استفاده از روش بوت استرپ[۱] در رگرسیون کمترین مربعات فازی

علیرضا عرب‌پور[۲] ، فرزانه مرادی[۳]

چکیده:

به منظور پیدا کردن رابطه خطی مناسب بین یک متغیر وابسته و چند متغیر مستقل در یک محیط فازی مدل‌های رگرسیون خطی فازی به کار برده می شوند. در این مقاله ابتدا پارامترهای مدل رگرسیون خطی فازی را با روش کمترین مربعات برآورد کرده ایم سپس بوت استرپ را به عنوان یک روش بازنمونه‌گیری برای بهبود برآوردهای مدل رگرسیون خطی فازی به کار برده‌ایم و در نهایت با چند مثال عددی کارایی این روش را با سایر سایر روش‌های رگرسیون کمترین مربعات فازی مورد مقایسه قرار داده‌ایم.

واژه‌های کلیدی: رگرسیون فازی، کمترین مربعات، بازنمونه‌گیری، بوت‌استرپ.

۱ مقدمه

مجموعه‌های فازی اولین بار توسط زاده[۴] [۲۴] معرفی شد. پس از آن اولین بار تاناکا[۵] و همکاران [۱۸] رگرسیون فازی را مورد بحث و بررسی قرار دادند. در این روش داده‌ها اعداد فازی مثلثی هستند و با مینیمم کردن یک شاخص فازی ضرایب رگرسیونی را تخمین زدند، این روش در سال های بعد توسط تاناکا و همکارانش بهبود یافت، در واقع اساس کار تاناکا روش‌های برنامه‌ریزی ریاضی بود. یاگر[۶] [۲۲] با رهیافتی متفاوت ساده ترین حالت رگرسیون فازی را مورد بحث قرار داد، وی با مشاهدات فازی مقدار متغیر وابسته را پیش‌بینی نمود.

چانگ و لی[۷] [۷] به برخی از ضعف های مدل پیشنهادی تاناکا اشاره کردند، به خصوص آنها نشان دادند که تحلیل رگرسیون فازی تاناکا همیشه برآوردهای مجاز نمی‌دهد. پیترز[۸] [۱۷] مدل رگرسیون خطی فازی دیگری بر مبنای روش تاناکا و با استفاده از روش برنامه‌ریزی خطی فازی معرفی کرد. کیم و بیشو[۹] [۱۴] مینیمم کردن اختلاف بین توابع عضویت اعداد فازی مشاهده شده و تخمین زده شده را ملاک کار قرار دادند.

در همان حین که بیشتر مدل‌های رگرسیون فازی با استفاده از برنامه‌ریزی ریاضی تحلیل می‌شدند جاجوگا[۱۰]

---

[۱] Bootstrap

[۲] بخش آمار، دانشگاه شهید باهنر کرمان

[۳] دانشجوی کارشناسی ارشد آمار ریاضی، بخش آمار، دانشگاه شهید باهنر کرمان

[۴] Zadeh

[۵] Tanaka

[۶] Yager

[۷] Chang and Lee

[۸] Peterz

[۹] Kim and Bishu

[۱۰] Jajuga

[۱۲] ضرایب رگرسیون خطی را با استفاده از یک نسخه تعمیم یافته از روش کمترین مربعات محاسبه کرد. سلمینس[۱۱] [۴ و ۵] روشی برای برازش دادن یک مدل فازی چند متغیره با مینیمم کردن یک تابع هدف کمترین مربعات پیشنهاد کرد و یک روش کمترین مربعات برای مدل‌های رگرسیون فازی ارائه داد. دیاموند[۱۲] [۹] یک متر بر مجموعه اعداد فازی معرفی کرد و از این متر برای تعریف معیار کمترین مربعات فازی استفاده کرد. چانگ و لی [۶] یک روش رگرسیون فازی برمبنای روش کمترین مربعات برای تخمین مقدار مراکز و پهناهای اعداد فازی $L-R$ پیشنهاد کردند. یانگ و لین[۱۳] [۲۳] یک روش کمترین مربعات برای مدل‌هایی با ورودی و خروجی فازی پیشنهاد دادند. کائو و چیو[۱۴] [۱۳] یک روش کمترین مربعات با استفاده از تابع عضویت مجموع مربعات خطا ارائه دادند. کاپی[۱۵] و همکاران [۸] یک روش کمترین مربعات برای تخمین یک مدل رگرسیون خطی با پاسخ فازی $L-R$ پیشنهاد دادند. عرب‌پور و تاتا[۱۶] [۳] با تعمیم و استفاده از فاصله‌ای که توسط دیاموند برای اعداد فازی تعریف شده است، پارامترهای مدل رگرسیون خطی فازی را برای داده های مثلثی و ذوزنقه‌ای به دست آوردند.

وانگ[۱۷] و همکاران [۲۱] روشی برای رگرسیون ناپارامتری فازی با ورودی غیرفازی و خروجی فازی $L-R$ برمبنای تکنیک هموارسازی خطی موضعی پیشنهاد دادند. در سال های اخیر برازش مدل رگرسیون فازی به سمت روش‌های عددی پیش رفته‌است. عبداله و باکلی[۱۸] [۲] روش مونت کارلو را در رگرسیون فازی به کار بردند.

از دیگر روش‌های عددی می‌توان بازنمونه‌گیری را نام برد. یکی از روش‌های بازنمونه‌گیری، بوت‌استرپ است. نمونه‌گیری بوت‌استرپ توسط افرون[۱۹] [۱۰] ارائه شد و مدت‌ها روی آن کار شد. لیانگ[۲۰] و همکاران [۱۵] بوت‌استرپ را در مدل رگرسیون خطی جزئی به کار بردند. پس از آن ولیلا[۲۱] [۲۰] از بوت‌استرپ برای تخمین توزیع مجانبی برآوردگر کمترین مربعات در مدل رگرسیون معمولی استفاده کرد. پاپارودیتیس[۲۲] و همکاران [۱۶] آزمون فرض بوت‌استرپ را در مدل‌های رگرسیون معمولی به کار بردند.

در ادامه به طور خلاصه رگرسیون کلاسیک و پس از آن رگرسیون فازی را در بخش ۲ ارائه داده و روش‌های متفاوت جهت برآورد پارامترهای مدل رگرسیون فازی را مورد بحث قرار می‌دهیم. در بخش ۳ به معرفی روش بوت‌استرپ و کاربرد آن در رگرسیون پرداختیم و پس از

---

[۱۱] Celmins
[۱۲] Diamond
[۱۳] Yang and Lin
[۱۴] Kao and Chio
[۱۵] Coppi
[۱۶] Arabpour and Tata
[۱۷] Wang
[۱۸] Abdallah and Buckley
[۱۹] Efron
[۲۰] Liang
[۲۱] Velilla
[۲۲] Paparoditis

آن استفاده از روش بوت‌استرپ در رگرسیون فازی را شرح داده‌ایم. در پایان نیز برای مقایسه کارایی این روش باسایر روش‌های رگرسیون خطی فازی به روش کمترین مربعات چند مثال عددی در بخش ۴ ارائه خواهیم داد.

۲ تحلیل رگرسیون کلاسیک و فازی

وظیفه کلی تحلیل رگرسیون به صورت شناسایی رابطه بین متغیرهای مستقل $\underline{x} = (x_۱, x_۲, ..., x_k)$ و متغیر وابسته $\underline{y}$ تعریف شده است، که k تعداد متغیرهای مستقل در مدل می‌باشد. مدل رگرسیون در این مورد به صورت زیر بیان می‌شود:

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

که $f(\underline{x})$ یک بردار از توابع $(f_۱(\underline{x}), f_۲(\underline{x}), ..., f_n(\underline{x}))^T$ و $\underline{\varepsilon} = (\varepsilon_۱, \varepsilon_۲, ..., \varepsilon_n)^T$ یک بردار از خطای تصادفی تخمین تابعی و n تعداد تکرارها است. مدل کلی فوق اغلب با فرض یک رابطه خطی بین متغیرهای مستقل و وابسته ساده‌تر می‌شود. متغیر وابسته چندگانه y به طور جداگانه n مدل رگرسیون مستقل ارائه می‌دهد. در ساده‌ترین حالت برای یک متغیر وابسته ساده مدل رگرسیون به فرم زیر داده می‌شوند:

$$y = b_۰ + b_۱ x + \varepsilon$$

که مدل رگرسیون خطی ساده نامیده می‌شود و مدل:

$$y = b_۰ + b_۱ x_۱ + ... + b_k x_k + \varepsilon$$

مدل رگرسیون خطی چندگانه نامیده می‌شود. هنگامی که پارامترهای مدل رگرسیون خطی به جای توزیع احتمالی با توابع عضویت فازی بیان شوند مدل‌های فوق مدل رگرسیون خطی فازی نامیده می‌شود. در واقع رگرسیون فازی تعمیمی از رگرسیون معمولی است.

به طور کلی سه روش برای تحلیل مدل‌های رگرسیون خطی فازی ارائه شده است: ۱− روش‌های کمترین مربعات فازی ۲− روش‌های برنامه‌ریزی ریاضی و ۳− روش‌های عددی، که در ادامه به طور مختصر به معرفی این سه روش می‌پردازیم.

۱.۲ روش‌های کمترین مربعات فازی

در این روش از مینیمم کردن مجموع مربعات خطا به عنوان یک معیار برازش استفاده می‌شود. روش‌های کمترین مربعات فازی بر اساس کمترین درجه اختلاف بین مقادیر مشاهده شده و مقادیر تخمین زده شده استوار هستند.

در این بخش به اولین و یکی از آخرین مدل‌ها بر اساس روش‌های کمترین مربعات فازی اشاره می‌کنیم.

۱.۱.۲ مدل سلمینس

این روش در سال ۱۹۸۷ ارائه شد و بر مبنای یک اندازه سازگار بین داده‌ها و مدل برازش داده شده است. اگر فرض کنیم A و B دو عدد فازی مثلثی باشند این اندازه به صورت زیر تعریف می‌شود:

$$\gamma(A, B) = \max_x \ \min[\mu_A(x), \mu_B(x)]$$

که $\mu_A(x)$ تابع عضویت A و $\mu_B(x)$ تابع عضویت B است. مقدار γ بین صفر و یک است، مقدار صفر را وقتی می‌گیرد که پهناهای دو عدد فازی هیچ همپوشانی نداشته

باشند و زمانی برابر یک است که مراکز دو کمیت فازی بر هم منطبق شوند.

اندازه مذکور در واقع اندازه سازگاری بین هر مقدار مشاهده شده شده و مقدار برآورد شده است. در این روش هدف، پیدا کردن مدلی است که سازگاری کلی بین داده‌ها و مدل برازش شده ماکسیمم شود که معادل با مینیمم سازی تابع زیر است:

$$W = \sum_{i=1}^{n} (1-\gamma_i)^2$$

به طوری که:

$$\widehat{y} = \widetilde{b_0} + \widetilde{b_1} \pm \sqrt{c_0^2 + 2c_{01}x + c_1^2 x^2}$$

بخش اول معادله فوق $(m_0 + m_1 x)$ خط مرکزی مدل رگرسیون فازی را نشان می‌دهد که ضرایب آن با استفاده از رگرسیون کمترین مربعات موزون به دست می‌آیند و از معکوس مربع ابهام هر داده به عنوان وزن آن داده استفاده می‌شود. قسمت دوم معادله $(\pm\sqrt{c_0^2 + 2c_{01}x + c_1^2 x^2})$ حدود فازی بالا و پایین مدل رگرسیون را مشخص می‌کند که c_0 و c_1 نیم‌پهناهای فازی ضرایب $\widetilde{b_0}$ و $\widetilde{b_1}$ هستند. بر طبق تعریف سلمینس مقدار c_{01} هماهنگی فازی بین ضرایب است. هماهنگی بین دو پارامتر فازی مفهوم مشابهی با کواریانس احتمالی بین دو پارامتر معمولی دارد.

۲.۱.۲ مدل عرب‌پور و تاتا

عرب‌پور و تاتا[۳] با استفاده از متر تعریف شده، توسط دایاموند برای برآورد پارامترهای مدل رگرسیون خطی فازی یک روش کمترین مربعات پیشنهاد دادند و آن را از اعداد فازی مثلثی به اعداد فازی ذوزنقه‌ای تعمیم داده‌اند که برای سه مدل رگرسیون خطی فازی با مشاهدات فازی مثلثی به شرح زیر می‌باشد:

۱ــ مدل رگرسیون خطی فازی ساده که در آن متغیر مستقل اعداد فازی مثلثی و پارامترها اعداد غیرفازی می‌باشند، به فرم زیر است:

$$\widetilde{y_i} = b_0 + b_1 \widetilde{x_i} + \widetilde{\varepsilon_i} \qquad (1)$$

که مشاهدات به صورت اعداد فازی مثلثی زیر می‌باشند:

$$\widetilde{y_i} = (y_{il}, y_i, y_{ir}) \quad , \quad \widetilde{x_i} = (x_{il}, x_i, x_{ir})$$

در این روش برای مدل فوق پارامترها به صورت زیر برآورد می‌شوند:

$$S(b_0, b_1) = \sum_{i=1}^{n} d^2(\widetilde{y_i}, b_0 + b_1 \widetilde{x_i})$$

$$= \sum_{i=1}^{n} [(y_i - b_0 - b_1 x_i)^2 + (y_{il} - b_0 - b_1 x_{il})^2 + (y_{ir} - b_0 - b_1 x_{ir})^2]$$

با مشتق‌گیری نسبت به پارامترها داریم:

$$\hat{b}_1 = \frac{\sum_{i=1}^{n} (x_{il}y_{il} + x_i y_i + x_{ir} y_{ir} - 3n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} (x_{il}^2 + x_i^2 + x_{ir}^2) - 3n\overline{x}^2}$$

$$\hat{b}_0 = \overline{y} - \hat{b}_1 \overline{x}$$

که در آن:

$$\overline{y} = \frac{\sum_{i=1}^{n} (y_{il} + y_i + y_{ir})}{3n}, \overline{x} = \frac{\sum_{i=1}^{n} (x_{il} + x_i + x_{ir})}{3n}$$

۲ــ مدلی که پارامترها اعداد فازی مثلثی و متغیرهای مستقل اعداد غیرفازی هستند:

$$\widetilde{y_i} = \widetilde{b_0} + \widetilde{b_1} x_i + \widetilde{\varepsilon_i} \qquad (2)$$

برای چنین مدلی خواهیم داشت:

$$S(b_\circ, b_۱) = \sum_{i=۱}^{n} d^۲(\widetilde{y}_i, \widetilde{b_\circ} + \widetilde{b_۱} x_i)$$

$$= \sum_{i=۱}^{n} [(y_i - b_\circ - b_۱ x_i)^۲ + (y_{il} - b_{\circ l} - b_{۱l} x_i)^۲ + (y_{ir} - b_{\circ r} - b_{۱r} x_i)^۲]$$

که با مشتق‌گیری نسبت به پارامترها به دست می آوریم:

$$\widehat{b}_{۱r} = \frac{\sum_{i=۱}^{n} x_i y_{ir} - n\overline{x}\,\overline{y}_r}{\sum_{i=۱}^{n} x_i^۲ - n\overline{x}^۲}\ ,\ \widehat{b}_{\circ r} = \overline{y}_r - \widehat{b}_{۱r}\overline{x} \quad (۳)$$

$$\widehat{b}_{۱l} = \frac{\sum_{i=۱}^{n} x_i y_{il} - n\overline{x}\,\overline{y}_l}{\sum_{i=۱}^{n} x_i^۲ - n\overline{x}^۲}\ ,\ \widehat{b}_{\circ l} = \overline{y}_l - \widehat{b}_{۱l}\overline{x}$$

$$\widehat{b}_۱ = \frac{\sum_{i=۱}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=۱}^{n} x_i^۲ - n\overline{x}^۲}\ ,\ \widehat{b}_\circ = \overline{y} - \widehat{b}_۱\overline{x}$$

که در آن:

$$\widehat{\widetilde{b}}_۱ = (\widehat{b}_{۱l}, \widehat{b}_۱, \widehat{b}_{۱r}) \quad , \quad \widehat{\widetilde{b}}_\circ = (\widehat{b}_{\circ l}, \widehat{b}_\circ, \widehat{b}_{\circ r})$$

و

$$\overline{y}_l = \frac{\sum_{i=۱}^{n} y_{il}}{n} \quad \overline{y} = \frac{\sum_{i=۱}^{n} y_i}{n} \quad \overline{y}_r = \frac{\sum_{i=۱}^{n} y_{ir}}{n}$$

۳– مدلی که هم پارامترها و هم متغیرهای مستقل اعداد فازی مثلثی هستند، داریم:

$$\widetilde{y}_i = \widetilde{b}_\circ + \widetilde{b}_۱ \widetilde{x}_i + \widetilde{\varepsilon}_i \qquad (۴)$$

و برای این مدل خواهیم داشت:

$$S(b_\circ, b_۱) = \sum_{i=۱}^{n} d^۲(\widetilde{y}_i, \widetilde{b}_\circ + \widetilde{b}_۱ \widetilde{x}_i)$$

$$= \sum_{i=۱}^{n} [(y_i - b_\circ - b_۱ x_i)^۲ + (y_{il} - b_{\circ l} - b_{۱l} x_{il})^۲ + (y_{ir} - b_{\circ r} - b_{۱r} x_{ir})^۲]$$

در این حالت نیز با مشتق‌گیری، برآورد پارامترها به صورت زیر به دست می آید:

$$\widehat{b}_{۱r} = \frac{\sum_{i=۱}^{n} x_{ir} y_{ir} - n\overline{x}_r\overline{y}_r}{\sum_{i=۱}^{n} x_{ir}^۲ - n\overline{x}_r^۲}\ ,\widehat{b}_{\circ r} = \overline{y}_r - \widehat{b}_{۱r}\overline{x}_r \quad (۵)$$

$$\widehat{b}_{۱l} = \frac{\sum_{i=۱}^{n} x_{il} y_{il} - n\overline{x}_l\overline{y}_l}{\sum_{i=۱}^{n} x_{il}^۲ - n\overline{x}_l^۲}\ ,\widehat{b}_{\circ l} = \overline{y}_l - \widehat{b}_{۱l}\overline{x}_l$$

$$\widehat{b}_۱ = \frac{\sum_{i=۱}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=۱}^{n} x_i^۲ - n\overline{x}^۲}\ ,\widehat{b}_\circ = \overline{y} - \widehat{b}_۱\overline{x}$$

که پارامترها مانند مدل قبل تعریف می‌شوند و:

$$\overline{y}_l = \frac{\sum_{i=۱}^{n} y_{il}}{n} \quad \overline{y} = \frac{\sum_{i=۱}^{n} y_i}{n} \quad \overline{y}_r = \frac{\sum_{i=۱}^{n} y_{ir}}{n}$$

$$\overline{x}_l = \frac{\sum_{i=۱}^{n} x_{il}}{n} \quad \overline{x} = \frac{\sum_{i=۱}^{n} x_i}{n} \quad \overline{x}_r = \frac{\sum_{i=۱}^{n} x_{ir}}{n}$$

۲.۲ روش‌های برنامه‌ریزی ریاضی

در این روش‌ها از مینیمم کردن ابهام مدل به عنوان معیار برازش استفاده می‌شود. مزیت روش‌های برنامه‌ریزی ریاضی سادگی آنها در برنامه‌ریزی و محاسبات است و عیب عمده این روش‌ها این است که مفهوم کمترین مربعات در نظر گرفته نشده است و در نتیجه پهنای مقادیر تخمین زده شده بسیار عریض است و همچنین با اضافه شدن یک مشاهده به داده‌ها تعداد قیدهای مساله دو یا چند برابر می‌شود.

در این بخش به اولین و یکی از آخرین مدل‌ها بر اساس روش‌های برنامه‌ریزی ریاضی اشاره می‌کنیم.

۱.۲.۲ مدل تاناکا و واتادا

تاناکا و واتادا [۱۹] مدل رگرسیون خطی فازی با خروجی فازی، ورودی غیرفازی و پارامترهای فازی را در نظر گرفتند:

$$\widetilde{y}_i = \widetilde{b}_\circ + \widetilde{b}_۱ x_{i۱} + ... + \widetilde{b}_k x_{ik} + \varepsilon_i \quad , \quad i = ۱, ۲, ..., n$$

که در آن پارامترها و مقدار متغیر مستقل در مشاهده iـام اعداد فازی مثلثی متقارن هستند که به ترتیب با $\widetilde{b}_j = (a_j, \alpha_j)\ ,\ j = ۰, ..., k$ و $\widetilde{y}_i = (y_i, e_i)\ ,\ i = ۱, ..., n$ نشان داده می‌شوند.
مدل فوق را به صورت زیر می‌توان بازنویسی کرد:

$$\widetilde{y}_i = (\underline{a}\ \underline{x}_i, \underline{\alpha}|\underline{x}_i|)$$

که در آن:

$$\underline{a} = (a_۰, a_۱, ..., a_k)^T\ ,\ \underline{\alpha} = (\alpha_۰, \alpha_۱, ..., \alpha_k)^T$$
$$\underline{x}_i = (۱, x_{i۱}, ..., x_{ik})^T \quad , \quad i = ۱, ..., n$$

تاناکا و واتادا [۱۹] پارامترهای فازی را به قسمی تعیین کردند که ابهام مدل که به صورت زیر تعریف می‌شود:

$$J(\underline{\alpha}) = \sum_{i=۱}^{n} \underline{\alpha}|\underline{x}_i|$$

مینیمم شود، به طوری که به ازای هر مقدار معین h، hـبرش مقادیر تخمین زده شده hـبرش مقادیر مشاهده شده را بپوشاند، یعنی $[\widetilde{y}_i]^h \subseteq [\widehat{\widetilde{y}}_i]^h$ که در آن :

$$[\widetilde{y}_i]^h = [y_i - l^{-۱}(h)e_i, y_i + l^{-۱}(h)e_i]$$

و

$$[\widehat{\widetilde{y}}_i]^h = [\underline{a}\ \underline{x}_i - l^{-۱}(h)\underline{\alpha}|\underline{x}_i|, \underline{a}\ \underline{x}_i + l^{-۱}(h)\underline{\alpha}|\underline{x}_i|]$$

تحلیل بالا منجر به مساله برنامه‌ریزی خطی زیر می‌شود:

$$\min J(\alpha) = \sum_{i=۱}^{n} \underline{\alpha}|\underline{x}_i|$$

به‌گونه‌ای که

$$\underline{a}\ \underline{x}_i - l^{-۱}(h)\underline{\alpha}|\underline{x}_i| \leq y_i - l^{-۱}(h)e_i\ , i = ۱, ..., n$$
$$-\underline{a}\ \underline{x}_i - l^{-۱}(h)\underline{\alpha}|\underline{x}_i| \leq -y_i - l^{-۱}(h)e_i, i = ۱, ..., n$$
$$\alpha_j \geq ۰\ ,\ a_j \in \Re\ ,\ j = ۰, ۱, ..., k\ ,\ ۰ \leq h \leq ۱$$

که مقدار h توسط تصمیم‌گیرنده مشخص می‌شود.

۲.۲.۲ مدل حجتی و همکاران

حجتی و همکاران [۱۱] روشی برای تخمین ضرایب رگرسیون ارائه دادند به طوری که مجموع انحراف نقاط بالایی hـبرش‌های مقادیر پیش‌بینی شده و مقادیر مشاهده شده و انحراف نقاط پایینی این hـبرش‌ها مینیمم شود. در این مدل فرض شده‌است که مقادیر مشاهده شده و ضرایب رگرسیون اعداد فازی مثلثی متقارن هستند. این مدل به صورت مساله برنامه‌ریزی زیر ارائه شده‌است:

$$\min \sum_{i=۱}^{n}(d_{iU}^{+} + d_{iU}^{-} + d_{iL}^{+} + d_{iL}^{-})$$

به‌گونه‌ای که

$$\sum_{j=۰}^{k}(a_j + (۱-h)\alpha_j)x_{ij} + d_{iU}^{+} - d_{iU}^{-} = y_i + (۱-h)e_i \ ,\ i = ۱, ..., n$$
$$\sum_{j=۰}^{k}(a_j - (۱-h)\alpha_j)x_{ij} - d_{iL}^{+} - d_{iL}^{-} = y_i - (۱-h)e_i \ ,\ i = ۱, ..., n$$
$$d_{iU}^{+}, d_{iU}^{-}, d_{iL}^{+}, d_{iL}^{-} \geq ۰ \ ,\ i = ۱, ..., n$$
$$\alpha_j \geq ۰\ , a_j \in \Re\ , j = ۰, ..., k\ , ۰ \leq h \leq ۱$$

قابل ذکر است که $|d_{iU}^{+} - d_{iU}^{-}|$ فاصله بین نقطه بالایی hـبرش مقدار پیش‌بینی شده و نقطه بالایی hـبرش مقدار مشاهده شده است و $|d_{iL}^{+} - d_{iL}^{-}|$ فاصله بین نقطه پایینی hـبرش مقدار پیش‌بینی شده و نقطه پایینی hـبرش مقدار مشاهده‌شده است. تابع هدف در حقیقت مینیمم سازی مجموع این دو فاصله است.

۳.۲ روش‌های عددی

از جمله روش‌های عددی که در رگرسیون فازی مورد استفاده گرفته است، می‌توان از روش مونت‌کارلو نام برد. عبداله و باکلی [۲] روش مونت‌کارلو را در رگرسیون خطی فازی به کار بردند. روش کار آن‌ها به این صورت است که ابتدا دو خطا روی مقادیر مشاهده شده و برآورد شده تعریف کردند سپس دنباله تصادفی از بردارهای غیرفازی برای ضرایب مدل تولید کردند و آن برداری را به عنوان راه حل مساله رگرسیون فازی ارائه دادند که خطاهای نامبرده را مینیمم کند.

۳ مدل رگرسیون فازی به روش بوت‌استرپ

واژه بوت‌استرپ از ماجرای «بارن مانچوسن[۲۳]» نوشته رادف اریک راسپه[۲۴] در اواخر قرن ۱۸ گرفته شده است که وقتی خودش را در ته دریاچه بدون هیچ کمک و وسیله‌ای می‌بیند با استفاده از بند پوتینش خود را بالا می‌کشد [۱]. در اینجا نیز موقعی که کمترین اطلاعات را داریم یعنی نه توزیع جامعه معلوم است نه اینکه داده‌ها به اندازه کافی هستند، سعی می‌کنیم با استفاده از نمونه کوچکی که در دست داریم استنباط خوب و معتبر ارائه دهیم. در این روش، آنالیز روی داده‌های واقعی است؛ نمونه‌های کوچک ولی معتبر و موثق قسمت مهم یک تحقیق است و در عمل بر پایه شبیه‌سازی است، اما در برخی مثال‌ها نیازی به شبیه‌سازی نیست. در واقع شبیه‌سازی از توزیع جامعه می‌باشد ولی اگر توزیع جامعه نامعلوم باشد به تعداد دلخواه نمونه از نمونه اصلی برداشته و میانگین آن‌ها را در نظر می‌گیریم، پس نیازی به توزیع جامعه نیست.

در رگرسیون به دو صورت می‌توانیم از بوت‌استرپ استفاده کنیم یکی اینکه از مشاهدات نمونه بوت‌استرپ بگیریم و دیگر اینکه از باقیمانده‌های مدل نمونه بوت‌استرپ انتخاب کنیم.

۱− بوت‌استرپ از مشاهدات: در این روش از نمونه‌ای که در دست داریم، نمونه باجایگذاری برمی‌داریم سپس به این نمونه‌ها مدل برازش می‌دهیم، سپس میانگین برآوردهای این مدل‌ها را به عنوان برآورد اصلی انتخاب می‌کنیم.

۲− بوت‌استرپ از باقیمانده‌ها: در این روش با نمونه‌ای که در دست داریم $((x_i,y_i)\ ,\ i=۱,۲,...,n)$ مدل را برازش می‌دهیم و باقیمانده‌های مدل را از رابطه $\varepsilon_i = y_i - \widehat{b}_۰ - \widehat{b}_۱ x_i\ ,\ i=۱,...,n$ به دست می آوریم، سپس از این باقیمانده‌ها نمونه باجایگذاری برمی‌داریم به طوری که احتمال انتخاب هر $\varepsilon_i\ ,\ i=۱,...,n$ در هر انتخاب $\frac{۱}{n}$ می‌باشد. فرض کنید این نمونه بوت‌استرپ $(\varepsilon_۱^*,\varepsilon_۲^*,...,\varepsilon_n^*)$ باشد. با اضافه کردن این باقیمانده‌های جدید، مقادیر پاسخ جدید را از رابطه $y_i^* = \widehat{b}_۰ + \widehat{b}_۱ x_i + \varepsilon_i^*\ ,\ i=۱,...,n$ به دست می آوریم. این مقادیر پاسخ جدید و متغیرهای مستقل، نمونه جدید $((x_i,y_i^*)\ ,\ i=۱,۲,...,n)$ را تشکیل می‌دهد و با این داده‌های جدید مدل برازش می‌دهیم. برآوردهای حاصل

[۲۳] Baron Manchosen
[۲۴] Radolf Erick Raspe

را $\widehat{b}^*_{\circ}$ و $\widehat{b}^*_{\backslash}$ می‌نامیم. این عمل را چندین بار تکرار می‌کنیم سپس میانگین برآوردها را به عنوان برآورد بوت‌استرپ ارائه می‌دهیم.

این روش بهتر از روش اول است زیرا حتی اگر باقیمانده‌ها نرمال نباشند می‌توان استنباط دقیق‌تری در مورد پارامترهای مدل انجام داد.

در این مقاله پس از انجام هر دو روش برای انتخاب نمونه بوت‌استرپ، به دلیل اینکه نمونه بوت‌استرپ از باقیمانده‌ها خطای کمتری داشت از روش دوم استفاده شد.

ابتدا پارامترهای مدل رگرسیون فازی را در هر یک از سه مدل با روابط (۱)، (۲) و (۴) با استفاده از روش عربپور—تاتا[۳] برآورد کرده‌ایم سپس در هر یک از این سه مدل روش بوت‌استرپ را به کار برده‌ایم به این صورت که در هر مدل B بار از باقیمانده‌های مدل نمونه‌گیری باجایگذاری انجام داده‌ایم و به روشی که توضیح داده شد پاسخ جدید تولید کردیم و با این داده‌های جدید به همان روش کمترین مربعات، مدل برازش دادیم و در هر مرحله برآورد پارامترها $(\widehat{b}^*_{\circ i},\widehat{b}^*_{\backslash i})\ ,\ i=\backslash,...,B$ را به دست آوردیم. در پایان B مرحله، میانگین این برآوردها را به عنوان برآورد پارامترهای مدل اصلی در نظر گرفته‌ایم.

۴ مثال‌ها

در این بخش با بیان چند مثال عددی نشان خواهیم داد در مدلی که مشاهدات اعداد مثلثی فازی هستند با اجرای بوت‌استرپ به شیوه‌ای که گفته شد برآوردهای بهتر با خطای کمتری به دست می‌آید.

مثال ۴—۱ کیم—بیشو [۱۴]، کائو—چیو [۱۳] و حجتی و همکاران [۱۱] از داده‌های جدول ۱ برای نشان دادن مدل‌های خود در مساله رگرسیون فازی با ورودی غیرفازی و خروجی فازی استفاده کردند. در این مثال ۵ جفت داده به صورتی که در جدول ۱ آمده است وجود دارد. در واقع این مدل از رابطه (۲) پیروی می‌کند که با استفاده از معادلات رابطه (۳) برآورد اولیه پارامترها به دست می‌آید. مدل رگرسیون خطی فازی که توسط کائو—چیو [۱۳] برازش شده به صورت زیر است:

$$\widetilde{y}_{KC}=(۱/۹۴, ۴/۹۵, ٦/۷۵)+۱/۷۱x$$

عربپور—تاتا [۳] با به کار بردن روش کمترین مربعات، مدل رگرسیون فازی را برای این داده‌ها به صورت زیر ارائه داده‌اند:

$$\begin{aligned}\widetilde{y}_{AT}= \quad & (۳/۱۱, ۴/۹۵, ٦/۷۹)+\\ & (۱/۵۵, ۱/۷۱, ۱/۸۷)x\end{aligned}$$

پس از اجرای روش بوت‌استرپ روی مدل بالا و استفاده از روش عرب‌پور—تاتا جهت برآورد پارامترها، خواهیم داشت:

$$\begin{aligned}\widetilde{y}_{Boot}= \quad & (۳/\circ ٦۹, ۴/۸۷۵, ٦/۷۲۹)\\ & +(۱/۵۵۴, ۱/۷\circ ٦, ۱/۸۷۷)x\end{aligned}$$

همان‌طور که در جدول ۲ مشاهده می‌کنید، مجموع خطاهای محاسبه شده برای این داده‌ها در مدل رگرسیون فازی با به کارگیری روش بوت‌استرپ مقدار ۱۰/۰۱۳۹ است که از خطای مدل‌های دیگر به طور قابل توجهی کمتر است.

جدول ۱. داده‌های مثال ۱.

$\widetilde{y}_i$	x_i
(۶/۲ ،۸/۰ ،۹/۸)	۱
(۴/۲ ،۶/۴ ،۸/۶)	۲
(۶/۹ ،۹/۵ ،۱۲/۱)	۳
(۱۰/۹ ،۱۳/۵ ،۱۶/۱)	۴
(۱۰/۶ ،۱۳/۰ ،۱۵/۴)	۵

جدول ۲. خطای برآورد شده برای داده های جدول ۱.

Boot	*AT*	*KC*	*Obs.*
۲/۲۷۸۱	۵/۴۶۶۸	۱۰/۰۹۳۷	۱
۲/۹۹۱۰	۱۱/۶۴۵۹	۷/۶۹۱۴	۲
۰/۹۸۸۴	۱/۱۶۶۰	۰/۴۱۳۷	۳
۲/۹۰۸۰	۸/۸۰۱۱	۱۳/۷۱۶۸	۴
۰/۸۴۸۴	۰/۸۶۵۲	۰/۲۷۲۱	۵
۱۰/۰۱۳۹	۲۷/۹۴۵۰	۳۲/۱۸۹۵	جمع

نکته ۱

روش بوت‌استرپ را روی برش صفر خط رگرسیون فازی به کار می‌بریم و آنگاه میانگین ضرایب بدست آمده را به عنوان برآورد اصلی در نظر می‌گیریم، به عبارتی بهترین برازش را در برش صفر خط رگرسیون به‌دست می آوریم یا در بدترین وضعیت ممکنه بهترین براازش را انجام می‌دهیم، به همین دلیل است که مدل بدست آمده در چنین حالتی دارای کمترین خطا می باشد.

نکته ۲

با توجه به اینکه مدل AT خطای کمتری نسبت به مدل های قبلی دارد [۳] بنابراین با اجرای روش بوت‌استرپ روی این مدل، مجموع خطاها از روش‌های قبلی نیز کمتر می‌شود.

مثال ۴—۲ برای مقایسه مدل ارائه شده با مدل کائو—چیو [۱۳] و مدل عربپور—تاتا [۳] از داده‌های جدول ۳ استفاده کردیم که شامل ۹ مشاهده با ورودی غیرفازی و خروجی فازی است. این داده‌ها نیز از مدلی با رابطه (۲) پیروی می‌کنند. مدل رگرسیون ارائه شده توسط کائو—چیو [۱۳] برای این داده‌ها به صورت زیر است:

$$\widetilde{y}_{KC} = -۴/۸۹۴۸۷ + ۱/۱۹۵۴۲x + (-۲/۹۹, ۰, ۱/۸۰)$$

مدل به دست آمده از روش عربپور—تاتا [۳] به فرم زیر است:

$$\widetilde{y}_{AT} = (-۵/۲۴۹, -۴/۸۹۵, -۴/۵۴۱) + (۱/۱۳۸, ۱/۱۹۵, ۱/۲۵۲)x$$

مدل برازش شده پس از به‌کارگیری بوت‌استرپ و استفاده از روش عربپور—تاتا [۳] به فرم زیر است:

$$\widetilde{y}_{Boot} = (-۵/۱۹۵, -۴/۸۹۴, -۴/۵۵۴) + (۱/۱۳۶, ۱/۱۹۳, ۱/۲۴۵)x$$

همان‌طور که در جدول ۴ مشاهده می‌کنید مجموع خطاهای محاسبه شده برای این داده‌ها با اجرای بوت‌استرپ مقدار۱۰/۰۹۵۵ است که از مجموع خطاهای مدل‌های دیگر کمتر است.

جدول ۳. داده‌های مثال ۴—۲.

$\widetilde{y}_i$	x_i
(−۲/۱، −۱/۶، −۱/۱)	۱
(−۲/۳، −۱/۸، −۱/۳)	۳
(−۱/۵، −۱، −۰/۵)	۴
(۰/۷، ۱/۲، ۱/۷)	۵/۶
(۱/۲، ۲/۲، ۳/۲)	۷/۸
(۵/۸، ۶/۸، ۷/۸)	۱۰/۲
(۹، ۱۰، ۱۱)	۱۱
(۹، ۱۰، ۱۱)	۱۱/۵
(۹، ۱۰، ۱۱)	۱۲/۷

جدول ۴. خطای برآورد شده برای داده های جدول ۳.

$Boot$	AT	KC	$Obs.$
۰/۸۷۴۹	۰/۹۱۱۳	۲/۸۸۷۵	۱
۰/۷۴۰۰	۰/۷۴۷۳	۱/۹۳۲۴	۲
۱/۰۱۵۶	۱/۰۴۷۱	۱/۹۹۴۳	۳
۰/۸۷۱۹	۰/۸۹۲۸	۱/۹۴۵۸	۴
۱/۷۴۳۴	۱/۷۹۹۳	۲/۶۱۴۳	۵
۰/۸۲۵۸	۰/۸۶۸۵	۱/۴۶۶۶	۶
۱/۹۰۶۷	۱/۹۵۳۷	۳/۰۰۷۷	۷
۱/۶۴۴۰	۱/۶۴۰۱	۲/۴۲۹۸	۸
۰/۴۷۳۲	۰/۵۳۴۲	۱/۴۲۵۴	۹
۱۰/۰۹۵۵	۱۰/۳۹۴۴	۱۹/۷۰۳۹	جمع

مثال ۴—۳ در این مثال برای نشان دادن تاثیر بوت‌استرپ روی برآوردهای مدل رگرسیون، این روش را روی داده‌هایی که توسط یانگ و لین مورد استفاده قرار گرفت به کار می‌بریم. در این مثال پارامترها و ورودی مدل اعداد فازی مثلثی هستند یعنی مدل به صورت رابطه (۴) است و برآوردهای اولیه از معادلات رابطه (۵) به دست می‌آیند. مدل رگرسیون ارائه شده برای این داده‌ها توسط یانگ—لین [۲۳] به صورت زیر است:

$$\widetilde{y}_{YL} = (۳/۲۰۵۲, ۳/۴۹۶۷, ۳/۷۸۸۲)$$
$$+(۰/۵۲۵۱, ۰/۵۲۹۳, ۰/۵۳۳۵)\widetilde{x}$$

برای ارزیابی مدل پیشنهادی، آن را با مدل رگرسیون فازی عربپور—تاتا [۳] که به صورت زیر است، مقایسه می‌کنیم:

$$\widetilde{y}_{AT} = (۳/۲۷۵۸, ۳/۵۷۲۴, ۳/۸۳۸۶)$$
$$+(۰/۵۱۸۸, ۰/۵۱۹۳, ۰/۵۲۳۵)\widetilde{x}$$

مدل برازش داده شده به روش بوت‌استرپ و استفاده از روش عربپور—تاتا[۳] جهت برآورد پارامترها به این صورت است:

$$\widetilde{y}_{Boot} = (۳/۲۸۸۷, ۳/۵۶۲۸, ۳/۸۱۶۹)$$
$$+(۰/۵۱۸۴, ۰/۵۲۰۷, ۰/۵۲۳۲)\widetilde{x}$$

خطاهای این مدل را به دست آوردیم و با مدل یانگ—لین [۲۳] و عربپور—تاتا[۳] مقایسه کردیم. داده‌ها و نتایج در جدول ۶ آمده است.

جدول ۵. داده‌های مثال ۴—۳.

$\widetilde{y}_i$	$\widetilde{x}_i$
(۳/۵، ۴، ۴/۵)	(۱/۵، ۲، ۲/۵)
(۵، ۵/۵، ۶)	(۳، ۳/۵، ۴)
(۶/۵، ۷/۵، ۸/۵)	(۴/۵، ۵/۵، ۶/۵)
(۶، ۶/۵، ۷)	(۶/۵، ۷، ۷/۵)
(۸، ۸/۵، ۹)	(۸، ۸/۵، ۹)
(۷، ۸، ۹)	(۹/۵، ۱۰/۵، ۱۱/۵)
(۱۰، ۱۰/۵، ۱۱)	(۱۰/۵، ۱۱، ۱۱/۵)
(۹، ۹/۵، ۱۰)	(۱۲، ۱۲/۵، ۱۳)

جدول ۶. خطای برآورد شده برای داده های جدول ۵.

$Boot$	AT	YL	$Obs.$
۰/۸۴۳۲۸	۰/۸۵۸۵۶	۰/۸۲۲۴۷	۱
۰/۲۲۷۱۵	۰/۲۱۵۸۰	۰/۲۷۸۲۲	۲
۱/۵۰۳۹۰	۱/۵۱۲۲۱	۱/۵۳۴۶۸	۳
۰/۹۱۳۱۰	۰/۹۴۱۶۴	۰/۹۵۱۱۱	۴
۰/۷۶۹۸۲	۰/۷۷۷۲۴	۰/۷۷۳۹۷	۵
۱/۴۴۹۸۵	۱/۴۷۹۸۳	۱/۵۱۵۴۶	۶
۰/۹۹۸۲۶	۱/۰۶۷۹۷	۱/۱۰۲۳۵	۷
۰/۸۰۸۰۲	۰/۸۳۷۶۵	۰/۸۸۸۴۷	۸
۷/۵۱۳۳۸	۷/۶۹۰۹۰	۷/۸۶۶۷۳	جمع

در این مدل نیز مجموع خطاهای برازش شده با به کارگیری بوت‌استرپ مقدار ۷/۵۱۳۳۸ است، که از مجموع خطاهای دو مدل دیگر کمتر است.

۵ نتیجه‌گیری

برای برآورد پارامترها در مدل رگرسیون خطی فازی روش‌های متعددی ارائه شده است. از این روش‌ها می‌توان روش کمترین مربعات را نام برد که با مینیمم کردن مربعات خطا (تفاضل بین مقادیر مشاهده شده و برآورد شده) مدل مناسبی برازش می‌دهد. در این مقاله برای بهبود برآوردهای به دست آمده از روش

کمترین مربعات در مدل رگرسیون خطی فازی از روش بوت‌استرپ استفاده شده است. با اجرای این روش روی چند مثال عددی تاثیر بوت‌استرپ روی کاهش خطاهای مدل به وضوح مشاهده شد. اجرای این روش با استفاده از کامپیوتر به آسانی امکان‌پذیر است و می‌توان مدل مناسب داده‌ها را پیدا کرد.

مراجع

[۱] دوست‌طلب، ص. (۱۳۸۵)، شبیه‌سازی و بازنمونه‌گیری، نشریه دانشجویی آمار، ندا، سال چهارم، شماره اول، ۳۳—۲۱.

[2] Abdalla, A. and Buckley, J.J. (2008). Monte Carlo methods in fuzzy linear regression II, *Soft Comput*, 12, 463-468.

[3] Arabpour, A. R. and Tata, M. (2008). Estimating the parameters of a fuzzy linear regression model, *Iranian Journal of Fuzzy Systems*, 5, 2, 1-19.

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

[5] Celmins, A. (1987). Multidimensional least-squares fitting of fuzzy models, *Mathematical Modelling*, 9, 669-690.

[6] Chang, P.T. and Lee, E.S. (1996). A generalied fuzzy weighted least squares regression, *Fuzzy Sets and Systems*, 82, 289-298.

[7] Chang, P.T. and Lee, E.S. (1994). Fuzzy linear regression with spreads unrestricted in sign, *Computers and Mathematics with Applications*, 28, 61-70.

[8] Coppi, R., D'Urso P., Giordani P. and Santoro, A. (2006). Least squares estimation of a linear regression model with L-R fuzzy response, *Computational Statistics and Data Analysis*, 51, 267-286.

[9] Diamond, P. (1988). Fuzzy least squares, *Infomation Sciences*, 46, 141-157.

[10] Efron, B. (1979). Bootstrap methods: another look at the Jacknife, *Analysis Statistics*, 101-118.

[11] Hojati, M., Bector, C.R. and Smimou, K. (2005). A simple method for computation of fuzzy linear regression, *European Journal of Operational Research*, 166, 172-184.

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

[13] Kao, C. and Chyu, C. (2002). A fuzzy linear regression model with better explanatory power, *Fuzzy Sets and Systems*, 126, 401-409.

[14] Kim, B. and Bishu, R.R. (1998). Evaluation of fuzzy linear regression models by comparing membership functions, *Fuzzy Sets and Systems*, 100, 343-353.

[15] Liang, H., Hardle, W. and Sommerfeld, V. (2000). Bootstrap approximation in a partially linear regression model, *Journal of Statistical Planning and Inference*, 91, 413-426.

[16] Paparoditis, E. and Politis, D.N. (2005). Bootstrap hypothesis testing in regression models, *Statistics and Probability Letters*, 74, 356-365.

[17] Peters, G. (1994), Fuzzy linear regression with fuzzy intervals, *Fuzzy Sets and Systems*, 63, 45-55.

[18] Tanaka, H., Uejima, S. and Asai, K. (1982). Linear regression analysis with fuzzy model, *IEEE Transactions on Systems Man Cybernet*, 12(6), 903-907.

[19] Tanaka, H. and Watada, J. (1988). Possibilistic systems and their application to the linear regression model, *Fuzzy Sets and Systems*, 27, 275-289.

[20] Velilla, S. (2001). On the bootstrap in misspecified regression models, *Computational Statistics and Data Analysis*, 36, 227-242.

[21] Wang, N., Zhang, W.X. and Mei, C.L. (2007). Fuzzy nonparametric regression based on local linear smoothing technique, *Information Sciences*, 177, 3882-3900.

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

[23] Yang, M. and Lin, T. (2001). Fuzzy least-squares linear regresion analysis for fuzzy input-output data, *Fuzzy Sets and Systems*, 126, 389-399.

[24] Zadeh, L.A. (1965). Fuzzy sets, *Information Control*, 8, 338-353.