



## پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک و تحلیل طیفی تکین هم‌پوشانی

زهرا حسن دوست<sup>۱</sup>

حمیدرضا وکیلی فرد<sup>۲\*</sup>

فریدون رهنمای رودپشتی<sup>۳</sup>

تاریخ دریافت: ۱۴۰۲/۰۳/۱۴ تاریخ پذیرش: ۱۴۰۲/۰۴/۲۶

### چکیده

پژوهش حاضر پیش‌بینی شاخص کل بورس اوراق بهادار تهران را با روش تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک و تحلیل طیفی تکین هم‌پوشانی مورد تحلیل می‌باشد. به لحاظ هدف کاربردی و از حیث روش، توصیفی-تحلیلی است. جامعه آماری آن، قیمت روزانه شاخص کل بورس اوراق بهادار تهران در بازه زمانی ده ساله (۱۳۸۸ تا ۱۳۹۷) و نمونه پژوهش نیز ۲۴۱۱ داده از بازه لگاریتمی شاخص مورد نظر می‌باشد. ابتدا الگوریتم ژنتیک در روش SSA بر روی شاخص پیاده‌سازی شد. سپس، با استفاده از روش Ov-SSA به منظور بهبود بازسازی و تفکیک‌پذیری مولفه‌ها، سری‌های زمانی اولیه و بزرگ به قسمت‌های کوچک و متوالی مشترک تقسیم‌گردیدند و روش تحلیل SSA استاندارد برای هر قسمت بکار گرفته شد. نتایج پژوهش نشان داد که روش تحلیل Ov-SSA با داشتن خطای قدرمطلق میانگین کمتر، عملکرد بالاتری نسبت به روش تحلیل GA-SSA دارد.

**کلمات کلیدی:** طول پنجره، سیگنال، نقطه‌ی برش، نویز، تحلیل طیفی تکین هم‌پوشانی.

۱. دانشجوی دکتری گروه مدیریت مالی، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

*Zahra.Hasandoost.62@gmail.com*

۲. گروه مدیریت مالی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران (نویسنده مسئول). *Vakilifard.phd@gmail.com*

۳. گروه مدیریت مالی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران. *Rahnama.Roodposhti@gmail.com*

## مقدمه

اکثر سهام داران به اطلاعات با کیفیت برای تصمیم‌گیری در مورد این که چگونه سهام‌شان را بخرند یا بفروشند نیاز دارند. نیازهای اطلاعاتی قیمت سهام بسیار ضروری هستند زیرا آن‌ها عوامل مختلفی هستند که می‌تواند بر نوسان قیمت سهام اثر بگذارند و از طرفی سهامداران به دنبال کسب سود هستند و انتظار کاهش سهم‌شان در بازار را ندارند (۴۱). بسیاری از سرمایه‌گذاران، شرکت‌ها، اعتباردهندگان و کلیه ذی‌نفعان جهت دستیابی به سودآوری و حتی تحلیل‌ها بازار سرمایه تمایل دارند از میزان دقیق یا درصد نسبتاً دقیقی اطلاع پیدا کنند. با تحلیل نوسانات می‌توان تحلیل روند گذشته، روند حال، حتی پیش‌بینی از روند آتی را در اختیار ذینفعان قرارداد و یکی از موارد بسیار مهم در نوسانات، مقدار نوسان می‌باشد که می‌تواند بسیاری از تصمیمات ذینفعان را تحت تأثیر قرار دهد و از این رو بسیار ضروری می‌باشد که میزان نوسان نیز با بهترین و دقیق‌ترین ابزار اندازه‌گیری گردد و این ابزارها در اختیار ذینفعان قرار داده شود تا با دستیابی به این اطلاعات راحت‌تر و بهتر تصمیم‌گیری نمایند (۵). نوسانات در بازارهای مالی از اهمیت ویژه‌ای برخوردار است. به‌طوری‌که بسیاری از مدل‌های تخصیص پورترفو، قیمت‌گذاری دارایی مالی و مدیریت ریسک، بر پایه میزان نوسانات و برآوردی که نوسان‌پذیری هر متغیر به دست می‌آید برپا شده‌اند (۴). نوسان‌پذیری ورودی و عامل کلیدی در ساختار پرتفوی، مدیریت و ریسک می‌باشد (۴۴). نوسانات بازار سهام در تصمیم‌گیری سرمایه‌گذاران بازار و مدیران شرکت‌ها نقش ویژه‌ای دارد (۶). مدیریت نوسان ریسک از زمان اولین بحران جهانی اخیر مالی به شدت در بازارهای مالی افزایش پیدا نموده است (۳۷). پیش‌بینی قیمت سهام، روش بسیار رایج در تحلیل قیمت سهام می‌باشد (۴۱).

## اهمیت و ضرورت پژوهش

گجرال و دیگران (۲۰۲۰) بیان نمودند که پیش‌بینی سری‌های زمانی، تکنیکی برای پیش‌بینی مقادیر بر مبنای تاریخی هستند و روش‌های سری‌های زمانی، در روش‌های مختلف طبقه‌بندی می‌شوند. برخی از دسته‌بندی‌های ممکن؛ قلمرو زمانی و فرکانس، تک‌متغیره و چندمتغیره، پارامتریک و ناپارامتریک می‌باشند. که در مطالعه

خود، هفت روش کلاسیک را انتخاب و بر مبنای قلمرو زمانی و فرکانس طبقه‌بندی نمودند. این ابزارها *PCA* (تحلیل مسیر اصلی)<sup>۱</sup>، شبکه‌های عصبی، مدل‌های *ARMA-ARIMA* (میانگین بازگشتی خودرگرسیون، میانگین برگشتی یکپارچه خودرگرسیون)، مدل‌های مارکوف، تحلیل فوریر و تحلیل طیفی هستند (۲۴). مدل‌های نوسان تصادفی اصطلاحاً به مدل‌هایی اطلاق می‌گردد که در آن‌ها نوسان به صورت تصادفی تغییر می‌کند. این تغییر تصادفی می‌تواند بر اساس یک معادله دیفرانسیل تصادفی و یا بر اساس یک فرایند تصادفی گسسته مدل شود (۳). مدل  $SV^2$  دربرگیرنده نوسان تصادفی با عامل پنهان می‌باشد و تفاوت عمده دیگر آن با مدل‌های  $GARCH^3$  که با بازگشت به میانگین (بازده) مشتق می‌شوند مدل‌های  $SV$  دو حرکت مشتق از بازده و نوسان را ارزیابی می‌نماید (۳۱). مدل‌سازی نوسان معمولاً با گشتاورهای مرتبه دوم مشخص می‌شود که جرقه آن قبل از قرن ۲۱ زده شده و یکی از زمینه‌های پر فعالیت تحقیق در ادبیات مالی شده است، دو رویکرد در این هدف به کار می‌رود: الف- مدل‌های آرچ<sup>۴</sup> ب- مدل‌های نوسان تصادفی ( $SV$ ) که نوسان‌پذیری یک فرایند نوسان نهفته برگرفته از حرکت جدید است که در فرایند نوسان‌پذیری نهادینه شده‌اند. سری‌های نوسان‌پذیری برای استخراج در مدل‌های  $SV$  نسبت به بسط‌های *ARCH* دشوارتر هستند. واریانس شرطی در مدل‌های مبتنی بر *ARCH* تابع دقیقی از مشاهدات گذشته هستند و می‌توانند دقیقاً به اندازه پارامترهای مربوطه (حداقل نمونه‌ای) شناخته شده شوند. ماهیت نهفته نوسانات در مدل‌های  $SV$ ، استخراجشان را دشوارتر نشان می‌دهند و روش‌های استخراج سیگنال معمولاً مورد نیاز هستند. مدل‌های  $SV$  غیرخطی هستند اما می‌توانند با به‌کارگیری الگوریتم‌های مربع، به صورت مجموع دو مؤلفه به خطی تبدیل شوند: بخش یا جزء نوسان پیشرو که جزء سیگنال است و نویز سفید را می‌افزاید (۱۵). مفهوم سازی تجزیه طیفی عملگر کوواریانس از فرآیندهای تصادفی توسط کارهونن و لوتیو حدود سال ۱۹۴۰ بنانهاده شد. *SSA* و *M-SSA* (چندکاناله)<sup>۵</sup> توسط بروم هد و کینگ و سپس توسط فردریک با مفهوم پویایی‌های غیرخطی معرفی شدند. بعد از آن حوالی سال ۱۹۹۰، تناسب‌هایی توسط جیل و ووتارد بین مسیر کینگ-بروم

فرض‌های محدودکننده‌ای مانند نرمال، خطی و ثابت بودن را برای داده‌ها در نظر می‌گیرند. این امر موجب محدود شدن کاربرد عملی مدل‌های سری‌های زمانی می‌شود، زیرا مجموعه‌های مالی غالباً به این فرضیات محدود نمی‌شوند. علاوه بر این، مؤلفه‌ی دیگری که می‌تواند ظرفیت پیش‌بینی مدل‌ها را محدود کند، وقوع نویز در داده‌ها است (۲۵). حسنی و توماکس (۲۰۱۰) فرض نمودند که به طور کلی، دو روش اصلی در مورد پیش‌بینی سری‌های زمانی با نویز وجود دارد. روش اول نویز را نادیده می‌گیرد و سپس مدلی را بر رویدادهای بدون نویز تنظیم می‌کند و در روش دوم که روش مؤثرتری است، سعی در تجزیه‌ی این سری به زیر مؤلفه‌ها، شناسایی مؤلفه‌ی نویز، استخراج آن و سپس پیش‌بینی اطلاعات جدید از سری فیلتر شده را دارد. در رویکرد دوم، می‌توان به روش تحلیل طیفی تکین اشاره کرد، که یک روش غیر پارامتری و قدرتمند برای تحلیل داده‌ها است که بر روی مسائل کاربردی متنوع مانند مطالعه سری‌های زمانی، هندسه چند متغیره، سیستم‌های دینامیکی و پردازش سیگنال مورد استفاده قرار می‌گیرد (۲۸). تحلیل طیفی تکین شامل دو ویژگی مهم می‌باشد که استفاده از آن را در تجزیه و تحلیل سری‌های زمانی مالی تبدیل به یک ابزار قوی می‌کند. اول، در مورد داده‌ها فرضی قرار نمی‌دهد و دوم، برخلاف سایر روش‌ها، می‌توان آن را روی نمونه‌های کوچک نیز اعمال کرد. این ویژگی امکان استفاده از آن را در طیف وسیعی از مطالعات فراهم می‌کند. اگرچه تحلیل طیفی تکین به ابزاری پر کاربرد در تجزیه و تحلیل و پیش‌بینی آب و هوا (ووتارد و گیل، ۱۹۸۹) (۴۲)، پزشکی (سنایی و حسنی، ۲۰۱۵) (۳۶)، پردازش تصویر دیجیتال و سایر زمینه‌های دانش مانند علوم اجتماعی و فیزیک (حسنی، ۲۰۱۰) (۲۷)، تبدیل شده است، باین حال کاربرد آن در مجموعه‌های اقتصادی و مالی هنوز محدود به مطالعات اخیر می‌باشد و هنوز در بسیاری از زمینه‌ها نیاز به مطالعات گسترده وجود دارد (۲۱). گجرال و دیگران (۲۰۲۰) بیان نمودند که پیش‌بینی سری‌های زمانی، تکنیکی برای پیش‌بینی مقادیر بر مبنای تاریخی هستند (۲۴). مسئله اساسی پژوهش این است که آیا در پیش‌بینی شاخص مورد نظر، روش تحلیل طیفی تکین هم پوشانی عملکرد بالاتری را در پیش‌بینی نسبت به روش

هد و تجزیه کارونن - لوئیو ترسیم شد که منجر به فرمول سازی SSA شد. برخی از کاربردهای مهم SSA: اول آن که شامل تخمینی از اندازه‌های آماری است، دوم آن که پدیده‌های اصلی به وسیله داده‌ها را تشریح می‌کند و سوم آن که ویژگی‌های نویزی داده‌ها را شفاف می‌سازد (۲۴). SSA، تنها یک روش تحلیل سری‌های زمانی نیست اما می‌تواند آماره‌های چند متغیره، هندسه چند متغیره و سیستم‌های پویا و فرآیند سیگنال را ادغام نماید. SSA، یک رویکرد ناپارامتریک است که به هیچ فرض آماری نیاز ندارد. SSA به واسطه تجزیه سری‌های زمانی به مجموعه‌ای از زیرسری‌ها که شامل روند، چرخه و یا نویز هستند عمل می‌کند (۴۱). تحلیل طیفی تکین، سیگنال‌های پیچیده را به چندین بخش شفاف شامل یک سری از نوسانات، روند و نویز کاهش می‌دهد و عملکرد بهینه استراتژی‌های SSA از استراتژی تحلیل حالت تجربی (EMD) بالاتر می‌باشد (۱۲). در SSA فرض‌های پارامتریک مربوط به نرمال بودن، ایستایی و خطی بودن، زمانی که از مدل‌های ناپارامتریک مانند تحلیل طیفی تکین استفاده می‌شود به کار نمی‌آید (۱۸). نوسان‌پذیری در سری‌های زمانی مالی دو ویژگی اصلی دارند: اول این که نوسان‌پذیری در طول زمان متغیر هستند و هم‌چنین در نوسان‌پذیری متغیرهای مالی پدیده خوشه‌ای وجود دارد. مدل‌های نوسان‌پذیری تصادفی برای در نظر گرفتن همین ویژگی در مالی مورد استفاده قرار گرفت (۴). در دهه‌ی اخیر، پیشرفت‌هایی در قیمت‌گذاری اختیار معامله، اهمیت نوسان‌پذیری تصادفی را نشان داده است. مخصوصاً مدل‌های نوسان تصادفی زمان پیوسته مانند مدل هستون و بسط‌های آن بسیار عمومیت یافته است (۳۰). مدل SV زمانی ایجاد می‌شود که گشتاور دوم شرطی فقط یک متغیر محسوب نمی‌شود بلکه دارای یک پویایی درونی نیز هست. با این رویکرد، هنگامی که یک نویز اضافی برای توضیح معادله واریانس شرطی به مدل اضافه می‌شود، مزایای اقتصادی سنجی مهمی به ارمغان می‌آورد. بنابراین مدل SV شامل دو فرآیند تصادفی جداگانه است، یکی مربوط به میانگین شرطی که بر بازده اثر می‌گذارد و دومی مربوط به فرآیند نوسان‌پذیری است (۴). علی‌رغم آن که در زمینه‌ی تحلیل سری‌های زمانی، تلاش‌های زیادی برای معرفی روش‌ها و مدل‌های توانمند شده است، اکثر این روش‌ها

تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک نشان می‌دهد؟  
 در پیش‌بینی شاخص کل بورس اوراق بهادار تهران، روش تحلیل طیفی تکین هم پوشانی عملکرد بالاتری را نسبت به روش تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک دارد.

### پیشینه تجربی پژوهش

نویسنده	خلاصه موضوع	نتیجه
حسن دوست و وکیلی فرد (۱۴۰۰)	مقایسه روش تحلیل طیفی تکین و تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک در پیش‌بینی داده‌های ده ساله شاخص کل بورس اوراق بهادار تهران در بازه زمانی ۱۳۹۷ تا ۱۳۸۸	روش تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک دقت و عملکرد بالاتری نسبت به روش تحلیل طیفی استاندارد دارد (۲).
اسکندری و محرم پور (۱۳۹۷)	استفاده از روش تحلیل طیفی تکین با تمرکز بر طول پنجره و نقطه برش	موفق به استخراج مولفه‌های سیگنال از سیگنال اصلی گردیدند (۱).
یارمحمدی و محمودوند (۱۳۹۵)	مدل‌سازی و پیش‌بینی نرخ روزانه دلار به ریال در بازه زمانی تیرماه ۱۳۹۲ تا شهریور ۱۳۹۴ با استفاده از روش تحلیل طیفی تکین و ارزیابی عملکرد آن با مدل آریما.	روش تحلیل طیفی تکین می‌تواند به عنوان یک روش قدرتمند در پیش‌بینی نرخ ارز مورد استفاده قرار گیرد (۸).
میرزاجانی و دیگران (۱۳۹۲)	بررسی قیمت سکه در بازه فروردین ۹۱ تا آبان ۹۲ با استفاده از روش تحلیل طیفی تکین بر روی ۳۸۶ داده قیمت بازار سکه‌ی شهر مشهد	این روش در استخراج هم‌زمان هم‌سازها و مؤلفه‌های روند به‌خوبی عمل می‌کند (۷).

### پیشینه پژوهش‌های خارجی

نویسنده	خلاصه موضوع	نتیجه
سولاندی و دیگران (۲۰۲۰)	SSA را با روش‌های دیگری برای بهبود عملکرد مدل پیش‌بینی برای سری‌های زمانی با الگوی پیچیده ترکیب نمودند.	با ترکیب این سه روش به نتایج دقیق‌تری در مقایسه با نتایج به‌دست آمده توسط هر روش به‌طور جداگانه، دست یافتند (۳۹).
لهمیری (۲۰۱۸)	ارائه مدلی جهت پیش‌بینی قیمت سهام با استفاده از تجزیه و تحلیل طیفی تکین و رگرسیون بردار پشتیبان همراه با بهینه‌سازی ازدحام ذرات	نتایج پیش‌بینی حاصل از شبیه‌سازی‌ها نشان می‌دهد که روش ارائه شده تا حد زیادی بهتر از روش تحلیل طیفی تکین استاندارد است (۳۲).
آرتچی و انریکز (۲۰۱۷)	تحلیل نوسان روزانه شاخص سهام صنعتی داوجونز و شاخص سهام <i>Ibex35</i> اسپانیا را با تحلیل طیفی تکین درون نمونه‌ای و کولموگروف وینر	روش تحلیل طیفی تکین عملکرد بالاتری دارد (۱۵).
عثمان زاده و دیگران (۲۰۱۷)	بررسی سهام تالینک‌گروپ با استفاده از روش تحلیل طیفی تکین برای	جهت پیش‌بینی با دقت بالا با استفاده از روش تحلیل طیفی تکین نیاز به مطالعه و بررسی بیشتر در مرحله‌ی بازسازی ماتریس‌های مسیراست (۳۵).
سیریمال و حسنی (۲۰۱۵)	استفاده از روش تحلیل طیفی تکین (SSA) جهت پیش‌بینی معاملات و ارزیابی اثر آن بر بحران مالی ۲۰۰۸ بر ایالات متحده	این روش می‌تواند پیش‌بینی دقیقی از معامله در مواقع بحران ایجاد نماید (۳۸).

### روش‌شناسی پژوهش

پژوهش حاضر از حیث روش، توصیفی و به لحاظ هدف کاربردی می‌باشد. دامنه مکانی آن محدود به بورس اوراق بهادار تهران و دامنه زمانی آن ده ساله (ابتدای فروردین ۱۳۸۸ تا پایان اسفند

می‌دهد. ماتریس مسیر سری‌های  $X$  به صورت رابطه زیر (۱) تعریف می‌شود:

$$X = [X_1, \dots, X_N] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_k \\ x_2 & x_3 & x_4 & \dots & x_{k+1} \\ x_3 & x_4 & x_5 & \dots & x_{k+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}$$

از آنجایی که تمامی مؤلفه‌های  $i+j$  آن مقداری ثابت است و تمامی مؤلفه‌های روی قطرهای فرعی نیز با هم برابرند. ماتریس مسیر  $X$  یک ماتریس هنگل<sup>۱۱</sup> است (۲۳).

مرحله اول: گام دوم: تجزیه مقدار تکین<sup>۱۲</sup>: در این گام، ماتریس

مسیر به دست آمده توسط گام اول، بر اساس مقادیر ویژه<sup>۱۳</sup>،

تجزیه و به ماتریس‌های پایه دوجه‌دو متعامد با رتبه‌ی یک تبدیل

می‌شود. ابتدا مقادیر ویژه  $XX^T$  محاسبه و سپس به ترتیب

نزولی  $0 \leq \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_L$  مرتب می‌شود. مجموعه‌ی

بردارهای متعامد  $U_1, \dots, U_L$  که ضرب دوجه‌دو آن‌ها برابر با

صفر می‌باشد را به عنوان بردارهای ویژه  $XX^T$  متناظر با این

مقادیر ویژه در نظر گرفته می‌شود، تعریف می‌شود:

$$d = \max(i, s.t \alpha_i \geq 0) = r(X) = \min\{L, N-L + 1\}$$

و  $v_i$  به صورت  $v_i = \frac{X^T U_i}{\sqrt{\alpha_i}}$  لحاظ می‌گردد. حال تجزیه مقدار تکین

ماتریس مسیر  $X$  را می‌توان همانند رابطه‌ی زیر (۲):

$$X = X_1 + X_2 + \dots + X_d$$

نوشت که در آن  $V_i^T X_i = \sqrt{\alpha_i} U_i$  و  $i=1, \dots, d$  می‌باشد.

ماتریس‌های  $X_i$  دارای رتبه‌ی یک هستند، بنابراین این

ماتریس‌ها، ماتریس‌های پایه<sup>۱۴</sup> هستند.  $U_i$  و  $V_i$  ها به ترتیب

بردارهای ویژه چپ (مؤلفه‌های اصلی<sup>۱۵</sup>) و راست ماتریس مسیر

و مجموعه‌ی  $(\sqrt{\alpha_i}, U_i, V_i^T)$  امین سه‌تایی ویژه<sup>۱۶</sup>

ماتریس  $X$  نامیده می‌شود و هم‌چنین مقادیر  $\sqrt{\alpha_i}$  مقادیر

تکین ماتریس  $X$  و مجموعه‌ی تمامی آنها را طیف<sup>۱۷</sup> ماتریس  $X$

می‌نامند (۲۳).

مرحله دوم: بازسازی<sup>۱۸</sup>: گام اول: گروه‌بندی<sup>۱۹</sup>: مرحله‌ی

گروه‌بندی به معنی تقسیم ماتریس‌های پایه  $X_i$  به چندین گروه

و جمع کردن ماتریس‌های درون هر یک از گروه‌ها با یکدیگر است.

این مرحله، در واقع به دنبال جداسازی داده‌های نویز از سیگنال

است. فرض کنید  $p, I = \{I_1, I_2, \dots, I_p\}$  گروه مجزا از

اندیس‌های  $\{1, \dots, d\}$  باشد، که در آن هر  $I_i$  شامل چندین

فرض کنید  $U_i V_i^T X_i = \sqrt{\alpha_i} I_i$  است. فرض کنید  $i_1, \dots, i_r$  مجموعه‌ی

نشانگرهای  $i_1, \dots, i_r$  باشد، بسط ماتریس  $X_{II}$  در گروه  $I_i$  به شکل

$$X_{II} = X_{I_1} + X_{I_2} + \dots + X_{I_r}$$

سال ۱۳۹۷) می‌باشد. از لحاظ نوع مطالعه نیز به صورت

کتابخانه ای می‌باشد که داده های مورد پژوهش از سایت

[www.tsetmc.com](http://www.tsetmc.com) که تحت نظارت شرکت مدیریت فناوری

بورس تهران می‌باشد، جمع‌آوری شده است. در نهایت تجزیه و

تحلیل اطلاعات نیز با استفاده از نرم‌افزار متلب جهت پیش‌بینی

و بررسی روش‌های تحقیق صورت گرفته است.

### جامعه و نمونه آماری

جامعه آماری، قیمت روزانه شاخص کل بورس اوراق بهادار تهران به

هفت روز هفته در بازه زمانی ده ساله (ابتدای فروردین ۱۳۸۸ تا آن

پایان اسفند سال ۱۳۹۷) می‌باشد که نمونه پژوهش نیز بازه

قیمت روزانه شاخص کل بورس اوراق بهادار تهران با حجم نمونه

۲۴۱۱ می‌باشد. ضمناً در روزهای متوالی که بازه لگاریتمی ثابت می‌باشد

داده مورد نظر حذف گردیده است.

### معرفی روش تحلیل طیفی تکین

دو چالش اصلی که روش تحلیل طیفی تکین در تلاش بر حل آن

ها است، یکی شناسایی مولفه‌هایی از سری‌های زمانی که قابل

جداشدن هستند و دیگری، مناسب‌ترین انتخاب برای پارامتر

طول پنجره و در واقع مناسب‌ترین گروه بندی برای استخراج

یک مولفه ی مطلوب است (۲۱). بدین منظور، الگوریتم این

روش شامل دو مرحله است: تجزیه و بازسازی. مرحله اول با

تبدیل داده‌ها به ماتریس مسیر و تجزیه آن‌ها به مقادیر تکین

اجرا می‌شود و گام دوم، به نوبه خود، به گروه‌بندی مؤلفه‌های

ماتریس مسیر تجزیه شده برای تشکیل زیرگروه‌هایی

می‌پردازد که سری جدید از آن‌ها ساخته می‌شود (۲۶). در

ادامه، مراحل الگوریتم تحلیل طیفی به‌طور کامل شرح داده

می‌شود:

مرحله اول: تجزیه<sup>۷</sup>: گام اول: نشانیدن<sup>۸</sup>: هدف از این مرحله،

محاسبه‌ی ماتریس مسیر<sup>۹</sup> از روی سری زمانی اولیه می‌باشد.

فرض کنید  $X = (x_1, \dots, x_N)$  یک سری زمانی با طول  $N$  که

در آن  $N$  بزرگ‌تر از ۲ است، باشد. فرض کنید  $L$  به‌طوری که  $L > 1$

$L < N$ ، یک عدد صحیح باشد که طول پنجره<sup>۱۰</sup> نامیده می‌شود

و  $K = N - L + 1$  را در نظر بگیرید. برای انجام عمل نشانیدن،

سری اولیه به مجموعه‌ای از بردارهایی به‌اندازه  $L$  تبدیل

می‌شود، به‌طوری که،  $X_i = (x_i, \dots, x_{i+L-1})^T$  ( $1 \leq i \leq K$ ).

در واقع مرحله‌ی نشانیدن را می‌توان همانند نگاشتی تصور کرد

که سری زمانی یک‌بعدی را به سری زمانی چند بعدی انتقال

می‌توان از آن برای مدل‌سازی یا پیش‌بینی، که در بخش بعد به آن اشاره می‌گردد، استفاده کرد.

### پیش‌بینی با روش تحلیل طیفی تکین

برای پیش‌بینی داده‌ها در روش تحلیل طیفی تکین از روابط بازگشتی خطی<sup>۲۳</sup> استفاده می‌شود، که دنباله‌ای است از سری‌های زمانی که به صورت بازگشتی بیان می‌شود. در این جا باید این موضوع را بیان نمود که سری‌های  $Y_T$  یک رابطه‌ی بازگشتی خطی از مرتبه  $d$  خواهند داشت، اگر اعداد  $a_1, a_2, \dots, a_d$  وجود داشته باشند به طوری که:

$$y_{i+d} = \sum_{k=1}^d a_k y_{i+d-k}, \quad 1 \leq i \leq N-d$$

این عبارت به این معنی هست که آخرین مؤلفه‌ی آن را می‌توان به صورت یک ترکیب خطی منحصر به فرد از سایر مؤلفه‌ها بیان کرد (۱۷). در نتیجه، هر دنباله‌ی گروه‌های ایجادشده که در شرط بالا (رابطه ۷) صدق می‌کند، توسط یک رابطه بازگشتی خطی می‌تواند ادامه یابد. حال فرض نمایید سری‌زمانی‌ای وجود دارد که در شرط (رابطه ۷) صدق می‌کند و هدف محاسبه‌ی  $h$  جمله‌ی دیگر سری است. در این صورت می‌توان از رابطه‌ی زیر (رابطه ۸) استفاده کرد:

$$y_i = \begin{cases} \tilde{y}_i & \text{for } i = 1, \dots, N \\ \sum_{j=1}^{L-1} a_j y_{i-j} & \text{for } i = N+1, \dots, N+h \end{cases}$$

که در اینجا  $\tilde{y}_i$  به ازای هر  $(i=1, \dots, N)$ ، مؤلفه‌های سری بازسازی‌شده هستند و  $y_{N+1}, \dots, y_{N+h}$  سری‌های جدیدی می‌باشند که توسط روش تحلیل طیفی تکین پیش‌بینی شده است [۱]. سؤال اساسی که در این مرحله مطرح می‌شود این است که ضرایب ماتریس به چه صورت محاسبه می‌شود. فرض کنید  $R = (a_1, \dots, a_{L-1})$  مجموعه‌ی ضرایب باشد، در این صورت از رابطه زیر (۹) می‌تواند محاسبه شود:

$$R = \frac{1}{1 - (\pi_1^2 + \dots + \pi_L^2)^2} \sum_{i=1}^L \pi_i U_i^{\nabla}$$

که در آن  $\{U_1, \dots, U_L\}$  یک پایه متعامد یک باشد و  $\pi_i$  آخرین مؤلفه‌ی بردار  $U_i$  هستند [۱۷].

### انتخاب پارامترهای روش تحلیل طیفی تکین

در روش تحلیل طیفی تکین دو پارامتر وجود دارد: یکی طول پنجره ( $L$ ) که در مرحله‌ی تجزیه مورد استفاده قرار می‌گیرد و پارامتر دوم تعداد مؤلفه‌ها ( $r$ ) برای بازسازی مجدد سری است. مقادیر  $L$  و  $r$  را می‌توان با استفاده از اطلاعات ارائه شده توسط سری‌های زمانی مورد مطالعه و یا از طریق شاخص‌های اضافی تعریف کرد (۲۵). انتخاب پارامترها بستگی به داده‌هایی دارد که

نتیجه‌ی تقسیم مجموعه اندیس‌های  $I=1, 2, \dots, d$  به زیرمجموعه‌های مجزای  $I_1, I_2, \dots, I_p$  ماتریس  $X$  را می‌توان به صورت رابطه (۳) زیر نمایش داد:

$$X = X_1 + X_2 + \dots + X_d = X_{I_1} + X_{I_2} + \dots + X_{I_p}$$

هدف گام گروه‌بندی، به دست‌آمده آوردن چندین گروه  $I_1, I_2, \dots, I_p$  است به طوری که ماتریس‌های  $X_{I_1}, X_{I_2}, \dots, X_{I_p}$  در رابطه (۱) صدق کنند و نزدیک به ماتریس‌های هنکل اصلی باشند. شیوه‌ی انتخاب مجموعه‌های  $I_1, I_2, \dots, I_p$  گروه‌بندی ویژه سه‌گانه<sup>۲۰</sup> نامیده می‌شود. اگر  $p=d$  به طوری که  $I_j = j$  به ازای هر  $j=1, \dots, d$  آنگاه این نوع گروه‌بندی به گروه‌بندی پایه معروف است. برای اندازه‌گیری میزان مشارکت مؤلفه‌ی  $X_I$  در رابطه (۳) به ازای یک گروه مشخص  $I$ ، از نسبت مقادیر ویژه متناظر یعنی  $\frac{\sum_{i \in I} \alpha_i}{\sum_{i=1}^d \alpha_i}$  استفاده می‌شود (۲۵). برای مثال، اگر  $P=2$  باشد، آن‌گاه گروه اول همان مؤلفه‌های اصلی (سیگنال) و گروه باقی‌مانده، نویز در نظر گرفته می‌شود (۲۳).

مرحله دوم: بازسازی گام دوم: میانگین‌گیری قطری<sup>۲۱</sup> (هنکل‌سازی<sup>۲۲</sup>): هر ماتریس هنکل یک رابطه‌ی یک‌به‌یک با یک سری زمانی دارد، این همان دلیلی است که در گام اول مرحله‌ی تجزیه، به راحتی سری اولیه را به ماتریس هنکل تبدیل می‌کند. ولی از آنجایی که گروه بندی مرحله‌ی قبل یک ماتریس هنکل نیست، پس باید مرحله‌ای وجود داشته باشد که ابتدا آن را به یک ماتریس هنکل تبدیل کند تا بتوان در ادامه آن را به یک سری زمانی برگرداند. یک روش بهینه برای این مشکل ارائه داده شده است که به میانگین‌گیری قطری معروف است. با اعمال روش هنکل‌سازی بر روی همه‌ی مؤلفه‌های ماتریس رابطه ۱، رابطه‌ی دیگری همانند رابطه زیر (۴) به دست‌آمده می‌آید:

$$X = \hat{X}_{I_1} + \hat{X}_{I_2} + \dots + \hat{X}_{I_p}$$

که در آن  $\hat{X}_{I_k} = HX_k$  است.

این عمل معادل با تجزیه‌ی سری اولیه  $(X=(x_1, \dots, x_N))$  به مجموع  $p$  سری در معادله زیر (رابطه ۵) است:

$$x_n = \sum_{k=1}^p \hat{x}_N^{(k)}$$

که در آن

$\hat{X}_N^{(k)} = (\hat{x}_1^{(k)}, \hat{x}_2^{(k)}, \dots, \hat{x}_p^{(k)})$ ،  $X_{I_k}$  ماتریس متناظر با ماتریس  $X_{I_k}$  است. یک گروه‌بندی مناسب، رابطه ۳، را به ماتریس‌های  $X_{I_k}$  که تقریباً هنکل هستند، تجزیه می‌کند (۲۱). با توجه به انتخاب مناسب  $r$  تا از  $m$  زیر سری، سری بازسازی‌شده از مرحله‌ی میانگین‌گیری قطری، یک سری با سطح نویز کم می‌باشد و لذا

در اختیار است و تجزیه و تحلیلی که باید بر روی آن‌ها صورت گیرد (۲۱).

### بررسی شکاف‌ها در طیف مقادیر ویژه و بردارهای ویژه:

سری نوین معمولاً دنباله‌هایی از مقادیر تکین، که به‌کندی کاهش می‌یابند را تولید می‌کند. به بیانی دیگر در صورتی که مقادیر تکین به صورت نزولی مرتب و رسم شوند، می‌توان دو قطعه متفاوت را در شکل حاصل از مقادیر تکین مشاهده کرد. یکی با شیب تند که نمایانگر مؤلفه‌های سیگنال است و دیگری بخش هموارتر که مؤلفه‌های نوین در سری‌های زمانی می‌باشد.

### بررسی ماتریس مقادیر قدر مطلق $W$ - همبستگی: در این

حالت همبستگی میان مؤلفه‌ها را می‌توان به‌عنوان معیار سنجشی بر تشخیص سیگنال از نوین دانست. مؤلفه‌هایی که همبستگی آن‌ها با سایر گروه مؤلفه‌ها (که بینشان همبستگی زیادی وجود دارد)، کمترین میزان است، مؤلفه‌های سیگنال و سایر مؤلفه‌ها، نوین در نظر گرفته می‌شود.

بررسی نمودار مؤلفه اصلی ویژه سه‌گانه‌ها: مؤلفه‌هایی که نمودار مؤلفه ویژه سه‌گانه‌های آن‌ها دارای تغییرات آرامی باشد، سیگنال در نظر گرفته می‌شود و اگر تغییرات زیادی مشاهده شود به عنوان مؤلفه نوین لحاظ می‌گردد (۳۶).

### انتخاب طول پنجره: اندازه ابعاد در مرحله‌ی نشانیدن یا همان

طول پنجره  $L$  یک پارامتر اساسی روش تحلیل طیفی تکین است. حسنی (۲۰۱۱) و مندز (۲۰۱۵)، هر کدام به‌طور جداگانه در مقالات خود ثابت کرده‌اند که بهترین انتخاب برای طول پنجره  $L$  با توجه به تعداد داده‌های سری زمانی، می‌تواند برابر با میان‌ی آن در نظر گرفته شود. گاهی سری‌های زمانی ممکن است مؤلفه‌ی دوره‌ای یا دوره‌ای صحیح داشته باشند، در این شرایط می‌توان طول پنجره را متناسب با دوره انتخاب کرد (۲۸).

به‌طور کلی، مقادیر زیاد  $L$ ، تجزیه منقطع تری را به اجزای اولیه ارائه می‌دهد و در نتیجه، قابلیت تفکیک بهتر را ایجاد می‌کند (۲۷). گلوپاندینیا و دیگران (۲۰۰۱) توصیه نمودند که  $L \sim \frac{N}{2}$  انتخاب شود (۲۱). حسنی (۲۰۰۷) با تمرکز بر قابلیت تفکیک و تفسیر اجزای به‌دست‌آمده،  $L = 24$  را نشان می‌دهد.

با این حال، هنگامی که هدف استخراج سیگنال و پیش‌بینی باشد،  $L = 14$  را پیشنهاد کردند (۲۹). سرانجام، همان‌طور که توسط گلوپاندینیا و ژینگلیاوسکی (۲۰۱۳) اشاره شد، هیچ قاعده‌ی کلی برای انتخاب طول پنجره وجود ندارد. با این حال، چندین اصل کلی برای انتخاب طول پنجره  $L$  وجود دارد که دارای مبانی نظری و عملی خاصی هستند و هر شرایطی که

باشد، همیشه ایده خوبی این است که چندین بار تحلیل طیفی تکین را با استفاده از مقادیر مختلف  $L$  آزمون نمودند (۲۳). اخیراً آبرئو و دیگران (۲۰۱۹) نیز در مقاله‌ای به بررسی هرچه بیشتر طول پنجره پرداخته‌اند و با این هدف که بهترین طول پنجره را به دست آورند (۱۰). از حداکثر مقدار پیشنهادی حسنی (۲۰۰۷) که برابر نصف تعداد داده‌های سری زمانی بود تا کمترین مقدار ممکن مورد بررسی قرار گرفته است. آن‌ها در این روش از ماتریس همبستگی وزنی بردارها استفاده می‌کنند و به این نتیجه رسیدند که هرچه طول پنجره بیشتر باشد، تأثیر زیادی در سری قابل تفکیک دارد (۲۵).

### انتخاب پارامتر در مرحله‌ی گروه‌بندی: مسئله دیگر مربوط

به انتخاب ماتریس‌های پایه است که برای مرحله‌ی بازسازی در مرحله‌ی گروه‌بندی استفاده می‌شود. رهنمودهای مربوط به انتخاب اندازه‌ی گروه‌بندی، به طور کلی در مقاله‌ی الهاری و حسنی (۲۰۱۶) یافت می‌شود (۱۳). در این زمینه، یک رویکرد باینری وجود دارد که مبتنی بر این فرض است که این سری توسط یک سیگنال به همراه نوین تشکیل می‌شود. از آنجایی که مؤلفه‌های سیگنال غالباً غالب هستند، تنها پارامتر گروه‌بندی تعداد مؤلفه‌های اصلی، یعنی اندازه‌ی گروه‌بندی است (۲۹). الکساندروف (۲۰۰۹) روشی را برای انتخاب گروه‌بندی خودکار بر اساس دوره نگار بردارهای ویژه برای استخراج روند و اجزای هارمونیک پیشنهاد داد (۱۱). آلسو و سالگادو (۲۰۰۸) از تکنیک‌های خوشه‌ای برای گروه‌بندی اجزای اولیه بر اساس روش خوشه‌بندی کی- میانگین استفاده کردند (۱۴).

### تفکیک‌پذیری سیگنال و نوین

اساسی‌ترین مرحله در روش تحلیل طیفی تکین، تعیین مقدار بهینه برای دو پارامتر طول پنجره و گروه‌بندی می‌باشد. اگر طول پنجره به‌درستی انتخاب نشده باشد مؤلفه‌ی سیگنال و نوین را نمی‌توان به‌خوبی از سیگنال اصلی استخراج کرد. حسنی و محمودوند (۲۰۱۰)، از مفهوم تفکیک‌پذیری بین مؤلفه‌ی سیگنال و نوین استفاده کردند و با بررسی‌های تحلیلی بر روی سری‌های زمانی نشان دادند که مناسب‌ترین مقدار برای طول پنجره، میان‌ی سری زمانی می‌باشد. در شرایط ایده آل می‌توان مؤلفه‌های نوین رو حذف کرد و عمل بازسازی رو به‌راحتی انجام داد ولی در شرایط واقعی بازسازی مؤلفه‌های سیگنال به‌طور کامل غیرممکن است، اما باین‌حال می‌توان نزدیک‌ترین  $T$  محاسبه کرد (۲۷).

مثل<sup>۲۴</sup> از عملگرهای ژنتیکی استفاده می‌شود. با تأثیر این عملگرها بر روی یک جمعیت، نسل<sup>۲۵</sup> بعدی آن جمعیت تولید می‌شود. عملگرهای انتخاب<sup>۲۶</sup>، ترکیب<sup>۲۷</sup> و جهش<sup>۲۸</sup> معمولاً بیشترین کاربرد را در الگوریتم‌های ژنتیکی دارند (۲۰). روش‌های انتخاب مورد نظر در این الگوریتم‌های ژنتیکی روش چرخ رولت می‌باشد بدین صورت که در نسخه اولیه الگوریتم ژنتیک که توسط جان هولند پیشنهاد شد، از انتخاب متناسب با برازندگی استفاده شده است که در آن نرخ انتظار هر فرد از تقسیم برازندگی آن فرد بر متوسط برازندگی‌های جمعیت به دست می‌آید. برای پیاده‌سازی انتخاب متناسب با برازندگی، روش‌های نمونه‌برداری متعددی از قبیل چرخ رولت<sup>۲۹</sup> پیشنهاد شد است. در این روش، به هر فرد قطعه‌ای<sup>۳۰</sup> از یک چرخ رولت مدور اختصاص داده می‌شود. اندازه‌ی این قطعه متناسب با برازندگی آن فرد است. چرخ<sup>۳۱</sup>  $N$  بار چرخانده می‌شود که  $N$  تعداد افراد در جمعیت است. در هر چرخش، فرد زیر نشانگر چرخ انتخاب می‌شود و در مخزن والدین نسل بعد قرار می‌گیرد. این روش می‌تواند به صورت (نرخ انتظار کل افراد جمعیت را جمع کرده و حاصل آن را  $T$  می‌نامند و با تکرار  $N$  بار از مراحل زیر، پیاده‌سازی می‌شود: یک عدد تصادفی  $r$  بین  $0$  و  $T$  انتخاب می‌شود. در میان افراد جمعیت جست‌وجو کرده و نرخ‌های انتظار آن‌ها با هم جمع می‌شود تا این که مجموع بزرگ‌تر یا مساوی  $r$  شود. فردی که نرخ انتظارش باعث بیشتر شدن حاصل جمع از این حد شود، به عنوان فرد برگزیده انتخاب می‌شود (۴۳).

### تحلیل طیفی تکین هم پوشانی

ایده اصلی در این پژوهش بهبود تحلیل طیفی تکین با کمک عمل تقسیم‌بندی بر روی سری زمانی و بررسی هم پوشانی بین قسمت‌های ایجاد شده، می‌باشد. بدیهی است که روش تقسیم‌بندی، سری‌های زمانی کوچکتری تولید می‌کند که برای تحلیل طیفی تکین مطلوب هستند. سری‌های زمانی اصلی، با طول  $N$ ، به بخش‌های کوچکتر و متوالی با طول ثابت  $Z$  تقسیم می‌شوند. برای هر بخش، الگوریتم استاندارد تحلیل طیفی تکین محاسبه می‌شود و یک سری زمانی محلی بازسازی می‌شود. مهمترین بخش در این مرحله چگونگی در کنار هم قرار دادن بخش‌های جدا شده است. از آنجایی که در تحلیل طیفی تکین اثرات مرزی وجود دارد، نقاط شدید در لبه‌های چپ و راست به منظور بازسازی معنی‌دار نیستند، بنابراین فقط یک زیرمجموعه داخلی از نمونه  $q$  برای نشان دادن سری زمانی محلی معنی‌دار تلقی می‌شود. برای حل این مشکل، رویکردی با نام روش

گلوپاندینا و دیگران (۲۰۰۷)، برای رفع این مشکل از معیاری بنام روابط همبستگی وزنی (معیاری برای میزان شباهت) استفاده نمودند. میزان شباهت بین دو سری زمانی  $Y_T^{(1)}$  و  $Y_T^{(2)}$  برابر است با رابطه زیر (۱۰): (۲۲).

$$\rho_{12}^{(w)} = \frac{(Y_T^{(1)}, Y_T^{(2)})_w}{\|Y_T^{(1)}\|_w \|Y_T^{(2)}\|_w}$$

که در آن

$$\|Y_T^{(i)}\|_w = \sqrt{(Y_T^{(i)}, Y_T^{(i)})_w} \quad \text{و} \quad (Y_T^{(i)}, Y_T^{(j)})_w = \sum_{p=1}^T w_p^L y_p^{(i)} y_p^{(j)}$$

$$\text{برای } (i, j=1, 2) \text{ و } w_p^L = \min\{p, L, T - p + 1\}0$$

اگر همبستگی وزنی بین  $Y_T^{(1)}$  و  $Y_T^{(2)}$  مینیمم باشد آنگاه سری‌های متناظر با این دو سری تقریباً متعامد هستند و در نتیجه تفکیک پذیری آن‌ها قوی است و می‌توان مؤلفه‌های سیگنال و نویز را با استفاده از این اصل به راحتی از هم تفکیک کرد. ولی اگر همبستگی وزنی بزرگ باشد، به طور ضعیفی قابل تفکیک هستند (۲۱).

با استفاده از قضیه بالا می‌توان رابطه همبستگی وزنی برای سیگنال و نویز را به صورت زیر بیان (رابطه ۱۱) کرد: (۲۱).

$$\text{Cost function} = \frac{(\text{signal}, \text{noise})_w}{\|\text{signal}\|_w \|\text{noise}\|_w} = \widehat{N}_L^r + \widehat{S}_L^r$$

$$(\text{signal}, \text{noise})_w = \sum_{k=1}^N w_k s_k n_k, w_k = \min\{K, L, N - K\}$$

در این پژوهش از این معیار به عنوان تابع هزینه برای محاسبه‌ی نقطه‌ی برش و طول پنجره‌ی بهینه با استفاده از الگوریتم ژنتیک استفاده شده است. الگوریتم ژنتیک با الهام از طبیعت که در قالب الگوریتمی با روند جستجوی تصادفی است، این جریان تکاملی طبیعت را تقلید و شبیه سازی می‌کند و با یک مجموعه از جواب‌ها که از طریق کروموزوم‌ها نشان داده می‌شوند، شروع می‌شود. این مجموعه جواب‌ها جمعیت اولیه نام دارد. در این الگوریتم جواب‌های حاصل از یک جمعیت برای تولید جمعیت بعدی استفاده می‌شوند. در این فرایند امید است که جمعیت جدید نسبت به جمعیت قبلی بهتر باشد. انتخاب بعضی از جواب‌ها از میان کل جواب‌ها (والدین) به منظور ایجاد جواب‌های جدید یا همان فرزندان بر اساس میزان برازندگی آن‌هاست. طبیعی است که جواب‌های مناسب‌تر، شانس بیشتری برای تولید مجدد داشته باشند. این فرایند تا برقراری شرطی که تعیین شده است (مانند تعداد جمعیت‌ها یا میزان بهبود جواب) ادامه پیدا می‌کند (۴۳). در الگوریتم‌های ژنتیکی، در طی مرحله‌ی تولید

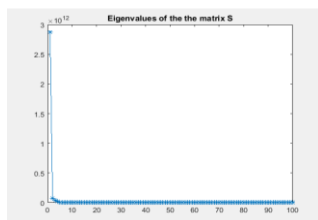


$$MAE = \frac{1}{N} \sum_{n=1}^N |x_n - y_n|$$

که در آن  $x_n$  یک نمونه از سری زمانی اصلی در زمان  $n$ ،  $y_n$  نمونه بازسازی شده سری زمانی در همان زمان و  $N$  نشان‌دهنده‌ی اندازه سری زمانی اصلی است (۲۱).

### نتایج تجربی تحلیل شاخص کل بازار با استفاده از تحلیل طیفی تکین

تعداد داده‌های در این بازه زمانی برای شاخص کل بازار برابر با ۲۴۱۱ می‌باشد که این بدین معنی است که طول اولیه سری زمانی برابر با  $N=2411$  است. در این جا طول پنجره را برابر با میانه سری زمانی  $N=2411$  که برابر با ۱۲۰۶ می‌باشد، در نظر گرفته می‌شود. ابتدا مقادیر ویژه و بردارهای ویژه متناظر با آن را محاسبه و نمودار هر کدام ترسیم می‌شود. همان‌طور که در شکل ۲ مشاهده می‌شود، ۳ مقدار ویژه ابتدا دارای شیب تندی هستند ولی در ادامه حرکت نمودار بسیار ملایم و حتی نزدیک به نمودار افقی می‌شود.

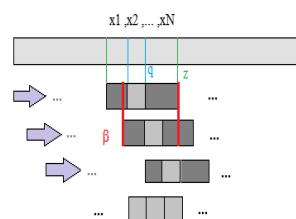


شکل ۲- نمودار نیمه لگاریتمی ۱۰۰ مقدار ویژه اول

از آنجایی که سری نویز

معمولاً دنباله‌ای از مقادیر ویژه را تولید می‌کند که به کندی کاهش می‌یابند، در شکل ۳ می‌توان دو قطعه متفاوت را مشاهده کرد که یکی با شیب تند که نمایشگر سیگنال است و دیگری بخش هموارتر که نمایشگر نویز در سری‌های زمانی می‌باشد که این قضیه نشان می‌دهد، ۳ مولفه ابتدا را می‌توان به عنوان مولفه‌های سیگنال در نظر گرفت و مابقی مولفه‌ها، نقش مولفه‌های نویز را برعهده می‌گیرند. حال زمان آن رسیده که نمودار بردارهای ویژه متناظر با این مقادیر ویژه ترسیم شود. به دلیل این که تعداد بردارهای ویژه بسیار زیاد است، در واقع به‌طور دقیق‌تر برابر با ۹۴۵ می‌باشد، هشت بردار ویژه ابتدا و انتها بررسی می‌شود.

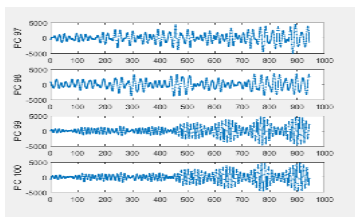
صرفه‌جویی در هم‌پوشانی<sup>۳۱</sup> ارائه شده است. بخش‌های متوالی با یکدیگر هم‌پوشانی دارند و نمونه‌های اولیه و نهایی یک بخش داده شده کنار گذاشته می‌شوند. این اجازه می‌دهد تا نمونه‌هایی که حفظ شده‌اند، به هم پیوند داده شوند و یک سری زمانی بازسازی شده با همان اندازه سری اصلی ایجاد شود. شکل ۱ می‌تواند در درک بهتر این روش به ما کمک نماید. یک بخش محلی را با طول  $Z$  در نظر بگیرید.



همه شکل ۱- تحلیل طیفی تکین تقسیم شده نمونه‌های این بخش برای محاسبه‌ی تحلیل طیفی تکین استفاده می‌شود اما فقط  $q$  نمونه معنی‌دار در نظر گرفته می‌شود (قسمت‌های خاکستری روشن‌تر)، در نتیجه  $\bar{L} = \frac{(Z-q)}{2}$  نمونه دور ریخته شده (قسمت‌های خاکستری تیره). توجه داشته باشید که بخش‌های محلی  $Z$  که برای تجزیه و تحلیل استفاده می‌شود، با درصد  $\beta = \frac{100(Z-q)}{Z}$  هم‌پوشانی دارند. با این حال، بخش‌های محلی  $q$  که بازسازی شده‌اند با هم تداخل ندارند و می‌توان به راحتی در توالی آن‌ها را، به هم پیوند داد و کل سری زمانی بازسازی شده را به صورت رنگ‌های خاکستری روشن که در پایین شکل ۱ رسم شده است، ایجاد کرد. در واقع زمانی که  $Z=N$  است، یعنی برابر با طول سری زمانی اصلی است، تحلیل طیفی تکین هم‌پوشانی نتایجی مشابه با تحلیل طیفی استاندارد را می‌دهد که این بدین معنی است که روش تحلیل طیفی تکین استاندارد حالت خاصی از روش تحلیل طیفی هم‌پوشانی است (۳۳ و ۳۴).

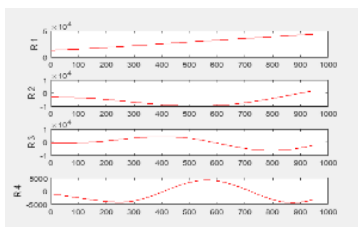
### بررسی عملکرد روش تحلیل طیفی تکین هم‌پوشانی

عملکرد بازسازی را می‌توان با معیارهای استاندارد خطا، از جمله خطای قدرمطلق میانگین<sup>۳۲</sup> بررسی کرد و هرچه میزان این خطا کمتر باشد، نشان از عملکرد مناسب این روش دارد. خطای مطلق میانگین به وسیله‌ی رابطه‌ی زیر (۱۲) محاسبه می‌شود:

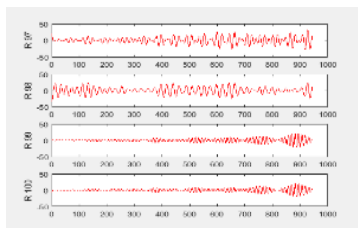


شکل ۶- مولفه‌های اصلی سری‌های زمانی برای ۴ مقدار

همان‌طور که در دو شکل ۸ و ۷ مشاهده می‌شود، عمل تجزیه و بازسازی تا حد زیادی در جداسازی سیگنال از نویز در نتایج تاثیر داشته است. اگر چهار نمودار ابتدای مولفه‌های اصلی را با چهار نمودار ابتدا در بردارهای ویژه مقایسه شود، کاهش تغییرات مولفه‌ها آشکار است و هم‌چنین این تغییرات در چهار مولفه انتهایی هم به وضوح دیده می‌شود.

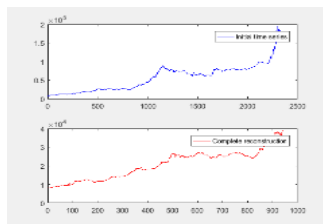


شکل ۷- مولفه‌های بازسازی شده سری‌های زمانی برای ۴ مقدار ویژه ابتدا

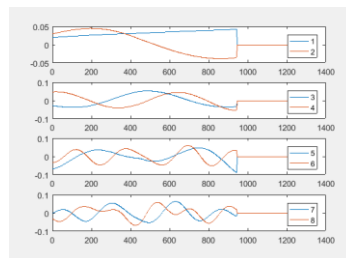


شکل ۸- مولفه‌های بازسازی شده سری‌های زمانی برای ۴ مقدار ویژه انتها

حال در شکل ۹، نمودار اولیه سری‌های زمانی با سری‌های زمانی ساخته شده مقایسه شده است که هماهنگی بسیاری بین آن دو مشاهده می‌شود. اکنون نوبت به آن رسیده است که با توجه به داده‌هایی که تاکنون به‌دست آمده، نتایج بعدی پیش‌بینی شود. در این مرحله، ۴۰ بازه زمانی که در این جا ۴۰ روز کاری تعبیر شده، پیش‌بینی می‌شود.

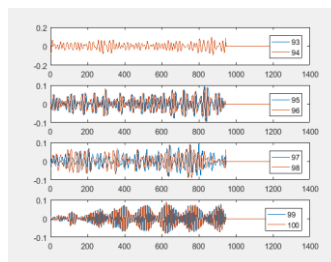


شکل ۹- مقایسه سری‌های زمانی اولیه و با ۳ مولفه‌های بازسازی شده



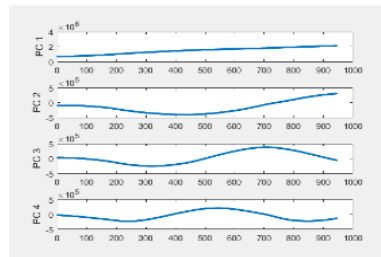
شکل ۳- بردارهای ویژه متناظر با هشت مقدار ویژه ابتدایی

شکل ۳ و ۴ به ترتیب بردارهای ویژه متناظر با هشت مقدار ویژه ابتدایی و بردارهای ویژه متناظر با هشت مقدار ویژه انتهایی را نشان می‌دهند. در ادامه ماتریس مسیر تجزیه می‌شود که بخشی از این تجزیه، مولفه‌های اصلی نامیده می‌شود. مجدد مولفه‌های اصلی برای چهار مقادیر ویژه ابتدا و انتها رسم می‌شود تا موجب اطمینان بیشتری در تشخیص مولفه‌های سیگنال از نویز شود. همانند نمودارهای بردارهای ویژه، مولفه‌های نویز برخلاف مولفه‌های سیگنال تغییرات بسیاری دارند.



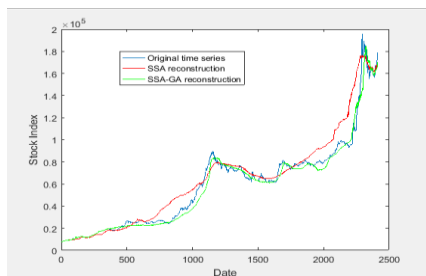
شکل ۴- بردارهای ویژه متناظر با هشت مقدار ویژه انتهایی

شکل ۵، مولفه‌های اصلی سری‌های زمانی برای چهار مقدار ویژه ابتدایی و شکل ۶، مولفه‌های اصلی سری‌های زمانی برای چهار مقدار ویژه انتهایی را نشان می‌دهند. برای تعیین مولفه‌های بازسازی شده ابتدا باید ماتریس  $L \times (N - L + 1)$  با استفاده از ترانهاده ماتریس مولفه‌های اصلی و ماتریس بردارهای ویژه محاسبه شود. سپس باید روش میانگین‌گیری یا همان هنکلسازی بر روی قطر فرعی این ماتریس که از مولفه‌های بازسازی شده برای سری اولیه ساخته شده، اعمال گردد.



شکل ۵- مولفه‌های اصلی سری‌های زمانی برای ۴ مقدار ویژه ابتدا

پس از محاسبه پارامترهای روش تحلیل طیفی تکین، مجدد الگوریتم روش تحلیل طیفی تکین بر روی سری زمانی شاخص کل بازار اجرا می‌شود و شکل ۱۱ نتایج حاصل را با سری زمانی بازسازی شده در مرحله قبل نشان می‌دهد. قابل توجه است که با در نظر گرفتن مقدار ۱۰۴۲ برای طول پنجره و ۶ مولفه در مرحله گروه‌بندی دقت بسیاری در نتایج حاصل می‌شود.



شکل ۱۱- نمودار مقایسه شاخص کل و الگوریتم ژنتیک

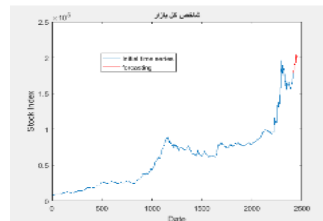
### بررسی شاخص کل بازار با استفاده از تحلیل طیفی تکین هم پوشانی

حال با استفاده از طول پنجره و نقطه‌ی برش به‌دست آمده به وسیله از الگوریتم ژنتیک، الگوریتم تحلیل طیفی تکین هم‌پوشانی بر روی سری‌زمانی حاصل از شاخص کل بازار در بازه زمانی ده ساله، بررسی می‌شود. در ابتدای شروع کار با روش تحلیل طیفی تکین هم‌پوشانی نیاز به مقداردهی برای پارامترهای اندازه تقسیم‌بندی  $Z$  و تعداد نمونه‌های جدا شده است. برای این کار سه مقدار متفاوت برای  $Z$  در نظر گرفته می‌شود و تحلیل طیفی هم‌پوشانی، بر روی این سه مقدار محاسبه می‌شود و میزان اشتراک میان بخش‌ها، ۸۰ درصد لحاظ می‌گردد که به این معنی می‌باشد،  $q$  متناسب با تغییر در اندازه بخش‌ها تغییر می‌کند. در این مرحله ابتدا سری‌زمانی با استفاده از روش تحلیل طیفی تکین هم‌پوشانی محاسبه می‌شود و سپس خطای قدرمطلق میانگین برای مقادیر  $Z$  و  $q$  محاسبه می‌گردد (جدول ۳).

جدول ۳- خطای قدرمطلق میانگین برای سری‌زمانی بازسازی شده

روش تحلیلی مورد استفاده	$Z=200$	$Z=300$	$Z=400$
تحلیل طیفی تکین هم‌پوشانی	۰/۲۴۹۶	۰/۲۷۲۲	۰/۲۹۴۱
تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک	۰/۲۹۹۲		

همان‌طور که در شکل ۱۰ نشان داده شده است، نتایج پیش‌بینی حکایت از صعود بعد از اصلاح شاخص بازار را دارد. حال زمان آن رسیده با استفاده از الگوریتم ژنتیک مناسب‌ترین طول پنجره و گروه‌بندی برای تجزیه طیفی تکین محاسبه شود.



شکل ۱۰- نمودار پیش‌بینی ۴۰ روز کاری شاخص کل بازار

در این جا هدف یافتن کمترین مقدار همبستگی بین سیگنال و نویز است پس تابع هدف مسئله به صورت رابطه زیر (۱۳) است:

$$Cost\ function = \frac{(signal, noise)_w}{\|signal\|_w \|noise\|_w} = \widehat{N}_L^T + \widehat{S}_L^T$$

$$(signal, noise)_w = \sum_{k=1}^N w_k s_k n_k ,$$

$$w_k = \min\{K, L, N - K\}$$

تحلیل شاخص کل بازار با استفاده از الگوریتم ژنتیک جدول ۱، پارامترهای الگوریتم ژنتیک را نشان می‌دهد و پس از اعمال الگوریتم ژنتیک با نرم‌افزار متلب مقادیر بهینه برای پارامترهای مسئله به صورت جدول ۲ می‌باشد.

جدول ۱- پارامترهای الگوریتم ژنتیک

پارامتر	مقدار
تعداد جمعیت نمونه	۱۰۰
تعداد تکرار	۵۱
نرخ تبادل	۰/۰۱
نرخ جهش	۰/۹
روش انتخاب	Roulette wheel

جدول ۲- پارامترهای بهینه با استفاده از الگوریتم ژنتیک

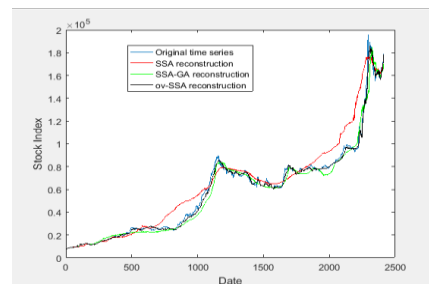
پارامتر	مقدار
مقدار بهینه تابع معیار	۲/۱۴۹
طول پنجره $L$	۱۰۴۲
نقطه‌ی برش $r$	$A = \{1, 2, \dots, 6\}$

با توجه به جدول ۲ طول پنجره بهینه تقریباً در نزدیکی میانه سری زمانی به‌دست آمده و مقادیر ویژه سه‌گانه  $A = \{1, 2, \dots, 6\}$  به‌عنوان مولفه‌های سیگنال و مابقی مولفه‌های نویز در نظر گرفته

### نتیجه‌گیری

در این مقاله، با استفاده از نرم‌افزار متلب، در ابتدا داده‌های ده ساله شاخص کل بورس اوراق بهادار تهران طی سال‌های ۱۳۸۸ تا ۱۳۹۷ روش تجزیه طیفی تکین مبتنی بر الگوریتم ژنتیک در قالب یک مسئله بهینه‌سازی توسط الگوریتم ژنتیک پیاده‌سازی شد. انتخاب تابع معیار یا همان تابع هزینه، گام مهمی در جهت بهبود روند الگوریتم ژنتیک است. برای هر کروموزوم در یک جمعیت مقدار تابع هزینه اندازه‌گیری شده و صحت عملکرد کنترل‌کننده توسط برازندگی این تابع سنجیده شد. سپس جهت مقایسه با روش قبلی، روش تحلیل طیفی تکین هم پوشانی به کار گرفته شد که این روش شامل تقسیم یک سری زمانی به بخش‌های کوچکتر، متوالی و همپوشان است. هم چنین نتایج حاصل از مقدار خطای قدرمطلق میانگین برای این دو روش نشان داد که تحلیل طیفی تکین هم پوشانی با دارا بودن خطای قدرمطلق میانگین کمتر، عملکرد بالاتری نسبت روش تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک دارد و بدین ترتیب فرضیه پژوهش مورد قبول واقع گردید. در واقع این نتایج نشان می‌دهد که روش تحلیل طیفی هم پوشانی در طی مرحله تقسیم‌بندی، انتقال اطلاعات شفاف‌تری را در بین بخش‌ها ایجاد می‌کند که این موجب افزایش دقت در این روش شده است و این روش را در میان روش‌های دیگر تقسیم‌بندی برجسته می‌کند. در واقع با استفاده از این استراتژی، توانایی جداسازی اجزای مربوطه از نویز یا اطلاعات بی‌معنی بهبود می‌یابد و سری‌های زمانی بازسازی شده با رویکرد پیشنهادی قابل اعتمادتر از سری بازسازی شده با روش تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک می‌باشد.

انتخاب طول پنجره  $L$  در روش تحلیل طیفی تکین در ایجاد سرعت یا کندی مرحله تجزیه تأثیر دارد. با افزایش  $L$ ، توانایی سازگاری سری‌زمانی بازسازی شده کاهش می‌یابد. از طرف دیگر، مقادیر کم  $L$  بر قابلیت گرفتن اجزای هارمونیک که دوره آن‌ها بیشتر از مقدار این پارامتر است، تأثیر می‌گذارد. این می‌تواند منجر به یک ایراد در روش پیشنهادی شود زیرا سری‌های زمانی با دوره‌های بسیار بالا، وقتی با بخش‌های متفاوت  $Z$  مقایسه می‌شوند، ممکن است فقط به عنوان روند تفسیر شوند. علاوه بر این، انتخاب پارامتر  $q$  را می‌توان به عنوان یک راه حل برای کاهش در هزینه محاسباتی در نظر گرفت. بنابراین، انتخاب پارامترهای  $Z$  و  $q$  شدیداً در ارتباط با نوع انتخاب پارامترهای روش تحلیل طیفی تکین استاندارد است. شکل ۱۲، سری بازسازی شده تحلیل طیفی استاندارد، تحلیل طیفی مبتنی بر الگوریتم ژنتیک و تحلیل طیفی هم پوشانی را با یکدیگر مقایسه می‌کند.



شکل ۱۲- مقایسه سری بازسازی شده

12. Algburi Riyadh Nazar Ali, Gao Hongli, and Al-Huda Zaid. *Implementation of Singular Spectrum Analysis in Industrial Robot to Detect Weak Position Fluctuations. Fluctuation and Noise Letters*. 2021. 20. 3. pp. 1-22.

13. Alharbi Nader and Hossein Hassani. *A new approach for selecting the number of the eigenvalues in singular spectrum analysis. Journal of the Franklin Institute*. 2016. 353.: pp. 1-16.

14. Alonso F. and D. Salgado. *Analysis of the structure of vibration signals for tool wear detection. Mechanical Systems and Signal Processing*. 2008. 22: pp. 735-748.

15. Artech Josu Josu and Javier García Enríquez. *Singular Spectrum Analysis for signal extraction in Stochastic Volatility models. Econometrics and Statistics*. 2017. 1: pp. 85-98.

16. Bonizzi Pietro, Joel M.H. Karel, Oliver Meste and Ralf L. M. Peeters. *Singular spectrum decomposition: A new method for time series decomposition. Advances in Adaptive Data Analysis*. 2014. 6: pp. 1450011.

17. Danilov Dmitri. and Anatoly Zhigljavsky. *Principal components of time series: the 'Caterpillar' method. St. Petersburg: University of St. Petersburg*. 1997: pp. 1-307.

18. Emmanuel Kimutai K, Wagala Adolphus and Muriithi Dennis K. *Singular Spectrum Analysis: An Application to Kenya's Industrial Inputs Price Index. EJ-MATH, European Journal of Mathematics and Statistics*. 2022. 3. 1. January. pp. 1-13.

19. Fenghua Wen, Xiao Jihong, He Zhifang and Gong Xu. *Stock price prediction based on SSA and SVM. Procedia Computer Science*. 2014. 31: pp. 625-631.

20. Goldberg David E., Bradley Korb and Kalyanmoy Deb. *Messy genetic algorithms: Motivation, analysis, and first results. Complex systems*. 1989. 3: pp. 493-530.

21. Golyandina Nina, Vladimir. Nekrutkin and Anatoly Zhigljavsky. *Analysis of time series structure: SSA and related techniques*. 2001.

22. Golyandina N. and E. Osipov. *The "Caterpillar"-SSA method for analysis of time series with missing values. Journal of Statistical planning and Inference*. 2007. 137: pp. 2642-2653.

23. Golyandina, Nina. and Anatoly Zhigljavsky. *Singular Spectrum Analysis for time series*. 2013.

24. Gujral Harshit, Kushwaha Ajay Kumar and Khurana Sukant. *Utilization of Time Series*

منابع :

۱. اسکندری شیما و علی محرم پور. تفکیک پذیری و تعیین طول پنجره در روش تحلیل طیفی تکین با الگوریتم ژنتیک. کنفرانس بین المللی پنجم مهندسی کامپیوتر و برق. ۲۰۱۸.

۲. حسن دوست زهرا و حمیدرضا وکیلی فرد. پیش بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش تحلیل طیفی تکین و الگوریتم ژنتیک. مهندسی مالی و مدیریت اوراق بهادار. ۱۴۰۰. ۴۹. زمستان. صص ۲۱۱-۱۸۳.

۳. سلیمی نسب سهیل، محمدعلی متذکری و سلمان سلیمی نسب. کاربرد سری فوریه در قیمت گذاری اختیارات معامله در بورس اوراق بهادار. اولین کنفرانس ملی ریاضیات صنعتی. ۱۳۹۳. تبریز. خرداد: صص ۵-۱.

۴. عباسی نژاد حسین، شاپور محمدی و سجاد ابراهیمی. مقایسه مدل های نوسان پذیری چندمتغیره در برآورد رابطه بین نرخ ارز و شاخص سهام. دانش سرمایه گذاری. ۱۳۹۳. سوم. یازدهم. پاییز: صص ۲۲۱-۲۰۱.

۵. محمدی سجاد. طراحی شاخص سهام سازگار با بورس اوراق بهادار تهران. پایان نامه کارشناسی ارشد. دانشکده معارف اسلامی و مدیریت دانشگاه امام صادق (ع). ۱۳۸۶.

۶. مومنی مانی. تحلیل نوسانات بازدهی در بازار سهام تهران. پایان نامه کارشناسی ارشد. دانشکده علوم اقتصادی و اداری دانشگاه مازندران. ۱۳۸۵.

۷. میرزاجانی بجستانی مسعود واحسان. ارمز. تحلیل مجموعه مقادیر تکین: مطالعه موردی بهای سکه. آمار(ندا). ۱۳۹۳. دوازدهم. دوم: صص ۳۹-۳۰.

۸. یارمحمدی مسعود و رحیم محمودوند. پیش بینی نرخ ارز با استفاده از روش تحلیل مجموعه ی مقادیر تکین. مطالعات اقتصادی کاربردی ایران. ۱۳۹۵. سال پنجم. ۱۸. تابستان: صص ۱۴۱-۱۳۳.

9. Abdollahzade Majid, Arash Miranian, Hossein Hassani and Hossein Iranmanesh. *A new hybrid enhanced local linear neuro-fuzzy model based on the optimized singular spectrum analysis and its application for nonlinear and chaotic time series forecasting. Information Sciences*. 2015. 295: pp. 107-125.

10. Abreu R.J., R.M. Souza and J.G. Oliveira. *Applying singular spectrum analysis and ARIMA-GARCH for forecasting*. 2019.

11. Alexandrov T heodore. *A method of trend extraction using singular spectrum analysis. Ar Xiv preprint arXiv: 0804.3367*. 2008.

35. Osanzade, Aytan. and P. Miidla, Singular spectrum analysis forecasting for financial time series. Master Thesis. University of Tartu. 2017.
36. Sanei Saeid and Hossein Hassani. Singular spectrum analysis of biomedical signals. 2015.
37. SengPun, Chi, Shing FungChung and Hoi Ying Wong. Variance swap with mean reversion, multifactor stochastic volatility and jumps. *European Journal of Operational Research*. 2015. 245: pp. 571-580.
38. Sirimal Silva Emmanuel, Hossein Hassani. n the use of singular spectrum analysis for forecasting U.S. trade before, during and after the 2008 recessio. *International Economics*. 2015. 141: pp. 34-49.
39. Sulandari. Winita, S.Subanar Muhammad, Hisyam Leec and Paulo Canas. Time series forecasting using singular spectrum analysis, fuzzy systems and neural networks. *MethodsX*. 2020. 7: pp. 101015.
40. Sulandari Winita, Subanar, Suhartono, Utami Herni, Lee Muhammad Hisyam and Rodrigues Paulo Canas. SSA-based hybrid forecasting models and application. *Bulletin of Electrical Engineering and Informatics*. 2020. 9. 5. October. pp. 2178-2188.
41. Syukur Abdul and Marjuni Airs. Stock Price Forecasting Using Univariate Singular Spectral Analysis through Hadamard Transform. *International Journal of Intelligent Engineering and Systems*. 2020. 13. pp. 96-107.
42. Vautard Robert and Michael Ghil. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*35.1989. 35(3): pp. 395-424.
43. Wilamowski B.M. and X. Li. Fuzzy system based maximum power point tracking for PV system. in *IEEE 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02*. 2002.
44. Zhu J. Pricing volatility of stock returns with volatile and persistent components. *Financ Mark Portf Manag*. 2009. 23: pp. 243-269.
- Tools in Life-sciences and Neuroscience. *Neuroscience Insights*. 2020. 15. pp. 1-18.
25. Hassani Hossein. Singular spectrum analysis: methodology and comparison. 2007. 5: pp. 239-257.
26. Hassani Hossein. Singular spectrum analysis based on the minimum variance estimator. 2010.
27. Hassani Hossein, Rahim Mahmoudvand and Masoud Yarmohammadi. Filtering and denoising in linear regression analysis. *Fluctuation and Noise Letters*. 2010. 9: pp. 343-358.
28. Hassani Hossein. and Dimitrios Thomakos. A review on singular spectrum analysis for economic and financial time series. *Statistics and its Interface*. 2010. 3: pp. 377-397.
29. Hassani Hossein, Z. Ghodsi, E.S. Silva and S. Heravi. From nature to maths: Improving forecasting performance in subspace-based methods using genetics Colonial Theory. *Digital Signal Processing*. 2016. 51: pp. 101-109.
30. Hu Jun and Juho Kannianen. Asymptotic expansion of European options with mean-reverting stochastic volatility dynamics. *Finance Research Letters*. 2015. 14: pp. 1-10.
31. Kim Jung mu, Yuen Jung Park and Doojin Ryu. Stochastic volatility of the futures prices of emission allowances: A Bayesian approach. *Physica A*. 2017. 465: pp. 714-724.
32. Lahmiri Salim. Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression. *Applied Mathematics and Computation*, 2018. 320: pp. 444-451.
33. Leles MCR. Sansao JP. Mozelli LA and Guimaraes HN. Improving reconstruction of time-series with singular spectrum analysis: A segmentation approach. *Digital Signal Processing*. 2017.77. june. pp. 63-76.
34. Leles MCR. Sansao JP. Mozelli LA and Guimaraes HN. A new algorithm in singular spectrum analysis framework: The Overlap-SSA (ov-SSA). *SoftwareX*. 2018. 8. July–December. pp. 26-32.

*Journal of Financial Knowledge of Securities Analysis*  
*Vol. 17 / No. 61, Spring 2024*

**Predicting the Overall Index of Tehran Stock Exchange  
Using Singular spectrum analysis Based on Genetic  
Algorithm and Overlap Singular spectrum analysis**

Zahra Hasandoost<sup>1</sup>  
Hamidreza Vakiliifar <sup>\*2</sup>  
Fraydoon Rahnamay Roodposhti<sup>3</sup>

**Abstract**

*The present study analyzes the prediction of the total index of the Tehran Stock Exchange with the singular spectral analysis method based on the genetic algorithm and overlapping singular spectral analysis. It is practical in terms of purpose and descriptive-analytical in terms of method. Its statistical population is the daily price of the total index of the Tehran Stock Exchange in the ten-year return (2009 to 2018) and the research sample is 2411 data from the logarithmic return of the target index. First, the genetic algorithm was implemented in the SSA method on the index. Then, using the Ov-SSA method in order to improve the reconstruction and resolution of the components, the initial and large time series were divided into small and common consecutive parts and the standard SSA analysis method was used for each part. The results of the research showed that the Ov-SSA analysis method has a higher performance than the GA-SSA analysis method with a lower mean absolute value error.*

**Key words:** Window Length, Signal, Break Point, Noise, Overlap Singular spectrum analysis.

1. Department of Financial Management, Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran. Zahra.Hasandoost.62@gmail.com

2. Department of Financial Management, Management and Economics, Science and Research, Branch, Islamic Azad University, Tehran, Iran. Vakilifard.phd@gmail.com

3. Department of Financial Management, Management and Economics, Science and Research ,Branch, Islamic Azad University, Tehran, Iran. Rahnama.Roodposhti@gmail.com

یادداشت‌ها:

---

1. *PCA: Principal Component Analysis*
2. *SV: Stochastic Volatility*
3. *GARCH: Generalized AutoRegressive Conditional Heteroskedasticity*
4. *ARCH: AutoRegressive Conditional Heteroscedasticity*
5. *M-SSA : multichannel SSA*
6. *EMD: Empirical Mode Decomposition*
7. *Decomposition*
8. *Embedding*
20. *The Trajectory Matrix*
10. *Length Window*
11. *Hankel Matrix*
12. *SVD: Singular Value Decomposition*
13. *Eigenvalues*
11. *Elementary Matrices*
15. *Principal components*
16. *Eigen Triple*
17. *The spectrum*
18. *Reconstruction*
19. *Grouping*
17. *Eigen triple Grouping*
21. *Diagonal Averaging*
22. *Hankelization*
23. *LRF: Linear Recurrent Formulate*
24. *Reproduction*
25. *Generation*
26. *Selection*
27. *Crossover*
28. *Mutation*
29. *Roulette Wheel*
30. *Slice*
31. *The overlap-save method*
32. *MAE: Mean Absolute Error*