Financial Research Journal

Print ISSN: 1024-8153
Online ISSN: 2423-5377

Paired Trading Strategy Optimization Using the Reinforcement Learning Method: Intraday Data of Tehran Stock Exchange

Saeid Fallahpour
Assistant Prof., Department of Financial Management, Faculty of Management, University of Tehran, Tehran, Iran. E-mail: falahpor@ut.ac.ir

Hasan Hakimian
*Corresponding author, MSc. Student, Department of Financial Engineering, Faculty of Management, University of Tehran, Tehran, Iran. E-mail: hasan.hakimian@ut.ac.ir

Abstract

Objective: Paired trading is among the most well-known and oldest algorithmic trading systems. The efficiency and profitability of this system have been demonstrated in many studies conducted so far in financial markets. Paired trading is principally based on long-run equilibrium relationships or reverting to the mean characteristic**.** In recent years, a large number of studies have been conducted on algorithmic trading using machine learning.

Methods: In this research, the reinforcement learning method - an appropriate method for modeling and optimizing problems involving different long-run relationships - was used in order to select appropriate trading thresholds and time windows for the purpose of maximizing efficiency and minimizing negative risks in paired trading through adopting the co-integration approach. Results are obtained by applying a combination of reinforcement learning method and co-integration approach in paired trading.

Results: Empirical results based on the intraday data of paired stocks showed that the reinforcement learning method used to design trading systems in paired trading had significant advantages over the other methods in previous works.

Conclusion: A pair trading strategy with the proposed algorithm can be used as a neutral market strategy in all market conditions, including prosperity and recession, by investors and individual and institutional traders. Also, for future research, it is possible to consider transaction costs in a pair trading strategy.

Keywords: Co-integration, Mean-Reverting Process, Pairs Trading, Reinforcement Learning, Sortino Ratio.

Citation: Fallahpour, S., & Hakimian, H. (2019). Paired Trading Strategy Optimization Using the Reinforcement Learning Method: Intraday Data of Tehran Stock Exchange. *Financial Research Journal*, 21(1), 19- 34. (*in Persian*)

Financial Research Journal, 2019, Vol. 21, No.1, pp. 19- 34
DOI: 10.22059/frj.2018.138913.1006099
Received: December 19, 2015; Accepted: February 09, 2018

فصلنامه تحقیقات مالی

شاپا چـــاپـــی: ۱۰۲۴-۸۱۵۳
شاپا الکترونیکی: ۲۴۲۳-۵۳۷۷

بهینه‌سازی استراتژی معاملات زوجی با استفاده از روش یادگیری تقویتی، با به‌کارگیری دیتاهای درون‌روزی در بورس اوراق بهادار تهران

سعید فلاح‌پور
استادیار، گروه مالی و بیمه، دانشکده مدیریت دانشگاه تهران، تهران، ایران. رایانامه: falahpor@ut.ac.ir

حسن حکیمیان
* نویسنده مسئول، دانشجوی کارشناسی ارشد، گروه مهندسی مـالی، دانشـکده مـدیریت، دانشـگاه تهـران، تهـران، ایـران. رایانامـه: hasan.hakimian@ut.ac.ir

چکیده

هدف: معاملات زوجی از معروف‌ترین و قدیمی‌ترین سیستم‌های معاملات الگوریتمی است که کارایی و سـودآوری آن در بسـیاری از پژوهش‌هایی که تاکنون در بازارهای مالی مختلف صورت گرفته است، اثبات و نشان داده شده اسـت. مهـم‌تـرین اصـل در معـاملات زوجی، وجود روابط تعادلی بلندمدت یا همان خاصیت بازگشت به میانگین است. از طرفی در سال‌های اخیر تحقیقـات شـایان تـوجهی روی معاملات الگوریتمی با استفاده از یادگیری ماشین صورت گرفته است.

روش: در این پژوهش از روش یادگیری تقویتی که برای مدل‌سازی و بهینه‌سازی مسائل با انواع مختلـف روابـط بلندمـدت مناسـب است، به‌منظور انتخاب آستانه‌های معاملاتی و پنجره‌های زمانی مناسب با هدف ماکزیمم‌سازی بازده و مینیمم‌سازی ریسک‌های منفی در معاملات زوجی با رویکرد هم‌انباشتگی استفاده شده است. پژوهش حاضر با به‌کارگیری ترکیبی از روش یادگیری تقویتی و رویکـرد هم‌انباشتگی در معاملات زوجی اجرا شده است.

یافته‌ها: نتایج آزمایش روی دیتاهای درون‌روزی زوج سهام منتخب، نشان می‌دهد که استفاده از روش یادگیری تقـویتی در طراحـی سیستم معاملات در معاملات زوجی نسبت به کارهای قبلی انجام‌شده، برتری چشمگیری دارد.

نتیجه‌گیری: استراتژی معاملات زوجی با الگوریتم پیشنهادی می‌تواند به‌عنوان استراتژی بازار خنثی در تمامی شـرایط بـازار اعـم از رونق و رکود توسط سرمایه‌گذاران و معامله‌گران حقیقی و حقوقی استفاده شود. همچنین می‌توان در نظر گرفتن هزینه‌های معـاملاتی در انجام معاملات در استراتژی معاملات زوجی را به‌عنوان موضعی برای پژوهش‌های آتی پیشنهاد کرد.

کلیدواژه‌ها: معاملات زوجی، یادگیری تقویتی، هم‌انباشتگی، نسبت سورتینو، فرایند بازگشت به میانگین.

استناد: فلاح‌پور، سعید؛ حکیمیان، حسن (۱۳۹۸). بهینه‌سازی استراتژی معاملات زوجـی بـا اسـتفاده از روش یـادگیری تقـویتی، بـا به‌کارگیری دیتاهای درون‌روزی در بورس اوراق بهادار تهران. *تحقیقات مالی*، ۲۱(۱)، ۱۹-۳۴.

تحقیقات مالی، ۱۳۹۸، دوره ۲۱، شماره ۱، صص. ۱۹-۳۴
DOI: 10.22059/frj.2018.138913.1006099
دریافت: ۱۳۹۴/۰۹/۲۸، پذیرش: ۱۳۹۶/۱۱/۲۰

مقدمه

معاملات زوجی[1] یک استراتژی سرمایه‌گذاری بازار خنثی[2] است که معامله‌گر را قادر می‌سازد در هر وضعیتی از بازار مانند نزولی، صعودی یا حرکت‌های مجانبی و حتی در دوره‌هایی با نوسان‌های بالا یا پایین، سود کسب کند (ویدیامورتی[3]، ۲۰۰۴). از آنجا که این استراتژی هم‌زمان خرید و فروش را روی دارایی‌های معادل لحاظ می‌کند و از مزایای تفاوت قیمتی بین آنها بهره برده و سود کسب می‌کند، در گروه استراتژی‌های آربیتراژ آماری قرار می‌گیرد.

پیدایش معاملات زوجی عموماً به کار گروهی از دانشمندان علوم رایانه، ریاضیات و فیزیک که در اوایل دهه ۱۹۸۰ میلادی در شرکت مورگان استنلی گرد هم آمده بودند، نسبت داده می‌شود. مطابق گاتو، گوتزمن و روئن‌هورست[4] (۲۰۰۶)، در معاملات زوجی، معامله‌گر هنگام مشاهده یک ضعف (گپ) در روابط همبستگی (تعادلی بلندمدت)[5] بین دو دارایی که به‌معنای انحراف قیمت دو دارایی از مقادیر تعادلی آنهاست، با ایجاد پرتفویی از دو دارایی یا یک «جفت»، به گرفتن موقعیت خرید در دارایی ارزان‌تر و گرفتن موقعیت فروش در آن سهمی که گران‌تر شده است، اقدام می‌کند. طبق مفهوم «بازگشت به میانگین»[6] که توسط هیلربراند[7] (۲۰۰۳) مطرح شد، هنگام بازگشت دو دارایی به تعادل بلندمدت خود با اتخاذ موقعیت‌های معاملاتی معکوس معامله را تکمیل می‌کنیم. سود این استراتژی صرف‌نظر از حرکت بازار، از تفاوت بین تغییرات قیمت در دو دارایی منتج می‌شود (ویدیامورتی، ۲۰۰۴).

در استراتژی‌های معاملات زوجی چهار رویکرد شناخته‌شده وجود دارد: رویکرد فاصله‌ای، رویکرد پیش‌بینی ترکیبی، رویکرد تفاضل قیمتی تصادفی و رویکرد هم‌انباشتگی. در این پژوهش از رویکرد هم‌انباشتگی استفاده شده است. هم‌انباشتگی در واقع مفهومی عمومی از ارتباطی مانا[8] بین متغیرهای ناماناست. در هم‌انباشتگی روابط اقتصادی بلندمدت بین دارایی‌ها برآورد و تحلیل می‌شود.

به‌طور کلی نخستین گام در اجرای این استراتژی، انتخاب یک زوج دارایی است که روابط آماری بلندمدتی دارند. پس از انتخاب زوج سهام از طریق بررسی وجود روابط مربوطه (روابط هم‌انباشتگی)، پارامترهای مدل تخمین زده می‌شود و در نهایت، در مرحله طراحی معاملات مقدار استاندارد شده تفاضل قیمتی[9] که حاصل از مابه‌التفاوت بین قیمت‌های دو دارایی است، برای انجام معاملات و اتخاذ موقعیت‌های مناسب ترسیم می‌شود. با توجه به خاصیت بازگشت به میانگین و روابط تعادلی بلندمدت، مقدار تفاضل قیمتی در دو طرف مقدار میانگین تغییر خواهد کرد. حال اینکه در چه حدی از مقدار تفاضل قیمتی باید موقعیت معاملاتی را باز کرد یا اینکه چه حدی را باید به‌عنوان حد ضرر[10] منظور کرد، به‌صورت دو نوع آستانه در دو طرف میانگین تفاضل قیمتی مشخص می‌شود. فاصله آستانه‌ها از میانگین تفاضل قیمتی، سود هر معامله و

1. Pairs trading
2. Neutral market
3. Vidyamurthy
4. Gatev, Goetzmann and Rouwenhorst
5. Long run equilibrium
6. Mean reversion
7. Hillebrand
8. Stationary
9. Spread
10. Stoploss

مدت زمان انجام هر معامله (بازکردن تا بستن موقعیت‌ها) را تعیین می‌کند. به‌طور کلی انتخاب پنجره‌های زمانی[1] و آستانه‌های[2] مناسب، نقش چشمگیری در میزان سودآوری این استراتژی دارد.

در این پژوهش پس از تشکیل پرتفویی از دو سهام کاندید، به‌منظور ارزیابی عملکرد این پرتفو با توجه به الگوریتم پیشنهادشده، از نسبت سورتینو به‌عنوان تابع هدف استفاده شده است. در واقع، به‌دنبال ماکزیمم‌سازی تابع هدف یعنی نسبت سورتینو با در نظرگرفتن یک سری محدودیت‌های غیرخطی و نهایتاً دستیابی به مدل بهینه هستیم. نسبت سورتینو مشابه نسبت شارپ، به محاسبهٔ بازده مورد انتظار به ازای هر واحد تغییر در ریسک سرمایه‌گذاری می‌پردازد با این تفاوت که نسبت شارپ تغییرپذیری بازده را بررسی می‌کند، در حالی که نسبت سورتینو تنها تغییرپذیری نامطلوب را مبنای ارزیابی قرار می‌دهد.

با گسترش سامانه‌های هوشمند، استفاده از ماشین‌هایی که همانند انسان قابلیت فراگیری داشته و می‌توانند با اکتشاف محیط، تجربه‌های خود را افزایش دهند، فراگیر شده است. یکی از این روش‌ها به یادگیری تقویتی[3] موسوم است. یادگیری تقویتی که از رفتار روان‌شناسی الهام می‌گیرد، ناحیه‌ای از یادگیری ماشین[4] محسوب می‌شود. ایده اصلی پشت یادگیری تقویتی، در واقع نوعی استراتژی جریمه ـ پاداش است. یادگیری تقویتی از دو مؤلفه اصلی عامل[5] و محیط[6] تشکیل شده است. عامل در محیط زندگی کرده و با توجه به بازخوردی که از محیط می‌گیرد، تجربه‌های خود را به‌روزرسانی می‌کند.

در این پژوهش از روش یادگیری تقویتی به‌منظور انتخاب بهینه پنجره تخمین، پنجره معاملاتی، آستانه‌های معاملاتی و حد ضرر با هدف ماکزیمم‌سازی نسبت سورتینو استفاده شده است. به بیان دیگر، استراتژی معاملات زوجی در قالب یکی از مسائل معروف در زمینه یادگیری تقویتی، یعنی مسئله N-arm bandit مدل شده است. نتایج آزمایش‌ها از قدرت شایان توجه این الگوریتم در بهبود بازده و کاهش هم‌زمان ریسک‌های منفی استراتژی نسبت به حالت‌های قبلی حکایت می‌کند.

پیشینه پژوهش

معاملات زوجی و یادگیری تقویتی دو ناحیه اصلی در پژوهش‌های صورت گرفته است که با پژوهش حاضر نیز ارتباط دارد. روش‌های کمی مختلفی برای توسعه و به‌کارگیری استراتژی معاملات زوجی در ادبیات مطرح شده است. همه این رویکردها تنها در اندازه‌گیری تفاضل قیمتی و شدت بازگشت به میانگین تفاوت دارند، با این حال ایده اصلی ورود به یک موقعیت، زمانی است که انحراف از میانگین تفاضل قیمتی به حد کافی قوی باشد، سپس بستن موقعیت زمانی که تفاضل قیمتی نزدیک به میانگینش است، رفتار مشترک همه رویکردهاست.

بسیاری از پژوهش‌هایی که تاکنون در زمینه معاملات زوجی صورت گرفته‌اند، بر کارایی و عملکرد استراتژی

1. Time windows
2. Thersholds
3. Reinforcement learning (RL)
4. Machine learning
5. Agent
6. Environment

معاملات زوجی در بازارهای خارجی متمرکز شده‌اند که از این بین می‌توان بـه گـاتو، گـوئتزمن و رون‌هورسـت[1] (۲۰۰۶) اشاره کرد. آنها برای آزمون استراتژی معاملات زوجی، از اطلاعات روزانه بازده سهام در بازه ۲۰۰۲– ۱۹۶۲ استفاده کردند و با یک قانون معاملاتی ساده، بازده اضافی سالیانه بالای ۱۱ درصد را برای هر سال از کل دوره نمونه به‌دست آوردند.

سالیان زیادی است که در زمینه طراحی قوانین و استراتژی‌های معاملاتی در سیستم‌های معاملات الگوریتمی بحث و تحقیق می‌شود. از جمله پژوهش‌هایی که در زمینه طراحی استراتژی در معاملات الگوریتمی و در ارتباط با پارامترهـای مربوط به آن (مانند انتخاب حدود آستانه‌های معاملاتی، محل به‌کارگیری حد ضرر و نحوه محاسبه بازده) صـورت گرفتـه است، می‌توان به ژانگ[2] (۲۰۰۱) اشاره کرد. وی در بررسی خود یک قانون فروش را بـر اسـاس دو سـطح آسـتانه، یـک قیمت هدف و یک حد ضرر، تعیین کرد و به‌منظور دستیابی به سطوح آستانه بهینه، به حل مجموعه‌ای از مسائل با مقادیر دو سطح آستانه پرداخت. ژو و ژانگ[3] (۲۰۰۵) تحت یک مدل با به‌کارگیری حرکات براونی به مطالعه قانون فروش بهینه پرداختند. آنها با حل مجموعه‌ای از معادلات جبری و استفاده از تکنیک هموارسازی، سطوح آستانه بهینه را تعیین کردند. دای، ژانگ و ژو[4] (۲۰۱۰) بر مبنای یک اندیکاتور احتمال شرطی به تعیین یـک قـانون پیـروی رونـد پرداختنـد و بـرای نشان‌دادن قانون معاملاتی بهینه به‌صورت دو منحنی آستانه، معادلات HJB[5] را حل کردند.

در ارتباط با سیستم معاملات زوجی، برترام[6] (۲۰۱۰) برای یک دارایی ترکیبی که از فرایند OU پیروی می‌کنـد، بـا در نظر گرفتن زمان و یک فرمول تحلیلی، به انتخاب باندهای مناسب پرداخت. وی مشاهده کرد که برای ماکزیمم‌سازی بازده در هر واحد زمانی و ماکزیمم‌سازی نسبت شارپ، باندهای بهینه به‌صورت متقارن اطراف میـانگین قـرار مـی‌گیـرد. ژنگ و لی[7] (۲۰۱۴) دربارهٔ تأثیر انتخاب باندهای باریک و پهن در میزان بازده و مدت زمان اجرای هر معامله بحث کردند و به بررسی باندهای بهینه در قالب تابعی از هزینه معاملات و پارامترهای فرایند OU با هدف مـاکزیمم‌سـازی میـانگین سود مورد انتظار بلندمدت پرداختند. آنها با پیاده‌سازی روش خود روی اطلاعات روزانـه سـهام شـرکت‌هـای کوکـاکولا و پپسی، مشاهده کردند که استراتژی جدید پیشنهادشده آنها عملکرد بهتری از تمرین‌های معمول دارد.

یادگیری تقویتی نوعی رویکرد محاسباتی اسـت کـه بـه‌صـورت خودکـار و بـا الهـام از محـیط، فهـم، یـادگیری و تصمیم‌گیری در جهت هدف را انجام می‌دهد. عامل طراحی‌شده با استفاده از یادگیری تقویتی، در حقیقت به جـای اینکـه به او گفته شود چه عملی برای انجام دادن مد نظر است، سعی می‌کند عمل مد نظر را که به حداکثر کـردن سـود منجـر می‌شود، با تجربه به‌دست آورد. این عامل می‌تواند سیاست‌های مختلف را برای انتخاب آن عمل در مواقع تصمیم‌گیـری، در نظر بگیرد. از جمله سیاست‌های متداول برای انتخاب عمل، سیاست $\epsilon - greedy$ است که در ϵ از موارد، عملـی را برمی‌گزیند که دارای بیشترین ارزش از نظر عامل بوده و در $1 - \epsilon$ موارد عملی را به‌صورت تصادفی و مستقل از میزان ارزش آن انتخاب می‌کند (ساتون و بارتو[8]، ۱۹۹۸).

1. Gatev, Goetzmann, and Rouwenhorst
2. Zhang
3. Gue and Zhang
4. Dai, Zhang and Zhu
5. Associated Hamilton-Jacobi-Bellman
6. Bertram
7. Zeng, & Lee
8. Sutton, & Barto

با توجه به قدرت یادگیری و اینکه الگوریتم یادگیری تقویتی قادر است یکسری از متغیرها را که هیچ شناخت و پیش تعریفی از محیط و ساختار حاکم بر آنها در دست نیست، به‌صورت ریاضی فرمول‌بندی کند، از آن به‌منظور مدل‌سازی و فرمول‌بندی بسیاری از مسائل در شاخه‌های مختلف علوم استفاده می‌شود. در مسائل و تحلیل‌های حوزه مالی نیز عدم قطعیت و پویایی از اجزای لاینفک آن محسوب شده و به‌کارگیری این الگوریتم می‌تواند بسیار مفید باشد (ساتون و بارتو، ۱۹۹۸).

ژاو و چان[1] (۲۰۰۰) ترکیبی از الگوریتم‌های Q-learning و ماکزیمم‌سازی نسبت شارپ برای معاملات و مدیریت پرتفولیو پیشنهاد کردند. آنها با به‌کارگیری سود مطلق و نسبت شارپ به‌عنوان تابع عملکرد به آزمون مدل خود پرداختند و در الگوریتم خود از مزایای هر دو قسمت بهره بردند. آنها سودمندی الگوریتم خود را با معاملات در بازارهای سهام خارجی نشان دادند. لی، پارک، او و هونگ[2] (۲۰۰۷) با هدف افزایش بیشتر عملکرد سیستم‌های برمبنای یادگیری تقویتی، چارچوبی برای معاملات سهام ارائه کردند. رویکرد پیشنهادی آنها با به‌کارگیری مزایای استفاده از چندین عامل Q-learning به‌وسیله تعریف نقش‌های لازم برای انجام توأم تصمیمات انتخاب و قیمت‌گذاری سهام، به انجام معاملات سهام پرداختند. وون لی[3] (۲۰۰۱) از یادگیری تقویتی برای مدل‌سازی و یادگیری انواع مختلف فعل و انفعالات در شرایط واقعی برای مسئله پیش‌بینی قیمت سهام استفاده کرد. وی مسئله پیش‌بینی قیمت سهام را به‌عنوان فرایند مارکوف که می‌توان آن را بر مبنای الگوریتم یادگیری تقویتی بهینه کرد، در نظر گرفت. وی در بررسی خود از روندهای قیمت سهام و تغییرات قیمتی پی‌درپی برای بیان حالت در یادگیری تقویتی استفاده کرد. مودی و سافل[4] (۲۰۰۱) روشی برای بهینه‌سازی پرتفولیو و سیستم‌های معاملاتی برمبنای یادگیری تقویتی مستقیم پیشنهاد دادند و به‌منظور کشف سیاست‌های سرمایه‌گذاری، الگوریتمی انطباقی که RRL[5] نامیده می‌شود را ارائه کردند. آنها در پژوهش خود نسبت شارپ و انحرافات منفی را برای بهینه‌سازی پرتفولیو به‌کار گرفتند. با توجه به تحقیقات صورت گرفته در زمینه طراحی استراتژی معاملات زوجی و همچنین کاربرد یادگیری ماشین به‌منظور بهینه‌سازی پرتفو و انجام معاملات الگوریتمی، در پژوهش حاضر از الگوریتم یادگیری تقویتی برای بهینه‌سازی استراتژی معاملات زوجی و انتخاب بهینه چهار پارامتر مؤثر در طراحی استراتژی معاملاتی استفاده شده است.

روش‌شناسی پژوهش

هم‌انباشتگی و معادلات تصحیح خطا

هم‌انباشتگی نخستین بار از مقاله اولیه گرنجر[6] (۱۹۸۱) به‌عنوان ابزاری استاندارد در روش‌های آماری و به‌منظور تحلیل مسائل اقتصادی مطرح شد. هم‌انباشتگی در واقع مفهومی عمومی از ارتباطی مانا بین متغیرهای نامانا‌ست. روش‌های متعددی برای آزمون هم‌انباشتگی به‌منظور انتخاب زوج مناسب در انجام معاملات زوجی وجود دارد که از این بین

1. Gao, & Chan
2. Lee, Park, Lee, & Hong
3. Won Lee
4. Moody and Saffell
5. Recurrent Reinforcement Learning
6. Granger

می‌توان به روش انگل ـ گرنجر[1] (۱۹۸۷) و روش یوهانسون[2] (۱۹۸۸) اشاره کرد. در این پژوهش از تست یوهانسون برای بررسی هم‌انباشتگی و شناسایی زوج دارایی استفاده شده است. به‌طور ساده می‌توان گفت روش یوهانسون (۱۹۸۸) در واقع تعمیم آزمون دیکی ـ فولر[3] به حالت چند متغیر است که از روش یک مرحله‌ای مبتنی بر رابطه بین رتبه ماتریس و ریشه‌های مشخصه آن به‌منظور بررسی هم‌انباشتگی استفاده می‌کند.

استراتژی معاملات زوجی با رویکرد هم‌انباشتگی

همان‌طور که بیان شد، در هم‌انباشتگی روابط اقتصادی بلندمدت بین دارایی‌ها برآورد و تحلیل می‌شود. در واقع ایده اصلی در تجزیه و تحلیل هم‌انباشتگی آن است که بسیاری از سری‌های زمانی اقتصادی نامانا بوده و روند تصادفی افزایشی یا کاهشی دارند، اما ممکن است در بلندمدت یک ترکیب خطی از این متغیرها، همواره مانا بوده و بدون روند تصادفی باشد. در این قسمت استراتژی معاملات زوجی با رویکرد هم‌انباشتگی در قالب سه قسمت اصلی ارائه می‌شود.

نخستین گام در اجرای استراتژی معاملات زوجی، انتخاب زوج سهام مناسب است. در این پژوهش با اعمال فیلترهایی بر اساس تحلیل‌های بنیادی و اطلاعات از روندهای گذشته، به انتخاب مناسب‌ترین سهام برای اجرای استراتژی اقدام شده است. از آنجا که احتمال وجود رابطه هم‌انباشتگی بین دو سهم از یک صنعت مشترک بیشتر است، در این پژوهش از بین سهام موجود در یک صنعت و با اعمال معیارهایی مانند میزان نقدشوندگی، میزان سرعت معاملات و حجم معاملات، به انتخاب زوج سهام کاندید پرداخته شده است. در پژوهش حاضر از اطلاعات قیمتی زوج سهام شرکت‌های ایران خودرو ـ سایپا، پالایش نفت لاوان ـ پالایش نفت تهران، بانک ملت ـ بانک صادرات، گسترش نفت و گاز پارسیان ـ پتروشیمی شپدیس و بانک صادرات ـ بانک انصار استفاده شده است.

پس از انتخاب زوج سهام کاندید در مرحله قبل، معادلات تصحیح خطای برداری به‌منظور تخمین پارامترها و معادله تفاضل قیمتی، تخمین زده می‌شود. بنا بر قضیه گرنجر، متناظر با هر رابطه بلندمدت اقتصادی باید رابطه کوتاه‌مدتی به‌صورت سازوکار تصحیح خطا[4] برای حصول به تعادل بلندمدت وجود داشته باشد؛ چرا که ECM چگونگی تعدیل متغیرهای دستگاه را در کوتاه‌مدت برای حصول به رابطه تعادلی بلندمدت نشان می‌دهد. در واقع اگر سازوکاری وجود نداشته باشد که متغیرها نسبت به عدم تعادل از رابطه تعادلی بلندمدت تعدیل شوند، چنین رابطه‌ای در بلندمدت میان متغیرها برقرار نمی‌شود. از این رو همان‌طور که نشان داده شد، هم‌انباشتگی مستلزم ECM است. یکی از ویژگی‌های اساسی بردارهای هم‌انباشتگی آن است که روند زمانی آنها تحت تأثیر انحرافاتی است که از تعادل بلندمدت وجود دارد. بر این اساس در صورتی یک سیستم به تعادل بلندمدت بازمی‌گردد که حداقل تغییرات برخی از متغیرها در جهت عکس عدم تعادل ایجادشده، باشد.

روابط تعادلی بلندمدت برای دو متغیر(دارایی یا سهام) A و B به صورت زیر است:

1. Engle and Granger
2. Johansen
3. Dickey fuller test
4. Error Correction Model (ECM)

$$\Delta A_t = \alpha_A(A_{t-1} - \beta A_{t-1} + \rho_A) + \cdots + \varepsilon_{At}$$ رابطه ۱)

$$\Delta B_t = \alpha_B(B_{t-1} - \beta B_{t-1} + \rho_B) + \cdots + \varepsilon_{Bt}$$ رابطه ۲)

در رابطه‌های ۱ و ۲، ε_{At} و ε_{Bt} به‌ترتیب فرایند نویز سفید مربوط به سری‌های زمانی متعلق به سهم A و B هستند. α_A و α_B به‌ترتیب میزان تصحیح خطا در معادلات فوق برای دو سری زمانی مربوط به سهم A و B را نشان می‌دهند. پارامتر ρ به‌عنوان محدودیت ثابت در معادلات فوق است و پارامتر β نیز ضریب هم‌انباشتگی را نشان می‌دهد.

عبارت داخل پرانتز در روابط بالا، به روابط تعادلی بلندمدت اشاره دارد که تفاضل قیمتی نامیده می‌شود و می‌توان آن را به‌صورت زیر، یعنی یک میانگین (μ) و یک عبارت نوفه سفید (ε_t) بازنویسی کرد:

$$A_{t-1} - \beta B_{t-1} + \rho = \mu + \varepsilon_t$$ رابطه ۳)

اکنون با توجه به خاصیت هم‌انباشتگی، تفاضل قیمتی دارای میانگین ثابتی طی زمان خواهد بود. از این رو، می‌توان مقدار استاندارد شده تفاضل قیمتی را به‌منظور ترسیم و اتخاذ موقعیت‌های معاملاتی روی آن، به‌صورت زیر تعریف کرد.

$$indicator = \frac{spread - mean(spread)}{STD\ (spread)}$$ رابطه ۴)

پس از انتخاب زوج سهام مناسب برای اجرای استراتژی و تخمین پارامترها، طراحی استراتژی معاملات زوجی و اجرای آن مدنظر خواهد بود. در واقع یکی از بخش‌های بسیار مهم در معاملات زوجی، طراحی استراتژی معاملات زوجی و تعیین مقادیر بهینه پارامترهای آن است.

۱. انتخاب پنجره زمانی مناسب برای تخمین مجدد پارامترها[1]: در واقع انتخاب تواترهای مناسب برای اجرای آزمون هم‌انباشتگی، تخمین پارامترها و ضرایب هم‌انباشتگی و تخمین مجدد معادله تفاضل قیمتی است.
۲. پنجره معاملاتی[2]: در صورت وجود هم‌انباشتگی، در طول مدت زمان پنجره معاملاتی، ضرایب هم‌انباشتگی، مقادیر پارامترها و معادله تفاضل قیمتی صادق خواهند بود.
۳. آستانه‌های معاملاتی (Δ): به آستانه‌هایی در بالا و پایین میانگین تفاضل قیمتی به‌منظور صدور سیگنال‌های معاملاتی و گرفتن موقعیت‌ها اشاره دارد.
۴. حد ضرر: آستانه‌هایی در دو طرف میانگین تفاضل قیمتی و پهن‌تر از آستانه‌های Δ که به‌منظور بستن موقعیت هنگام ضرر و جلوگیری از ضرر بیشتر استفاده می‌شوند. یکی از ریسک‌های استراتژی معاملات زوجی در رانده شدن طولانی مدت تفاضل قیمتی از میانگین بلندمدتش است که این اهمیت تعیین بهینه موقعیت حد ضرر را نشان می‌دهد.

1. Estimation window
2. Trading window

در مطالعات قبلی اغلب از مقادیری ثابت و مشخص برای این چهار پارامتر استفاده شده، در حالی که در این پژوهش از یادگیری تقویتی برای انتخاب مقادیر بهینه این پارامترها و با هدف ماکزیمم‌سازی نسبت سورتینو استفاده شده است. در واقع با به‌کارگیری RL به‌جای تخصیص مقادیری ثابت برای هر یک از پارامترهای طراحی معاملات، عامل با به‌کارگیری تجربه‌های خود در فرایند پیش‌پردازش و استفاده از اطلاعات گذشته، در هر بار باز و بسته شدن پرتفو، پارامترهای بهینه را با توجه به تجربه‌ها و اکتشاف‌های خود انتخاب می‌کند. واضح است که این چهار پارامتر به‌صورت پویا می‌توانند مقداردهی شوند. توضیحات کامل دربارهٔ چگونگی به‌کارگیری الگوریتم یادگیری تقویتی و عملکرد آن در قسمت بعدی ارائه شده است.

در اجرای استراتژی با انحراف تفاضل قیمتی از مقدار تعادل بلندمدت خود، در صورتی که تفاضل قیمتی یکی از آستانه‌های Δ بالا یا پایین مربوط به آن دوره زمانی را لمس کند، موقعیت‌های معاملاتی را برای پرتفوی دارایی‌های خود (در اینجا دو دارایی) بازکرده و پس از وقوع یکی از دو حالت اصلاح انحراف و بازگشت تفاضل قیمتی به میانگین خود یا با عبور از حد ضرر با اتخاذ موقعیت‌های معکوس، موقعیت معاملاتی بسته می‌شود. حال بار دیگر اطلاعات و تابع ارزش الگوریتم به‌روز شده و الگوریتم دوباره برای مقداردهی به چهار پارامتر یاد شده تصمیم‌گیری می‌کند. همان‌طور که اشاره شد، در این پژوهش از نسبت سورتینو به‌عنوان تابع هدف و معیاری برای سنجش عملکرد الگوریتم با در نظر گرفتن توأم بازده و ریسک استفاده شده است.

یادگیری تقویتی از دو مؤلفه اصلی عامل و محیط تشکیل شده است. عامل در محیط زندگی کرده و با توجه به بازخوردی که از محیط می‌گیرد، تجربه‌های خود را به‌روزرسانی می‌کند. به‌طور رسمی در یادگیری تقویتی، عامل در هر لحظه در حالت s قرار داشته، با انتخاب عملی (a) از فضای اعمال خود و انجام آن به حالت بعدی (s') انتقال می‌یابد و پاداشی (r) از محیط دریافت می‌کند. سپس بر اساس پاداش به‌دست آمده، تجربه خود از حضور در حالت s و انجام عمل a را به‌روزرسانی می‌کند. به‌روزرسانی تجربه عامل تخمین ارزشی از حالت مد نظر و عمل انجام‌شده در آن حالت (تخمین ارزش حالت ـ عمل) است. برای به‌روزرسانی تخمین ارزش حالت ـ عمل رابطه زیر به‌کار گرفته می‌شود:

تخمین جدید ⟵ تخمین قبلی + نرخ یادگیری × پاداش به دست آمده

میزان یادگیری (α) عددی بین [۰,۱] بوده و با افزایش زمان کاهش می‌یابد. در ابتدای زندگی عامل، مقدار α نزدیک به ۱ است، چرا که عامل در ابتدا تجربه‌ای نداشته و باید میزان اهمیت عامل در تخمین ارزش حالت ـ عمل خود به بازخوردی که از محیط دریافت می‌کند، زیاد باشد. با افزایش زمان، از آنجا که عامل در هر تغییر حالت، بازخوردی از محیط می‌گیرد، بهتر است به تخمین ارزش حالت ـ عمل خود اهمیت بیشتری داده و تأثیر بازخورد از محیط را در به‌روزرسانی تخمین‌های جدید کاهش دهد. بنابراین عامل می‌تواند در طول زندگی خود در محیط، رفتار محیط را به‌خوبی شناسایی کرده و در حالت‌های مختلف، بهترین تصمیم را برای بیشینه‌کردن پاداش خود اتخاذ کند.

یکی از مسائل معروف در زمینه یادگیری تقویتی، مسئله n-arm bandit است. N-arm bandit در واقع نوعی

وسیله بازی است که از N اهرم تشکیل شده است. با هر دفعه کشیدن اهرم مربوط به دستگاه بازی، امتیازی به‌عنوان جایزه توسط دستگاه به بازیکن تعلق‌گرفته و نمایش داده می‌شود. از این مسئله در یادگیری تقویتی برای طراحی یک عامل فراگیرنده استفاده می‌شود. به بیان دیگر با مدل کردن مسئله n-arm bandit در قالب یک محیط، عاملی طراحی می‌کند تا بتواند بیشترین امتیاز ممکن از بازی با دستگاه را به دست آورد. در این مسئله فضای حالت، فقط از یک حالت تشکیل می‌شود و فضای اعمال، تعداد اهرم‌های دستگاه هستند. در هر مرحله عامل پس از انتخاب اهرم، کشیدن آن و گرفتن امتیاز، دوباره به حالتی بازمی‌گردد که می‌تواند اهرم دیگری را برای کشیدن انتخاب کرده، بکشد و امتیاز دریافت کند. از آنجا که امتیازی دریافتی از توزیع خاصی پیروی می‌کند که برای عامل مشخص نیست، عامل در هر مرحله سعی می‌کند با اکتشاف محیط و به‌دست آوردن تجربه‌های کافی، اهرمی را که میانگین میزان امتیاز بیشتری به دست می‌دهد، انتخاب کند. با افزایش تعداد تکرارها، میزان یادگیری عامل کاهش‌یافته و اهمیت انتخاب اهرمی که تاکنون تخمین ارزش بیشتری دارد، افزایش می‌یابد.

می‌توان از مسئله N-arm bandit ایده گرفت و برای بیشینه کردن سود به‌دست آمده در معاملات زوجی یک عامل، یادگیری تقویتی طراحی کرد. از آنجا که تغییر رفتار دو سهم در طول زمان دستخوش عوامل غیرقطعی فراوانی است و مدل‌کردن این عوامل به‌صورت نویزی ساده خیلی کارآمد نیست، به نظر می‌رسد طراحی یک عامل یادگیری تقویتی که بتواند عوامل غیرقطعی مختلف را مستقل از نوع تأثیر آن تا حد خوبی مدل کند، کارآمدتر از روش‌های سنتی باشد.

در این حالت زمانی که یک پرتفو باز می‌شود، عامل بایستی مقدار مناسبی برای پارامترهای بیان‌شده انتخاب کند و پس از بسته‌شدن پرتفو، میزان نسبت سورتینو به‌دست آمده را به‌عنوان پاداش در نظر بگیرد. بنابراین حالت عامل، زمان باز شدن هر پرتفوی است که در حقیقت همگی یک حالت را تشکیل می‌دهند. فضای اعمال عامل، شامل چهار پارامتر بیان شده است که عامل در هنگام باز شدن پرتفو بایستی براساس تخمین ارزش حالت ـ عمل خود مقداری برای هر یک از پارامترها انتخاب کرده و پس از بسته شدن پرتفو، با استفاده از نسبت سورتینو به‌دست آمده، تخمین ارزش حالت ـ عمل خود را به‌روزرسانی کند. به بیان دیگر، در زمان باز شدن هر پرتفو مشابه n-arm bandit بایستی مقداری برای هر یک از چهار پارامتر بیان‌شده انتخاب کند و زمانی که پرتفو بسته می‌شود، ارزش انتخاب این مقادیر برای پارامترهای بیان شده به‌روزرسانی می‌شود و عامل دوباره به حالت اولیه خود بازمی‌گردد و منتظر می‌ماند که دوباره پرتفویی بازشده و چهار مقدار دیگر را انتخاب کند. همان‌طور که بیان شد، مقادیر دو پارامتر اول گسسته و مقادیر دو پارامتر دوم پیوسته‌اند. از آنجا که در مسئله n-arm bandit، فضای اعمال گسسته است، مقادیر دو پارامتر دلتا و پارامتر حد ضرر را با در نظر گرفتن دقتی (به‌طور مثال ۰/۵ واحد) گسسته می‌کنیم. حال مقادیر تمام چهار پارامتر بیان‌شده، گسسته است. هر یک از جایشگت‌های تمام مقدار پارامترها، یک عمل محسوب می‌شود (در واقع نماینده یک اهرم در مسئله n-arm bandit هستند). با مدل‌سازی مسئله معادلات زوجی مشابه مسئله n-arm bandit ، عامل را طراحی کرده و شبیه‌سازی می‌کنیم. نتایج به‌دست آمده، مؤثر بودن روش پیشنهادی طراحی عامل یادگیری تقویتی نسبت به سایر روش‌های قبلی را نشان می‌دهد. الگوریتم این روش در زیر آورده شده است:

Algorithm 1 (RL Pairs Trading)

Input: Stock A, Stock B

1. Initialization:
2. $actionspace_{n\times 4} \leftarrow [\delta_{n\times 1}, stoploss_{n\times 1}, estimation\ window_{n\times 1}, tradingwindow_{n\times 1}]$
   $numofiteration \leftarrow maxnumberofiteration$
3. **For** $episode \leftarrow 1\ to\ numofiteration$
4. $action \leftarrow$ choose action based on epsilon greedy policy
5. $reward \leftarrow$ performaction$(action, A, B)$
6. action update$(action, reward)$
7. **end**
8. $return\ sortino\ ratio.$

Output: Sortino Ratio

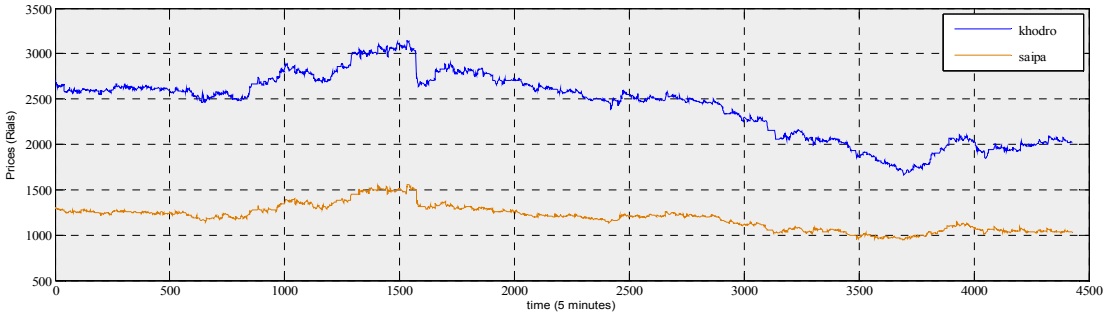
در الگوریتم فوق، ابتدای فضای اعمال بر اساس آنچه اشاره شد، مقداردهی اولیه می‌شود. این فضا شامل چهار پارامتر پنجره تخمین، پنجره اجرا، مقدار دلتا و مقدار حد ضرر است. سپس تعداد تکرارهای کافی برای آموزش عامل انتخاب می‌شود. در ادامه برای هر دوره[1] اجرایی، ابتدا بر اساس سیاست تصمیم‌گیری $\epsilon - greedy$ و بر اساس تجربه عامل، عمل مناسب یعنی انتخاب یک مقدار از چهار پارامتر فوق در قالب یک عمل عامل انتخاب می‌شود. سپس بر اساس متدولوژی ارائه‌شده، مقدار نسبت سورتینو مد نظر در قالب پاداش محاسبه‌شده و به‌عنوان بازخورد محیط پیرامون عامل به آن بازگردانده می‌شود. در ادامه عامل با استفاده از تابع actionupdate و مقدار پاداش دریافت شده در مرحله قبل، تجربه عامل از مقدار ارزش ـ عمل هر یک از چهار پارامتر اشاره شده به‌روزرسانی می‌شود. در انتها با خاتمه الگوریتم، مقدار نسبت سورتینو کل به‌دست آمده به ازای بررسی تمام داده‌های ورودی به عنوان خروجی بازگردانده می‌شود.

در این پژوهش به منظور ارزیابی عملکرد الگوریتم پیشنهادی و اجرای استراتژی معاملات زوجی از داده‌های قیمتی درون‌روزی[2] سهام شرکت‌های منتخب در بورس اوراق بهادار تهران، شامل سهام زوج شرکت‌های ایران خودروـ سایپا، پالایش نفت لاوان ـ پالایش نفت تهران، بانک ملت ـ بانک صادرات، گسترش نفت و گاز پارسیان ـ پتروشیمی شپدیس و بانک صادرات ـ بانک انصار مربوط به بازه زمانی شش ماهه ۱۳۹۴/۰۲/۳۰ تا ۱۳۹۴/۰۸/۳۰ استفاده شده است.

با توجه به اینکه تخمین پارامترها و در نهایت محاسبه معادله تفاضل قیمتی از مقایسه روند قیمتی دو سری زمانی مربوط به قیمت دو دارایی حاصل می‌شود، این دو سری زمانی را با تبدیل آنها به فواصل زمانی پنج دقیقه و به‌کارگیری آخرین قیمت در هر پنج دقیقه به‌عنوان نماینده آن، به سری‌هایی گسسته و مقایسه‌پذیر تبدیل کرده و از این سری‌های گسسته به‌منظور انجام آزمایش‌ها استفاده می‌کنیم. شایان ذکر است که اطلاعات قیمتی مربوط به آخرین قیمت برای هر سهم در فواصل زمانی پنج دقیقه از نرم‌افزار ره‌آورد نوین استخراج شده است.

در شکل زیر سری قیمت‌های درون روزی مربوط به زوج سهام شرکت‌های ایران‌خودرو و سایپا نشان داده شده است. همان‌طور که مشاهده می‌شود، روند حرکتی زوج سهام به‌صورت تاریخی یکدیگر را دنبال می‌کنند.

1. Episode
2. Intraday price



شکل ۱. قیمت‌های درون‌روزی سهام شرکت‌های ایران‌خودرو و سایپا

یافته‌های پژوهش

در این قسمت با انجام آزمایش‌های طراحی شده، به ارزیابی عملکرد الگوریتم پیشنهادی و مقایسه نتایج حاصل از آن با حالت بدون استفاده از الگوریتم RL در معاملات زوجی پرداخته می‌شود. بدین منظور الگوریتم پیشنهادشده در قسمت قبل را با در نظر گرفتن شرایط و پارامترهای مشخص روی زوج سهام مختلف آزمایش می‌کنیم. در ادامه نخست شرایط انجام آزمایش‌ها تشریح شده و پس از آن به انجام آزمایش‌ها و آزمون پایایی به‌منظور اعتبارسنجی الگوریتم پیشنهادی پرداخته می‌شود. در انتهای این قسمت نیز نتایج به‌دست آمده بررسی و تحلیل خواهد شد.

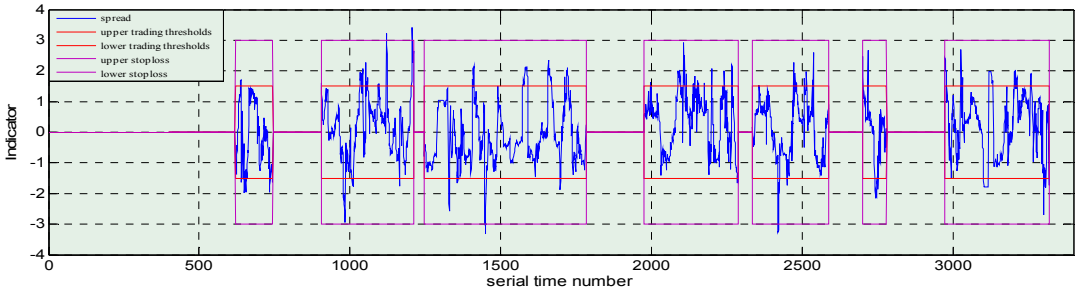
همان‌طور که بیان شد، هدف این پژوهش پیشنهاد یک الگوریتم برای بهینه‌سازی استراتژی معاملات زوجی با استفاده از الگوریتم یادگیری تقویتی است که تابع هدف آن ماکزیمم‌سازی نسبت سورتینو است.

ما از الگوریتم یادگیری تقویتی برای انتخاب پویای مقادیر چهار پارامتر آستانه‌های معاملاتی، حدضرر، پنجره معاملاتی و پنجره تخمین استفاده کرده‌ایم. مقادیر این چهار پارامتر که به‌عنوان ورودی در الگوریتم یادگیری تقویتی به‌کار می‌روند، به‌صورت زیر است:

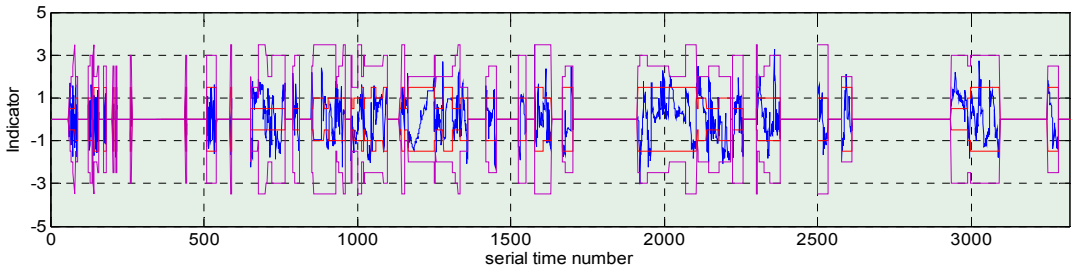
- آستانه‌های معاملاتی در فاصله ۲/۵ انحراف معیار در دو طرف میانگین تفاضل قیمتی و با دقت ۰/۵ انحراف معیار.
- آستانه‌های حدضرر در فاصله چهار انحراف معیار در دو طرف میانگین تفاضل قیمتی و با دقت ۰/۵ انحراف معیار.
- پنجره معاملاتی در بازه ۵ تا ۱۲۰ دقیقه و با دقت ۱۰ دقیقه.
- پنجره تخمین در بازه ۶۰ تا ۶۰۰ دقیقه و با دقت ۱۰ دقیقه.

همان‌طور که اشاره شد، الگوریتم یادگیری تقویتی در هر بار تکرار (train) با انتخاب یک مقدار برای هر یک از چهار پارامتر، به ارزیابی و شناسایی محیط می‌پردازد. با دو پارامتر α و ϵ چگونگی استراتژی RL در به‌کارگیری تجربه‌ها (یادگیری‌ها) و اکتشاف‌ها و نسبت آنها تعیین می‌شود. در انجام تمامی شبیه‌سازی‌های این مقاله از مقادیر ϵ=۱, α=۰/۵ برای الگوریتم یادگیری تقویتی استفاده شده است. با توجه به ماهیت الگوریتم یادگیری تقویتی که در آن مفاهیم آموزش عامل و کسب تجربه آن در محیط به‌منظور اتخاذ تصمیم برای انجام معاملات، از اهمیت بسیار زیادی برخوردار است، در تمام آزمایش‌ها از ۷۵ درصد داده‌ها به‌عنوان داده آموزش (insample) و ۲۵ درصد باقی‌مانده برای انجام معاملات (outsample) استفاده شده است.

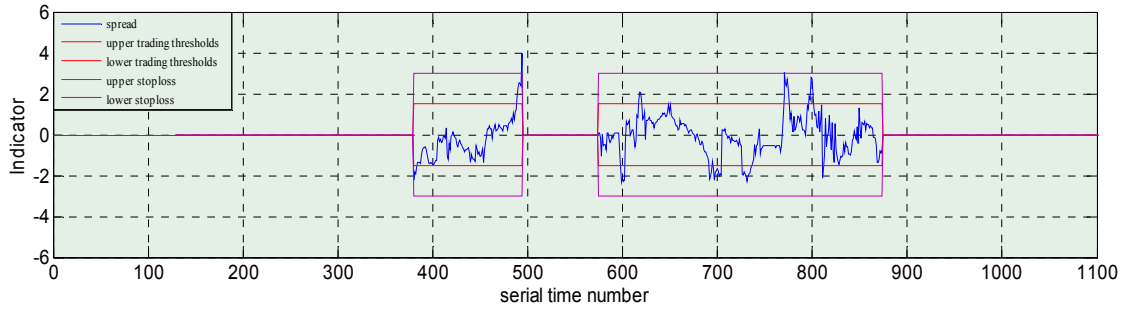
همان‌طور که اشاره شد، در این پژوهش از چند زوج داده که همگی از شناخته‌شده‌ترین سهم‌های موجود در بورس اوراق بهادار تهران هستند، استفاده شده است. در ادامه نتایج حاصل از اجرای آزمایش‌ها روی زوج سهام ایران‌خودرو ـ سایپا در بازه زمانی ۱۳۹۴/۰۲/۳۰ تا ۱۳۹۴/۰۸/۳۰ نشان داده شده است. شکل‌های ۲ و ۳ نتایج اجرای استراتژی معاملات زوجی در قسمت insample با دو روش استفاده از الگوریتم پیشنهادی و با مقادیر ثابت پنجره تخمین ۳۸۰ دقیقه، پنجره معاملاتی ۸۰ دقیقه و مقادیر ۱/۵ و ۳ انحراف معیار را به‌ترتیب برای آستانه‌های Δ و حد ضرر نشان می‌دهد. در ادامه نیز به‌ترتیب شکل‌های ۴ و ۵ نتایج اجرای استراتژی معاملات زوجی با مقادیر ثابت و با استفاده از الگوریتم یادگیری تقویتی را در قسمت outsample نشان می‌دهد.



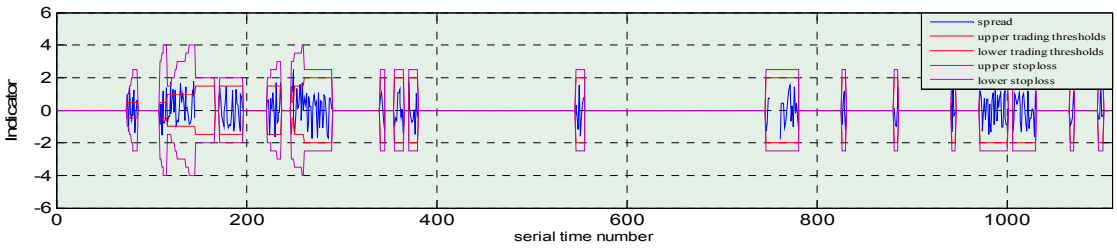
شکل ۲. اندیکاتور و موقعیت‌های معاملاتی روی داده‌های آموزش (insample)
برای زوج سهام ایران‌خودرو ـ سایپا با روش پارامترهای ثابت



شکل ۳. اندیکاتور و موقعیت‌های معاملاتی روی داده‌های آموزش (insample)
برای زوج سهام ایران‌خودرو ـ سایپا با روش پیشنهادی



شکل ٤. اندیکاتور و موقعیت‌های معاملاتی روی داده‌های outsample
برای زوج سهام ایران‌خودرو ـ سایپا با روش پارامترهای ثابت



شکل ۵. اندیکاتور و موقعیت‌های معاملاتی روی داده‌های outsample برای زوج سهام ایران‌خودرو ـ سایپا با روش پیشنهادی

همان‌طور که در شکل‌های ۳ و ۵ مشاهده می‌شود در حالت استفاده از الگوریتم RL طول پنجره‌های معاملاتی و همین طور فاصله آستانه‌های معاملاتی و حدضررها در هر دوره متغیر بوده و توسط عامل با توجه به بازخوردی که دریافت کرده (در جهت ماکزیمم‌سازی تابع هدف که در واقع ماکزیمم‌سازی نسبت سورتینو است) تغییر می‌کند.

خلاصه نتایج حاصل از اجرای آزمایش‌ها در جدول ۱ بیان شده است.

جدول ۱. نتایج شبیه‌سازی روی زوج سهام ایران‌خودرو ـ سایپا با استفاده از الگوریتم پیشنهادی و روش پارامترهای ثابت

	حالت معمولی با پارامترهای ثابت		حالت با استفاده از الگوریتم RL	
طول پنجره تخمین (دقیقه)	۳۸۰		متغیر در بازه (۶۰،۶۰۰)	
اندازه پنجره معاملاتی (دقیقه)	۸۰		متغیر در بازه (۵،۱۲۰)	
Δ (انحراف معیار)	۱/۵		متغیر در بازه (۲/۵ ، ۰/۵)	
حدضرر (انحراف معیار)	۳		متغیر در بازه (۱،۴)	
	insample	outsample	insample	outsample
بازده	٪ ۱۰/۵	٪ ۵/۴	٪ ۳۹/۲	٪ ۱۹
نوسانات منفی	۰/۸۹	۰/۲	۰/۲۰۲	۰/۰۸۹
بازده سالیانه	٪ ۲۸	٪ ۴۳/۲	٪ ۱۰۴/۵	٪ ۱۵۲
نسبت سورتینو	۰/۲۱	۱/۱۳	۶/۷۸	۷/۹
تعداد معاملات	۹۷	۳۸	۲۲۳	۱۲۱
میانگین بازده در هر معامله	٪ ۰/۰۰۱	٪۰/۰۰۱۴	٪۰/۰۰۱۸	٪۰/۰۰۱۶
میانگین زمان هر معامله (دقیقه)	۱۷۰	۱۴۵	۷۵	۴۵/۵

همان طور که در جدول ۱ مشاهده می‌شود، نتایج برتری چشمگیر اجرای معاملات زوجی با استفاده از الگوریتم پیشنهادی را نسبت به حالت معمولی با پارامترهای ثابت نشان می‌دهد؛ به‌طوری که مقدار بازده حاصل با استفاده از الگوریتم پیشنهادی در هر دو حالت insample و outsample برتری محسوسی نسبت به حالت با پارامترهای ثابت دارد. نسبت سورتینو نیز که توأمان بازده و ریسک‌های منفی را نشان می‌دهد، برای RL در حالت insample برابر با ۶/۷۸ است، درحالی‌که برای حالت معمولی (با پارامترهای ثابت) ۰/۲۱ به دست آمده است. در حالت outsample نیز این نسبت برای الگوریتم پیشنهادی ۷/۹ و برای حالت معمول ۱/۱۳ است. نکته شایان توجه این است که در بازه زمانی بررسی، بازدهی سهام شرکت‌های ایران‌خودرو و سایپا به‌ترتیب ۲۶/۴- و ۲۰/۷- درصد بوده است.

جدول ۲. نتایج به‌دست آمده از اجرای استراتژی معاملات زوجی با روش پیشنهادی و روش پارامترهای ثابت روی زوج سهام‌های مختلف

	خودرو – سایپا			وبملت – وبصادر			شتران – شاوان			پارسان – شپدیس			وبصادر – وانصار		
	پارامتر ثابت		الگوریتم پیشنهادی	پارامتر ثابت		الگوریتم پیشنهادی	پارامتر ثابت		الگوریتم پیشنهادی	پارامتر ثابت		الگوریتم پیشنهادی	پارامتر ثابت		الگوریتم پیشنهادی
	In sample	Out sample	In sample	In sample	Out sample	In sample	In sample	Out sample	In sample	In sample	Out sample	In sample	In sample	Out sample	In sample
بازده	٪ ۱۰/۵	٪۵/۴	٪۳۹/۲	٪۱۳/۱	٪ ۸	٪ ۴۵	٪۹/۵	٪ -۶/۲	٪ ۳۳	٪۲۲/۵	٪۱۱/۳	٪۵۰/۱	٪ ۳۳	٪ -۱۸/۱	٪ ۷۵
نوسانات منفی	۰/۸۹	۰/۲	۰/۲۰۲	۰/۷۱	۰/۴۶	۰/۱۱	۰/۲۴	۰/۰۹	۰/۳۱	۰/۵۲	۰/۲۴	۰/۱۱	۰/۶۶	۰/۳۱	۰/۱۹۵
بازده سالیانه (٪)	٪ ۲۸	٪۳۳/۲	۱۰۴/۵	۳۵	۶۴	۱۲۰/۲	۲۵/۴	-۴۹/۶	۸۸/۱	۶۰	۹۰/۴	۱۳۳/۷	۸۸/۱	-۱۴۴/۸	۲۰۰/۲
نسبت سورتینو	۰/۲۱	۱/۱۳	۶/۷۸	۰/۵۴	۱/۰۸	۴/۹۵	۰/۴۶	-۱/۲۱	۳/۵۶	۰/۷۱۸	۱/۵۲	۸/۴۷	۱/۰۸	-۲/۴۴	۱۰/۵۹
تعداد معاملات	۹۷	۳۸	۲۲۳	۳۸	۲۸	۱۸۵	۵۸	۴۰	۱۵۲	۱۱۵	۸۴	۲۰۱	۴۶	۱۸	۱۵۲
میانگین زمان هر معامله (دقیقه)	۱۷۰	۱۴۵	۷۵	۴۳۴	۱۹۶/۴	۸۹/۱	۲۸۴/۵	۱۳۷/۵	۱۰۸/۵	۱۳۳/۳	۶۵/۵	۸۲	۳۵۹	۳۰۵/۵	۱۰۸/۵

از طرفی این نکته حائز اهمیت است که در حالت استفاده از الگوریتم یادگیری تقویتی، تعداد معاملات افزایش چشمگیری داشته است که این خود به‌دلیل متغیر بودن آستانه‌ها و پنجره‌های زمانی در اجرای معاملات است.

باید به این نکته نیز توجه کرد که ما در تمام آزمایش‌ها از ۱۰۰ بار تکرار برای یادگیری عامل به‌کمک داده‌های insample استفاده کرده‌ایم، بدین معنا که عامل ۱۰۰ حالت از کل فضای حالت را مشاهده و ارزیابی کرده است. با در نظر گرفتن بازه مقادیر هر یک از چهار پارامتر بررسی‌شده و دقت آنها، کل فضای حالت مسئله برای انتخاب‌های عامل در هر بار تکرار، حدود ۵۰۰۰۰ حالت است که به‌طبع با افزایش تعداد تکرارها، عامل به درک و شناخت بهتری از محیط رسیده و نتایج بهبود می‌یابد.

آزمون پایایی

به‌طور کلی آزمون پایایی درجه‌ای است که یک سیستم می‌تواند در شرایط ورودی‌ها و محیط‌های مختلف به درستی عمل کند. در واقع هدف از انجام آزمون پایایی توسعه موردها و محیط‌های انجام آزمایش به‌منظور بررسی قدرت سیستم است. بدین منظور در این مقاله از پنج زوج داده برای انجام آزمایش‌ها و مقایسه نتایج با حالت بدون استفاده از الگوریتم RL (با پارامترهای ثابت) و بررسی قدرت الگوریتم استفاده شده است. نتایج به‌دست آمده در جدول ۲ مشاهده می‌شود.

این جدول نشان می‌دهد که در تمامی زوج‌ها، نسبت سورتینو و بازده حاصل از انجام معاملات زوجی با استفاده از الگوریتم RL، نسبت به حالت معمولی برتری داشته که این خود نشان‌دهنده پایایی الگوریتم پیشنهادی در خصوص زوج سهم‌های مختلف و از گروه‌های صنعت مختلف است.

نتیجه‌گیری

همان‌طور که اشاره شد، در تحقیقات قبلی اغلب از مقادیر ثابت و مشخص برای چهار پارامتر آستانه‌های معاملاتی، حد ضرر، پنجره معاملاتی و پنجره تخمین استفاده شده است، در حالی‌که در این پژوهش برای انتخاب مقادیر بهینه این پارامترها و با هدف ماکزیمم‌سازی نسبت سورتینو، از یادگیری تقویتی استفاده شده است. در واقع با به‌کارگیری روش یادگیری تقویتی، به‌جای تخصیص مقدار ثابت برای هر یک از پارامترهای طراحی معاملات، عامل با به‌کارگیری تجربه‌های خود در فرایند پیش‌پردازش و استفاده از اطلاعات گذشته، در هر بار باز و بسته شدن پرتفو، پارامترهای بهینه را با توجه به تجربه‌ها و اکتشاف‌های خود انتخاب می‌کند. نتایج آزمایش‌های انجام‌شده، عملکرد چشمگیر این الگوریتم در اجرای استراتژی معاملات زوجی را نسبت به پژوهش‌های قبلی صورت گرفته در اجرای این استراتژی با پارامترهای ثابت نشان می‌دهد.

استراتژی معاملات زوجی با الگوریتم پیشنهادی می‌تواند به‌عنوان استراتژی بازار خنثی در تمامی شرایط بازار اعم از رونق و رکود توسط سرمایه‌گذاران و معامله‌گران حقیقی و حقوقی استفاده شود. به‌منظور ارائه پیشنهاد برای پژوهش‌های آتی نیز می‌توان به در نظر گرفتن هزینه‌های معاملاتی در انجام معاملات در استراتژی معاملات زوجی اشاره کرد.

همچنین می‌توان عملکرد رویکرد فازی را در انتخاب مقادیر بهینه برای پارامترهای طراحی معاملات زوجی بررسی کـرده و نتایج را با الگوریتم پیشنهاد شده در این پژوهش مقایسه کرد.

References

Bertram, W., (2010). Analytic solutions for optimal statistical arbitrage trading. *Physica* A, 2010, 389(11), 2234–2243.

Dai, M., Zhang, Q., & Zhu, Q. J. (2010). Trend following trading under a regime switching model. *SIAM Journal on Financial Mathematics*, *1*(1), 780-810.

Engle, R. F., and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. Econometrica: *journal of the Econometric Society*, 251-276.

Gao, X., & Chan, L. (2000). An algorithm for trading and portfolio management using Q-learning and sharpe ratio maximization. *In Proceedings of the international conference on neural information processing* (pp. 832-837).

Gatev, E., Goetzmann, W. N., and Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies*, 19(3), 797-827.

Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, 16(1), 121-130.

Guo, X., & Zhang, Q. (2005). Optimal selling rules in a regime switching model. *IEEE Transactions on Automatic Control*, 50, 1450–1455.

Hillebrand, E. (2003). A mean-reversion theory of stock-market crashes. *Journal of Finance*, 41, 591-601.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2), 231-254.

Lee, J. W., Park, J., Lee, J., & Hong, E. (2007). A multiagent approach to Q-learning for daily stock trading. Systems, Man and Cybernetics, Part A: Systems and Humans, *IEEE Transactions* on, 37(6), 864-877.

Moody, J., and Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, 12(4), 875–889.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. MIT Press.

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

Won Lee, J. (2001). Stock price prediction using reinforcement learning. In Industrial Electronics, 2001. Proceedings. ISIE 2001. *IEEE International Symposium* on (Vol. 1, pp. 690-695). IEEE.

Zeng, Z., & Lee, C. G. (2014). Pairs trading: optimal thresholds and profitability. *Quantitative Finance*, 14(11), 1881-1893.

Zhang, Q. (2001). Stock trading: An optimal selling rule. *SIAM Journal on Control and Optimization*, 40(1), 64-87.