تحلیل مجموعه‌ی مقادیر تکین: مطالعه‌ی موردی بهای سکه

مسعود میرزاجانی بجستانی، احسان ارمز
گروه آمار، دانشگاه آزاد اسلامی واحد مشهد

چکیده

تحلیل مجموعه‌ی مقادیر تکین (SSA) تکنیکی برای تحلیل سری‌های زمانی است که از عناصر تحلیل سری‌زمانی کلاسیک، آمار چندمتغیره، هندسه چندمتغیره، سیستم‌های پویا و پردازش سیگنال‌ها استفاده می‌کند. هدف SSA تجزیه سری اصلی به مجموعه‌ای کوچک از اجزای تفسیر پذیر مانند روند، مؤلفه نوسانی و نوفه بدون ساختار است. مبنای SSA تجزیه مقدار تکین ماتریس خاص ساخته شده بر روی سری‌زمانی است. این روش به مدل پارامتری و شرایطی مانند مانایی نیاز نداشته و از این رو مستقل از مدل است. در واقع SSA روشی ناپارامتری است که با سری‌های زمانی دلخواه، خطی یا غیر خطی، ایستا یا ناایستا، نرمال یا غیرنرمال به خوبی کار می‌کند. در این مقاله ضمن معرفی اجمالی این روش و ویژگی‌های آن، به بررسی قیمت سکه در بازه فروردین ۹۱ تا آبان ۹۲ با استفاده از SSA پراخته شده است.

واژه‌های کلیدی: سری‌زمانی، تحلیل مجموعه‌ی مقادیر تکین.

۱ مقدمه

در سال‌های اخیر روش تحلیل مجموعه‌ی مقادیر تکین (SSA[۱]) به عنوان تکنیکی قوی در تحلیل سری‌های زمانی مورد استفاده قرار گرفته، توسعه یافته و در بسیاری از مسائل کاربردی استفاده شده است.

SSA ابزار مفیدی است که می‌تواند برای حل مسائل زیر نیز مورد استفاده قرار گیرد:

- پیش‌بینی،
- هموارسازی،
- شناسایی روند،
- استخراج مؤلفه‌های فصلی و دوره‌ای،

[۱] Singular Spectrum Analysis

- شناسایی ساختار سری‌های زمانی کوتاه مدت،
- آزمون علیت.

برای به‌دست آوردن قابلیت‌های SSA نیازی به دانستن پارامترهای مدل سری‌زمانی مورد نظر نیست، به‌علاوه این روش چندین بسط اساسی دارد. نسخه چندمتغیره این روش، بسط همزمان چند سری را ممکن می‌سازد؛ همچنین ایده‌های SSA به چندین روش پیش‌بینی سری‌زمانی منجر می‌شود.
سری‌های زمانی واقعی اغلب با داده‌های گمشده توأم هستند که می‌توانند مانع تحلیل شده و دقت نتایج را کاهش دهند. روش‌های مختلفی بر اساس SSA وجود دارد که می‌توانند برای جایگزین کردن داده‌های گمشده مورد استفاده قرار گیرند. جنبه مهم دیگر SSA این است که برخلاف اکثر روش‌های دیگر، این روش حتی برای نمونه‌های با اندازه کوچک عملکرد مناسبی دارد.

۲ تاریخچه

شروع تحلیل مجموعه‌ی مقادیر تکین را معمولاً وابسته به انتشار مقاله‌ی بروم‌هد و کینگ [۲] می‌دانند، در حالی‌که ایده‌های SSA مستقلاً در روسیه (سنت پترزبورگ و مسکو) و نیز توسط چندین گروه در آمریکا و انگلستان توسعه داده شده بود. تاکنون چند صد مقاله در مورد جنبه‌های روش‌شناسی و کاربرد SSA منتشر شده‌اند. توصیفی کامل از مبانی نظری و کاربردی روش SSA (همراه با چندین مثال) را می‌توان در دانیلوف و ژیگلافسکی [۳] پیدا کرد. همچنین مقدمه ساده‌ای از مطالب در الزنر و سونیس [۴] ارائه شده است.
حسنی [۶] بر اساس مطالعه‌ای کاربردی نشان داد که SSA می‌تواند نسبت به روش‌هایی مانند SARIMA، ARAR و هولت وینتر پیش‌بینی‌های بهتری فراهم کند.
همچنین حسنی [۷] تعریف مفیدی از روش SSA را به همراه کاربردهای مختصری از آن ارائه کرده است. حسنی و ژیگلافسکی [۸] نیز به صورت کاربردی نشان داده‌اند که SSA روشی قدرتمند در تحلیل و پیش‌بینی سری‌های زمانی اقتصادی است، آن‌ها کاربرد روش SSA را برای تحلیل و پیش‌بینی حساب‌های ملی ایران با استفاده از داده‌های به دست آمده از بانک مرکزی ایران در نظر گرفته‌اند.

۳ مروری بر روش تحلیل مجموعه‌ی مقادیر تکین

هدف SSA تجزیه سری‌های اصلی به مجموعه‌ای کوچک از اجزای تفسیرپذیر مانند روند، اجزای نوسانی و نوفه بدون ساختار، بر اساس تجزیه مقادیر تکین ماتریس خاص ساخته شده بر روی سری‌زمانی است. این روش به مدل پارامتری و شرایطی مانند مانایی نیاز نداشته و از این رو مستقل از مدل است. سری زمانی غیر صفر $Y_T = (y_1, \cdots, y_T)$ را در نظر بگیرید. روش SSA در دو مرحله صورت می‌پذیرد که هریک شامل دو گام مجزای زیر هستند: در ادامه به تشریح مراحل فوق می‌پردازیم.

مرحله ۱ : تجزیه
- نشانیدن
- تجزیه مقدار تکین

مرحله ۲ : بازسازی
- گروه‌بندی
- میانگین‌گیری قطری

۱.۳ مرحله ۱: تجزیه

۱.۱.۳ گام اول: نشانیدن

نشانیدن را می‌توان به عنوان نگاشتی در نظرگرفت که سری‌زمانی یک بعدی Y_T را به سری چند بعدی $X_1, X_2, \cdots, X_K$ انتقال می‌دهد. نتیجه این گام، ماتریس مسیر زیر است.

$$\mathbf{X} = [X_1, X_2, \cdots, X_K] = (x_{ij})_{i,j=1}^{L,K}$$
$$= \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_k \\ y_2 & y_3 & y_4 & \cdots & y_{k+1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{pmatrix}. \qquad (1)$$

در این ماتریس پارامتر L را طول پنجره می‌نامند که عددی صحیح و $0 \leq L \leq T$ است. همچنین مقدار K برابر است با $T - L + 1$. توجه داشته باشید که ماتریس مسیر $\mathbf{X}$ یک ماتریس هنکل [2] است، که بدین معناست که تمامی مؤلفه‌های $i + j$ مقداری ثابت دارند. اعضای روی قطر اصلی مساوی هستند و تمامی مؤلفه‌های روی قطرهای فرعی نیز با هم برابرند.

۲.۱.۳ گام دوم: تجزیه مقدار تکین (SVD[3])

گام دوم، تجزیه ماتریس مسیر را بر اساس مقادیر ویژه انجام داده و این عمل را به عنوان مجموعی از ماتریس‌های مقدماتی دو به دو متعامد با رتبه ۱ نشان می‌دهد. مقادیر ویژه XX^T را به ترتیب نزولی $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$ مرتب و دستگاه متعامد یکه $U_1, \cdots, U_L$ را به عنوان بردارهای ویژه XX^T متناظر با این مقادیر ویژه در نظر می‌گیریم. تعریف کنید:

$$d = \max(i, s.t. \lambda_i > 0) = rank(X).$$

با استفاده از نماد $V_i = \dfrac{X^T U_i}{\sqrt{\lambda_i}}$ ، SVD ماتریس مسیر را می‌توان به شکل

$$X = X_1 + X_2 + \cdots + X_d, \qquad (2)$$

نوشت که $X_i = \sqrt{\lambda_i} U_i V_i^T$. ماتریس‌های X_i دارای رتبه یک و بنابراین ماتریس‌های مقدماتی هستند، U_iها (که در ادبیات SSA آنها را EOF می‌نامند) و V_iها (که اغلب مؤلفه‌های اصلی نامیده می‌شوند) به ترتیب بردارهای ویژه چپ و راست ماتریس مسیر هستند. مجموعه $(U_i, V_i, \sqrt{\lambda_i})$ iامین سه‌تایی ویژه ماتریس نامیده می‌شود و $\sqrt{\lambda_i}$ مقادیر ویژه ماتریس X_i است. همچنین مجموعه کلیه مقادیر $\sqrt{\lambda_i}$ طیف ماتریس نیز نامیده می‌شود. چنانچه تمامی مقادیر ویژه همنهشت یک باشند، بسط (۲) به شکل یکتائی تعریف می‌شود. SVD (۲)، از این نظر بهینه است که در بین کلیه ماتریس‌های $X^{(r)}$ با رتبه $r < d$ بهترین تقریب ماتریس مسیر $\mathbf{X}$ توسط $\sum_{i=1}^{r} X_i$ به دست می‌آید، به‌طوری‌که $\| X - X^{(r)} \|$ مینیمم است.
بنابراین می‌توانیم نسبت $\frac{\lambda_i}{\sum_{i=1}^{d} \lambda_i}$ را به عنوان مشخصه میزان مشارکت ماتریس X_iدر بسط (۲) در نظر بگیریم. لذا $\frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$ ، به عبارت دیگر مجموع r نسبت اول، مشخصه تقریب بهینه ماتریس مسیر به وسیله ماتریس‌هایی با رتبه r است. مشخصه بهینه دیگر SVD به خصوصیات جهت‌های مشخص شده توسط بردارهای ویژه $U_1, U_2, \cdots, U_d$ باز می‌گردد. به‌طور خاص، اولین بردار ویژه (U_1) تعیین کننده جهتی است که تغییرات

[2] Hankel

[3] Singular Value Decomposition

تصاویر بردارهای تأخیر در این جهت ماکسیمم است. هر بردار ویژه بَعدی تعیین کننده جهتی خواهد بود که عمود بر جهات قبلی است و تغییرات تصویر بردارهای تأخیر در این جهت نیز ماکسیمم است. بنابراین طبیعی است که جهت iامین بردار ویژه را iامین جهت اصلی بنامیم. توجه داشته باشید که ماتریس‌های مقدماتی X_i از تصویر بردارهای تأخیر بر روی iامین جهت خاص ساخته شده‌اند. این طرز نگاه به SVD ماتریس مسیر و ارتباط آن با تحلیل مؤلفه‌های اصلی، منجر به نام‌گذاری‌های زیر می‌گردد. بردار U_i را iامین بردار ویژه و V_i را iامین بردار عامل می‌نامیم.

۳.۱.۳ گزینش طول پنجره L

طول پنجره L تنها پارامتر مرحله تجزیه است. با اطلاع از این مطلب که سری‌های زمانی ممکن است یک مؤلفه دوره‌ای با دوره‌ای صحیح داشته باشند، به منظور به‌دست آوردن تجزیه‌پذیری بهتری از این مؤلفه‌ی دوره‌ای، توصیه می‌شود که طول پنجره را متناسب با این دوره انتخاب کنیم. به عنوان مثال با فرض اینکه یک تناوب سالانه در سری‌ها وجود دارند پیشنهاد می‌شود که مقدار L مضربی از ۱۲ باشد.

۲.۳ مرحله ۲: بازسازی

۱.۲.۳ گام اول: گروه‌بندی

مرحله گروه‌بندی متناظر با تقسیم ماتریس‌های مقدماتی به چندین گروه و جمع‌کردن ماتریس‌های درون هر یک از گروه‌ها است. فرض کنید $I = \{i_1, i_2. \cdots, i_p\}$ گروهی از اندیس‌ها باشند، آنگاه ماتریس X_I متناظر با گروه I به شکل $X_I = X_{i_1} + X_{i_2} + \cdots + X_{i_p}$ تعریف خواهد شد. تقسیم مجموعه اندیس‌های $J = 1, 2, \cdots, d$ به زیرمجموعه‌های مجزای $I_1, I_2, \cdots, I_m$ متناظر با نمایش مجدد ماتریس X به شکل زیر است.

$$X = X_{I_1} + X_{I_2} + \cdots + X_{I_m}.$$

رویه انتخاب مجموعه‌های $I_1, I_2, \cdots, I_m$ را گروه‌بندی سه‌تایی‌های ویژه می‌نامیم. به ازای یک گروه معلوم I مشارکت مؤلفه X_I در بسط (۱) به وسیله نسبت مقادیر ویژه متناظر یعنی $\frac{\sum_{i \in I} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$ اندازه‌گیری می‌شود. چنانچه ماتریس X_I یک ماتریس هنکل باشد، آنگاه سری‌های $Y_T^{(1)}$ و $Y_T^{(2)}$ وجود دارند، به‌طوری‌که: $Y_T = Y_T^{(1)} + Y_T^{(2)}$ و ماتریس‌های مسیر این سری‌ها به ترتیب X_I و $X_{j.I}$ هستند. در این حالت می‌گوییم که سری‌های مورد نظر تقریباً جداشدنی هستند. برای جزئیات بیشتر به گولیاندینا و همکاران [۵] مراجعه کنید. در نتیجه هدف گام گروه‌بندی، به‌دست آوردن چندین گروه $I_1, I_2, \cdots, I_m$ است به‌طوری‌که ماتریس‌های $X_{I_1}, X_{I_2}, \cdots, X_{I_m}$ در رابطه (۱) صدق کنند و نزدیک به ماتریس‌های هنکل مشخص باشند. مرحله گروه‌بندی بر اساس تحلیل سه‌تایی‌های ویژه در بسط SVD است. برخی از توصیه‌های کاربردی برای گروه‌بندی در بخش ۱-۶ گولیاندینا و همکاران [۵] شرح داده شده‌اند.

۲.۲.۳ گام دوم: میانگین‌گیری قطری

هدف میانگین‌گیری قطری تبدیل یک ماتریس به فرم یک ماتریس هنکل است که متعاقباً می‌تواند به یک سری‌زمانی تبدیل شود. چنانچه مؤلفه‌های ماتریس Z را با z_{ij} نشان دهیم آنگاه kامین جمله سری، از میانگین‌گیری z_{ij} بر روی کلیه i و jهایی که در شرط $i+j = k+1$ صدق می‌کنند، به‌دست می‌آید. این رویه را میانگین‌گیری قطری

یا هنکل‌سازی ماتریس Z می‌نامیم. نتیجه هنکل‌سازی یک ماتریس Z، ماتریس هنکل HZ خواهد بود. توجه داشته باشید که فرآیند هنکل‌سازی از این نظر بهینه است که ماتریس HZ از لحاظ فرم ماتریس در بین کلیه ماتریس‌های هنکل دارای اندازه متناظر، به Z نزدیک‌تر است. (بخش ۶-۲ گولیاندینا و همکاران [۵] را ملاحظه کنید.) ماتریس هنکل HZ به نوبه خود به طور یکتایی یک سری‌زمانی را توصیف می‌کند. این عمل به کمک مرتبط کردن مقادیر موجود در قطرها با مقادیر سری صورت می‌پذیرد. با اعمال رویه هنکل‌سازی بر روی تمامی مؤلفه‌های ماتریس (۱) بسط زیر را به‌دست خواهیم آورد.

$$X = \hat{X}_{I_1} + \hat{X}_{I_2} + \cdots + \hat{X}_{I_m},$$

که در آن $\hat{X}_{I_1} = HX$. این عمل معادل تجزیه سری اولیه $Y_T = (y_1, \cdots, y_T)$ به مجموع m سری زیر است:

$$y_t = \sum_{k=1}^{m} \hat{y}_T^{(k)},$$

که در آن $\hat{Y}_T^{(k)} = (\hat{y}_1^{(k)}, \hat{y}_2^{(k)}, \cdots, \hat{y}_m^{(k)})$ متناظر با ماتریس X_{I_k} است. یک گروه‌بندی منطقی منجر به تجزیه (۱) ای خواهد شد که ماتریس‌های X_{I_k} حاصل از آن تقریباً هنکل هستند. این مطلب متناظر با تجزیه‌پذیری تقریبی است و باعث خواهد شد تا حاصل‌ضرب‌های نقطه‌ای دو به دوی ماتریس‌های مختلف X_{I_k}، کوچک باشند. رویه محاسبه سری‌زمانی $\hat{y}_T^{(k)}$ (یعنی ساختن گروه‌های I_k و میانگین‌گیری قطری ماتریس‌های X_{I_k}) بازسازی سری $\hat{Y}_T^{(k)}$ به وسیله سه‌تایی‌های ویژه با اندیس‌های I_k نامیده می‌شود. در رابطه با روش گروه‌بندی باید به این نکته توجه داشت که چنانچه L به اندازه کافی بزرگ باشد، بردارهای ویژه رفتاری مشابه مؤلفه‌های سری‌های زمانی متناظر از خود نشان خواهند داد. به ویژه روند سری، متناظر با بردارهای ویژه با تغییرات کم خواهد بود. مؤلفه همساز (هارمونیک) زوجی از بردارهای ویژه همساز چپ (و راست) تولید می‌کند که فرکانس مشابهی دارند و غیره. آنچه بدان اشاره شد را SSA پایه می‌نامند. SSA پایه علاوه بر پیش‌بینی می‌تواند برای هموارسازی، پالایش، کاهش نوفه، استخراج تناوب‌ها در شکل همسازهای تعدیل شده، پرکردن شکاف و... استفاده شود. همچنین SSA پایه می تواند به شیوه‌های متنوعی اصلاح شده و توسعه یابد.

۳.۳ پیش‌بینی به کمک SSA

در این بخش مراحل نحوه‌ی پیش‌بینی آینده به کمک روش SSA را شرح می‌دهیم.

۱.۳.۳ رابطه بازگشتی خطی (LRF[۴])

پیش‌بینی به کمک روش SSA به طور تقریبی با استفاده از LRF به کار می‌رود. سری‌های Y_T یک LRF از مرتبه d دارد اگر اعداد $a_1, a_2, \cdots, a_d$ وجود داشته باشند به‌طوری‌که

$$y_{i+d} = \sum_{k=1}^{d} a_k\, y_{i+d-k}, \qquad 1 \leq i \leq T-d. \quad (3)$$

یعنی LRF معادله (۳) برای سری‌های Y_t معادل با نماینده آن به عنوان مجموعی نمایی‌ها، چندجمله‌ای‌ها و همسازها است یعنی

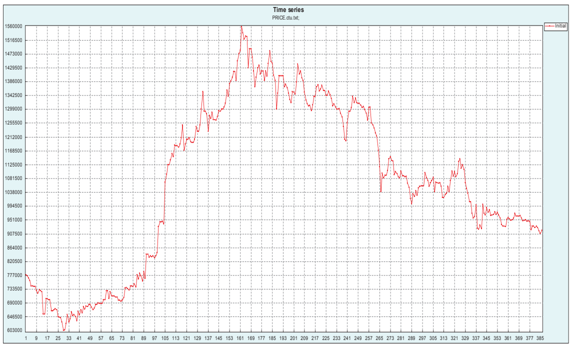
$$y_t = \sum_{k=1}^{q} \alpha_k(t) e^{\mu_k t} \sin(2\pi\omega_k t + \phi_k), \qquad (4)$$

[۴] Linear Recurrent Formulae

که در آن $\alpha_k(t)$ چندجمله‌ای‌ها هستند، μ_k، ω_k و ϕ_k پارامترهای دلخواه هستند. تعداد جملات مستقل خطی q در معادله (۴) کمتر یا مساوی d است. رده‌ی سری‌هایی که می‌توانند توسط رابطه‌ی بازگشتی خطی (۳) (یا به طور معادل معادله (۴) با تعداد جملات کمتر) مورد بررسی قرار گیرند نسبتاً وسیع بوده و در مسائل کاربردی مهم می‌باشند.

۴ کاربرد: پیش‌بینی بهای سکه

می‌خواهیم روش تحلیل مجموعه‌ی مقادیر تکین را به کمک نرم‌افزار Caterpillar-SSA شرح دهیم. سری‌زمانی به نام PRICE که در بر گیرنده بهاء سکه در بازه زمانی فروردین ۱۳۹۱ تا آبان ۱۳۹۲ به صورت روزانه است را در نظر بگیرید. این داده‌ها از آرشیو روزنامه شهرآرا (بازار سکه مشهد) در سایت این روزنامه [۱] بدست آمده است. سری کامل که شامل N=۳۸۶ داده است را بررسی خواهیم کرد. لازم به ذکر است که در برخی از روزها به دلیل نوسان شدید بازار، داده‌ای وجود ندارد مثلاً در ماه‌های مهر و آبان ۱۳۹۱ داده‌ای برای قیمت سکه گزارش نشده است. شکل‌های ارائه شده مفاهیم تئوری را شرح می‌دهند و می‌توانند در به‌دست آوردن درک شهودی مناسبی از روش SSA مفید باشند. نمودار پراکنش سری‌زمانی در شکل ۱ نشان داده شده است. این نمودار نشان می‌دهد که سری ما یک روند دارد. هدف، تجزیه این سری‌زمانی به ۳ مؤلفه‌ی روند، مؤلفه خطی و یک نوفه است.

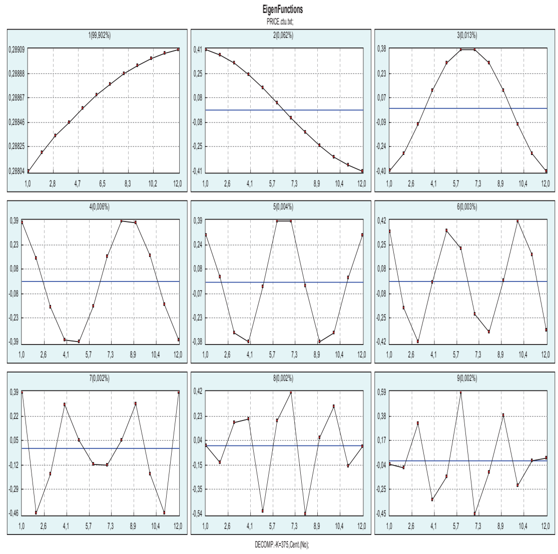


شکل ۱: نمودار سری‌زمانی بهاء سکه

۱.۴ تجزیه: روند، مؤلفه همساز و نوسان

- **انتخاب طول پنجره**

انتخاب طول پنجره مناسب به مساله مورد بررسی و اطلاعات اولیه سری‌زمانی بستگی دارد. می‌دانیم که L را باید به اندازه کافی بزرگ انتخاب کنیم، به‌طوری‌که از $\frac{T}{2}$ بزرگتر نباشد. پس از بررسی طول پنجره‌های مختلف و با توجه به مشاهده نوسانات فصلی در سری‌زمانی، بهترین طول پنجره برای سری‌زمانی مورد نظر ۱۲ انتخاب می‌شود. بنابراین بر اساس این طول پنجره و SVD ماتریس مسیر، ۱۲ سه‌تایی ویژه داریم که بر اساس میزان مشارکت آن‌ها در تجزیه مرتب شده‌اند. اولین سه‌تایی ویژه تمرکز کلی سری‌زمانی را شرح می‌دهد. با توجه به اینکه در اغلب موارد سه‌تایی‌های ویژه‌ای که سهم اندکی دارند با مؤلفه نوفه سری مرتبط هستند، لذا بایستی مجموعه سه‌تایی‌های ویژه برتر را مشخص کنیم. می‌خواهیم نتایج گام SVD را مد نظر قرار دهیم. شکل ۲ مؤلفه‌های اصلی (بردارهای ویژه چپ) مرتبط با ۹ سه‌تایی ویژه نخست را نشان می‌دهد. توجه داشته باشید که شکل بردارهای عامل (بردارهای ویژه راست) مشابه شکل مؤلفه‌های اصلی نیست زیرا $L = ۱۲$ نزدیک $K = ۳۷۵$ نیست.



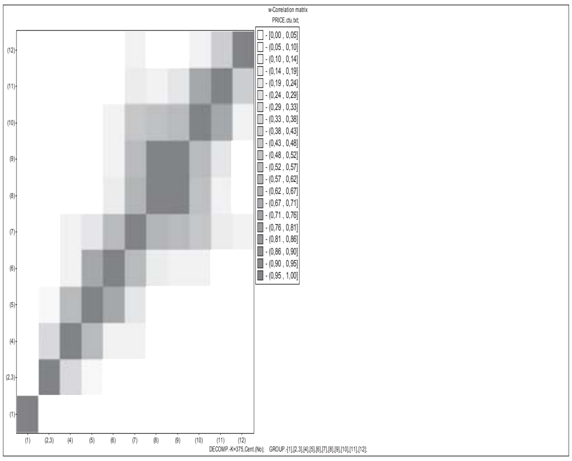
شکل ۲: نمودارهای یک بعدی بردارهای ویژه

- **تفکیک پذیری**

تجزیه SSA سری Y_T تنها در صورتی می‌تواند موفقیت آمیز باشد که مؤلفه‌های جمعی حاصل به طور تقریبی از یکدیگر تفکیک پذیر باشند. کمیت زیر که همبستگی وزن دار شده یا ω–همبستگی نامیده می‌شود، اندازه‌ای طبیعی از وابستگی بین دو سری $Y_T^{(1)}$ و $Y_T^{(2)}$ است.

$$\rho_{12}^{(\omega)} = \frac{\left(Y_T^{(1)}, Y_T^{(2)}\right)}{\|Y_T^{(1)}\|_\omega \|Y_T^{(2)}\|_\omega},$$

ماتریس اندازه‌های مطلق ω–همبستگی‌های متناظر با تجزیه کامل ابزار مفیدی برای تعریف گروه‌های سه‌تایی‌های ویژه، است که در آن هر گروه متناظر با یکی از مؤلفه‌های ماتریس SVD است. چنانچه دو مؤلفه‌ی تجدید ساختار شده، ω–همبستگی صفر داشته باشند نتیجه می‌گیریم که این دو مؤلفه تفکیک‌پذیر هستند. چنانچه اندازه‌ی مطلق ω–همبستگی کوچک باشد آنگاه سری‌های متناظر تقریباً ω–متعامد هستند و چنانچه بزرگ باشد آنگاه دو سری از ω–متعامد بودن فاصله داشته و لذا تفکیک‌پذیری مناسبی نخواهند داشت. مقادیر همبستگی بین مؤلفه‌های تجدید ساختار شده متناظر با مقادیر مطلق همبستگی از صفر تا یک در یک مقیاس ۲۰ درجه‌ای خاکستری از سفید تا سیاه نشان داده می‌شوند. در شکل ۳ مؤلفه‌های ۳ به بَعد تفکیک‌پذیری مناسبی ندارند و می‌توان آن‌ها مرتبط با نوفه دانست، لذا بلوک سه‌تایی‌های ویژه باقیمانده را به عنوان نوفه در نظر می‌گیریم. فرم ماتریس ω–همبستگی نشان می‌دهد که چگونه گروه‌بندی مناسب را انجام دهیم. سه‌تایی ویژه اول کاملاً متناظر با روند است (این وضعیت در مسائل کاربردی بسیار متداول است) سه‌تایی‌های ویژه بَعدی متناظر با همسازها هستند.



شکل ۳: همبستگی‌های ۱۲ مؤلفه بازسازی شده ماتریس

۲.۴ بازسازی: گروه‌بندی و میانگین‌گیری قطری

- **شناسایی روند**

فرض کنید سری‌زمانی تنها متشکل از روند، باشد. مسائل کاربردی نشان می‌دهند که در این حالت یکی یا بیشتر

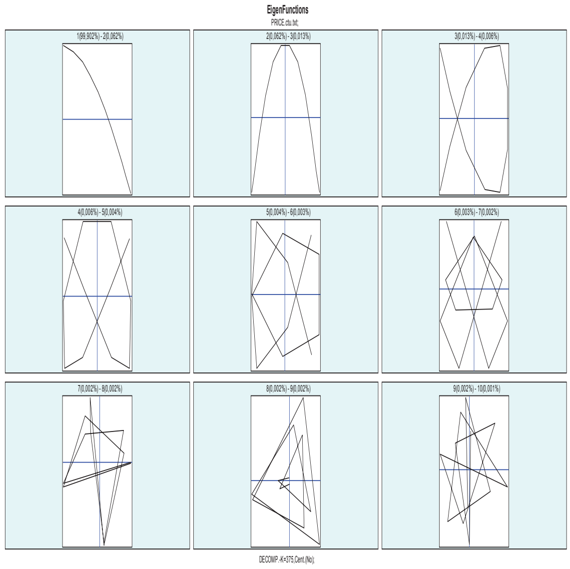
از یکی از سه‌تایی‌های ویژه‌ی ابتدایی نیز به کندی تغییر خواهند کرد. در این مثال بردار ویژه اول به طور آشکار دارای شکل مورد نظر است. اما سه‌تایی‌های ۲ و ۳ به این صورت نیستند بنابراین، روند تنها توسط سه‌تایی ویژه اول توصیف می‌گردد این بدان معناست که روند سری اصلی توسط یک تابع نمایی تقریب زده می‌شود.

- **تشخیص مؤلفه همساز**

مسئله اساسی در این بخش شناسایی و تفکیک مؤلفه‌های نوسانات سری است که نقشی در روند ایفا نمی‌کنند. در عمل مقادیر ویژه‌ی دو سه‌تایی ویژه یک سری همساز، بسیار به یکدیگر نزدیک هستند و این حقیقت تشخیص تصویری مؤلفه‌های همساز را ساده می‌سازد. تحلیل دوبه‌دوی نمودارهای پراکنش بردارهای ویژه نیز به شناخت تصویری سه‌تایی‌های ویژه‌ای که با مؤلفه‌های همساز سری متناظر هستند، کمک می‌کند؛ مشروط بر آنکه این مؤلفه‌ها از مؤلفه باقیمانده تفکیک‌پذیر باشند. شکل ۴ نمودارهای پراکنش بردارهای عامل زوج‌شده داده‌های بهاء سکه در شهر مشهد متناظر با همسازهای دارای تعداد فرکانس‌های کوچک را نشان می‌دهد. این شکل نمودارهای دوبُعدی‌ای را نشان می‌دهد که مسیرهای دوبُعدی آن‌ها در منحنی‌های مارپیچ خود رأس‌هایی دارند. این مطلب نشان می‌دهد که این زوج‌های بردارهای ویژه، توسط مؤلفه‌های همساز تعدیل‌یافته سری‌زمانی اولیه تولید شده‌اند. بدین طریق سه‌تایی‌های ویژه ۳-۲ متناظر با دوره ۴ با فرکانس $\frac{۱}{۴} = \frac{۳}{۱۲}$ هستند.

- **تفکیک سیگنال از نوفه**

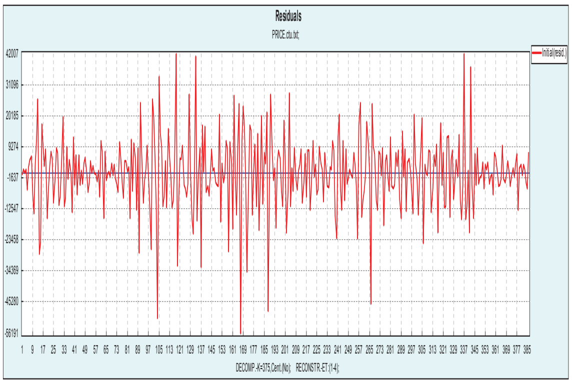
حال می‌خواهیم بر روی تفکیک مؤلفه‌های متناظر با سیگنال از مؤلفه‌های نوفه تمرکز کنیم. رفتار نامنظم بردارهای ویژه می‌تواند تعلق آن‌ها به مجموعه‌ای که توسط



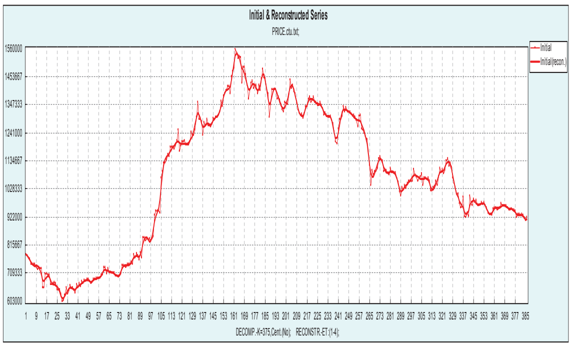
شکل ۴: نمودارهای دوبعدی مؤلفه‌های اصلی

مؤلفه‌ی نوفه تولید شده است را نشان دهد. با این حال بایستی بین این بی‌نظمی و ترکیب مؤلفه‌ها، که به واسطه عدم تفکیک‌پذیری قوی آن‌ها رخ می‌دهد، تمایز قایل گردیم. مرز بین مؤلفه‌های سیگنال و نوفه می‌تواند به وسیله‌ی کاهش کند (تقریباً بدون پرش) مقادیر ویژه (از جایی به بعد) تأیید شود. همچنین مجموعه بزرگی از سه‌تایی‌های ویژه که توسط مؤلفه‌های به یکدیگر همبسته بازسازی شده تولید می‌شوند با احتمال زیاد به یک نوفه تعلق دارند. (شکل ۳ چنین بلوکی از سه‌تایی‌های ویژه با شماره‌های ۴-۱۲ را در بر می‌گیرد). این حقیقت که ω-همبستگی بین سری تجدید ساختار یافته (سه‌تایی‌های ویژه ۱-۳) و باقیمانده‌ها (سه‌تایی‌های ویژه ۴-۱۲) برابر ۰.۰۰۵ است، نیز تأیید می‌کند که این گروه‌بندی بسیار منطقی است. شکل ۵ باقیمانده‌ها را پس از استخراج روند و مؤلفه‌ی فصلی نشان می‌دهد. چنانچه روند و باقیمانده‌ها را با یکدیگر جمع کنیم به سری اصلی دست خواهیم یافت که برای نوسانات فصلی تعدیل شده است و

در صورتی که سری‌های شکل‌های ۴، ۵، ۶ را جمع کنیم به سری اصلی (شکل۱) دست خواهیم یافت.



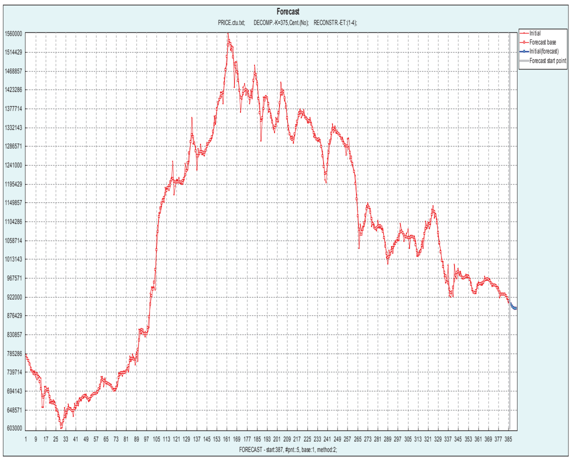
شکل ۵: سری باقیمانده‌ها



شکل ۶: نمودار سری بازسازی شده و اصلی

۳.۴ پیش‌بینی

در مساله پیش‌بینی، مدل SSA می‌تواند به کمک یک رابطه‌ی بازگشتی خطی توصیف شود. در این مثال پنج داده بَعدی (بهاء سکه ۵ روز نخست آذر ۱۳۹۲) را پیش‌بینی می‌کنیم. از طول پنجره ۱۲، ۴ سه‌تایی ویژه اول را برای تجزیه سری اولیه مورد استفاده قرار داده و از آنچه الگوریتم برداری نامیده می‌شود (گولیاندینا و همکاران [۵] را مشاهده کنید) برای مثال اخیر استفاده کرده‌ایم. شکل ۷ سری اولیه و پیش‌بینی پنج روز نخست آذر ۱۳۹۲ را نشان می‌دهد. جدول ۱ نیز قیمت واقعی و قیمت پیش‌بینی شده سکه را در پنج روز نخست آذر ۱۳۹۲ نشان می‌دهد.



شکل ۷: سری اولیه و پیش‌بینی پنج روز اول آذر ۱۳۹۲

جدول ۱: داده‌های اصلی و پیش‌بینی شده پنج روز اول آذر ۱۳۹۲

روزها	قیمت واقعی (ریال)	قیمت پیش‌بینی شده (ریال)
اول آذر ۱۳۹۲	۹۰۰۰۰۰۰	۹۰۸۸۹۸۰
دوم آذر ۱۳۹۲	۸۹۸۰۰۰۰	۹۰۷۱۸۹۰
سوم آذر ۱۳۹۲	۸۸۱۰۰۰۰	۹۰۲۰۶۳۰
چهارم آذر ۱۳۹۲	۸۵۴۰۰۰۰	۸۹۶۹۳۶۰
پنجم آذر ۱۳۹۲	۸۲۵۰۰۰۰	۹۰۰۳۵۴۰

۵ نتیجه‌گیری

در این مقاله روش‌شناسی SSA را شرح دادیم و نشان دادیم که این روش برای تحلیل و پیش‌بینی سری‌زمانی

مورد نظر مناسب است. همچنین نشان دادیم که این روش در استخراج همزمان همسازها و مؤلفه‌های روند به خوبی عمل می‌کند. ما از ۳۸۶ داده قیمت بازار سکه شهر مشهد برای توصیف روش SSA استفاده کردیم. نتایج نشان داد که سری PRICE مثالی با ساختار پیچیده است که به راحتی توسط روش SSA تحلیل می‌شود و می‌تواند به خوبی کاربرد SSA را شرح دهد.
توجه داشته باشید که برخلاف روش‌های استاندارد برای تحلیل سری‌زمانی اقتصادی، SSA به مدل پارامتری یا تبدیل داده‌ها به مقیاس لگاریتمی نیازی ندارد.

مراجع

[۱] روزنامه‌ی شهرآرا:
URL:http://www.shahrara.com.

[2] Broomhead, D. S. and King, G. (1986). Extracting qualitative dynamics from experimental data. Physica D, 20, 217-236.

[3] Danilov, D. and Zhigljavsky, A. (1997). (Eds.) Principal components of time series: the ‘caterpillar’ method. University of St. Petersburg, St. Petersburg, (In Russian).

[4] Elsner, J. B. and Tsonis, A. A. (1996). Singular spectrum analysis, a new tool in time series analysis. Plenum Press, New York and London.

[5] Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2001). Analysis of time series structure: SSA and related techniques. Chapman and Hall/CRC, New York – London.

[6] Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. Journal of Data Science, 5(2), 239-257.

[7] Hassani, H. (2010). A brief introduction to singular spectrum analysis. UK-China workshop on singular spectrum analysis and its applications.

[8] Hassani, H. and Zhigljavsky, A. (2009). Singular spectrum analysis: methodology and application to economics data. Journal of System Science and Complexity, 22(3), 372-394.