



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

زهرا حسن دوست^۱

تاریخ دریافت مقاله: ۱۴۰۰/۰۱/۱۷ تاریخ پذیرش مقاله: ۱۴۰۰/۰۳/۱۱

حمیدرضا وکیلی فرد^۲

چکیده

نوسانات در بازارهای مالی با سیگنال و نویز همراه می‌باشد. در این مقاله علاوه بر تجزیه و تحلیل طیفی تکین، برای پیدا کردن طول پنجره و نقطه برش بهینه از الگوریتم ژنتیک استفاده شده است که تابع هدف آن، یافتن حداقل مقدار برای تابع همبستگی میان مؤلفه‌های سیگنال و نویز می‌باشد. بدین خاطر ابتدا داده‌های ده‌ساله شاخص کل بورس اوراق بهادار تهران طی سال‌های ۱۳۸۸ تا ۱۳۹۷ با استفاده از روش تجزیه طیفی تکین در سه پیاده‌سازی شد. سپس در قالب یک مسئله بهینه‌سازی توسط الگوریتم ژنتیک حل شد. نتایج حاصل از فرضیه اول نشان داد که تفکیک‌پذیری سیگنال و نویز در روش تحلیل طیفی تکین امکان‌پذیر می‌باشد. همچنین با توجه به نتایج حاصل در تحقیق، تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک با داشتن خطای قدر مطلق میانگین کمتر، بهبود در دقت پیش‌بینی را نشان داد. در نهایت نیز با توجه به یافتن کمترین همبستگی وزنی بین مؤلفه‌های سری زمانی جهت تفکیک سیگنال و نویز (یافتن نقطه‌ی برش) و سپس با دست آوردن طول پنجره‌ی بهینه در تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک، گویای این واقعیت است که تغییر در مقدار پارامترها می‌تواند در بهبود عملکرد روش تحلیل طیفی مفید واقع شود.

کلمات کلیدی

شاخص کل بازار، طول پنجره، نقطه‌ی برش، سیگنال، نویز.

۱- گروه مدیریت مالی، 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

نا اطمینانی، شرایطی است که در آن پیشامدهای ممکن که در آینده اتفاق می‌افتند، مشخص و معلوم نیستند و یا این‌که، اگر این پیشامدها مشخص و معلوم باشند، احتمال‌های مربوط به وقوع این پیشامدها در دسترس می‌باشند و وقتی که هر کدام یا هر دوی این موارد پیش می‌آید، تصمیم‌گیری نسبت به آینده، پیچیده و مشکل می‌شود و از این رو، فضای نااطمینانی بر تصمیم‌ها حاکم می‌شود [۵]. نوسانات در بازارهای مالی از اهمیت ویژه‌ای برخوردار است. به طوری که بسیاری از مدل‌های تخصیص پورتفو، 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 بسیاری از کاربردهای مالی شامل؛ ساختار پرتفوی، مدیریت و ریسک می‌باشد. اگرچه نوسان‌پذیری به خودی خود قابل مشاهده نیست اما با به کارگیری آن می‌توان دلیل منطقی در دارایی‌های مشاهده نمود [۳۹]. نوسانات بازار سهام در تصمیم‌گیری سرمایه‌گذاران بازار و مدیران شرکت‌ها نقش ویژه‌ای دارد [۷]. 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 [۶]. نوسان‌پذیری بازار دارایی نیز نقش مهمی را در سیاست‌های پولی بازی می‌کند. عواقب بحران مالی اخیر بر اقتصاد جهانی نشان داد که چقدر نوسان‌پذیری بازارهای مالی در اجرای سیاست پولی مؤثر، مهم است. در بازارهای مالی، نوسان‌پذیری قیمت معیاری، برای سنجش نوسانات قیمت یک ابزار مالی در طول زمان است. نوسان‌پذیری را نمی‌توان به طور مستقیم مشاهده کرد، بلکه آن را باید با استفاده از معیارهای مناسب یا به عنوان جزئی از یک مدل قیمت‌گذاری دارایی تصادفی^۱ تخمین زد [۲]. مدیریت نوسان ریسک از زمان اولین بحران جهانی اخیر مالی به شدت در بازارهای مالی افزایش پیدا نموده است [۳۴]. 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 $GARCH^2$ که با بازگشت به میانگین (بازده) مشتق می‌شوند

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

مدل‌های SV دو حرکت مشتق از بازده و نوسان را ارزیابی می‌نمایند [۳۰]. 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^۴ ب- مدل‌های نوسان تصادفی (SV) که نوسان پذیری یک فرآیند نوسان نهفته برگرفته از حرکت جدید است که در فرآیند نوسان پذیری نهادینه شده‌اند. سری‌های نوسان پذیری 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]. نوسان پذیری در سری‌های زمانی مالی دو ویژگی اصلی دارند اول این‌که نوسان پذیری در طول زمان متغیر هستند و هم‌چنین در نوسان‌پذیری متغیرهای مالی پدیده خوشه‌ای وجود دارد. مدل‌های نوسان پذیری تصادفی برای در نظر گرفتن همین ویژگی در مالی مورد استفاده قرار گرفت [۴]. در دهه اخیر، پیشرفت‌هایی در قیمت‌گذاری اختیار معامله، اهمیت نوسان پذیری تصادفی را نشان داده است. مخصوصاً مدل‌های نوسان تصادفی زمان پیوسته مانند مدل هستون و بسط‌های آن به علت تحلیل قابل تیمارشان (طبق نظر کاربر) بسیار عمومیت یافته است [۲۹]. 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 نوسان پذیری است [۴]. علی‌رغم آن‌که در زمینه‌ی تحلیل سری‌های زمانی، تلاش‌های زیادی برای معرفی روش‌ها و مدل‌های توانمند شده است، اکثر این روش‌ها 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 [۲۴]. حسنی و توماکس در مقاله‌ی خود در سال ۲۰۱۰ فرض نمودند که به‌طور کلی، دو روش اصلی در مورد پیش‌بینی سری‌های زمانی با نویز وجود دارد. روش اول نویز را نادیده می‌گیرد و سپس مدلی را

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و نه / زمستان ۱۴۰۰

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 [۲۷].

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. به عنوان مثال، حسنی و ژینگلیاوسکی در سال ۲۰۰۹ نشان دادند که استفاده از تحلیل طیفی تکین برای پیش‌بینی سری اقتصاد کلان کوچک و بزرگ بسیار سودمند است. اگرچه تحلیل طیفی تکین به ابزاری پر کاربرد در تجزیه و تحلیل و پیش‌بینی آب و هوا (ووتارد و گیل، ۱۹۸۹) [۳۷]، پزشکی (سنایی و حسنی، ۲۰۱۵) [۳۳]، پردازش تصویر دیجیتال و سایر زمینه‌های دانش مانند علوم اجتماعی و فیزیک (حسنی، ۲۰۱۰) [۲۶]، تبدیل شده است، با این 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. تحلیل طیفی تکین می‌تواند با شیوه‌های مختلف ادغام شود و در نتیجه تبدیل به روش‌هایی مانند تحلیل طیفی تکین تک، دوکاناله و حتی سه کاناله شود. ادغامی که اغلب در مطالعات و نشریات مربوط به تحلیل طیفی تکین به‌وفور یافت می‌شود، تحلیل طیفی تکین چندکاناله است. این ادغام در مواردی اعمال می‌شود که سری زمانی با چندین متغیر مرتبط با یکدیگر تشکیل شده باشد [۲۱]. سوالات اساسی این تحقیق، آیا روش تحلیل طیفی تکین روشی است که بتوان با آن تفکیک‌پذیری سیگنال و نویز را در شاخص کل بورس اوراق بهادار تهران مشخص نمود؟ آیا می‌توان از الگوریتم ژنتیک در بهبود دقت پیش‌بینی روش تحلیل طیفی تکین استفاده کرد؟

فرضیه‌های پژوهش

فرضیه اول: تفکیک‌پذیری سیگنال و نویز در روش تحلیل طیفی تکین امکان‌پذیر است.

فرضیه دوم: الگوریتم ژنتیک باعث بهبود دقت پیش‌بینی روش تحلیل طیفی تکین می‌شود.

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

پیشینه تجربی پژوهش

میرزاجانی و همکاران (۱۳۹۲) به بررسی قیمت سکه در بازار شهر مشهد در بازه فروردین ۱۳۹۱ تا آبان ۱۳۹۲ با استفاده از روش تحلیل طیفی تکین پرداختند. و در بررسی های خود نشان دادند که این روش در استخراج هم زمان هم سازها و مؤلفه‌های روند به خوبی عمل می‌کند [۸]. اسکندری و محرم پور ۱۳۹۷ بررسی‌های خود را بر روی پارامترهای روش تحلیل طیفی تکین متمرکز نمودند و در جستجوی بهترین مقدار برای دو پارامتر این روش که همان طول پنجره و نقطه‌ی برش است، از الگوریتم ژنتیک کمک گرفتند. هدف اصلی آنان استخراج بهترین مؤلفه‌های سیگنال از سیگنال اصلی بوده است [۱]. یارمحمدی و محمودوند در سال ۱۳۹۵ روش تحلیل طیفی تکین را برای مدل‌سازی و پیش‌بینی نرخ روزانه‌ی دلار به ریال در بازه زمانی تیرماه ۱۳۹۲ تا شهریور ۱۳۹۴ مورد استفاده قرار دادند و برای ارزیابی عملکرد آن، مدل آرما را به کار برده و با استفاده از مثال‌های تجربی به این نتیجه رسیدند که روش تحلیل طیفی تکین می‌تواند به عنوان یک روش قدرتمند در پیش‌بینی نرخ ارز مورد استفاده قرار گیرد [۹]. فنگویا و دیگران در تحقیق خود (۲۰۱۴) با استفاده از روش تحلیل طیفی تکین، قیمت سهام بازار شانگهای چین در بازه زمانی ۲۰۰۹ تا سال ۲۰۱۳، از نظر روند، نوسان بازار و نویز با ویژگی‌های اقتصادی مختلف در افق‌های زمانی مختلف تجزیه کرده و سپس این ویژگی‌ها را در دستگاه بردار پشتیبانی برای پیش‌بینی قیمت معرفی نمودند و به نتایج بهتری در قیاس با روش تحلیل طیفی تکین رسیدند [۱۸]. سیریمال و حسنی (۲۰۱۵) در مقاله خود ضمن معرفی قدرت تکنیک پیش‌بینی و تحلیل سری‌های زمانی نا پارامتریک تحلیل طیفی تکین، جهت پیش‌بینی معامله با کاربردش که ارزیابی اثرش بر بحران مالی ۲۰۰۸ بر ایالات متحده است را بررسی نمودند و نتیجه گرفتند که این روش می‌تواند پیش‌بینی دقیقی از معامله در مواقع بحران ایجاد نماید [۳۵]. عثمان زاده و همکاران (۲۰۱۷) از روش تحلیل طیفی تکین برای بررسی بر روی سهام تالینک گروپ^۵ استفاده کردند. آنها از مقادیر متفاوت برای طول پنجره و تعداد مؤلفه‌ها بهره بردند و بهینه‌ترین مقدار را در نظر گرفتند و در آخر نشان دادند که برای یک مقدار پیش‌بینی با دقت بالا با استفاده از روش تحلیل طیفی تکین نیاز به مطالعه و بررسی بیشتر در مرحله‌ی بازسازی ماتریس‌های مسیراست [۳۲]. آرتچی و گارسیا انریکز (۲۰۱۷) در پژوهش خود نوسان روزانه شاخص سهام صنعتی داو جونز و شاخص سهام اسپانیا را با تحلیل طیفی تکین درون نمونه‌ای و سپس کولموگروف وینر تحلیل نمودند و به این نتیجه رسیدند که روش تحلیل طیفی تکین عملکرد بالاتری دارد و ضمناً از تحلیل طیفی تکین خارج نمونه‌ای نیز برای نشان دادن پایداری و ثبات ارزش در معرض استفاده نمودند [۱۵]. له‌میری (۲۰۱۸) مدلی را برای پیش‌بینی قیمت سهام داخل روز ارائه می‌دهند که از تجزیه

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و نه / زمستان ۱۴۰۰

و تحلیل طیفی تکین و رگرسیون بردار پشتیبان^۶ همراه با بهینه‌سازی ازدحام ذرات^۷ استفاده می‌کنند. با استفاده از روش تحلیل طیفی تکین، سری‌های زمانی قیمت سهام را به تعداد کمی از اجزای مستقل که به‌عنوان پیش‌بینی کننده استفاده می‌شوند، تجزیه می‌کند و سپس رگرسیون بردار پشتیبان را برای مرحله‌ی پیش‌بینی اعمال می‌شود و در آخر از روش بهینه‌سازی ازدحام ذرات برای بهینه‌سازی پارامترهای رگرسیون بردار پشتیبان استفاده می‌شود. نتایج پیش‌بینی حاصل از شبیه‌سازی‌ها نشان می‌دهد که روش ارائه‌شده ی، تجزیه و تحلیل طیفی تکین و رگرسیون بردار پشتیبان همراه با بهینه‌سازی ازدحام ذرات (SSA-PSO-SVR) تا حد زیادی بهتر از روش تحلیل طیفی تکین استاندارد است [۳۱]. سولاندی و همکاران (۲۰۲۰) به بررسی پیش‌بینی سری‌های زمانی پرداختند و روش تحلیل طیفی تکین را با شبکه‌های عصبی و سیستم‌های فازی مقایسه کردند و سپس با ترکیب این سه روش به نتایج دقیق‌تری در مقایسه با نتایج به‌دست‌آمده توسط هر روش به‌طور جداگانه، دست یافتند [۳۶]. در ادامه مقاله، روش تحلیل طیفی تکین و چگونگی پیش‌بینی سری‌های زمانی با این روش و به دنبال آن، بهبود در روند روش تحلیل طیفی تکین با الگوریتم ژنتیک معرفی می‌شود و در پایان نیز، نتایج تحلیل، بر رویدادهای حاصل از شاخص کل بورس اوراق بهادار تهران نشان داده شده است.

روش‌شناسی پژوهش

پژوهش حاضر از پژوهش‌های توصیفی و به لحاظ هدف کاربردی است. داده‌های این پژوهش، قیمت روزانه شاخص کل بورس اوراق بهادار تهران می‌باشد که از سال ۱۳۸۸ تا ۱۳۹۷ در یک دوره زمانی ده ساله از سایت 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.

معرفی روش تحلیل طیفی تکین

اساسی‌ترین هدف روش تحلیل طیفی تکین تبدیل یک سری به یک مجموع اجزای قابل تفسیر است، به‌طوری‌که هر یک از مؤلفه‌های آن می‌تواند به‌عنوان یک روند، اجزای نوسانی یا نویز شناخته شوند [۲۴]. 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 زمانی جدید از آن‌ها ساخته می‌شود [۲۵]. در ادامه، مراحل الگوریتم تحلیل طیفی به‌طور کامل شرح داده می‌شود:

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

مرحله اول: تجزیه^۸:

گام اول: نشانیدن^۹:

هدف از این مرحله، محاسبه‌ی ماتریس مسیر^{۱۰} 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 2 است، باشد. فرض کنید L به طوری که $1 < L < N$ ، یک عدد صحیح باشد که طول پنجره^{۱۱} نامیده می‌شود و $K=N-L+1$ را در نظر بگیرید. برای انجام عمل نشانیدن، سری اولیه به مجموعه‌ای از بردارهایی به اندازه L تبدیل می‌شود، به طوری که، $X_i = (x_i, \dots, x_{i+L-1})^T (1 \leq i \leq K)$ ، در واقع مرحله‌ی نشانیدن را می‌توان همانند نگاهی تصور کرد که سری زمانی یک‌بعدی را به سری زمانی چند بعدی انتقال 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 هنکل^{۱۲} است [۲۳].

گام دوم: تجزیه مقدار تکین^{۱۳}

در این گام، ماتریس مسیر به دست آمده توسط گام اول، بر اساس مقادیر ویژه^{۱۴}، تجزیه و به ماتریس‌های پایه دوبه‌دو متعامد با رتبه‌ی یک تبدیل می‌شود. ابتدا مقادیر ویژه XX^T محاسبه و سپس به ترتیب نزولی $0 \leq \alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_L \geq 0$ مرتب می‌شود. مجموعه‌ی بردارهای متعامد یکه U_1, \dots, U_L که ضرب دوبه‌دو آن‌ها برابر با صفر می‌باشد را به عنوان بردارهای ویژه XX^T متناظر با این مقادیر ویژه در نظر گرفته می‌شود، تعریف می‌شود:

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

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 $U_i V_i^T X_i = \sqrt{\alpha_i}$ و $i=1, \dots, d$ می‌باشد. ماتریس‌های X_i دارای رتبه‌ی یک هستند، بنابراین این ماتریس‌ها، ماتریس‌های پایه^{۱۵} هستند. U_i و V_i ها به ترتیب بردارهای ویژه چپ (مؤلفه‌های

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و نه / زمستان ۱۴۰۰

اصلی^{۱۶}) و راست ماتریس مسیر و مجموعه‌ی $(i, \sqrt{\alpha_i}, U_i, V_i^T)$ امین سه‌تایی ویژه^{۱۷} ماتریس X نامیده می‌شود و هم‌چنین مقادیر $\sqrt{\alpha_i}$ ، مقادیر تکین ماتریس X و مجموعه‌ی تمامی آنها را طیف^{۱۸} ماتریس X می‌نامند [۲۳].

مرحله دوم: بازسازی^{۱۹}

گام اول: گروه‌بندی^{۲۰}

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 $X_{I_i} = \sqrt{\alpha_i} V_i^T X_i$ است. فرض کنید $I_i = \{i_1, \dots, i_r\}$ مجموعه‌ی نشانگرهای i_1, \dots, i_r باشد، بسط ماتریس X_{I_i} در گروه I_i به شکل $X_{I_i} = X_{I_{i_1}} + X_{I_{i_2}} + \dots + X_{I_{i_r}}$ تعریف خواهد شد. در 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^{۲۱} 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_j} 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 [۲۴]. برای مثال، اگر $p=2$ باشد، آن‌گاه گروه اول همان مؤلفه‌های اصلی (سیگنال) و گروه باقی‌مانده، نویز در نظر گرفته می‌شود. در این حالت برای مثال r تا از بزرگ‌ترین مقادیرهای تکین و بردارهای ویژه متناظر، برای تقریب سری اصلی انتخاب‌شده و بقیه مقادیرهای تکین به عنوان مؤلفه نویز لحاظ می‌شود [۲۳].

گام دوم: میانگین‌گیری قطری^{۲۲} (هنکل‌سازی^{۲۳})

هر ماتریس هنکل یک رابطه‌ی یک‌به‌یکی 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

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

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. این روش بدین صورت می‌باشد که اگر Z_{ij} را به عنوان یک مؤلفه از ماتریس Z در نظر گرفته شود، آنگاه k امین جمله سری، از میانگین‌گیری Z_{ij} بر روی تمامی i و j هایی که در شرط $i+j=k+1$ صدق می‌کنند، به دست آمده می‌آید. این روش را میانگین‌گیری قطری یا هنکل‌سازی ماتریس Z می‌نامند. نتیجه‌ی هنکل‌سازی یک ماتریس Z ، ماتریس هنکل HZ خواهد بود، که ماتریس مسیر متناظر با سری‌های به دست آمده در نتیجه‌ی میانگین‌گیری قطری است. توجه به این نکته ضروری هست که فرایند هنکل‌سازی از این جهت بهینه است که ماتریس HZ از لحاظ فرم ماتریس در بین تمامی ماتریس‌های هنکل دارای اندازه متناظر، بیشترین نزدیکی به Z را دارد. از آنجایی که ماتریس هنکل HZ به طور یکتا توصیف کننده یک سری زمانی است، مقادیر موجود در قطرها را می‌توان با مقادیر سری زمانی مرتبط دانست. با اعمال روش هنکل‌سازی بر روی همهی مؤلفه‌های ماتریس رابطه ۱، رابطه‌ی دیگری همانند رابطه زیر (۴) به دست آمده 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

$$x_n = \sum_{k=1}^p \hat{x}_n^{(k)} \quad \text{سری در معادله زیر (رابطه ۵) است:}$$

که در آن، $\hat{X}_N^{(k)} = (\hat{x}_1^{(k)}, \hat{x}_2^{(k)}, \dots, \hat{x}_p^{(k)})$ 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 بازسازی شده از مرحله‌ی میانگین‌گیری قطری، یک سری با سطح نویز اندک هست و لذا می‌توان از آن برای مدل‌سازی یا پیش‌بینی، که در بخش بعد به آن اشاره می‌گردد، استفاده کرد. بیان ریاضی مطالب ارائه شده، در ادامه شرح داده می‌شود: [۲۳]. فرض کنید که $Y_{K \times L}$ یک ماتریس با درایه‌های y_{ij} که در آن $1 \leq i \leq L$ و $k \leq j \leq 1$ باشد و در شرایط زیر (رابطه ۶) صدق کند:

$$K^* = \max(L, K), L^* = \min(L, K), \quad \begin{cases} y_{ij}^* = y_{ij} & \text{for } L < K \\ y_{ij}^* = y_{ji} & \text{otherwise} \end{cases}$$

روش میانگین‌گیری ماتریس $Y_{K \times L}$ با استفاده از روش زیر به سری‌های y_1, y_2, \dots, y_N تبدیل می‌شود.

$$y_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k y_{m, k-m+1}^* & \text{for } 1 \leq k < L^* \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+1}^* & \text{for } L^* \leq k \leq K^* \\ \frac{1}{N-K+1} \sum_{m=k-K^*+1}^{N-K^*+1} y_{m, k-m+1}^* & \text{for } K^* < k \leq N \end{cases}$$

پیش‌بینی با روش تحلیل طیفی تکین

حال زمان آن رسیده است که روش تحلیل طیفی تکین برای پیش‌بینی داده‌ها مورد استفاده قرار گیرد. در روش تحلیل طیفی تکین از روابط بازگشتی خطی^{۳۴} 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 ایجاد شده در شرط بالا (رابطه ۷) صدق می‌کند، توسط یک رابطه بازگشتی خطی می‌تواند ادامه یابد. حال فرض نماییم سری زمانی‌ای 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 بازسازی شده هستند و y_{N+1}, \dots, y_{N+h} سری‌های جدیدی می‌باشند که توسط روش تحلیل طیفی تکین پیش‌بینی شده است [۱۸]. 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:

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

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

انتخاب پارامترهای روش تحلیل طیفی تکین

انتخاب پارامترها بستگی به داده‌هایی دارد که در اختیار است و آنالیزی که باید بر روی آن‌ها صورت گیرد [۲۱]. همان‌طور که بیان شد در روش تحلیل طیفی تکین دو پارامتر وجود دارد: یکی طول پنجره L که در مرحله‌ی تجزیه مورد استفاده قرار می‌گیرد و پارامتر دوم r تعداد مؤلفه‌ها 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 [۲۴]. چهار روش مجزا را می‌توان برای بررسی گزینه‌های

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

مناسب برای مقادیر طول پنجره و نقطه‌ی برش^{۲۵} در نظر گرفت؛ اول) تجزیه و تحلیل طیف مقادیر تکین، دوم) بررسی پراکندگی دوبه‌دو بردارهای ویژه، سوم) تجزیه و تحلیل دوره نگار سری اصلی (بردارهای ویژه) و چهارم) تجزیه و تحلیل ماتریس همبستگی وزنی، که یک معیار معمول قابل تفکیک است

- 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 [۳۳]. در ادامه به بررسی دقیق‌تر پارامترهای موردنیاز، پرداخته می‌شود.
- 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)، هرکدام به‌طور جداگانه در مقالات خود ثابت کرده‌اند که بهترین انتخاب برای طول پنجره L با توجه به تعداد داده‌های سری زمانی، می‌تواند برابر با میانه‌ی 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، که این می‌تواند کمک شایانی را در تجزیه‌پذیری بهتر ماتریس مسیر نماید. به عنوان مثال، اگر سری زمانی یک تناوب سالانه داشته باشد، می‌توان از ضرب‌های ۱۲ استفاده نمود [۲۷]. به‌طور کلی، مقادیر زیاد L ، تجزیه با انعطاف بیشتری را به اجزای اولیه ارائه می‌دهد و در نتیجه، قابلیت تفکیک بهتر را ایجاد می‌کند. اگر هیچ اطلاعات قبلی در نظر گرفته نشده باشد، دستورالعمل‌های کلی در ادبیات روش تحلیل طیفی تکین در مورد انتخاب L ارائه شده است [۲۶]. میانه سری زمانی اصلی را به‌عنوان گزینه‌ای مناسب برای طول پنجره پیشنهاد داده‌اند. گلوپاندینیا و همکاران (۲۰۰۱) توصیه نمودند که $L \sim \frac{N}{2}$ انتخاب شود [۲۱]. حسنی (۲۰۰۷) با تمرکز بر قابلیت تفکیک و تفسیر اجزای

به دست آمده، $L = 24$ را نشان می‌دهد. با این حال، هنگامی که هدف استخراج سیگنال و پیش‌بینی باشد، $L = 14$ را پیشنهاد کردند [۲۸]. سرانجام، همان‌طور که توسط گلوباندینیا و ژیکلیاوسکی (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 ابرئو و همکاران نیز در سال ۲۰۱۹، طی مقاله‌ای به بررسی هرچه بیشتر طول پنجره پرداخته‌اند و با این هدف که بهترین طول پنجره را به دست آورند. در این مقاله چندین تجزیه از طریق آزمون و خطا انجام شده است [۱۱]. 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 [۲۴].

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 نویز تشکیل می‌شود. از آنجایی که مؤلفه‌های سیگنال غالباً غالب هستند، تنها پارامتر گروه‌بندی تعداد مؤلفه‌های اصلی، یعنی اندازه‌ی گروه‌بندی است. از طرف دیگر، حسنی و همکاران (۲۰۱۶) نشان دادند که تفکیک باینری ممکن است به نتایج مطلوب منجر نشود. آن‌ها رویکردی را در نظر می‌گیرند که تمام مقادیر ویژه را تجزیه و تحلیل می‌کند و آن‌هایی را که اطلاعات بیشتری دارند در نظر می‌گیرد [۲۸]. الکساندروف (۲۰۰۹) روشی را برای انتخاب گروه‌بندی خودکار بر اساس دوره نگار بردارهای ویژه برای استخراج روند و اجزای هارمونیک پیشنهاد داد [۱۲]. آلسو و سالگادو (۲۰۰۸) از تکنیک‌های خوشه‌ای برای گروه‌بندی اجزای اولیه بر اساس روش خوشه‌بندی کی-میانگین استفاده کردند [۱۴].

انتخاب خودکار: روش‌هایی هم برای انتخاب خودکار پارامترهای تحلیل طیفی تکین در سال‌های اخیر ارائه شده است. بونیزی و همکاران (۲۰۱۴) تجزیه طیفی تکین را پیشنهاد کردند. روشی کاملاً خودکار برای انتخاب پارامترها با تمرکز بر طیف فرکانس [۱۶]. عبدالله زاده و همکاران (۲۰۱۵)، برای تنظیم دقیق پارامترهای روش تحلیل طیفی تکین از الگوریتم بهینه‌ساز یا ازدحام ذرات^{۲۶} استفاده کردند [۱۰]. حسنی و همکاران (۲۰۱۶)، هم‌چنین از نظریه‌ی استعمار^{۲۷} به‌عنوان یک روش بهینه‌سازی برای انتخاب پارامترهای

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

تحلیل طیفی تکین استفاده کردند. برای هر L ، الگوریتم مقادیر ویژه R بهینه را تابع هزینه داده‌شده توسط خطای مربع میانگین ریشه بازسازی جستجو می‌کند. در واقع در این روش بهینه‌سازی هر L بررسی می‌شود [۲۸]. در جدول ۱، روش‌های مطرح‌شده فوق‌الذکر را، به‌طور خلاصه بیان شده است:

جدول ۱- خلاصه روش‌های انتخاب پارامتر روش تحلیل طیفی تکین

طول پنجره	گروه‌بندی خودکار
اطلاعات قبلی ($L=kT$) مقادیر ویژه طیفی	تجزیه طیفی تکین
تفکیک‌پذیری ($L \sim \frac{N}{2}$) دوره نگار بردارهای ویژه	بهینه‌سازی ازدحام ذرات
استخراج سیگنال $w \ll \frac{N}{2}$ - L همبستگی	نظریه استعمار

تفکیک‌پذیری سیگنال و نویز

همان‌طور که تا به این مرحله بررسی شد، 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 نویز را نمی‌توان به‌خوبی از سیگنال اصلی استخراج کرد. حسنی و محمودوند (۲۰۱۰)، از مفهوم تفکیک‌پذیری بین مؤلفه‌ی سیگنال و نویز استفاده کردند و با بررسی‌های تحلیلی بر روی سری‌های زمانی نشان دادند که مناسب‌ترین مقدار برای طول پنجره، میانه سری زمانی می‌باشد. آن‌ها در این پروژه سری زمانی Y_T را به صورت رابطه $Y_T = S_T + e_T$ که در آن T تعداد داده‌های سری زمانی، S_T مؤلفه‌های سیگنال و e_T مؤلفه‌های نویز است، لحاظ نمودند. آن‌ها هم‌چنین نشان دادند که هدف از مرحله بازسازی در تحلیل طیفی تکین جستجوی یک تخمین برای مؤلفه‌های سیگنال S_T یعنی \hat{S}_T است. در شرایط ایده آل می‌توان مؤلفه‌های نویز را حذف کرد و عمل بازسازی را به‌راحتی انجام داد ولی در شرایط واقعی بازسازی مؤلفه‌های سیگنال به‌طور کامل غیرممکن است، اما با این حال می‌توان نزدیک‌ترین \hat{S}_T را محاسبه کرد [۲۶]. گلوپاندینا و همکاران (۲۰۰۷)، برای رفع این مشکل از معیاری بانام 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

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

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

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 تفکیک‌پذیری 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. حال برای درک بهتر مطلب فرض نمایید که فقط دو مؤلفه سیگنال و نویز وجود دارد. در نتیجه، مقدار همبستگی‌شان می‌دهد که چگونه مؤلفه سیگنال بازسازی شده با استفاده از روش تحلیل طیفی تکین از مؤلفه نویز جدا می‌شود [۲۱].

قضیه: فرض کنید \widehat{S}_L^T سری‌های بازسازی شده r مقدار ویژه اول ماتریس مسیر X باشد، بنابراین:

$$\widehat{S}_L^T = \widehat{S}_K^T, \widehat{N}_L^T = \widehat{N}_K^T \text{ که در آن } \widehat{N}_K^T = Y_T - \widehat{S}_K^T.$$

نتیجه‌گیری: فرض کنید $\rho_{L,r}^{(w)}$ همبستگی بین دو سری زمانی \widehat{S}_L^T و \widehat{N}_L^T باشد، آنگاه $\rho_{L,r}^{(w)}$ است که از این عبارت می‌توان چنین نتیجه‌گیری داشت که فقط نیاز است که طول پنجره‌های در فاصله $L \in \left\{ 2, \dots, \left\lfloor \frac{T+1}{2} \right\rfloor \right\}$ در نظر گرفته شود. این قضیه به ظاهر ساده کمک بسیاری در کاهش تعداد بررسی‌ها در جستجوی مقدار بهینه برای طول پنجره می‌کند [۲۶]. با استفاده از قضیه بالا می‌توان رابطه

همبستگی وزنی برای سیگنال و نویز را به صورت زیر بیان (رابطه ۱۱) کرد: [۲۱]

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

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

در این پژوهش از این معیار به عنوان تابع هزینه برای محاسبه نقطه‌ی برش و طول پنجره‌ی بهینه با استفاده از الگوریتم ژنتیک استفاده شده است. انتخاب تابع معیار یا همان تابع هزینه مناسب، کمک شایانی در جهت بهبود روند الگوریتم ژنتیک می‌کند. در ادامه به بررسی الگوریتم ژنتیک پرداخته می‌شود. الگوریتم‌های ژنتیک قسمتی از محاسبات تکاملی هستند که به عنوان زیرمجموعه‌ای در زمینه‌ی هوش مصنوعی به سرعت در حال رشد می‌باشند. همان‌طور که از نام آن نیز می‌توان حدس زد الگوریتم ژنتیک، الهام گرفته از طبیعت است که در قالب الگوریتمی با روند جستجوی تصادفی، این جریان تکاملی طبیعت را تقلید و شبیه‌سازی می‌کند. هنگامی که در جریان حل مسئله مفروض از یک یا مجموعه‌ای از جواب‌ها، به جواب یا مجموعه‌ای از جواب‌های بهتر حرکت کند، در واقع روند تکاملی جواب‌ها را، نیل به پاسخ بهینه طی کرده است. جذابیت زیاد الگوریتم‌های ژنتیک این است که نتایج نهایی قابل ملاحظه‌تری دارند. الگوریتم ژنتیک با یک مجموعه از جواب‌ها که از طریق کروموزوم‌ها نشان داده می‌شوند، شروع

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

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 [۳۸]. در الگوریتم‌های ژنتیکی، در طی مرحله‌ی تولید مثل^{۳۸} از عملگرهای ژنتیکی استفاده می‌شود. با تأثیر این عملگرها بر روی یک جمعیت، نسل^{۳۹} بعدی آن جمعیت تولید می‌شود. عملگرهای انتخاب^{۴۰}، ترکیب^{۴۱} و جهش^{۴۲} معمولاً بیشترین کاربرد را در الگوریتم‌های ژنتیکی دارند. در این بخش، هر یک از عملگرهای فوق به‌صورت جداگانه معرفی می‌شود:

۱. عملگر انتخاب: این عملگر از بین کروموزوم‌های موجود در یک جمعیت، تعدادی کروموزوم را برای تولیدمثل انتخاب می‌کند. کروموزوم‌های برازنده‌تر، شانس بیشتری دارند تا برای تولیدمثل انتخاب شوند [۲۰].

۲. عملگر ترکیب: عملگر ترکیب بر روی یک زوج کروموزوم از نسل والد عمل کرده و یک زوج کروموزوم جدید تولید می‌کند. عملگرهای ترکیب متعددی از قبیل، ترکیب تک نقطه‌ای^{۴۳} و ترکیب دونقطه‌ای^{۴۴} وجود دارد. در ترکیب تک نقطه‌ای، یک موقعیت تصادفی بین دو ژن در نظر گرفته می‌شود. سپس تمامی ژن‌های طرف راستی اطراف چپ این موقعیت در کروموزوم‌های والد با یکدیگر جابجا می‌شوند تا کروموزوم‌های جدید به دست آیند. در ترکیب دونقطه‌ای، دو موقعیت به‌صورت تصادفی انتخاب می‌شود و تمامی ژن‌های بین این دو موقعیت در کروموزوم‌های والد با یکدیگر جابجا می‌شوند. لازم به ذکر است که ترکیب معمولاً بر روی همه زوج کروموزوم‌های انتخاب‌شده برای جفت‌گیری به‌کاربرده نمی‌شود. معمولاً احتمال ترکیب برای هر زوج کروموزوم بین $0/6$ تا $0/95$ در نظر گرفته می‌شود که به این عدد نرخ ترکیب^{۴۵} یا احتمال ترکیب^{۴۶} گفته می‌شود و با P_c نمایش داده می‌شود. در صورتی که بر روی روی یک زوج کروموزوم عمل ترکیب صورت نگیرد، فرزندان با تکرار نمودن والدین تولید می‌شوند [۲۰].

۳. عملگر جهش: پس از اتمام عمل ترکیب، عملگر جهش بر روی کروموزوم‌ها اثر داده می‌شود. این عملگر یک ژن از یک کروموزوم را به‌طور تصادفی انتخاب نموده و سپس محتوای آن ژن را تغییر می‌دهد. اگر ژن از جنس اعداد دودویی باشد، آن را به وارونش تبدیل می‌کند و چنان چه متعلق به یک مجموعه باشد، مقدار یا عنصر دیگری از آن مجموعه را به‌جای آن ژن قرار می‌دهد. احتمال انجام عمل جهش بر روی هر کروموزوم را نرخ جهش^{۴۷} یا احتمال جهش^{۴۸} می‌گویند و با P_m نمایش می‌دهند. معمولاً این

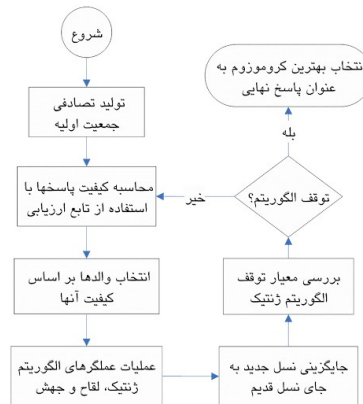
فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و نه / زمستان ۱۴۰۰

عدد را بسیار کوچک مثلاً (۰/۰۰۱) در نظر می‌گیرند. پس از اتمام عمل جهش، کروموزوم‌های تولیدشده به‌عنوان نسل جدید شناخته شده و برای دور بعد اجرای الگوریتم ارسال می‌شوند [۲۰]. روش‌های انتخاب متعددی برای استفاده در الگوریتم‌های ژنتیکی پیشنهاد شده‌اند که در ادامه این بخش، برخی از این روش‌ها معرفی می‌شوند:

۱. انتخاب متناسب با برازندگی: 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^{۳۹} پیشنهاد شد است [۳۸].

۲. نمونه‌برداری به روش چرخ رولت: در این روش، به هر فرد قطعه‌ای^{۴۰} از یک چرخ رولت مدور اختصاص داده می‌شود. اندازه‌ی این قطعه متناسب با برازندگی آن فرد است. چرخ 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 به‌صورت (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 [۳۸]. در شکل ۱، مراحل الگوریتم ژنتیک به‌طور خلاصه نشان داده شده است:



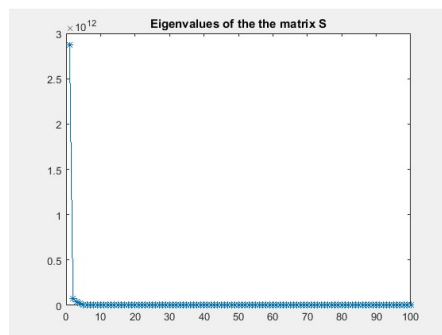
شکل ۱- مراحل الگوریتم ژنتیک [۱].

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

نتایج تجربی

این بخش شامل نتایج عددی با استفاده از الگوریتم تحلیل طیفی تکین و الگوریتم ژنتیک بر روی داده‌های مورد نظر است. آزمایش‌ها، شامل دو مرحله می‌باشد. در مرحله اول از روش تحلیل طیفی تکین برای تحلیل داده‌ها استفاده می‌شود و در مرحله دوم با استفاده از الگوریتم ژنتیک بهینه‌ترین مقادیر برای پارامترهای روش تحلیل طیفی تکین به دست می‌آید و داده‌ها مجدد با این روش بررسی می‌گردد و سرانجام نتایج حاصل از این دو روش باهم مقایسه می‌شود. هم‌چنین لازم به ذکر است که کلیه تحلیل‌ها با نرم‌افزار متلب انجام گردیده است.

الف: تحلیل شاخص 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 اینجا 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 محاسبه و نمودار هر کدام ترسیم می‌شود.

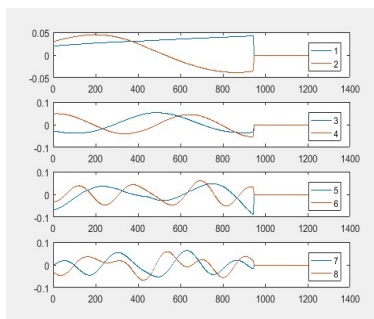


شکل ۲- نمودار نیمه لگاریتمی ۱۰۰ مقدار ویژه اول

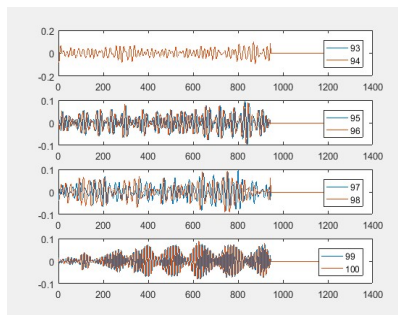
همان‌طور که در شکل ۲، مشاهده می‌شود، ۳ مقدار ویژه ابتدا دارای شیب تندی هستند ولی در ادامه حرکت نمودار بسیار ملایم و حتی نزدیک به نمودار افقی می‌شود. از آنجایی که سری نویز معمولاً دنباله‌ای از مقادیر ویژه را تولید می‌کند که به کندی کاهش می‌یابند، و در شکل ۲، می‌توان دو قطعه متفاوت را مشاهده کرد که یکی با شیب تند که نمایشگر سیگنال است و دیگری بخش هموارتر که نمایشگر نویز در سری‌های زمانی می‌باشد که این قضیه نشان می‌دهد، ۳ مؤلفه ابتدا را می‌توان به عنوان مؤلفه‌های سیگنال در نظر گرفت و مابقی مؤلفه‌ها، نقش مؤلفه‌های نویز را بر عهده می‌گیرند. در ادامه به بررسی هرچه دقیق‌تر این شمارش پرداخته می‌شود، حال زمان آن فرارسیده است که نمودار بردارهای ویژه متناظر با این مقادیر ویژه ترسیم شود. به دلیل این که تعداد بردارهای ویژه بسیار زیاد است، در واقع

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و نه / زمستان ۱۴۰۰

به‌طور دقیق‌تر برابر با ۹۴۵ می‌باشد، هشت بردار ویژه ابتدا و انتها بررسی می‌شود. شکل ۳ و شکل ۴، به ترتیب بردارهای ویژه متناظر با هشت مقدار ویژه ابتدایی و بردارهای ویژه متناظر با هشت مقدار ویژه انتهایی را نشان می‌دهند.



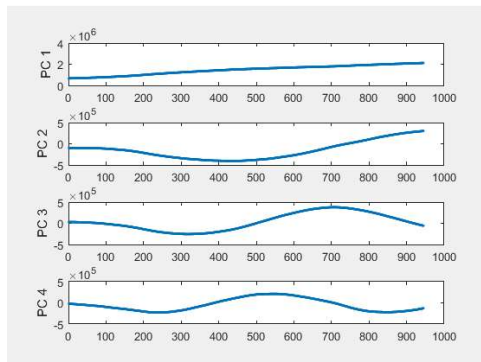
شکل ۳- بردارهای ویژه متناظر با هشت مقدار ویژه ابتدایی کل



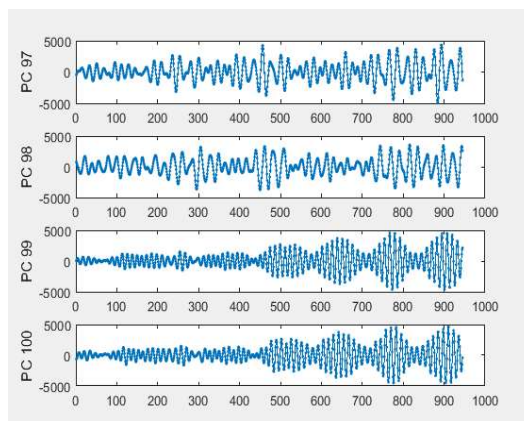
۴- بردارهای ویژه متناظر با هشت مقدار ویژه انتهایی

در ادامه ماتریس مسیر تجزیه می‌شود و همان‌طور که در بخش سوم بیان شد، بخشی از این تجزیه، مؤلفه‌های اصلی نامیده می‌شود. مجدداً مؤلفه‌های اصلی برای چهار مقادیر ویژه ابتدا و انتها رسم می‌شود تا موجب اطمینان بیشتری در تشخیص مؤلفه‌های سیگنال از نویز شود. همانند نمودارهای بردارهای ویژه، مؤلفه‌های نویز برخلاف مؤلفه‌های سیگنال تغییرات بسیاری دارند. شکل ۵، مؤلفه‌های اصلی سری‌های زمانی برای چهار مقدار ویژه ابتدایی و شکل ۶، مؤلفه‌های اصلی سری زمانی برای چهار مقدار ویژه انتهایی را نشان می‌دهند.

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

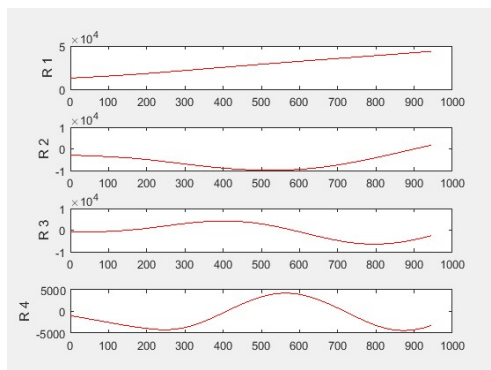


شکل ۵- مؤلفه‌های اصلی سری‌های زمانی برای ۴ مقدار ویژه ابتدا

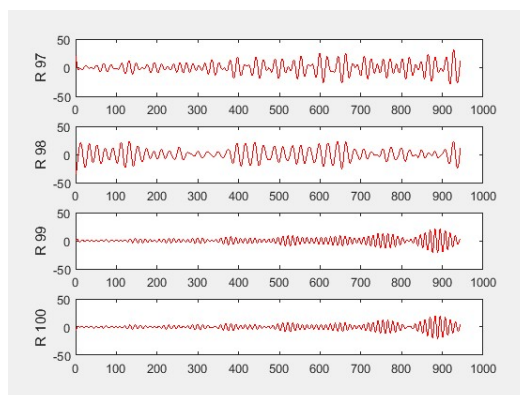


شکل ۶- مؤلفه‌های اصلی سری‌های زمانی برای ۴ مقدار ویژه انتها

برای تعیین مؤلفه‌های بازسازی‌شده ابتدا باید ماتریس $(N - L + 1) \times L$ با استفاده از ترانهاده ماتریس مؤلفه‌های اصلی و ماتریس بردارهای ویژه محاسبه شود. سپس باید روش میانگین‌گیری یا همان هنکل‌سازی بر روی قطر فرعی این ماتریس که از مؤلفه‌های بازسازی‌شده برای سری زمانی اولیه ساخته شده، اعمال گردد. همان‌طور که در دو شکل ۷ و ۸ مشاهده می‌شود، عمل تجزیه و بازسازی تا حد زیادی در جداسازی سیگنال از نویز در نتایج تأثیر داشته است. اگر چهار نمودار ابتدای مؤلفه‌های اصلی را با چهار نمودار ابتدا در بردارهای ویژه مقایسه شود، کاهش تغییرات مؤلفه‌ها آشکار است و هم‌چنین این تغییرات در چهار مؤلفه انتهایی هم به‌وضوح دیده می‌شود.



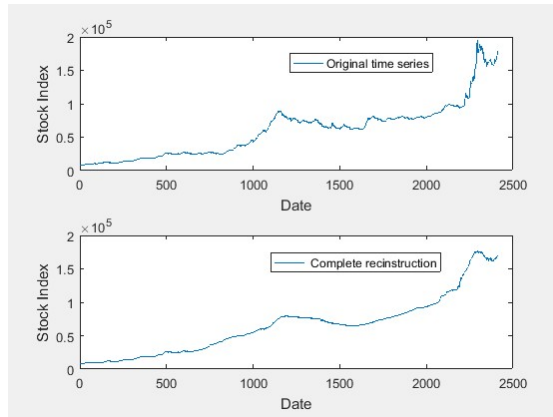
شکل ۷- مؤلفه‌های بازسازی شده سری‌های زمانی برای ۴ مقدار ویژه ابتدا



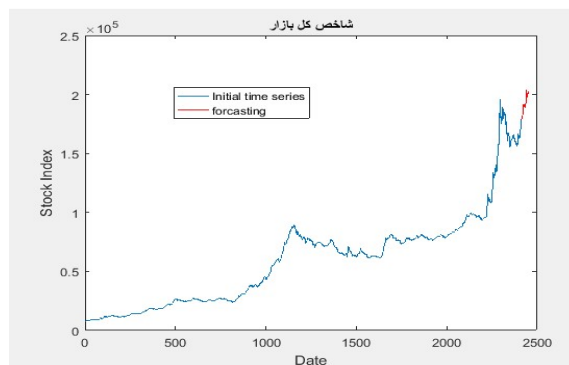
شکل ۸- مؤلفه‌های بازسازی شده سری‌های زمانی برای ۴ مقدار ویژه انتها

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 به دست آمده، نتایج بعدی پیش‌بینی شود. در این مرحله، ۴۰ بازه زمانی، که در اینجا ۴۰ روز کاری تعبیر شده (از تاریخ ۹۸/۱/۵ الی ۹۸/۳/۵)، پیش‌بینی می‌شود. همان‌طور در شکل ۱۰ نشان داده شده است، نتایج پیش‌بینی حکایت از صعود بعد از اصلاح را در شاخص نشان می‌دهد.

پیش بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد



شکل ۹- مقایسه سری های زمانی اولیه و با ۳ مؤلفه های بازسازی شده



شکل ۱۰- نمودار پیش بینی ۴۰ روز کاری شاخص کل بازار

حال زمان آن رسیده با استفاده از الگوریتم ژنتیک مناسب ترین طول پنجره و گروه بندی برای تجزیه طیفی تکین محاسبه شود. در اینجا هدف یافتن کمترین مقدار همبستگی بین سیگنال و نویز است پس تابع هدف مسئله به صورت زیر (رابطه ۱۲) می باشد [۲۱]:

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

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

ب- تحلیل شاخص با استفاده از الگوریتم ژنتیک: 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

فصلنامه مهندسی مالی و مدیریت اوراق بهادار / شماره چهل و نه / زمستان ۱۴۰۰

اندازه‌گیری شده و صحت عملکرد کنترل‌کننده توسط برازندگی این تابع سنجیده می‌شود. هدف یافتن کمترین همبستگی وزنی بین مؤلفه‌های سری زمانی جهت تفکیک سیگنال و نویز (یافتن نقطه‌ی برش) و سپس به دست آوردن طول پنجره‌ی بهینه می‌باشد. لذا تابع هزینه‌ی تعریف‌شده همبستگی وزنی بین مؤلفه‌های سیگنال و نویز است. جدول ۲، مقدار پارامترهای الگوریتم ژنتیک را نشان می‌دهد.

جدول ۲- پارامترهای الگوریتم ژنتیک

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

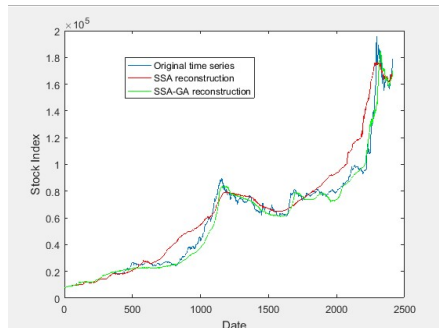
پس از اعمال الگوریتم ژنتیک با نرم‌افزار متلب مقادیر بهینه برای پارامترهای مسئله به صورت جدول ۳ می‌باشد.

جدول ۳- پارامترهای بهینه با استفاده از الگوریتم ژنتیک

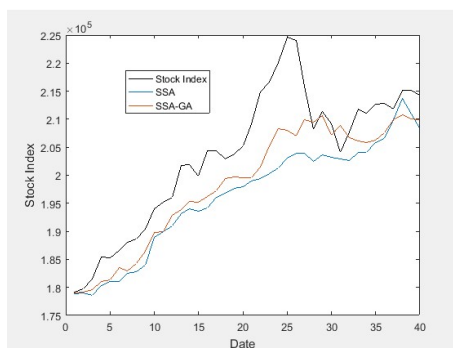
پارامتر	مقدار
مقدار بهینه تابع معیار	۲/۱۴۹
طول پنجره L	۱۰۴۲
نقطه‌ی برش r	$A=\{۱,۲,...,۶\}$

با توجه به جدول ۳، طول پنجره بهینه تقریباً در نزدیکی میانه سری زمانی به دست آمده و مقادیر ویژه سه‌گانه $A=\{۱,۲,...,۶\}$ به عنوان مؤلفه‌های سیگنال و مابقی مؤلفه‌های نویز در نظر گرفته می‌شود. 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 ۱۲، نتایج حاصل را با سری زمانی بازسازی شده در مرحله قبل مقایسه می‌کند. قابل توجه است که با در نظر گرفتن مقدار ۱۰۴۲ برای طول پنجره و ۶ مؤلفه در مرحله گروه‌بندی دقت بسیاری در نتایج حاصل می‌شود. حال با استفاده از مقادیر به دست آمده توسط الگوریتم ژنتیک مقادیر ۴۰ روز آتی شاخص کل بازار محاسبه و با نتایج حاصل از مرحله اول مقایسه می‌شود (شکل ۱۳).

پیش بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد



شکل ۱۲- نمودار مقایسه شاخص کل بازار و الگوریتم ژنتیک



شکل ۱۳- نمودار مقایسه تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک بر روی شاخص کل بازار در

۴۰ روز کاری

عملکرد بازسازی رامی توان با معیارهای استاندارد خطا، از جمله خطای مطلق میانگین^{۴۱} بررسی کرد و هرچه این خطا مقادیر کم را محاسبه کند، نشان از عملکرد مناسب این روش هست. خطای مطلق میانگین به وسیلهی رابطه‌ی زیر (۱۳) محاسبه می‌شود:

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

که در آن x_n یک نمونه از سری زمانی اصلی در زمان n ، y_n نمونه‌ی بازسازی شده سری زمانی در همان زمان و N نشان‌دهنده‌ی اندازه سری زمانی اصلی است [۲۱]. خطای مطلق میانگین سری بازسازی شده برای بررسی عملکرد دو روش در جدول ۴ نشان داده می‌شود:

جدول ۴- خطای قدر مطلق میانگین

خطای مطلق میانگین	روش استفاده شده
.۳۸۶۲	تحلیل طیفی تکین
.۲۹۹۲	تحلیل طیفی با الگوریتم ژنتیک

نتیجه گیری

در این مقاله، با استفاده از نرم افزار متلب، داده های ده ساله شاخص کل بورس اوراق بهادار تهران طی سال های ۱۳۸۸ تا ۱۳۹۷ با استفاده از روش تجزیه طیفی تکین در سه مرحله پیاده سازی گردید و نشان داده شد که این روش تا حدودی در پیش بینی شاخص مناسب است ولی نیاز به بررسی های بیشتر بر روی پارامترهای روند تجزیه که همان طول پنجره و نقطه ی برش هست، دارد. سپس در قالب یک مسئله بهینه سازی توسط الگوریتم ژنتیک حل 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 اندازه گیری شده و صحت عملکرد کنترل کننده توسط برازندگی این تابع سنجیده شد. نهایتاً نیز جهت صحت بیشتر موضوع، پیش بینی داده های ۴۰ روز کاری شاخص کل بازار با روش های تحلیل طیفی تکین و تحلیل طیفی مبتنی بر الگوریتم ژنتیک صورت گرفت و نتایج حاصل از خطای قدر مطلق میانگین نشان داد با بهبود در عملکرد روش تحلیل طیفی با الگوریتم ژنتیک، افزایش دقت در محاسبات به روشنی دیده می شود. به طور کلی نتایج حاصل از این تحقیق نشان داد که؛ تفکیک پذیری سیگنال و نویز با روش تحلیل طیفی تکین امکان پذیر است و با توجه به نتایج حاصل، تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک با داشتن خطای قدر مطلق میانگین کمتر، بهبود در دقت پیش بینی را به همراه داشت. با توجه به هدف یافتن کمترین همبستگی وزنی بین مؤلفه های سری زمانی جهت تفکیک سیگنال و نویز (یافتن نقطه ی برش) و سپس به دست آوردن طول پنجره ی بهینه که در تحلیل طیفی تکین مبتنی بر الگوریتم ژنتیک یافت می شود در واقع تابع هزینه ی تعریف شده همبستگی وزنی بین مؤلفه های سیگنال و نویز است گویای این واقعیت است که تغییر در مقدار پارامترها تا چه اندازه می تواند در بهبود عملکرد روش تحلیل طیفی مفید واقع شود

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

منابع

- ۱) اسکندری شیما و علی محرم‌پور. تفکیک‌پذیری و تعیین طول پنجره در روش تحلیل طیفی تکین با الگوریتم ژنتیک. کنفرانس بین‌المللی پنجم مهندسی کامپیوتر و برق. ۲۰۱۸.
- ۲) رهنمای‌رودپشتی‌فریدون ومهدیه کلانتری‌دهقی. مدل‌های مولتی‌فرکتال درعلوم‌مالی: ریشه، ویژگی‌ها و کاربردهای آنها. دانش مالی تحلیل اوراق بهادار. ۱۳۹۳. چهارم. بیست‌وهفتم. زمستان: صص ۴۷-۲۵.
- ۳) سلیمی نسب سهیل، محمدعلی متذکری و سلمان سلیمی نسب. کاربرد سری فوریه در قیمت‌گذاری اختیارات معامله در بورس اوراق بهادار. اولین کنفرانس ملی ریاضیات صنعتی. ۱۳۹۳. تبریز. خرداد: صص ۵-۱.
- ۴) عباسی‌نژاد حسین، شاپور محمدی و سجاد ابراهیمی. مقایسه مدل‌های نوسان‌پذیری چندمتغیره در برآورد رابطه بین نرخ ارز و شاخص سهام. دانش سرمایه‌گذاری. ۱۳۹۳. سوم. یازدهم. پاییز: صص ۲۲۱-۲۰۱.
- ۵) فرزین‌وش اسدالله و موسی عباسی. بررسی ارتباط بین تورم و نااطمینانی تورمی در ایران با استفاده از مدل‌های GARCH و حالت - فضا. تحقیقات اقتصادی. ۱۳۸۴. ۷۴: صص ۵۵-۲۵.
- ۶) محمدی سجاد. طراحی شاخص سهام سازگار با بورس اوراق بهادار تهران. پایان‌نامه کارشناسی ارشد. دانشکده معارف اسلامی و مدیریت دانشگاه امام صادق (ع). ۱۳۸۶.
- ۷) موتنی مانی. تحلیل نوسانات بازدهی در بازار سهام تهران. پایان‌نامه کارشناسی ارشد. دانشکده علوم اقتصادی و اداری دانشگاه مازندران. ۱۳۸۵.
- ۸) میرزاجانی بجستانی مسعود واحسان. رمز. تحلیل مجموعه مقادیر تکین: مطالعه موردی بهای سکه. آمار(ندا). ۱۳۹۳. دوازدهم. دوم: صص ۳۹-۳۰.
- ۹) یارمحمدی مسعود و رحیم محمودوند. پیش‌بینی نرخ ارز با استفاده از روش تحلیل مجموعه‌ی مقادیر تکین. مطالعات اقتصادی کاربردی ایران. ۱۳۹۵. سال پنجم. ۱۸. تابستان: صص ۱۴۱-۱۳۳.
- 10) 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.
- 11) Abreu R.J., R.M. Souza and J.G. Oliveira. Applying singular spectrum analysis and ARIMA-GARCH for forecasting. 2019.
- 12) Alexandrov T heodore. A method of trend extraction using singular spectrum analysis. Ar Xiv preprint arXiv: 0804.3367. 2008.
- 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) 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.
- 19) Goldberg David E., Bradley Korb and Kalyanmoy Deb. Messy genetic algorithms: Motivation, analysis, and first results. *Complex systems*. 1989. 3: pp. 493-530.
- 20) Goldberg, David E., *Genetic Algorithms in Search, “Optimization and Machine Learning”*, Addison Wesley Longman Publishing Co. Inc. Boston, MA, USA, 1989.
- 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) Hassani Hossein. *Singular spectrum analysis: methodology and comparison*. 2007. 5: pp. 239-257.
- 25) Hassani Hossein. *Singular spectrum analysis based on the minimum variance estimator*. 2010.
- 26) Hassani Hossein, Rahim Mahmoudvand and Masoud Yarmohammadi. Filtering and denoising in linear regression analysis. *Fluctuation and Noise Letters*. 2010. 9: pp. 343-358.
- 27) 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.
- 28) 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.

پیش‌بینی شاخص کل بورس اوراق بهادار تهران با استفاده از روش.../حسن دوست و وکیلی فرد

- 29) Hu Jun and Juho Kannianen. Asymptotic expansion of European options with mean-reverting stochastic volatility dynamics. Finance Research Letters. 2015. 14: pp. 1-10.
- 30) 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.
- 31) 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.
- 32) Osmanzade, Aytan. and P. Miidla, Singular spectrum analysis forecasting for financial time series. Master Thesis. University of Tartu. 2017.
- 33) Sanei Saeid and Hossein Hassani. Singular spectrum analysis of biomedical signals. 2015.
- 34) 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.
- 35) 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.
- 36) 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.
- 37) Vautard Robert and Michael Ghil. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Physica D35.1989. 35(3): pp. 395-424.
- 38) 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.
- 39) Zhu J. Pricing volatility of stock returns with volatile and persistent components. Financ Mark Portf Manag. 2009. 23: pp. 243-269.

یادداشت‌ها:

-
۱. Stochastic
 ۲. SV: Stochastic Volatility
 ۳. GARCH: Generalized AutoRegressive Conditional Heteroskedasticity

۴. ARCH: AutoRegressive Conditional Heteroscedasticity
۵. AS Tallink Group
۶. SVR: Support Vector Regression
۷. PSO: Particle Swarm Optimization
۸. Decomposition
۹. Embedding
20. The Trajectory Matrix
۱۱. Length Window
۱۲. Hankel Matrix
۱۳. SVD: Singular Value Decomposition
۱۴. Eigenvalues
11. Elementary Matrices
۱۶. Principal components
۱۷. Eigen Triple
۱۸. The spectrum
۱۹. Reconstruction
۲۰. Grouping
17. Eigentriple Grouping
۲۲. Diagonal Averaging
۲۳. Hankelization
۲۴. LRF: Linear Recurrent Formulate
۲۵. Break Point
۲۶. PSO: Particle Swarm Optimization
۲۷. SSA-CT: Colonial Theory
۲۸. Reproduction
۲۹. Generation
۳۰. Selection
۳۱. Crossover
۳۲. Mutation
۳۳. One-Point Crossover
۳۴. Two-Point Crossover
۳۵. Crossover Rate
۳۶. Crossover Probability

- ۳۷. Mutation Rate
- ۳۸. Mutation Probability
- ۳۹. Roulette Wheel
- ۴۰. Slice
- ۴۱. MAE: Mean Absolute Error