

تحلیل مخاطره سرمایه‌گذاری در شرکت‌های پذیرفته شده در بورس اوراق بهادار ایران با استفاده از الگوریتم تقویتی

صدیقه زمانی مهربان^{۱*} و ناصر اقبالی^۲

^۱ گروه آمار، دانشکده علوم پایه، دانشگاه بین‌المللی امام خمینی (ره)، قزوین، ایران

^۲ گروه حسابداری، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران

تاریخ دریافت: ۱۴۰۴/۰۲/۱۰

تاریخ پذیرش: ۱۴۰۴/۰۵/۰۱

چکیده:

در سال‌های اخیر بسیاری از پژوهشگران به پژوهش روی مدل‌های پیش‌بینی قیمت سهام و بازده بازار تمرکز کرده‌اند. مدل‌های پیش‌بینی آماری مانند مدل‌های خودبازگشتی، خودبازگشتی-میانگین متحرک و واریانس ناهمگن شرطی خودبازگشتی تعمیم‌یافته به‌طور گسترده مورد استفاده قرار گرفته است. توسعه پردازنده‌های کامپیوتری موجب معرفی الگوریتم‌های جدید از جمله الگوریتم‌های یادگیری ماشین جهت پیش‌بینی داده‌های مالی شده است. در این مقاله، به‌منظور تحلیل مخاطره سرمایه‌گذاری در شرکت‌های پذیرفته شده در بورس اوراق بهادار ایران، داده‌های مربوط به حجم معاملاتی شرکت‌های پذیرفته شده در بورس در نظر گرفته شده و مدل بهینه در مدل‌های خودبازگشتی برداری با استفاده از الگوریتم تقویتی برازش داده می‌شود. مدل پیشنهادی با استفاده از الگوریتم تقویتی با مدل‌های آماری معمول مقایسه می‌شود. نشان می‌دهیم که الگوریتم تقویتی در برازش مدل عملکرد مناسب‌تری دارد.

واژه‌های کلیدی: الگوریتم تقویتی، حجم معاملات، مدل خودبازگشتی برداری، مدل واریانس ناهمگن شرطی خودبازگشتی تعمیم‌یافته.

۱ مقدمه

بتوانند با اطمینان در جهت روند به معامله بپردازند. روندها در بازار بورس با استفاده از حجم معاملات تأیید می‌شوند و در صورتی که حجم معاملات، یک روند را در بازار تعیین کند عملاً نشان از وجود شرایط غیرعادی در آن روند خواهد داشت.

حجم معاملات نشان دهنده قوی یا ضعیف بودن روند مورد بررسی است؛ به این صورت که افزایش روند صعودی حجم معاملات تأیید کننده قوی بودن روند صعودی بوده، به این معنا که قدرت در دست خریداران است و کاهش این حجم نشان‌دهنده ضعیف شدن قدرت روند صعودی بوده و خطاری برای بازگشت روند و نزولی شدن آن محسوب می‌شود. همچنین، افزایش روند نزولی حجم معاملات تأیید کننده قوی بودن روند نزولی بوده، به عبارت دیگر قدرت در دست فروشندگان است و کاهش این حجم کاهش نشان‌دهنده ضعیف شدن قدرت روند نزولی است. برای مطالعه بیشتر به نجار زاده و زیودار [۱۰] مراجعه کنید. تعیین حجم معاملات یکی از اصول اساسی در معامله‌گری است که مشخص می‌کند چه مقدار از سرمایه به یک معامله اختصاص داده شود.

حجم معاملات به‌عنوان یک شاخص آماری برای تشخیص قدرت خریداران و فروشندگان استفاده می‌شود که اهمیت بسیار زیادی در تحلیل سهم‌های مختلف و شاخص بورس دارد. به تعداد معاملات انجام شده در سهم شاخص کل و شاخص صنعت در بازه زمانی مورد بررسی، حجم معاملات می‌گویند. حجم معاملات یک سهم در بازار بورس در واقع به مجموعه معاملات انجام پذیرفته در آن سهم در یک روز معاملاتی اطلاق می‌شود. یکی از اطلاعات بسیار مهمی که همراه با حجم روزانه معاملات، می‌توان از آن برای تحلیل شرایط مختلف استفاده نمود و به آن نیاز خواهیم داشت متوسط حجم معاملات آن سهم در ۹۰ روز معاملاتی است. حجم معاملات نشان‌دهنده قوت یا ضعف روند مورد بررسی است و تأثیر زیادی بر افزایش سود و کاهش ریسک معاملاتی دارد. حجم معاملات می‌بایست روند موجود در بازار را تأیید کند تا معامله‌گران

میانگین متحرک (MA)، خودبازگشتی-میانگین متحرک ($ARMA$) و خودبازگشتی-میانگین متحرک تلفیقی ($ARIMA$) را در پیش‌بینی شاخص سهام شرکت‌های بورس اوراق بهادار تهران مقایسه کردند. غلامیان و داوودی [۷] به پیش‌بینی روند قیمت در بازار سهام با استفاده از روش‌های طبقه‌بندی هوش مصنوعی از جمله الگوریتم جنگل تصادفی پرداختند و نشان دادند که روش پیشنهادی در برآورد روند بازار نسبت به دو روش رگرسیون لجستیک و روش کاملاً تصادفی از دقت بالاتری برخوردار است.

حکیمیان و احمدی [۱] به بررسی تأثیر شوک‌های ناشی از هر یک از متغیرها بر بازدهی شاخص کل بورس اوراق بهادار تهران با استفاده از مدل $VAR - GARCH$ پرداختند. ذوالفقاری و همکاران [۲] مدلی بر اساس مدل‌های ترکیبی شبکه یادگیری عمیق و مدل‌های خانواده $GARCH$ جهت پیش‌بینی بازده شاخص کل بورس اوراق بهادار پیشنهاد دادند. رستمی و همکاران [۵] به مطالعه مدل‌سازی و تخمین بازده بورس اوراق بهادار تهران با استفاده از مدل‌های پویا پرداختند. کیانی‌زاده [۹] به مقایسه دقت مدل‌های منتخب یادگیری ماشین از جمله شبکه عصبی، رگرسیون لجستیک، k -نزدیک‌ترین همسایه، ماشین بردار پشتیبان و اعتبارسنجی متقابل جهت پیش‌بینی قیمت سهام برای ۱۲ شرکت منتخب بورس اوراق بهادار تهران که از طریق روش حذف سیستماتیک انتخاب شده‌اند در قالب مدل‌های یادگیری ماشین پرداختند.

در این مقاله، مدل سری زمانی بهینه با استفاده از الگوریتم تقویتی بر داده‌های مربوط به حجم معاملاتی شرکت‌های پذیرفته شده در بورس اوراق بهادار ایران برآزش داده و عملکرد این الگوریتم با روش‌های معمول برآزش سری زمانی مقایسه می‌شود.

مقاله به شرح زیر سازماندهی شده است: در بخش ۲، مبانی نظری مدل‌های پیشنهادی از جمله مدل خودبازگشتی برداری، مدل واریانس‌ناهمگن شرطی خودبازگشتی تعمیم‌یافته و مدل خودبازگشتی-واریانس‌ناهمگن شرطی خودبازگشتی تعمیم‌یافته ارائه شده است. در بخش ۳، مبانی نظری روش‌های انتخاب متغیر و انتخاب مدل از جمله آزمون علیت گرانجر و الگوریتم تقویتی به‌عنوان الگوریتم یادگیری ماشین بیان شده است. در بخش ۴، به مدل‌سازی داده‌های روزانه حجم معاملاتی شرکت‌های پذیرفته شده در بورس اوراق بهادار ایران از تاریخ ۲۶ فوریه ۲۰۲۳ تا ۱۰ سپتامبر ۲۰۲۳ با استفاده از الگوریتم تقویتی و روش‌های معمول سری زمانی پرداخته شده است. نتیجه‌گیری در بخش ۵ ارائه شده است.

اندازه هر معامله مستقیماً بر میزان سود و مهم‌تر از آن، میزان سرمایه‌ای که در معرض ریسک قرار می‌گیرد، تأثیر دارد. بنابراین، تعیین حجم معاملات تنها به کسب سود مربوط نمی‌شود، بلکه به حفظ سرمایه و جلوگیری از زیان‌های بزرگ نیز کمک می‌کند. با تعیین دقیق مقدار ریسک در هر معامله، معامله‌گران می‌توانند میزان زیان‌ها را محدود کرده و دوره‌های ضرر اجتناب‌ناپذیر را مدیریت کنند. در مقابل، تعیین نادرست حجم معاملات یکی از دلایل اصلی شکست در معامله‌گری است، زیرا منجر به زیان‌هایی می‌شود که جبران آن‌ها دشوار است. حجم معاملات به شدت نتایج معامله‌گری را تحت تأثیر قرار می‌دهد. اگر حجم معامله بیش از حد بزرگ باشد، یک حرکت کوچک بازار به سمت مخالف می‌تواند زیان سنگینی به همراه داشته باشد و حتی کل سرمایه را به خطر بیندازد. از طرف دیگر، اگر حجم معامله بیش از حد کوچک باشد، ممکن است سود معاملات برنده نتواند زیان‌های قبلی را جبران کند. در این رابطه نوسانات بازار نیز مهم است؛ دارایی‌های با نوسان بالا نیاز به حجم معاملات کوچک‌تری دارند تا سطح ریسک کنترل شود. برای مطالعه بیشتر به نجارزاده و زیودار [۱۰] مراجعه کنید. در مدل‌های $SAIH$ ^۱ ارائه شده توسط جنینگ و همکاران [۱۷] یک رابطه مثبت همزمان و علی بین حجم معاملات و قدرمطلق تغییرات قیمت مشاهده شده است. لی و ریو [۱۸] با استفاده از تعداد خبرهای وارده به بازار، قدرمطلق بازدهی و حجم معاملات به بررسی نوسانات بازده ورود اطلاعات در طول یک روز معاملاتی پرداخته‌اند و بحث U -شکل بودن را توضیح دادند. آنها نتیجه گرفتند که افزایش حجم معاملات، تغییر قیمت و تعداد اخبار رسیده به بازار در دوره‌های آغاز و پایان معاملات در هر روز دلیل بر ارتباط مثبت این متغیرها است.

مشیری و مروت [۱۱] داده‌های روزانه و هفتگی شاخص کل بازدهی سهام تهران را در بازه زمانی سال ۱۳۸۲ تا ۱۳۸۴ با استفاده از مدل‌های سری زمانی خطی و غیرخطی و شبکه‌های عصبی، پیش‌بینی کردند. آنها نشان دادند که مدل شبکه‌های عصبی مصنوعی عملکرد بهتری در پیش‌بینی‌های روزانه و هفتگی دارد. فلاح‌پور و علی‌پور ریکنده [۸] به مطالعه شاخص کل سهام بورس اوراق بهادار تهران با استفاده از مدل‌های مختلف شبکه‌های عصبی موجکی پرداختند و نشان دادند که شبکه عصبی موجکی در پیش‌بینی شاخص کل سهام، سطح خطای کمتری دارد و دارای عملکرد بهتری نسبت به شبکه عصبی است. رادمهر و شمس قارنه [۴] با استفاده از سری زمانی فازی بر اساس تعریف نرخ بازده به پیش‌بینی شاخص بورس تهران پرداختند. سرلک و همکاران [۶] دقت مدل‌های سری زمانی خودبازگشتی (AR)،

¹Hypothesis Information of Arrival Sequential

۲ مدل‌های پژوهش

۱.۲ مدل‌های خودبازگشتی و خودبازگشتی

برداری

از جمله مدل‌های سری زمانی خطی می‌توان به مدل‌های خودبازگشتی، میانگین متحرک و خودبازگشتی-میانگین متحرک اشاره کرد. مدل خودبازگشتی $AR(p)$ ،

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

را در نظر بگیرید که در آن p مرتبه مدل خودبازگشتی، ϕ_i برای $i = 1, \dots, p$ ضرایب خودبازگشتی و ε_t خطاهای مستقل و هم‌توزیع با میانگین صفر و واریانس متناهی σ_ε^2 هستند. به عبارت دیگر، مقدار جاری سری y_t ترکیب خطی از p تا از جدیدترین مقادیرهای گذشته سری به اضافه یک مؤلفه ε_t است. متغیر ε_t هر چیز جدید سری در زمان t را که توسط مقادیرهای گذشته تبیین نمی‌شود، وارد مدل می‌کند. با فرض مستقل بودن ε_t از $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ ، مدل خودبازگشتی مرتبه p مانا است اگر و تنها اگر هر یک از p ریشه معادله مشخصه AR از AR حث قدمطلق بزرگتر از ۱ باشد.

مدل خودبازگشتی برداری مجموعه‌ای از k متغیر به نام متغیرهای درون‌زا را در طول زمان توصیف می‌کند. مدل خودبازگشتی برداری $VAR(p)$ ،

$$y_t = C + \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \varepsilon_t, \quad t = 1, \dots, n, \quad n > 1, \quad (2)$$

را در نظر بگیرید که در آن Φ_i ها، ماتریس $(k \times k)$ بعدی از ضرایب هستند و ε_t یک بردار $(k \times 1)$ بعدی از خطاها با میانگین صفر و ماتریس کوواریانس Ω است. ماتریس کوواریانس Ω ماتریس قطعی مثبت است و

$$E(\varepsilon_{it} \varepsilon_{js}') = \begin{cases} \sigma_{ij} & t = s \\ 0 & t \neq s \end{cases}$$

مدل خودبازگشتی برداری (۲) را می‌توان به صورت

$$\Phi(B)y_t = C + \varepsilon_t$$

بازنویسی کرد که در آن $\Phi(B) = I_k - \Phi_1 B - \dots - \Phi_p B^p$ ، I_k و $B^j y_t = y_{t-j}$ است. مدل خودبازگشتی برداری (۲) یک مدل مانا است اگر و تنها اگر برای $|B| \leq 1$ ، $|\Phi(B)| \neq 0$ باشد.

مدل (۲) را در نظر بگیرید که در آن ε_t ها مستقل و هم‌توزیع هستند و از توزیع نرمال چندمتغیره، $N_k(0, \Omega)$ ، با تابع چگالی احتمال توأم

$$f_{y_t|z_t}^{\varepsilon_t}(y_t|z_t) = (2\pi)^{-\frac{k}{2}} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (y_t - \Pi' z_t)' \Omega^{-1} (y_t - \Pi' z_t)\right)$$

پیروی می‌کنند که در آن

$$\gamma = (C', \text{vec}(\Phi_1)', \dots, \text{vec}(\Phi_p)', \text{vec}(\Omega)')'$$

$$z_t = \begin{pmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}_{(kp+1) \times 1}$$

و

$$\Pi' = \begin{pmatrix} C & \Phi_1 & \Phi_2 & \dots & \Phi_p \end{pmatrix}_{k \times (kp+1)}$$

برآوردگرهای درستنمایی ماکسیمم پارامترهای مجهول مدل، با حل

معادلات درستنمایی به صورت

$$\hat{\Pi}'_n = \left(\sum_{t=1}^n y_t z_t' \right) \left(\sum_{t=1}^n z_t z_t' \right)^{-1}$$

و

$$\hat{\Omega}_n = \frac{1}{n} \sum_{t=1}^n \varepsilon_t \varepsilon_t'$$

قابل محاسبه هستند که در آن j -امین سطر از $\hat{\Pi}'_n$ برابر است با

$$u_j' \hat{\Pi}'_n = u_j' \left(\sum_{t=1}^n y_t z_t' \right) \left(\sum_{t=1}^n z_t z_t' \right)^{-1}$$

و u_j برداری از ۰ و ۱ است که عنصر j -ام این بردار ۱ است. برای مطالعه بیشتر مدل‌های خودبازگشتی برداری به وی [۲۳] مراجعه کنید.

۲.۲ مدل واریانس ناهمگن شرطی خودبازگشتی

تعمیم یافته

چون در بعضی از موارد واریانس شرطی سری به شرط معلوم بودن گذشته آن ثابت نیست، بنابراین با وجود اینکه سری ممکن است خوشه‌بندی ناپایداری را نمایش دهد اما اغلب دنباله‌ای به‌طور پیاپی ناهمبسته با میانگین صفر است. انگل [۱۳] مدل واریانس ناهمگن شرطی خودبازگشتی، $ARCH$ ، را برای مدل‌بندی تغییرپذیری واریانس پیشنهاد داد. مدل $ARCH(1)$ ،

$$y_t = \sigma_{t|t-1} \varepsilon_t$$

$$\sigma_{t|t-1} = \omega + \alpha y_{t-1}^2$$

را در نظر بگیرید که در آن $\alpha \geq 0$ و $\omega \geq 0$ پارامترهای مجهول هستند و $\{\varepsilon_t\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با میانگین صفر و واریانس یک است که به‌عنوان نوآوری‌ها نیز شناخته می‌شود. لینگ و مک آلیر [۲۰] نشان دادند که شرط $0 \leq \alpha < 1$ برای مانایی ضعیف مدل

درست‌نمایی ماکسیم پارامترهای مدل را بررسی کردند. مدل خودبازگشتی با مانده‌های واریانس‌ناهمگن شرطی خودبازگشتی تعمیم‌یافته $AR-GARCH$ ، را در نظر بگیرید که در آن

$$\sigma_{t-1}^2 = \omega + \alpha_1 y_{t-1} + \beta_1 \sigma_{t-1|t-2}^2$$

ϕ_0 و ϕ_1 پارامترهای مدل خودبازگشتی $\sigma_t^2 = \omega$ ، $\alpha_1 \geq 0$ و $\beta_1 \geq 0$ هستند.

از جمله مشکلات در تحلیل این نوع از مدل‌های سری زمانی، برآورد پارامترهاست. چون برآوردگرهای درست‌نمایی ماکسیم برای مدل $AR-GARCH$ جواب صریحی ندارند همچنین بعضی از خواص جانبی این برآوردگرها حفظ نمی‌شوند، روش‌های برآورد دیگری همچون روش گشتاوری، روش درست‌نمایی ماکسیم تعدیل‌یافته، EM -الگوریتم و غیره را می‌توان استفاده کرد.

۳ انتخاب متغیر و مرتبه مدل

انتخاب مدل در مسائل مختلف آماری از جمله تعیین تعداد عامل‌ها در تحلیل عاملی، انتخاب متغیر در مدل‌های رگرسیونی و انتخاب مرتبه در مدل‌های سری زمانی مورد استفاده قرار می‌گیرد. در تعیین یک مدل سری زمانی برای برازش به داده‌های مشاهده شده، شناخت مدل، انتخاب مرتبه مدل و توزیع خطاها از اهمیت فراوانی برخوردار است. با رسم نمودار داده‌ها معمولاً اطلاعاتی در مورد این‌که آیا سری زمانی شامل روند، نمود فصلی، نقاط بیرونی، واریانس غیرثابت و نامانایی است، حاصل می‌شود. این اطلاعات معمولاً مبنایی برای ایجاد یک تبدیل بر روی داده‌ها را فراهم می‌سازد. معمول‌ترین این تبدیلات که مورد استفاده قرار می‌گیرد تبدیل پایدار کردن واریانس (تبدیل باکس و کاکس) و تفاضلی کردن است. توابع خودهمبستگی و خودهمبستگی جزئی از ابزارهای مهم برای شناخت مدل و تعیین وابستگی بین مشاهدات هستند. برای انتخاب مدل بهینه در سری‌های زمانی، معمولاً فرض می‌شود که خطاها از توزیع نرمال با میانگین صفر و واریانس σ_ε^2 پیروی می‌کنند و انتخاب مرتبه مدل با استفاده از معیارهای اطلاع یا آزمون فرضیه‌ها انجام می‌شود. از جمله معیارهای اطلاع برای انتخاب مدل بهینه می‌توان به معیار اطلاع آکائیکه، AIC ، اشاره کرد که در مدل خودبازگشتی با مرتبه p و فرض توزیع نرمال، $N(0, \sigma_\varepsilon^2)$ ، برای مانده‌ها به صورت

$$AIC = n \log(\hat{\sigma}_\varepsilon^2) + 2(p+1)$$

$ARCH(1)$ لازم و کافی است. در مدل $ARCH(1)$ ، پیش‌بینی واریانس شرطی آینده تنها شامل جدیدترین توان دوم سری است. بولرسلف [۱۵] با وارد کردن p تأخیر از واریانس شرطی در مدل $ARCH(q)$ ، مدل واریانس‌ناهمگن شرطی خودبازگشتی تعمیم‌یافته $GARCH(p, q)$ ،

$$y_t = \sigma_{t-1} \varepsilon_t$$

$$\sigma_{t-1}^2 = \omega + \beta_1 \sigma_{t-1|t-2}^2 + \dots + \beta_p \sigma_{t-p|t-p-1}^2 + \alpha_1 y_{t-1}^2 + \dots + \alpha_q y_{t-q}^2$$

را معرفی کرد.

مدل‌های $GARCH$ در زمینه تجزیه و تحلیل سری‌های زمانی مالی توجه ویژه‌ای به خود جلب کرده‌اند. داده‌های سری‌های زمانی مالی مانند قیمت سهام و بازده‌ها شامل نوسانات زیاد و دم‌های سنگین هستند و فرکانس زیاد، میانگین و واریانس غیرثابت را نشان می‌دهند. فرآیندهایی با دم‌های سنگین را می‌توان برای برازش داده‌های سری‌های زمانی بازده سهام استفاده کرد. معمولاً ساختار مدل $GARCH$ بر این فرض استوار است که سری‌های زمانی مالی از توزیع نرمال (گاوسی) پیروی می‌کنند. در ادامه محاسبات برآوردگرهای درست‌نمایی ماکسیم برای حالت مدل $GARCH(1, 1)$ تشریح می‌شود. مدل $GARCH(1, 1)$ ،

$$y_t = \sigma_{t-1} \varepsilon_t$$

$$\sigma_{t-1}^2 = \omega + \beta_1 \sigma_{t-1|t-2}^2 + \alpha_1 y_{t-1}^2$$

را در نظر بگیرید. تابع چگالی احتمال شرطی برابر است با

$$f(y_t | y_{t-1}, \dots, y_1) = \frac{1}{\sqrt{2\pi\sigma_{t-1}^2}} \exp\left\{-\frac{y_t^2}{2\sigma_{t-1}^2}\right\},$$

بنابراین لگاریتم تابع درست‌نمایی شرطی عبارت است از:

$$L(\omega, \beta_1, \alpha_1) = -\frac{1}{2} \log(2) - \frac{1}{2} \sum_{t=1}^T \left\{ \log(\sigma_{t-1}^2) + \frac{y_t^2}{\sigma_{t-1}^2} \right\}.$$

با مشتق گرفتن از تابع $L(\omega, \beta_1, \alpha_1)$ نسبت به پارامترهای مجهول مدل و معادل قرار دادن این معادلات با صفر، برآوردگرهای درست‌نمایی ماکسیم پارامترهای مجهول مدل قابل محاسبه است. برآوردگرهای درست‌نمایی ماکسیم پارامترهای ω ، β_1 و α_1 فرم بسته‌ای ندارند، اما می‌توان آن‌ها را با استفاده از روش‌های عددی محاسبه کرد.

۳.۲ مدل خودبازگشتی با مانده‌های واریانس‌ناهمگن

شرطی خودبازگشتی تعمیم‌یافته

لینگ و لی [۱۸، ۱۹] مدل‌های خودبازگشتی-میانگین متحرک با مانده‌های $GARCH$ را پیشنهاد دادند و رفتار حدی برآوردگرهای

فرض کنید داده‌هایی که مدل برای یادگیری از آن‌ها استفاده می‌کند $\{(x_1, y_1), \dots, (x_n, y_n)\}$ باشد و هدف از یادگیری، کمینه کردن یک تابع زیان مانند $L(y, F(x))$ باشد. به عبارت دیگر

$$\hat{F} = \operatorname{argmin}_F E_{XY} (L(y, F(x))).$$

در روش L^2 -تقویتی، $L(y, F(x)) = (y - F(x))^2$ است. در این روش هدف، یادگیری مدلی مانند $F(x)$ برای کمینه کردن میانگین مربعات خطا است. الگوریتم L^2 -تقویتی، از روش حداقل مربعات برای برازش مدل مناسب بر متغیر وابسته y تنها با یک متغیر توضیحی x_j استفاده می‌کند. سپس به طور مکرر مانده‌های به دست آمده از مدل رگرسیون را به عنوان متغیر وابسته جدید y در نظر می‌گیرد و متغیر وابسته جدید را با تنها یک متغیر توضیحی x_j برازش می‌دهد. روند کلی روش L^2 -تقویتی در الگوریتم زیر توضیح داده شده است.

الگوریتم تقویتی- L^2 (بوهملمن و یو [۱۶]):

گام اول: با داده‌های آموزشی y_i شروع کنید.

گام دوم: برای $m = 1, \dots, M$

الف. برای $j = 1, \dots, k$ (برای هر متغیر)

(i) مدل رگرسیونی $y_i = \beta_{m,j} x_{ji} + u_i$ را با استفاده از روش حداقل مربعات برازش دهید.

(ii) $err_{m,j} = 1 - R_{m,j}^2$ را محاسبه کنید که در آن $R_{m,j}^2$ ضریب تعیین مدل رگرسیونی به دست آمده از روش حداقل مربعات است.

ب. $\hat{j}_m = \operatorname{argmin}_j err_{m,j}$ را پیدا کنید.

ج. قرار دهید $\hat{y}_i = \hat{\beta}_{m, \hat{j}_m} x_{\hat{j}_m, i}$ ، $i = 1, \dots, n$

گام سوم: مدل نهایی $\hat{F}_M(x) = \sum_{m=1}^M \hat{\beta}_{m, \hat{j}_m} x_{\hat{j}_m, i}$ است.

پارامتر توقف M با استفاده از روش‌های اعتبارسنجی یا برخی از معیارهای اطلاع در عمل انتخاب می‌شود.

۳.۳ الگوریتم تقویتی در مدل‌های خودبازگشتی برداری

مدل خودبازگشتی برداری مرتبه p ارائه شده در معادله (۲) را در نظر بگیرید. این مدل را می‌توان به صورت مدل رگرسیون چندمتغیره

$$y_t = \mathbf{x}_t \Phi + \varepsilon \quad (3)$$

نوشت که

$$\Phi = [C, \Phi_1, \dots, \Phi_p]', \mathbf{Xx}_t = [1, y_{t-1}, \dots, y_{t-p}]'. \quad (4)$$

محاسبه می‌شود که در آن $\hat{\sigma}_\varepsilon^2$ برآوردگر درست‌نمایی ماکسیمم σ_ε^2 است. شیباتا [۲۲] نشان داد که معیار AIC مرتبه مدل‌های خودبازگشتی را بیش برآورد می‌کند. آکائیکه [۱۲، ۱۳] تعمیم بیزی معیار اطلاع آکائیکه را به صورت

$$BIC = n \log(\hat{\sigma}_\varepsilon^2) - (n - p - 1) \log\left(1 - \frac{p+1}{n}\right) + (p+1) \log n + \log\left(\frac{\hat{\sigma}_\varepsilon^2 - 1}{\hat{\sigma}_\varepsilon^2 + 1}\right)$$

محاسبه کرد که در آن $\hat{\sigma}_\varepsilon^2$ واریانس نمونه‌ای سری است. در ادامه برخی از روش‌های انتخاب متغیر و انتخاب مرتبه مدل ارائه شده است.

۱.۳ آزمون علیت گرانجر

آزمون علیت گرانجر، یک آزمون فرضیه آماری برای تشخیص علیت میان سری‌های زمانی است. این آزمون بر این اصل استوار است که علت از نظر زمانی بر معلولش مقدم باشد. بنابراین هرگاه مقادیر گذشته سری زمانی $\{x_t\}$ در پیش‌بینی مقادیر آینده سری زمانی دیگر $\{y_t\}$ به طور معناداری کمک کند بیشتر از آنچه مقادیر گذشته خود $\{y_t\}$ می‌تواند کمک کند گوییم فرآیند $\{x_t\}$ در معیار گرانجر، علت فرآیند $\{y_t\}$ است. به عبارت دیگر، فرآیند $\{x_t\}$ علت فرآیند $\{y_t\}$ است اگر

$$\sigma^2(Y_t | z_t) > \sigma^2(Y_t | z_t, x_t)$$

که در آن $z_t = \{y_{t-1}, \dots, y_{t-p}\}$ و $x_t = \{x_{t-1}, \dots, x_{t-p}\}$ است. در عمل برای تشخیص علیت گرانجر میان دو سری زمانی $\{x_t\}$ و $\{y_t\}$ دو رگرسیون خطی

$$y_t = \sum_{i=1}^L \alpha_i y_{t-i} + \varepsilon_t$$

و

$$y_t = \sum_{i=1}^L \alpha_i y_{t-i} + \sum_{i=1}^L \beta_i x_{t-i} + \varepsilon_t$$

ایجاد می‌شود. اگر مدل دوم به طور معناداری مدل بهتری برای پیش‌بینی سری زمانی $\{y_t\}$ باشد، گوییم فرآیند $\{x_t\}$ علت فرآیند $\{y_t\}$ است. آزمون علیت گرانجر برای بیش از دو سری زمانی توسط مدل خودبازگشتی برداری انجام می‌گیرد، به گرانجر [۱۷] مراجعه کنید.

۲.۳ الگوریتم L^2 -تقویتی

در یادگیری ماشین، هدف پیش‌بینی یا رده‌بندی است. مانند سایر روش‌های تقویت، روش تقویت گرادیان ترکیب خطی از یک سری مدل‌های ضعیف برای ایجاد یک مدل قوی و کارآمد است.

۱) مدل خودبازگشتی مرتبه p ، $\hat{\mathbf{R}}_t^{(m)} = \mathbf{x}_{(j)} \Phi'_{(j)} + \boldsymbol{\varepsilon}_t$ را با استفاده از روش درست‌نمایی ماکسیمم برازش دهید که در آن اگر $m = 1$ باشد، آنگاه $\hat{\mathbf{R}}_t^{(1)} = \mathbf{y}_t = \mathbf{y}_t$.

۲) $err_{m_j} = AIC_{m_j}$ را محاسبه کنید که در آن AIC_{m_j} معیار اطلاع آکائیکه مدل به دست آمده با استفاده از روش درست‌نمایی ماکسیمم است.

ب. پیدا کنید $\hat{j}_m = \operatorname{argmin}_j AIC_{m_j}$.

ج. قرار دهید $\hat{\mathbf{R}}^{(m+1)} \leftarrow \hat{\mathbf{R}}^{(m)} - \mathbf{x}_{(\hat{j}_m)} \Phi'_{(\hat{j}_m)}$.

د. اگر معیار اطلاع آکائیکه مدل جدید، $F_{m+1}(x)$ ، از معیار اطلاع آکائیکه مدل قبلی، $F_m(x)$ ، کوچک‌تر باشد، مؤلفه جدید را به مدل قبلی اضافه کنید و به گام دوم بروید در غیر این صورت به گام سوم بروید.

در گام سوم: مدل نهایی $F_M(x) = \sum_{m=1}^M \mathbf{x}_{(\hat{j}_m)} \Phi'_{(\hat{j}_m)}$ استفاده شده است.

درگام الف (i) مدل خودبازگشتی برداری تنها با یک متغیر توضیحی با مرتبه بهینه بر اساس معیار آکائیکه برآورد می‌شود و در گام الف (ii) مدل خودبازگشتی برداری بهینه (متغیرتوضیحی بهینه) انتخاب می‌شود. در گام (د)، اگر معیار اطلاع مدل خودبازگشتی برداری با مولفه جدید کمتر از معیار اطلاع آکائیکه مدل خودبازگشتی قبلی باشد مولفه جدید به مدل اضافه می‌شود در غیر این صورت الگوریتم متوقف می‌شود. اگر چه در الگوریتم‌های تقویتی از معیارهای اطلاع و اعتبارسنجی مختلفی همچون معیار اطلاع آکائیکه، معیار اطلاع بیز، میانگین حداقل مربعات و غیره برای انتخاب پارامتر توقف استفاده می‌شود اما در این الگوریتم پارامتر توقف با استفاده از معیار اطلاع آکائیکه تعیین می‌شود.

۴ تحلیل داده‌ها

در این مقاله جامعه آماری، کلیه شرکت‌های پذیرفته شده در بورس اوراق بهادار ایران است و ۴۰ شرکت به عنوان نمونه آماری به صورت تصادفی انتخاب شده است. نام شرکت‌های انتخاب شده به عنوان نمونه در پیوست ارائه شده است. این داده‌ها در سایت

<http://www.tsetmc.com/Loader.aspx?ParTree=15>

وجود دارند که شامل اطلاعاتی همچون نماد (نام) شرکت، تاریخ، تعداد، حجم معاملات، ارزش، اولین قیمت، بیشترین قیمت، کمترین قیمت و قیمت روز قبل‌تر هستند. داده‌ها به صورت روزانه از تاریخ ۲۶ فوریه

ایده اصلی الگوریتم تقویتی در مدل‌های خودبازگشتی، انتخاب بازگشتی ستونی از متغیرهای توضیحی در \mathbf{x}_t است که بهترین برازش را برای \mathbf{y}_t در اولین مرحله تقویت یا بر مانده‌های فعلی در مراحل تقویت بعدی ارائه می‌دهد. در هر گام تقویت، می‌توان ستونی از \mathbf{x}_t را انتخاب کرد که بهترین برازش را برای ستون انتخابی از \mathbf{y}_t یا ماتریس مانده‌ها ارائه دهد و عنصری از ماتریس Φ را با استفاده از روش حداقل مربعات به‌روزرسانی کند، به لوتز و بوهلمن [۲۱] مراجعه کنید. از طرف دیگر، می‌توان از الگوریتم تقویتی لوتز و بوهلمن [۲۱] نیز برای انتخاب ستونی در \mathbf{x}_t استفاده کرد که بهترین برازش را برای تمامی ماتریس \mathbf{y}_t یا ماتریس مانده‌ها ارائه دهد و یک ردیف از ورودی‌ها Φ را به‌روزرسانی کند.

در این مقاله مشابه با الگوریتم تقویتی لوتز و بوهلمن [۲۱]، الگوریتم تقویتی به منظور انتخاب متغیر معرفی شده با این تفاوت که در این الگوریتم از تابع درست‌نمایی ماکسیمم به منظور برآورد پارامترها و معیار اطلاع آکائیکه به منظور انتخاب پارامتر توقف استفاده شده است. این الگوریتم در هر گام الگوریتم، متغیری با تمام تأخیرهای انتخاب می‌شود که بهترین برازش را ایجاد کند. چون هر متغیر با تمام تأخیرهایش را به عنوان یک گروه انتخاب می‌کنیم، تنظیم مجدد ستون‌ها در معادله (۳) مفید خواهد بود، به طوری که آن‌ها بر اساس متغیرهای

$$\mathbf{x}_g = [\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(j)}, \dots, \mathbf{x}_{(k)}]_{n \times kp}$$

گروه‌بندی می‌شوند که در آن $\mathbf{x}_{(j)}$ ماتریسی $n \times p$ بعدی است. فرآیند $VAR(p)$ ارائه شده در معادله (۳) را می‌توان به صورت

$$\mathbf{y}_t = \mathbf{x}_g \Phi_g + \boldsymbol{\varepsilon}_t$$

نوشت که در آن $\Phi_g = [\Phi_{(1)}, \dots, \Phi_{(j)}, \dots, \Phi_{(k)}]'$ و $\Phi_{(j)}$ ماتریس ضرایب $p \times k$ بعدی هست. فرض کنید m نشان‌دهنده مرحله تکرار، $m = 1, \dots, M$ پارامتر توقف باشد که با استفاده از معیار اطلاع آکائیکه مشخص می‌شود. چون تمام تأخیرهای یک متغیر در هر گام انتخاب می‌شود، بنابراین مدل در مرحله m برای متغیر j به صورت

$$\hat{\mathbf{R}}_t^{(m-1)} = \mathbf{x}_{(j)} \Phi'_{(j)} + \boldsymbol{\varepsilon}_t$$

که در آن اگر $m = 1$ باشد، آنگاه $\hat{\mathbf{R}}_t^{(0)} = \mathbf{y}_t$ است. روند کلی این الگوریتم در زیر توضیح داده شده است.

گام اول: با داده‌های آموزشی \mathbf{y}_i شروع کنید.

گام دوم: برای $m = 1, \dots, M$

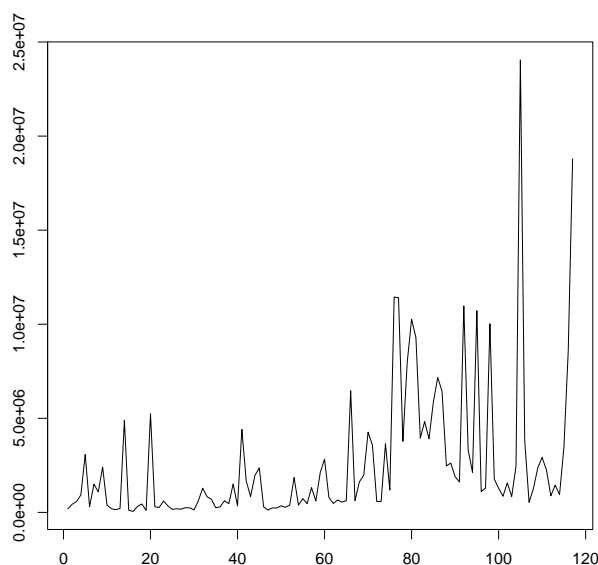
الف. برای $j = 1, \dots, k$ (برای هر متغیر)

غیرشرطی است که از نظر آماری با صفر تفاوت معناداری دارند. همچنین سری دارای چولگی مثبت است و در نهایت، سری دارای دم‌های سنگین است چون مقدار ضریب چولگی این سری مقدار مثبتی است. با توجه به اینکه p -مقدار آزمون شاپیرو کوچک‌تر از 0.05 است، بنابراین فرض نرمال بودن داده‌ها رد می‌شود. نمودار سری زمانی این داده‌ها در شکل ۱ ارائه شده است که نشان دهنده نامانای بودن سری است.

۱۰ سپتامبر ۲۰۲۳ تا ۲۰۲۳ با تعداد $n = 117$ مشاهده، جمع‌آوری شده‌اند. صندوق سرمایه‌گذاری انارنماد ارزش-درسهام (انار) به‌عنوان شرکت هدف انتخاب شده است. هدف از این پژوهش، تحلیل داده‌های حجم معاملاتی این شرکت با استفاده از مدل‌های سری زمانی و روش‌های انتخاب متغیر و مرتبه مدل است. حجم معاملاتی شرکت انتخاب شده با نماد y_t و حجم معاملاتی سایر شرکت‌های نمونه با $x_{1,t}, \dots, x_{40,t}$ نمایش داده شده است. آماره‌های توصیفی سری y_t در جدول ۱ نمایش داده شده است. از جدول ۱ مشاهده می‌شود که سری دارای میانگین

جدول ۱. آماره‌های توصیفی سری y_t .

مینیم	چارک اول	میانه	میانگین	چارک سوم	ماکسیمم	ضریب چولگی	ضریب کشیدگی	مقدار احتمال آزمون شاپیرو
۵۱۷۹۹	۳۹۱۷۱۰	۱۱۰۶۹۶۳	۲۵۴۵۳۸۰	۲۹۱۹۷۵۸	۲۴۰۴۶۰۱۵	۲/۹۴۳۲	۱۰/۸۵۱۹	< 0.05

شکل ۱. نمودار سری زمانی داده‌های حجم معاملاتی y_t .

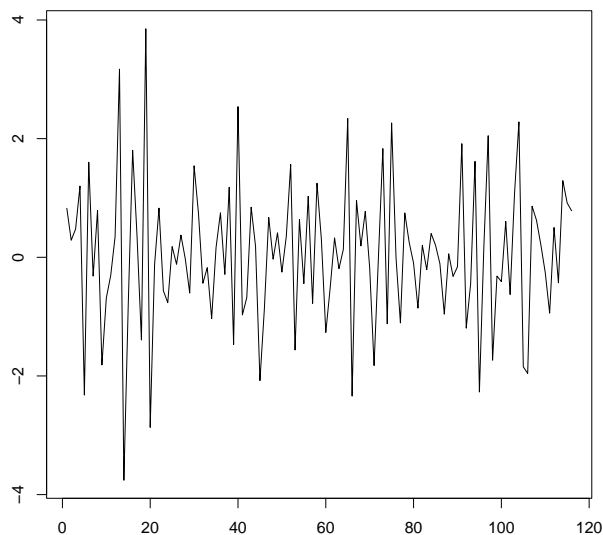
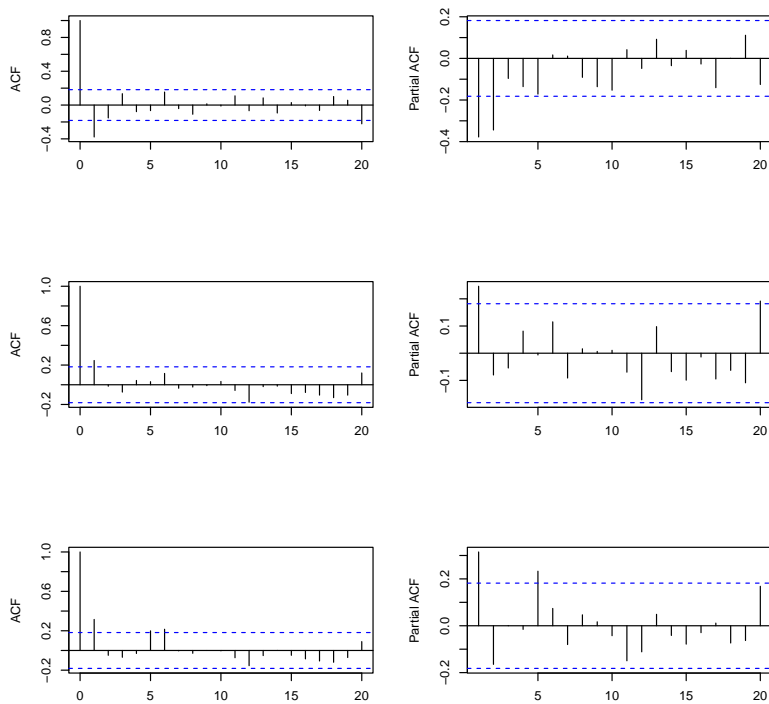
دارای ضریب چولگی نزدیک به صفر است. p -مقدار آزمون شاپیرو بزرگ‌تر از 0.05 است، بنابراین دلیلی برای رد فرض نرمال بودن داده‌ها نداریم. p -مقدار آزمون دیکی فولر برابر با 0.01 است، به‌عبارت دیگر فرض مانا بودن سری تبدیل‌یافته را می‌پذیریم.

با توجه به شکل ۲ مشاهده می‌شود که تلاطم در بعضی دوره‌ها زیاد و در بعضی دوره‌ها کم است که نشان‌دهنده ویژگی خوشه‌بندی تلاطم در سری زمانی تبدیل‌یافته است و نیاز به مدل‌بندی تلاطم دارد. همچنین روند ثابت مقادیر سری در طول زمان نشان دهنده وجود مانایی در میانگین سری است.

فرآیند $y_t^* = \nabla \log y_t$ را در نظر بگیرید که در آن y_t نشان دهنده حجم معاملاتی صندوق سرمایه‌گذاری انارنماد ارزش-درسهام و y_t^* نشان دهنده داده‌های تبدیل‌یافته مربوط به حجم معاملاتی صندوق سرمایه‌گذاری انارنماد ارزش-درسهام هستند و $\nabla y_t = y_t - y_{t-1}$. آماره‌های توصیفی و نمودار سری زمانی داده‌های تبدیل‌یافته، y_t^* به‌ترتیب در جدول ۲ و شکل ۲ نمایش داده شده‌اند. با استفاده از جدول ۲ مشاهده می‌شود که سری دارای میانگین غیرشرطی است که از نظر آماری با صفر تفاوت معناداری ندارند، p -مقدار آزمون تی جهت بررسی فرضیه صفر بودن میانگین سری برابر با 0.7292 است. همچنین سری

جدول ۲. آماره‌های توصیفی سری y_t^* .

مقدار احتمال آزمون شاپیرو	ضریب کشیدگی	ضریب چولگی	ماکسیمم	چارک سوم	میانگین	میانه	چارک اول	مینیمم
۰/۴۶۹۹	۰/۷۴۹۹	۰/۰۱۹۵	۳/۸۵۱۶	۰/۷۵۵۵	۰/۰۳۹۶	۰/۰۳۶۰	-۰/۶۴۰۲	-۳/۷۵۷۷

شکل ۲. نمودار سری زمانی داده‌های تبدیل یافته y_t^* .شکل ۳. همبستگی‌نگار (نمودارهای سمت چپ) و نمودار خودهمبستگی جزئی (نمودارهای سمت راست) سری زمانی، قدرمطلق و توان دوم سری تبدیل یافته y_t^* (به ترتیب از بالا به پایین).

است، می‌توان دریافت که همبستگی بین مقادیر توان دوم سری وجود دارد که مبین نیاز به مدل‌بندی تلاطم است. بنابراین یکی از مدل‌های پیشنهادی می‌تواند مدل $AR(2) + GARCH(1, 1)$ باشد.

مدل $AR(2) + GARCH(1, 1)$ ،

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \sigma_{t|t-1} \varepsilon_t$$

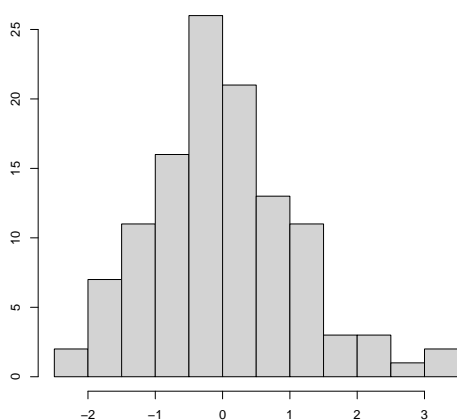
$$\sigma_{t|t-1}^2 = \omega + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1|t-2}^2$$

را در نظر بگیرید که در آن ϕ_1 ، ϕ_2 پارامترهای مدل خودبازگشتی، $\omega > 0$ ، $\alpha_1 \geq 0$ ، $\beta_1 \geq 0$ هستند و ε_t از توزیع نرمال استاندارد پیروی می‌کند. پارامترهای مدل با استفاده از روش درست‌نمایی ماکسیمم برآورد و نتایج مقادیر برآوردگرها، مقادیر معیار اطلاع آکائیکه، AIC ، میانگین توان دوم خطاها، MSE و p -مقدار آزمون کلموگروف-اسمیرنف، KS ، مدل پیشنهاد شده در جدول ۳ ارائه شده است. p -مقدار آزمون عدم معناداری پارامترهای مدل در پراونت گزارش شده است. مشاهده می‌شود که همه p -مقدارها بجز ϕ_0 کمتر از ۰/۰۵ است، بنابراین تمام پارامترها معنادار هستند.

همبستگی‌نگار و نمودار خودهمبستگی جزئی سری زمانی تبدیل‌یافته y_t^* (نمودارهای بالا)، همبستگی‌نگار و نمودار خودهمبستگی جزئی قدرمطلق سری زمانی تبدیل‌یافته y_t^* (نمودارهای وسط) و همبستگی‌نگار و نمودار خودهمبستگی جزئی توان دوم سری زمانی تبدیل‌یافته y_t^* (نمودارهای پایین) در شکل ۳ نمایش داده شده است. همبستگی‌نگار و نمودار خودهمبستگی جزئی سری زمانی تبدیل‌یافته y_t^* (نمودارهای بالا در شکل ۳) نشان‌دهنده وجود همبستگی در سری زمانی y_t^* است. با توجه به این‌که مقادیر خودهمبستگی جزئی سری زمانی تبدیل‌یافته y_t^* بعد از تأخیر ۲ داخل باند اطمینان قرار گرفته است، بنابراین یکی از مدل‌های پیشنهادی می‌تواند مدل خودبازگشتی $AR(2)$ باشد. در قسمت وسط و پایین شکل ۳ به ترتیب همبستگی‌نگار و نمودار خودهمبستگی جزئی قدرمطلق و توان دوم سری زمانی تبدیل‌یافته y_t^* نمایش داده شده است. با توجه به شکل مشهود است که همبستگی بین آن‌ها وجود دارد. مدل‌های تلاطم چنین همبستگی در سری زمانی را بیان می‌کند. برای تشخیص وجود مدل تلاطم نیز از آزمون لیونگ-باکس استفاده می‌شود. با توجه به این‌که p -مقدار آزمون لیونگ-باکس برای توان دوم سری زمانی تبدیل‌یافته y_t^* ، ۰/۰۰۹۵، کوچک‌تر از ۰/۰۵

جدول ۳. برآورد پارامترهای مدل $AR(2) + GARCH(1, 1)$.

KS	MSE	AIC	$\hat{\beta}_1$	$\hat{\alpha}_1$	$\hat{\omega}$	$\hat{\phi}_2$	$\hat{\phi}_1$	$\hat{\phi}_0$	
۰/۳۰۵۸	۱/۱۴۰۴	۳۵۰/۵۰۶	۰/۷۹۴۷	۰/۰۳۳۰	۰/۱۹۳۲	-۰/۳۵۲۳	-۰/۵۱۶۷	۰/۰۵۳۰	برآورد
			۰/۲۸۴۸	۰/۰۶۲۴	۰/۳۰۳۸	۰/۰۸۷۲	۰/۰۸۷۲	۰/۰۹۹۲	انحراف استاندارد
			(۰/۰۰۵۹)	(۰/۰۴۲۶)	(۰/۰۴۸۷)	(۵/۳۶e-۰۵)	(۳/۰۹e-۰۹)	(۰/۰۵۹۳)	p -مقدار



شکل ۴. هیستوگرام مانده‌های مدل $AR(2) + GARCH(1, 1)$.

از آزمون لیونگ-باکس جهت تشخیص ناهمبسته بودن مانده‌ها و از نمودار هیستوگرام و آزمون کلموگروف-اسمیرنف جهت تشخیص نرمال بودن مانده‌ها استفاده شده است. با توجه به p -مقدار آزمون لیونگ-باکس برای مانده‌ها، ۰/۵۰۱۷، می‌توان دریافت که همبستگی بین مقادیر سری مانده‌ها وجود ندارد. همچنین با توجه به p -مقدار آزمون کلموگروف-اسمیرنف، ۰/۳۰۵۸، می‌توان دریافت که مانده‌ها از توزیع نرمال پیروی می‌کنند. شکل ۴ این نتیجه را تأیید می‌کند. به‌منظور انتخاب متغیر، ضریب همبستگی بین سری y_t^* و سری‌های $x_{1,t}, \dots, x_{4,t}$ در نمودار ۵ ارائه شده است. نتایج آزمون ضریب همبستگی نشان می‌دهد که همبستگی بین سری y_t^* و سری‌های

$$x_{3,t}, x_{4,t}, x_{5,t}, x_{6,t}, x_{8,t}, x_{13,t}, x_{16,t}, x_{19,t}, x_{30,t}$$

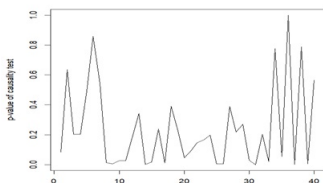
$$x_{22,t}, x_{23,t}, x_{34,t}, x_{35,t}$$

وجود دارد.

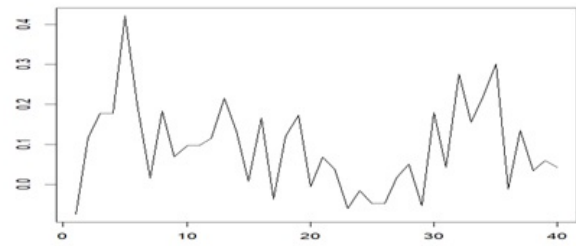
و ε_t برداری از مانده‌ها است. مرتبه‌های $p = 1, 2, \dots, 12$ به‌عنوان مرتبه‌های پیشنهادی در نظر گرفته شده‌اند. پارامترهای مجهول مدل‌های پیشنهادی با استفاده از روش درست‌نمایی ماکسیمم برآورد و مدل بهینه با استفاده از معیار اطلاع آکائیکه، AIC ، انتخاب شده است. مدل خودبازگشتی برداری برآورد شده مرتبه چهارم، $VAR(4)$ ، دارای کوچک‌ترین مقدار معیار اطلاع آکائیکه است، بنابراین مدل خودبازگشتی برداری $VAR(4)$ با استفاده از معیار اطلاع آکائیکه به‌عنوان مدل بهینه انتخاب می‌شود. نتایج مقادیر پارامترهای برآورد شده و مقادیر معیار اطلاع آکائیکه $VAR(4)$ در جدول ۴ ارائه شده‌اند. از جدول ۴ مشاهده می‌شود که متغیرهای $x_{t-1,30}$ ، $x_{t-2,33}$ و $x_{t-4,8}$ معنادار هستند. در جدول ۴ پارامترهای معنادار با علامت * در ستون آخر مشخص شده‌اند. مدل برازش داده شده به‌صورت

$$y_t^* = -0.7499y_{t-1}^* + 0.7766x_{t-1,30} - 0.5139y_{t-2}^* + 0.6592x_{t-2,33} - 0.2970x_{t-4,8}$$

حاصل می‌شود که در آن مانده‌ها از توزیع نرمال با میانگین 0.0000 و واریانس 0.7615 پیروی می‌کنند. از مقایسه جدول‌های ۳ و ۴ مشاهده می‌شود که مدل $VAR(4)$ برآورد شده نسبت به مدل $GARCH(1,1) + AR(2)$ برآورد شده دارای مقادیر AIC و MSE کوچک‌تری هستند.



شکل ۶. نمودار p -مقدار آزمون علیت y_t^* و سری‌های $x_{1,t}, \dots, x_{40,t}$.



شکل ۵. نمودار ضریب همبستگی بین سری y_t^* و سری‌های

$$x_{1,t}, \dots, x_{40,t}$$

در این بخش، علاقه‌مند به بررسی آزمون علیت گرنجر بین سری y_t^* و سری‌های $x_{1,t}, \dots, x_{40,t}$ هستیم. مشاهده می‌شود که علیت معناداری در سطح ۵ درصد بین سری y_t^* و سری‌های

$$x_{8,t}, x_{9,t}, x_{10,t}, x_{14,t}, x_{15,t}, x_{17,t}, x_{20,t},$$

$$x_{25,t}, x_{26,t}, x_{30,t}, x_{31,t}, x_{33,t}, x_{37,t}, x_{39,t} \quad (5)$$

وجود دارد. توجه کنید که این آزمون برای مدل‌های خودبازگشتی مرتبه اول، دوم و سوم در نظر گرفته شده است. نتایج این آزمون برای هر سه مدل خودبازگشتی یکسان بوده‌اند. p -مقدار آزمون علیت برای مدل خودبازگشتی مرتبه دوم در شکل ۶ ارائه شده است. با توجه به شکل‌های ۵ و ۶ مشاهده می‌شود که از سری‌های (۵) $x_{8,t}, x_{9,t}, x_{10,t}, x_{14,t}, x_{15,t}, x_{17,t}, x_{20,t}, x_{25,t}, x_{26,t}, x_{30,t}, x_{31,t}, x_{33,t}, x_{37,t}, x_{39,t}$ باعث ایجاد سری y_t^* شده‌اند، بنابراین مدل خودبازگشتی برداری، $VAR(p)$

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + \varepsilon_t, \quad t = 1, \dots, n,$$

برای برازش به این داده‌ها در نظر گرفته که در آن

$$Y_t = (Y_t^*, x_{8,t}, x_{9,t}, x_{10,t}, x_{14,t}, x_{15,t}, x_{17,t}, x_{20,t},$$

$$\Phi_i = (\phi_{i,1}, \phi_{i,8}, \phi_{i,9}, \phi_{i,10}, \phi_{i,14}, \phi_{i,15}, \phi_{i,17}, \phi_{i,20},$$

جدول ۴. مقادیر پارامترهای برآورد شده و مقدار معیار اطلاع آکائیکه مدل $VAR(4)$.

مقدار p	آماره آزمون	انحراف استاندارد	برآورد ضرایب	
۰/۷۲۵۲	۰/۳۵۲۰	۰/۰۸۹۶	۰/۰۳۱۶	ϕ_0
* < ۰/۰۰۵	-۷/۲۹۷۰	۰/۱۰۲۷	-۰/۷۴۹۹	$\phi_{1,1}$
۰/۱۵۵۳	۱/۴۳۲۰	۰/۰۹۲۸	۰/۱۳۳۰	$\phi_{1,2}$
* ۰/۰۰۵۱	۲/۸۶۶۰	۰/۲۷۱۰	۰/۷۷۶۶	$\phi_{1,3}$
۰/۲۶۵۵	۱/۱۲۰۰	۰/۲۳۲۶	۰/۲۶۰۴	$\phi_{1,33}$
* ۰/۰۰۰۲	-۳/۸۳۴۰	۰/۱۳۴۰	-۰/۵۱۳۹	$\phi_{2,1}$
۰/۸۶۹۵	-۰/۱۶۵۰	۰/۱۱۷۲	-۰/۰۱۹۳	$\phi_{2,2}$
۰/۲۱۷۹	۱/۲۴۰۰	۰/۲۷۳۰	۰/۳۳۸۵	$\phi_{2,3}$
* ۰/۰۱۳۵	۲/۵۱۶۰	۰/۲۶۲۰	۰/۶۵۹۲	$\phi_{2,33}$
۰/۲۸۸۲	-۱/۰۶۸۰	۰/۱۲۹۷	-۰/۱۳۸۵	$\phi_{3,1}$
۰/۰۵۲۹	-۱/۹۶۰۰	۰/۱۱۸۵	-۰/۲۳۲۳	$\phi_{3,2}$
۰/۹۵۸۹	۰/۰۵۲۰	۰/۲۶۶۹	۰/۰۱۳۷	$\phi_{3,3}$
۰/۵۶۱۴	۰/۵۸۳۰	۰/۲۷۶۴	۰/۱۶۱۰	$\phi_{3,33}$
۰/۹۵۱۲	-۰/۰۶۱۰	۰/۱۰۱۲	-۰/۰۰۶۲	$\phi_{4,1}$
* ۰/۰۰۰۲۶	-۳/۰۸۹۰	۰/۰۹۶۱	-۰/۲۹۷۰	$\phi_{4,2}$
۰/۲۱۹۲	۱/۲۳۷۰	۰/۲۵۷۹	۰/۳۱۹۰	$\phi_{4,3}$
۰/۱۵۱۶	۱/۴۴۵۰	۰/۲۴۲۶	۰/۳۵۰۷	$\phi_{4,33}$
$AIC = 322,3245 \quad MSE = 0,7547 \quad KS = 0,6725$				

واریانس $0/8161$ پیروی می‌کنند.

از مقایسه جدول‌های ۴ و ۵ مشاهده می‌شود که مدل $VAR(2)$ برآورد شده با استفاده از الگوریتم تقویتی دارای مقادیر AIC و MSE کوچک‌تری هستند.

از آزمون لیونگ-باکس جهت تشخیص ناهمبسته بودن مانده‌ها و از نمودار هیستوگرام و آزمون کلموگروف-اسمیرنف جهت تشخیص نرمال بودن مانده‌ها استفاده شده است. با توجه به p -مقدار آزمون لیونگ-باکس برای مانده‌ها، $0/7397$ ، می‌توان دریافت که همبستگی بین مقادیر سری مانده‌ها وجود ندارد. همچنین با توجه به p -مقدار آزمون کلموگروف-اسمیرنف، $0/97543594$ ، می‌توان دریافت که مانده‌ها از توزیع نرمال پیروی می‌کنند. جهت تشخیص وجود مدل تلاطم از آزمون لیونگ-باکس برای قدرمطلق و توان دوم سری مانده‌ها استفاده شده است. با توجه به این که p -مقدار آزمون لیونگ-باکس به ترتیب برای قدرمطلق، $0/9754$ ، و توان دوم سری مانده‌ها، $0/3941$ ، بزرگ‌تر از $0/05$ است، می‌توان دریافت که همبستگی بین مقادیر توان دوم سری وجود ندارد، بنابراین نیاز به مدل‌بندی تلاطم نیست.

در ادامه با استفاده از الگوریتم تقویتی، مدل با مرتبه بهینه انتخاب می‌شود. با استفاده از این الگوریتم به انتخاب متغیر پرداخته و مدل خودبازگشتی برداری بهینه برازش داده می‌شود، به طوری که مدل خودبازگشتی برداری $VAR(2)$ با متغیرهای $x_{10,t}, x_{30,t}, x_{33,t}, x_{31,t}, x_{10,t-1}, x_{30,t-1}, x_{33,t-1}, x_{31,t-2}, x_{31,t-2}$ به عنوان مدل بهینه انتخاب و نتایج مدل برآورد شده در جدول ۵ ارائه شده است. در هر گام الگوریتم، متغیری با تمام تأخیرهایش انتخاب می‌شود که بهترین برازش را ایجاد می‌کند. انتخاب مرتبه مدل همزمان با انتخاب متغیر به عنوان بخشی از فرآیند انتخاب الگوریتم تقویتی بوده است. از جدول ۵ مشاهده می‌شود که ۵ پارامترهای معنادار با علامت * در ستون آخر مشخص شده‌اند. مدل برازش داده شده به صورت

$$y_t^* = -0/6812y_{t-1}^* + 0/4828x_{t-1,10} + 0/5582x_{t-1,30} - 0/4426y_{t-2}^* + 0/5765x_{t-2,33} - 0/2889x_{t-2,31}$$

نوشته می‌شود که در آن مانده‌ها از توزیع نرمال با میانگین $0/0294$ و

جدول ۵. مقادیر پارامترهای برآورد شده و مقدار معیار اطلاع آکائیکه مدل $VAR(2)$.

برآورد ضرایب	انحراف استاندارد	آماره آزمون	p -مقدار	
$\phi_{1,1}$	۰/۰۸۷۵	-۷۷۸۳۰	< ۰/۰۵	*
$\phi_{1,10}$	۰/۱۶۸۴	۲,۸۷۳۰	۰/۰۰۴۹	*
$\phi_{1,30}$	۰/۲۵۱۱	۲,۲۲۳۰	۰/۰۲۸۳	*
$\phi_{1,33}$	۰/۲۱۸۳	۱,۵۷۴۰	۰/۱۱۸۵	
$\phi_{1,31}$	۰/۰۸۶۹	۰,۸۸۴۰	۰/۳۷۸۸	
$\phi_{2,1}$	۰/۰۸۶۰	-۵,۱۴۵۰	< ۰/۰۵	*
$\phi_{2,10}$	۰/۱۶۹۴	۰,۷۱۶۰	۰/۴۷۵۳	
$\phi_{2,30}$	۰/۲۵۴۰	۰,۸۹۴۰	۰/۳۷۳۵	
$\phi_{2,33}$	۰/۲۱۵۴	۲,۶۷۶۰	۰/۰۰۸۶	*
$\phi_{2,31}$	۰/۰۸۹۷	۳,۲۲۰۰	۰/۰۰۱۷	*
$AIC = ۳۲۱,۴۷۹۷$ $MSE = ۰,۷۰۹۸$ $KS = ۰,۳۵۹۴$				

۵ بحث و نتیجه‌گیری

ایران بر حجم معاملاتی شرکت مورد مطالعه است. انتخاب مدل خودبازگشتی برداری به‌عنوان مدل بهینه برای حجم معاملاتی شرکت هدف می‌تواند به این مفهوم باشد که حجم معاملاتی شرکت‌های مورد مطالعه بازار بورس می‌تواند تأثیر زیادی بر استراتژی‌های سرمایه‌گذاری در شرکت هدف داشته باشند. دلایل احتمالی این تأثیرگذاری می‌تواند تعلق شرکت هدف و شرکت‌های مورد مطالعه به صنعتی خاص، لیدر بودن برخی از این شرکت‌ها در بازار و غیره باشند. سرمایه‌گذاران می‌توانند با توجه به تغییرات در حجم معاملات شرکت هدف و شرکت‌های مورد مطالعه، استراتژی‌های خود را تنظیم کنند. اخبار کلان اقتصادی و تغییرات در سیاست‌های پولی شرکت‌های مورد مطالعه می‌تواند تأثیر زیادی بر حجم معاملاتی شرکت هدف داشته باشند. به‌عنوان مثال، اعلام نرخ بهره جدید یا تغییرات در سیاست‌های مالی شرکت‌های مورد مطالعه می‌تواند باعث افزایش یا کاهش حجم معاملاتی شرکت هدف شود.

در این مطالعه داده‌های مربوط به حجم معاملاتی شرکت‌های پذیرفته شده در بورس اوراق بهادار ایران در نظر گرفته شد. مدل‌های خودبازگشتی-واریانس‌ناهمگن شرطی خودبازگشتی تعمیم‌یافته $(AR + GARCH)$ و خودبازگشتی برداری به‌عنوان مدل‌های پیشنهادی در نظر گرفته شدند. با استفاده از روش‌های معمول آماری و همچنین روش‌های یادگیری ماشین از جمله الگوریتم تقویتی به انتخاب متغیر و مدل پرداخته شد. با مقایسه مقادیر معیار اطلاع آکائیکه و میانگین توان دوم خطاهای مدل‌های برازش داده شده، مشاهده گردید که مدل خودبازگشتی برداری برآورد شده با استفاده از الگوریتم تقویتی دارای کم‌ترین مقادیر معیار اطلاع آکائیکه و میانگین توان دوم خطاها است. به عبارت دیگر الگوریتم تقویتی عملکرد مناسب‌تری در تحلیل این سری از داده‌ها دارد. انتخاب مدل خودبازگشتی برداری به‌عنوان مدل بهینه نشان دهنده تأثیرگذاری حجم معاملات برخی از شرکت‌های پذیرفته شده در بورس اوراق بهادار

مراجع

- [۱] حکیمیان، حسن؛ احمدی، زانبار. (۱۳۹۸). بررسی تأثیر بلندمدت بازده بازارهای موازی بر بازده شاخص کل بورس اوراق بهادار تهران. فصلنامه بورس اوراق بهادار، ۱۲(۴۵)، ۶۸-۹۴.
- [۲] ذوالفقاری، مهدی؛ سبحانی، بهرام؛ بختیاری، محمدجواد. (۱۳۹۹). طراحی مدلی جهت پیش‌بینی بازده شاخص کل بورس اوراق بهادار با تأکید بر مدل‌های ترکیبی شبکه یادگیری عمیق و مدل‌های خانواده $GARCH$. مهندسی مالی و مدیریت اوراق بهادار (مدیریت پرتفوی)، ۱۱(۴۲)، ۱۳۸-۱۷۱.

- [۳] رادمهر، فرید؛ شمس قارنه، ناصر. (۱۳۹۱). پیش‌بینی شاخص بورس تهران با استفاده از سری زمانی فازی بر اساس تعریف نرخ بازده. فصلنامه علمی پژوهشی دانش سرمایه‌گذاری، ۱(۴)، ۱۷۵-۱۹۳.
- [۴] رستمی، ژیلا؛ فتاحی، شهرام؛ سهیلی، کیومرث. (۱۴۰۱). مدل‌سازی و تخمین بازده بورس اوراق بهادار تهران با استفاده از مدل‌های پویا. فصلنامه اقتصاد مالی، ۱۷(۱)، ۱۸۵-۲۱۶.
- [۵] سرلک، احمد؛ محمدی، محمد؛ برزگر، محمد. (۱۳۹۵). ارزیابی مقایسه‌ای دقت مدل‌های سری زمانی AR ، MA ، $ARMA$ ، $ARIMA$ در پیش‌بینی شاخص سهام شرکت‌های بورس اوراق بهادار تهران. فصلنامه پژوهش‌های جدید در مدیریت و حسابداری، ۲(۳).
- [۶] غلامیان، الهام؛ داوودی، سید محمدرضا. (۱۳۹۷). پیش‌بینی روند قیمت در بازار سهام با استفاده از الگوریتم جنگل تصادفی. مهندسی مالی و مدیریت اوراق بهادار، ۳۵، ۳۰۱-۳۲۲.
- [۷] فلاح‌پور، سعید؛ علی‌پور ریکنده، جواد. (۱۳۹۳). پیش‌بینی شاخص سهام با استفاده از شبکه‌های عصبی موجکی در بورس اوراق بهادار تهران. فصلنامه علمی پژوهشی راهبرد مدیریت مالی، ۲(۴)، ۱۵-۳۱.
- [۸] کیانی‌زاده، حسین؛ باغانی، علی؛ حمیدیان، محسن. (۱۴۰۲). مقایسه دقت مدل‌های منتخب یادگیری ماشین جهت پیش‌بینی قیمت سهام در بورس اوراق بهادار. فصلنامه بورس اوراق بهادار، ۱۶(۶۲)، ۷۵-۱۰۲.
- [۹] مشیری، سعید؛ مروت، حبیب. (۱۳۸۴). پیش‌بینی شاخص کل بازدهی سهام تهران با استفاده از الگوی خطی و غیرخطی. پژوهشنامه بازرگانی، ۱۱(۴۱)، ۲۴۵-۲۷۵.
- [۱۰] نجارزاده، رضا؛ زیودار، مهدی. (۱۳۸۵). بررسی رابطه تجربی بین حجم معاملات و بازده سهام در بازار بورس اوراق بهادار تهران. فصلنامه پژوهش‌های اقتصادی، ۶(۲)، ۵۹-۷۹.
- [11] Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, **30**(1), 9-14.
- [12] Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, **66**, 237-242.
- [13] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Econometrics*, **31**, 307-327.
- [14] Buhlmann, P., & Yu, B. (2003). Boosting with the L^2 loss: regression and classification. *Journal of the American Statistical Association*, **98**(462), 324-339.
- [15] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987-1007.
- [16] Granger, C. W. J. (2004). Time series analysis, cointegration and applications. *American Economic Review*, **94**(3), 421-425.
- [17] Jennings, R. H., Starks, L., & Fellingham, J. (1981). An equilibrium model of asset trading with sequential information arrival. *Journal of Finance*, **36**, 143-161.
- [18] Lee, B. S., & Rui, O. M. (2002). The dynamic relationship between stock returns and trading volume: domestic and cross-country evidence. *Journal of Banking & Finance*, **26**, 51-78.

- [19] Ling, S., & Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroskedasticity. *Journal of the American Statistical Association*, **9**, 1184-1194.
- [20] Ling, S., & Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. *Annals of Statistics*, **26**, 84-125.
- [21] Ling, S., & McAleer, M. (2002). Stationarity and the existence of moments of a family of GARCH processes. *Econometrics*, **106**, 109-117.
- [22] Lutz, R. W., & Buhlmann, P. (2006). Boosting for high-multivariate responses in high-dimensional linear regression. *Statistica Sinica*, **16**(2), 471-494.
- [23] Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, 117-126.
- [24] Wei, W. S. (2006). *Time series analysis: Univariate and multivariate methods* (2nd ed.). Boston: Pearson.

پیوست

- X_۱: صندوق سرمایه‌گذاری افرا نماد پایدار- ثابت (افران)
- X_۲: صندوق سرمایه‌گذاری صنایع آگاه-۱۰- بخشی (پتروآگاه)
- X_۳: صندوق سرمایه‌گذاری نوع دوم افق آتی- ثابت (آفاق)
- X_۴: اختیارخ اهرم ۲۲/۱۲/۱۴۰۱ (ضهرم ۱۲۲۲)
- X_۵: صندوق سرمایه‌گذاری شاخصی آرام مفید (آرام)
- X_۶: صندوق سرمایه‌گذاری شاخصی بازارآشنا (وبازار)
- X_۷: صندوق سرمایه‌گذاری ارمغان ایرانیان (ارمغان)
- X_۸: صندوق سرمایه‌گذاری سهام آوای معیار-س (آوا)
- X_۹: صندوق سرمایه‌گذاری آسمان آرمانی سهام-س (آساس)
- X_{۱۰}: صندوق سرمایه‌گذاری آوند مفید-د (آوند)
- X_{۱۱}: صندوق سرمایه‌گذاری انارنماد ارزش-درسهام (انار)
- X_{۱۲}: صندوق اندیشه ورزان صباتامین -د (اوصتا)
- X_{۱۳}: صندوق سرمایه‌گذاری برلیان- سهام (برلیان)
- X_{۱۴}: صندوق سرمایه‌گذاری سهامی کاریزما- اهرمی (اهرم)
- X_{۱۵}: صندوق سرمایه‌گذاری هستی بخش آگاه-س (آگاس)
- X_{۱۶}: صندوق شاخص ۳۰ شرکت فیروزه- سهام (فیروزه)
- X_{۱۷}: صندوق سرمایه‌گذاری اعتماد داریک-د (داریک)
- X_{۱۸}: صندوق سرمایه‌گذاری اعتماد هامرز-ثابت (هامرز)
- X_{۱۹}: صندوق سرمایه‌گذاری همای آگاه-ثابت (همای)
- X_{۲۰}: صندوق سرمایه‌گذاری اعتماد آفرین پارسیان-د (اعتماد)
- X_{۲۱}: صندوق سرمایه‌گذاری طلای سرخ نوویرا (نهال)
- X_{۲۲}: صندوق سرمایه‌گذاری با درآمد ثابت مانی (مانی)
- X_{۲۳}: صندوق سرمایه‌گذاری با درآمد ثابت کمند (کمند)
- X_{۲۴}: صندوق سرمایه‌گذاری کامیاب آشنا-د (کامیاب)
- X_{۲۵}: صندوق سرمایه‌گذاری کارا -د (کارا)
- X_{۲۶}: صندوق سرمایه‌گذاری کارا -د (کارا۲)
- X_{۲۷}: صندوق سرمایه‌گذاری پارند پایدار سپهر (پارند)
- X_{۲۸}: صندوق سرمایه‌گذاری خاتم ایساتیس پویا-ثابت (خاتم)
- X_{۲۹}: صندوق سرمایه‌گذاری نوع دوم رایکا- ثابت (رایکا)
- X_{۳۰}: صندوق پالایشی یکم-سهام (پالایش)
- X_{۳۱}: صندوق سرمایه‌گذاری هوش تصمیم -س (هوشیار)
- X_{۳۲}: صندوق سرمایه‌گذاری سروسودمند مدبران-سهام
- X_{۳۳}: صندوق سرمایه‌گذاری گروه زعفران سحرخیز (سحرخیز)
- X_{۳۴}: صندوق سرمایه‌گذاری سپر سرمایه بیدار- ثابت (سپر)
- X_{۳۵}: صندوق سرمایه‌گذاری سلام فارابی- سهام (سلام)
- X_{۳۶}: صندوق سرمایه‌گذاری ثبات ویستا -د (ثبات)
- X_{۳۷}: صندوق سرمایه‌گذاری توسعه اطلس مفید-س (اطلس)
- X_{۳۸}: صندوق سرمایه‌گذاری با درآمد ثابت تصمیم (تصمیم)
- X_{۳۹}: صندوق سرمایه‌گذاری اهرمی توان مفید-س (توان)
- X_{۴۰}: صندوق سرمایه‌گذاری ترمه- سهام (ترمه)

Analysis of Investment Risk in Companies Listed on the Iran Stock Exchange Using a Boosting Algorithm

S. Zamani Mehreyan^{1*} and N. Eghbalifar²

¹ Department of Statistics, Imam Khomeini International University, Qazvin, Iran

² Department of Accounting, Yasouj Branch, Islamic Azad University, Yasuj, Iran

Abstract:

In recent years, stock price prediction and market return models have attracted significant research attention. Traditional statistical forecasting models, such as autoregressive (AR), autoregressive moving average (ARMA) and generalized autoregressive conditional heteroskedasticity (GARCH) models, have been widely used. Advances in computer processing power have facilitated the introduction of new algorithms, including machine learning techniques, for predicting financial data. Consequently, this study analyzes investment risk in companies listed on the Iran Stock Exchange by examining data on their trading volume. An optimal vector autoregressive (VAR) model is fitted using a boosting algorithm. The performance of this boosting-based model is compared to that of models fitted using conventional statistical methods. The results demonstrate the superior performance of the boosting algorithm.

Keywords: Boosting algorithm, Trading volume, Vector autoregressive (VAR) model, Generalized autoregressive conditional heteroskedasticity (GARCH) model.