر است:*

$$\mu_{\tilde{A}}(x)=\begin{cases}1-\left(\frac{a-x}{c_L}\right) & a-c_L\leq x\leq a\\ 1-\left(\frac{x-a}{c_R}\right) & a<x\leq a+c_R\\ 0 & o.w\end{cases}$$

که در آن a مرکز عدد فازی و c_L و c_R به ترتیب پهنای چپ و راست عدد فازی است. مجموعه اعداد فازی مثلثی را با $T(\mathbb{R})$ نشان می‌دهیم. در صورتی که $c_L=c_R$، عدد فازی مثلثی متقارن نامیده می‌شود و با نماد $\tilde{A}=(a,c)_T$ نشان داده می‌شود.

قضیه ۲.۲. *فرض کنید $\tilde{A}_1=(a_1,c_{L1},c_{R1})_T$ و $\tilde{A}_2=(a_2,c_{L2},c_{R2})_T$ اعداد فازی مثلثی و k یک عدد حقیقی باشد. در این صورت داریم:*

$$\tilde{A}_1+\tilde{A}_2=(a_1+a_2,c_{L1}+c_{L2},c_{R1}+c_{R2})_T$$

$$k\tilde{A}_1=\begin{cases}(ka_1,kc_{L1},kc_{R1})_T & k\geq 0\\ (ka_1,-kc_{R1},-kc_{L1})_T & k<0\end{cases}$$

۳ رگرسیون ریج

در فضای رگرسیون احتمالی با حضور همخطی، دترمینان ماتریس $X'X$ به سمت صفر میل می‌کند و این امر باعث نامطلوب شدن پارامترهای برآوردشده می‌گردد. وجود چندهمخطی آثار جدی متعددی بر برآوردهای حداقل مربعات ضرایب رگرسیونی دارد. به عنوان مثال فرض کنید فقط دو متغیر رگرسیونی $X_۱$ و $X_۲$ وجود داشته باشند. با این فرض که $X_۱$ و $X_۲$ و Y به طول واحد مقیاس-سازی شده باشند، مدل را به صورت

$$Y = \beta_۱ X_۱ + \beta_۲ X_۲ + \epsilon$$

در نظر می-گیریم و معادلات حداقل مربعات نرمال عبارتند از:

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

یا

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

که در آن $r_{۱۲}$ ضریب همبستگی پیرسن بین$X_۱$ و $X_۲$ و برای هر $j = ۱, ۲$، r_{jy} ضریب همبستگی پیرسن بین X_j و Y می‌باشد . معکوس ماتریس $X'X$ عبارتست از:

$$C = (X'X)^{-۱} = \begin{pmatrix} \frac{۱}{۱-r_{۱۲}^۲} & \frac{-r_{۱۲}}{۱-r_{۱۲}^۲} \\ \frac{-r_{۱۲}}{۱-r_{۱۲}^۲} & \frac{۱}{۱-r_{۱۲}^۲} r_{۲y} \end{pmatrix} \qquad (۱.۳)$$

و برآوردگرهای ضرایب رگرسیونی عبارتند از :

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

اگر بین $X_۱$ و $X_۲$ همخطی شدید موجود باشد، در این صورت ضریب همبستگی بزرگ خواهد بود و با توجه به رابطه (۱.۳) اگر $|r_{۱۲}| \to ۱$ آنگاه داریم:

$$V(\hat{\beta}_j) = \frac{۱}{۱ - r_{۱۲}^۲} \times \sigma^۲ \to \infty$$

بسته به اینکه $r_{۱۲} \to ۱$ یا $r_{۱۲} \to -۱$ خواهیم داشت:

$$Cov(\hat{\beta}_۱, \hat{\beta}_۲) = \frac{-r_{۱۲}}{۱ - r_{۱۲}^۲} \times \sigma^۲ \to \pm\infty$$

بنابراین همخطی زیاد بین $X_۱$ و $X_۲$ منجر به واریانس‌ها و کواریانس‌های بزرگ برای برآوردگرهای حداقل مربعات ضرایب رگرسیونی خواهد شد. وقتی که بیش از دو متغیر رگرسیونی وجود داشته باشد، همخطی اثرات مشابهی ایجاد می-کند. همخطی همچنین به ایجاد برآوردهای حداقل مربعات $\hat{\beta}_j$ که از نظر قدرمطلق خیلی بزرگ می‌باشند، منجر خواهد شد. (رجوع شود به دراپر و اسمیت (۱۹۸۱))

یکی از روش‌های مقابله با مشکل همخطی، رگرسیون ریج است. در رگرسیون ریج به ماتریس $X'X$ یک پارامتر λ اضافه می‌شود، به عبارت دیگر ماتریس $(X'X - \lambda I)$ جانشین ماتریس $X'X$ می‌گردد، در نتیجه به منظور برآورد پارامترها رابطه زیر به دست می‌آید:

$$\hat{\beta} = (X'X - \lambda I)^{-۱} X'Y$$

هر چه λ بیشتر باشد واریانس پارمترها کمتر اما اریبی آن‌ها بیشتر می‌گردد. هدف مقدار کوچک قابل قبول از λ است که در آن برآورد ریج پایدار شود. مقدار λ معمولاً توسط محقق تعیین می‌گردد و معمولاً عددی بین صفر و یک است. اگرچه در روش رگرسیون ریج،

ضرایب برآوردشده نسبتاً اریب می‌باشند، ولی در مقایسه با برآوردهای کمترین مربعات، خطای کوچکتر و پایداری بیشتری دارند. در واقع برآورد ریج یک تبدیل خطی از برآورد حداقل مربعات است.

۴ مدل رگرسیون ریج فازی پیشنهادی

دونوسو و همکاران (۲۰۰۷) با تعمیم روش هاستی و همکاران (۲۰۰۵)در محیط فازی توانستند یک مسئله مینیمم‌سازی برای برآورد پارامترهای مدل رگرسیون ریج ارائه دهند. روش آن‌ها زمانی که متغیرهای پیشگو، دقیق و متغیر پاسخ، فازی فرض هستند قابل استفاده است. اما زمانی که متغیرهای پیشگو دارای ابهام باشند، روش آن‌ها کارایی خود را از دست می‌دهد. از این رو باید به مدل‌هایی روی آورد که علاوه بر فازی بودن متغیر پاسخ، متغیرهای پیشگو نیز فازی در نظر گرفته شوند. در ادامه سعی کردیم با ارائه روشی جدید مشکل این گونه مدل ها را برطرف کنیم.

مدل رگرسیون زیر را در نظر بگیرید:

$$\tilde{Y} = A_0 + A_1\tilde{X}_1 + \dots + A_p\tilde{X}_p \tag{۱.۴}$$

به طوری که برای هر $i = 1, 2, \dots, n$ و $j = 1, 2, \dots, p$ مقادیر $\tilde{X}_{ij} = (x_{ij}, c_{x_{ij}})_T$، متغیرهای پیشگو، $\tilde{Y}_i = (y_i, c_{y_i})_T$ متغیر پاسخ فازی متقارن و A_0 و A_j پارامترهای مدل رگرسیونی هستند. ضرب بین پارامترهای دقیق مدل و متغیرهای پیشگوی فازی زمانی که اعداد مثلثی فازی باشند با استفاده از قضیه ۲.۲ قابل تعریف است. بنابراین مدل رگرسیون (۱.۴) به صورت زیر بازنویسی می‌شود:

$$\begin{aligned}\tilde{Y}_i &= A_0 + A_1\tilde{X}_1 + \dots + A_p\tilde{X}_p \\ &= \left(A_0 + \sum_{j=1}^{p} A_j x_{ij}, \sum_{j=1}^{p} |A_j| c_{x_{ij}}\right)\end{aligned} \tag{۲.۴}$$

حال با توجه به مدل رگرسیون (۲.۴) و تعمیم روش دونوسو و همکاران (۲۰۰۷) برای حالت متغیرهای پیشگو و پاسخ فازی و ضرایب دقیق، تابع هدف به صورت زیر خواهد بود:

$$\begin{aligned}&k_1 \sum_{i=1}^{n} \left(y_i - (A_0 + \sum_{j=1}^{p} A_j x_{ij})\right)^2 + k_2 \Bigg(\sum_{i=1}^{n} \left(y_i - c_{y_i} - (A_0 + \sum_{j=1}^{p} A_j x_{ij} - \sum_{j=1}^{p} |A_j| c_{x_{ij}})\right)^2 \\ &+ \sum_{i=1}^{n} \left(y_i + c_{y_i} - (A_0 + \sum_{j=1}^{p} A_j x_{ij} + \sum_{j=1}^{p} |A_j| c_{x_{ij}})\right)^2 \Bigg) + \lambda(A_0^2 + \sum_{j=1}^{p} A_j^2)\end{aligned} \tag{۳.۴}$$

که در آن k_1 و k_2 اعداد حقیقی بین صفر و یک هستند که توسط محقق تعیین می‌شوند. در صورتی که $k_1 > k_2$ اهمیت مرکز بیشتر از پهنای پاسخ‌ها در نظر گرفته می‌شود و اگر $k_1 < k_2$ پهنای پاسخ‌ها از اهمیت بیشتری برخودار می‌گردد. همچنین در صورتی که $k_1 = k_2$، اهمیت مرکز و پهنا یکسان فرض شده است. پارامتر λ نیز همان پارامتر ریج می‌باشد. با حل مسئله مینیمم‌سازی (۳.۴)، پارامترهای مدل رگرسیون (۱.۴) برآورد می‌شوند. در ادامه برای توضیح بیشتر، مدل رگرسیونی فوق را بر روی مشاهدات واقعی اجرا می‌کنیم.

۱.۴ معیار نیکویی برازش

در رگرسیون فازی برای مقایسه مدل‌های رگرسیون، معیارهای نیکویی برازش متنوعی وجود دارد. در این بخش، دو معیار رایج در ارزیابی مدل‌های رگرسیونی را بیان می‌کنیم.

تعریف ۱.۴. *(کیم و بیشو (۱۹۹۸)) فرض کنید که $\tilde{Y}_i$، پاسخ مشاهده‌شده و $\hat{\tilde{Y}}_i$ پاسخ برآوردشده iام باشد. اختلاف توابع عضویت پاسخ‌های برآوردشده مدل و مشاهدات متناظر آن‌ها را به صورت زیر تعریف می‌گردد:*

$$D = \frac{1}{n}\sum_{i=1}^{n} D_i = \frac{1}{n}\sum_{i=1}^{n}\int_{S_{\tilde{Y}_i}\cup S_{\hat{\tilde{Y}}_i}} |\mu_{\tilde{Y}_i}(y) - \mu_{\hat{\tilde{Y}}_i}(y)|dy,$$

به طوری که $S_{\tilde{Y}_i}$ و $S_{\hat{\tilde{Y}}_i}$ به ترتیب تکیه‌گاه پاسخ‌های مشاهده‌شده و برآوردهای متناظر آن (طبق تعریف ؟؟) است. این معیار در واقع مساحت غیرمشترک زیر نمودار تابع عضویت پاسخ مشاهده‌شده و برآورد متناظر آن است . هرچه مقدار D کوچکتر باشد، مدل رگرسیون برازش داده شده به داده‌ها از کارایی بیشتری برخوردار است.

تعریف ۲.۴. *(گیلده وگین (۲۰۰۲)) فرض کنید $\tilde{Y}_i$ و $\hat{\tilde{Y}}_i$، به ترتیب پاسخ مشاهده‌شده و برآورد متناظر آن است. در این صورت داریم:*

$$\begin{aligned} D^2_{p,q} &= \frac{1}{n}\sum_{i=1}^{n} D^2_{p,q}(i) \\ &= \frac{1}{n}\sum_{i=1}^{n}((1-q)\int_0^1 ([\hat{\tilde{Y}}_i]^-_\alpha - [\tilde{Y}_i]^-_\alpha)^p d\alpha + q\int_0^1 ([\hat{\tilde{Y}}_i]^+_\alpha - [\tilde{Y}_i]^+_\alpha)^p d\alpha), \end{aligned}$$

که در آن $[\tilde{Y}_i]^-_\alpha$ و $[\tilde{Y}_i]^+_\alpha$، به ترتیب حدود پایین و بالای α- برش پاسخ‌های مشاهده‌شده می‌باشد و به طور مشابه برای پاسخ‌های برآوردشده نیز تعریف می‌شوند. در این معیار به طور معمول مقادیر p و q، به ترتیب، ۲ و ۰/۵ انتخاب می‌شوند.

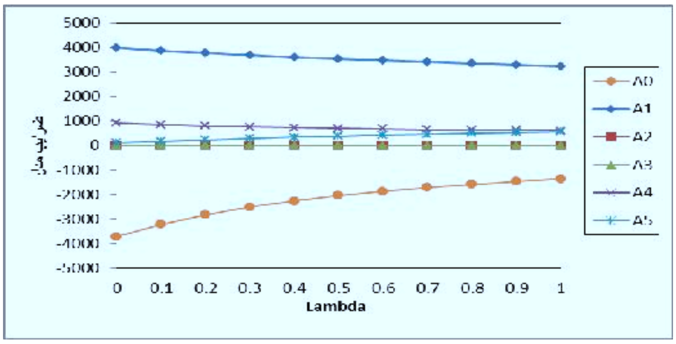
مثال ۳.۴. *در این مثال از داده‌های تاناکا و همکاران (۱۹۸۲) استفاده کردیم. این داده‌ها در جدول ۱ آمده است، که در آن متغیر پاسخ، قیمت فروش خانه (yen ۱۰۰۰)، و متغیرهای پیشگو به ترتیب، کیفیت مصالح ساختمانی (به ترتیب کیفیت پایین، متوسط، بالا)، مساحت طبقه اول (m^2)، مساحت طبقه دوم(m^2)، تعداد اتاق‌ها و تعداد اتاق‌های به سبک ژاپنی می‌باشد. در ابتدا به بررسی*

جدول ۱: داده‌های مقاله تاناکا و همکاران

ردیف	$\tilde{Y}_i$	X_1	X_2	X_3	X_4	X_5	ردیف	$\tilde{Y}_i$	X_1	X_2	X_3	X_4	X_5
۱	$(۶۰۶۰, ۵۵۰)_T$	۱	۳۸/۰۹	۳۶/۴۳	۵	۱	۹	$(۱۰۹۲۰, ۶۰۰)_T$	۲	۷۶/۱۲	۴۳/۰۶	۷	۲
۲	$(۷۱۰۰, ۵۰)_T$	۱	۶۲/۱۰	۲۶/۵۰	۶	۱	۱۰	$(۱۲۰۳۰, ۱۰۰)_T$	۲	۹۰/۲۶	۴۲/۶۴	۷	۲
۳	$(۸۰۸۰, ۴۰۰)_T$	۱	۶۳/۷۶	۴۴/۷۱	۷	۱	۱۱	$(۱۳۹۴۰, ۳۵۰)_T$	۳	۸۵/۷۰	۳۱/۳۳	۶	۳
۴	$(۸۲۸۰, ۱۵۰)_T$	۱	۷۴/۵۲	۳۸/۰۹	۸	۱	۱۲	$(۱۴۲۰۰, ۲۵۰)_T$	۳	۹۵/۲۷	۲۷/۶۴	۶	۳
۵	$(۸۶۵۰, ۷۵۰)_T$	۱	۷۵/۳۸	۴۱/۴۰	۷	۲	۱۳	$(۱۶۰۱۰, ۳۰۰)_T$	۳	۱۰۵/۹۸	۲۷/۶۴	۶	۳
۶	$(۸۵۲۰, ۴۵۰)_T$	۲	۵۲/۹۹	۲۶/۴۹	۴	۲	۱۴	$(۱۶۳۲۰, ۵۰۰)_T$	۳	۷۹/۲۵	۶۶/۸۱	۶	۳
۷	$(۹۱۷۰, ۷۰۰)_T$	۲	۶۲/۹۳	۲۶/۴۹	۵	۲	۱۵	$(۱۶۹۹۰, ۶۵۰)_T$	۳	۱۲۰/۵۰	۳۲/۲۵	۶	۳
۸	$(۱۰۳۱۰, ۲۰۰)_T$	۲	۷۲/۰۴	۳۳/۱۲	۶	۳							

همخطی بین متغیرهای پیشگو می‌پردازیم. مقدار آماره VIF برای بررسی همخطی بین مشاهدات پیشگو به‌صورت $VIF_1 = ۹/۵۰$،

$VIF_۲ = ۶/۰۵$، $VIF_۳ = ۱/۶۷$، $VIF_۴ = ۳/۸۷$ و $VIF_۵ = ۶/۱۳$ به‌دست آمده است. با توجه به مقدار بیشتر از ۵ آماره VIF برای متغیرهای $X_۱$، $X_۲$ و $X_۵$ مشکل همخطی وجود دارد. از این رو بهتر است از روش رگرسیون ریج برای برآورد پارامترهای مدل رگرسیون فازی استفاده شود. با توجه به این‌که متغیرهای پیشگو، دقیق هستند، برای دو متغیر $X_۲$ و $X_۳$ (این دو متغیر پیوسته هستند، درنتیجه امکان خطا در اندازه‌گیری وجود دارد. بنابراین آن‌ها را به صورت فازی در نظر می‌گیریم) پهناهایی متناسب با مقدار مشاهده‌شده در نظرگرفته می‌شود. برای این منظور، با مشورت متخصص، به ازای هر $i = ۱, ۲, ..., n$ و $j = ۲, ۳$، $S_{ij}X_{ij}$ برای پهنای مشاهده مورد نظر انتخاب شدند. به طوری‌که $۰ \leq S_{ij} \leq ۱$ و مقادیر تصادفی باشند (تمامی مشاهدات اعداد فازی مثلثی متقارن هستند). مسئله بهینه‌سازی برای $k_۱ = k_۲ = ۱$ و λ های متفاوت از صفر تا یک با گام ۰/۱ انجام می‌گیرد، پارامترهای برآورد شده در شکل ۱ نشان داده شده است. با توجه به شکل ۱ ، برآوردها به طور تقریبی از $\lambda = ۰/۸$ هموار شده‌اند. بنابراین مدل رگرسیون



شکل ۱: ضرایب برای λ های متفاوت

نهایی به صورت زیر است:

$$\tilde{Y} = -۱۵۸۰/۸۶ + ۳۳۴۳/۳۷\tilde{X}_۱ + ۸/۸۵\tilde{X}_۲ + ۱۰/۸۰\tilde{X}_۳ + ۶۳۳/۵۱\tilde{X}_۴ + ۵۰۰/۹۷\tilde{X}_۵$$

برای پیش‌بینی از مدل فوق، فرض‌کنید بردار $\{(۱, ۰)_T, (۳, ۰)_T, (۵۴/۸۴, ۰/۲۵)_T, (۸۰/۲۱, ۰/۵۰)_T, (۱, ۰)_T\}$ بردار مشاهدات متغیرهای پیشگو باشد، با جایگذاری در مدل، مقدار $\hat{\tilde{Y}} = (۵۴۶۶/۱۴, ۷/۱۲)_T$ برای قیمت خانه پیش‌بینی می‌شود.

مثال ۴.۴. در جدول ۲ ، مجموعه ای از داده‌ها مربوط به درصد تبدیل n-هپتان به استیلن و سه متغیر توضیحی نشان داده شده است. (کانوجی (۱۹۶۱))

برای این داده‌ها، $VIF_۱ = ۱۲/۲۲۵$، $VIF_۲ = ۱/۰۶۲$ و $VIF_۳ = ۱۲/۳۲۵$ محاسبه شده است. با توجه به این مقادیر، در بین داده‌ها مشکل همخطی وجود دارد، از این رو لازم است از روش‌های رگرسیون مناسب در حضور همخطی استفاده شود.

پهنای چپ و راست متغیرها مطابق مثال قبل شبیه‌سازی می‌شوند. نتیجه محاسبات در ۱۱ بار اجرای برنامه بهینه‌سازی برای $k_۱ = k_۲ = ۱$ و λ های متفاوت از صفر تا یک با گام ۰/۱ در شکل ۲ نشان داده شده است.

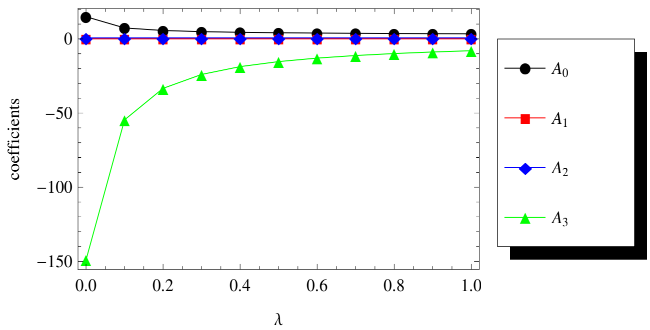
با توجه به شکل ۲ ، برآوردها به طور تقریبی از $\lambda = ۰/۱$ هموار شده‌اند. بنابراین مدل رگرسیون نهایی به صورت زیر است:

$$\tilde{Y} = ۴/۹۳۹ + ۰/۰۲۴\tilde{X}_۱ + ۰/۶۸۵\tilde{X}_۲ - ۴۰/۴۹۹\tilde{X}_۳$$

برای پیش‌بینی از مدل فوق، اگر بردار $\{(۰/۰۳۵, ۰/۰۰۰۲, ۰/۰۰۰۲)_T, (۱۵, ۰/۷۰۵, ۶/۸۲۵)_T, (۱۲۰۰, ۹۹۰, ۹۱۱)_T\}$ بردار مشاهدات متغیرهای پیشگو باشد، مقدار $\hat{\tilde{Y}} = (۴۲/۴۴۴, ۲۴/۱۲۵, ۲۶/۴۳۲)_T$ برای تبدیل n-هپتان به استیلن پیش‌بینی

جدول ۲: داده‌های استیلن

مشاهدات	تبدیل n-هپتان به استیلن Y	دمای راکتور (^oC) $X_۱$	نسبت $H_۲$ به n-هپتان (نسبت مولی) $X_۲$	زمان برخورد (ثانیه) $X_۳$
۱	۴۹	۱۳۰۰	۷٫۵	۰٫۰۱۲
۲	۵۰٫۲	۱۳۰۰	۹	۰٫۰۱۲
۳	۵۰٫۵	۱۳۰۰	۱۱	۰٫۰۱۱۵
۴	۴۸٫۵	۱۳۰۰	۱۳٫۵	۰٫۰۱۳
۵	۴۷٫۵	۱۳۰۰	۱۷	۰٫۰۱۳۵
۶	۴۴٫۵	۱۳۰۰	۲۳	۰٫۰۱۲
۷	۲۸	۱۲۰۰	۵٫۳	۰٫۰۴
۸	۳۱٫۵	۱۲۰۰	۷٫۵	۰٫۰۳۸
۹	۳۴٫۵	۱۲۰۰	۱۱	۰٫۰۳۲
۱۰	۳۵	۱۲۰۰	۱۳٫۵	۰٫۰۲۶
۱۱	۳۸	۱۲۰۰	۱۷	۰٫۰۳۴
۱۲	۳۸٫۵	۱۲۰۰	۲۳	۰٫۰۴۱
۱۳	۱۵	۱۱۰۰	۵٫۳	۰٫۰۸۴
۱۴	۱۷	۱۱۰۰	۷٫۵	۰٫۰۹۸
۱۵	۲۰٫۵	۱۱۰۰	۱۱	۰٫۰۹۲
۱۶	۲۹٫۵	۱۱۰۰	۱۷	۰٫۰۸۶



شکل ۲: ضرایب برای λ های متفاوت مثال ۴.۴

می‌شود. مقادیر D و D_{pq} برای این مثال، به ترتیب برابر است با ۱۷٫۲۹۸ و ۱۵۱٫۰۵.

بحث و نتیجه‌گیری

در این مقاله تلاش کردیم برای مشاهدات فازی که با مشکل همخطی روبرو هستند، مدل رگرسیون ریج فازی را برازش دهیم. برای به انجام رساندن این هدف، با تعمیم یک مدل رگرسیون ریج فازی، یک مسئله برنامه‌ریزی درجه دوم پیشنهاد گردید، به طوری که علاوه بر غلبه بر مشکل همخطی، برای متغیرهای پیشگوی فازی نیز قابل استفاده باشد

مراجع

Asai H., Tanaka S. and Uegima K. (1982), Linear Regression Analysis With Fuzzy Model, *IEEE Trans. Systems Man Cybern* 12, 6, 903-907.

Donoso, S., Nicolás M., and Vila M. A., (2007), Fuzzy ridge regression with non symmetric membership functions and quadratic models. *Intelligent Data Engineering and Automated Learning-IDEAL 2007*. Springer Berlin Heidelberg, 135-144.

Draper N. R. and Smith H. (1981), *Applied Regression Analysis*. John Wiley & Sons, Inc, New York.

Dubois, D., and Prade, H. (1980), Fuzzy Sets and Systems: Theory and Applications. *Mathematics in Science and Engineering*. Vol. 144.

Gildeh B. and Gien D.A. J. (2002), A goodness of fit index to reliability analysis in fuzzy model, *3rd WSEAS International Conference on Fuzzy Sets and Fuzzy Systems. Iterlaken, Switzerland* 11-14.

Hastie T., Tibshirani, R. Friedman, J. and Franklin, J. (2005), The elements of statistical learning: data mining, inference and prediction,*The Mathematical Intelligencer*, 27, 2, 83-85.

Kim B. and Bishu R.R. (1998), Evaluation of fuzzy linear regression models by comparing membership functions, *Fuzzy Sets and Systems* 100, 1, 343-352.

Kunugi T., Tamura T. and Naito T. (1961), New acetylene process uses hydrogen dilution, *Chemical Engineering Progress 57* 43-49.