



مدل بندی سری های زمانی ناپارامتری بر اساس داده های فازی

فائزه ترکیان^۱ مسعود یارمحمدی^{۱*}، غلامرضا حسامیان^۱ محمدقاسم اکبری^۲

(۱) گروه آمار دانشگاه پیام نور، صندوق پستی ۱۹۳۹۵-۴۶۹۷ تهران، ایران
(۲) گروه آمار، دانشکده علوم ریاضی، دانشگاه بیرجند، خراسان جنوبی، ایران

دبیر مسئول: رحیم چینی پرداز

تاریخ پذیرش: ۱۴۰۰/۵/۱۲

تاریخ دریافت: ۱۴۰۰/۱/۱۴

چکیده: در این مقاله، یک مدل سری زمانی ناپارامتری بر اساس مشاهدات فازی ارائه شده است. با استفاده از تعمیم روش نادارایا-واتسون در محیط فازی، برآورد مقادیر فازی پیش بینی شده بدست می آید. در این راستا الگوریتمی جهت دستیابی به مرتبه اتورگرسیو و پهنای باند بهینه بیان شده و سپس معیارهایی برای بررسی ارزیابی این پیش بینی معرفی می شود. در ادامه با استفاده از داده های واقعی کارایی مدل پیشنهادی مورد بررسی و تحلیل قرار می گیرد. همچنین، تأثیر مدل سری زمانی پیشنهادی در پیش بینی با سایر مدل های سری زمانی با داده های فازی مورد مقایسه قرار می گیرد.

واژه های کلیدی: مدل سری زمانی فازی ناپارامتری، روش کرنل، پهنای باند بهینه، مرتبه اتورگرسیو.

رده بندی ریاضی: 91A30, 91B16, 46N10, 91G70

۱ مقدمه

سری زمانی مجموعه ای از مشاهدات است که در فواصل زمانی مساوی و منظمی جمع آوری شده است. طبیعت ذاتی یک سری زمانی، همبسته بودن مشاهدات است [۴۷]. مهم ترین هدف از تحلیل سری های زمانی در قلمرو زمان، پیش بینی مقادیر آینده است. این پیش بینی بر اساس بررسی توصیفی و مدل بندی رفتار سری زمانی و همچنین لحاظ کردن تغییرات مقطعی و دراز مدت عناصر سری انجام می شود. در این راستا، برای مطالعه وابستگی های موجود بین عناصر سری تحلیل توابع خودهمبستگی و خودهمبستگی جزئی مورد توجه می باشد. تجزیه و تحلیل سری زمانی به طور گسترده ای در بسیاری از زمینه های علمی از جمله مباحث اقتصاد [۳۲، ۴۲]، نجوم [۴۰]، مالی [۲۷، ۳۴] و علوم محیط زیست [۸، ۳۰] مورد استفاده قرار می گیرد. بر اساس یک دسته بندی مدل های سری زمانی به دو دسته پارامتری و ناپارامتری تقسیم می شوند [۱۲]. مدل های سری زمانی پارامتری در صورت مشخص بودن مدل، ابزارهای قدرتمندی را برای تجزیه و تحلیل داده های سری زمانی ارائه می دهند. در یک خانواده با ساختار مشخص، فرایند مدل سازی پارامتری به طور معمول شامل شناسایی مدل، برآورد پارامترها، بازرسی تشخیصی مدل پیشنهادی و در صورت مناسب بودن آن، پیش بینی است. از طرف دیگر، اگر از چارچوب ساختاری داده ها اطلاعات کمتری در

*نویسنده مسئول مقاله

رایانامه: (M. Yarmohammadi) masyar@pnu.ac.ir ، (F. Torkian) fatorkian29@gmail.com ، (M. G. Akbari) gzakbari@birjand.ac.ir ، (G. Hesamian) gh.hesamian@pnu.ac.ir

اختیار داشته باشیم، معمولاً از تکنیک‌های مختلف هموارسازی ناپارامتری برای استخراج ویژگی‌های سری زمانی استفاده می‌شود که معمولاً مدل‌های ناپارامتری دارای ساختار ساده‌تری هستند. به عبارتی مدل‌های ناپارامتری فرایند را بدون در نظر گرفتن ساختاری مشخص برای آن تخمین می‌زنند.

در تمامی مدل‌های پارامتری و ناپارامتری فرض می‌شود مشاهدات نمونه مقادیر دقیقی هستند در حالی که در بسیاری از کاربردهای واقعی با اطلاعات نادقیق روبه‌رو هستیم. بنابراین مدل‌های سری زمانی کلاسیک توانایی لازم برای مدل‌بندی و پیش‌بینی این داده‌ها را ندارند. در چنین حالتی نظریه مجموعه فازی ابزار مناسبی را برای مدل‌بندی و پیش‌بینی این داده‌ها را فراهم می‌کند. با توجه به توانایی نظریه مجموعه فازی در برخورد با ابهام و عدم دقت در مشاهدات، در سال‌های اخیر استفاده از این نظریه مورد توجه بسیار قرار گرفته است که علاقه‌مندان می‌توانند به مراجع [۲، ۲۵] و [۴۴] مراجعه نمایند.

پایه و اساس تئوری سری زمانی فازی[†] که در سال‌های اخیر مورد استفاده گسترده قرار گرفته است، توسط سونگ و چیسوم [۳۷] گذاشته شده است. آنها مدلی را براساس عدم اطمینان و دانش نادقیق موجود در داده‌های سری زمانی توسعه دادند. مدل‌های سری زمانی فازی شامل سه مرحله اصلی است. در مرحله اول داده‌های دقیق جمع‌آوری می‌شوند. در مرحله دوم فرض بر این است که پیش‌بینی مقادیر آینده مواردی مانند قیمت دلار، تعداد دانشجویان ورودی و غیره مقادیر غیردقیق و مبهم است. لذا پیش‌بینی مقادیر آینده به صورت مقادیر غیردقیق و بر اساس مجموعه‌های فازی صورت می‌گیرد. تکنیک‌های متفاوتی برای تبدیل داده‌های دقیق به داده‌های فازی وجود دارد که برای مثال می‌توان به روش‌های گروه منطق فازی و ماتریس رابطه فازی [۳، ۵، ۱۱] و [۳۳]، تکنیک‌های آماری [۱۹، ۲۹، ۴۱] و روش‌های مختلف هوش مصنوعی [۱، ۴، ۴۸] اشاره کرد. مرحله سوم غیرفازی‌سازی یعنی تبدیل مشاهدات فازی به مقادیر دقیق است [۲۰، ۳۱، ۴۹]. اما در بسیاری از موارد در سری زمانی فازی، مشاهدات اساساً نادقیق بوده و می‌بایست به صورت اعداد فازی گزارش شوند. به عنوان مثال قیمت مسکن و یا تغییرات نرخ ارز که به علت شرایط سیاسی و اقتصادی و تغییرات روزانه در واقع به صورت مشاهدات نادقیق می‌باشند. در زمینه مدل‌های سری زمانی فازی با داده‌های نادقیق تنها دو مطالعه صورت گرفته است که در زیر بیان می‌شوند.

اولین مواجهه با مدل‌های سری زمانی فازی با داده‌های نادقیق، توسط حسامیان و اکبری [۱۵] معرفی شد. در این روش یک مدل آماری بر اساس داده‌های فازی به محیط فازی تعمیم داده شد. برای این منظور یک مدل سری زمانی فازی نیمه پارامتریک بر اساس داده‌های فازی معرفی شد. آنها با استفاده از تعریف جدیدی از متغیر تصادفی فازی به برآورد پارامترهای دقیق و فازی مدل پرداختند. پس از پیش‌بینی مدل بر اساس معیارهایی بر مبنای خطای برآورد بین مقادیر مشاهده شده و برآورد شده به ارزیابی عملکرد مدل پرداختند.

زارعی و همکاران [۵۰] نیز مدل حسامیان و اکبری [۱۵] را درحالتی خاص وقتی که مشاهدات سری زمانی فازی به صورت اعداد فازی مثلثی باشند، بررسی کردند. در هر دو مطالعه انجام شده، مدل‌ها به صورت نیمه پارامتریک می‌باشند. در مدل‌های پارامتری چون باید پیش‌زمینه‌ای از مدل در نظر گرفته شود بنابراین لازم است حدس‌هایی در مورد مدل داشته باشیم که ممکن است با اصل داده‌ها سازگار نباشد. بنابراین شاید بهتر است مدل به صورت ناپارامتری در نظر گرفته شود تا بدون گذاشتن هیچ قید و شرط و پیش‌فرضی برای مدل به شناسایی ساختار موجود در داده‌ها پرداخت.

هدف اصلی این مقاله، معرفی رویکردی جدید در تعمیم مدل سری زمانی آماری بر اساس داده‌های فازی است. در این راستا، یک مدل سری زمانی ناپارامتری بر اساس مشاهدات فازی تعریف می‌شود. سپس الگوریتمی ترکیبی برای تعیین پارامترهای مدل یعنی مرتبه خودهمبستگی و پهنای باند معرفی می‌شود. میزان پیش‌بینی‌پذیری مدل در مقایسه با دیگر مدل‌های موجود با دو مثال کاربردی بر اساس چند معیار متداول ارزیابی مورد بررسی و تحلیل قرار می‌گیرد. این مقاله به صورت زیر تنظیم شده است.

در بخش ۲ مفاهیم مورد نیاز در رابطه با مجموعه‌های فازی بیان می‌شوند. مروری بر سری زمانی ناپارامتری در حالت کلاسیک در بخش ۳ ارائه می‌شود. بخش ۴ شامل معرفی مدل سری زمانی ناپارامتری بر اساس مشاهدات نادقیق و برآورد تابع هموارسازی خطی با توجه به تعمیم روش نادارایا-واتسون در محیط فازی می‌باشد. فرآیند به دست آوردن پارامترهای بهینه مدل نیز در بخش ۵ معرفی می‌شود. در بخش ۶ معیارهایی جهت بررسی و مقایسه پیش‌بینی مدل ارائه شده با روش‌های موجود معرفی می‌شود. کارایی مدل معرفی شده، در بخش ۷ با استفاده از دو مثال مورد تحلیل قرار می‌گیرد.

۲ مفاهیم اولیه

در این بخش، برخی مفاهیم اساسی مجموعه‌های فازی که در این مقاله مورد استفاده قرار گرفته است، بیان می‌شوند [۱۰، ۲۲]. یک مجموعه فازی \tilde{A} از مجموعه اعداد حقیقی \mathbb{R} بر اساس تابع عضویت $\mu_{\tilde{A}} : \mathbb{R} \rightarrow [0, 1]$ تعریف می‌شود. برای هر $\alpha \in (0, 1]$ α -برش یک مجموعه فازی \tilde{A} به صورت مجموعه $\tilde{A}[\alpha] = \{x \in \mathbb{R} : \mu_{\tilde{A}}(x) \geq \alpha\}$ تعریف می‌شود. این بازه به صورت $\tilde{A}[\alpha] = [\tilde{A}_{\alpha}^L, \tilde{A}_{\alpha}^U]$ نمایش داده می‌شود که $\tilde{A}_{\alpha}^L = \inf\{x : x \in \tilde{A}[\alpha]\}$ و $\tilde{A}_{\alpha}^U = \sup\{x : x \in \tilde{A}[\alpha]\}$ همچنین مجموعه $\tilde{A}[0] = \{x \in \mathbb{R} : \mu_{\tilde{A}}(x) > 0\}$ بستار مجموعه فازی \tilde{A} نامیده می‌شود. یک مجموعه فازی \tilde{A} از مجموعه \mathbb{R} یک عدد فازی نامیده می‌شود اگر شرایط زیر را دارا باشد:

[†]Fuzzy Time Series

(۱) برای هر $\alpha \in [0, 1]$ بازه‌ای فشرده و مخالف تهی در \mathbb{R} می‌باشد.

(۲) نرمال باشد یعنی عدد یکتایی مانند $x_A^* \in \mathbb{R}$ وجود داشته باشد به طوری که $\mu_{\tilde{A}}(x_A^*) = 1$.

(۳) برای α و $\beta \in [0, 1]$ که $\alpha < \beta$ داشته باشیم $\tilde{A}[\beta] \subseteq \tilde{A}[\alpha]$.

زمانی که امکان دسترسی به مقادیر دقیق فراهم نباشد، می‌توان از ارزیابی تقریبی که توسط کارشناسان ارائه می‌شود و مورد قبول تصمیم گیرندگان است، استفاده کرد. این ارزیابی تقریبی به عنوان عدد فازی در نظر گرفته می‌شود که به خاطر نمایش ساده‌تر و کاربردی‌تر، به شکل پارامتری به نام عدد فازی LR ارائه می‌شود [۱۶]. یک عدد فازی LR به صورت $\tilde{A} = (a; l_a, r_a)_{LR}$ نمایش داده شده و $l_a, r_a > 0$ دارای تابع عضویت زیر می‌باشد:

$$\mu_{\tilde{A}}(x) = \begin{cases} L\left(\frac{a-x}{l_a}\right), & x \leq a, \\ R\left(\frac{x-a}{r_a}\right), & x > a, \end{cases} \quad (۱.۲)$$

به طوری که توابع L و R توابعی اکیداً نزولی و پیوسته از بازه $[0, 1]$ به بازه $[0, 1]$ هستند و دارای دو شرط زیر می‌باشند.

$$L(0) = R(0) = 1, \quad L(1) = R(1) = 0$$

بر کاربردترین اعداد فازی LR ، اعداد فازی مثلثی هستند که به صورت زیر تعریف می‌شوند. مجموعه فازی \tilde{A} را یک عدد فازی مثلثی[‡] گویند، اگر تابع عضویت آن به صورت زیر باشد:

$$\mu_{\tilde{A}}(x) = \begin{cases} \frac{x-(a-l_a)}{l_a} & a-l_a \leq x \leq a, \\ \frac{a+r_a-x}{r_a} & a \leq x \leq a+r_a, \\ 0 & x \in \mathbb{R} - [a-l_a, a+r_a]. \end{cases} \quad (۲.۲)$$

عدد فازی مثلثی را به صورت $\tilde{A} = (a; l_a, r_a)_T$ نشان می‌دهند. برخی عملگرهای معمول بین دو عدد فازی LR $\tilde{A} = (a; l_a, r_a)_{LR}$ و $\tilde{B} = (b; l_b, r_b)_{LR}$ به صورت زیر بیان می‌شوند [۲۴]:

$$\tilde{A} \oplus \tilde{B} = (a+b; l_a+l_b, r_a+r_b)_{LR} \quad (۱)$$

$$\lambda \otimes \tilde{A} = \begin{cases} (\lambda a; \lambda l_a, \lambda r_a)_{LR}, & \text{if } \lambda > 0, \\ (\lambda a; -\lambda r_a, -\lambda l_a)_{RL}, & \text{if } \lambda < 0. \end{cases} \quad (۲)$$

علاوه بر این، معیار فاصله D بین دو عدد فازی \tilde{A} و \tilde{B} به صورت زیر تعریف می‌شود:

$$D(\tilde{A}, \tilde{B}) = \sqrt{0.9|a-b|^2 + 0.5(|l_a-l_b|^2 + |r_a-r_b|^2)}, \quad (۳.۲)$$

این فاصله به صورت ترکیب خطی محدب از تفاضل میانه‌ها و تفاضل پهناها است که چون پهناها میزان ابهام عدد فازی را نشان می‌دهند، وزن کمتری گرفته‌اند. برای هر عدد فازی \tilde{A} ، \tilde{B} و \tilde{C} فاصله D دارای ویژگی‌های زیر است:

$$\tilde{A} = \tilde{B} \text{ اگر و تنها اگر } D(\tilde{A}, \tilde{B}) = 0 \cdot$$

$$D(\tilde{A}, \tilde{B}) = D(\tilde{B}, \tilde{A}) \cdot$$

$$D(\tilde{A}, \tilde{C}) \leq (D(\tilde{A}, \tilde{B}) + D(\tilde{B}, \tilde{C})) \cdot$$

[‡]Triangular Fuzzy Number

فاصله ارائه شده در رابطه (۳.۲) را برای برآورد پارامترهای مدل سری زمانی فازی در بخش بعد به کار می‌بریم. همچنین به منظور مقایسه نتایج پیش‌بینی مدل پیشنهادی با موارد مشابه، علاوه بر معیارهای ارزیابی، از نمودار انطباق پیش‌بینی با مقادیر نمونه نیز استفاده خواهیم کرد. برای این منظور مقادیر فازی با استفاده از معیار غیرفازی ساز سوگنو [۲۸] به مقادیر دقیق تبدیل می‌شوند. معیار سوگنو که به مرکز ثقل معروف است، برای عدد فازی $\tilde{A} = (a; l_a, r_a)_{LR}$ به صورت زیر تعریف می‌شود:

$$M_{\tilde{A}} = \frac{\int_{\mathbb{R}} x \mu_{\tilde{A}}(x) dx}{\int_{\mathbb{R}} \mu_{\tilde{A}}(x) dx} \quad (۴.۲)$$

نتیجه ۱.۲. اگر عدد فازی $\tilde{A} = (a; l_a, r_a)_{LR}$ را با توابع $L(x) = 1 - x$ و $R(x) = 1 - x^3$ در نظر بگیریم، مقادیر غیرفازی شده بر اساس روش مرکز ثقل برابر است با:

$$M_{\tilde{A}} = a + \frac{-3l_a^3 + 10r_a^3}{15(l_a + 2r_a)}. \quad (۵.۲)$$

و در یک حالت خاص، اگر $\tilde{A} = (a; l_a, r_a)_T$ یک عدد فازی مثلثی باشد، مرکز ثقل برابر است با:

$$M_{\tilde{A}} = a + (r_a - l_a)/3. \quad (۶.۲)$$

در بخش ۳ مروری بر مدل سری زمانی ناپارامتری در حالت کلاسیک داشته و سپس با استفاده از مفاهیم ارائه شده در این بخش، به معرفی مدل سری زمانی فازی پیشنهاد شده می‌پردازیم.

۳ مدل سری زمانی ناپارامتری در حالت کلاسیک

در این بخش، ابتدا یک مدل سری زمانی ناپارامتری بر اساس مشاهدات دقیق را در نظر گرفته و روش نادارایا-واتسون را برای برآورد تابع هموارساز بیان می‌کنیم. در حالت کلاسیک با در اختیار داشتن مشاهدات سری x_1, x_2, \dots, x_n هدف، پیش‌بینی مقدار مشاهده نشده x_{n+l} برای $l \geq 1$ است. در یک روش متداول فرض می‌شود که سری زمانی از یک فرآیند اتورگرسیو مرتبه p به صورت زیر پیروی می‌کند.

$$x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-p}) + \epsilon_t$$

که ϵ_t خطای فرآیند و f نیز تابعی حقیقی مقدار است [۱۲]. با توجه به اینکه مهم‌ترین هدف از تحلیل سری‌های زمانی در قلمرو زمان، پیش‌بینی مقادیر آینده بر اساس بررسی توصیفی و مدل‌بندی رفتار سری زمانی است، بنابراین برای پیش‌بینی سری ابتدا باید تابع f برآورد شود. یکی از معمول‌ترین روش‌های برآورد تابع f در حالت کلاسیک استفاده از تابع هموارساز است [۲۳، ۴۵]. از جمله توابع هموارساز، برآوردگرهای کرنل هستند که نخستین بار توسط نادارایا [۲۸] و واتسون [۴۶] معرفی شدند. پیشنهاد آنها این بود که هر هموارساز خطی ترکیبی خطی از مشاهدات بوده و هر مشاهده در برآورد تابع f دارای یک وزن است. در این صورت برآورد تابع رگرسیون \hat{f} که تابع هموارساز خطی نامیده می‌شود، به صورت زیر معرفی می‌شود:

$$\hat{f}(x) = \sum_{t=p+1}^n w_t(x) x_t$$

که برای هر x_t بردار $w(x_t) = (w_1(x_t), \dots, w_n(x_t))$ بردار وزن‌ها است به طوری که $\sum_{t=1}^n w_t(x) = 1$ برای یافتن برآورد تابع هموارساز، پیشنهاد نادارایا [۲۸] و واتسون [۴۶] این بود که وزن‌ها در برآورد تابع هموارساز می‌توانند به وسیله تابعی متقارن تک مدی حول صفر که مقادیر آن در دو طرف صفر توسط یک پارامتر مقیاس کنترل می‌شود، جایگزین شود. نخستین انتخاب آنها توابع چگالی احتمال به عنوان توابع کرنل بود. بنابراین، وزن‌ها در برآوردگر کرنل نادارایا-واتسون با استفاده از تابع کرنل K به صورت زیر تعریف می‌شود:

$$w_t(x_t) = \frac{K\left(\frac{x_t - x_j}{h}\right)}{\sum_{j=p+1}^n K\left(\frac{x_t - x_j}{h}\right)}$$

در برآورد تابع هموارساز یافتن وزن‌های مناسب دارای اهمیت هستند. برای تعیین وزن‌های مناسب باید به دو نکته توجه داشت. نکته اول اینکه در برآورد تابع $f(x_t)$ به نقاط نزدیکتر به مشاهده x_t وزن‌های بزرگتر و به نقاط دورتر وزن‌های کوچکتر اختصاص یابد. بنابراین معیاری

برای تعیین دوری و نزدیکی مورد نیاز است. این معیار، پارامتر هموارسازی یا پهنای باند نام دارد و با h نشان داده می شود. برآورد پارامتر پهنای باند را می توان از کمینه کردن معیاری به نام اعتباریابی متقابل تعمیم یافته $^{\S} (GCV)$ به صورت زیر به دست آورد.

$$\hat{h}_p = \arg \min_{h > 0} GCV_p(h) = \arg \min_{h > 0} \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \hat{f}(x_{i-1}, \dots, x_{i-p})}{1 - V} \right)^2 \quad (1.3)$$

که در آن $V = \frac{1}{n} tr(W)$ و $tr(W)$ ترانهاده ماتریس وزن ها است. انتخاب پارامتر هموارسازی در برآورد ناپارامتری بسیار مهم است. زیرا اگر مقدار کوچکی برای این پارامتر در نظر گرفته شود، برآورد تابع ناپارامتری دقیق تر اما ناهموار می شود و برعکس با انتخاب مقدار بزرگ برای این پارامتر برآورد تابع ناپارامتری نادقیق اما هموارتر می شود [۴۵].

نکته دوم در تعیین وزن های مناسب، مربوط به تعداد تاخیرهای مورد استفاده است که معمولاً توصیه می شود زیرمجموعه ای از تاخیرها به کار برده شود که منجر به بهترین پیش بینی گردد [۳۹]. برای تعیین تعداد تاخیرها و به عبارت دیگر مرتبه ای اتورگرسو سری زمانی، از معیار میانگین خطای پیش بینی $^{\dagger} (MFE)$ استفاده می شود. بنابراین مرتبه اتورگرسو بهینه به صورت زیر به دست می آید:

$$\hat{p} = \arg \min_p MFE_p = \frac{1}{n-p} \sum_{i=p+1}^n (x_i - \hat{f}(x_{i-1}, \dots, x_{i-p}))^2 \quad (2.3)$$

پس با به دست آوردن دو پارامتر پهنای باند و مرتبه اتورگرسو می توان تابع هموارساز را برآورد کرده و بر اساس مدل حاصل به پیش بینی پرداخت. بنابراین، در مدل سری زمانی ناپارامتری، پیش بینی X_{n+k} با تأخیر زمانی $k \in \mathbb{N}$ به صورت زیر نتیجه می شود:

$$\hat{x}_{n+k} = \sum_{t=p+1}^n \frac{K\left(\frac{x_{n+k-i} - x_{t-i}}{h}\right)}{\sum_{t=p+1}^n K\left(\frac{x_{n+k-i} - x_{t-i}}{h}\right)} \times x_{t+k-1} \quad (3.3)$$

در ادامه مدل سری زمانی ناپارامتری فازی با استفاده از روش کرنل معرفی می شود.

۴ مدل سری زمانی ناپارامتری فازی با استفاده از روش کرنل

در این بخش، به تعمیم مدل سری زمانی ناپارامتری در حالتی که مشاهدات مقادیر نادقیق و فازی هستند، می پردازیم. ابتدا توجه کنید که دو دیدگاه در مورد متغیر تصادفی فازی وجود دارد در رویکرد اول، یک متغیر تصادفی فازی یک متغیر تصادفی معمولی با مقادیر فازی در نظر گرفته می شود. در این دیدگاه، مفهوم متغیر تصادفی فازی کاملاً رسمی و زیربنای بسیاری از مطالعات احتمالی و آماری اخیر است که شامل داده های به دست آمده از یک آزمایش تصادفی است که توابع عضویت آنها بر اساس ترکیبی از روش های آماری و فازی تعیین می شود. از جمله مطالعات انجام شده با این مفهوم از متغیر تصادفی فازی می توان به [۹، ۲۶، ۳۵] اشاره کرد. در دیدگاه دوم منظور از یک نمونه تصادفی نادقیق و فازی، درک مبهمی از یک نمونه تصادفی معمولی است [۱۷]. به عبارت دیگر فازی بودن یک متغیر تصادفی بر اساس نظر یک کارشناس است که برای متغیر تصادفی کران هایی را به عنوان کم ترین و بیشترین مقدار ابهام و عدم دقت در نظر می گیرد [۱۳، ۱۴]. در این مقاله ما از دیدگاه دوم متغیر تصادفی فازی استفاده می کنیم. همچنین منظور از سری زمانی فازی نیز درک مبهمی از سری زمانی معمولی است. به عبارت دیگر یک سری زمانی معمولی با مشاهدات نادقیق و مبهم به عنوان سری زمانی فازی در نظر گرفته می شود. مجموعه $\tilde{x}_T = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n\}$ یک سری زمانی فازی نامیده می شود اگر هر \tilde{x}_i یک عدد فازی باشد که مفهوم ابهام 'حدودا' x_i را بیان کند [۷].

تعریف ۱.۴. مدل سری زمانی ناپارامتری فازی بر اساس سری زمانی فازی \tilde{x}_T به صورت زیر تعریف می شود:

$$\tilde{x}_t = \tilde{f}(\tilde{x}_{t-1}, \dots, \tilde{x}_{t-p}) \oplus \tilde{\epsilon}_t, \quad (1.4)$$

به طوری که

$$\tilde{f}(\tilde{x}_{t-1}, \dots, \tilde{x}_{t-p}) = (f(x_{t-1}, \dots, x_{t-p}); l_f(l_{t-1}, \dots, l_{t-p}), r_f(r_{t-1}, \dots, r_{t-p}))_{LR} \cdot \tilde{f} \quad ۱$$

یک عدد فازی است.

§ Generalized Cross Validation

† Mean Forecast Error

$$۲. \quad \tilde{\epsilon}_t = (\epsilon_t; l_{\epsilon_t}, r_{\epsilon_t})_{LR} \text{ خطای فازی است.}$$

۳. مرتبه اتورگرسیو است.

حال بر اساس یک مجموعه سری زمانی فازی، روش نادارایا-واتسون را برای برآورد تابع هموارساز فازی با استفاده از تعریف زیر به محیط فازی، تعمیم می‌دهیم.

تعریف ۲.۴. مدل سری زمانی ناپارامتری فازی معرفی شده در تعریف ۱.۴ را در نظر بگیرید. برآورد نادارایا-واتسون تعمیم‌یافته در نقاط فازی $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_p$ به صورت زیر تعریف می‌شود:

$$\begin{aligned} \tilde{f}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_p) &= \bigoplus_{t=p+1}^T (w_t(j; h, p) \otimes \tilde{x}_t)_{LR} \\ &= \bigoplus_{t=p+1}^T (w_t(j; h, p)x_t; w_t(j; h, p)l_{x_t}, w_t(j; h, p)r_{x_t})_{LR} \\ &= \left(\sum_{t=p+1}^T w_t(j; h, p)x_t; \sum_{t=p+1}^T w_t(j; h, p)l_{x_t}, \sum_{t=p+1}^T w_t(j; h, p)r_{x_t} \right)_{LR}, \end{aligned}$$

به طوری که

$$w_t(j; h, p) = \frac{\prod_{i=1}^p K\left(\frac{D(\tilde{x}_{j-i}, \tilde{x}_{t-i})}{h}\right)}{\sum_{t=p+1}^T \prod_{i=1}^p K\left(\frac{D(\tilde{x}_{j-i}, \tilde{x}_{t-i})}{h}\right)},$$

که K تابع کرنل و $h > 0$ است [۳۶، ۴۵].

به منظور بررسی تاثیر کرنل انتخابی در میزان پیش‌بینی مدل سری زمانی فازی ارائه شده در تعریف ۱.۴ از سه کرنل متداول نرمال، سه وزنی و ایپانچنیکف استفاده می‌کنیم (جدول ۱).

جدول ۱: برخی توابع کرنل	نوع تابع کرنل
شکل تابع کرنل	نرمال
$K(u) = \frac{1}{\sqrt{\pi}} \exp(-\frac{1}{2}u^2), \quad u \in R$	
$K(u) = \frac{3}{4}(1 - u^2), \quad u \leq 1$	ایپانچنیکف
$K(u) = \frac{35}{32}(1 - u^2)^2, \quad u \leq 1$	سه وزنی

ملاحظه ۳.۴. با توجه به فرضیات تعریف ۲.۴ در مدل سری زمانی ناپارامتری فازی ارائه شده، پیش‌بینی \tilde{x}_{T+k} با تأخیر زمانی $k \in \mathbb{N}$ یک عدد فازی است که به صورت زیر بیان می‌شود:

$$\tilde{x}_{T+k} = \bigoplus_{t=p+1}^T \left(\frac{\prod_{i=1}^p K\left(\frac{D(\tilde{x}_{T+k-i}, \tilde{x}_{t-i})}{h}\right)}{\sum_{t=p+1}^T \prod_{i=1}^p K\left(\frac{D(\tilde{x}_{T+k-i}, \tilde{x}_{t-i})}{h}\right)} \otimes \tilde{x}_{t+k-1} \right). \quad (۲.۴)$$

۵ انتخاب پهنای باند و مرتبه اتورگرسیو بهینه

حال برای به کارگیری مدل ارائه شده در تعریف ۱.۴ لازم است ابتدا به تعیین پهنای باند بهینه، مرتبه اتورگرسیو و تابع کرنل پرداخت. بدین منظور، با استفاده از تعمیم روش کلاسیک به حالت فازی، برای یافتن پهنای باند و مرتبه اتورگرسیو بهینه فرآیند چند مرحله‌ای زیر را انجام می‌دهیم.

۱. ابتدا مقدار $p = ۱$ را اختیار می کنیم.

۲. بر اساس روش ارزیابی متقابل تعمیم یافته با استفاده از رابطه زیر، پهنای باند بهینه را به دست می آوریم.

$$\hat{h}_p = \arg \min_{h > 0} GCV_p(h) = \arg \min_{h > 0} \frac{1}{T-p} \sum_{t=p+1}^T \left(\frac{D(\tilde{x}_t, \tilde{x}_t)}{1 - \frac{tr(W)}{T-p}} \right)^2, \quad (۱.۵)$$

که \tilde{x}_t مقدار پیش بینی \tilde{x}_t به ازای $t = p+1, \dots, T$ است و بر اساس نمونه فازی $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_T$ به دست می آید. همچنین $tr(W)$ ترانهاده ماتریس وزن ها است که درایه های آن به صورت زیر می باشند:

$$w_t(j; h, p) = \frac{\prod_{i=1}^p K\left(\frac{D(\tilde{x}_{j-i}, \tilde{x}_{t-i})}{h}\right)}{\sum_{t=p+1}^T \prod_{i=1}^p K\left(\frac{D(\tilde{x}_{j-i}, \tilde{x}_{t-i})}{h}\right)}$$

۳. مقدار p را یک واحد افزایش داده و به مرحله دوم باز می گردیم. مرتبه اتورگرسیو بهینه برابر است با:

$$\hat{p} = \arg \min_p MFE_p. \quad (۲.۵)$$

که در آن

$$MFE_p = \frac{1}{T-p} \sum_{i=p+1}^T D^2(\tilde{x}_i, \tilde{x}_i). \quad (۳.۵)$$

به عبارت دیگر برای $p = ۱, ۲, ۳$ مقادیر h_p و MFE_p را محاسبه می کنیم. مرتبه اتورگرسیو بهینه مقداری است که به ازای آن MFE_p کمینه شود. با داشتن مرتبه اتورگرسیو بهینه چون با استفاده از معیار GCV پهنای باند بهینه متناظر با آن محاسبه شده است، بنابراین پارامترهای بهینه مشخص می شوند. این فرآیند برای هر یک از توابع کرنل به کار گرفته می شود.

ملاحظه ۱.۵. در روش حسامیان و اکبری [۱۵]، یک مدل سری زمانی فازی نیمه پارامتریک به صورت زیر معرفی شده است:

$$\tilde{x}_i = \oplus_{l=1}^p (\theta_l \otimes \tilde{x}_{i-l}) \oplus \tilde{f}(t_i) \oplus \tilde{\epsilon}_i, \quad t_i = \frac{i}{T}$$

در این روش، حسامیان و اکبری [۱۵] یک الگوریتم بهینه سازی ترکیبی برای برآورد پارامترهای غیرفازی و تابع هموارساز فازی جهت پیش بینی مدل معرفی کردند.

زارعی و همکاران [۵۰] نیز با ارائه مدل سری زمانی نیمه پارامتری، حالت خاصی از مدل حسامیان و اکبری [۱۵] را بررسی کردند. بنابراین برای ارزیابی روش پیشنهادی، تنها با روش حسامیان و اکبری [۱۵] مقایسه صورت گرفته است.

۶ معیارهای ارزیابی

معیارهای ارزیابی متعددی برای بررسی توانایی پیش بینی مدل سری زمانی در تحقیقات انجام شده، معرفی شده است. در این بخش، به معرفی برخی از این معیارها می پردازیم [۲۱، ۶]. اگر خطای برآورد بین \tilde{x}_i و $\tilde{\hat{x}}_i$ به صورت زیر تعریف شود:

$$E_{\backslash}(\tilde{x}_i, \tilde{\hat{x}}_i) = \int_R \frac{|\tilde{x}_i(y) - \tilde{\hat{x}}_i(y)|}{\int_R \tilde{x}_i(y) dy} dy, \quad i = p+1, \dots, T \quad (۱.۶)$$

$$E_{\vee}(\tilde{x}_i, \tilde{\hat{x}}_i) = \int_R |\tilde{x}_i(y) - \tilde{\hat{x}}_i(y)| dy, \quad i = p+1, \dots, T \quad (۲.۶)$$

بر اساس روابط (۱۶) و (۲۶) از معیارهای زیر برای ارزیابی مدل استفاده می‌شود.

$$ME = \frac{1}{T-p} \sum_{i=p+1}^T E_1(\tilde{x}_i, \hat{x}_i). \quad (۳۶)$$

$$G_2 = \frac{1}{T-p} \sum_{i=p+1}^T \frac{1}{1 + E_2(\tilde{x}_i, \hat{x}_i)}. \quad (۴۶)$$

بر اساس معیار ME ، مدل مناسب‌تر، مدلی با مقدار کمتر ME است. بر اساس معیار $G_2 \in [0, 1]$ مدلی انتخاب می‌شود که این معیار مقادیر بیشتری را اختیار کند. اندازه مشابهت معیار دیگری است که برای اندازه‌گیری میزان نزدیکی مقدار فازی مشاهده شده و مقدار برآورد فازی متناظر با آن استفاده می‌شود. این معیار مقادیر بین صفر و یک را اختیار می‌کند و هرچه به یک نزدیک‌تر باشد، خطای برآورد در پیش‌بینی کمتر و در نتیجه مدل بهتر است. اندازه مشابهت به صورت زیر تعریف می‌شود [۱۸]:

$$MSM = \frac{1}{T-p} \sum_{i=p+1}^T S_{UI}(\tilde{x}_i, \hat{x}_i), \quad (۵۶)$$

به طوری که

$$S_{UI}(\tilde{x}_i, \hat{x}_i) = \frac{\int \min(\tilde{x}_i(y), \hat{x}_i(y)) dy}{\int \max(\tilde{x}_i(y), \hat{x}_i(y)) dy},$$

۷ مثال کاربردی

در این بخش، میزان توانایی پیش‌بینی مدل سری زمانی فازی پیشنهادی در مقایسه با روش‌های موجود برای دو مثال عددی مورد بحث و بررسی قرار می‌گیرد.

مثال ۱.۷. در این مثال، متوسط دمای ماهانه شهر اصفهان از سال ۲۰۰۰ تا ۲۰۱۸ بر اساس داده‌های اداره کل هواشناسی استان اصفهان، در نظر گرفته شده است. این داده‌ها با استفاده از روش باکلی [۲] به اعداد فازی مثلثی متقارن تبدیل شده‌اند. به این ترتیب که یک دمای هوای فازی بوسیله عدد فازی مثلثی متقارن $\tilde{x}_i = (\tilde{x}_i; t_{1,0.1} \frac{s_i}{\sqrt{n}})$ نمایش داده می‌شود که در آن s_i به ترتیب میانگین و انحراف استاندارد دمای هوا در سال‌های $i = 2000, \dots, 2018$ و $t_{v,\alpha}$ توزیع استودنت با v درجه آزادی است. به عبارتی بر اساس ۱۲ داده هرسال، یک داده فازی مثلثی متقارن خواهیم داشت. داده‌ها در جدول ۳ نشان داده شده‌اند. بر اساس مدل سری زمانی ناپارامتری فازی ارائه شده در تعریف ۱.۴ مدل زیر به داده‌های جدول ۳ برازش داده می‌شود.

$$\tilde{x}_t = \tilde{f}(\tilde{x}_{t-1}, \dots, \tilde{x}_{t-p}) \oplus \tilde{\epsilon}_t, \quad t = p+1, \dots, 19.$$

بر اساس الگوریتم معرفی شده در بخش ۵، تابع کرنل نرمال را در نظر می‌گیریم. ابتدا قرار می‌دهیم $p = 1$. سپس در مرحله دوم مقدار بهینه پهنای باند بر اساس کمینه کردن ارزیابی متقابل تعمیم یافته (GCV) محاسبه می‌شود. این دو مرحله برای $p = 2, 3$ نیز تکرار می‌شود. در مرحله بعد معیار MFE_p برای هر یک از مقادیر مرتبه اتورگرسیو با توجه به پهنای باند بهینه به دست آمده، محاسبه می‌شوند. مرتبه اتورگرسیو بهینه مقداری از p است که به ازای آن MFE_p کمینه شود. بنابراین برای کرنل نرمال پارامترهای بهینه به صورت $\hat{p} = 3$ و $\hat{h} = 0.18$ به دست می‌آیند. در پایان، معیارهای ارزیابی را برای مدل مشخص شده محاسبه می‌کنیم. روش حسامیان و اکبری [۱۵] را نیز به داده‌های دمای ماهانه هوای شهر اصفهان برازش می‌دهیم. در این روش بر اساس یک فرآیند دو مرحله‌ای شامل روش کمترین مربعات خطا و ارزیابی متقابل (CV) به یافتن پهنای باند، مرتبه اتورگرسیو و پارامترهای θ_l بر اساس مدل زیر می‌پردازیم.

$$\tilde{x}_i = \bigoplus_{l=1}^p (\theta_l \otimes \tilde{x}_{i-l}) \oplus \tilde{f}(t_i) \oplus \tilde{\epsilon}_i, \quad t_i = \frac{i}{19}$$

و سپس معیارهای ارزیابی را نیز به دست می‌آوریم. نتایج ارائه شده در جدول ۲ نشان می‌دهد که معیار اندازه مشابهت برای روش پیشنهادی مقدار 0.2749 بوده در حالی که این معیار برای روش حسامیان و اکبری [۱۵] مقدار 0.8398 نتیجه شده است. مقادیر این معیار برای دو روش نشان می‌دهد که در روش پیشنهادی مقادیر مشاهده شده و مقادیر برازش داده شده متناظر، نزدیکی و سازگاری بیشتری نسبت به

روش حسامیان و اکبری [۱۵] دارند و در نتیجه بر اساس این معیار با توجه به تابع کرنل نرمال روش پیشنهادی برازش بهتری را نسبت به مشاهدات انجام داده است. همچنین مقادیر معیارهای G_2 و ME نیز نشان از برازش بهتر روش ارائه شده دارند. پس بر اساس این سه معیار، با تابع کرنل نرمال روش پیشنهادی برازش مناسبتری به مشاهدات دارد و تصویر گویاتری از ساختار موجود بین مشاهدات را ارائه می‌دهد. به همین ترتیب مراحل برازش مدل را برای توابع کرنل ایپانچنیکف و سه وزنی نیز انجام می‌دهیم. نتایج پایانی برازش با پهنای باند و مرتبه اتورگرسو بهینه در جدول ۲ نشان داده شده است. محاسبات مربوط به یافتن پارامترهای بهینه با استفاده از نرم افزار ممتیکا انجام گردیده است. با مقایسه نتایج به دست آمده، در بین سه تابع کرنل، برای روش پیشنهادی توابع کرنل ایپانچنیکف و سه وزنی کارایی یکسانی را در برازش به داده‌ها نشان داده و بهتر از تابع کرنل نرمال عمل می‌کنند. هر سه معیار ME ، MSM و G_2 توابع کرنل ایپانچنیکف و سه وزنی را به عنوان توابع کرنل با نتایج بهتر نشان می‌دهند. مدل به دست آمده به صورت زیر می‌باشد:

$$\hat{x}_t = \hat{f}(\tilde{x}_{t-1}, \tilde{x}_{t-2}, \tilde{x}_{t-3}), \quad t = 4, 5, \dots, 19.$$

نتایج به دست آمده از روش حسامیان و اکبری [۱۵] کرنل ایپانچنیکف را به عنوان تابع کرنل مناسب برای این داده‌ها معرفی می‌کند. مدل به دست آمده برابر است با

$$\hat{x}_i = -0.05514 \otimes \tilde{x}_{i-1} \oplus -0.114388 \otimes \tilde{x}_{i-2} \oplus \hat{f}(t_i), \quad i = 3, \dots, 19$$

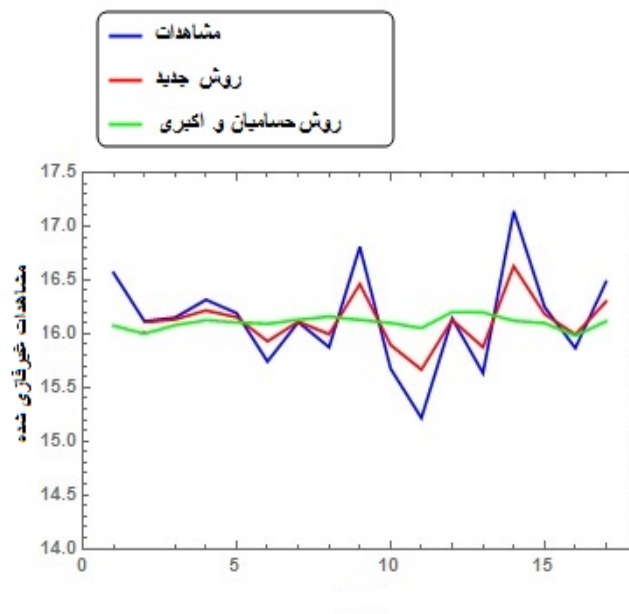
کلیه معیارهای ME ، MSM و G_2 ارائه شده در جدول ۲ نشان می‌دهند که هر سه تابع کرنل انتخاب شده بر اساس روش پیشنهادی نتایج بهتری را نسبت به روش حسامیان و اکبری [۱۵] ارائه می‌دهند. نمودار مقادیر غیرفازی شده مشاهدات فازی و برآورد متناظر آنها با استفاده از روش پیشنهادی و روش حسامیان و اکبری [۱۵] بر اساس تابع کرنل ایپانچنیکف در شکل ۱ نشان داده شده است. همان گونه که در شکل مشخص است، درحالی که روش حسامیان و اکبری [۱۵] برازشی کاملاً هموار ارائه داده است، روش پیشنهادی ساختار موجود در داده‌ها را به صورت واضح‌تری مشخص کرده و همخوانی بیشتری با شکل واقعی مشاهدات دارد. از آنجایی که روش زارعی و همکاران [۵۰] نیز مشابه روش حسامیان و اکبری [۱۵] می‌باشد، در مقایسه و برازش مدل این روش استفاده نشده است.

جدول ۲: معیارهای نیکویی برازش بر اساس روش پیشنهادی و روش حسامیان و اکبری در مثال ۱.۷

روش	تابع کرنل	\hat{p}	\hat{h}	MSM	ME	G_2	نتایج
روش حسامیان و اکبری	نرمال	۳	۰/۱۸	۰/۹۰۲۷۴	۰/۱۰۵۰۹	۰/۷۴۰۵۰	
روش پیشنهادی	ایپانچنیکف	۳	۰/۰۱	۰/۹۱۹۷۹	۰/۰۸۶۳۰	۰/۷۷۹۱۲	
سه وزنی	سه وزنی	۳	۰/۰۱	۰/۹۱۹۷۹	۰/۰۸۶۳۰	۰/۷۷۹۱۲	
روش حسامیان و اکبری	نرمال	۱	۰/۳۵	۰/۸۳۹۰۸	۰/۱۸۴۷۹	۰/۶۴۰۱۲	$\hat{\theta} = -0.04587$
روش حسامیان و اکبری	ایپانچنیکف	۲	۰/۶۳	۰/۸۴۶۹۰	۰/۱۷۴۴۷	۰/۶۴۷۸۰	$\hat{\theta}_1 = -0.05514$ $\hat{\theta}_2 = -0.11439$
سه وزنی	سه وزنی	۳	۰/۷۹	۰/۸۳۸۳۴	۰/۱۸۵۲۴	۰/۶۳۳۰۳	$\hat{\theta}_1 = -0.20299$ $\hat{\theta}_2 = -0.09890$ $\hat{\theta}_3 = -0.05577$

مثال ۲.۷. قیمت جهانی نفت خام را از اکتبر ۲۰۱۸ تا جولای ۲۰۲۰ در نظر می‌گیریم. داده‌ها از طریق آدرس <https://fred.stlouisfed.org> قابل دسترسی می‌باشند. به علت تغییر در شرایط سیاسی و اقتصادی و در نتیجه تغییرات روزانه قیمت جهانی نفت خام، این مشاهدات به صورت مشاهدات نادقیق می‌باشند. بر اساس تفاوت قیمت نفت خام هر ماه با کمترین و بیشترین قیمت در همان ماه، مشاهدات به صورت اعداد فازی LR با توابع $L(x) = 1 - x^3$ و $R(x) = 1 - x$ در نظر گرفته شده‌اند. مشاهدات در جدول ۵ نشان داده شده است. مشابه مثال قبل، بر اساس الگوریتم پیشنهادی با استفاده از سه تابع کرنل مدل زیر را به داده‌ها برازش می‌دهیم.

$$\tilde{x}_t = \hat{f}(\tilde{x}_{t-1}, \dots, \tilde{x}_{t-p}) \oplus \tilde{\epsilon}_t, \quad t = p + 1, \dots, 22.$$



شکل ۱: نمودار $M_{\hat{x}}$ و $M_{\bar{x}}$ با تابع کرنل ایپانچنیکف در مثال ۱.۷

جدول ۳: داده‌های دمای هوا اصفهان در مثال ۱.۷

$(x_i; l_i)_T$	سال	$(x_i; l_i)_T$	سال
$(۱۶/۸۰۸۳۳; ۳/۶۲۸۱۰)_T$	۲۰۱۰	$(۱۶/۴۶۶۶۷; ۳/۹۷۸۹۲)_T$	۲۰۰۰
$(۱۵/۶۷۵۰۰; ۴/۰۹۸۶۱)_T$	۲۰۱۱	$(۱۷/۱۴۱۶۷; ۳/۸۷۶۶۵)_T$	۲۰۰۱
$(۱۵/۲۱۶۶۷; ۳/۸۹۶۵۳)_T$	۲۰۱۲	$(۱۶/۵۶۶۶۷; ۳/۸۱۴۲۳)_T$	۲۰۰۲
$(۱۶/۱۴۱۶۷; ۳/۸۵۶۸۳)_T$	۲۰۱۳	$(۱۶/۱۱۶۶۷; ۳/۹۲۲۶۲)_T$	۲۰۰۳
$(۱۵/۶۳۳۳۳; ۴/۱۱۰۱۶)_T$	۲۰۱۴	$(۱۶/۱۵۰۰۰; ۳/۷۷۳۱۲)_T$	۲۰۰۴
$(۱۷/۱۴۱۶۷; ۳/۸۷۶۶۵)_T$	۲۰۱۵	$(۱۶/۳۱۶۶۷; ۳/۹۰۹۴۵)_T$	۲۰۰۵
$(۱۶/۲۵۰۰۰; ۳/۷۶۱۷۰)_T$	۲۰۱۶	$(۱۶/۱۹۱۶۷; ۴/۰۱۴۴۸)_T$	۲۰۰۶
$(۱۵/۸۶۶۶۷; ۳/۸۵۷۲۶)_T$	۲۰۱۷	$(۱۵/۷۴۱۶۷; ۳/۹۷۳۳۶)_T$	۲۰۰۷
$(۱۶/۴۸۳۳۳; ۳/۷۳۸۹۳)_T$	۲۰۱۸	$(۱۶/۱۰۸۳۳; ۴/۲۱۲۲۲)_T$	۲۰۰۸
		$(۱۵/۸۷۵۰۰; ۳/۸۷۶۸۹)_T$	۲۰۰۹

نتایج برازش روش پیشنهادی به این داده‌ها در جدول ۴ ارائه شده است. در روش پیشنهادی، معیار اندازه مشابهت تابع کرنل ایپانچنیکف را به عنوان تابع کرنل با برازش بهتر معرفی می‌کند. معیارهای ME و G_2 نیز تابع کرنل نرمال را به عنوان تابع کرنل مناسب در نظر می‌گیرند. بنابراین مدل مناسب برای داده‌ها با توجه به این تابع کرنل به صورت زیر نتیجه می‌شود:

$$\tilde{x}_t = \tilde{f}(\tilde{x}_{t-1}, \tilde{x}_{t-2}, \tilde{x}_{t-3}) \oplus \tilde{\epsilon}_t, \quad t = 4, \dots, 22.$$

نتایج برازش روش حسامیان و اکبری [۱۵] ارائه شده در جدول ۴ نشان می‌دهد که در این روش مرتبه اتورگرسیو بهینه برای تمامی توابع کرنل مقدار $\hat{p} = 2$ به دست آمده است. در این روش معیارهای ارزیابی نشان می‌دهند که دو تابع کرنل نرمال و سه وزنی عملکرد یکسانی دارند و نتایج بهتری را نسبت به تابع کرنل ایپانچنیکف ارائه می‌دهند. مدل نهایی بر اساس تابع کرنل نرمال برابر است با:

$$\tilde{x}_i = 1.01657 \otimes \tilde{x}_{i-1} \oplus -0.43379 \otimes \tilde{x}_{i-2} \oplus \tilde{f}(t_i), \quad t_i = \frac{i}{22}, \quad i = 3, 4, \dots, 22$$

در روش پیشنهادی برای تابع کرنل نرمال با پارامترهای بهینه $\hat{h} = 3.91$ و $\hat{p} = 3$ معیار میانگین خطا برابر است با $ME = 1.08858$ در حالی که در روش حسامیان و اکبری [۱۵] مقدار این معیار برابر با $ME = 1.63474$ به دست آمده و نشان از میانگین خطای بیشتر در این مدل و در نتیجه کمتر بودن خطا در مدل پیشنهادی دارد. همچنین روش پیشنهادی مقدار بیشتری از معیار G_2 و معیار اندازه مشابهت را در مقایسه با روش حسامیان و اکبری [۱۵] اختیار کرده که نشان از برازش بهتر و نزدیک‌تر به مشاهدات دارد. به همین ترتیب، برای توابع کرنل ایپانچنیکف و سه وزنی، با مرتبه اتورگرسیو بهینه $\hat{p} = 1$ مشاهده می‌شود که بر پایه معیارهای ME ، MSM و G_2 روش پیشنهادی برازش مناسب‌تری در مقایسه با برازش مدل حسامیان و اکبری [۱۵] دارد.

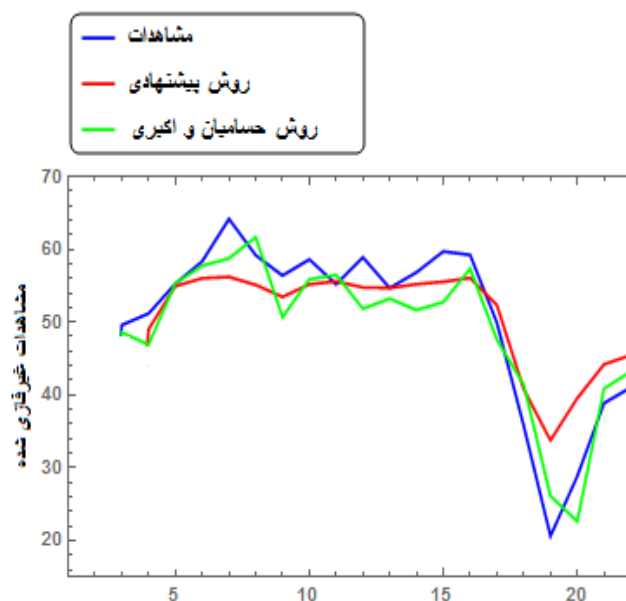
جدول ۴: معیارهای نیکویی براساس روش پیشنهادی و روش حسامیان و اکبری در مثال ۲.۷

روش	تابع کرنل	\hat{p}	\hat{h}	MSM	ME	G_2	نتایج
روش حسامیان و اکبری	ایپانچنیکف	۱	۰.۷۹	۰.۳۸۸۱۲	۱.۳۸۴۹۴	۰.۱۰۷۹۷	روش پیشنهادی
	سه وزنی	۱	۰.۶۳	۰.۳۷۹۶۸	۱.۴۱۲۰۷	۰.۱۰۵۵۷	روش پیشنهادی
	نرمال	۲	۰.۲۷	۰.۳۴۶۶۹	۱.۶۳۴۷۴	۰.۰۷۱۳۳	نرمال
							$\hat{\theta}_1 = 1.01657$ $\hat{\theta}_2 = -0.43379$
حسامیان و اکبری	ایپانچنیکف	۲	۰.۵۴	۰.۳۰۱۶۲۲	۱.۶۶۲۲۹	۰.۰۷۰۱۹	حسامیان و اکبری
	سه وزنی	۲	۰.۷۲	۰.۳۴۶۶۹	۱.۶۳۹۳۰	۰.۰۷۱۳۲	سه وزنی
							$\hat{\theta}_1 = 1.01225$ $\hat{\theta}_2 = -0.43980$

نمودار مقادیر غیرفازی شده مشاهدات فازی و برآوردهای متناظر آنان بر اساس روش پیشنهادی و روش حسامیان و اکبری [۱۵] بر اساس تابع کرنل نرمال در شکل ۲ نشان داده شده است. نمودار نشان می‌دهد که مقادیر برازش داده شده در روش پیشنهادی نزدیکی بیشتری به مقادیر مشاهده شده متناظر خود دارند.

۸ نتیجه گیری

در روش‌های تحلیل سری زمانی به تجزیه و بررسی مجموعه‌ای از مشاهدات همبسته جمع‌آوری شده در زمان، پرداخته می‌شود. این روش‌ها به دو دسته پارامتری و ناپارامتری تقسیم می‌شوند. در روش‌های پارامتری لازم است فرضیاتی در مورد مدل پیشنهادی داشته باشیم که اگر برقرار نباشند امکان استفاده از این روش‌ها وجود ندارد. اما روش‌های ناپارامتری اجازه می‌دهند بدون در نظر گرفتن قیود و پیش فرضی برای مدل به شناسایی ساختار موجود در داده‌ها پرداخته شود. همچنین از آنجائیکه در کاربردها و واقعیت ممکن است مشاهدات به صورت نادقیق و



شکل ۲: نمودار $M_{\hat{x}}$ و $M_{\tilde{x}}$ با تابع کرنل نرمال در مثال ۲.۷

جدول ۵: داده‌های قیمت نفت خام در مثال ۲.۷

$(x_i; l_i, r_i)_{LR}$	زمان	$(x_i; l_i, r_i)_{LR}$	زمان
$(54/07; 1/23, 9/31)_{LR}$	سپتامبر ۲۰۱۹	$(65/31; 0/50, 11/59)_{LR}$	اکتبر ۲۰۱۸
$(54/18; 3/19, 2/74)_{LR}$	اکتبر ۲۰۱۹	$(50/93; 1/52, 14/46)_{LR}$	نوامبر ۲۰۱۸
$(55/17; 1/10, 3/57)_{LR}$	نوامبر ۲۰۱۹	$(45/41; 3/05, 9/14)_{LR}$	دسامبر ۲۰۱۸
$(61/06; 5/71, 1/28)_{LR}$	دسامبر ۲۰۱۹	$(53/79; 9/44, 1/58)_{LR}$	ژانویه ۲۰۱۹
$(51/56; 0/59, 14/09)_{LR}$	ژانویه ۲۰۲۰	$(57/22; 5/99, 0/59)_{LR}$	فوریه ۲۰۱۹
$(44/76; 0/19, 9/74)_{LR}$	فوریه ۲۰۲۰	$(60/14; 5/62, 0/59)_{LR}$	مارس ۲۰۱۹
$(20/48; 1/21, 28/18)_{LR}$	مارس ۲۰۲۰	$(63/91; 3/78, 2/69)_{LR}$	آوریل ۲۰۱۹
$(19/56; 13/01, 9/56)_{LR}$	آوریل ۲۰۲۰	$(53/50; 0/45, 10/43)_{LR}$	می ۲۰۱۹
$(35/49; 17/44, 0/40)_{LR}$	می ۲۰۲۰	$(58/47; 7/87, 1/46)_{LR}$	ژوئن ۲۰۱۹
$(39/27; 5/00, 2/35)_{LR}$	ژوئن ۲۰۲۰	$(58/58; 3/86, 2/36)_{LR}$	جولای ۲۰۱۹
$(41/06; 2/52, 1/45)_{LR}$	جولای ۲۰۲۰	$(55/10; 4/58, 2/89)_{LR}$	اگوست ۲۰۱۹

فازی گزارش شوند، در این مقاله، بر اساس مشاهدات فازی و نادقیق یک مدل سری زمانی فازی به صورت ناپارامتری معرفی گردید. سپس با استفاده از تعمیم روش نادارایا-واتسون در محیط فازی، تابع هموارساز را برآورد نموده و با ارائه یک الگوریتم بهینه سازی، پارامترهای مدل برای دستیابی به پیش‌بینی دقیق‌تر برآورد شدند. در این الگوریتم، ابتدا یک تابع کرنل انتخاب می‌شود سپس برای هر مقدار از مرتبه اتورگرسیو، پهنای باند بهینه متناظر بر اساس ارزیابی متقابل تعمیم یافته به دست می‌آید. مرتبه اتورگرسیو بهینه مقداری است که معیار MFE_p را کمینه کند. الگوریتم بهینه سازی معرفی شده نسبت به روش‌های موجود ساده‌تر و دارای محاسبات راحت‌تر است. همچنین برای بررسی ارزیابی مدل و جهت انتخاب یک تابع کرنل با برازش بهتر به داده‌ها از معیارهای نیکویی برازش مدل استفاده می‌شود. جهت مقایسه کارایی مدل با روش حسامیان و اکبری [۱۵] نیز این معیارها به کار گرفته می‌شوند. در ادامه مقادیر غیرفازی شده مشاهدات فازی و برآوردهای متناظر آنها رسم می‌شوند تا نحوه برازش مدل به مشاهدات بهتر به تصویر کشیده شود. توانایی پیش‌بینی مدل در مقایسه با روش‌های موجود در دو مثال کاربردی، با استفاده از معیارهای ارزیابی بررسی شد. نتایج به دست آمده حاکی از آن است که روش ارائه شده نسبت به روش‌های موجود، به پیش‌بینی دقیق‌تری می‌انجامد. در این تحقیق برای برازش مدل از تعمیم روش ناپارامتری نادارایا-واتسون در محیط فازی استفاده شد. می‌توان برآورد و پیش‌بینی مدل را با استفاده از تعمیم سایر روش‌های ناپارامتری در محیط فازی نیز دنبال کرد. حساسیت مدل به داده‌های دورافتاده و نحوه شناسایی و برخورد با این مشاهدات نیز به عنوان آینده این تحقیق محل بررسی خواهد بود. در نظر گرفتن مدل سری زمانی ناپارامتری با داده‌های فازی و متغیر کمی دقیق یا فازی نیز می‌تواند یک زمینه مناسب پژوهشی دیگر باشد. همچنین می‌توان با استفاده از تعریف دیگر متغیر تصادفی فازی رویکرد جدیدی را در رابطه با مدل سری زمانی ناپارامتری فازی در نظر گرفت و به تعمیم توابع اتوکواریانس و خودهمبستگی در محیط فازی پرداخت تا بتوان به شناخت دقیق‌تری از ساختار داده‌ها دست یافت.

فهرست منابع

- [1] Aladag, C. H. Yolcu U. and Egrioglu, E.(2010) A high order fuzzy time series forecasting model based on adaptive expectation and artificial neural network, *Math. Comput. Simul*, 81 875-882.
- [2] Buckley, J.J. (2006), *Fuzzy Statistics, Studies in Fuzziness and Soft Computing*, Springer-Verlag, Berlin.
- [3] Bulut, E.(2014) Modeling seasonality using the fuzzy integrated logical forecasting (FILF) approach, *Expert Syst. Appl*, 41 1806-1812.
- [4] Chen, M. Y. (2014) A high-order fuzzy time series forecasting model for internet stock trading, *Future Gener. Comput. Syst*, 37, 461-467.
- [5] Cheng, S. H. Chen, S. M. and Jian, W. S. (2016) Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures, *Inf. Sci*, 327 272-287.
- [6] Chen, S.P. and Dang, J.F. (2008) A variable spread fuzzy linear regression model with higher explanatory power and forecasting accuracy, *Information Sciences*, 178, 3973-3988.
- [7] Denoeux, T. Masson, M.H. Herbert, P.H. (2005) Non-parametric rank-based statistics and significance tests for fuzzy data, *Fuzzy Sets Syst*, 153, 1-28.
- [8] Diks, C. and Mudelsee, M. (2000) Redundancies in the Earth's climatological time series, *Physics Letters*, A 275(5-6), 407-414.
- [9] Dominguez-Menchero, J. S. Lopez-Diaz, M. Colubi, A. and D. A. Ralescu.(2001) On the formalization of fuzzy random variables. *Information Science*, 133, 3-6.
- [10] Dubois, D. and Prade, H.(1980) *Fuzzy Sets and Systems, Theory and Applications*, Academic Press, New York.
- [11] Efendi, R. Ismail Z. and Deris, M. M. (2015) A new linguistic out-sample approach of fuzzytime series for daily forecasting of Malaysian electricity load demand, *Appl. Soft Comput*, 28, 422-430.

- [12] Fan, J. and Yao, Q.(2002) Nonlinear time series nonparametric and parametric methods, *Springer-Verlag New York Berlin Heidelberg*
- [13] Grzegorzewski, P. (2009) K-sample median test for vague data, *International Journal of Intelligent Systems*, 24, 529–539.
- [14] Grzegorzewski, P. (2020) Two-sample dispersion problem for fuzzy data, *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 1239, 82–96.
- [15] Hesamian, G. and Akbari, M.G. (2018) A semi-parametric model for time series based on fuzzy data, *IEEE Transactions on Fuzzy Systems*, 26, 2953 - 2966.
- [16] Hesamian, G. and Akbari, M.G. (2019) Fuzzy Lasso regression model with exact explanatory variables and fuzzy responses, *International Journal of Approximate Reasoning*, 115, 290-300.
- [17] Hesamian, G. and Shams, M. (2015), Parametric testing statistical hypotheses for fuzzy random variables, *Soft Computing*, 20, 1537–1548.
- [18] 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.
- [19] Holicapek M. and Nguyen, L.(2016) Suppression of high frequencies in time series using fuzzy transform of higher degree, in: 16th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2016, Eindhoven, The Netherlands, 705-716.
- [20] Huang, Y. L. Horng, S. J. He, M. Fan, P. Kao, T. W. Khan, M. K. Lai J. L. and Kuo, I.H. (2011) A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization, *Expert Syst. Appl*, 38, 8014-8023.
- [21] Kim, B. and Bishu, R.R.(1998), Evaluation of fuzzy regression models by comparing membership functions, *Fuzzy Sets and Systems*, 100, 343–352.
- [22] Klir, V.E., and Yuan, B.(1995), Fuzzy sets and Fuzzy logic: Theory and application. Prentice-Hall Inc.
- [23] Kvam, P.H. and Vidakovic, B. (2007), Nonparametric statistics with applications to science and engineering, John Wiley & Sons
- [24] Lee, K.H. (2005), First course on fuzzy theory and applications, Springer-Verlag, Berlin.
- [25] Liu, B. (2013), Uncertainty Theory. Tsinghua University-Beijing 100084-China.
- [26] Lpez-Daz, M. Gil, M.A. and Ralescu, D.A. (2006), Overview on the development of fuzzy random variables, *Fuzzy Sets and Systems*, 157, 2546–2557.
- [27] Lütkepohl, H. and Krattzig, M.(2004), Applied time series econometrics, New York, Cambridge University Press.
- [28] Nadaraya, E. A. (1964), On Estimating Regression, *Theory of Probability and Its Applications*, 9 (1), 141–142.
- [29] Nguyen, L. and Novák, V. (2019), Forecasting seasonal time series based on fuzzy techniques, *Fuzzy Sets Syst*, 361, 114–129.

- [30] Olatayo, T.O. and Taiwo, A.I. (2014), Statistical modelling and prediction of rainfall time series data, *Global Journal of Computer Science and Technology*, 14, 1-9.
- [31] Peng, H. W. Wu, S. F. Wei C. C. and S. J. Lee, (2015) Time series forecasting with a neuro-fuzzy modeling scheme, *Appl. Soft Comput*, 32, 481-493.
- [32] Rashid, R. and Hossen, S.S.(2020) Effect of foreign direct investment on bangladesh economy: a Time series analysis from 1972 to 2013. Preprints 2020, doi: 10.20944/preprints202012.0611.v1
- [33] Sadaei, H. J. Enayatifar, R. Lee M.H. and Mahmud, M. (2016) A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting, *Appl. Soft Comput*, 40, 132-149.
- [34] Sezer, O. B. Gudelek, M. U. and Ozbayoglu, A. M. (2020), Financial time series forecasting with deep learning : A systematic literature review: 2005–2019, *Applied Soft Computing*, 90, doi:10.1016/j.asoc.2020.106181.
- [35] Shapiro, F.A. (2009) Fuzzy random variables, *Insurance: Journal of Mathematical Economics*, 44, 307–314.
- [36] Simonoff, J.(1996), Smoothing methods in statistics. Springer, New York.
- [37] Song, Q. Chissom, B.S. (1993), Fuzzy time series, its models, *Fuzzy Sets Syst*, 54, 269–277.
- [38] Sugeno, M.(1985), An introductory survey of fuzzy control, *Information Sciences*, 36, 59–83.
- [39] Tjostheim, D. and Auestad, B. (1994) Non-parametric identification of non-linear time series: selecting significant lags, *Journal of the American Statistical Association*, 89, 1410-1419.
- [40] Tong, H. (1990), Non-linear Time Series: A Dynamical Systems Approach, Oxford University Press, Oxford.
- [41] Torbat, S. Khashei M. and Bijari, M. (2018), A hybrid probabilistic fuzzy ARIMA model for consumption forecasting in commodity markets, *Econ.Anal. Policy*, 58, 22-31.
- [42] Tsay, R. S.(2002), Analysis of financial time series, John Wiley & Sons.
- [43] Uslu, V. R. Bas, E. Yolcu U. and Egrioglu ,E.(2014) A fuzzy time series approach based on weights determined by the number of recurrences of fuzzy relations, *Swarm Evol. Comput*, 15, 19-26.
- [44] Viertl, R.(2011), Statistical methods for fuzzy data. Wiley, Chichester.
- [45] Wasserman, L. (2007), All of nonparametric statistics, New York, Springer.
- [46] Watson, G. S. (1964), Smooth regression analysis, *Sankhyā: The Indian Journal of Statistics*, Series A. 26 (4), 359–372.
- [47] Wei, W.S. (2006), Time series analysis: univariate and multivariate methods, New York, Pearson Addison Wesley.
- [48] Yolcu O. C. and Lam,H. K.(2017) A combined robust fuzzy time series method for prediction of time series, *Neurocomputing*, 247, 87-101.
- [49] Yu, H. (2005), Weighted fuzzy time series models for TAIEX forecasting, *ELSEVIER Journal of Physica A*.

- [50] Zarei, R. Akbari, M.Gh. and Chachi, J. (2020), Modeling autoregressive fuzzy time series data based on semi-parametric methods, *Soft Computing*, 24(10), 7295-7304.



Nonparametric Time Series Modeling based on Fuzzy Data

Faezeh Torkian¹, Masoud Yarmohammadi¹ , Gholamreza Hesamian¹, Mohammad Ghasem Akbari²

⁽¹⁾ Department of Statistics, Payame Noor University, Tehran 19395-4697, Iran.

⁽²⁾ Department of Statistics, University of Birjand, Birjand 615-97175, Iran.

Received: 2021/4/3

Accepted: 2021/8/3

Communicated by: Rahim Chinipardaz

Abstract: In this paper, a nonparametric time series model based on fuzzy observations is presented. Fuzzy prediction values are estimated using the generalization of the Nadaraya-Watson method in fuzzy environment. An algorithm for achieving autoregressive order and optimal bandwidth is stated and then criteria are introduced to evaluate the predicted values. In the following the performance of the proposed model is examined and analyzed using real data. The effectiveness of the proposed model is also compared with the other time series models with fuzzy data.

Keywords: Fuzzy nonparametric time series, kernel method, optimal bandwidth, autoregressive order.



©2021 Shahid Chamran University of Ahvaz, Ahvaz, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0 license) (<http://creativecommons.org/licenses/by-nc/4.0/>).

¹ Corresponding author.

E-mail addresses: masyar@pnu.ac.ir (M. Yarmohammadi), fatorkian29@gmail.com, (F.Torkian), gh.hesamian@pnu.ac.ir, (G. Hesamian), g_z_akbari@birjand.ac.ir, (M. G. Akbari).