مهندسی صنایع و مدیریت شریف، (تابستان ۱۴۰۰)
دوره‌ی ۱-۳۷، شماره‌ی ۱، ص. ۳-۱۲، (پژوهشی)

Original Article

بهبود عملکرد روش‌های خودرگرسیون میانگین متحرک انباشته کلاسیک با استفاده از تکنیک‌های تجزیه‌ی تجمعی به عوامل اصلی

پگاه امینی (دانشجوی کارشناسی ارشد)
مهدی خاشعی * (استادیار)
دانشکده‌ی صنایع و برنامه‌ریزی سیستم‌ها، دانشگاه صنعتی اصفهان

مدل‌های خودرگرسیون میانگین متحرک انباشته از شناخته شده‌ترین روش‌های آماری هستند. در ادبیات موضوع تلاش‌های فراوانی برای رفع نقایص و محدودیت‌های این‌گونه از مدل‌ها ارائه شده است. در این نوشتار، روشی برای مقابله با محدودیت ساختارهای پیچیده و چندگانه با استفاده از تکنیک‌های تجزیه‌ی تجمعی به عوامل اصلی ارائه شده است. در روش پیشنهادی، ابتدا سری زمانی مورد مطالعه که اساساً پیچیده و شامل چندین ساختار همزمان متفاوت است، به اجزاء تشکیل‌دهنده‌ی خود که اصولاً پیچیدگی کمتری دارند و ساختارهای کمتری را نیز شامل می‌شوند، تجزیه می‌شود. سپس هریک از این ساختارهای ساده‌سازی شده با استفاده از روش خودرگرسیون میانگین متحرک انباشته، پیش‌بینی می‌شود. نهایتاً نیز پیش‌بینی هریک از اجزاء اصلی به‌منظور تشکیل پیش‌بینی‌های نهایی با یکدیگر ترکیب می‌شود. نتایج حاصله از به‌کارگیری روش پیشنهادی در پیش‌بینی قیمت جهانی نفت خام که از پیچیده‌ترین سری‌های زمانی در بازارهای مالی هستند، بیانگر کارآمدی روش پیشنهادی است.

واژگان کلیدی: مدل‌های خودرگرسیون میانگین متحرک انباشته کلاسیک، تکنیک‌های تجزیه‌ی تجمعی به عوامل اصلی، سری‌های زمانی پیچیده‌ی چندگانه، پیش‌بینی بازارهای مالی، قیمت نفت خام.

pegah_14_amini@yahoo.com
khashei@cc.iut.ac.ir

۱. مقدمه

پیش‌بینی سری‌های زمانی یکی از مهمترین زمینه‌های مدل‌سازی است که در آن مشاهدات گذشته‌ی یک متغیر جمع‌آوری و به‌منظور به دست آوردن روابط اساسی بین مشاهدات و تعیین یک مدل توصیفی، تجزیه و تحلیل می‌شود؛ سپس از مدل حاصله برای برون‌یابی سری‌های زمانی استفاده می‌شود. این روش مدل‌سازی اصولاً زمانی مورد استفاده قرار می‌گیرد که اطلاعات کافی درمورد فرایند اساسی تولید داده‌ها در دسترس نباشد یا هنگامی که هیچ مدل توضیحی رضایت بخشی که متغیر وابسته را به سایر متغیرهای توضیحی مرتبط سازد، وجود نداشته باشد.[۱] در چند دهه‌ی اخیر، تلاش‌های فراوانی برای توسعه و بهبود مدل‌های پیش‌بینی سری‌های زمانی صورت گرفته است. مدل‌های خودرگرسیون میانگین متحرک انباشته، یکی از مهمترین و شناخته شده ترین مدل‌های آماری سری‌های زمانی هستند.

عمومیت این‌گونه از مدل‌ها به دلیل خواص آماریشان و همچنین متدولوژی معروف باکس ـ جنکینز[۲] در فرایند مدل‌سازی این‌گونه از مدل‌ها است. علاوه بر این، مدل‌های هموارسازی نمایی متعددی می‌توانند توسط این‌گونه از مدل‌ها به‌کارگرفته شوند. مدل‌های خودرگرسیون میانگین متحرک انباشته همچنین کاملاً انعطاف‌پذیرند، چرا که می‌توانند مدل‌های بسیار متفاوتی از سری‌های زمانی همچون سری‌های خودرگرسیون، میانگین متحرک و ترکیب هر دو آنها را تشریح کنند. این‌گونه مزایای منحصر به‌فرد روش‌های خودرگرسیون میانگین متحرک انباشته باعث شده تا از گذشته‌های بسیار دور تا به امروز به کرات در حوزه‌های مختلف به‌کار گرفته شوند.

ادیگر و همکاران[۳] در سال ۲۰۰۶ تولید منابع سوخت‌های فسیلی را در ترکیه با استفاده از روش رگرسیون مقایسه‌یی و مدل خودرگرسیون میانگین متحرک انباشته پیش‌بینی کردند. دیاز روبلز و همکاران[۴] در سال ۲۰۰۸ از یک روش ترکیبی خودرگرسیون میانگین متحرک انباشته و مدل شبکه‌های عصبی برای پیش‌بینی ریزگردها در مناطق شهری (شهر تموکو شیلی) استفاده کردند. کادناس و ریورا[۵] در

* نویسنده مسئول
تاریخ: دریافت ۱۳۹۸/۳/۱۲، اصلاحیه ۱۳۹۹/۱۱/۵، پذیرش ۱۳۹۹/۱۲/۱۲.
DOI:10.24200/J65.2021.53326.1993

سال ۲۰۱۰ از مدل ترکیبی خودرگرسیون میانگین متحرک انباشته و مدل شبکه‌های عصبی به‌منظور پیش‌بینی سرعت باد در سه منطقه از مکزیک استفاده کردند. هان و همکاران[۶] در سال ۲۰۱۰، روش خودرگرسیون میانگین متحرک انباشته را به‌منظور پیش‌بینی خشک‌سالی به کار گرفتند. همچنین لی و تانگ[۷] در سال ۲۰۱۱، پیش‌بینی سری‌های زمانی را با استفاده از روش خودرگرسیون میانگین متحرک انباشته و برنامه‌نویسی ژنتیک ارائه کردند. سن و همکاران[۸] در سال ۲۰۱۶ از روش خودرگرسیون میانگین متحرک انباشته، مصرف انرژی کارخانه‌های تولیدکننده آهن خام در هند را پیش‌بینی کردند. هیکیچی و همکاران[۹] در سال ۲۰۱۷ تعداد گواهی‌های ایزو ۱۴۰۰۱ را در آمریکا با استفاده از مدل خودرگرسیون میانگین متحرک انباشته پیش‌بینی کردند. گارسیا نیتو و همکاران[۱۰] در سال ۲۰۱۷ به‌منظور پیش‌بینی فاکتور آلودگی در شمال اسپانیا از روش خودرگرسیون میانگین متحرک انباشته استفاده کردند. موسایل و همکاران[۱۱] در سال ۲۰۱۸ پیش‌بینی تقاضای الکتریسیته کوتاه‌مدت را با استفاده از روش خودرگرسیون میانگین متحرک انباشته و با استفاده از جمع‌آوری داده‌های تقاضای الکتریسیته کوتاه‌مدت کوئینزلند استرالیا انجام دادند.

اما علی‌رغم تمامی مزایای منحصربه‌فردی که برای مدل‌های خودرگرسیون میانگین متحرک انباشته در نظر گرفته شده است، این‌گونه از مدل‌ها معایبی نیز دارند که باعث شده تا کاربردهای آنها در برخی از مسائل و زمینه‌های خاص محدود شود. مهمترین این محدودیت‌ها عبارت‌اند از: ۱. محدودیت تعداد داده‌های مورد نیاز؛ ۲. محدودیت خطی بودن؛ ۳. محدودیت قطعی بودن؛ ۴. محدودیت ساختارهای پیچیده و چندگانه و غیره.[۱۲] تلاش‌های بسیاری در ادبیات موضوع به‌منظور مرتفع ساختن یا تعدیل این‌گونه محدودیت‌ها صورت پذیرفته است. تسنگ و همکاران[۱۳] مدلی بهبود یافته از مدل‌های خودرگرسیون میانگین متحرک انباشته با منطق و مجموعه‌های فازی ارائه کرده‌اند تا بتوانند حتی‌الامکان محدودیت تعداد داده‌های مورد نیاز را در این‌گونه از مدل‌ها مرتفع سازند. خاشعی و همکاران[۱۴] با ترکیب شبکه‌های هوشمند غیرخطی با مدل‌های خودرگرسیون میانگین متحرک انباشته، سعی در مرتفع ساختن محدودیت خطی بودن در مدل‌سازی توسط این مدل‌ها را داشته‌اند. خاشعی و همکاران[۱۵] مدلی ترکیبی با عنوان خودرگرسیون میانگین متحرک انباشته‌ی فازی احتمالی را به‌منظور مدل‌سازی الگوهای غیرقطعی موجود در داده‌ها و رفع محدودیت قطعی بودن در این‌گونه از مدل‌ها ارائه کرده‌اند.

هدف اصلی نوشتار حاضر نیز مرتفع ساختن محدودیت ساختارهای چندگانه و پیچیده در مدل‌های خودرگرسیون میانگین متحرک انباشته است. در ادبیات موضوع، ایده‌های متنوعی به‌منظور رفع این نقیصه پیشنهاد شده است. این تکنیک‌ها مبتنی هستند بر استفاده از مفهوم جداسازی و تجزیه‌ی سری‌های زمانی پیچیده و چندگانه به چندین سری تشکیل‌دهنده، به قسمی که تعداد ساختارهای موجود در این اجزا تشکیل‌دهنده یا پیچیدگی موجود در آنها کمتر از سری زمانی اولیه مورد مطالعه باشد. به‌عبارت دیگر، در این‌گونه تکنیک‌ها یک سری زمانی به اجزای اصلی تشکیل‌دهنده خود تجزیه می‌شود تا بدین وسیله بتوان به داده‌هایی با پیچیدگی کمتر و ساختارهای ساده‌تر دست یافت. ایده‌ی اصلی این‌گونه از تکنیک‌ها بر پایه‌ی اصل تجزیه و تجمع بنا نهاده شده است که برای تجزیه‌ی یک سری زمانی غیرخطی و غیرایستا همراه با نوسانات، پیچیدگی‌ها و فراریت بالا به اجزاء اصلی تشکیل‌دهنده‌ی آن سری زمانی به کار برده می‌شود.[۱۶] به‌عبارت دیگر، هدف اصلی تجزیه‌ی داده، ساده‌سازی داده‌ها به‌منظور مدل‌سازی و پیش‌بینی است.[۱۷]

تکنیک تجزیه به عوامل اصلی[۱] یکی از شناخته شده ترین روش‌های تجزیه است که اولین بار توسط هوانگ و همکاران[۱۸] در سال ۱۹۹۸ معرفی شده است. این تکنیک به‌کرات در مدل‌سازی بازارهای مالی که اساساً داده‌هایی با نوسانات بسیار بالا و پیچیدگی زیاد و ساختاری چندگانه را شامل می‌شوند، مورد استفاده قرار گرفته است. به‌کارگیری این تکنیک همراه با سایر روش‌های پیش‌بینی، تأثیرات به‌سزایی در بهبود دقت پیش‌بینی سری‌های زمانی با ویژگی‌های مذکور داشته است. این بهبود دقت حاصله، باعث شده تا محققان متعددی از این روش استفاده کنند. لیو و همکاران[۱۹] یک روش ترکیبی بر پایه‌ی مدل خودرگرسیون میانگین متحرک انباشته و تکنیک تجزیه به عوامل اصلی برای پیش‌بینی سرعت باد پیشنهاد کرده‌اند. یو و همکاران[۲۰] یک روش ترکیبی بر پایه‌ی تکنیک تجزیه به عوامل اصلی برای پیش‌بینی قیمت نفت خام پیشنهاد داده‌اند. وانگ و همکاران[۲۱] با استفاده از تکنیک تجزیه، یک روش ترکیبی به‌منظور پیش‌بینی سرعت باد پیشنهاد کرده‌اند. چنگ و همکاران[۲۲] نیز از این روش به‌منظور پیش‌بینی قیمت سهام در بورس تایوان استفاده کرده‌اند.

اگرچه، تکنیک تجزیه به عوامل اصلی تا حدودی توانسته بر مشکلات مترتب به پیش‌بینی سری‌های زمانی با نوسانات و پیچیدگی‌های بالا و همچنین ساختار چندگانه غلبه کند، اما این تکنیک نیز با اشکالاتی در تجزیه‌ی سری‌های زمانی مواجه است. یکی از مهمترین این مشکلات، پدیده‌یی است که آن را اصطلاحاً با نام «اختلاط مد[۲]» در ادبیات موضوع می‌شناسند. در این پدیده برخلاف ایده‌ی اصلی تجزیه به عوامل اصلی تشکیل‌دهنده، ممکن است چندین مد اختلاطی در یک بسامد خاص از داده‌ها موجود باشد. این پدیده به‌دلیل ناهمگونی که در عوامل تشکیل‌دهنده ایجاد می‌کند، عملکرد نهایی مدل پیش‌بینی را تحت تأثیر قرار خواهد داد. بر این اساس، وو و هوانگ[۲۳] روشی با عنوان روش تجزیه‌ی تجمعی اجزاء اصلی تشکیل‌دهنده را در سال ۲۰۰۹ ارائه کردند. این روش که اساساً روشی تعمیم‌یافته از همان روش کلاسیک است، پدیده‌ی اختلاط مد را بر پایه‌ی اضافه کردن یک عبارت خطای تصادفی در هر تکرار مرتفع کند.

بر این اساس، هدف نهایی روش پیشنهادی در این نوشتار ترکیب تکنیک تجزیه‌ی تجمعی اجزاء اصلی با روش خودرگرسیون میانگین متحرک انباشته است. در روش پیشنهادی، ابتدا سری زمانی مورد مطالعه با استفاده از تکنیک تجزیه‌ی تجمعی به تعدادی از اجزاء تشکیل‌دهنده‌ی مستقل تجزیه می‌شود. سپس در مرحله بعد، هریک از این اجزاء تشکیل‌دهنده به‌طور جداگانه با استفاده از روش خودرگرسیون میانگین متحرک انباشته پیش‌بینی می‌شود؛ و نهایتاً این پیش‌بینی‌ها با یکدیگر ترکیب می‌شوند تا پیش‌بینی نهایی حاصل شود. به‌منظور ارزیابی عملکرد روش پیشنهادی، دو مجموعه داده از داده‌های معیار بازارهای مالی مربوط به قیمت نفت خام، شامل قیمت نفت خام برنت و قیمت نفت خام تگزاس مورد استفاده قرارگرفته است. علت انتخاب این داده‌ها، وجود نوسانات بسیار زیاد، فراریت و پیچیدگی بالا با ساختار چندگانه است. نتایج حاصله از به‌کارگیری روش پیشنهادی نشان می‌دهد که این مدل نه‌تنها توانسته است محدودیت ساختارهای پیچیده و چندگانه‌ی موجود در داده‌ها را مرتفع سازد، بلکه توانسته است نتایجی دقیق‌تر از شبکه‌های عصبی و مدل‌های هوشمند نیز به‌دست آورد. سایر قسمت‌های این مقاله بدین صورت مرتب شده‌اند: در بخش دوم روش پیشنهادی توصیف شده است. در بخش سوم، داده‌های معیار قیمت نفت خام استفاده شده تشریح شده است. در بخش چهارم نتایج عددی حاصل از به‌کارگیری روش پیشنهادی در پیش‌بینی قیمت نفت خام شرح داده شده است. در قسمت پنجم، عملکرد روش پیشنهادی با سایر روش‌های کلاسیک و هوش محاسباتی مقایسه شده است. در نهایت نیز نتیجه‌گیری بیان شده است.

۲. مدل ترکیبی پیشنهادی

چنان که اشاره شد، هدف نهایی روش پیشنهادی ترکیب خصوصیات منحصر به فرد تکنیک‌های تجزیه‌ی تجمعی به عوامل اصلی تشکیل‌دهنده با مدل‌های خودرگرسیون میانگین متحرک انباشته کلاسیک است. بر این اساس، فرایند مدل‌سازی روش پیشنهادی را می‌توان در سه گام اصلی بیان کرد: ۱. تجزیه‌ی سری زمانی مورد مطالعه به اجزاء اصلی تشکیل‌دهنده؛ ۲. پیش‌بینی هریک از اجزاء تشکیل‌دهنده به‌صورت تکی؛ ۳. ترکیب پیش‌بینی‌های تکی با یکدیگر.

- **مرحله‌ی اول (تجزیه به عوامل اصلی):** در این مرحله با استفاده از تکنیک تجزیه‌ی تجمعی به عوامل اصلی تشکیل‌دهنده، سری زمانی مورد مطالعه به تعدادی از اجزاء تشکیل‌دهنده‌ی ساده‌تر تجزیه می‌شود. در حالت کلی، تعداد این اجزاء تشکیل‌دهنده به میزان پیچیدگی موجود در داده‌ها و تعداد ساختارهای موجود در آنها بستگی دارد. بر این اساس، هرچه سری زمانی مورد مطالعه پیچیده‌تر باشد، تعداد اجزاء تشکیل‌دهنده‌ی آن نیز بیشتر خواهد بود. به‌عبارت دیگر، فرایند تجزیه‌ی داده‌ها در تکنیک‌های تجزیه به عوامل اصلی تشکیل‌دهنده تا زمانی ادامه پیدا خواهد کرد که آخرین جزء تشکیل‌دهنده، کم‌ترین پیچیدگی ممکنه را داشته باشد. بر این اساس، یک جزء تشکیل‌دهنده، تابعی است که دو شرط اساسی زیر را برآورده کند:

۱. در مجموعه‌ی داده‌ها تعداد اکسترمم‌ها و تعداد نقاط صفر باید برابر، یا حداکثر یک واحد اختلاف داشته باشند؛

۲. در هر نقطه، مقدار میانگین بیشینه‌ی محلی و کمینه‌ی محلی باید برابر با صفر باشد. به‌عبارت دیگر، میانگین پوش بالا و پوش پایین در هر نقطه باید برابر با صفر باشد.

بر این اساس، اگر سری زمانی مورد مطالعه را با $x(t)$ نمایش دهیم، در ابتدای فرایند تجزیه به عوامل اصلی، باید تمامی نقاط اکسترمم داده‌ها را شناسایی کرد. سپس، بر اساس تابع مشتق، نوع هریک از اکسترمم‌های معین شده، مشخص می‌شود. به‌عبارت دیگر، در این مرحله می‌بایست تمامی نقاط بیشینه و همچنین نقاط کمینه‌ی تابع مترتب به سری زمانی مورد مطالعه به تفکیک مشخص شود. در ادامه، تابعی به نام «تابع اسپلاین» به نقاط بیشینه و همچنین نقاط کمینه برازش داده می‌شود و به این ترتیب پوش بالای داده‌ها و پوش پایین داده‌ها به دست می‌آید. سپس میانگین پوش بالا و پوش پایین محاسبه می‌شود؛ اگر میانگین پوش بالا و پوش پایین را با m_1 نمایش دهیم، کاندیدای اولین مؤلفه‌ی اصلی تشکیل‌دهنده که اصولاً با نماد h_1 بدان اشاره می‌شود را می‌توان از تفاضل بین داده و میانگین پوش بالا و پایین، مطابق رابطه‌ی ۱ به دست آورد:

$$h_1 = x(t) - m_1 \qquad (1)$$

حال به‌صورت ایده‌آل، خود h_1 ممکن است یک مؤلفه‌ی اصلی تشکیل‌دهنده باشد، اما از آن‌جا که در عمل این فرایند ممکن است در مورد بعضی از داده‌ها، نقاط اکسترمم جدید تولید کند، برای برقراری شرط دوم و متقارن‌تر شدن h_1 غالباً باید این مرحله بارها و بارها تکرار شود. بر این اساس و مطابق با نمادهای قبلی، در مرحله‌ی تکرار اول، h_1 به‌عنوان داده‌ی ورودی در نظرگرفته می‌شود و سپس مراحل بالا مجدداً روی آن اجرا می‌شود تا h_{11} مطابق رابطه‌ی ۲ به دست آید:

$$h_{11} = h_1 - m_{11} \qquad (2)$$

این مراحل آن‌قدر تکرار می‌شود تا h_{1k} بتواند شروط یاد شده را مرتفع سازد. این مراحل تکراری را اصطلاحاً «غربال‌گری» می‌نامند. در حالت کلی، برای تضمین این که h_{1k} شروط مذکور را به‌طور کامل برآورده کند، باید شرط توقفی برای فرایند غربال‌گری در نظرگرفته شود. بدین منظور، انحراف معیار بین نتیجه‌ی دو غربال کردن متوالی، مطابق رابطه‌ی ۳ به عنوان شرط توقف در نظرگرفته می‌شود:

$$SD = \sum_{t=1}^{T} \left[\frac{|(h_{1(k-1)}(t) - h_{1k}(t)|^2}{h_1^2(k-1)^{(t)}} \right] \qquad (3)$$

اصولاً زمانی که انحراف معیار بین ۰٫۲ و ۰٫۳ باشد، فرایند غربال‌گری متوقف می‌شود. بدین ترتیب، h_{1k} به‌عنوان اولین مؤلفه‌ی اصلی تشکیل‌دهنده (c_1) در نظر گرفته می‌شود:

$$c_1 = h_{1k} = h_{1(k-1)} - m_{1k} \qquad (4)$$

در حالت کلی، c_1 حاوی کوچکترین مقیاس‌ها یا بزرگترین فرکانس‌هاست. حال، با تفاضل‌گیری c_1 از داده‌ی اولیه، مقدار باقیمانده‌ی مرحله‌ی اول (r_1) مطابق رابطه‌ی ۵ به دست می‌آید.

$$r_1 = x(t) - c_1 \qquad (5)$$

بر همین اساس، مقدار باقیمانده به‌عنوان داده‌ی جدید در نظرگرفته می‌شود و دوباره مراحل غربال‌گری روی آن صورت می‌گیرد تا مؤلفه‌ی تشکیل‌دهنده‌ی اصلی بعدی از آن استخراج شود. این فرایند برای باقیمانده‌های بعدی نیز تکرار می‌شود تا تمامی مؤلفه‌های اصلی از آن استخراج شود:

$$r_2 = r_1 - c_2,\ r_2 = r_2 - c_3, \ldots\ldots, r_n = r_{n-1} - c_n \qquad (6)$$

این فرایند تا زمانی که آخرین باقیمانده حداکثر دو اکسترمم داشته باشد، ادامه خواهد یافت. بر این اساس داریم:

$$x(t) = \sum_{i=1}^{n}(c_i) + r_n \qquad (7)$$

رابطه‌ی ۷ نشان می‌دهد که سری زمانی مورد مطالعه در این فرایند به n مؤلفه‌ی اصلی $(c_i \ \ i = 1, 2, \ldots, n)$ و یک مقدار باقیمانده (r_n) تجزیه خواهد شد. در نهایت نیز به‌منظور رفع مشکل مدهای اختلاطی در فرایند تجزیه‌ی کلاسیک، کافی است تا اصلاحات زیر به‌ترتیب روی داده‌ها اجراء شوند:

۱. اضافه کردن یک جزء اختلال تصادفی به داده‌های اولیه؛

۲. تجربه کردن داده‌ها همراه با جزء اختلال تصادفی اضافه شده؛

۳. تکرار چندین باره‌ی مراحل اول و دوم با اجزاء اختلال تصادفی متفاوت؛

۴. به دست آوردن میانگین مؤلفه‌های مشابه از مرحله‌ی سوم و در نظر گرفتن میانگین آنها به‌عنوان مؤلفه‌های نهایی.

- **مرحله‌ی دوم (مدل‌سازی عوامل اصلی):** پس از این که سری زمانی به روش تجزیه‌ی تجمعی به مؤلفه‌های اصلی ساده‌تر تجزیه شد، هرکدام از مؤلفه‌ها به صورت جداگانه و با استفاده از روش خودرگرسیون میانگین متحرک انباشته پیش‌بینی می‌شود. در این مرحله از مدل‌سازی، بر خلاف قبل، این اطمینان وجود دارد که این مؤلفه‌ها به‌ساده‌ترین حالت ممکن و با کمترین پیچیدگی هستند، لذا روش خودرگرسیون میانگین متحرک انباشته می‌تواند بیشترین عملکرد ممکن را داشته باشد. حال برای هریک از این مؤلفه‌ها می‌توان مطابق رابطه‌ی ۸ پیش‌بینی کرد:

$$\phi(B)(1-B)^d(c_{it}-\mu)=\theta(B)a_t \qquad (8)$$

به‌طوری که $\phi(B)=1-\phi_1B-\phi_2B^2-.....-\phi_pB^p$ و $\theta(B)=1-\theta_1B-\theta_2B^2-.....-\theta_qB^q$ و p و q بوده، B یک عملگر پسرو، اعداد q,d,p اعداد صحیح و $\{c_{it}\}$ بیانگر مقادیر مشاهده شده‌ی مؤلفه‌ی اصلی iام هستند. به‌طورکلی، چنین فرض می‌شود که جمله‌ی خطای خالص a_t متغیری تصادفی با توزیع نرمال با میانگین صفر و واریانس σ^2 و مستقل از مشاهدات است. همچنین ریشه‌های معادله‌ی $\phi(C)=0$ و $\theta(C)=0$ همگی بزرگتر از ۱ هستند. در این مرحله، یکی از مهمترین قسمت‌های مدل‌سازی خودرگرسیون میانگین متحرک انباشته، تعیین صحیح مرتبه‌ی مدل (p,q) است. متدولوژی بوکس ـ جنکینز[۲] سه مرحله‌ی تکراری ـ شناسایی مدل، تخمین پارامترها و کنترل تشخیصی ـ است. ایده‌ی اصلی شناسایی مدل نیز بر این اصل استوار است که اگر یک سری زمانی از یک فرایند خودرگرسیون میانگین متحرک انباشته تولید شده باشد، باید برخی خواص خودهمبستگی نظری را داشته باشد. لذا از تطبیق الگوهای خودهمبسته‌ی تجربی با الگوهای خودهمبسته‌ی نظری، غالباً شناسایی یک یا چند مدل بالقوه برای سری‌های زمانی امکان‌پذیر خواهد بود. پیشنهاد باکس ـ جنکینز[۲] استفاده از تابع خودهمبستگی و خودهمبستگی جزئی داده‌های نمونه به‌عنوان ابزار پایه برای شناسایی مرتبه‌ی مدل است.

- **مرحله‌ی سوم (ترکیب پیش‌بینی ها):** در این مرحله برای دستیابی به پیش‌بینی های نهایی، تمامی پیش‌بینی‌های انفرادی مربوط به هریک از مؤلفه‌های اصلی تشکیل‌دهنده با یکدیگر ترکیب می‌شود. بر این اساس خواهیم داشت:

$$\hat{F}_t=\sum_{i=1}^{n}\hat{\phi}(B)(1-B)^d(c_{it}-\mu)-\hat{\theta}(B) \qquad (9)$$

به قسمی که $\hat{F}_i$ پیش‌بینی نهایی مدل پیشنهادی است.

۳. داده‌ها، معیارهای سنجش و تست‌های استفاده شده

۱.۳. داده‌های استفاده شده

در این نوشتار، به‌منظور ارزیابی عملکرد روش پیشنهادی در مقایسه با روش خودرگرسیون میانگین متحرک انباشته و همچنین سایر روش‌های معمول پیش‌بینی، دو مجموعه داده‌ی معیار مربوط به قیمت نفت خام تگزاس و قیمت نفت خام برنت در نظر گرفته شده است. این دو مجموعه داده شامل قیمت‌های هفتگی نفت خام تگزاس از اول ژانویه‌ی سال ۱۹۸۶ تا یازدهم جولای ۲۰۱۴ و همچنین قیمت‌های هفتگی نفت خام برنت از سوم ژانویه‌ی سال ۱۹۸۷ تا یازدهم جولای ۲۰۱۴ است. نمودار قیمت‌های هفتگی نفت خام تگزاس برحسب دلار برگالن در شکل ۱ نشان داده شده است. هریک از این دو مجموعه داده به‌منظور مدل‌سازی و همچنین ارزیابی عملکرد مدل‌های پیش‌بینی به‌ترتیب به دو دسته داده‌های آموزش و داده‌های آزمایش تقسیم شده‌اند.

در هر تقسیم‌بندی ۸۰٪ از داده‌ها به‌منظور آموزش و ۲۰٪ باقیمانده‌ی داده‌ها برای آزمایش مورد استفاده قرار گرفته است. بدین ترتیب، از ۱۴۸۹ داده‌ی مربوط به قیمت نفت خام تگزاس، ۱۱۹۱ داده به‌عنوان داده‌ی آموزش و ۲۹۸ داده‌ی نهایی به‌عنوان داده‌ی آزمون در نظر گرفته شده است. بر همین اساس، از ۱۴۱۸ داده مربوط به قیمت نفت خام برنت، ۱۱۳۴ داده به‌عنوان داده‌ی آموزش و ۲۸۴ داده‌ی نهایی به‌عنوان داده‌ی آزمون در نظر گرفته شده است.

شکل ۱. قیمت‌های هفتگی نفت خام تگزاس.

۲.۳. معیارهای سنجش استفاده شده

در این مقاله، همچنین پنج معیار سنجش عملکرد برای محاسبه‌ی میزان دقت پیش‌بینی‌های هریک از مدل‌های زیر در نظرگرفته شده است. این معیارهای عملکردی شامل: میانگین قدرمطلق خطا (MAE)، میانگین مربعات خطا (MSE)، میانگین قدرمطلق درصدی خطا (MAPE)، مجذور میانگین مربعات خطا (RMSE) و نیکویی برازش (R^2) هستند که طبق روابط ۱۰ تا ۱۴ محاسبه می‌شوند:

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|A_i-P_i| \qquad (10)$$

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(A_i-P_i)^2 \qquad (11)$$

$$MAPE=\frac{1}{N}\sum_{i=1}^{N}\frac{|A_i-P_i|}{|A_i|} \qquad (12)$$

$$RMSE=\left(\frac{1}{N}\sum_{i=1}^{N}(A_i-P_i)^2\right)^{\frac{1}{2}} \qquad (13)$$

$$R^2=1-\frac{\sum_{i=1}^{N}(A_i-P_i)^2}{\sum_{i=1}^{N}(A_i-M)^2} \qquad (14)$$

به‌قسمی که A_i مقدار واقعی سری زمانی برای مشاهده‌ی iام، P_i مقدار پیش‌بینی شده توسط مدل برای مشاهده‌ی iام، N تعداد داده‌ها و M میانگین داده‌های سری زمانی مورد مطالعه است.

۳.۳. آزمون‌های استفاده شده

در این نوشتار همچنین از دو آزمون برای صحت‌سنجی و بررسی کفایت مدل پیشنهادی مورد استفاده قرار گرفته است. این آزمون‌ها عبارت‌اند از: آزمون جارک ـ برا[۳] و آزمون Q [۴] هستند که در ادامه شرح مختصری برای آزمون جارک ـ برا ارائه خواهیم داد.

۱.۳.۳. آزمون جارک ـ برا

آزمون جارک ـ برا یکی از آزمون‌های نیکویی برازش است که اساساً به‌منظور بررسی نرمال بودن باقیمانده‌ها در سری‌های زمانی و رگرسیون مورد استفاده قرار می‌گیرد. اساس عملکرد این آزمون بر تقارن چولگی[۵] و کشیدگی[۶] یک تابع توزیع بنا نهاده شده است. این آزمون در سال ۱۹۷۸ و توسط دو اقتصاددان مکزیکی (کارلوس جارک) و آمریکایی (آنیل برا) معرفی شده است. چنان که گفته شد، این آزمون بر اساس

اندازه‌های مربوط به تقارن عمودی و افقی توزیع نرمال عمل میکند که توسط چولگی و کشیدگی اندازه‌گیری می‌شوند. بر این اساس، فرض کنید که S میزان چولگی نمونه و K نیز میزان کشیدگی نمونه برای مشاهدات $r_1, r_2, \ldots, r_n$ باشد. بر این اساس خواهیم داشت که:

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{1/n \sum_{i=1}^{n}(r_i - \bar{r})^3}{(1/n \sum_{i=1}^{n}(r_i - \bar{r})^2)^{3/2}} \quad (15)$$

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{1/n \sum_{i=1}^{n}(r_i - \bar{r})^4}{(1/n \sum_{i=1}^{n}(r_i - \bar{r})^2)^2} \quad (16)$$

بدین ترتیب آماره‌ی آزمون جارک برحسب این دو پارامتر تعریف می‌شود:

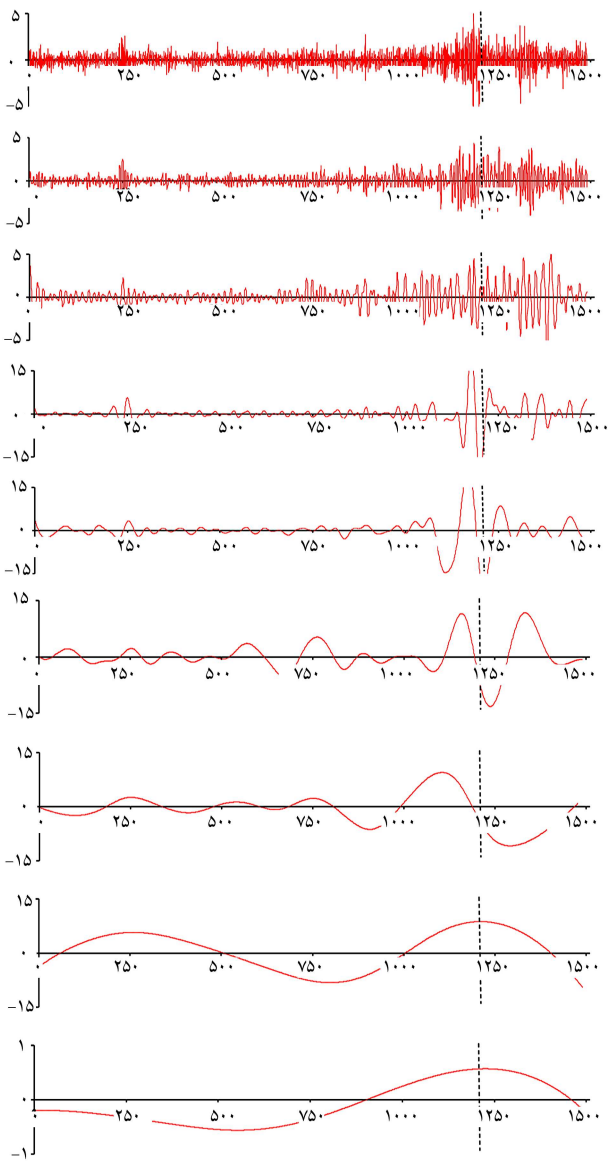
$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \quad (17)$$

که در آن μ_3 و μ_4 به ترتیب گشتاورهای مرکزی مرتبه‌ی سوم و چهارم و $\hat{\mu}_3$ و $\hat{\mu}_4$ نیز برآوردهای این گشتاورها هستند.

۴. نتایج عددی به‌کارگیری روش پیشنهادی در پیش‌بینی قیمت نفت خام

در این قسمت، فرایند مدل‌سازی روش پیشنهادی به‌منظور پیش‌بینی قیمت نفت خام تشریح شده است. به‌منظور تشریح بهتر مراحل اجرایی روش پیشنهادی، هریک از این مراحل به‌عنوان نمونه بر روی مجموعه داده‌ی معیار اول، یعنی قیمت هفتگی نفت خام تگزاس تشریح شده‌اند. توضیحاً در این نوشتار برای هر دو مجموعه داده، تعداد تجمع برابر با ۱۰۰ و انحراف معیار جمله اختلال وارد شده به مدل برابر با ۰/۲ در نظر گرفته شده است. بر این اساس و مطابق روش پیشنهادی، خواهیم داشت:

- **مرحله‌ی اول (تجزیه به عوامل اصلی):** مطابق فرایند مدل پیشنهادی، در مرحله‌ی اول، عوامل اصلی تشکیل‌دهنده‌ی سری زمانی مورد مطالعه با استفاده از تکنیک تجزیه‌ی تجمعی به عوامل اصلی به تعدادی از اجزاء تشکیل‌دهنده‌ی ساده‌تر تجزیه می‌شود. چنان که اشاره شد تعداد این اجزاء تشکیل‌دهنده به میزان پیچیدگی موجود در داده‌ها و تعداد ساختارهای موجود در آنها بستگی دارد. نتایج حاصل از تجزیه‌ی داده‌های مربوط به قیمت نفت خام تگزاس نشان می‌دهد که این سری زمانی از ده مؤلفه‌ی اصلی تشکیل شده است. این تعداد مؤلفه‌ی اصلی مؤید این موضوع است که مطابق آن‌چه که در ادبیات موضوع نیز بدان اشاره شد، قیمت‌های نفت خام از پیچیدگی بالایی برخوردار است. قیمت‌های نفت خام معمولاً به‌عنوان سری‌های زمانی غیرخطی و غیرایستا در ادبیات موضوع در نظرگرفته می‌شود که تحت تأثیر فاکتورهای زیادی نیز هستند. نمودارهای مربوط به این ۹ مؤلفه‌ی اصلی به‌ترتیب از مؤلفه‌ی اول تا نهم در شکل ۲ ارائه شده است. چنان که مشاهده می‌شود این مؤلفه‌های اصلی دارای بیشترین پیچیدگی در مؤلفه‌ی اول و کمترین پیچیدگی در مؤلفه‌ی نهم هستند. به‌عبارت دیگر، داده‌ها به‌ترتیب از بالاترین فرکانس به کمترین فرکانس تجزیه می‌شود. در حالت کلی، می‌توان انتظار داشت که میزان پیش‌بینی‌پذیری به‌ترتیب از مؤلفه‌ی اول تا مؤلفه‌ی نهم، افزایشی غیراکید داشته باشد.
- **مرحله‌ی دوم (مدلسازی عوامل اصلی):** در این مرحله، مؤلفه‌های اصلی به‌دست آمده از روش تجزیه‌ی تجمعی برای قیمت نفت خام تگزاس به‌صورت جداگانه توسط روش خودرگرسیون میانگین متحرک انباشته پیش‌بینی می‌شود.



شکل ۲. تجزیه‌ی سری زمانی قیمت‌های نفت خام تگزاس به مؤلفه‌های اصلی.

بدین منظور، مرتبه‌ی مناسب برای هریک از این مؤلفه‌های اصلی با استفاده از توابع خودهمبستگی و خودهمبستگی جزئی تشخیص داده شده و مدل نهایی با استفاده از معیار آکائیک تعیین می‌شود. در انتها نیز مقادیر مترتب به هر یک از پارامترهای مدل با استفاده از نرم‌افزار ایویوز محاسبه شده است. نتایج عملکردی روش خودرگرسیون میانگین متحرک انباشته برای هریک از این مؤلفه‌های اصلی در مجموعه داده‌های آموزش در جدول ۱ ارائه شده است. همچنین نتایج عملکردی حاصل از روش خودرگرسیون میانگین متحرک انباشته برای هریک از این مؤلفه‌های اصلی در مجموعه داده‌های آزمون در جدول ۲ ارائه شده است. نتایج حاصل از این قسمت نشان می‌دهد که رابطه‌ی غیراکید افزایشی در دقت پیش‌بینی‌های مربوط به هریک از مؤلفه‌ها در داده‌های آموزش برقرار است. به‌طور مثال، میانگین قدر مطلق خطای مؤلفه‌ی اول برابر با ۰/۵۰۹ است، در صورتی که این مقدار برای مؤلف های دوم تا نهم به‌ترتیب برابر با ۰/۱۹۱، ۰/۰۶۲، ۰/۰۲۷، ۰/۰۱۰، ۰/۰۰۲، ۰/۰۰۰، ۰/۰۰۰، ۰/۰۰۰ بوده است.

جدول ۱. عملکرد روش خودرگرسیون میانگین متحرک انباشته برای هر یک از مؤلفه‌های اصلی (داده‌های آموزش).

	معیارهای سنجش عملکرد				
	میانگین قدرمطلق		میانگین	جذر میانگین	نیکویی برازش
	خطا	درصدی خطا	مربعات خطا	مربعات خطا	
مؤلفه‌ی اصلی اول	۰٫۵۰۹	٪۲۸۰٫۷۰۰	۰٫۵۰۵	۰٫۷۱۱	۰٫۲۷۰
مؤلفه‌ی اصلی دوم	۰٫۱۹۱	٪۷۶٫۱۸۰	۰٫۰۷۲	۰٫۲۶۹	۰٫۸۷۲
مؤلفه‌ی اصلی سوم	۰٫۰۶۲	٪۴٫۴۳۳	۰٫۰۰۸	۰٫۰۸۹	۰٫۹۹۱
مؤلفه‌ی اصلی چهارم	۰٫۰۲۷	٪۱۷٫۲۱۶	۰٫۰۰۱	۰٫۰۳۷	۰٫۹۹۹
مؤلفه‌ی اصلی پنجم	۰٫۰۱۰	٪۰٫۳۶۲	۰٫۰۰۰	۰٫۰۱۳	۰٫۹۹۹
مؤلفه‌ی اصلی ششم	۰٫۰۰۲	٪۵٫۲۱۳	۰٫۰۰۰	۰٫۰۰۳	۰٫۹۹۹
مؤلفه‌ی اصلی هفتم	۰٫۰۰۰	٪۰٫۰۰۹	۰٫۰۰۰	۰٫۰۰۰	۱٫۰۰۰
مؤلفه‌ی اصلی هشتم	۰٫۰۰۰	٪۰٫۰۰۵	۰٫۰۰۰	۰٫۰۰۰	۱٫۰۰۰
مؤلفه‌ی اصلی نهم	۰٫۰۰۰	٪۰٫۰۰۱	۰٫۰۰۰	۰٫۰۰۰	۱٫۰۰۰

جدول ۲. عملکرد روش خودرگرسیون میانگین متحرک انباشته برای هر یک از مؤلفه‌های اصلی (داده‌های آزمون).

	معیارهای سنجش عملکرد				
	میانگین قدرمطلق		میانگین	جذر میانگین	نیکویی برازش
	خطا	درصدی خطا	مربعات خطا	مربعات خطا	
مؤلفه‌ی اصلی اول	۰٫۸۴۳	٪۲۲۷٫۷۲۳	۱٫۱۳۱	۱٫۰۶۴	۰٫۳۰۹
مؤلفه‌ی اصلی دوم	۰٫۳۸۴	٪۳۳٫۳۳۲	۰٫۲۲۲	۰٫۴۷۱	۰٫۹۰۱
مؤلفه‌ی اصلی سوم	۰٫۱۵۱	٪۶۳٫۹۲۰	۰٫۰۳۶	۰٫۱۹۰	۰٫۹۹۲
مؤلفه‌ی اصلی چهارم	۰٫۰۶۳	٪۰٫۸۰۳	۰٫۰۰۶	۰٫۰۸۱	۰٫۹۹۹
مؤلفه‌ی اصلی پنجم	۰٫۰۱۶	٪۰٫۴۴۴	۰٫۰۰۰	۰٫۰۲۲	۰٫۹۹۹
مؤلفه‌ی اصلی ششم	۰٫۰۰۵	٪۰٫۱۹۸	۰٫۰۰۰	۰٫۰۰۶	۰٫۹۹۹
مؤلفه‌ی اصلی هفتم	۰٫۰۴۸	٪۲۸٫۹۳۰	۰٫۰۹۱	۰٫۳۰۱	۰٫۹۹۳
مؤلفه‌ی اصلی هشتم	۰٫۰۰۰	٪۰٫۰۲۲	۰٫۰۰۰	۰٫۰۰۰	۱٫۰۰۰
مؤلفه‌ی اصلی نهم	۰٫۰۰۰	٪۰٫۰۰۷	۰٫۰۰۰	۰٫۰۰۰	۱٫۰۰۰

جدول ۳. عملکرد روش خودرگرسیون میانگین متحرک انباشته (داده‌های آموزش و آزمون).

	مجموعه داده‌ها	
	آموزش	آزمون
میانگین قدرمطلق خطا	۰٫۷۴۸	۱٫۱۳۶
میانگین قدرمطلق درصدی خطا	٪۲٫۷۵۲	٪۱٫۵۲۵
میانگین مربعات خطا	۱٫۰۰۰	۲٫۰۹۱
جذر میانگین مربعات خطا	۱٫۰۰۰	۱٫۴۴۶
نیکویی برازش	۰٫۹۶۰	۰٫۹۴۷

- **مرحله‌ی سوم (ترکیب پیش‌بینی‌ها):** در این مرحله، پیش‌بینی‌های صورت گرفته برای هریک از مؤلفه‌های اصلی در مرحله‌ی قبل، مطابق رابطه‌ی ۹ با یکدیگر ترکیب می‌شوند تا پیش‌بینی‌های نهایی به دست آید. نتایج حاصل از این مرحله در پیش‌بینی قیمت نفت خام تگزاس در جدول ۳ آورده شده است. همچنین نمودار مقادیر واقعی در مقابل مقادیر پیش‌بینی شده برای کل داده‌ها و داده‌های آزمون به‌ترتیب در شکل‌های ۳ و ۴ آورده شده است.

۵. مقایسه‌ی عملکرد روش پیشنهادی با سایر روش‌ها

در این قسمت عملکرد روش پیشنهادی در پیش‌بینی قیمت نفت خام تگزاس و همچنین پیش‌بینی قیمت نفت خام برنت با روش خودرگرسیون میانگین متحرک انباشته و سایر روش‌های هوشمند پیش‌بینی همچون پرسپترون‌های چندلایه، شبکه‌های عصبی شعاعی‌محور و شبکه‌های عصبی رگرسیون عمومی مقایسه شده است. نتایج عملکرد روش‌های مختلف در پیش‌بینی قیمت نفت خام تگزاس برای مجموعه داده‌های آزمون در جدول ۴ آورده شده است. همچنین میزان بهبود روش پیشنهادی در تقابل با هریک از روش‌های مذکور در پیش‌بینی این سری زمانی در جدول ۵ ارائه شده است. نتایج حاصله نشان می‌دهد که روش پیشنهادی توانسته عملکرد روش خودرگرسیون میانگین متحرک انباشته کلاسیک را به میزان ٪۴۶٫۹۹، ٪۷۳٫۷۹، ٪۴۵٫۸۳ و ٪۴۸٫۸۰ به‌ترتیب در میانگین قدرمطلق خطا، میانگین مربعات خطا، میانگین قدرمطلق درصدی خطا و مجذور میانگین مربعات خطا بهبود بخشد.

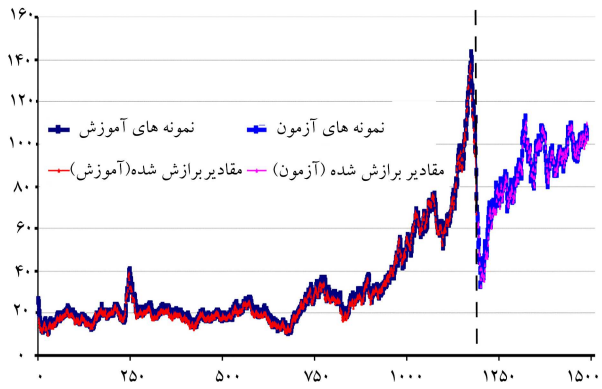
همچنین این نتایج حاکی است که روش پیشنهادی دقت پیش‌بینی‌های سایر روش‌های هوشمند را نیز افزایش داده است. به‌طور مثال، میانگین قدر مطلق خطای

جدول ۴. عملکرد روش‌های متفاوت در پیش‌بینی قیمت نفت خام تگزاس (داده‌های آزمون).

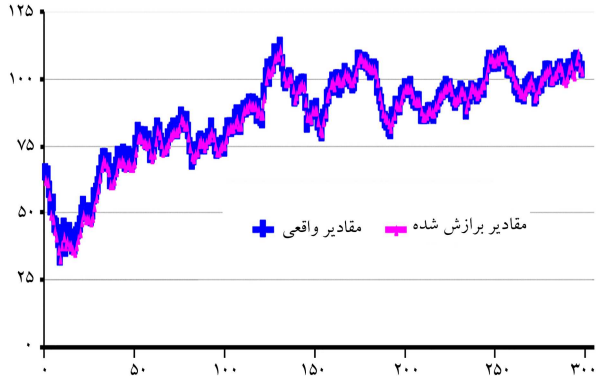
	معیارهای سنجش عملکرد			
	میانگین قدرمطلق		میانگین	جذر میانگین
	خطا	درصدی خطا	مربعات خطا	مربعات خطا
خودرگرسیون میانگین متحرک انباشته	۲٫۱۴۳	٪۲٫۸۱۵	۷٫۹۷۷	۲٫۸۲۴
پرسپترون‌های چندلایه	۲٫۱۳۹	٪۲٫۸۰۲	۷٫۰۱۵	۲٫۶۴۹
شبکه‌های عصبی شعاعی‌محور	۲٫۱۴۱	٪۲٫۸۰۷	۷٫۴۲۳	۲٫۷۲۵
شبکه‌های عصبی رگرسیون عمومی	۲٫۱۴۱	٪۲٫۸۰۷	۷٫۴۲۳	۲٫۷۲۵
روش پیشنهادی	۱٫۱۳۶	٪۱٫۵۲۵	۲٫۰۹۱	۱٫۴۴۶

جدول ۵. درصد بهبود روش پیشنهادی در تقابل با سایر روش‌ها در پیش‌بینی قیمت نفت خام تگزاس (داده‌های آزمون).

	معیارهای سنجش عملکرد			
	میانگین قدرمطلق		میانگین	جذر میانگین
	خطا	درصدی خطا	مربعات خطا	مربعات خطا
خودرگرسیون میانگین متحرک انباشته	٪۴۶٫۹۹	٪۴۵٫۸۳	٪۷۳٫۷۹	٪۴۸٫۸۰
پرسپترون‌های چندلایه	٪۴۶٫۸۹	۴۵٫۵۷	٪۷۰٫۱۹	٪۴۵٫۴۱
شبکه‌های عصبی شعاعی‌محور	٪۴۶٫۹۴	٪۴۵٫۶۷	٪۷۱٫۸۳	٪۴۶٫۹۴
شبکه‌های عصبی رگرسیون عمومی	٪۴۶٫۹۴	٪۴۵٫۶۷	٪۷۱٫۸۳	٪۴۶٫۹۴



شکل ۳. داده‌های قیمت نفت خام تگزاس در مقابل مقادیر پیش‌بینی شده (کل داده‌ها).



شکل ۴. داده‌های قیمت نفت خام تگزاس در مقابل مقادیر پیش‌بینی شده (داده‌های تست).

روش‌های پرسپترون‌های چندلایه، شبکه‌های عصبی شعاعی‌محور و شبکه‌های عصبی رگرسیون عمومی به‌ترتیب برابر با ۲٫۱۳۹، ۲٫۱۴۱ و ۲٫۱۴۱ بوده است که به‌ترتیب به‌میزان ٪۴۶٫۸۹، ٪۴۶٫۹۴ و ٪۴۶٫۹۴ توسط روش پیشنهادی بهبود یافته است.

نتایج عملکرد روش‌های مختلف در پیش‌بینی قیمت نفت خام برنت نیز در جدول ۶ آورده شده است. همچنین میزان بهبود روش پیشنهادی در تقابل با هریک از روش‌های مذکور در پیش‌بینی این سری زمانی در جدول ۷ ارائه شده است. نتایج حاصله نشان می‌دهد که روش پیشنهادی توانسته است عملکرد روش خودرگرسیون میانگین متحرک انباشته کلاسیک را به‌میزان ٪۵۹٫۰۲، ٪۸۴٫۱۷، ٪۵۸٫۸۶، ٪۶۰٫۲۴ به‌ترتیب در میانگین قدر مطلق خطا، میانگین مربعات خطا، میانگین قدرمطلق درصدی خطا و مجذور میانگین مربعات خطا بهبود بخشد. همچنین این نتایج نشان می‌دهند که روش پیشنهادی دقت پیش‌بینی‌های سایر روش‌های هوشمند را نیز افزایش داده است. به‌طور مثال، میانگین قدرمطلق خطای روش‌های پرسپترون‌های چندلایه، شبکه‌های عصبی شعاعی‌محور و شبکه‌های عصبی رگرسیون عمومی به‌ترتیب برابر با ۱٫۹۸۱، ۱٫۹۹۷ و ۱٫۹۹۵ بوده است که به‌ترتیب به‌میزان ٪۵۸٫۵۱، ٪۵۸٫۸۴ و ٪۵۸٫۸۰ کمتر از روش پیشنهادی بوده است. همچنین به‌منظور بررسی کفایت مدل پیشنهادی نتایج مربوط به آزمون جارک ـ برا و آزمون Q مربوط به باقیمانده‌های مدل پیشنهادی برای مجموعه داده قیمت نفت خام برنت به ترتیب در جداول ۸ و ۹ آورده شده است. نتایج حاصل از بررسی باقیمانده‌های روش پیشنهادی نشان می‌دهد که این روش به خوبی توانسته است الگوهای موجود در داده‌ها را مدل‌سازی کند. به عبارت دیگر، وجود الگوهای قابل مدل‌سازی خطی در باقیمانده‌های روش پیشنهادی معنی‌دار نیست.

۶. نتیجه‌گیری

دقت پیش‌بینی‌ها به ویژه پیش‌بینی سری‌های زمانی از مهمترین چالش‌های فراروی تصمیم‌گیرندگان و مدیران عملیاتی و مالی است. اما ادبیات موضوع نشان می‌دهد

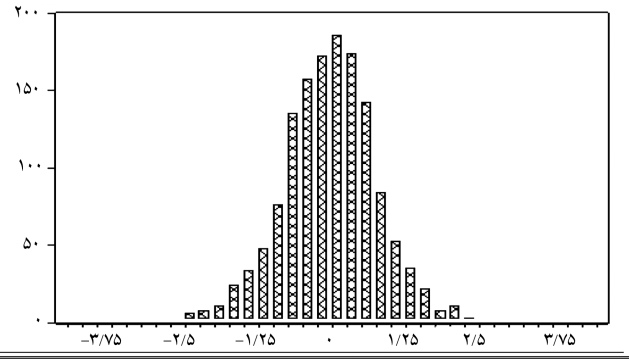
جدول ۶. عملکرد روش‌های متفاوت در پیش‌بینی قیمت نفت خام برنت (داده‌های آزمون).

	معیارهای سنجش عملکرد			
	میانگین قدرمطلق		میانگین	جذر میانگین
	خطا	درصدی خطا	مربعات خطا	مربعات خطا
خودرگرسیون میانگین متحرک انباشته	۲/۰۰۶	٪۱/۹۴۷	۶/۸۸۱	۲/۶۲۳
پرسپترون‌های چندلایه	۱/۹۸۱	٪۱/۹۲۴	۶/۷۰۴	۲/۵۸۹
شبکه‌های عصبی شعاعی‌محور	۱/۹۹۷	٪۱/۹۳۹	۶/۸۲۱	۲/۶۱۲
شبکه‌های عصبی رگرسیون عمومی	۱/۹۹۵	٪۱/۹۳۷	۶/۸۱۱	۲/۶۱۰
روش پیشنهادی	۰/۸۲۲	٪۰/۸۰۱	۱/۰۸۹	۱/۰۴۳

جدول ۷. درصد بهبود روش پیشنهادی در تقابل با سایر روش‌ها در پیش‌بینی قیمت نفت خام برنت (داده‌های آزمون).

	معیارهای سنجش عملکرد			
	میانگین قدرمطلق		میانگین	جذر میانگین
	خطا	درصدی خطا	مربعات خطا	مربعات خطا
خودرگرسیون میانگین متحرک انباشته	۵۹/۰۲	۵۸/۸۶	۸۴/۱۷	۶۰/۲۴
پرسپترون‌های چندلایه	۵۸/۵۱	۵۸/۳۷	۸۳/۷۶	۵۹/۷۱
شبکه‌های عصبی شعاعی‌محور	۵۸/۸۴	۵۸/۶۹	۸۴/۰۳	۶۰/۰۷
شبکه‌های عصبی رگرسیون عمومی	۵۸/۸۰	۵۸/۶۵	۸۴/۰۱	۶۰/۰۴

جدول ۸. نتایج آزمون جارک ـ برا برای آزمایش نرمال بودن باقیمانده‌ها (داده‌های هفتگی قیمت نفت خام برنت).



سری زمانی: قیمت نفت خام برنت
تعداد مشاهدات: ۱۴۱۸
میانگین: ۰/۰۰۰۷۱۵
میانه: ۰/۰۳۹۷۵۲
بیشینه: ۴/۳۹۵۳۹۰
کمینه: -۴/۲۵۲۲۵۸
انحراف معیار: ۰/۸۳۶۰۱۹
چولگی: -۰/۲۲۳۲۰۲
کشیدگی: ۴/۷۵۹۹۶۳
آماره جارک- برا: ۱۹۲/۳۱۰
احتمال ۰/۰۰۰۰۰۰

که بهبود دقت پیش‌بینی‌ها بالاخص در محیط‌های پیچیده و مبهم امروزی اصلاً کار راحتی نیست. این دلیل اصلی اشتیاق محققان و پژوهش‌گران به‌منظور بسط و گسترش مدل‌های موجود، یا ارائه‌ی مدل‌های جدید پیش‌بینی است که حتی‌الامکان قابلیت بالاتری در مدل‌سازی پیچیدگی‌ها و ابهامات موجود در داده‌ها و نتیجتاً حصول نتایج دقیق‌تر را داشته باشند. در این مقاله، به‌منظور مدل‌سازی هرچه دقیق‌تر سری‌های زمانی مخصوصاً در محیط‌های مالی که دارای بیشترین سطح پیچیدگی بوده و همچنین به‌صورت هم‌زمان شامل چندین ساختار متفاوت هستند، یک روش ترکیبی پیش‌بینی ارائه شده است. هدف نهایی مدل پیشنهادی مرتفع ساختن محدودیت مدل‌سازی ساختارهای چندگانه و پیچیده در مدل‌های خودرگرسیون میانگین متحرک انباشته است. بدین منظور، در مرحله‌ی اول از مدل پیشنهادی سری‌های زمانی مورد مطالعه با استفاده از تکنیک‌های تجزیه‌ی تجمعی به تعدادی از مؤلفه‌های اصلی که اساساً ساختارهای کمتر و پیچیدگی‌های پایین‌تری نسبت به داده‌های اولیه دارند، تجزیه می‌شوند. بدین ترتیب، دیگر داده‌های ورودی به مرحله‌ی مدل‌سازی، به‌اندازه‌ی قبل مشکل ساختارهای پیچیده‌ی چندگانه را نخواهد داشت و لذا انتظار میرود نتایج بهتری حاصل شود. بر این اساس، در مرحله‌ی دوم از مدل پیشنهادی، هریک از مؤلفه‌های اصلی با استفاده از مدل خودرگرسیون میانگین متحرک انباشته مدل می‌شود. سپس تمامی پیش‌بینی‌های مربوط به مؤلفه‌ها با یکدیگر ترکیب می‌شوند تا پیش‌بینی‌های نهایی حاصل شوند. نتایج حاصل از به‌کارگیری روش پیشنهادی در پیش‌بینی قیمت‌های هفتگی نفت خام تگزاس و برنت بیان‌گر کارآمدی روش پیشنهادی بوده است. روش پیشنهادی در حالت کلی، دقت مدل خودرگرسیون میانگین متحرک انباشته را به‌طور میانگین به‌میزان ٪۵۹/۷۱ و دقت مدل‌های هوشمند را به‌طور میانگین به‌میزان ٪۵۸/۹۳ افزایش داده است.

جدول ۹. نتایج آزمون Q به‌منظور تست نیکویی برازش مدل پیشنهادی (داده‌های هفتگی قیمت نفت خام برنت).

Autocorrelation	Partial Correlation		AC	PAC	Q − Stat	Prob
|**|	|**|	۱	۰٫۱۹۷	۰٫۱۹۷	۵۴٫۶۹۲	۰٫۰۰۰
| |	*| |	۲	-۰٫۰۵۳	-۰٫۰۹۶	۵۸٫۷۰۳	۰٫۰۰۰
*| |	*| |	۳	-۰٫۱۴۹	-۰٫۱۲۴	۸۹٫۹۱۹	۰٫۰۰۰
| |	| |	۴	-۰٫۰۵۴	-۰٫۰۰۳	۹۴٫۰۳۲	۰٫۰۰۰
| |	| |	۵	۰٫۰۰۸	۰٫۰۰۳	۹۴٫۱۳۲	۰٫۰۰۰
| |	| |	۶	۰٫۰۰۴	-۰٫۰۲۲	۹۴٫۱۵۴	۰٫۰۰۰
| |	| |	۷	-۰٫۰۰۳	-۰٫۰۰۷	۹۴٫۱۷۰	۰٫۰۰۰
| |	| |	۸	۰٫۰۲۸	۰٫۰۳۳	۹۵٫۲۴۷	۰٫۰۰۰
| |	| |	۹	-۰٫۰۱۱	-۰٫۰۲۸	۹۵٫۴۲۶	۰٫۰۰۰
| |	| |	۱۰	۰٫۰۳۲	۰٫۰۴۲	۹۶٫۸۴۷	۰٫۰۰۰
| |	| |	۱۱	-۰٫۰۲۶	-۰٫۰۳۷	۹۷٫۸۳۱	۰٫۰۰۰
| |	| |	۱۲	۰٫۰۴۰	۰٫۰۵۷	۱۰۰٫۱۱	۰٫۰۰۰
| |	| |	۱۳	۰٫۰۱۳	-۰٫۰۰۲	۱۰۰٫۳۳	۰٫۰۰۰
| |	| |	۱۴	۰٫۰۲۸	۰٫۰۲۷	۱۰۱٫۴۱	۰٫۰۰۰
| |	| |	۱۵	-۰٫۰۰۲	-۰٫۰۰۲	۱۰۱٫۴۱	۰٫۰۰۰
| |	| |	۱۶	-۰٫۰۴۳	-۰٫۰۳۸	۱۰۴٫۰۲	۰٫۰۰۰
| |	| |	۱۷	-۰٫۰۱۸	۰٫۰۰۶	۱۰۴٫۴۸	۰٫۰۰۰
| |	| |	۱۸	-۰٫۰۳۸	-۰٫۰۴۵	۱۰۶٫۵۵	۰٫۰۰۰
| |	| |	۱۹	-۰٫۰۰۶	۰٫۰۰۴	۱۰۶٫۶۰	۰٫۰۰۰
| |	| |	۲۰	۰٫۰۴۰	۰٫۰۲۹	۱۰۸٫۹۰	۰٫۰۰۰
| |	| |	۲۱	۰٫۰۰۱	-۰٫۰۱۹	۱۰۸٫۹۰	۰٫۰۰۰
| |	| |	۲۲	۰٫۰۵۴	۰٫۰۵۶	۱۱۳٫۰۶	۰٫۰۰۰
| |	| |	۲۳	۰٫۰۱۴	۰٫۰۰۵	۱۱۳٫۳۵	۰٫۰۰۰
| |	| |	۲۴	۰٫۰۵۷	۰٫۰۶۱	۱۱۷٫۹۴	۰٫۰۰۰
| |	| |	۲۵	-۰٫۰۲۹	-۰٫۰۴۴	۱۱۹٫۱۶	۰٫۰۰۰
| |	| |	۲۶	-۰٫۰۱۳	۰٫۰۲۰	۱۱۹٫۳۹	۰٫۰۰۰
| |	| |	۲۷	-۰٫۰۳۲	-۰٫۰۲۹	۱۲۰٫۸۹	۰٫۰۰۰
| |	| |	۲۸	-۰٫۰۱۳	-۰٫۰۰۴	۱۲۱٫۱۵	۰٫۰۰۰
| |	| |	۲۹	-۰٫۰۲۱	-۰٫۰۲۵	۱۲۱٫۸۰	۰٫۰۰۰
| |	| |	۳۰	۰٫۰۳۴	۰٫۰۳۸	۱۲۳٫۴۷	۰٫۰۰۰
| |	| |	۳۱	-۰٫۰۰۱	-۰٫۰۱۷	۱۲۳٫۴۷	۰٫۰۰۰
| |	| |	۳۲	-۰٫۰۲۸	-۰٫۰۴۳	۱۲۴٫۶۳	۰٫۰۰۰

پانوشت‌ها

1. empirical mode decomposition (EMD)
2. mode mixing
3. Jarque-Bera Test
4. Q-Test
5. skewness
6. kotusis

منابع (References)

1. Abramson, B. and Finizza, A. "Probabilistic forecasts from probabilistic models: a case study in the oil market", *International Journal of Forecasting*, **11**, pp. 63-72 (1995).
2. Box, G. and Jenkins, G., *Times Series Analysis Forecasting and Control*, Holden-Day San Francisco (1970).
3. Ediger, V.Ş., Akar, S. and Uğurlu, B. "Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model", *Energy Policy*, **34**, pp. 3836-3846 (2006).
4. Diaz-Robles, L.A. and et al. "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of temuco, chile", *Atmospheric Environment*, **42**, pp. 8331-8340 (2008).
5. Cadenas, E. and Rivera, W. "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model", *Renewable Energy*, **35**, pp. 2732-2738 (2010).

6. Han, P. and et al. "Drought forecasting based on the remote sensing data using ARIMA models", *Mathematical and Computer Modelling*, **51**, pp. 1398-1403 (2010).
7. Lee, Y.S. and Tong, L.-I. "Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming", *Knowledge-Based Systems*, **24**, pp. 66-67 (2011).
8. Sen, P., Roy, M., and Pal, P. "Application of ARIMA for forecasting energy consumption and GHG emission: a case study of an Indian pig iron manufacturing organization", *Energy*, **116**, pp. 1031-1038 (2016).
9. Hikichi, S.E., Salgado, E.G. and Beijo, L.A. "Forecasting number of ISO 14001 certifications in the Americas using ARIMA models", *Journal of Cleaner Production*, **147**, pp. 242-253 (2017).
10. Garcia, P., Sánchez, F., Garcia- Gonzalo, E. and et al. "PM10 concentration forecasting in the metropolitan area of Oviedo (northern Spain) using models based on SVM, MLP, VARMA and ARIMA: a case study", *Science of The Total Environment*, **621**, pp. 753-761 (2018).
11. Al-Musaylh, M.S. and et al. "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia", *Advanced Engineering Informatics*, **35**, pp. 1-16 (2018).
12. Khashei, M., Bijari, M. and Raissi Ardali, G. "Improvement of auto-regressive integrated moving average model using fuzzy logic and artificial neural networks", *Neurocomputing*, **72**, PP. 956-976 (2009).
13. Tseng, F.M., Tzeng, G.H., Yu, H.C. and et al. "Fuzzy ARIMA model for forecasting the foreign exchange market", *Fuzzy Sets and Systems*, **118**, pp. 9-19 (2001).
14. Khashei, M. and Bijari, M. "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting", *Applied Soft Computing*, **11**, pp. 2664-2675 (2011).
15. Khashei, M., Bijari, M. and Raissi, G.A. "Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks", *Computers & Industrial Engineering*, **63**, pp. 37-45 (2012).
16. Yu, L., Wang, S. and Lai, K.K. "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm", *Energy Economics*, **30**, pp. 2623-2635 (2008).
17. Tang, L., Yu, L., Wang, S. and et al. "A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting", *Applied Energy*, **93**, pp. 432-443 (2012).
18. Huang, N.E., Shen, Z., Long, S.R. and et al. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Paper Presented at The Proceedings of The Royal Society of London A: Mathematical, Physical and Engineering Sciences* (1998).
19. Liu, H. Tian, H. and Li, Y. "An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system", *Journal of Wind Engineering and Industrial Aerodynamics*, **141**, pp. 27-38 (2015).
20. Yu, L., Wang, S., and Kin Keung Lai, "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm", *Energy Economic*, **30**, pp. 2623-2635 (2008).
21. Wang, J., Zhang, W., Li, Y. and et al. "Forecasting wind speed using empirical mode decomposition and elman neural network", *Applied Soft Computing*, **23**, pp. 452-459 (2014).
22. Cheng, C.-H. and Wei, L.-Y. "A novel time-series model based on empirical mode decomposition for forecasting TAIEX", *Economic Modelling*, **36**, pp. 136-141 (2014).
23. Wu, Z. and Huang, N.E. "Ensemble empirical mode decomposition: a noise-assisted data analysis method", *Advances in Adaptive Data Analysis*, **1**, pp. 1-41 (2009).