An Improved Hybrid Model with Automated Lag Selection to Forecast Stock Market

Moien Nikusokhan
MSc., Department of Financial Management, Faculty of Management and Accounting, Shahid Beheshti University, Tehran, Iran. E-mail: mnikusokhan@gmail.com

Abstract

Objective: In general, financial time series such as stock indexes have nonlinear, mutable and noisy behavior. Structural and statistical models and machine learning-based models are often unable to accurately predict series with such a behavior. Accordingly, the aim of the present study is to present a new hybrid model using the advantages of the GMDH method and Non-dominated Sorting Genetic Algorithm II (NSGA II) to, more accurately, predict the trend of movement and volatility of Tehran Stock Exchange Price Index, and to compare its ability with the ARIMA model based on RMSE, MAPE, and TIC error assessment criterions.

Methods: For this purpose, the data of Tehran Stock Exchange Dividend and Price Index (TEDPIX) was used during the period of October 2008 to September 2013. The hybrid model NSGA II - GMDH utilizes the GMDH network as a model resistant to non-stationary and noisy data for prediction and uses the NSGA II multi-objective optimization algorithm to minimize predictive error and select the optimal input variables.

Results: The results of the proposed hybrid model in this study indicated a lower error and more prediction accuracy compared to ARIMA model for out-of-sample data based on all three error criterions.

Conclusion: The empirical findings of the study showed that the proposed model has higher flexibility and capability in covering unstable changes in the total index movement trend.

Keywords: Financial time series, Group method of data handling, Hybrid model, Non-dominated sorting genetic algorithm II, Stock market forecasting.

Citation: Nikusokhan, M. (2018). An Improved Hybrid Model with Automated Lag Selection to Forecast Stock Market. *Financial Research Journal*, 20(3), 389-408. (*in Persian*)

Financial Research Journal, 2018, Vol. 20, No.3, pp. 389-408
DOI: 10.22059/frj.2018.257176.1006656
Received: December 30, 2017; Accepted: June 26, 2018

ارائه یک مدل ترکیبی بهبودیافته با انتخاب وقفه‌های خودکار برای پیش‌بینی بازار سهام

معین نیکوسخن
کارشناس ارشد، گروه مدیریت مالی، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی، تهران، ایران. رایانامه: mnikusokhan@gmail.com

چکیده

هدف: به‌طور کلی سری‌های زمانی مالی مانند شاخص سهام، رفتار غیرخطی، بی‌ثبات و نویزی دارند. مدل‌های ساختاری و آماری و مدل‌های مبتنی بر یادگیری ماشین، اغلب توانایی پیش‌بینی دقیق سری‌هایی با این گونه رفتار را ندارند. بر این اساس، هدف تحقیق حاضر، ارائه یک مدل ترکیبی جدید با بهره‌مندی از مزایای روش گروهی مدل‌سازی داده‌ها (GMDH) و الگوریتم ژنتیک با مرتب‌سازی نامغلوب (NSGA II) برای پیش‌بینی دقیق‌تر روند حرکت و تغییرات شاخص کل بورس اوراق بهادار تهران و مقایسه توانایی آن با مدل ARIMA بر اساس معیارهای سنجش خطا شامل RMSE، MAPE و TIC است.

روش: برای دستیابی به هدف پژوهش، از داده‌های شاخص کل قیمت و بازده نقدی در بورس اوراق بهادار تهران (TEDPIX) طی دوره زمانی مهر ۱۳۸۷ تا شهریور ۱۳۹۷ استفاده شده است. مدل ترکیبی NSGA II-GMDH، شبکه GMDH را به‌عنوان مدلی مقاوم در برابر داده‌های نویزی و نامانا برای پیش‌بینی به‌کار می‌گیرد و از الگوریتم بهینه‌سازی چندهدفه NSGA II برای کمینه‌سازی خطای پیش‌بینی و انتخاب متغیرهای ورودی بهینه استفاده می‌کند.

یافته‌ها: نتایج به‌دست آمده از مدل ترکیبی ارائه‌شده در این پژوهش، بر اساس هر سه معیار سنجش خطا، بیان‌کننده خطای کمتر و دقت پیش‌بینی بیشتر آن در مقایسه با مدل ARIMA برای داده‌های خارج از نمونه است.

نتیجه‌گیری: با توجه به یافته‌های تجربی می‌توان نتیجه گرفت که مدل پیشنهادی در پوشش تغییرات ناپایدار روند حرکت شاخص کل، از انعطاف‌پذیری و توانایی بیشتری برخوردار است.

کلیدواژه‌ها: الگوریتم ژنتیک با مرتب‌سازی نامغلوب، پیش‌بینی بازار سهام، روش گروهی مدل‌سازی داده‌ها، سری‌های زمانی مالی، مدل ترکیبی.

استناد: نیکوسخن، معین (۱۳۹۷). ارائه یک مدل ترکیبی بهبودیافته با انتخاب وقفه‌های خودکار برای پیش‌بینی بازار سهام. *فصلنامه تحقیقات مالی*، ۲۰(۳)، ۳۸۹ – ۴۰۸.

فصلنامه تحقیقات مالی، ۱۳۹۷، دوره ۲۰، شماره ۳، صص. ۳۸۹ – ۴۰۸
DOI: 10.22059/frj.2018.257176.1006656
دریافت: ۱۳۹۶/۱۰/۰۹، پذیرش: ۱۳۹۷/۰۴/۰۵

مقدمه

امروزه عوامل بسیاری از جمله رویدادهای سیاسی، وضعیت کلی اقتصاد، انتظاراتی که فعالان بازار دارند و... عملکرد بازارهای مالی را تحت تأثیر قرار داده و بر پیچیدگی‌های آن می‌افزایند. این پیچیدگی‌ها، شاخص بازار سهام را که یکی از معیارهای سنجش عملکرد بازارهای مالی شناخته می‌شود با درجه بالایی از بی‌قاعدگی‌ها و فرایندهای غیرخطی، نامانایی و نویزی روبه‌رو می‌کند. در نتیجه، پیش‌بینی سری‌های زمانی شاخص بازار سهام همواره با مسئله غیرخطی، نامانایی و نویزی بودن همراه است (یو، وانگ و لایی[1]، ۲۰۰۹). از آنجا که این مسائل پیش‌بینی دقیق سری زمانی شاخص‌های بازار سهام را با اختلال روبه‌رو کرده و دقت پیش‌بینی را تضعیف می‌کند، کلیدی‌ترین مسائل در پیش‌بینی سری‌های زمانی مالی شناخته شده‌اند (لو، لی و چیو[2]، ۲۰۰۹).

در حالی‌که نتیجه تحقیقات تجربی حاکی از رفتار بی‌نظم، غیرخطی، نامانا، نویزی و ناپارامتریک بازارهای مالی است (گارلگی، شاری و شفیقی[3]، ۲۰۱۴ و رافیوزامن[4]، ۲۰۱۴)، روش‌های ساختاری و کلاسیک پیش‌بینی بازار سهام همواره به‌دنبال ارائه مدل خطی بر پایه تئوری‌های اقتصادی میان قیمت سهام و متغیرهای اقتصادی و مالی هستند، مانند مدل‌های تحلیل رگرسیونی و میانگین متحرک خودرگرسیون انباشته (ARIMA)[5] که تنها می‌توانند ویژگی‌های خطی داده‌های مالی را در نظر بگیرند. به‌دنبال ناموفق بودن روش‌های ساختاری و کلاسیک در پیش‌بینی شاخص‌های بازار سهام، در دو دهه گذشته گرایش به استفاده از مدل‌های ناپارامتریک، غیرخطی و مبتنی بر هوش مصنوعی برای پیش‌بینی شاخص‌های بازار سهام، به‌طور شایان توجهی افزایش یافته است. مدل‌های مبتنی بر هوش مصنوعی مانند شبکه‌های عصبی، به‌دلیل تفاوت‌های ذاتی آنها با رویکردهای خطی و پارامتریک، از قابلیت احتساب پویایی غیرخطی و پیچیدگی‌های داده‌های مالی برخوردارند (گالشچوک[6]، ۲۰۱۶). اما این گروه از مدل‌ها نیز، هنگام مواجه با سری‌های زمانی نویزی عملکرد مناسبی از خود نشان نداده و نویز موجود در این سری‌ها به بیش‌برازشی این طبقه از مدل‌ها منجر می‌شود (جایلز، لاورنس و تسوی[7]، ۲۰۰۱). مسئله دیگری که عملکرد اغلب این مدل‌ها را با چالش جدی روبه‌رو کرده است، مسئله انتخاب متغیرهای توضیحی به‌عنوان متغیرهای ورودی به مدل است. به استثنای مدل‌های ساختاری، در سایر مدل‌های پیش‌بینی، متغیرهای توضیحی به‌واسطه نظر محقق یا به‌صورت آزمون خطا در سطح محدودی تعیین می‌شوند که این نمی‌تواند روش اصولی و درستی برای یافتن مناسب‌ترین متغیرهای ورودی به مدل باشد.

از این رو در پژوهش حاضر مدل ترکیبی جدید NSGA II-GMDH برای پیش‌بینی شاخص کل بورس اوراق بهادر تهران با هدف غلبه هم‌زمان بر مسئله غیرخطی، نامانایی و نویزی بودن سری زمانی شاخص و مسئله انتخاب بهترین متغیرهای توضیحی ارائه شده است. مدل ترکیبی ارائه‌شده، از روش گروهی مدل‌سازی داده‌ها (GMDH)[8] به‌عنوان نوعی روش داده‌کاوی خودسازمانده اکتشافی برای مدل‌سازی و پیش‌بینی شاخص بازار استفاده می‌کند. مزیت اصلی این روش

1. Yu, Wang, & Lai
2. Lu, Lee, & Chiu
3. Gharleghi, Shaari, & Shafighi
4. Rafiuzzaman
5. Autoregressive Integrated Moving Average
6. Galeshchuk
7. Giles, Lawrence, & Tsoi
8. Group Method of Data Handling

مقاومت آن در برابر داده‌های نویزی و ناماناست؛ به‌طوری که برای سطح معینی از داده‌های نویزی و نامانا، روش GMDH عملکرد بهتری از سایر روش‌های خطی و غیرخطی از خود نشان می‌دهد (مولر و لمکه[1]، ۲۰۰۰؛ واس و هاولند[2]، ۲۰۰۳؛ یانگ و همکاران[3]، ۲۰۰۹ و ژو، هی و لیاتسیس[4]، ۲۰۱۲). این مدل ترکیبی، هم‌زمان از ویژگی بهینه‌سازی چندهدفه الگوریتم ژنتیک با مرتب‌سازی نامغلوب نسخه دوم (NSGA II)[5] برای بهره‌مندی از توانایی خودانتخابی متغیرهای توضیحی استفاده می‌کند؛ به‌گونه‌ای که این الگوریتم با جست‌وجو میان وقفه‌های تأخیری[6] مختلف، وقفه‌های مؤثر را شناسایی کرده و به‌عنوان متغیرهای ورودی برای کاهش هزینه و زمان پردازش و بهبود کارایی مدل GMDH انتخاب می‌کند. سرانجام عملکرد مدل ترکیبی NSGA II-GMDH و مدل ARIMA با مجموعه داده‌های سری زمانی شاخص کل بورس اوراق بهادار مقایسه می‌شود.

در ادامه مرور کوتاهی از ادبیات مربوطه ارائه شده و روش‌شناسی و داده‌های پژوهش به تفصیل توصیف می‌شود. پس از آن، به بحث درباره نتایج تجربی و یافته‌های به‌دست از مقایسه مدل‌ها پرداخته شده و در انتها نتیجه‌گیری ارائه می‌شود.

پیشینه پژوهش

تحلیل‌گران تکنیکی، نخستین گروهی بودند که به پیش‌بینی بازار پرداختند. آنها بر این باورند که مطالعه رفتار گذشته قیمت‌ها، به کشف الگوی تغییر آنها و در نهایت پیش‌بینی قیمت‌های آتی منجر می‌شود. در مقابل، بنیادگرایان معتقدند که قیمت سهم، تابعی از ارزش ذاتی آن سهم است. آنها ارزش ذاتی را به‌وسیله عوامل مؤثر در سطح شرکت، صنعت و اقتصاد تعیین می‌کنند. اما تحقیقات تجربی با ارائه شواهد سازگار با فرضیه بازار کارای (EMH)[7] فاما[8] (۱۹۷۰) نشان می‌دهد متغیرهای اقتصادی، بنیادی و اطلاعات گذشته، محتوای پیش‌بینی‌کنندگی ناچیزی در خصوص رفتار بازار سهام دارند و هیچ مدلی برای پیش‌بینی بازار، بهتر از مدل گام تصادفی عمل نمی‌کند (گالوتی و اسکینتاولی[9]، ۱۹۹۴؛ لاورنس[10]، ۱۹۹۷ و تیمرمن و گرنجر[11]، ۲۰۰۴). از سوی دیگر، رویکردهای اقتصادسنجی برای پیش‌بینی سری‌های زمانی مانند مدل ARIMA، اغلب روابط را به‌صورت خطی توصیف می‌کنند، از این رو نسبت به پویایی سری‌های اقتصادی و مالی قادر به پاسخگویی مناسب نیستند (تیمرمن و گرنجر، ۲۰۰۴). به علاوه طی دو دهه اخیر، پیشرفت سریع و گسترده علم و فناوری و تلاش محققان برای اصلاح روش‌ها و مدل‌های پیش‌بینی، موجب افزایش شایان توجه گرایش به استفاده از تکنیک‌های پیشرفته مبتنی بر هوش مصنوعی، به‌منظور توضیح رفتار قیمت‌ها و پیش‌بینی آنها در بازار سرمایه شده است. این مدل‌ها به‌طور عمده از توانایی احصای فرایندهای غیرخطی، نامانا و نویزی برخوردارند.

در چند سال اخیر، برای بهره‌مندی هم‌زمان از مزایای چندین مدل، علاقه به استفاده از مدل‌های ترکیبی در مسئله

1. Müller, & Lemke
2. Voss, & Howland
3. Yang, Liao, Chen, Huang, Huang, Huang, & Chung,
4. Zhu, He, & Liatsis
5. Non-dominated Sorting Genetic Algorithm II
6. Lag
7. Efficient-market hypothesis
8. Fama
9. Galotti, & Schiantavelli
10. Lawrence
11. Timmermann, & Granger

پیش‌بینی سری‌های زمانی مالی، به‌طور گسترده‌ای افزایش یافته است. ابود و دشپند[1] در سال ۲۰۰۸ تلاش کردند الگوریتم پایه GMDH را به‌وسیله الگوریتم ژنتیک (GA) و روش بهینه‌سازی ازدحام ذرات (PSO) بهبود دهند. ایده اصلی آنها استفاده از روش بهینه‌سازی بهتر در تخمین چندجمله‌ای‌های درجه دوم الگوریتم GMDH و ایجاد تابعی است که رابطه ورودی ـ خروجی را دقیق‌تر توصیف کند تا دقت پیش‌بینی الگوریتم GMDH بهبود یابد. آنها عملکرد روش GAGMDH و SPOGMDH را با GMDH معمولی مقایسه کردند و نتیجه گرفتند که الگوریتم‌های ترکیبی GMDH، اثربخش‌تر، کارآمدتر و دقیق‌تر از الگوریتم GMDH معمولی است.

تسای و وانگ[2] (۲۰۰۹) پس از ترکیب مدل درخت تصمیم (DT) و شبکه‌های عصبی مصنوعی (ANN)، تلاش کردند با ارائه قواعدی برای توصیف تصمیمات پیش‌بینی، عملکرد ANN را بهبود بخشند. نتیجه کار آنها به افزایش دقت ۷۷درصدی مدل ترکیبی DTANN در پیش‌بینی قیمت سهام بازار تایوان انجامید و از دقت پیش‌بینی مدل‌های DT و ANN به‌صورت جداگانه بیشتر بود.

تسای و هوآنگ[3] (۲۰۰۹) برای پیش‌بینی آتی شاخص قیمت بازار سهام تایوان از ترکیب رگرسیون بردار پشتیبان (SVR) با الگوریتم نگاشت‌های خودسازمانده ویژگی[4] (SOFM) استفاده کردند. این سیستم ترکیبی ابتدا با استفاده از الگوریتم SOFM به خوشه‌بندی نمونه‌های آموزش می‌پردازد و در نهایت از SVR برای پیش‌بینی استفاده می‌کند. نتایج کار آنها نشان داد که مدل SOFM-SVR میانگین دقت پیش‌بینی و زمان آموزش را نسبت به مدل SVR استاندارد بهبود داده است.

چنگ، چن و وی[5] (۲۰۱۰) یک مدل پیش‌بینی ترکیبی با استفاده از شاخص‌های تکنیکال برای پیش‌بینی روند بازار سهام پیشنهاد کردند. مدل آنها ابتدا با استفاده از الگوریتم تئوری مجموعه راف (RST)[6]، قواعد کلامی را از مجموعه شاخص‌های تکنیکال استخراج می‌کرد، سپس از الگوریتم ژنتیک (GA) برای اصلاح قواعد استخراج شده برای دستیابی به پیش‌بینی دقیق‌تر بهره می‌برد. نتایج تجربی آنها نشان داد که دقت پیش‌بینی مدل SRT-GA نسبت به هر یک از مدل‌های GA و SRT به تنهایی بهبود یافته است؛ به‌گونه‌ای که سود خلق شده به وسیله مدل پیشنهادشده به مراتب بیشتر از مدل‌های دیگر بود.

سان و ژآئو[7] (۲۰۱۵) برای استفاده هم‌زمان از قابلیت جست‌وجوی جهانی الگوریتم بهینه‌سازی ازدحام ذرات (PSO) و مزایای جست‌وجوی محلی شبکه عصبی پس‌انتشار خطای ترکیبی (HBP) در پیش‌بینی بازار سهام، مدل ترکیبی HBP-PSO را ارائه کردند. مدل پیشنهاد شده، الگوریتم PSO را برای آموزش وزن‌های ارتباطی و آستانه‌های شبکه BP به‌کار می‌گیرد، سپس از شبکه BP آموزش دیده برای پیش‌بینی قیمت سهام استفاده می‌کند. نتایج تجربی آنها با استفاده از داده‌های بورس اوراق بهادار شانگهای، نشان از انعطاف‌پذیری و اثربخشی این روش دارد.

1. Abbod, & Deshpande
2. Tsai, & Wang
3. Tsai, & Huang
4. Self-organizing feature map
5. Cheng, Chen, & Wei
6. Rough Set Theory
7. Sun, & Gao

گریگوریان[1] (۲۰۱۶) برای پیش‌بینی شاخص‌های بورس اوراق بهادار بخارست و بازار سهام بالتیک، از مدل پیش‌بینی انباشته‌ای مبتنی بر ماشین بردار پشتیبان (SVM) و تکنیک تحلیل مؤلفه‌های مستقل (ICA)[2] با نام SVM-ICA استفاده کرد. رویکرد ارائه شده ابتدا با استفاده از تکنیک ICA، متغیرهای مهم را از میان داده‌های به‌دست آمده از تحلیل تکنیکی و بنیادی استخراج می‌کند، سپس با الگوریتم SVM به پیش‌بینی سری زمانی بازار سهام می‌پردازد. تجزیه و تحلیل تطبیقی وی نشان می‌دهد که مدل پیشنهاد شده SVM-ICA در پیش‌بینی سری‌های زمانی نامانا بهتر از مدل SVM عمل می‌کند.

عبادتی و مرتضوی[3] (۲۰۱۸) یک روش ترکیبی از الگوریتم ژنتیک (GA) و شبکه عصبی مصنوعی (ANN) پس‌انتشار خطا، برای پیش‌بینی سری زمانی شاخص نزدک[4] و پنج شرکت بزرگ از شرکت‌های تشکیل‌دهنده آن ارائه کردند. در روش پیشنهاد شده، مقادیر خروجی GA به‌منظور ثابت‌کردن مقدار خطا در یک نقطه مشخص به‌وسیله الگوریتم ANN توسعه داده می‌شود. نتایج آنها نشان داد که عملکرد، سرعت و دقت پیش‌بینی الگوریتم پیشنهاد شده GA-ANN در مقایسه با روش‌های سنتی بهبود یافته است.

مهرآرا، معینی، احراری و هامونی (۱۳۸۸) با تلفیق الگوریتم ژنتیک (GA) و رویکرد شبکه GMDH، به پیش‌بینی شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران و شناخت متغیرهای مؤثر بر آن پرداختند. بدین منظور آنها از داده‌های ماهانه متغیرهای کلان اقتصادی مرتبط با بازار سرمایه در حد فاصل بهمن ۱۳۷۴ تا اسفند ۱۳۸۵ به‌عنوان متغیر وابسته استفاده کردند. نتایج پژوهش آنها گویای اثر معنادار شاخص قیمت زمین، هزینه مسکن، شاخص قیمت مصرف‌کننده، پایه پولی، کرایه مسکن استیجاری و قیمت جهانی نفت خام بر شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران است.

شمس و ناجی زواره (۱۳۹۴) به‌منظور پیش‌بینی قیمت قرارداد آتی سکه طلا در بورس کالای ایران، مدلی ترکیبی بر اساس سیستم ژنتیک فازی (GFS) و شبکه عصبی مصنوعی (ANN) ارائه کردند. مدل پیشنهاد شده آنها در گام اول متغیرهای مؤثر را با استفاده از روش رگرسیون گام‌به‌گام شناسایی کرده و در گام بعد با استفاده از شبکه عصبی خودسازمانده، متغیرهای شناسایی شده را به K خوشه دسته‌بندی می‌کند و در نهایت این دسته‌ها به GFS وارد شده و پیش‌بینی انجام می‌شود. نتایج آنها نشان داد خطای پیش‌بینی مدل ترکیبی ارائه‌شده نسبت به روش خطی ARIMA کمتر است. درودی و ابراهیمی (۱۳۹۵) نیز، در تلاش برای پیش‌بینی دقیق‌تر روند حرکتی و تغییرات شاخص کل قیمت سهام، مدل هیبریدی نوینی ارائه دادند. مدل پیشنهاد شده این دو محقق، در مرحله اول بر اساس تکنیک تبدیل موجک، سری زمانی شاخص را به شش سری زمانی مجزا با ویژگی‌های غیرخطی و متلاطم تفکیک می‌کند و در مرحله بعد، سری‌های با رفتار غیرخطی را با استفاده از ترکیب مدل SVR و PSO و سری‌های زمانی مبتنی بر رفتار متلاطم را با بهره‌گیری از مدل GJR پیش‌بینی می‌کند. در ادامه با تجمیع این نتایج سری زمانی شاخص کل برآورد می‌شود. نتایج پژوهش آنها بر اساس داده‌های ۱۳۸۷ تا ۱۳۹۴ شاخص کل، نشان می‌دهد مدل هیبریدی پیشنهاد شده در مقایسه با سایر روش‌های پیش‌بینی، خطای کمتری داشته و از دقت بیشتری برخوردار است.

1. Grigoryan
2. Independent Component Analysis
3. Ebadati, & Mortazavi
4. NASDAQ

فخاری، ولی‌پور خطیر، موسوی (۱۳۹۶) با طراحی مدل پیش‌بینی بر اساس شبکه عصبی بیزین، به مقایسه عملکرد آن با مدل‌های کلاسیک و معرفی مدل مناسب برای پیش‌بینی قیمت روز آتی سهام پرداختند. در این راه، آنها از داده‌های قیمت روزانه بازار و شاخص‌های تکنیکی مالی دوره زمانی ۱۳۹۰ تا ۱۳۹۳ استفاده کردند. یافته‌های تحقیق آنها گویای کارایی بیشتر شبکه عصبی بیزین در استفاده از فرصت‌های سرمایه‌گذاری کوتاه‌مدت بازار در مقایسه با مدل ARIMA است که می‌تواند به سرمایه‌گذاران در انتخاب پرتفوی مناسب و کسب بازده بیشتر کمک کند.

بر این اساس پژوهش حاضر به‌دنبال آزمون فرضیه زیر است: عملکرد مدل ترکیبی NSGA II-GMDH برای پیش‌بینی شاخص کل بورس اوراق بهادار در خارج از نمونه، بهتر از مدل ARIMA است.

روش‌شناسی پژوهش

مدل ARIMA

در تحلیل سری زمانی، مدل ARIMA یکی از مؤثرترین مدل‌ها در پیش‌بینی سری‌های زمانی نامانا در نظر گرفته می‌شود. مدل ARIMA$(p.d.q)$ ترکیبی از درجه انباشتگی d با مدل خودرگرسیونی (AR) با مرتبه p و مدل میانگین متحرک (MA) با مرتبه q است.

رابطه ۱)

$$ARIMA(p.d.q):\ y_t = \sum_{i=1}^{p} \beta_i y_{t-i} \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j+}\ \varepsilon_t$$

$$y_t = \Delta^d y_t = (1-L)^d y_t$$

جایی که ε_t جزء خطای نوفه سفید و L عملگر وقفه است.

برای تخمین مدل ARIMA$(p.d.q)$ از روش باکس ـ جنکینز[1] استفاده می‌شود که دارای چهار مرحله شناسایی، برآورد، آزمون کنترل تشخیصی و پیش‌بینی است. تعداد وقفه‌های معنادار اجزای خودرگرسیون (p) و میانگین متحرک (q) معمولاً با استفاده از توابع خودهمبستگی[2] (AC) و خود همبستگی جزئی (PAC)[3] بر اساس روش باکس ـ جنکینز تعیین می‌شود، اما از آنجا که ممکن است مدل‌های بهینه دیگری وجود داشته باشند که بر الگوی یاد شده ترجیح داده شوند، این مدل‌ها توسط معیار اطلاعات آکائیک (AIC) بازبینی می‌شوند.

مدل GMDH

نخستین بار ایواکننکو[4] (۱۹۶۸) از الگوریتم GMDH برای مدل‌سازی سیستم‌های پیچیده استفاده کرد. این شبکه با تشکیل یک مدل خودسازمانده، امکان تشخیص الگوها و پیش‌بینی آنها را فراهم می‌کند؛ به‌گونه‌ای که نرون‌های مؤثر (متغیرها) در لایه‌های پنهان و ساختار مدل بهینه به‌صورت خودکار تعیین می‌شود.

1. Box-jenkins
2. Auto Coloration
3. Partial Auto Coloration
4. Ivakhnenko

رویکرد ایواکننکو برای انطباق میان متغیرهای ورودی و خروجی در شبکه GMDH و تشکیل تابع غیرخطی بی‌نهایت به نام سری ولترا یا چندجمله‌ای ولترا ـ کولموگروف ـ گابور (VKG) است (رابطه ۲).

$$y = \alpha_0 + \sum_{i=1}^{m} \alpha_i X_i + \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_{ij} X_i X_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_{ijk} X_i X_j X_k + \cdots$$ رابطه ۲)

جایی که $X(x_1, x_2, \ldots, x_m)$ بردار متغیرهای ورودی و $A(\alpha_1, \alpha_2, \ldots, \alpha_R)$ بردار ضرایب یا وزن‌هاست. همان‌طور که مشخص است، هدف الگوریتم GMDH تخمین بردار ضرایب در سری فوق است، اما تعداد ضرایب تخمینی در این چندجمله‌ای با افزایش تعداد ورودی‌ها به‌صورت نمایی افزایش می‌یابد؛ به‌طوری که برای داشتن m ورودی، 2^m ضریب باید تخمین‌زده شود. ایده الگوریتم GMDH، تجزیه سری ولترا پیچیده فوق به چندجمله‌ای‌های درجه دوم دومتغیره (رابطه ۳) با تعداد مجهولات کمتر است.

$$y_{ij} = f(X_i, X_j) = \alpha_0 + \alpha_1 X_i + \alpha_2 X_j + \alpha_3 X_i^2 + \alpha_4 X_j^2 + \alpha_5 X_i X_j$$ رابطه ۳)

حال ضرایب α با روش حداقل مربعات معمولی (OLS) برای هر جفت از متغیرهای ورودی X_i و X_j تخمین زده می‌شود.

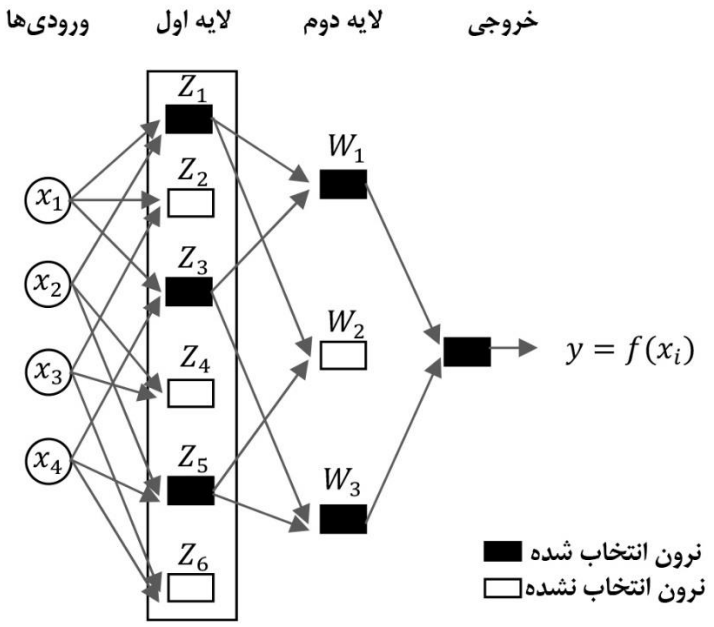
در شکل ۱، یک شبکه GMDH سه لایه به‌عنوان شبکه پیش‌خور[1] به نمایش درآمده است. ابتدا متغیرهای ورودی X به‌منظور برازش چندجمله‌ای‌های درجه دوم و دومتغیره ایواکننکو (رابطه ۳) به‌صورت زوج مرتب در هر نرون قرار می‌گیرند. بر این اساس برای m متغیر ورودی، ترکیب $\binom{m}{2}$ نرون در لایه اول شکل می‌گیرد؛ بدین ترتیب که هر نرون دارای دو ورودی و یک خروجی است که رابطه میان آنها با استفاده از چندجمله‌ای درجه دوم ایواکننکو در قالب شش ضریب مجهول و یک جمله خطا بیان می‌شود. پس از تخمین ضرایب نامعلوم این چندجمله‌ای‌ها، ضرایب محاسبه شده در داخل هر نرون جایگذاری می‌شود تا مقدار متغیر جدید Z که حاصل چندجمله‌ای‌های متشکل از متغیرهای ورودی X است، به‌دست آید (رابطه ۴).

$$Z_{ij} = \hat{y}_{ij} = \alpha_0 + \alpha_1 X_i + \alpha_2 X_j + \alpha_3 X_i^2 + \alpha_4 X_j^2 + \alpha_5 X_i X_j$$ رابطه ۴)

از آنجا که در این شبکه، نرون‌های هر لایه مولد نرون‌های لایه بعد هستند، خروجی لایه اول بردار Z، به‌عنوان ورودی لایه بعد، جایگزین متغیرهای اولیه می‌شوند؛ اما به‌منظور بهبود شبکه و دستیابی به مدل بهینه، تنها Zهایی که بهترین برآوردکننده بردار y هستند، به لایه بعد راه می‌یابند و مابقی آنها حذف می‌شوند. در این رابطه، بردار متغیرهای Z بر اساس معیار جذر میانگین مربعات خطا (رابطه ۵) مرتب می‌شوند، سپس ۷۰ درصد ابتدایی که با مقادیر واقعی y بیشترین قرابت را دارند، به‌عنوان متغیر ورودی به لایه بعد وارد شده و مابقی نرون‌ها حذف می‌شوند. این کار از واگرایی شبکه نیز جلوگیری می‌کند.

1. Feed-forward

$$RMSE_j = \sqrt{\sum(y_{ij} - Z_{ij})^2}\ j = 1,2,\cdots,\binom{m}{2}$$ رابطه ۵)



شکل ۱. نمایی از یک شبکه GMDH

حال در لایه بعد، متغیرهای جدید Z بار دیگر به‌صورت دوبه‌دو در قالب هر نرون برای برازش چندجمله‌ای‌های ایواکننکو و ایجاد نسل جدید W ترکیب می‌شوند و این فرایند تا زمانی که شرایط خاتمه محقق نشده، تکرار می‌شود تا در نهایت مدل نهایی به‌دست آید. شایان ذکر است که شبکه GMDH مدنظر دارای حداکثر ۱۵ لایه بوده و هر لایه حداکثر ۳۰ نرون را در خود جای خواهد داد.

مدل ترکیبی NSGA II-GMDH

الگوریتم پیشنهاد شده در این مقاله، به‌عنوان ترکیبی از نسخه دوم الگوریتم ژنتیک با مرتب‌سازی نامغلوب (NSGA II) و شبکه GMDH، تلاش می‌کند که در کنار بهره‌مندی از توانایی شبکه GMDH در مدل‌سازی و پیش‌بینی سرهای زمانی پیچیده با ساختار نامشخص، از قدرت الگوریتم NSGA II به‌عنوان روش بهینه‌سازی چندهدفه برای یافتن وقفه‌های تأخیری بهینه استفاده کند. الگوریتم NSGA II نوعی روش بهینه‌سازی و جست‌وجوی اکتشافی بر پایه تکامل طبیعی است که توسط دِب، پراتاپ، آگراوال و میاریوان[1] (۲۰۰۲) معرفی شده است. این الگوریتم برخلاف سایر روش‌های

1. Deb, Pratap, Agrawal, & Meyarivan

بهینه‌سازی، نیازی به توصیف ریاضی مسئله بهینه‌سازی ندارد و برازندگی جواب‌ها را به‌وسیله رتبه‌بندی بر اساس نامغلوب‌بودن و فاصله ازدحامی[1] تعیین می‌کند. بدین‌ترتیب الگوریتم ترکیبی NSGA II-GMDH می‌تواند با به‌کارگیری ویژگی‌های هر دو الگوریتم، بر مشکل نویزی بودن و نامانایی سری‌های زمانی و مسئله تعیین وقفه‌های مؤثر فائق آید. معماری الگوریتم NSGA II-GMDH در شکل ۲ به تصویر کشده شده است. روند شکل‌گیری و کارکرد این الگوریتم را می‌توان در شش گام زیر بیان کرد:

گام نخست: در نخستین گام، جمعیت اولیه از افرادی که در واقع تعدادی از جواب‌های ممکن مسئله است، به‌صورت تصادفی ایجاد می‌شود تا الگوریتم کار خود را آغاز کند.

گام دو: به‌منظور انتخاب مناسب‌ترین افراد، به یک تابع هزینه برای ارزیابی هر عضو از جمعیت نیاز است که در این رابطه از شبکه GMDH استفاده شده است. به‌دلیل وجود دو تابع هدف در مسئله پژوهش حاضر، یعنی کمینه‌سازی هم‌زمان خطای پیش‌بینی (RMSE) حاصل از شبکه GMDH و تعداد وقفه‌های تأخیری به‌عنوان متغیرهای ورودی، تابع هزینه به‌صورت یک مسئله بهینه‌سازی دو هدفه که شامل برداری از خطای شبکه GMDH و تعداد متغیرهای ورودی است، تعریف می‌شود.

گام سه: از آنجا که معمولاً جمعیت اولیه حائز شرایط خاتمه الگوریتم نیست، باید جمعیت بهتر (نسل جدید) از جمعیت اولیه ایجاد شود. برای ایجاد جمعیت جدید، ابتدا باید افرادی از جمعیت اولیه که شایستگی دارند، به‌عنوان والدین انتخاب شوند. برای انتخاب والدین لازم است مجموعه جواب‌های مسئله بر اساس مفهوم غلبه و فاصله ازدحامی مرتب شوند. چون در مسئله پژوهش دو تابع هدف مدنظر است، در بیشتر مواقع جواب‌هایی وجود دارند که هیچ‌یک بر دیگری برتری ندارند و نمی‌توان آنها را با یکدیگر مقایسه کرد. به همین دلیل برای مقایسه جواب‌ها از مفهوم غلبه استفاده می‌شود که بیان می‌کند جواب x_1 بر جواب x_2 غالب است، اگر: الف) جواب x_1 در هیچ‌یک از اهداف بدتر از x_2 نباشد و ب) جواب x_1 دست کم در یک هدف اکیداً بهتر از جواب x_2 باشد. بدین صورت به هر جواب بر اساس تعداد مغلوب‌شدن نسبت به سایر جواب‌ها رتبه‌ای تعلق می‌گیرد که به آن مرتب‌سازی نامغلوب[2] می‌گویند. بنابراین جواب‌هایی که در جبهه نخست قرار دارند و توسط هیچ‌یک از جواب‌های دیگر مغلوب نشده‌اند، در رتبه یک دسته‌بندی می‌شوند و جواب‌هایی که توسط حداقل یکی از جواب‌های جبهه نخست مغلوب می‌شوند، در رتبه دوم قرار می‌گیرند و... . پس از مرتب‌سازی جواب‌ها بر اساس غلبه، به‌منظور ایجاد نظم در مجموعه جواب‌های بهینه، باید آنها را براساس فاصله ازدحامی مرتب کرد. فاصله ازدحامی برای انتخاب جواب‌های بهتر و یکنواخت‌تر از نظر پراکندگی روی یک جبهه استفاده می‌شود که بر اساس رابطه‌های ۶ و ۷ محاسبه می‌شود.

$$d_i^j = \frac{\left|f_j^{i+1} - f_j^{i-1}\right|}{f_j^{max} - f_j^{min}} \qquad \text{رابطه ۶)}$$

1. Crowding distance
2. Non-dominated sorting

$$d_i = \sum_{j=1}^{m} d_i^j$$ رابطه ۷)

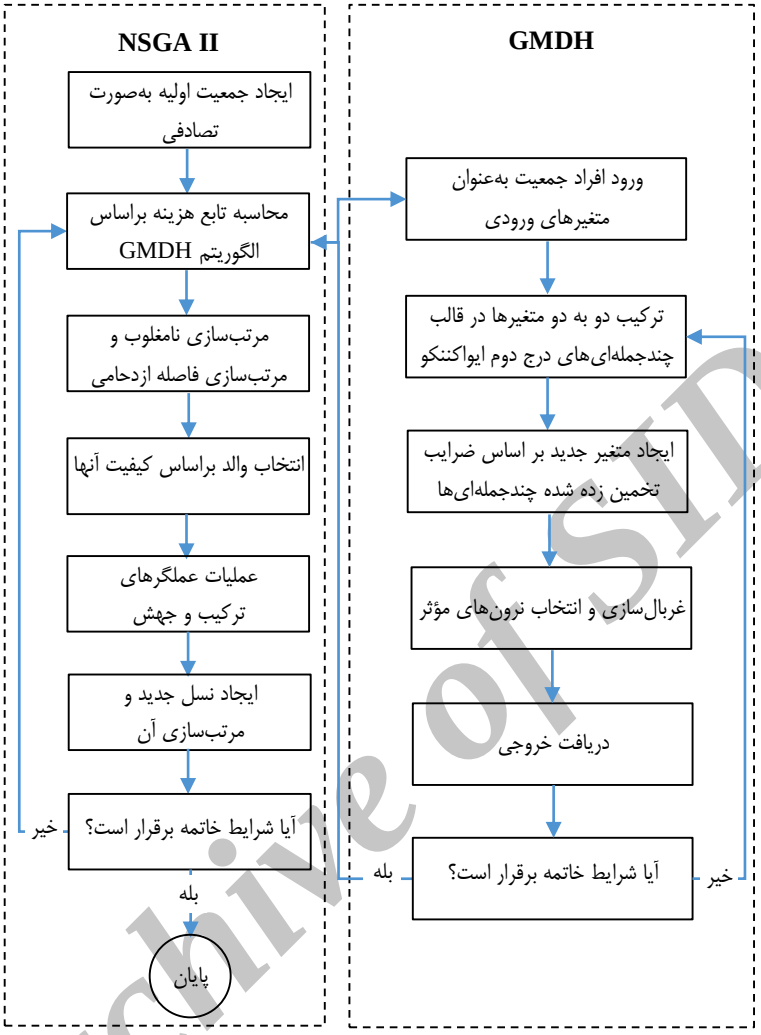
f_j^{i+1} مقدار تابع هر jام در جواب $i+1$، f_j^{i-1} مقدار تابع هر jام در جواب $i-1$، f_j^{max} بیشترین مقدار تابع هدف در جبهه jام، f_j^{min} کمترین مقدار تابع هدف در جبهه jام، d_i^j فاصله ازدحامی جواب iام در جبهه jام و فاصله ازدحامی جواب iام در اهداف i است. سپس با استفاده از روش انتخاب تورنمنت چند هدفه، مجموعه‌ای از افراد به‌عنوان والدین برای تولید نسل جدید انتخاب می‌شوند؛ به این صورت که هر چه رتبه جواب کمتر و دارای فاصله ازدحامی بیشتری باشد، احتمال انتخاب‌شدن آن برای تولید نسل بعدی بیشتر است.

گام چهار: برای ایجاد جمعیت جدید از والدین انتخاب‌شده در گام قبل، از دو عملگر تقاطع[1] و جهش[2] استفاده می‌شود. عملگر تقاطع به‌عنوان عملگر اصلی تولید نسل جدید، بین افراد انتخاب‌شده از نسل قبل، فرایند تولید مثل ایجاد می‌کند. بدین ترتیب افراد انتخاب‌شده به‌طور تصادفی و دوتایی با هم ترکیب‌شده تا دو فرد (دو فرزند) جدید ایجاد شود. برای ترکیب کروموزوم والدین به‌منظور تولید فرزندان، سه عملگر برش تک‌نقطه‌ای[3]، برش دونقطه‌ای[4] و برش یکنواخت[5] استفاده می‌شود که تفاوت عملکرد آنها در انتخاب نقاط متفاوت از رشته کروموزوم والدین برای برش به‌منظور ترکیب و ایجاد فرزندان است. برای بهره‌گیری از مزایای هر سه روش تقاطع، انتخاب از میان آنها به چرخ رولتی[6] سپرده می‌شود که برای تولید فرزندان ۱۰ درصد از برش یک‌نقطه‌ای، ۲۰ درصد از برش دونقطه‌ای و ۷۰ درصد از برش یکنواخت استفاده می‌کند. بدین‌ترتیب جمعیت فرزندان ایجاد می‌شود. شایان ذکر است که اندازه جمعیت فرزندان ۸۰ درصد اندازه جمعیت اولیه تعیین شده است. برای برانگیختگی بیشتر در روند ایجاد نسل جدید، از عملگر جهش نیز استفاده می‌شود. این عملگر طی یک فرایند تصادفی، بیت‌هایی از رشته کروموزوم یک والد را انتخاب کرده و آن را با بیت‌هایی از کروموزوم والد دیگر که به‌طور تصادفی انتخاب‌شده، جابه‌جا می‌کند. این سازوکار به افزایش سرعت همگرایی الگوریتم کمک می‌کند و مانع سلطه افراد مشابه بر یک نسل می‌شود. اندازه جمعیت جهش‌یافته معادل ۲۰ درصد اندازه جمعیت اولیه بوده و مقدار جهش ۵ درصد است.

گام پنج: به‌منظور ایجاد جمعیت جدید، در این گام جمعیت قدیم، جمعیت فرزندان و جمعیت جهش‌یافتگان با یکدیگر ادغام می‌شوند. افراد جمعیت تازه تشکیل‌شده ابتدا برحسب رتبه و بعد برحسب فاصله ازدحامی مرتب‌سازی می‌شوند و به اندازه جمعیت اولیه، افراد از بالای فهرست مرتب‌شده به‌عنوان جمعیت جدید انتخاب می‌شوند.

گام شش: در پایان بررسی می‌شود که آیا شرایط خاتمه محرز شده است یا نه. اگر شرایط خاتمه محقق شده باشد، الگوریتم به پایان می‌رسد و چنانچه شرایط خاتمه ایجاد نشود، الگوریتم بار دیگر از گام دوم تکرار می‌شود.

1. Crossover
2. Mutation
3. Single point crossover
4. Double point crossover
5. Uniform crossover
6. Roulette wheel



شکل ۲. نمایی از الگوریتم NSGA II-GMDH

معیار ارزیابی عملکرد

برای ارزیابی کارایی و دقت پیش‌بینی مدل‌ها، از سه معیار ریشه میانگین مربعات خطا (RMSE)، درصد میانگین قدر مطلق خطا (MAPE) و ضریب نابرابری تایل(TIC)[1] استفاده شده است که به‌صورت زیر محاسبه می‌شوند.

$$RMSE = \sqrt{1/N \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad \text{رابطه ۸)}$$

$$MAPE = 100/N \sum_{i=1}^{N} |y_i - \hat{y}_i|/y_i \quad \text{رابطه ۹)}$$

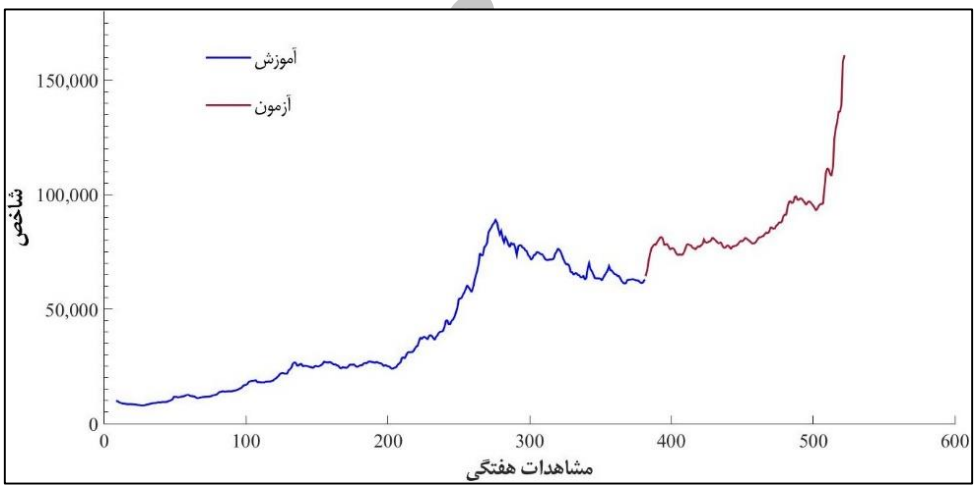
1. Theil inequality coefficient

$$TIC = \frac{\sqrt{\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{N}}}{\sqrt{\frac{\sum_{i=1}^{N}(y_i)^2}{N}}+\sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i)^2}{N}}}$$ رابطه ۱۰)

N مجموع تعداد مشاهدات، y_i مقادیر واقعی و $\hat{y}_i$ مقادیر پیش‌بینی شده هستند.

یافته‌های پژوهش

داده‌های این پژوهش شامل سری زمانی داده‌های هفتگی شاخص کل بورس اوراق بهادار تهران (TEDPIX)[1] برای دوره ۱۰ ساله (از مهر ۱۳۸۷ تا شهریور ۱۳۹۷) است. این داده‌ها از پایگاه داده بورس اوراق بهادار تهران[2] استخراج شده‌اند. شکل ۳، سری شاخص کل بورس را نشان می‌دهد. این سری حاوی ۵۲۲ مشاهده است که ۵۲ مشاهده اول صرف ایجاد وقفه‌های مختلف به‌عنوان متغیرهای ورودی بالقوه شده است، پس از آن، ۷۰ درصد ابتدایی داده‌ها (۳۲۹ مشاهده اول) در فرایند آموزش و برازش مدل‌ها به‌کار رفته و ۳۰ درصد باقی مانده داده‌ها (۱۴۱ مشاهده آخر) برای آزمون و ارزیابی مدل‌ها استفاده شده است.



شکل ۳. سری زمانی شاخص کل بورس اوراق بهادار تهران از سال ۸۷ تا ۹۷

جدول ۱، خلاصه آمار توصیفی و آزمون ریشه واحد سری شاخص کل بورس را نشان می‌دهد. در حالی که میانگین سری شاخص کل بورس در پانل الف این جدول عدد ۵۱،۲۵۲ را نشان می‌دهد، نیمی از مشاهدات کمتر از مقدار ۶۱،۳۰۷ است. این توزیع نامتقارن مشاهدات به سمت مقادیر بزرگ‌تر، نشان از چولگی مثبت مجموعه داده‌ها دارد. کشیدگی توزیع سری شاخص کل بورس نیز مقدار اندکی کمتر از توزیع نرمال است. نتایج آزمون ریشه واحد دیکی ـ فولر تعمیم یافته

1. Tehran Stock Exchange Dividend and Price Index (TEDPIX)
2. www.tse.ir

(ADF)[1] در پانل ب حاکی از آن است که سری شاخص کل بورس در سطح مانا نیست، اما تفاضل مرتبه اول سری با اطمینان ۹۹ درصد ماناست. از این رو مشکلی از بابت به‌کارگیری مدل ARIMA برای مدل‌سازی حرکت شاخص و پیش‌بینی آن وجود نخواهد داشت. شایان ذکر است که کلیه مدل‌ها در نرم‌افزار متلب برنامه‌نویسی و اجرا شده است.

جدول ۱. آمار توصیفی و آزمون ریشه واحد سری زمانی شاخص کل بورس

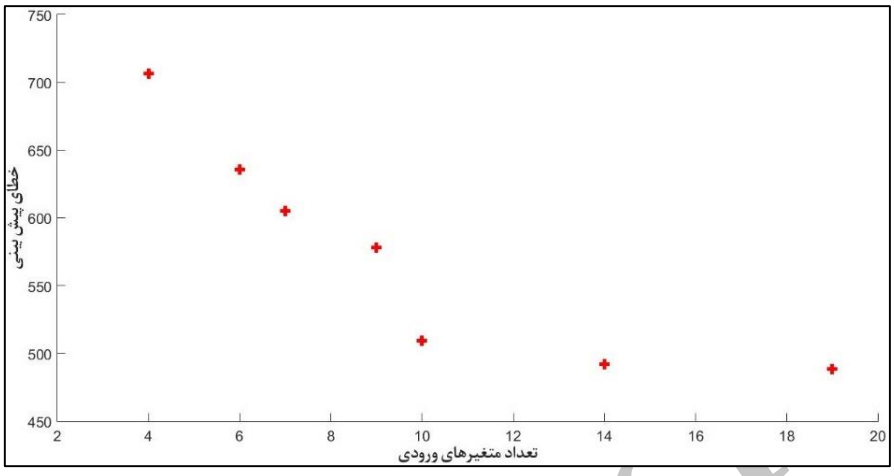
پانل الف. آمار توصیفی			
میانگین	۵۱۲۵۲	میانه	۶۱۳۰۷
چارک اول	۲۳۷۵۶	چارک سوم	۷۷۶۱۷
حداقل	۷۹۶۶	حداکثر	۱۶۰۵۳۸
کشیدگی	۲/۲۱	چولگی	۰/۲۶
انحراف معیار	۳۱۷۴۴	تعداد مشاهدات	۵۲۲

پانل ب. آزمون ریشه واحد		
آماره آزمون	**در سطح**	**تفاضل مرتبه اول**
دیکی ـ فولر تعمیم یافته	(۱/۹۷-)	(۲/۵۹-)***

*معناداری در سطح ۱۰ درصد **معناداری در سطح ۵ درصد ***معناداری در سطح ۱ درصد

در برازش مدل ARIMA$(p.d.q)$ برای شناسایی الگو، از نمودارهای AC و PAC استفاده شده است. برای ارزیابی الگوی برآوردشده، مناسب‌ترین الگو بر اساس معیار اطلاعات آکائیک انتخاب می‌شود. بدین ترتیب الگوی ARIMA(۹،۱،۱۰) به‌عنوان مدلی با کمترین میزان معیار اطلاعات آکائیک (۱۶/۴۴۸) انتخاب شده است. مدل ترکیبی NSGA II–GMDH با بهره‌گیری از یک تابع هزینه دوهدفه برای بهینه‌سازی هم‌زمان میزان خطای پیش‌بینی (RMSE) و وقفه‌های متناظر با آن، پس از ۱۰۰ بار تکرار با جمعیت اولیه ۲۰۰، به جبهه جواب بهینه رسیده است. در شکل ۴، جبهه پارتوی به‌دست آمده از مدل ترکیبی که نشان‌دهنده خطای پیش‌بینی به ازای تعداد وقفه‌ها به‌عنوان متغیر ورودی است، به تصویر کشیده شده است. همان‌طور که مشخص است، هیچ‌یک از جواب‌های این جبهه نسبت به یکدیگر مغلوب نیستند. با توجه به بده‌بستان دقت پیش‌بینی و تعداد ورودی‌ها، به نظر می‌رسد جواب متشکل از ۱۰ وروری با خطای پیش‌بینی ۵۰۹/۴ بهترین جواب از میان جبهه پارتوی باشد. بدین ترتیب متغیرهای ورودی مدل ترکیبی NSGA II–GMDH شامل وقفه‌های اول، سوم، هفتم، نهم، دوازدهم، بیستم، بیست و چهارم، سی‌ام، سی و ششم، چهل و هفتم می‌شود.

1. Augmented Dickey–Fuller unit root tests



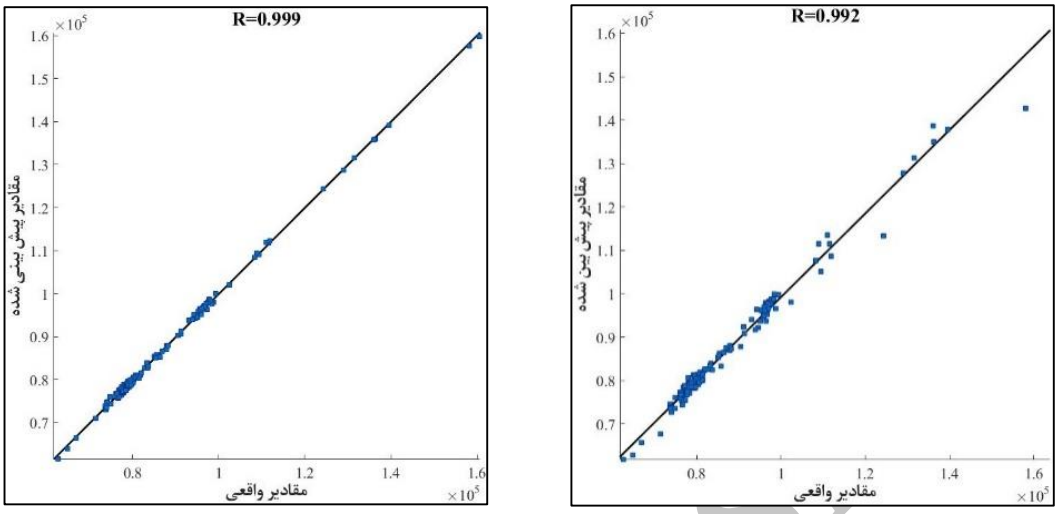
شکل ٤. جبهه پارتوی جواب بهیه مدل ترکیبی NSGA II-GMDH

شکل ۵ داده‌های پیش‌بینی شده به‌وسیله هر مدل را در برابر داده‌های واقعی در خارج از نمونه نمایش می‌دهد. این نمودار نشان می‌دهد که در گروه آزمون، مقادیر پیش‌بینی شده به‌وسیله الگوریتم NSGA II–GMDH به مراتب تطبیق بیشتری با داده‌های واقعی دارند. نتایج حاکی از این است که مشاهدات مدل NSGA II–GMDH پراکندگی کمتری نسبت به خط رگرسیون تجربه کرده و ضریب R^2 این مدل، بالاتر از مدل ARIMA است که قدرت بیشتر این مدل در پیش‌بینی رفتار شاخص کل را نشان می‌دهد.

جدول ۲ دقت پیش‌بینی هر یک از مدل‌های NSGA II–GMDH و ARIMA را بر اساس معیارهای ارزیابی عملکرد RMSE، MAPE و TIC در خارج از نمونه ارائه می‌کند. همسو با نتایج قبلی، مدل ترکیبی NSGA II–GMDH نسبت به مدل ARIMA از دقت پیش‌بینی بهتری در گام آزمون برخوردار است؛ به‌طوری که MAPE پیش‌بینی شاخص کل بورس برای مدل ترکیبی برابر ۰/۷۵درصد است. مقدار RMSE مدل NSGA II–GMDH در مقایسه با مدل ARIMA، ۱۳۵۳ واحد کمتر بوده و TIC کمتر این مدل نیز مؤید دقت پیش‌بینی بالاتر آن است. این حقیقت که مدل پیشنهاد شده NSGA II–GMDH پیش‌بینی دقیق‌تری ارائه می‌کند، از به‌کارگیری هم‌زمان توانایی شبکه خودسازمانده GMDH برای مقاومت در برابر نویز و ویژگی جست‌وجوی جهانی و بهینه‌سازی چند هدفه الگوریتم NSGA II در کمینه‌سازی خطای پیش‌بینی و یافتن متغیرهای ورودی بهینه نشئت می‌گیرد.

جدول ۲. نتایج عمکرد مدل ها

مدل	RMSE	MAPE	TIC
ARIMA	۲۰۶۵/۷	۱/۲۹	۰/۰۱۱۷
NSGA II-GMDH	۷۱۳/۲	۰/۷۵	۰/۰۰۴۰



مدل ARIMA مدل ترکیبی NSGA II–GMDH

شکل ۵. مقادیر پیش‌بینی در برابر مقادیر واقعی سری زمانی شاخص کل

همین ویژگی‌ها به تولید نتایج بهتر با دقت پیش‌بینی ۹۹/۲۵ درصدی برای شاخص کل بورس منجر شده است. همسو با تحقیقات تجربی، نتایج پژوهش حاضر نشان می‌دهد که مدل ترکیبی درک مناسب‌تری از رابطه غیرخطی و پیچیده حاکم بر حرکات شاخص ایجاد می‌کند (ابود و دشپند، ۲۰۰۸؛ تاسی و وانگ، ۲۰۰۹؛ چنگ و همکاران، ۲۰۱۰؛ سان و ژآئو، ۲۰۱۵ و گریگوریان، ۲۰۱۶). این موضوع که مدل ترکیبی NSGA II–GMDH تنها با استفاده از وقفه‌های خود شاخص به این سطح از دقت پیش‌بینی رسیده است، علاوه بر توانایی مدل می‌تواند به‌دلیل وجود همبستگی متوالی میان عناصر سری زمانی شاخص باشد. بنابراین در تقابل با نتایج گالوتی و اسکینتاولی (۱۹۹۴)، لاورنس (۱۹۹۷) و تیمرمن و گرنجر (۲۰۰۴)، وجود این روند غیرتصادفی در سری زمانی شاخص کل، می‌تواند حاکی از ناکارایی بورس اوراق بهادار تهران در سطح ضعیف و رد فرضیه گشت تصادفی قیمت‌ها باشد.

با اینکه نتایج تجربی پژوهش حاضر نشان‌دهنده بهبود عملکرد پیش‌بینی سری زمانی شاخص کل به‌وسیله مدل ترکیبی NSGA II–GMDH است، برای آزمون فرضیه پژوهش، اثبات اهمیت آماری این نتایج نیز لازم است. به‌منظور آزمون این فرضیه که عملکرد مدل ترکیبی NSGA II-GMDH برای پیش‌بینی شاخص کل بورس اوراق بهادار در خارج از نمونه بهتر از مدل ARIMA است، به پیروی از ژانگ، هی و لیاتسیس[1] (۲۰۱۲) از آزمون‌های ناپارامتریک فریدمن[2] و ویلکاکسون[3] استفاده شده است. جدول ۳ نتایج این دو آزمون را نشان می‌دهد. پانل الف این جدول، نتایج آزمون فریدمن را که با فرضیه صفر برابری میزان دقت پیش‌بینی مدل‌ها صورت می‌پذیرد، ارائه می‌کند. در ستون اول پانل الف، میانگین رتبه دو مدل و در ستون دوم، آماره کای ـ دو آزمون و معناداری مطابق آن درج شده است. نتایج آزمون فریدمن نشان می‌دهد در سطح معناداری ۱ درصد، فرضیه .H، یعنی برابری دقت پیش‌بینی دو مدل رد شده و می‌توان با اطمینان ۹۹

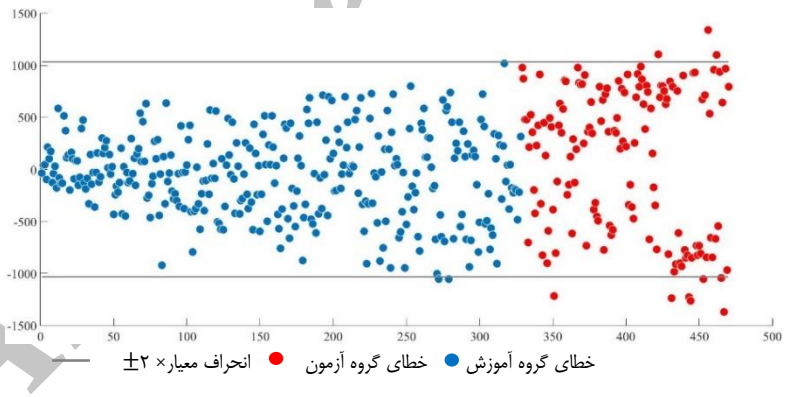
1. Zhang, He, & Liatsis
2. Friedman test
3. Wilcoxon test

درصد ادعا کرد که دقت پیش‌بینی دو مدل متفاوت است. نیز نتایج آزمون ویلکاکسون برای بررسی اینکه کدام مدل عملکرد بهتری دارد، در پانل ب جدول ۳ ارائه شده است. نتایج حاکی از معناداری آماره Z متعلق به این جفت مدل در سطح ۱درصد بوده و فرضیه صفر آزمون (برابری دقت پیش‌بینی) رد می‌شود. بنابراین می‌توان تأیید کرد که در سطح خطای ۱ درصد، مدل NSGA II–GMDH در پیش‌بینی شاخص کل بورس اوراق بهادار عملکرد بهتری نسبت به مدل ARIMA از خود نشان می‌دهد. با توجه به اطمینان از عملکرد بهتر مدل NSGA II–GMDH، در ادامه نتایج این مدل تحلیل می‌شود.

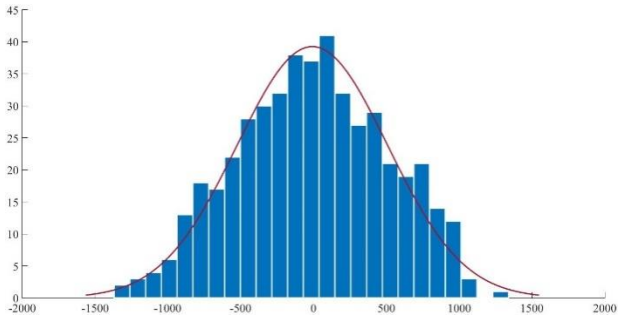
جدول ۳. آزمون‌های ناپارامتریک فریدمن و ویلکاکسون

پنل الف. آزمون فریدمن		
مدل	**میانگین رتبه**	**آماره کای ـ مربع**
ARIMA	۱/۶۴۱	۹۸/۷۴***
NSGA II-GMDH	۱/۲۵۴	
پنل ب. آزمون ویلکاکسون		
مدل	**آماره Z**	
NSGA II-GMDH Vs. ARIMA	-۶/۰۴***	

*معناداری در سطح ۱۰ درصد **معناداری در سطح ۵ درصد ***معناداری در سطح ۱درصد



پانل الف. پراکندگی جملات‌خطا

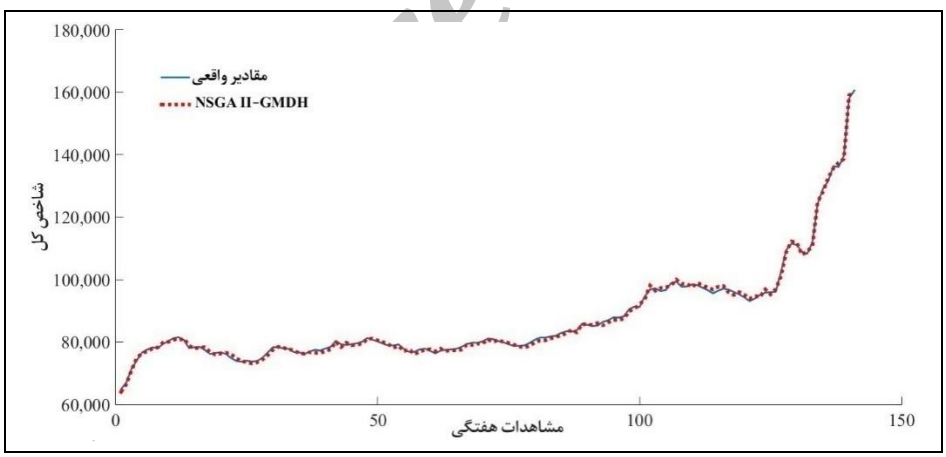


پانل ب. توزیع جملات‌خطا

شکل ٦. پراگندکی و توزیع جملات‌خطا مدل NSGA II-GMDH

پراکندگی و توزیع جملات خطای مدل NSGA II–GMDH در شکل ۶ ارائه شده است. پانل الف این نمودار پراکندگی مقادیر پیش‌بینی شده را در حد فاصل مثبت/منفی ۲ انحراف معیار خطا نشان می‌دهد. با توجه به برازش مدل به‌وسیله داده‌های گروه آموزش، مطابق انتظار پراکندگی خطا در این گروه چندان چشمگیر نیست؛ اما با وجود افزایش نوسان‌های شاخص کل در دوره آزمون، پراکندگی مدل در این دوره چندان افزایش نیافته و به‌ندرت از دو انحراف معیار خطا تجاوز می‌کند. با وجود این، تمرکز خروجی گروه آزمون در اطراف خط ۲± انحراف معیار نشان‌دهنده پراکندگی بیشتر مدل در دوره آزمون نسبت به دوره آموزش است. پانل ب نیز با به تصویرکشیدن هیستوگرام باقی‌مانده‌ها نشان می‌دهد توزیع جملات خطا فاصله شایان توجهی با توزیع نرمال نداشته و گسستگی چندانی در میله‌های نمودار مشاهده نمی‌شود.

در شکل ۷ مقایسه مقادیر پیش‌بینی‌شده مدل NSGA II–GMDH و مقادیر واقعی شاخص کل بورس اوراق بهادار تهران برای داده‌های خارج از نمونه به تصویر کشیده شده است. واضح است که مدل پیشنهاد شده به‌خوبی قادر به پوشش تغییرات ناپایدار روند حرکت و الگوسازی رفتار شاخص کل است. با وجود نوسان‌های زیاد شاخص در دامنه ۴۶ تا ۱۶۰ هزار واحدی طی دوره آزمون و نامانایی مشهود سری، این مدل همچنان قدرت پیش‌بینی‌کنندگی بالایی از خود نشان می‌دهد. دقت پیش‌بینی بالا برای داده‌های خارج از نمونه، مؤید توانایی این مدل در تعمیم نتایج گروه آموزش به داده‌هایی است که پیش از این آنها را تجربه نکرده است.



شکل ۷. مقادیر واقعی و پیش بینی شده مدل NSGA II-GMDH در خارج از نمونه

نتیجه‌گیری

موفقیت مدل‌های ترکیبی در پیش‌بینی و مدل‌سازی سیستم‌های پیچیده طی چند سال اخیر، توجه محققان را به این بخش از ادبیات موضوعی جلب کرده است (چنگ و همکاران، ۲۰۱۰؛ سان و ژائو، ۲۰۱۵ و گریگوریان، ۲۰۱۶). چنین توجهی ممکن است تا حدی مربوط به این واقعیت باشد که مدل‌های ترکیبی علاوه بر بهره‌مندی هم‌زمان از ویژگی‌ها و مزیت‌های چند مدل، ضعف‌های یکدیگر را پوشش می‌دهند. در مقاله حاضر، ضمن ارزیابی عملکرد مدل ترکیبی جدید NSGA II– GMDH متشکل از شبکه GMDH و الگوریتم NSGA II، در پیش‌بینی شاخص کل بورس اوراق بهادار تهران، قدرت

پیش‌بینی آن با مدل ARIMA مقایسه شد. نتایج پژوهش نشان داد که دقت پیش‌بینی مدل ترکیبی NSGA II–GMDH در داده‌های خارج از نمونه به مراتب بیشتر از مدل ARIMA است و مدل پیشنهاد شده در این پژوهش، توانایی زیادی در پوشش تغییرات ناپایدار روند حرکت شاخص کل از خود نشان می‌دهد؛ زیرا این مدل، هم‌زمان از مزیت‌های شبکه GMDH در انعطاف‌پذیری و مقاومت در برابر داده‌های نویزی و نامانا در کنار ویژگی جست‌وجوی جهانی و بهینه‌سازی چندهدفه الگوریتم NSGA II در کمینه‌سازی خطای پیش‌بینی و یافتن متغیرهای ورودی بهینه بهره می‌گیرد.

منابع

فخاری، حسین؛ ولی‌پور خطیر، محمد؛ موسوی، سیده مائده (۱۳۹۶). بررسی عملکرد شبکه عصبی بیزین و لونبرگ مارکوات در مقایسه با مدل‌های کلاسیک در پیش‌بینی قیمت سهام شرکت‌های سرمایه‌گذاری. *فصلنامه علمی ـ پژوهشی تحقیقات مالی*، ۱۹(۲)، ۳۱۸-۲۹۹.

درودی، دیاکو؛ ابراهیمی، سید بابک (۱۳۹۵). ارائه روش هیبریدی نوین برای پیش‌بینی شاخص کل قیمت بورس اوراق بهادار. *فصلنامه علمی ـ پژوهشی تحقیقات مالی*، ۱۸(۴)، ۶۱۳-۶۳۲.

شمس، شهاب‌الدین؛ ناجی زواره، مرضیه (۱۳۹۴). بررسی مقایسه‌ای بین مدل ترکیبی سیستم ژنتیک فازی ـ عصبی خودسازمانده و مدل خطی در پیش‌بینی قیمت توافقی قراردادهای آتی سکه طلا. *فصلنامه علمی ـ پژوهشی تحقیقات مالی*، ۱۷(۲)، ۲۳۹-۲۵۸.

مهرآرا، محسن؛ معینی، علی؛ احراری، مهدی؛ هامونی، امیر (۱۳۸۸). الگوسازی و پیش‌بینی شاخص بورس اوراق بهادار تهران و تعیین متغیرهای مؤثر بر آن. *فصلنامه پژوهش‌ها و سیاست‌های اقتصادی*، ۱۷(۵۰)، ۳۱-۵۱.

References

Abbod, M. & Deshpande, K. (2008). Using Intelligent Optimization Methods to Improve the Group Method of Data Handling in Time Series Prediction. In M. Bubak, G. D. V. Albada, J. Dongarra & P. M. A. Sloot (Eds.), *International Conference on Computational Science* (pp.16–25). Poland: June.

Cheng, C.H., Chen, T.L. & Wei, L.Y. (2010). A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Information Sciences*, 180(9), 1610–1629.

Deb, K., Pratap, A., Agrawal, S. & Meyarivan, T. (2002). Fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transaction on Evolutionary Computation*, 6(2), 182–197.

Dorodi, D., Abrahimi, S. B. (2017). Presenting a new hybrid method for predicting the Stock Exchange price inde. *Financial Research Journal*, 18(4), 613-632. (*in Persian*)

Ebadati, O. M. & Mortazavi, M. (2016). An efficient hybrid machine learning method for time series stock market forecasting. *Neural Network World*, 28(1), 41–55.

Fakhari, H., Valipour Khatir, M., Mousavi, M. (2017). Investigating Performance of Bayesian and Levenberg-Marquardt Neural Network in Comparison Classical Models in Stock Price Forecasting. *Financial Research Journal*, 19(2), 299-318. (*in Persian*)

Galeshchuk, S. (2016). Neural networks performance in exchange rate prediction. *Neurocomputing*, 172(8), 446–452.

Galotti, M. & Schiantavelli, F. (1994). Stock mare volatility and investment: Do only fundamental matters? *Economica*, 61(242), 144–165.

Gharleghi, B., Shaari, A.H. & Shafighi, N. (2014). Predicting exchange rates using a novel "cointegration based neuro-fuzzy system". *International Economics*, 137(1), 88–103.

Giles, C.L., Lawrence, S. & Tsoi, A.C. (2001). Noisy time series prediction using recurrent neural networks and grammatical inference. *Machine Learning*, 44(1-2), 161–183.

Grigoryan, H. (2016). Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA). *Database Systems Journal*, 7(1), 12–21.

Ivakhnenko, A.G. (1968). Group Method of Data Handling - a rival of the method of stochastic approximation. *Soviet Automatic Control*, 3(1), 43–55.

Lawrence, R. (1997). *Using Neural Networks to Forecast Stock Market Prices*. Canada: University of Manitoba.

Lu, C.J., Lee, T.S. & Chiu, C.C. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, 47(2), 115–125.

Mehrara, M., Moeini, A., Ahrari, M. & Hamony, A. (2009). Modeling stock market prices based on GMDH Neural Network: a case study for Iran. *Quarterly Journal of Economic Research and Policies*, 17(50), 31–51. (*in Persian*)

Müller, J.A. & Lemke, F. (2000). *Self-Organising Data Mining: Extracting Knowledge from Data*. Hamburg: BoD.

Rafiuzzaman, M. (2014). Forecasting Chaotic Stock Market Data using Time Series Data Mining. *International Journal of Computer Applications*, 101(10), 27–34.

Shams, Sh., Naji Zavareh, M. (2015). Comparison Between the Hybrid Model of Genetic Fuzzy and Self - Organizing Systems and Linear Model to Predict the Price of Gold Coin Futures Contracts. *Financial Research Journal,* 17(2), 239-258. (*in Persian*)

Sun, Y. & Gao, Y. (2015). An improved hybrid algorithm based on PSO and BP for stock price forecasting. *The Open Cybernetics & Systemics Journal*, 9(1), 2565–2568.

Timmermann, A. & Granger, C. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20(1), 15–27.

Tsai, C. F. & Wang, S.P. (2009). Stock Price Forecasting by Hybrid Machine Learning Techniques. *In A. SIO-IONG, O. Castillo, C. Douglas, D. Dagan-Feng & L. Jeong-A(Eds.), Proceedings of the International Multi Conference of Engineers and Computer Scientists (pp.57–63).* Hong Kong: March.

Tsai, C.Y. & Huang, C.L. (2009). A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. *Expert Systems with Applications*, 36(2), 529–1539.

Voss, M.S. & Howland, J.C. (2003). Financial modelling using social programming. *In M. H. Hamza (Eds.), Financial Engineering and Applications (pp.16–25).* Canada: July.

Yang, C.H., Liao, M.Y., Chen, P.L., Huang, M.T., Huang, C.W., Huang, J.S. & Chung, J.B. (2009). Constructing financial distress prediction model using group method of data handling technique. *In Proceedings of the eighth, International conference on machine learning and cybernetics (pp. 2897–2902).* China: July.

Yu, L., Wang, S.Y. & Lai, K.K. (2009). A neural-network-based nonlinear metamodeling approach to financial time series forecasting. *Applied Soft Computing*, 9(2), 563–574.

Zhang, M., He, C. & Liatsis, P. (2012). A D-GMDH model for time series forecasting. *Expert Systems with Applications*, 39(5), 5711–5716.

Zhu, B., He, C. & Liatsis, P. (2012). A robust missing value imputation method for noisy data. *Applied Intelligence*, 36(1), 61–74.