رگرسیون فازی:
مروری بر چند رویکرد

شهره میرزایی یگانه [۱]، ناصررضا ارقامی [۲]

$arghami\_nr@yahoo.com$, $shohreh\_my@yahoo.com$

چکیده

مبحث رگرسیون فازی از اوائل دهه ۱۹۸۰ و حدوداً هجده سال بعد از معرفی نظریه مجموعه‌های فازی مورد توجه و بررسی قرار گرفت. می‌توان گفت که بیشترین مطالعات در تلفیق روش‌های کلاسیک آماری و روش‌های مبتنی بر مجموعه‌ها و منطق فازی، در زمینه رگرسیون انجام گرفته است. تا به امروز شیوه‌های متعدد و متنوعی از سوی محققان تحت عنوان رگرسیون فازی ارائه گردیده است. در این مقاله به بیان و تشریح مبانی و ایده کلی برخی از این روش‌ها پرداخته می‌شود.

واژه‌های کلیدی: رگرسیون فازی، رگرسیون امکانی، روش مینیمم مربعات فازی، رگرسیون بازه‌ای، عملگرهای حافظ شکل.

۱ مقدمه

رگرسیون یکی از کارآمدترین ابزارهای آماری مورد استفاده در بسیاری از زمینه‌های تحقیقاتی می‌باشد. مدل متداول برای رگرسیون (خطی) آماری به‌صورت زیر است

$$y = a_0 + a_1 x + \ldots + a_p x_p + e \qquad (1)$$

که e، جمله‌ی خطای تصادفی می‌باشد. در رگرسیون آماری سه نوع کمیت قابل اندازه گیری یا قابل برآورد وجود دارد که عبارت‌اند از: متغیر وابسته، متغیر(های) مستقل و پارامترها. در رگرسیون آماری کلاسیک فرض می‌شود که متغیرها و مشاهدات مربوط به آنها دقیق هستند و خطای مدل نیز به خطاهای تصادفی و عدم حضور برخی متغیرها که اثرات جزئی بر متغیر وابسته دارند و در مدل درنظر گرفته نشده‌اند نسبت داده می‌شود. به‌علاوه مفروضاتی نیز در مورد خطاهای مدل در نظر گرفته می‌شود، به عنوان مثال فرض می‌شود که این خطاهای تصادفی، نرمال، ناهمبسته و دارای میانگین صفر و واریانس ثابت می‌باشند. اما در بسیاری موارد ممکن است برخی از این مفروضات برقرار نباشند. در این موارد باید از شیوه‌های جایگزین آماری استفاده نمود.

از سوی دیگر ممکن است مشاهدات مربوط به یک یا چند متغیر و یا اساساً ارتباط میان متغیرها نادقیق و مبهم باشد. در چنین مواردی روش‌های رگرسیون کلاسیک قابلیت کافی جهت مدل‌سازی داده‌ها را نخواهند داشت.

یکی از شیوه‌های جایگزین رگرسیون کلاسیک در این‌گونه موارد، رگرسیون فازی می‌باشد. رگرسیون فازی و به‌عبارت دقیق‌تر، رگرسیون در محیط فازی، تنوعی بسیار دارد.

بسته به اینکه هریک از کمیت‌های مدل (۱) دقیق یا فازی درنظر گرفته شوند، چهار رده مدل رگرسیونی مختلف زیر حاصل می‌شوند؛ در مدل‌های زیر حروف بزرگ نشان دهنده اعداد فازی و حروف کوچک نشان دهنده اعداد دقیق

[۱] دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان
[۲] گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

می‌باشند.

۱. $(\underline{x}, y, \underline{A})$، حالتی که تنها ارتباط بین متغیرها (ضرائب معادله رگرسیونی) فازی درنظر گرفته می‌شود. این مدل‌ها به مدل‌های رگرسیونی امکانی نیز موسوم‌اند.

۲. $(\underline{X}, Y, \underline{a})$، مدل با ورودی و خروجی فازی، به این مدل، مدل رگرسیونی الزامی [۳] نیز گفته می‌شود.

۳. $(\underline{x}, Y, \underline{A})$، مدل با ضرائب و خروجی فازی. اکثر مقالات ارائه شده در رگرسیون فازی در ارتباط با این‌گونه مدل‌ها می‌باشند.

۴. $(\underline{X}, Y, \underline{A})$، مدل با ورودی، خروجی و ضرائب فازی. رویکرد به این مدل به تازگی بیشتر مورد توجه قرار گرفته است.

علاوه بر تقسیم‌بندی یاد شده، انواع رگرسیون فازی با درنظر گرفتن جنبه‌های دیگر به‌صورت زیر تقسیم‌بندی می‌شوند

— از نظر نوع الگو

مشابه با حالت رگرسیون آماری، انواع مدل‌های رگرسیون خطی فازی [۴]، رگرسیون چندجمله‌ای فازی و امثال آن ارائه شده‌اند.

— از نظر نوع اعداد فازی

عامل دیگری که به تنوع رگرسیون فازی می‌افزاید، نوع توابعی است که برای نشان دادن تابع عضویت کمیت های فازی انتخاب می‌شوند. انواع اعداد فازی عبارت‌اند از: اعداد فازی مثلثی، اعداد فازی سهموی، اعداد فازی نرمال، اعداد فازی بازه‌ای و غیره.

نظر به سادگی محاسبات و ویژگی‌های مطلوب دیگر، در رگرسیون فازی، معمولاً از اعداد فازی **LR**، بویژه اعداد فازی مثلثی و نرمال استفاده می‌شود.

— از نظر روش

مشابه با رگرسیون آماری، اصل یا اصولی که برآورد پارامترهای مدل بر اساس آنها انجام می‌پذیرند نیز عاملی در ایجاد تنوع در رگرسیون فازی است. در رگرسیون آماری معمولاً از اصل مینیمم مربعات جهت برآورد پارامترها استفاده می‌شود. رگرسیون فازی از نظر روش به انواع مختلف از جمله موارد زیر تقسیم می‌شود

i) مینیمم‌سازی ابهام کل مدل (مجموع ابهام خروجی‌های برآورد شده)، با شرط شمول Y_i در $\hat{Y}_i$. این روش، روشی است که در رگرسیون امکانی بکار می رود.

ii) مینیمم‌سازی ابهام کل مدل، با شرط شمول $\hat{Y}_i$ در Y_i، این روش به روش رگرسیون الزامی موسوم است.

iii) تلفیق رگرسیون امکانی و الزامی.

iv) مینیمم‌سازی مجموع مربعات (مینیمم کردن مجموع فواصل Y_i و $\hat{Y}_I$) که رگرسیون مینیمم مربعات فازی (FLS) نامیده می‌شود.

v) تلفیق رگرسیون امکانی و رگرسیون مبتنی بر روش مینیمم مربعات.

گفتنی است که بیشتر روش های بهینه‌سازی در رگرسیون امکانی و الزامی، مبتنی بر حل مسائل برنامه‌ریزی خطی (LP) و یا برنامه‌ریزی غیر خطی (QP) می‌باشند.

— از نظر تعامل ضرائب

وجود یا عدم وجود تعامل بین ضرائب فازی مدل، رگرسیون فازی را به دو نوع رگرسیون فازی با تأثیر متقابل (Interactive) و رگرسیون فازی بدون تأثیر متقابل (Non-Interactive) تقسیم می‌کند.

با توجه به تقسیم‌بندی چهارگانه که قبلاً بیان شد و نیز تقسیم‌بندی‌های فوق آشکار است که تنوع بسیاری از روش‌های رگرسیونی در محیط فازی وجود داشته باشد. در

---

[۳] Necessity Regression

[۴] Fuzzy Linear Regression

این مقاله و در بخش‌های آینده، برخی از این روش‌ها را مرور می‌کنیم. در بیان شیوه‌های مختلف رگرسیون فازی تنها به اصول کلی اشاره شده است و خوانندگان می‌توانند برای مطالعه جزئیات بیشتر به اصل مقالات مراجعه نمایند. در این مقاله فرض بر این است که خواننده با مفاهیم مقدماتی مجموعه‌های فازی آشنا است.

۲ پیش‌بینی فازی بر اساس رگرسیون آماری

در بسیاری موارد، گرچه رابطه بین متغیرهای مستقل و وابسته (یعنی ضرائب مدل) بر اساس داده‌های دقیق برآورد می‌شوند، اما هنگام استفاده از رابطه رگرسیونی برای پیش‌بینی متغیر وابسته، مقدار دقیقی از متغیر مستقل در دسترس نیست. فرض کنید که مدل رگرسیون آماری به‌صورت

$$y = a_\circ + a_۱ x_۱ + ... + a_p x_p$$

باشد که در آن همه ضرائب و متغیرها اعداد دقیق بوده و هدف، پیش‌بینی مقدار y، به ازای $x_{\circ j} = X_{\circ j}$ است که $j = ۱, ..., p$ و $X_{\circ j}$ها، مشاهدات فازی از متغیرهای مستقل، به صورت اعداد فازی **LR** هستند.
یاگر(۱۹۸۲) پیش‌بینی فازی بر اساس رگرسیون آماری را به صورت زیر پیشنهاد کرده است

$$\hat{Y}_\circ = \hat{a}_\circ + \hat{a}_۱ X_{\circ ۱} + ... + \hat{a}_p X_{\circ p}$$

که در آن $\hat{a}_j$ها برآوردهای آماری مینیمم مربعات ضرائب هستند. وی، در صورتی که مدل دارای فقط یک متغیر مستقل x باشد، عدد فازی

$$S = s\sqrt{۱ + \frac{۱}{n} + \frac{(X_\circ - \bar{x})^۲}{\sum_{i=۱}^{n} (x_i - \bar{x})^۲}}$$

را، که در آن $s^۲$ برآورد معمولی واریانس جمله خطا می‌باشد، به‌عنوان خطای استاندارد $\hat{Y}_\circ$ پیشنهاد نموده است. با توجه به روابط جمع اعداد فازی **LR** پیش‌بینی به‌دست آمده، طبعاً یک عدد فازی **LR** خواهد بود. گرچه روش فوق از نخستین روش‌های پیشنهاد شده در رگرسیون فازی است، اما در عمل کاربرد چندانی نداشته و به ندرت مورد توجه قرار گرفته‌است.

۳ رگرسیون فازی $(\underline{\mathrm{x}},\mathrm{y},\underline{\mathrm{A}})$

این مدل رگرسیونی (مدل رگرسیونی امکانی)، نخستین بار توسط تاناکا و همکاران (۱۹۸۲) معرفی شده است. در این مدل رگرسیونی، متغیرهای مدل اعداد دقیق و ضرائب مدل اعداد فازی درنظر گرفته می‌شوند.
فرض کنید مشاهدات $(y_i, x_{۱_i}, ..., x_{p_i}), i = ۱, ..., n$، که همگی اعداد دقیق می‌باشند را در اختیار داریم. صورت کلی مدل مورد بحث این‌گونه است

$$\hat{Y}_i = A_\circ + A_۱ x_{i_۱} + ... + A_p x_{i_p} \quad , \; i = ۱, ..., n \qquad (۲)$$

پارامترهای فازی $A_\circ, ..., A_p$ به‌گونه‌ای برآورد می‌شوند که مدل، بر اساس برخی معیارهای نیکویی برازش، بهترین برازش را به داده‌ها داشته باشد. به‌عنوان حالتی خاص، و البته متداول، فرض کنید ضرائب مدل اعداد فازی مثلثی متقارن $A_j = (\alpha_j, \beta_j)_T \;, j = ۱, ..., p$ باشند، در این حالت مدل رگرسیونی (۲) را می‌توان به‌صورت زیر نوشت

$$\begin{aligned}\hat{Y}_i &= (\alpha_\circ, \beta_\circ)_T + (\alpha_۱, \beta_۱)_T x_{i_۱} + ... + (\alpha_p, \beta_p)_T x_{i_p} \\ &= (\alpha_\circ + \alpha_۱ x_{i_۱} + ... + \alpha_p x_{i_p}, \beta_\circ + \beta_۱ |x_{i_۱}| + ... \\ &\quad + \beta_p |x_{i_p}|)_T \\ &= (\hat{y}_i, \hat{e}_i)_T \\ &= (\underline{\alpha}' x_i, \underline{\beta}' |x_i|)_T \qquad , \quad i = ۱, ..., n\end{aligned}$$

β_jها و α_jها به‌گونه‌ای تعیین می‌شوند که، اولاً، مجموع ابهام خروجی‌های فازی برآوردشده، یعنی

$$J=\sum_{i=1}^{n}\beta'|x_i|=\sum_{i=1}^{n}\sum_{j=1}^{p}\beta_j|x_{ij}|$$

مینیمم گردد و ثانیاً، درجه عضویت خروجی‌های مشاهده شده (y_i ها) در $\hat{Y}_i$ دست کم به بزرگی h باشد، یعنی

$$h_i=\hat{Y}_i(y_i)\geq h \Longleftrightarrow y_i\in[\hat{Y}_i]_h=\hat{Y}_i \quad \text{برش}-h$$
$$\Longleftrightarrow \hat{y}_i-(1-h)\hat{e}_i\leq y_i\leq \hat{y}_i+(1-h)\hat{e}_i$$

بدین ترتیب، مسأله‌ی برآورد پارامترهای فازی مدل (۲)، به حل مسأله برنامه‌ریزی خطی زیر می‌انجامد

$$Min\ J=\sum_{i=1}^{n}\sum_{j=1}^{p}\beta_j|x_{ij}|$$

$$s.t. \qquad \hat{Y}_i(y_i)\geq h \quad ,i=1,...,n \qquad (3)$$

مورد استفاده رگرسیون امکانی فوق زمانی است که داده‌ها غیرفازی باشند اما نابرابری $y\neq f(x)$ که در رگرسیون معمولی از طریق وجود خطای تصادفی توجیه می‌شود، از طریق فازی بودن ضرائب قابل توجیه باشد. همچنین در مواردی که تعداد داده‌ها کم باشد، این شیوه می‌تواند جایگزین مناسبی برای رگرسیون آماری باشد. از معایب این روش می‌توان به حساسیت آن نسبت به نقاط پرت و افزایش طول بازه‌ها (ابهام کل مدل) با زیاد شدن n (حجم نمونه) اشاره کرد؛ به‌علاوه، در برنامه‌ریزی خطی، اغلب بعضی ضرائب دقیق به‌دست می‌آیند.

برخی مطالعات دیگر در مورد رگرسیون امکانی که در جهت کاهش مفروضات یا رفع معایب فوق انجام شده است، عبارت‌اند از: ساکاوا و یانو (۱۹۹۰)، ین و همکاران (۱۹۹۹)، نصرآبادی (۲۰۰۴) چوی و دنگ (۲۰۰۴)، گو و تاناکا (۲۰۰۶).

۴ روش مینیمم مربعات فازی (FLS)

این روش برای حالتی که متغیرها فازی و پارامترها دقیق ($\underline{X}$,Y,$\underline{a}$)، و نیز حالتی که متغیر وابسته و پارامترها فازی، و متغیر مستقل دقیق ($\underline{x}$,Y,$\underline{A}$) می‌باشند، نخستین بار توسط دیاموند (۱۹۸۷) و کلمینس (۱۹۸۷) مورد بررسی قرار گرفت. فرض کنید مشاهدات فازی زیر را در اختیار داشته باشیم

$$Y_i=(y_i,\underline{e_i},\overline{e_i})_T$$
$$X_i=(x_i,\underline{c_i},\overline{c_i})_T \qquad ,i=1,...,n$$

که X_iها و Y_iها اعداد فازی مثلثی هستند. مدل رگرسیونی

$$\hat{Y}_i=\alpha_0+\alpha_1X_i \qquad (4)$$

را که در آن α_0 و α_1 پارامترهای حقیقی مجهول می باشند در نظر بگیرید. بر اساس حساب اعداد فازی مدل (۴) را می‌توان به شکل زیر نوشت

$$\hat{Y}_i=(\alpha_0+\alpha_1x_i,|\alpha_1|\underline{c_i},|\alpha_1|\overline{c_i})_T$$

اکنون بر پایه‌ی مشاهدات، α_0 و α_1 را به‌گونه‌ای برآورد می‌کنیم که مجموع فاصله Y_i و $\hat{Y}_i$ها، یعنی

$$r(\alpha_0,\alpha_1)=\sum_{i=1}^{n}d(\hat{Y}_i,Y_i) \qquad (5)$$

مینیمم گردد، که d نوعی فاصله بین دو عدد فازی است. به‌عنوان مثال می‌توان از متر زیر که توسط پوری و همکاران (۱۹۸۶) معرفی شده‌است استفاده نمود

$$d(F,G)=[(f-L_F)-(g-L_G)]^2+[f-g]^2+[(f+U_f)-(g-U_G)]^2$$

که در آن

$$F=(f,L_F,U_F)_T,G=(g,L_G,U_G)_T$$

با مشتق‌گیری از رابطه (۵) نسبت به α_0 و α_1، برآوردهای این دو پارامتر حاصل می‌شود.

۵ یک روش مینیمم مربعات دیگر $(\underline{x}, Y, \underline{A})$

فرض کنید مشاهدات مربوط به متغیر(های) توضیحی، دقیق باشند، ولی مشاهدات متغیر وابسته به‌صورت اعداد فازی $Y_i = (y_i, e_i), i = ۱, ..., n$ باشند. مدل رگرسیونی با ضرائب فازی زیر را درنظر می‌گیریم

$$\hat{Y}_i = A_۰ + A_۱ x_{i۱} + ... + A_p x_{ip} \qquad i = ۱, ..., n$$

و

$$A_i = (\alpha_i, \beta_i)_T$$

در این صورت $\hat{Y}_i$ یک عدد فازی مثلثی به شکل زیر خواهد بود

$$\begin{aligned}\hat{Y}_i &= (\alpha_۰ + \alpha_۱ x_{i۱} + ... + \alpha_p x_{ip}, \beta_۰ + \beta_۱ |x_{i۱}| \\ &\quad + ... + \beta_p |x_{ip}|)_T \\ &= (\underline{\alpha}' x_i, \underline{\beta}' |x_i|)_T\end{aligned}$$

پارامترهای مدل را به‌گونه‌ای برآورد می‌کنیم که $\sum_{i=۱}^{n} d(\hat{Y}_i, Y_i)$ مینیمم گردد، که مشابه روش قبل، d نوعی فاصله بین دو عدد فازی است.

برای مثال می‌توان از متر زیر استفاده نمود

$$d(F_۱, F_۲) = (f_۱ - f_۲)^۲ + (c_۱ - c_۲)^۲$$

که $F_۱ = (f_۱, c_۱)_T, F_۲ = (f_۲, c_۲)_T$

بر این اساس مجموع مربعات خطاها به‌صورت

$$\sum_{i=۱}^{n} d(\hat{Y}_i, Y_i) = \sum_{i=۱}^{n} (y_i - \underline{\alpha}' \underline{x})^۲ + \sum_{i=۱}^{n} (e_i - \underline{\beta}' |\underline{x_i}|)^۲$$

خواهد بود. با مشتق‌گیری از رابطه فوق برآوردهای α_i و β_iها به‌دست می‌آیند. لازم به ذکر است که در این روش امکان قرار نگرفتن بعضی y_iها در تکیه‌گاه $\hat{Y}_i$ وجود دارد، ولی مسأله حساسیت نسبت به نقاط پرت و یا دقیق به‌دست آمدن بعضی ضرائب در این شیوه مطرح نیست. به‌علاوه، مجموع پهناهای خروجی‌های مدل (ابهام کل مدل) نیز با بزرگ شدن حجم نمونه افزایش نمی‌یابد.

عطف به مطالب این بخش و بخش قبل، لازم به ذکر است که بارگیلا و همکاران (۲۰۰۷) رگرسیون چندگانه را به حالتی که داده‌ها فازی باشند، بر پایه معیار کمترین مربعات خطا، تعمیم داده‌اند و یک الگوریتم تکراری برای محاسبات مربوطه ارائه نموده‌اند. روش کمترین مربعات فازی در مدل رگرسیونی بر پایه مجموعه‌های فازی تصادفی، توسط کرشمر (۲۰۰۶) نیز بررسی شده است. همچنین، ویژگی‌های مجانبی برآوردگرهای کمترین مربعات برای مدل‌های رگرسیونی با ورودی و خروجی فازی و جملات خطای تصادفی توسط کیم و همکاران (۲۰۰۷) مورد مطالعه قرار گرفته‌اند.

۶ تلفیق رگرسیون امکانی و رگرسیون مبتنی بر روش کمترین مربعات فازی

این شیوه نخستین بار توسط تاناکا و لی (۱۹۹۷) ارائه شده است. مجدداً مدل رگرسیون فازی به صورت زیر درنظر گرفته می‌شود

$$\hat{Y}_i = A_۰ + A_۱ x_{i۱} + ... + A_p x_{ip} = (\underline{\alpha}' x_i, \underline{\beta}' |\underline{x}_i|)_T$$

در این شیوه، تابع هدف به صورت زیر تشکیل می‌شود

$$\begin{aligned}J &= k_۱ \sum_{i=۱}^{n} (y_i - \underline{\alpha}' \underline{x}_i)^۲ + k_۲ \sum_{i=۱}^{n} (\underline{\beta}' |\underline{x}_i|)^۲ \\ &= k_۱ \sum_{i=۱}^{n} (y_i - \underline{\alpha}' \underline{x}_i)^۲ + k_۲ \sum_{i=۱}^{n} \underline{\beta}' \underline{x}_i \underline{x}_i' \underline{\beta} \\ &= k_۱ \sum_{i=۱}^{n} (y_i - \underline{\alpha}' \underline{x}_i)^۲ + k_۲ \underline{\beta}' [\sum_{i=۱}^{n} \underline{x}_i \underline{x}_i'] \beta\end{aligned}$$

که $k_۱$ و $k_۲$ ضرائب وزنی مثبت هستند. ملاحظه می‌کنید که تابع هدف، در واقع ترکیبی از مربع خطاها (جملات اول) و ابهام مدل (جملات دوم) است. تابع هدف با توجه به محدودیت‌های زیر مینیمم می‌شود

$$y_i \in [\hat{Y}_i]_h \Longleftrightarrow \hat{y}_i - (۱ - h)\hat{e}_i \leq y_i \leq \hat{y}_i + (۱ - h)\hat{e}_i$$
$$, i = ۱, ..., n$$

در این حالت تابع هدف درجه دو، و قیدها درجه اول هستند. لذا با مسأله برنامه‌ریزی درجه دو (QP) روبرو هستیم. امتیاز این روش بر روش رگرسیون امکانی (روش تاناکا و همکاران (۱۹۸۲)) این است که در رگرسیون امکانی، اغلب بعضی از ضرائب دقیق به‌دست می‌آیند، اما در این روش با چنین مسأله‌ای مواجه نخواهیم بود. البته، هنوز مسأله حساسیت مدل نسبت به نقاط پرت در این شیوه نیز وجود دارد.

۷ رگرسیون بازه ای

۱.۷ رگرسیون بازه ای

شیوه رگرسیون بازه‌ای، به دو روش برنامه‌ریزی خطی و برنامه‌ریزی درجه دو، توسط تاناکا و لی (۱۹۹۸) پیشنهاد شده است. مدل رگرسیون بازه‌ای به‌صورت

$$Y_i = A_0 + A_1 x_{i1} + \dots + A_p x_{ip} \quad , i = 1, \dots, n$$

بیان می‌شود که در آن x_iها مقادیر دقیق و متغیر وابسته و ضرائب، بازه‌هایی به‌صورت زیر هستند

$$\begin{array}{ll} Y_i = [\underline{y_i}, \overline{y_i}] = (y_i, e_i)_I & , i = 1, \dots, n \\ A_j = [\underline{a_j}, \overline{a_j}] = (\alpha_j, \beta_j)_I & , j = 1, \dots, p \end{array}$$

بر این اساس، و با توجه به حساب بازه‌ای، خواهیم داشت

$$\begin{array}{ll} \hat{Y}_i = (\alpha_0, \beta_0)_I + (\alpha_1, \beta_1)_I x_{i1} + \dots + (\alpha_p, \beta_p)_I x_{ip} & \\ \quad = (\underline{\alpha}' x_i, \underline{\beta}' |\underline{x}_i|)_I & , i = 1, \dots, n \end{array}$$

حال می‌توان دو برنامه مینیمم‌سازی زیر را درنظر گرفت

$$Min\ J = \sum_{i=1}^{n} \underline{\beta}' |\underline{x}_i| \quad s.t. \ \ y_i \in \hat{Y}_i \ , \ i = 1, \dots, n \quad (6)$$

$$MinJ = \underline{\beta}' (\sum_{i=1}^{n} |\underline{x}_i| |\underline{x}_i|') \underline{\beta} \quad s.t. y_i \in \hat{Y}_i, i = 1, \dots, n \quad (7)$$

که اولی با به کار گیری روش‌های LP و دومی توسط روش‌های QP حل می‌شود. لازم به ذکر است که مورد استفاده روش فوق زمانی است که محقق از حدود متغیر Y با خبر باشد، یعنی بداند که y_i در چه بازه‌ای قرار دارد و اطلاع از اینکه ضرائب در چه حدودی قرار دارند برای وی کفایت کند.

۲.۷ تلفیق رگرسیون بازه‌ای و روش کمترین مربعات

در این شیوه، ضرائب بازه‌ای، توسط روش مینیمم مربعات و با حل برنامه‌ریزی درجه دوی زیر برآورد می‌شوند

$$\begin{array}{l} Min\ J = k_1 \sum_{i=1}^{n} (y_i - \underline{\alpha}' \underline{x}_i)^2 \\ \quad + k \underline{\beta}' (\sum_{i=1}^{n} |\underline{x}_i| |\underline{x}_i|') \underline{\beta} \\ s.t. \\ \quad Y_i \subset \hat{Y}_i \qquad , i = 1, \dots, n \end{array}$$

که در آن $\sum_{i=1}^{n} (y_i - \underline{\alpha}' \underline{x}_i)^2$، مطابق با برآورد کمترین مربعات و k_1 و k_2 ضرائب وزنی مثبت می‌باشند. این تابع هدف مجموع مربعات پهناهای خروجی‌های برآوردشده و مجموع مربعات فواصل بین مراکز خروجی‌های برآوردشده و خروجی‌های مشاهده شده را مینیمم می‌کند. برای مطالعه بیشتر در مورد این روش، به تاناکا و لی (۱۹۹۸) مراجعه نمایید.

۳.۷ تلفیق رگرسیون امکانی و رگرسیون الزامی

در این روش، دو مدل رگرسیونی بالایی و پائینی به‌صورت زیر تعریف می‌شوند (تاناکا و لی (۱۹۹۸))

$$\begin{array}{ll} Y_i^* = A_0^* + A_1^* x_{i1} + \dots + A_p^* x_{ip} & , i = 1, \dots, n \\ Y_{*i} = A_{*0} + A_{*1} x_{i1} + \dots + A_{*p} x_{ip} & , i = 1, \dots, n \end{array}$$

می‌توان ثابت کرد که اگر $A_{*j} \subset A_j^*, j = 1, \dots, p$، آن‌گاه $\hat{Y}_*(\underline{x}) \subset \hat{Y^*}(\underline{x})$. با در نظر گرفتن ضرائب رگرسیون بالایی و پائینی به‌صورت زیر

$$A_{*j} = (\alpha_j, \beta_j)_I \quad , \quad A_j^* = (\alpha_j, \beta_j + d_j)_I$$

مسأله به‌دست آوردن ضرائب بازه‌ای A_{*j} و A_j^*، به مسأله LP زیر تبدیل می‌شود

$$Min\ J^* = \sum_{i=1}^{n} (\underline{\beta} + \underline{d})' |\underline{x}_i| \ \ and \ \ Max\ J_* = \sum_{i=1}^{n} \underline{\beta}' |\underline{x}_i|$$

$$s.t. \qquad \hat{Y}_{*i} \subset Y_i \subset \hat{Y}_i^* \quad , i = 1, ..., n$$

ضرائب مدل از حل مسأله QP زیرنیز قابل حصول می‌باشند

$$MinJ = J^* - J_*$$

$$s.t. \quad \hat{Y}_{*i} \subset Y_I \subset \hat{Y}_i^* \ \ , i = 1, ..., n$$

به‌طوری که $J = \underline{d}'(\sum_{i=1}^{n} |x_i||x_i|')\underline{d}$.

از مزیت‌های مسأله QP نسبت به LP این‌است که در روش LP، ممکن است بعضی ضرائب دقیق برآورد شوند.

در پایان این بخش لازم به ذکر است که، رگرسیون بازه‌ای با بازه‌های نامتقارن، توسط زو وهمکاران (۲۰۰۸) مورد مطالعه قرار گرفته‌است. رویکرد دیگری در ارتباط با مدل رگرسیونی مبتنی بر داده‌های بازه‌ای، توسط چوانگ (۲۰۰۷) بررسی شده‌است.

۸ رگرسیون فازی با تأثیر متقابل

۱.۸ ضرائب با تأثیر متقابل

در این زیر بخش مقدماتی در ارتباط با بردارهای فازی و بردارهای فازی با تأثیر متقابل ارائه می‌گردد.

تعریف ۱ زیرمجموعه‌ی فازی $\tilde{B}$ از $\mathbf{R^r}$ یک بردار فازی نامیده می‌شود اگر هر h− برش از آن، $h \in [0, 1]$، یک زیر مجموعه محدب از $\mathbf{R^r}$ باشد.

تعریف ۲ تصویر بردار فازی $\tilde{B}$ بر بعد j ام فضای $\mathbf{R^r}$ عبارت است از عدد فازی B_j، که تابع عضویت آن به‌صورت زیر است

$$Bj(u) = Sup_{b_j = u} \tilde{B}(\underline{b})$$

تعریف ۳ توزیع امکانی بردار $\tilde{B}$، و نیز توزیع امکانی $B_1, ..., B_r$، بدون تأثیر متقابل نامیده می‌شوند هر گاه

$$\tilde{B}(\underline{b}) = min\{B_1(b_1), ..., B_r(b_r)\}$$

به عبارت دیگر $\tilde{B}$ بدون تأثیر متقابل است اگر هر h− برش آن یک مجموعه حاصل‌ضرب باشد.

تعریف ۴ اگر تابع عضویت عدد فازی A به ازای هر مقدار a به‌صورت زیر باشد

$$A(a) = e^{-(\frac{a-\alpha}{\beta})^2} \qquad a \in R$$

آن گاه A را یک عدد فازی نرمال گوییم و به‌صورت

$$A = (\alpha, \beta)_N$$

نمایش می‌دهیم.

تعریف ۵ اگر توزیع امکانی بردار $\tilde{A}$ به‌صورت

$$\tilde{A}(\underline{a}) = e^{-(\underline{a}-\underline{\alpha})' D^{-1} (\underline{a}-\underline{\alpha})}$$

باشد که D یک ماتریس معین مثبت است، در این صورت بردار $\tilde{A}$ یک بردار فازی نرمال نامیده می‌شود و آن را به‌صورت زیر نشان می‌دهیم

$$\tilde{A} = (\underline{\alpha}, D)_N$$

نکته: اگر ماتریس D یک ماتریس قطری باشد، آن‌گاه بردار فازی $\tilde{A}$ بدون تأثیر متقابل است.

۲.۸ رگرسیون فازی با تأثیر متقابل

بردار مشاهدات را به‌صورت زیر درنظر بگیرید

$$(y_i, x_{i1}, ..., x_{ip}) \quad , \quad i = 1, ..., n$$

مدل رگرسیونی با ضرائب فازی به‌صورت زیر درنظر گرفته می‌شود

$$Y = A_0 + A_1 x_1 + ... + A_p x_p = \tilde{A}\underline{x}$$

که $\underline{x}$ بردار مشاهدات (ورودی ها) و $\tilde{A} = (A_1, ..., A_p)'$ بردار ضرائب فازی بر اساس تعریف (۵) است

$$\tilde{A} = (A_0, A_1, ..., A_p)' = (\underline{\alpha}, D)_N$$

با استفاده از اصل گسترش به‌راحتی ثابت می‌شود که

$$\hat{Y}_i = \tilde{A}\underline{x}_i = (\alpha'\underline{x}_i, \underline{x}_i' D \underline{x}_i)_N$$

روشی که تاناکا و همکاران (۱۹۹۵) برای برآورد پارامترها پیشنهاد نموده‌اند بدین صورت است که ابتدا برآورد $\underline{\alpha}$، $\underline{\alpha}^*$ را از طریق رگرسیون آماری زیر به‌دست می آوریم

$$\begin{array}{l} y_i = \alpha_0 + \alpha_1 x_{i1} + ... + \alpha_p x_{ip} \\ \Rightarrow \hat{\alpha}_j = \alpha_j^* \qquad j = 1, ..., p \end{array}$$

و سپس D از حل مسأله LP زیر به‌دست می آید

$$\begin{array}{ll} Min \sum_{i=1}^{n} \underline{x}_i' D \underline{x}_i & \\ \quad s.t. & y_i \in [\hat{Y}_i]_\alpha, \ D > 0 \\ \Longleftrightarrow s.t. & \mu_{\hat{Y}_i}(yi) \geq h \\ \Longleftrightarrow s.t. & e^{-(y_i - \underline{\alpha}^{*\prime}\underline{x}_i)^2/(\underline{x}_i' D \underline{x}_i)} \geq h \\ \Longleftrightarrow s.t. & (y_i - \underline{\alpha}^{*\prime}\underline{x}_i)^2 \leq (-lnh)(\underline{x}_i' D x_i) \\ \Longleftrightarrow s.t. & \underline{\alpha}^{*\prime}\underline{x}_i - \sqrt{(-lnh)(\underline{x}_i' D \underline{x}_i)} \leq y_i \leq \underline{\alpha}^{*\prime}\underline{x}_i \\ & \qquad + \sqrt{(-lnh)\underline{x}_i' D \underline{x}_i} \end{array}$$

بدین ترتیب می‌توان به ازای هر مشاهده مربوط به متغیر توضیحی، مقدار متغیر وابسته را به‌صورت یک عدد فازی برآورد نمود.

[5] Fuzzy Clustering

۹ رگرسیون آماری با خوشه‌بندی فازی

۱.۹ خوشه‌بندی فازی[5]

همان‌گونه که می‌دانیم، در خوشه‌بندی کلاسیک، هدف افراز مجموعه‌ای از n شی به k گروه از اشیا همگن است. اما در خوشه‌بندی فازی، این تقسیم‌بندی یک تقسیم‌بندی منعطف و نادقیق است.

تعریف ٦ خوشه‌بندی فازی عبارت‌است از تعیین درجه عضویت هریک از واحدهای نمونه در هریک از k خوشه، بر اساس $r = p + 1$ متغیر؛ که اگر درجه عضویت واحد i ام نمونه در خوشه jام را با f_{ij} نشان دهیم، f_{ij} ها در روابط زیر صدق می کنند

$$\begin{array}{l} 0 \leq f_{ij} \leq 1, \forall i, j \\ \sum_{j=1}^{k} f_{ij} = 1 \quad i = 1, ..., n \\ \sum_{i=1}^{n} f_{ij} > 0 \quad , j = 1, ..., k \end{array}$$

جهت تعیین f_{ij} ها، کمیت زیر را

$$L = \sum_{i=1}^{n} \sum_{j=1}^{k} f_{ij}^2 d_{ij}$$

تحت محدودیت

$$\sum_{j=1}^{k} f_{ij} = 1 \quad \forall i$$

مینیمم می کنیم که d_{ij} فاصله واحد i از خوشه jام می‌باشد. به‌عنوان مثال می‌توان از متر زیر

$$d_{ij} = (\underline{t}_i - \bar{t}_j)' M_j (\underline{t}_i - \bar{t}_j)$$

که توسط ماهالانوبیس (۱۹۳٦) معرفی شده‌است، استفاده نمود که $\bar{t}_j$ و M_j به‌ترتیب میانگین و ماتریس کوواریانس داده‌های خوشه jام هستند. خوشه‌بندی فازی می‌تواند بر اساس تعاریف دیگری از d_{ij} نیز به‌دست آید. خوشه‌بندی فازی برای حالتی که داده‌ها فازی باشند نیز قابل تعریف است.

۲.۹ رگرسیون آماری با خوشه‌بندی فازی

در بعضی موارد، داده‌های مربوط به متغیرها، داده‌های همگنی نیستند، یعنی، ممکن است داده‌ها از الگوهای متفاوتی آمده باشند، در این‌گونه موارد بهتر است که قبل از انجام رگرسیون، داده‌ها خوشه‌بندی شوند. رگرسیون آماری با خوشه‌بندی فازی توسط جاجوگا (۱۹۸۶) معرفی شد. مشاهدات مربوط به $r = p + 1$ متغیر را به‌صورت زیر درنظر بگیرید $(y_i, x_{i1}, ..., x_{ip}),\ i = 1, ..., n$. فرض کنید k، تعداد خوشه‌ها، عددی معین باشد، در این صورت پس از انجام خوشه‌بندی فازی و تعیین مقادیر f_{ij} در هر یک از k خوشه، با انجام یک رگرسیون آماری تعمیم یافته کمیت‌های زیر را به‌دست می آوریم

$$\begin{aligned}
&\alpha_j = (X'_0 F_j X_0)^{-1} X'_0 F_j \underline{y} \\
&\hat{e}_j = \underline{\hat{y}}_j - \underline{y}_j \\
&F_j = diag(f_{1j}, ..., f_{nj}), \qquad j = 1, ..., k \\
&R_J^2 = 1 - \frac{\sum f_{ij}(\hat{y}_{ij} - \bar{y}_{j.})^2}{\sum f_{ij}(y_i - \bar{y}_{j.})^2}\ , \bar{y}_j = \frac{1}{\sum_{i=1}^n f_{ij}} \sum_{i=1}^n f_{ij} y_i \\
&R^2 = \sum_{j=1}^k b_j R_j^2 \qquad , b_j = \frac{\sum_{i=1}^n f_{ij}}{\sum_{j=1}^k \sum_{i=1}^n f_{ij}}
\end{aligned}$$

مقدار کمیت R^2 را به ازای kهای مختلف، $k = 1, 2, ...$، محاسبه می کنیم و هرگاه افزایش قابل توجهی در مقدار R^2 ایجاد نشود، k متناظر را به عنوان تعداد خوشه‌ها درنظر می‌گیریم. ملاحظه می‌کنید که در این روش هیچ‌گونه شرط شمولی درنظر گرفته نشده‌است. مؤلفین دیگر، ازجمله یانگ و کو (۱۹۹۷) استفاده از خوشه بندی فازی را به رگرسیون با داده‌های فازی تعمیم داده‌اند.

۱۰ مقایسه رگرسیون فازی و رگرسیون آماری ناپارامتری

از آنجا که در رگرسیون ناپارامتری (NP) هیچ‌گونه ساختار احتمالی برای متغیرها فرض نمی‌شود، این روش جانشین مناسبی برای رگرسیون فازی (FZ) محسوب می‌شود. کیم و چن (۱۹۹۷) مقایسه‌ای بین رگرسیون خطی فازی و رگرسیون آماری خطی ناپارامتری انجام داده‌اند. عملکرد توصیفی و پیش بینی در دو رگرسیون با استفاده از شبیه‌سازی مقایسه شده‌اند، تا مشخص شود که تحت چه شرایطی کدام رگرسیون عملکرد بهتری خواهد داشت. مقایسه این دو روش با توجه به عوامل مختلف در جدول زیر آمده است

عامل		توصیف	پیش بینی
حجم نمونه	کم	$FZ > NP$	$FZ = NP$
	زیاد	$NP > FZ$	$NP > FZ$
کیفیت داده ها	خوب	$NP > FZ$	$NP > FZ$
	بد	$FZ > NP$	$NP > FZ$
صحت الگو	درست	$NP > FZ$	$NP > FZ$
	نادرست	$FZ > NP$	$NP > FZ$
ناهمپراشی	وجود دارد	$NP > FZ$	$NP > FZ$
	وجود ندارد	$FZ > NP$	$NP > FZ$
خودهمبستگی	منفی	$NP > FZ$	$NP > FZ$
	صفر	$FZ > NP$	$NP > FZ$
	مثبت	$FZ > NP$	$NP > FZ$

نتایج نشان می‌دهند که رگرسیون ناپارامتری از نظر قابلیت پیش‌بینی بر رگرسیون فازی ارجحیت دارد. در حالی که قابلیت توصیفی دوروش بسته به عوامل مختلف متفاوت است. هنگامی که حجم نمونه اندک باشد، جملات خطا پراکندگی کمی داشته باشند و یا هنگامی که روابط بین متغیرها به خوبی بیان نشده باشد رگرسیون فازی از نظر توصیفی عملکرد بهتری نسبت به رگرسیون ناپارامتری خواهد داشت. نتایج این مقایسه در انتخاب روش رگرسیون مناسب تحت شرایط معین جهت اهداف پیش بینی یا توصیفی، مفید خواهد بود.

یک نوع مدل رگرسیون فازی ناپارامتری بر اساس روش هموارسازی خطی موضعی توسط ونگ و همکاران (۲۰۰۷) در حالتی که ورودی‌ها اعداد دقیق و خروجی‌ها اعداد فازی هستند مطالعه شده‌است.

۱۱ رگرسیون کمترین مربعات فازی با استفاده از عملگرهای حافظ شکل[6]

مسأله‌ای که تا به اینجا در ارتباط با رگرسیون فازی به چشم می‌خورد، ابداع روشی است که در آن علاوه بر پارامترها، متغیرهای مستقل نیز فازی فرض شوند. می‌دانیم که حاصلضرب دو عدد فازی **LR**، بر اساس T− نرم‌های متداول، لزوماً یک عدد فازی **LR** نمی‌باشد و در نتیجه در این موارد برآورد پارامترها مستلزم حل مسائل بهینه‌سازی پیچیده‌ای خواهد بود. هانگ و دو (۱۹۹۷) نشان دادند که ضرب اعداد فازی بر پایه T_W (ضعیف ترین T−نرم) حافظ شکل اعداد ضرب شونده است. در حقیقت، T_W تنها T− نرمی است که بر اساس آن حاصل‌ضرب دو عدد فازی **LR**، یک عدد فازی **LR** از نوع اعداد ضرب شونده خواهد بود. هانگ و همکاران (۲۰۰۱) روشی را برای انجام رگرسیون فازی با ورودی، خروجی و ضرائب فازی بر پایه عملگرهای حافظ شکل ارائه داده‌اند. در این شیوه، اعمال حسابی بر روی اعداد فازی **LR** به‌صورت زیر تعریف می‌شود

$$A \oplus B = (a+b, max(\alpha_A, \alpha_B), max(\beta_A, \beta_B))_{LR}$$
$$A \ominus B = (a-b, max(\alpha_A, \beta_B), max(\beta_A, \alpha_B))_{LR}$$

که $A=(a,\alpha_A,\beta_A)_{LR}\ ,\ B=(b,\alpha_B,\beta_B)_{LR}$. و

$$A \otimes B = (ab, max(\alpha_A(b), \alpha_B|a|))_L$$

که $A=(a,\alpha_A)_L\ ,\ B=(b,\alpha_B)_L$ اعداد فازی متقارن هستند.

فرض کنید زوج مشاهدات $(X_i,Y_i), i=1,...,n$ که $X_i=(x_i,\zeta_i)_L$ ها و $Y_i=(Y_i,\eta_i)_L$ ها اعداد فازی متقارن می‌باشند را در اختیار داریم. مدل رگرسیونی را به‌صورت زیر درنظر می‌گیریم

$$Y = A \oplus B \otimes X$$

که در آن $A=(a,\alpha)_L$ و $B=(b,\beta)_L$.

جهت برآورد پارامترهای مدل، مسأله بهینه‌سازی کمترین مربعات زیر را حل می‌کنیم

$$min\ r(A,B) = \sum_{i=1}^{n} D(A \oplus B \otimes X_i, Y_i)^2$$

در رابطه‌ی فوق، D، متر معرفی شده توسط دیاموند (۱۹۸۸) است که به‌صورت زیر تعریف می‌شود

$$\begin{aligned} d(X,Y) = (m_x,m_y)^2 + ((m_x-\alpha_x)-(m_y-\alpha_y))^2 \\ +((m_x+\beta_x)-(m_y+\beta_y))^2 \end{aligned}$$

در این تعریف $X=(m_x,\alpha_x,\beta_x)_T$ و $Y=(m_y,\alpha_y,\beta_y)_T$ دو عدد فازی مثلثی هستند.

با استفاده از متر فوق خواهیم داشت

$$\begin{aligned} r(A,B) = f(a,b) + \textstyle\sum_{i=1}^{n}[2(max(\alpha,|b|\zeta_i,|x_i|\beta)^2 \\ -4\eta_i max(\alpha,|b|\zeta_i,|x_i|\beta)] \end{aligned} \quad (8)$$

که f تابعی فقط از a و b می‌باشد. با مشتق گیری از f(a,b) نسبت به a، برآورد a به‌صورت زیر به‌دست می آید

$$\hat{a} = \bar{y} - b\bar{x}$$

که $\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i$ و $\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i$.

برآورد پارامترهای دیگر، با حل یک مسأله برنامه‌ریزی ترکیبی درجه دو، یعنی مینیمم‌سازی رابطه (۸) حاصل می‌شوند.

۱۲ دورنمای تحقیقات آینده

در پایان، به برخی مباحث و زمینه‌هایی از رگرسیون در محیط فازی که پیش‌بینی می‌شود در آینده بیشتر مورد توجه قرار گیرند، اشاره می‌کنیم

[6] Shape Preserving Operations

— تعریف و بررسی ضریب همبستگی بین متغیرها در مدل‌های مختلف رگرسیون فازی.
— انجام مقایسه‌های دقیق‌تر بین رگرسیون فازی و رگرسیون آماری و نیز بین انواع روش‌های رگرسیون فازی، و بررسی توانایی‌های هر روش و مزیت‌ها و معایب آنها.
— تحقیقات بیشتر در رابطه با رگرسیون فازی ناپارامتری.
— تهیه نرم افزارهای مناسب جهت انجام انواع روش‌های رگرسیون فازی.

مراجع

[۱] اخلاقی، سعید؛ تقی‌زاده کاخکی، حسین؛ ارقامی، ناصررضا (۱۳۸۳)؛ رگرسیون حداقل مربعات فازی، اندیشه آماری، شماره ۹: ۶۲—۵۳.

[۲] ارقامی، ناصررضا (۱۳۸۱)؛ مروری بر رگرسیون فازی، گزارش نخستین سمینار مجموعه‌های مشکک و کاربردهای آن، دانشگاه شهید باهنر کرمان، ۱۸—۱.

[۳] طاهری، سید محمود؛ ماشین‌چی، ماشاالله (۱۳۸٦)؛ مقدمه‌ای بر آمار و احتمال فازی، انتشارات دانشگاه شهید باهنر کرمان.

[۴] ماشین‌چی، ماشاالله (۱۳۷۹)؛ رگرسیون با استفاده از پایگاه اطلاعاتی مشکک، اندیشه آماری، شماره ۵: ۱۹—۱۳.

[۵] مجدی، سعید؛ طاهری، سید محمود؛ علامت‌ساز، محمدحسین (۱۳۸۱)؛ رگرسیون خطی با ضرائب فازی، گزارش ششمین کنفرانس آمار ایران، دانشگاه تربیت مدرس، ۳۳۴—۳۱۹.

[6] Bargiela, A., Pedrycz, W., Nakashima, T. (2007), "Multiple regression with fuzzy data", Fuzzy Sets and Systems 158, 2169-2188.

[7] Celmins, A. (1987), "Least squares model fitting to fuzzy vector data", Fuzzy Sets and Systems 22, 245-269.

[8] Choi, S.H., Dong, K.H. (2004), "Note on fuzzy regression model", In: Proc. of the 7th Iranian Statistical Conference, Allameh-Tabatabaei Univ., pp. 51-55.

[9] Chuang, C. (2007), "Extended support vector interval regression networks for interval input-output data", Information Sciences 178, 871-891.

[10] Diamond, P. (1987), "Fuzzy least squares", Information Sciences 46, 141-157.

[11] Guo, P., Tanaka, H. (2006), "Dual models for possibilistic regression analysis", Comput. Stat. and Data Anal. 51, 253-266.

[12] Hong, D.H. (1997), "Fuzzy system reliability analysis by the use T_W (the weakest t-norm) on fuzzy number arithmetic operations", Fuzzy Sets and Systems 90, 307-316.

[13] Hong, D.H., Lee, S., Do, H.Y. (2001), "Fuzzy-least squares linear regression analysis using shape preserving operations", Information Sciences 138, 185-193.

[14] Jajuga, K. (1986), "Linear fuzzy regression", Fuzzy Sets and Systems 20, 343-353.

[15] Kim, K. J., Chen, H. R. (1997), "A comparison of fuzzy and non-parametric linear regression", Computers and Operational Researches Vol 24, 505-519.

[16] Kim, H.K., Yoon, J.H., Li, Y. (2007), "Asymptotic properties of least squares estimation with fuzzy observations", Information Sciences 178, 439-451.

[17] Krätschmer, V. (2006), "Least-squares estimation in linear regression models with vague concepts", Fuzzy Sets and Systems 157, 2579-2592.

[18] Krätschmer, V. (2006), "Strong consistency of least squares estimation in linear regression models with vague concepts", J. Multivariate Analysis 97, 633-654.

[19] Mahalanobis, P. (1936), "On the generalized distance in statistics", Proceedings of National Institute of Science 12, 49-55.

[20] Nasrabadi, M.M., Nasrabadi, E. (2004), "A mathematical-programing opproach to fuzzy linear regression analysis", Applied Mathematics and Computing 155, 873-881.

[21] Sakawa, M., Yano, H. (1990), "Multiobjective fuzzy linear regression analysis and its application", Fundamental Electronic Science 73, 1-10.

[22] Tanaka, H., Uejima, S., Asai, K., (1980), "Fuzzy linear regression model", International Congress on Applied Systems and Cybernetics, Vol. VI., Aeapuleo, Mexico, 2933-2938.

[23] Tanaka, H., Uejima, S., Asai, K., (1982), "Linear regression analysis with fuzzy model", IEEE Trans. Systems Man Cybernet. 12, 903-907.

[24] Tanaka, H., Hayashi, I. and Watada, J. (1989), "Poissibilistic linear regression analysis for fuzzy data" European J. of Operational Researches 40, 383-396.

[25] Tanaka, H., Ishibuchi, H. and Yoshikawa, S. (1995), "Exponential possibilistic regression analysis", Fuzzy Sets and Systems 69, 305-318.

[26] Tanaka, H. and Lee, H. (1997), "Fuzzy linear regression combining central tendency and possiblistic properties", Proc. FUZZY-IEEE$'$ 97, Barcelona, Spain, Vol. 1, pp. 63-68.

[27] Tanaka, H. and Lee, H. (1998), "Interval regression analysis by quadratic programming approach", IEEE Trans. On Fuzzy Systems, Vol. 6, No. 4, 473-482.

[28] Tanaka, H. and Lee, H. (1999), "Exponential pssibility regression analysis by identification method of possibilistic coefficients", Fuzzy Sets and Systems 106, 155-165.

[29] Wang, N., Zhang, W.X., Mei, C.l. (2007), "Fuzzy nonparametric regression based on local linear smooting technique", Information Sciences 177, 3882-3900.

[30] Xu, S., Luo, Q., Xu, G., Zhang, L. (2008), "Asymmetrical interval regression using extended ε-SVM with robust algorithm", Fuzzy Sets and Systems (to apear).

[31] Yager, R. (1982), "Fuzzy prediction based on regression models", Information Sciences 26 45-63.

[32] Yang, M. S., Ko, C. H., (1997), "On cluster-wise fuzzy regression analysis", IEEE Trans. Syst. Man Cybernet., B 27(1), 1-13.

[33] Yen, K.K., Ghoshray, G., Roig, G. (1999), " A linear regression model using triangular fuzzy number coefficients", Fuzzy Sets and Systems 106, 167-177.