فصلنامــه علمی پژوهشی دانش سرمایه‌گــذاری
سال هشتم/ شماره سی‌ودوم/ زمستان ۱۳۹۸

استراتژی آربیتراژ آماری بر مبنای مدل‌های عاملی قیمت سهام در بورس تهران

فریماه مخاطب رفیعی
دانشیار دانشگاه تربیت مدرس، دانشکده مهندسی صنایع و سیستم‌ها
f.mokhatab@modares.ac.ir

کامیار نوربخش
دانش آموخته کارشناسی ارشد مهندسی صنایع گرایش سیستم‌های مالی، دانشگاه تربیت مدرس، (نویسنده مسئول)
kamyarnourbakhsh@yahoo.com

تاریخ دریافت: ۹۶/۰۸/۲۰ تاریخ پذیرش: ۹۶/۱۱/۱۷

چکیده

استراتژی‌های آربیتراژ آماری به دنبال کشف فرصت‌های سودده با استفاده از روش‌های آماری هستند. ما در این مقاله از رویکرد جدیدی برای طراحی یک استراتژی آربیتراژ آماری متناسب با بورس تهران استفاده می‌کنیم. در این رویکرد پیشنهادی جدید، بجای اینکه قیمت سهام مختلف به صورت مستقل مدل‌سازی و پیش‌بینی شود، آن‌ها را به صورت یک کل در نظر می‌گیریم و الگوهای حرکتی مشترک بین آن‌ها که می‌تواند بیانگر حرکات کلی بازار باشد، را با استفاده از روش تحلیل اجزای اصلی شناسایی می‌کنیم. بعد از آن رفتار این الگوها (عامل‌ها) را مدل‌سازی و پیش‌بینی کرده و از طریق آن‌ها، بازدهی سهام را در آینده پیش‌بینی می‌کنیم. در نهایت سبدی از سهام منتخب در هر دوره تشکیل می‌دهیم. نتایج تجربی حاصل از این مقاله نشان‌دهنده سودده بودن این استراتژی‌ها است به طوری که استراتژی منتخب با پنجره زمانی ۱۰۰ روزه و افق پیش‌بینی ۱ روزه، بدون در نظر گرفتن هزینه معاملاتی توانست بازدهی متوسط سالیانه ۱۱۵٪ را فراهم کند.

واژه‌های کلیدی: آربیتراژ آماری، مدل‌های عاملی، تحلیل اجزای اصلی.

۱- مقدمه

پیش‌بینی حرکت قیمت دارایی موضوع چالش برانگیزی است و بر طبق فرض بازار کارا، قیمت‌ها باید از فرآیند گام تصادفی پیروی کنند. اما مطالعات زیادی به منظور کشف فرصت‌های سودآور که مدرکی بر ناکارآمدی بازار است، انجام شده است. برخی از این مطالعات سعی بر این داشته‌اند تا با استفاده از مدل‌های ریاضی و آماری به طراحی استراتژی‌هایی به منظور تولید بازدهی از حرکت‌های سیستماتیک در قیمت سهام بپردازند. این استراتژی‌ها که بعدها استراتژی‌های آربیتراژ آماری نام گرفتند، مدت زمان بسیاری است که توسط سرمایه‌گذاران حرفه‌ای مورد استفاده قرار می‌گیرند. بطوریکه اولین بار در سال ۱۹۸۰ تیمی متشکل از فیزیک‌دانان، ریاضی‌دانان و متخصصین کامپیوتر در شرکت مورگان استنلی[1] به منظور کشف این فرصت‌های سودآور تشکیل شد و نتایج رضایت بخشی برای آن‌ها در برداشت. معاملات زوجی[2] ساده‌ترین و شناخته شده‌ترین استراتژی آربیتراژ آماری است که بر اساس رابطه بین سطح قیمت‌های دو دارایی مالی عمل می‌کند. با گذشت زمان، استراتژی‌های آربیتراژ آماری پیچیدگی‌های بیشتری پیدا کردند و از مدل‌های ریاضیاتی و آماری به منظور کشف الگوهای موجود در حرکت سهام بهره بردند.

در این مقاله ما از رویکرد جدیدی برای طراحی یک استراتژی معاملاتی متناسب با بورس تهران استفاده می‌کنیم و نشان می‌دهیم که این رویکرد جدید روش قدرتمندی برای پیاده سازی و مطالعه استراتژی‌های آربیتراژ آماری است. مقالات زیادی به منظور طراحی استراتژی‌های معاملاتی، به صورت مستقیم به مدل‌سازی قیمت سهام مختلف پرداخته و با استفاده از این مدل‌ها، سعی کرده‌اند قیمت را در آینده پیش‌بینی کنند. ما در این مقاله با رویکردی متفاوت، سهام موجود در بازار را به صورت یک کل و در ارتباط با یکدیگر در نظر می‌گیریم و الگوهای حرکتی مشترک بین آن‌ها را به صورت عامل‌های مجازی استخراج می‌کنیم. هر یک از این عامل‌ها (الگوها) می‌تواند در شرایط و زمان‌های مختلف، تفسیر متفاوتی داشته باشد. به عنوان مثال عامل اول که بیشترین حرکات مشابه در بین سهام را منعکس می‌کند، می‌توان به عنوان سبد بازار در نظر گرفت که رفتاری مشابه با شاخص کل بورس از خود نشان می‌دهد (اولاندا و لی[3]، ۲۰۱۰). عامل های بعدی نیز به ترتیب به دنبال کشف سایر الگوهای حرکتی در بین سهام هستند. برای تخمین این عامل‌ها از روش اجزای اصلی[4] که یک تکنیک مفید آماری برای یافتن الگوها در داده‌های با ابعاد بالا است و کاربردهای زیادی در حوزه‌های مختلف پیدا کرده است، استفاده می‌شود. این تکنیک، از دل مجموعه داده‌ای شامل تعداد زیادی متغیر، تعدادی متغیر جدید ایجاد می‌کند. به طوریکه این متغیرهای جدید، بیشترین اطلاعات موجود از متغیرهای اصلی را در برداشته باشند. در ادامه بجای مدل‌سازی مستقیم قیمت سهام، این عامل‌ها یا همان متغیرهای تخمین زده شده، که نمایان‌گر الگوهای پنهان حرکتی بین سهام مختلف هستند را مدل‌سازی و برای دوره‌های زمانی آینده پیش‌بینی می‌کنیم. در نهایت بر اساس پیش‌بینی رفتار این الگوها، قیمت و بازدهی سهام را برآورد می‌کنیم و سبدی از سهام منتخب تشکیل می‌دهیم. در نظر داشته باشید که در تشکیل سبد، تنها از موقعیت خرید استفاده شده است که با شرایط بورس تهران هماهنگ باشد. در ادامه بازده حاصل از استراتژی مطرح شده در مقاله را با بازده

استراتژی خرید و نگهداری[5] سهام با وزن یکسان در ابتدای دوره مقایسه می‌کنیم و نشان می‌دهیم استراتژی آربیتراژ آماری بسیار بهتر از استراتژی مورد مقایسه عمل می‌کند.

ادامه مقاله به این ترتیب ساماندهی شده است که ابتدا به مرور خلاصه‌ای از ادبیات موضوع پیرامون آربیتراژ آماری و مدل‌های عاملی می‌پردازیم. پس از آن داده‌ها و مدل‌های مورد استفاده در بخش دوم مطرح می‌شود و استراتژی‌های تشکیل سبد معاملاتی تشریح می‌گردد. نتایج بدست آمده در بخش سوم بیان می‌شود و در بخش انتهایی نتیجه‌گیری و جمع‌بندی انجام می‌گیرد.

پیشینه پژوهش

بینش نهفته در آربیتراژ آماری، بهره‌گیری از فاصله میان بازده‌های مورد انتظار سبدهای دارایی‌های مالی است. پیش‌گامان استراتژی‌های آربیتراژ آماری، استراتژی‌های معاملات زوجی هستند که به کشف دارایی‌هایی که قیمت آنها تقریبا مسیر یکسانی را طی می‌کند، می‌پردازند. ویدیامورفی[6] (۲۰۰۴) و پل[7] (۲۰۰۷) به معرفی کلیات معاملات زوجی پرداخته‌اند. مطالعات بسیاری در حوزه معاملات زوجی در داخل و خارج از کشور انجام شده است. به طوری که گیتیو و همکاران[8] (۲۰۰۶) نشان دادند که بازدهی تولید شده از استراتژی‌های آربیتراژ آماری به علت استفاده فراوان توسط معامله‌گران، کاهش یافته است. بنابراین توسعه مدل‌ها و تکنیک‌های جدید مورد توجه قرار گرفت.

معاملات زوجی بعدها به آربیتراژ آماری توسعه یافت. یکی از انواع استراتژی‌های آربیتراژ آماری، آربیتراژ ریسک یا آربیتراژ خنثی نسبت به بازار است که بر مبنای تئوری کلاسیک تعادل عمل می‌کند. بیشترین مدل‌های بازار خنثی بر روی مدل قیمت‌گذاری دارایی سرمایه‌ای (CAPM) که توسط شارپ[9] (۱۹۶۴)، لینتر[10] (۱۹۶۵) و بلک[11] (۱۹۷۲) توسعه یافت، ساخته شدند. بعد از فراهم آمدن تعریف دقیقی از آربیتراژ آماری، هوگان و همکاران[12] (۲۰۰۴) تستی برای کارایی بازار ارائه کردند که این تست توسط جارو و همکاران[13] (۲۰۱۲) توسعه یافت. آنها به صورت تجربی استراتژی‌های تکانه و معامله ارزش را بررسی کردند و به این نتیجه رسیدند که این استراتژی‌ها، آربیتراژ آماری تولید می‌کنند. احمدزاده و همکاران (۱۳۹۳) کارایی بورس تهران را بر مبنای آربیتراژ آماری بررسی کردند و به این نتیجه رسیدند که بورس تهران کارا نیست. اولاندا و لی[14] (۲۰۱۰) استراتژی‌های آربیتراژ آماری را در فضای سهام S&P500 بررسی کردند. آنها بازدهی سهام مختلف را به دو طریق مدل‌سازی کردند. در مدل اول از روش اجزای اصلی و برآورد عامل اول از درون ماتریس همبستگی سهام منتخب، استفاده کردند. آنها این عامل اول را تخمینی از حرکت کل بازار در نظر گرفتند. سپس بازدهی سهام را با استفاده از این عامل مدل کردند. مدل دوم مشابه مدل اول رفتار بازدهی سهام را مدل می‌کرد. با این تفاوت که در ساخت آن بجای عامل اصلی اول، از بازدهی صندوق‌های قابل معامله استفاده کردند. پس از مدل‌سازی بازدهی سهام، یک استراتژی خنثی نسبت به بازار طراحی کردند که از موقعیت خرید و فروش به طور همزمان استفاده می‌کرد. نتایج حاصل نشان‌دهنده سودده بودن این استراتژی بود.

در میزهای معاملاتی پر بسامد[15]، روزانه هزاران معامله بر روی صدها سهم انجام می‌شود. در نتیجه انجام روش‌هایی مثل شبیه‌سازی برای محاسبه پارامترهای استراتژی، به اندازه کافی سریع نیستند. برترام[16] (۲۰۱۰) برای حل این مشکل، راه‌حلی تحلیلی با استفاده از مدل‌های ریاضیاتی برای معاملات آربیتراژ آماری بر روی سهامی که خاصیت بازگشت به میانگین از خود نشان می‌دادند، بدست آورد. او برای این منظور از بهینه سازی بر روی بازده مورد انتظار و نسبت شارپ استفاده کرد. فوکاردی و همکاران[17] (۲۰۱۶) رویکرد جدیدی برای استراتژی‌های آربیتراژ آماری بر مبنای مدل‌های عاملی پویای قیمت سهام ارائه کردند. آنها بازده و قیمت سهام را به طور همزمان با استفاده از ۱۵ عامل اصلی، که با روش اجزای اصلی تخمین زده می‌شدند، مدل کردند و به این نتیجه رسیدند که مدل‌های بر مبنای قیمت به علت وجود خاصیت بازگشت به میانگین، عملکرد بهتری از مدل‌های بر مبنای بازدهی دارند. در این مقاله آنها با رویکرد جدیدی، در عوض پیش‌بینی قیمت سهام، به پیش‌بینی عامل‌های اصلی تخمین زده شده پرداختند و سعی کردند با استفاده از این پیش‌بینی‌ها، قیمت سهام را در آینده تخمین بزنند.

در این پژوهش نیز از تحلیل اجزای اصلی[18] برای تخمین عامل‌ها و مدل‌سازی قیمت سهام بر روی آنها استفاده شده است. تحلیل اجزای اصلی یک روش چند متغیره آماری است که در علوم مختلف استفاده می‌شود. این تکنیک از دل مجموعه داده‌های متشکل از تعداد زیادی متغیر، متغیرهای جدیدی را شناسایی می‌کند که بیانگر الگوهایی از رفتار کلی متغیرهای اصلی هستند. تحلیل عاملی از فنون مربوط به تحلیل اجزای اصلی استفاده می‌کند. مدل‌های عاملی پویا اولین بار در سال ۱۹۷۷ توسط گوک[19] (۱۹۷۷) و سارجنت و سیمز[20] (۱۹۷۷) معرفی شد. پنا و باکس[21] (۱۹۸۷) مدل‌های عاملی پویا را در یک چهارچوب فضای حالت فرموله کردند بطوریکه متغیرها بر روی عامل‌ها رگرس می‌شدند و عامل‌ها از یک فرایند میانگین متحرک خود رگرسیو[22] پیروی می‌کردند. اکلاند و کاپتانیوس[23] (۲۰۰۸) روش‌های پیش‌بینی برای مجموعه داده‌های بزرگ را بازبینی کردند. برهومی و همکاران[24] (۲۰۱۳) بازبینی ای از مقالات مربوط به مدل‌های عاملی پویا فراهم کردند.

۲- روش شناسی تحقیق

در این بخش در رابطه با داده ها و مدل‌های استفاده شده در این پژوهش بحث می‌کنیم. مراحل انجام این استراتژی به این صورت است که با استفاده از قیمت‌های سهام در یک پنجره زمانی، عامل‌هایی را با استفاده از روش تحلیل اجزای اصلی تخمین می‌زنیم که نمایانگر رفتار و الگوهای مشترک بین این سهام در همان پنجره زمانی است. در ادامه با پیشنهاد رویکرد جدیدی، بجای پیش‌بینی مستقیم قیمت سهام، رفتار این عامل‌ها را پیش‌بینی می‌کنیم. در نهایت در یک فرآیند معکوس، قیمت سهام را بر روی عامل‌های موجود مدل کرده و با استفاده از عامل‌های پیش‌بینی شده، قیمت سهام را پیش‌بینی می‌کنیم و سبدی از سهام منتخب تشکیل می‌دهیم.

داده ها

داده های استفاده شده در این پژوهش، قیمت روزانه آخرین معامله ۱۶۱ سهم فعال در بورس تهران از تاریخ ۱۳۹۰/۱/۶ تا ۱۳۹۵/۱۲/۲۸ می‌باشد. این قیمت‌ها نسبت به افزایش سرمایه و توزیع سود هر سهم تعدیل شده اند. در انتخاب سهام سعی شده است از سهامی استفاده شود که در اکثر روزهای سال فعال باشند و در هر پنجره زمانی در نظر گرفته شده برای استراتژی، حداکثر ۲۵ روز بدون معامله باشند. برای درک بهتر داده‌های مورد استفاده در این پژوهش، آمار توصیفی سهام مورد استفاده به تفکیک صنایع مختلف را می‌توانیم در جدول ۱ مشاهده کنیم.

جدول ۱- آمار توصیفی سهام مورد استفاده در پژوهش به تفکیک صنایع

صنعت	تعداد سهام	متوسط بازدهی سالانه	انحراف معیار بازدهی سالانه	کمترین بازدهی سالانه	بیشترین بازدهی سالانه
خدمات فنی و مهندسی	۲	۰,۵۴	۱,۲	-۰,۳۶	۲,۹
رایانه و فعالیت‌های وابسته به آن	۶	۰,۳	۰,۵۸	-۰,۲۴	۱,۳۲
انبوه‌سازی، املاک و مستغلات	۹	۰,۲۱	۰,۴۸	-۰,۲۱	۱,۰۶
بیمه و صندوق بازنشستگی به جز تامین اجتماعی	۷	۰,۱۷	۰,۴۷	-۰,۱۴	۱,۰۹
مخابرات	۱	۰,۲۷	۰,۲۱	-۰,۱۴	۰,۵۲
حمل و نقل، انبارداری و ارتباطات	۲	۰,۴۴	۰,۷	-۰,۳۶	۱,۶۵
سایر واسطه‌گری‌های مالی	۶	۰,۱۸	۰,۳۹	-۰,۲۳	۰,۷۸
بانک‌ها و موسسات اعتباری	۹	۰,۱۷	۰,۵۴	-۰,۲	۱,۲۱
سرمایه‌گذاری‌ها	۹	۰,۲۱	۰,۳۷	-۰,۱۹	۰,۸۹
هتل و رستوران	۱	۰,۹۸	۱,۵۹	-۰,۴۲	۳,۳۳
سایر محصولات کانی غیر فلزی	۵	۰,۶۴	۰,۹۶	۰,۰۶	۲,۵۱
سیمان، آهک و گچ	۵	۰,۲	۰,۷۸	-۰,۴۵	۱,۷۴
کاشی و سرامیک	۲	۰,۴۲	۰,۸۸	-۰,۳۵	۲,۰۵
محصولات شیمیایی	۸	۰,۳۷	۰,۷۶	-۰,۴۲	۱,۷۵
مواد و محصولات دارویی	۵	۰,۳۷	۰,۸۶	-۰,۱۷	۲,۰۹
محصولات غذایی و آشامیدنی به جز قند و شکر	۸	۰,۳۶	۰,۵۲	-۰,۲۸	۱,۲۳
شرکت‌های چند رشته‌ای صنعتی	۳	۰,۳	۰,۷۵	-۰,۱۷	۱,۷۹
قند و شکر	۸	۰,۸۸	۲,۰۴	-۰,۵۵	۴,۹۸
خودرو و ساخت قطعات	۲۲	۰,۴	۰,۸۳	-۰,۳۳	۱,۹
ماشین آلات و دستگاه‌های برقی	۳	۰,۲۷	۰,۵۶	-۰,۱۹	۱,۲۹

صنعت	تعداد سهام	متوسط بازدهی سالانه	انحراف معیار بازدهی سالانه	کمترین بازدهی سالانه	بیشترین بازدهی سالانه
ماشین‌آلات و تجهیزات	۷	۰,۴۴	۰,۶۶	۰,۰۱۴	۱,۷۳
ساخت محصولات فلزی	۲	۰,۴۴	۰,۷۹	-۰,۲۲	۱,۸۳
فلزات اساسی	۱۱	۰,۲۷	۰,۴	-۰,۲۵	۰,۹۷
لاستیک و پلاستیک	۵	۰,۴۱	۰,۶۱	-۰,۰۹۶	۱,۲۱
فرآورده‌های نفتی، کک و سوخت هسته‌ای	۶	۰,۵	۰,۷۴	-۰,۴۵	۱,۶۷
محصولات کاغذی	۱	۰,۵۳	۱,۰۹	-۰,۳۲	۲,۷۲
استخراج کانه‌های فلزی	۷	۰,۱۹	۰,۳۴	-۰,۳۹	۰,۵۱
استخراج نفت، گاز و خدمات جنبی جز اکتشاف	۱	۰,۲۸	۰,۶۱	-۰,۳۱	۱,۳۸

تحلیل اجزای اصلی

تحلیل اجزای اصلی یک تکنیک مفید آماری است که کاربردهای زیادی در حوزه‌های مختلف پیدا کرده است و روشی برای یافتن متغیرهای زیربنایی در داده‌های با ابعاد بالاست. این تکنیک از دل مجموعه داده‌های متشکل از تعداد زیادی متغیر، متغیرهای جدیدی را شناسایی می‌کند که بیانگر الگوهایی از رفتار کلی متغیرهای اصلی هستند. بنابراین هنگامی که داده‌های در اختیار ما شامل N متغیر است و هر متغیر به صورت یک سری زمانی از T مشاهده تشکیل شده است، می‌توان از دل این N متغیر، F سری زمانی جدید با همان تعداد مشاهده (T) استخراج کرد. در نظر داشته باشید که تعداد عامل‌های استخراج شده (F) کمتر از متغیرهای اصلی مساله است. این عامل‌ها، از ترکیب خطی متغیرهای حقیقی بدست می‌آیند و ویژگی اصلی آن‌ها به اینگونه است که اولین عامل تخمین زده شده، که به آن اولین عامل اصلی می‌گوییم، بیشترین واریانس ممکن از متغیرهای اصلی را در بر دارد. دومین عامل اصلی مقدار دیگری از واریانس باقی مانده را جذب می‌کند که کمتر از عامل اول است و به همین ترتیب تا آخرین عامل ادامه دارد.

در این پژوهش مطابق مقاله فوکاردی و همکاران[۲۵] (۲۰۱۶)، ما از لگاریتم قیمت سهام استفاده می‌کنیم. فرض کنید که N سری زمانی از لگاریتم قیمت N سهم مختلف در اختیار داریم. برای تخمین عامل‌ها طبق روش تحلیل اجزای اصلی، ابتدا ماتریس کوواریانس لگاریتم قیمت سهام را محاسبه کرده و سپس مقادیر و بردارهای ویژه[۲۶] این ماتریس را محاسبه می‌کنیم. پس از این مرحله برای محاسبه سری زمانی مربوط به عامل اول، بردار ویژه مربوط به بیشترین مقدار ویژه (که نشان دهنده بیشترین واریانس قابل جذب توسط یک متغیر از تمامی داده ها است)، را در ماتریس داده‌های اصلی به صورت زیر ضرب خطی می‌کنیم:

$$X_{(T\times N)} \times EV_{(N\times 1)}^{(1)} = F_{(T\times 1)}^{(1)} \qquad \text{رابطه ۱}$$

که در رابطه ۱، $X_{(T\times N)}$ لگاریتم قیمت سهام، $EV_{(N\times 1)}$ بردار ویژه مربوط به بیشترین مقدار ویژه ماتریس کوواریانس لگاریتم قیمت سهام و $F^{(1)}_{(T\times 1)}$ بردار عامل اصلی اول است. برای محاسبه عامل دوم، همین عملیات را برای بردار ویژه مربوط به دومین مقدار ویژه بزرگ محاسبه شده تکرار می‌کنیم و این کار را تا تخمین تمام عامل‌های مورد نظر ادامه می‌دهیم (چگونگی تعیین تعداد عامل‌ها در ادامه آورده شده است).

پس از محاسبه عامل‌ها، در فرآیندی معکوس می‌توانیم لگاریتم قیمت هر سهم را به صورت تابعی از این عامل‌ها مدل کنیم و رفتار آن را تخمین بزنیم. این کار به ما این امکان را می‌دهد که با داشتن پیش‌بینی از آینده عامل‌ها، قیمت سهام را با استفاده از آن‌ها برآورد کنیم. برای این منظور از فرمول تئوری زیر استفاده می‌کنیم:

$$P_i(t) = a_{i1}F_1(t) + \ldots + a_{ik}F_k(t) + \xi_i(t) \qquad i = 1,\ldots,N \qquad \text{(رابطه ۲)}$$

در رابطه ۲، $P_i(t)$ لگاریتم قیمت سهم i در زمان t، $F_k(t)$ عامل اصلی Kام در زمان t، a_{ij} حساسیت لگاریتم قیمت سهم i نسبت به عامل j و $\xi_i(t)$ بخشی از لگاریتم قیمت سهم است که مستقل از عامل‌ها حرکت می‌کند. عامل‌ها $(F_1,\ldots,F_k)$ برای تمامی سهام مشترک بوده و فرض می‌شود مقادیر $\xi_i(t)$ غیر همبسته هستند.

پس از تخمین مدل مربوط به رابطه ۲، نیاز داریم که عامل‌ها را برای دوره زمانی fh روز در آینده پیش‌بینی کنیم. عامل‌های مختلف، رفتار متفاوتی نسبت به یکدیگر دارند. بنابراین برای پیش‌بینی عامل‌ها باید از مدلی استفاده کنیم که بتواند برای تمامی آن‌ها پیش‌بینی مناسبی را ارائه دهد. برای این منظور از معیار Bayesian Information Criterion یا به اختصار BIC استفاده می‌کنیم. این معیار برای انتخاب بهترین مدل استفاده می‌شود به نحوی که بین پیچیدگی و کارایی مدل تعادلی برقرار می‌کند و به صورت زیر محاسبه می‌شود:

$$BIC = \ln(n)k - 2\ln(\widehat{L}) \qquad \text{(رابطه ۳)}$$

که n تعداد مشاهدات، k تعداد پارامترهای مدل و $\widehat{L}$ بیشترین مقدار تابع درست نمایی[27] برای هر مدل است. در نتیجه به منظور انتخاب بهترین مدل پیش‌بینی، مدل‌های مختلف سری زمانی را برای تمام عامل‌ها اجرا کرده و مقدار BIC را برای هر کدام محاسبه می‌کنیم. در ادامه برای هر مدل، مقدار BIC محاسبه شده برای تمامی عامل‌ها را با هم جمع کرده و مدلی انتخاب می‌شود که کمترین مقدار را برای این حاصل جمع داشته باشد. بنابراین با کمی تقریب، مدل پیش‌بینی ARIMA(1,0,1) انتخاب می‌شود. پس از این مرحله، مدل اریمای انتخاب شده را برای هر عامل به صورت جداگانه تخمین می‌زنیم و با استفاده از آن، مقدار عامل را برای fh دوره در آینده پیش‌بینی می‌کنیم. پس از پیش‌بینی تمامی عامل‌ها، لگاریتم قیمت سهام را برای fh دوره در آینده با استفاده از رابطه ۲ تخمین می‌زنیم. سپس بازده مورد انتظار با استفاده از رابطه ۴ محاسبه می‌شود.

$$fR_i(t) = \frac{\exp(fp_i(t+fh)) - P_i(t)}{P_i(t)} \qquad \text{(رابطه ۴)}$$

در رابطه ۴، $fR_i(t)$ بازدهی پیش‌بینی شده سهم i در زمان t برای fh روز در آینده، $fp_i(t+fh)$ لگاریتم قیمت پیش‌بینی شده سهم i در زمان t برای fh روز در آینده و $p_i(t)$ قیمت واقعی سهم در زمان t است. در نظر بگیرید برای اینکه بازدهی سهم را محاسبه کنیم نیاز داریم بجای لگاریتم قیمت سهم، از خود قیمت استفاده کنیم. به همین منظور از عدد نپر به توان لگاریتم قیمت پیش‌بینی شده ($\exp(fp_i(t+fh))$) استفاده می‌کنیم تا لگاریتم قیمت تبدیل به قیمت شود.

تعداد عامل‌ها

هرچه تعداد عامل‌های تخمین زده شده به روش تحلیل اجزای اصلی بیشتر باشد، رابطه شماره ۲، بر رفتار حقیقی سهم بیشتر منطبق می‌شود و مقدار خطای آن کاهش می‌یابد. به طوریکه اگر تعداد این عامل‌ها با تعداد سهام (N) برابر شود، مدل تشکیل شده کاملا بر روی قیمت سهام منطبق می‌گردد. اما این کار در حالی که پیچیدگی مدل را افزایش می‌دهد، لزوما باعث بهبود پیش‌بینی انجام شده توسط مدل نمی‌شود. به همین علت تعیین تعداد مناسب عامل‌ها، از اهمیت بالایی برخوردار است. برای این منظور همانند قسمت قبل از معیار BIC استفاده می‌کنیم چرا که می‌خواهیم بین پیچیدگی و کارایی مدل تعادل برقرار شود. برای تخمین این معیار بازه زمانی در اختیار را به سه بخش تقسیم کرده و برای هر بخش با ساخت مدل‌های مختلف با تعداد عامل‌های ۵، ۱۰ و ۱۵ عامل و با ۰، ۱ و ۲ لگ مقدار این آماره را برای هر مدل محاسبه و بهترین مدل را انتخاب کردیم. همانطور که از جدول ۲ مشخص است، بهترین مدل که کمترین مقدار BIC را دارد، شامل ۱۵ عامل و صفر لگ است.

جدول ۲- مقدار معیار BIC برای مدل‌های مختلف

پنجره زمانی								
سوم			دوم			اول		
BIC	لگ مدل	تعداد عامل‌ها	BIC	لگ مدل	تعداد عامل‌ها	BIC	لگ مدل	تعداد عامل‌ها
-۱۰۱۲۳۴	۰	۵	-۴۶۹۸۲	۰	۵	-۴۶۷۶۴	۰	۵
-۱۳۰۹۶۲	۰	۱۰	-۷۳۱۶۷	۰	۱۰	-۶۳۱۹۳	۰	۱۰
-۱۴۸۴۷۲	۰	۱۵	-۹۳۳۶۴	۰	۱۵	-۷۴۰۸۷	۰	۱۵
-۹۹۸۱۱	۱	۵	-۴۷۰۳۹	۱	۵	-۴۵۸۶۹	۱	۵
-۱۲۷۳۶۰	۱	۱۰	-۷۰۶۷۰	۱	۱۰	-۶۱۴۱۹	۱	۱۰
-۱۴۲۴۸۳	۱	۱۵	-۸۹۲۵۰	۱	۱۵	-۷۰۹۰۰	۱	۱۵
-۹۸۱۲۸	۲	۵	-۴۶۴۹۳	۲	۵	-۴۴۸۷۵	۲	۵
-۱۲۳۳۴۰	۲	۱۰	-۶۸۰۲۳	۲	۱۰	-۵۹۴۵۹	۲	۱۰
-۱۳۶۳۲۰	۲	۱۵	-۸۴۹۷۹	۲	۱۵	-۶۷۷۴۲	۲	۱۵

۳- یافته‌های پژوهش

برای تست استراتژی طراحی شده، از داده‌های مربوط به سهام موجود در بورس تهران از ابتدای سال ۱۳۹۰ تا انتهای سال ۱۳۹۵ استفاده کردیم. به این صورت که از ابتدای دوره، یک پنجره زمانی متحرک با طول ew در نظر گرفته و با استفاده از مقادیر موجود در این پنجره، عامل‌ها و مدل‌های مورد نظر را تخمین می‌زنیم. در هر مرحله هیچ نگاه رو به جلویی فراتر از پنجره زمانی نداریم و با پیش‌بینی عامل‌ها برای fh روز در آینده، قیمت و سپس بازدهی سهام را پیش‌بینی می‌کنیم. در ادامه برای استفاده از این پیش‌بینی، سبدی شامل k درصد از سهامی که بیشترین بازده پیش‌بینی شده را دارند، تشکیل می‌دهیم. بعد از آن، پنجره زمانی را به fh روز در آینده منتقل کرده و با استفاده از داده‌های جدید، بازدهی حقیقی سبد را محاسبه و سهام موجود در سبد را به روز رسانی می‌کنیم و این کار را تا انتهای دوره مورد بررسی انجام می‌دهیم. در نظر داشته باشید که در طراحی این استراتژی از سه پارامتر استفاده شده است که تغییر در مقادیر آن‌ها می‌تواند بر روی عملکرد استراتژی تاثیر بگذارد. به همین دلیل برای تست نیرومندی[۲۸] استراتژی نسبت به تغییر این پارامترها، عملکرد استراتژی را با ۲۷ ترکیب مختلف از پارامترهای ew، fh و k بررسی می‌کنیم و برای هر استراتژی، میانگین بازده سالانه و نسبت شارپ را محاسبه می‌کنیم. بدین منظور سه مقدار ۱۰۰، ۲۰۰ و ۵۰۰ روز را برای پنجره زمانی تخمین، سه مقدار ۱، ۵ و ۱۰ روز را برای افق پیش‌بینی و سه مقدار ۱۰٪، ۲۰٪ و ۳۰٪ را برای پارامتر k در نظر می‌گیریم. بنابراین هر استراتژی را به صورت ew-fh-k نمایش می‌دهیم. در نهایت برای بررسی بهتر عملکرد این استراتژی، بازدهی آن را با استراتژی خرید و نگهداری[۲۹] از ابتدای دوره تا انتهای آن با وزن‌های برابر مقایسه می‌کنیم. در جدول ۳ می‌توانید مقادیر میانگین بازدهی سالانه[۳۰] مربوط به استراتژی‌های طراحی شده را مشاهده کنید.

جدول ۳- مقادیر میانگین بازدهی سالانه مربوط به استراتژی‌های مختلف

میانگین بازدهی سالانه به درصد	استراتژی	میانگین بازدهی سالانه به درصد	استراتژی	میانگین بازدهی سالانه به درصد	استراتژی
۶۶	۵۰۰-۱-۰,۱	۷۸	۲۰۰-۱-۰,۱	۱۱۵	۱۰۰-۱-۰,۱
۱۰۸	۵۰۰-۱-۰,۲	۸۹	۲۰۰-۱-۰,۲	۶۴	۱۰۰-۱-۰,۲
۸۰	۵۰۰-۱-۰,۳	۵۸	۲۰۰-۱-۰,۳	۵۲	۱۰۰-۱-۰,۳
۷۶	۵۰۰-۵-۰,۱	۳۴	۲۰۰-۵-۰,۱	۵۰	۱۰۰-۵-۰,۱
۸۷	۵۰۰-۵-۰,۲	۵۰	۲۰۰-۵-۰,۲	۴۹	۱۰۰-۵-۰,۲
۴۸	۵۰۰-۵-۰,۳	۴۹	۲۰۰-۵-۰,۳	۵۱	۱۰۰-۵-۰,۳
۴۵	۵۰۰-۱۰-۰,۱	۳۸	۲۰۰-۱۰-۰,۱	۴۵	۱۰۰-۱۰-۰,۱
۵۲	۵۰۰-۱۰-۰,۲	۴۱	۲۰۰-۱۰-۰,۲	۳۶	۱۰۰-۱۰-۰,۲
۴۹	۵۰۰-۱۰-۰,۳	۵۳	۲۰۰-۱۰-۰,۳	۴۱	۱۰۰-۱۰-۰,۳

همانطور از جدول 3 مشخص است بیشترین متوسط بازدهی سالانه مربوط به استراتژی با پارامتر های ew=۱۰۰ ، fh=۱ و k=٪۱۰ برابر با مقدار سالانه ٪۱۱۵ است.

در ادامه برای این منظور که علاوه بر بازدهی حاصل از هر استراتژی، میزان ریسک موجود در آن‌ها را نیز در نظر بگیریم، از معیار نسبت شارپ[31] استفاده می‌کنیم. این معیار را به صورت زیر محاسبه می‌کنیم:

$$SR_i = \frac{\overline{R}_i}{\sigma(R_i)} \quad \text{(رابطه ۵)}$$

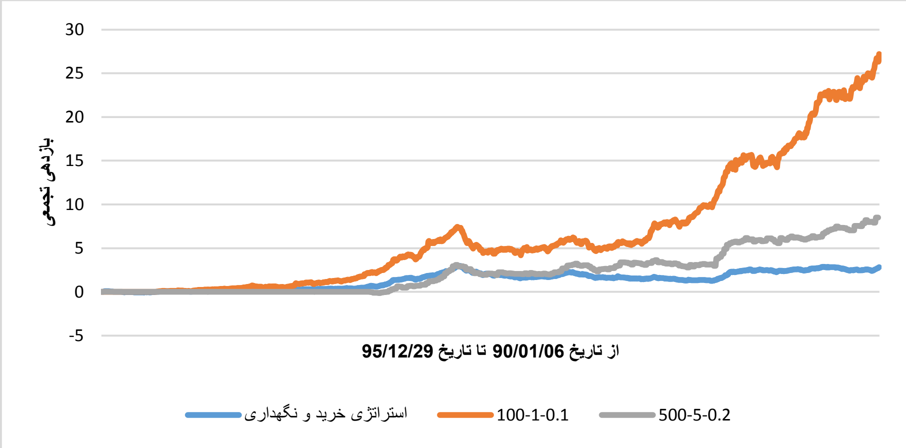
که $\overline{R}$ متوسط بازدهی های کسب شده توسط هر استراتژی و $\sigma(R_i)$ انحراف معیار این بازدهی‌ها است. مقدار این معیار را برای پارامترهای مختلف می‌توان در جدول 4 مشاهده کرد.

جدول ۴- مقادیر نسبت شارپ برای استراتژی‌های مختلف

نسبت شارپ	استراتژی	نسبت شارپ	استراتژی	نسبت شارپ	استراتژی
۰,۰۹۶	۵۰۰-۱-۰,۱	۰,۱۴۷	۲۰۰-۱-۰,۱	۰,۱۷۷	۱۰۰-۱-۰,۱
۰,۱۵۵	۵۰۰-۱-۰,۲	۰,۱۵۷	۲۰۰-۱-۰,۲	۰,۱۷۵	۱۰۰-۱-۰,۲
۰,۱۳۳	۵۰۰-۱-۰,۳	۰,۱۴۷	۲۰۰-۱-۰,۳	۰,۱۶۳	۱۰۰-۱-۰,۳
۰,۲۱۵	۵۰۰-۵-۰,۱	۰,۱۳۸	۲۰۰-۵-۰,۱	۰,۱۹۷	۱۰۰-۵-۰,۱
۰,۲۷۹	۵۰۰-۵-۰,۲	۰,۱۸۶	۲۰۰-۵-۰,۲	۰,۲۰۲	۱۰۰-۵-۰,۲
۰,۲۳۹	۵۰۰-۵-۰,۳	۰,۲۰۲	۲۰۰-۵-۰,۳	۰,۲۱۸	۱۰۰-۵-۰,۳
۰,۱۹۶	۵۰۰-۱۰-۰,۱	۰,۲۰۸	۲۰۰-۱۰-۰,۱	۰,۲۰۷	۱۰۰-۱۰-۰,۱
۰,۲۴۱	۵۰۰-۱۰-۰,۲	۰,۲۰۸	۲۰۰-۱۰-۰,۲	۰,۱۹۴	۱۰۰-۱۰-۰,۲
۰,۲۳۷	۵۰۰-۱۰-۰,۳	۰,۲۶۱	۲۰۰-۱۰-۰,۳	۰,۲۱۳	۱۰۰-۱۰-۰,۳

همانطور که در جدول ۴ مشاهده می‌کنید بهترین مقدار برای این معیار ۰,۲۷۹ است که با استراتژی با پارامترهای ew=۵۰۰ ، fh=۵ و k=٪۲۰ بدست می‌آید. اگرچه این استراتژی نسبت به استراتژی منتخب در حالت قبل بازدهی کل و متوسط بازدهی سالانه پایین‌تری دارد، اما مقدار ریسک محاسبه شده برای آن از استراتژی حالت قبلی کمتر است و این موضوع باعث شده که از نظر معیار نسبت شارپ، استراتژی مناسب‌تری باشد.

برای درک بهتر عملکرد این دو استراتژی، آنها را با استراتژی ساده خرید و نگهداری[32] سهام از ابتدا تا انتهای دوره مقایسه می‌کنیم. این استراتژی به این صورت است که در ابتدای دوره مورد بررسی، تمامی سهام موجود در بازار را با وزن‌های برابر خریداری می‌کنیم و در انتهای دوره ۶ ساله بازدهی حاصل از این استراتژی را محاسبه می‌کنیم. برای درک بهتر و مقایسه کامل‌تر بین دو استراتژی منتخب و استراتژی خرید و نگهداری، نمودار بازدهی تجمعی سالانه این سه استراتژی را می‌توانید در شکل ۱ ببینید.



شکل ۱- بازدهی تجمعی استراتژی‌های برتر در مقایسه با استراتژی خرید و نگهداری

همانطور که در شکل ۱ مشاهده می‌کنید، عملکرد استراتژی ۰,۱-۱-۱۰۰ در تمام روزها بهتر از استراتژی خرید و نگهداری سهام بوده است. استراتژی ۰,۲-۵-۵۰۰ در ابتدا عملکردی ضعیف تر از استراتژی خرید و نگهداری دارد، که می‌توان این ضعف را تا حدودی به شروع به معامله دیرتر این استراتژی به علت نیاز به اطلاعات ۵۰۰ روز گذشته برای تصمیم‌گیری نسبت داد، اما در ادامه بازدهی بیشتری از این استراتژی کسب می‌کند.

۴- نتیجه‌گیری و بحث

در این مقاله ما رویکرد جدیدی برای مطالعه رفتار سهام و طراحی استراتژی‌های معاملاتی ارائه کردیم. در این رویکرد پیشنهادی، بجای استفاده مستقیم از قیمت سهام به دنبال کشف روابط پنهان بین آن‌ها به صورت الگوهای حرکتی هستیم. به عنوان مثال اولین الگوی استخراج شده که به عامل اصلی اول معروف است، رفتاری شبیه به شاخص کل بازار از خود نشان می‌دهد. بنابراین با شناسایی این الگوها، می‌توان به علت‌هایی که منجر به حرکت قیمت سهام می‌شوند و روابط بین آن‌ها پی برد و با پیش‌بینی این الگوها، آینده سهام را پیش‌بینی کرد.

نتایج تجربی حاصل از این مقاله نشان‌دهنده سودده بودن استراتژی‌های آربیتراژ آماری بر مبنای مدل‌های عاملی قیمت سهام است و این مدل‌ها می‌توانند چهارچوبی برای پیاده سازی استراتژی‌های آربیتراژ آماری باشند. استراتژی‌های ساخته شده در این مقاله بدون در نظر گرفتن هزینه معاملاتی، بازدهی‌های نسبتا بالایی دارند. به طوریکه استراتژی منتخب توانست بازدهی متوسط سالیانه ۱۱۵٪ را فراهم کند که این بازدهی از استراتژی خرید و نگهداری بلند مدت سهام بسیار بیشتر بود. هزینه معاملاتی در کشور ما نسبتا مقدار بالایی است و برای هر معامله تقریبا برابر با ۱,۵ درصد است. از آنجایی که این استراتژی بر مبنای تعداد معاملات بالا و دوره نگهداری کوتاه با هدف کسب بازدهی‌های اندک عمل می‌کند، در صورت کاهش هزینه معاملاتی سازمان بورس اوراق

بهادار تهران، می‌تواند چهارچوب مناسبی برای طراحی و پیاده‌سازی استراتژی‌های معاملاتی برای کسب سود از بازار باشد.

فهرست منابع

* احمدزاده، عزیز؛ یاوری، کاظم و صالح آبادی، علی (۱۳۹۳). آزمون آربیتراژ آماری در بورس اوراق بهادار تهران. فصل‌نامه پژوهش‌ها و سیاست‌های اقتصادی. شماره ۷۰: ۲۶۸-۲۴۷
* Alexander, C., (1999). Optimal hedging using cointegration. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 357(1758):2039-2058.
* Alexander, C. and Dimitriu, A., (2002). The cointegration alpha: Enhanced index tracking and long-short equity market neutral strategies.
* Avellaneda, M. and Lee, J.H., (2010). Statistical arbitrage in the US equities market. Quantitative Finance, 10(7):761-782.
* Barhoumi, K., Darné, O. and Ferrara, L., (2013). Dynamic factor models: A review of the literature.
* Bertram, W.K., (2010). Analytic solutions for optimal statistical arbitrage trading. Physica A: Statistical Mechanics and its Applications, 389(11):2234-2243.
* Black, F., (1972). Capital market equilibrium with restricted borrowing. The journal of business, 45(3):444-455.
* Eklund, J. and Kapetanios, G., (2008). A review of forecasting techniques for large data sets. National Institute Economic Review, 203(1):109-115.
* Focardi, S.M., Fabozzi, F.J. and Mitov, I.K., (2016). A new approach to statistical arbitrage: Strategies based on dynamic factor models of prices and their performance. Journal of Banking & Finance, 65:134-155.
* Geweke, J., (1977). The dynamic factor analysis of economic time series. Latent variables in socio-economic models.
* Hogan, S., Jarrow, R., Teo, M. and Warachka, M., (2004). Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. Journal of Financial economics, 73(3): 525-565.
* Jarrow, R.A., Teo, M., Tse, Y.K. and Warachka, M., (2005). Statistical arbitrage and market efficiency: Enhanced theory, robust tests and further applications.
* Jarrow, R., Teo, M., Tse, T.Y., Warachka, M., (2012). An improved test for statistical arbitrage. Journal of Financial Markets 15:47–80.
* Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J.P., (2000). Random matrix theory and financial correlations. International Journal of Theoretical and Applied Finance, 3(03):391-397.
* Lintner, J., (1965). Security prices, risk, and maximal gains from diversification. The journal of finance, 20(4):587-615.
* Pena, D. and Box, G.E., (1987). Identifying a simplifying structure in time series. Journal of the American Statistical Association, 82(399): 836-843.
* Pole, A.,(2007) Statistical Arbitrage: Algorithmic Trading Insights and Techniques.
* Sargent, T.J. and Sims, C.A., (1977). Business cycle modeling without pretending to have too much a priori economic theory. New methods in business cycle research, 1:145-168.
* Sharpe, W.F., (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The journal of finance, 19(3):425-442.

* Vidyamurthy, G., (2004). Pairs Trading: quantitative methods and analysis (Vol. 217). John Wiley & Sons

یادداشت‌ها

[1] Morgan Stanley
[2] Pair Trading
[3] Avellaneda & Lee
[4] Principal Components
[5] Buy and Hold
[6] Vidyamurthy
[7] Pol
[8] Gatev et al
[9] Sharp
[10] Lintner
[11] Black
[12] Hogan et al
[13] Jarrow et al
[14] Avellaneda & Lee
[15] High Frequency Trading Desks
[16] Bertram
[17] Focardi et al
[18] Principal Component Analysis
[19] Geweke
[20] Sargent & Sims
[21] Pena & Box
[22] ARMA
[23] Eklund & Kapetanios
[24] Barhoumi et al
[25] Focardi et al
[26] Eigen values And Vectors
[27] Maximum Likelihood
[28] Robustness
[29] Buy And Hold
[30] Average Annual Return
[31] Sharp Ratio
[32] Buy and Hold