مجله بررسی‌های حسابداری دوره ۳، شماره ۱۰، بهار ۱۳۹۵، صفحات ۱۹- ۴۰

پیش‌بینی ریسک سیستماتیک شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از الگوریتم‌های کلونی مورچه‌ها و لارس

محسن حمیدیان[۱]، سیدجواد حبیب‌زاده‌بایگی[۲]، مریم سلمانیان[۳]، سیدحسام وقفی*[۴]

چکیده

تصمیم‌گیری‌های مالی و اقتصادی به دلیل وجود عدم اطمینان آتی، همواره با ریسک مواجه هستند. در این راستا، گزارشگری مالی می‌تواند دربرگیرندهٔ اطلاعات مفیدی دربارهٔ پیش‌بینی ریسک باشد. هدف این پژوهش بررسی پیش‌بینی ریسک سیستماتیک بر اساس الگوریتم کلونی مورچگان و الگوریتم لارس طی سال‌های ۱۳۸۹ تا ۱۳۹۳ برای شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می‌باشد. بر این اساس ۱۵۴ شرکت (۷۷۰ سال- شرکت) به عنوان نمونهٔ آماری انتخاب شده است. در تحقیق حاضر برای سنجش میزان ریسک سیستماتیک شرکت از عامل بتا استفاده شده است. نتایج این تحقیق حاکی از آن است که الگوریتم کلونی مورچگان با خطای ۱/۲۵۲ درصد و الگوریتم لارس با ۱/۵۶۳ درصد توانایی پیش‌بینی ریسک سیستماتیک را دارند. در واقع می‌توان اذعان نمود که الگوریتم‌های به کارگرفته شده با دقت بالایی توانایی کشف ریسک سیستماتیک را دارد.

واژه‌های کلیدی: پیش‌بینی ریسک سیستماتیک، الگوریتم کلونی مورچگان، الگوریتم لارس.

۱. استادیار حسابداری، دانشگاه آزاد واحد تهران جنوب
۲. دانشجوی دکتری حسابداری، دانشگاه آزاد واحد تهران جنوب
۳. دانشجوی دکتری حسابداری، دانشگاه آزاد واحد تهران جنوب
۴. مربی حسابداری و دانشجوی دکتری دانشگاه پیام نور تهران- ایران

تاریخ دریافت مقاله: ۹۵/۱/۳۰
تاریخ پذیرش مقاله: ۹۵/۳/۹
*نویسندهٔ مسئول: سیدحسام وقفی
h.vaghfi2012@gmail.com

مقدمه

در بازار سرمایه، سرمایه‌گذاران و تصمیم‌گیران مالی سعی دارند به شکلی سرمایه‌گذاری کنند که بیش‌ترین بازدهی را برای آنان به ارمغان آورد و در این راستا به ریسک مربوط به سرمایه‌گذاری نیز توجه دارند و در صورتی تحمل ریسک را پذیرا می‌شوند که بابت آن چیزی عایدشان شود و این امر چیزی جز سود بیش‌تر سرمایه‌گذاری‌ها نخواهد بود. از سوی دیگر، شرکت‌ها نیز تلاش دارند به گونه‌ای فعالیت کنند که افزایش ارزش شرکت را محقق نمایند (پیری و همکاران، ۱۳۹۲). هدف سرمایه‌گذاران از انجام سرمایه‌گذاری به حداکثر رساندن ثروتشان می‌باشند. بر این اساس اگر سرمایه‌گذاران قادر باشند ریسک و بازده اوراق بهادار را پیش‌بینی کنند. در این صورت می‌تواند در مورد خرید، نگهداری و یا فروش اوراق بهادار تصمیم‌گیری کنند. از آنجا که سرمایه‌گذاران منطقی به دنبال بیش‌تر کردن بازده خود در سطح قابل قبولی از ریسک و سرمایه‌گذاری در شرکت‌ها با ارزش بالاتر هستند، به معیارهایی نیاز دارند تا بتوانند عملکرد سرمایه‌گذاری خود را پیش‌بینی کنند (هرست و همکاران، ۲۰۰۸). سرمایه‌گذاران در فرایند تصمیم‌گیری توجه خاصی به ریسک سیستماتیک دارند. در این راستا، به نظر می‌رسد که آن‌ها در فرایند تصمیم‌گیری ابتدا نیازمند اطلاعاتی هستند تا بتوانند از آن طریق ریسک سیستماتیک مربوط به سرمایه‌گذاری خود را تعیین کنند (ملایی، ۱۳۹۰).

تصمیم‌گیری‌های مالی و اقتصادی به دلیل وجود نااطمینانی‌های آتی، همواره با ریسک مواجه هستند. بنابراین می‌توان از طریق ارائه الگوی مناسب به پیش‌بینی ریسک سرمایه‌گذاری کمک نمود. بر این اساس تصمیم‌گیری‌هایی که مطابق با این پیش‌بینی انجام می‌شود صحیح‌تر است. سرمایه گذاران به دلیل ریسک جدانشدنی سرمایه‌گذاری، انتظار بازده دارند. لذا برای سرمایه گذاران، آگاهی از ریسک و بازده سرمایه‌گذاری از اهمیت ویژه‌ای برخوردار است (قائمی و رحیم‌پور، ۱۳۸۹). ریسک سیستماتیک نیز به عنوان ریسک غیر قابل کنترل با تأثیرگذاری بر میزان سودآوری و ارزش بنگاه نقش مهمی را در تصمیم‌گیری‌های مالی ایفا می‌کند (بریمل، ۲۰۰۳). برای پیش‌بینی ریسک سیستماتیک می‌توان از مدل‌های مختلفی استفاده نمود. مدل‌های رگرسیونی الگوهایی فراهم می‌آورند که می‌توان بر پایهٔ آن‌ها ارتباط بین مجموعه‌ای از متغیرها را بررسی کرد. این متغیرها شامل متغیرهای مستقل (توضیحی/ پیش‌بینی‌کننده) و وابسته (پاسخ) هستند. در این مدل‌ها، بر پایهٔ مشاهده‌ها مربوط به متغیرهای مستقل و وابسته، تابعی به منظور پیش‌بینی و کنترل متغیر وابسته بنا می‌کنیم در رگرسیون معمولی فرض می‌شود که متغیرهای مورد مطالعه متغیرهای دقیق هستند و مشاهده‌ها مربوط به متغیرها نیز مشاهده‌های دقیق می‌باشند. اختلاف بین مقدار مشاهده شده برای متغیر وابسته و مقداری که از طریق مدل حاصل

می‌شود، و کلاً خطای مدل، به خطاهای تصادفی مربوط به مشاهده‌ها و اندازه‌گیری‌ها، عدم حضور برخی از متغیرها و... نسبت داده می‌شود. دربارهٔ جملات خطا و توزیع احتمالی آن، فرضیه‌هایی (مانند نرمال بودن، ناهمبسته بودن، ثبات واریانس و...) در نظر گرفته می‌شود، به گونه‌ای که بتوان بر پایهٔ این فرضیه‌ها تجزیه و تحلیل‌های آماری را دربارهٔ مدل انجام داد، مانند برآورد پارامترها، پیش‌بینی مقدار متغیر وابسته، آزمون‌های فرض مرتبط با مدل و غیره در این باره بکارگیری تبدیلات مختلف روی داده‌ها، کمک می‌کند تا بتوان احتمالاً به فرضیه‌ها لازم نزدیک شد. اما در بسیاری اوقات ممکن است یک یا چند فرض از فرض‌های بالا برقرار نباشد یا اینکه (مثلاً به دلیل حجم نمونهٔ اندک) نتوان از درستی بعضی فرض‌ها اطمینان حاصل کرد. مسلماً در این موارد مدل‌های رایج اعتبار و کارآیی لازم ندارند. در این موارد باید شیوه‌های جدیدی را جایگزین شیوه‌های کلاسیک نمود یکی از این شیوه‌های جایگزین، داده کاوی با استفاده از الگوریتم‌های هوشمند است. با توجه به رشد روزافزون داده‌های پیرامون بشر، به کارگیری ابزارهایی برای تحلیل داده‌ها و دستیابی به دانش نهفتهٔ درون آن‌ها الزامی می‌باشد. اخیراً داده‌کاوی به این دلیل که بشر دیدگاه ذاتی و بصری غیرقابل درکی درباره مسائل با ابعاد بالا و یا سایزهای بزرگ پایگاه داده دارد، از اهمیت بالایی برخوردار شده‌اند. داده کاوی به بررسی و تجزیه و تحلیل مقادیر زیادی از داده‌ها به منظور کشف الگوها و قوانین معنی‌دار اطلاق می‌شود. داده کاوی عمدتاً با ساختن مدل‌ها مرتبط است. یک مدل اساساً الگوریتم یا مجموعه‌ای از ورودی‌ها را با هدف یا مقصد خاصی مرتبط می‌نماید. هدف از داده کاوی یافتن بهترین الگوهای جذابی است که برآمده از میلیون‌ها بایت داده است. البته تنها یافتن الگوها کافی نیست، بلکه باید بتوان به الگوها پاسخ داده و بر اساس آن‌ها عمل نمود و در نهایت داده را به اطلاعات، اطلاعات را به عمل و عمل را به ارزش تبدیل کرد. این همان چرخهٔ تعالی داده کاوی است. چرخه تعالی داده کاوی یک فرآیند یادگیری تکراری و مرحله‌ای است که بر اساس نتایج در طول زمان تهیه می‌گردد. اگرچه این الگوریتم‌ها مهم‌اند، اما راه حل‌های داده کاوی چیزی فراتر از مجموعه‌ای از تکنیک‌ها و ساختارهای داده‌ای قوی است. این تکنیک‌ها را باید در جای مناسب و با داده‌های صحیح به کار برد. در این تحقیق نیز تلاش می‌گردد با استفاده از رویکرد داده کاوی و با به کارگیری الگوریتم‌های کلونی مورچه‌ها و لارس به پیش‌بینی ریسک سیستماتیک پرداخته شود. یکی از مهم‌ترین اهداف سرمایه‌گذاری در اوراق بهادار کسب سود سالیانه به همراه منفعت سرمایه می‌باشد. سرمایه‌گذاران جهت تحقق این هدف سعی می‌کنند در سهامی که کم‌ترین ریسک و بالاترین بازده را دارد، سرمایه‌گذاران کنند (یحیی‌زاده و حجابی، ۱۳۸۹). هدف از ارائهٔ این تحقیق کمک به سرمایه‌گذاران در جهت انتخاب مناسب سرمایه‌گذاری می‌باشد. در این

راستا، یکی عوامل مؤثری که بر تصمیم‌گیری سرمایه‌گذاران تأثیر می‌گذارد آگاهی از میزان ریسک شرکت‌ها می‌باشد. بنابراین باید روشی مشخص شود که بر آن اساس بتوان ریسک را به طور دقیق پیش‌بینی نمود. این پژوهش با استفاده از الگوریتم کلونی مورچگان و الگوریتم لارس سعی در پیش‌بینی مناسب ریسک دارد.

مبانی نظری

یکی از الزامات رشد و توسعه اقتصادی توجه بیش‌تر به بازارهای مالی است. فعالان در این بـازار همواره به دنبال کسب بازده بیش‌تر با ریسک معقول می‌باشند. شـرکت‌هـا بـرای افـزایش ارزش خود، سعی در کاهش ریسک سرمایه‌گذاری دارند. ریسک سرمایه‌گذاری عبارت است از نوسان در بازدهی‌های تحقق یافته و مورد انتظار سرمایه‌گذاری (پیری و همکاران، ۱۳۹۲).

در حالت کلی دو نوع ریسک شامل ریسک سیستماتیک و ریسک غیرسیستماتیک وجود دارد. ریسک غیرسیستماتیک فقط منحصر به یک دارایی می‌باشد برای اینکه آن ریسک در ارتباط با بخشی از بازده یک دارایی می‌باشد. این میزان از ریسک مختص یک شرکت یا یک صنعت می‌باشد و ناشی از عوامل و پدیده‌هائی مانند اعتصابات‌کارگری، عملکرد مدیریت، رقابت تبلیغاتی، تغییر در سلیقهٔ مصرف‌کنندگان و غیره می‌باشد. ریسک سیستمایتک آن قسمت از ریسک می‌باشد که به شرایط عمومی بازار مربوط است. تغییر نرخ بهره، نرخ برابری پول ملی در مقابل استعمار خارجی، نرخ تورم، سیاست‌های پولی و مالی، شرایط سیاسی و غیره از منابع ریسک سیستماتیک می‌باشند. هر تغییر در عوامل فوق‌الذکر بر روی شرایط کلی بازار تأثیر می‌گذارد. تغییرات متغیرهای کلان اقتصادی از منابع اصلی ریسک سیستماتیک می‌باشد. ریسک نرخ بهره یکی از اجرای ریسک سیستماتیک است. ریسک نرخ بهره به عنوان تغییرپذیری بالقوه در بازده که ناشی از تغییرات نرخ بهرهٔ بازار است تعریف می‌شود. ریسک تورم جزء دیگر ریسک سیستماتیک است. ریسک تورم به عدم اطمینان نسبت به سطح نرخ تورم در سال‌های آینده اطلاق می‌گردد. ریسک تورم به عبارت دیگر به انحراف نرخ تورم واقعی مورد انتظار بستگی دارد. ریسک کشور منبع دیگر ریسک سیستماتیک است. ریسک کشور یا ریسک سیاسی به ثبات یک کشور از ابعاد اقتصادی و سیاسی مربوط می‌شود. این نوع ریسک در کشورهایی که از ثبات سیاسی و اقتصادی بالایی برخوردارند، پائین است. ریسک بازار نیز از منابع ریسک سیستماتیک است. تغییرپذیری در بازده اوراق که ناشی از انتظارات سرمایه‌گذاران است، ریسک بازار نام دارد. از دیدگاه سرمایه‌گذاران، هر چه کیفیت و کمیت اطلاعات منتشر شده توسط شرکت بالاتر باشد، عدم تقارن اطلاعاتی شرکت پایین‌تر و در نتیجه دورنمای عملکرد آتی شرکت روشن‌تر است. در

چنین حالتی سرمایه‌گذاران، برآوردهای دقیق‌تری در خصوص بازدهی سرمایه‌گذاری دارند و تعدیلات کم‌تری در برآوردهای انجام شده صورت می‌گیرد. این امر، ریسک سهام شرکت‌های دارای اطلاعات با کیفیت بالا را کاهش می‌دهد (حیدرپور و خواجه‌محمود، ۱۳۹۳). وجود یا عدم وجود اطلاعات در خصوص شرکت‌ها و سهام آن‌ها، می‌تواند بر ریسک کلی شرکت تأثیرگذار باشد چنانچه مدیریت به عنوان یک فرد مطلع، آینده شرکت را از طریق پیش‌بینی سود هر سهم برای سهامداران ترسیم نماید، ریسک اطلاعاتی کاهش یافته و سرمایه‌گذاران در ارزیابی جریانات نقدی آتی مطمئن‌تر عمل می‌نمایند (فرانسیس و همکاران، ۲۰۰۸). داشتن یا نداشتن اطلاعات در خصوص شرکت و سهام می‌توانند بر ریسک کلی تأثیرگذار باشد به همین جهت هزینهٔ سرمایه کاهش می‌یابد، ریسک اطلاعاتی کم می‌شود و سرمایه‌گذاران در ارزیابی جریان‌های نقدی آتی مطمئن‌تر عمل می‌کنند (فیرر و لاکس، ۲۰۰۷؛ راجرز و همکاران، ۲۰۰۹). با توجه به این که مدیران نسبت به افراد خارج از شرکت اطلاعات دقیق‌تری در اختیار دارند که افراد خارج از شرکت به آن دسترسی ندارند، می‌توانند با پیش‌بینی‌های دقیق و معتبر ریسک شرکت را کاهش دهند. بهبود کیفیت گزارشگری مالی و عدم تقارن اطلاعاتی دربارهٔ عملکرد شرکت، نوسانات بازده سهام و ریسک را کاهش می‌دهد (راج‌گوپال و موهان، ۲۰۱۱).

جورج و همکاران (۲۰۰۹) بیان می‌کند که کاهش اطلاعات نامتقارن مطلوب است، چرا که باعث کاهش ریسک پیش‌بینی‌ها می‌شود و هزینهٔ سرمایه را کاهش می‌دهد. بهبود افشاها و کیفیت گزارشگری مالی، نبود تقارن اطلاعاتی در زمینهٔ عملکرد شرکت و نوسانات قیمت سهام را کاهش می‌دهد و افزایش نوسان‌های بازده سهام برای افزایش نامتقارن اطلاعاتی است که جزئی از هزینهٔ سرمایه است. ایسلی و اوهاری (۲۰۰۴) معتقدند که کیفیت گزارشگری مالی می‌تواند ریسک سیستماتیک اطلاعاتی و ریسک غیرسیستماتیک و هزینهٔ سرمایه را تحت تأثیر قرار می‌دهد.

پیشینهٔ تحقیق

صالح‌نژاد و وقفی (۱۳۹۵) در مقاله خود نشان دادند که دقت پیش‌بینی سود هر سهم توسط مدیریت بر ارزش شرکت تأثیر معناداری دارد اما با ریسک سیستماتیک شرکت‌ها رابطهٔ معناداری ندارد. نتایج بررسی زلقی و همکاران (۱۳۹۳) نشان می‌دهد که بین دقت و دفعات پیش‌بینی سود اعلام شده از سوی مدیریت با ریسک غیر سیستماتیک شرکت ارتباط منفی و معناداری وجود دارد. نتایج تحقیق پیری و همکاران (۱۳۹۲) نشان می‌دهد که با استفاده از روش داده‌های تابلویی پویا رابطهٔ منفی و معکوسی بین ریسک سیستماتیک و ارزش افزودهٔ اقتصادی وجود دارد.

نتایج بررسی بهزادی (۱۳۹۱) نشان می‌دهد که مدل شبکهٔ عصبی مصنوعی در پیش‌بینی ریسک سیستماتیک، به طور معناداری نسبت به مدل‌های آماری رگرسیون از دقت پیش‌بینی بیش‌تری برخوردار است. مشایخی و مطمئن (۱۳۹۱) در مقاله خود نشان می‌دهد که میان ریسک سیستماتیک و حسابداری محافظه‌کاری مشروط رابطهٔ منفی وجود دارد. مشکی و عاصر ربانی (۱۳۹۰) نشان می‌دهد که رابطهٔ مثبت و خطی بین خطای پیش‌بینی سود و بازده غیر عادی وجود دارد همچنین بین خطای پیش‌بینی سود و ریسک سیستماتیک رابطه‌ای مشاهده نشده است. نتایج بررسی یحیی‌زاده و حجابی (۱۳۸۹) نشان می‌دهد که رابطهٔ بین نسبت قیمت به سود و بازده سهام در سه سطح شرکت، صنعت و بازار مثبت معنادار می‌باشد. همچنین نسبت قیمت به سود و ریسک سیستماتیک در سطح شرکت و صنایع مختلف رابطهٔ مثبت معنادار وجود دارد. حیدرپور و نایب (۱۳۸۸) نشان دادند که میان ارزش افزوده اقتصادی و ریسک سیستماتیک شرکت‌ها بورس اوراق بهادارهمبستگی بالا و قابل توجهی وجود ندارد. نتایج بررسی دستگیر و بزاززاده (۱۳۸۵) نشان می‌دهد که میزان افشای اطلاعات در صورت‌های مالی سالانه از سوی شرکت‌ها، تأثیری بر ریسک سیستماتیک آن‌ها ندارد. بررسی ایسلی و اوهاری (۲۰۰۴) نشان می‌دهد که کیفیت گزارشگری مالی می‌تواند ریسک اطلاعاتی و ریسک غیرسیستماتیک و هزینهٔ سرمایه را تحت تأثیر قرار دهد، بنابراین، داشتن یا نداشتن اطلاعات بر ریسک شرکت‌ها تأثیر گذار است. عبدالقانی (۲۰۰۵) نشان می‌دهد که چهار متغیر اندازه دارایی، نسبت جاری، رشد سود و نسبت سود سهام پرداختی دارای ارتباط با اهمیتی با ریسک سیستماتیک هستند و متغیرهای نسبت اهرمی، تغییرپذیری سود و بتای سود رابطهٔ ضعیفی با ریسک سیستماتیک دارند. چان و همکاران (۲۰۰۶) بیان می‌کنند که علی رغم رابطهٔ عکس بین ریسک غیرسیستماتیک و کیفیت گزارشگری مالی، رابطهٔ بین کیفیت گزارشگری مالی و ریسک غیرسیستماتیک قوی است. آشباق و همکاران (۲۰۰۹) نشان می‌دهند که شرکت‌هایی که نقص عمدهٔ کنترل‌های داخلی خود را افشا می‌کنند، ریسک سیستماتیک، غیرسیستماتیک و هزینهٔ سرمایهٔ بالایی دارند. بررسی فیلیپون و ریچارد (۲۰۱۰) نشان می‌دهد که بین اهرم مالی شرکت‌ها و ریسک سیستماتیک رابطه معناداری دارد. تحقیق راج‌گوپال و موهان (۲۰۱۱) در رابطه با بررسی کیفیت گزارشگری مالی و ریسک غیر سیستماتیک نشان می‌دهد که پایین آمدن کیفیت سود با افزایش ریسک غیرسیستماتیک رابطهٔ معنادار دارد. آچاریا و استیفن (۲۰۱۲) در تحقیق خود با عنوان" تجزیه و تحلیل ریسک سیستماتیک از بخش بانکداری اروپا" به این نتیجه رسیدند که منابع ایجاد کنندهٔ بدهی در بانک‌ها یکی از عوامل اصلی افزایش ریسک سیستماتیک در بانک‌ها اروپایی می‌باشد. ویس و همکاران (۲۰۱۳) در مقاله خود با عنوان "ادغام بانک‌ها و ریسک سیستماتیک" به این نتیجه

رسیده است که با ادغام بانک‌ها و افزایش سهم آن‌ها در بازار، ریسک سیستماتیک بانک کاهش می‌یابد. باقس و کیمان (۲۰۱۴) نشان دادند با توجه به پیشرفت‌های اخیر در اقتصاد مالی در رابطه با اندازه‌گیری ریسک سیستماتیک، استفاده از برنامه‌های کاربردی داده کاوی برای تجزیه و تحلیل و قیمت‌گذاری از ریسک سیستماتیک در حال حاضر مزایای قابل توجهی حداقل در سطح مفهومی و عملی دارا می‌باشد. صالحی و همکاران (۲۰۱۵) در تحقیق خود نشان دادند که روش رگرسیون مبتنی بر هسته با دقت پیش‌بینی بیش‌تر از ۹۰ درصد نسبت به الگوریتم لارس وآدابوست توانایی بالاتری جهت پیش‌بینی ریسک سیستماتیک دارد. سنسوی (۲۰۱۶) به بررسی ریسک سیستماتیک در بازارهای سهام متعارف و بازارهای سهام اسلامی پرداخته است. نتایج این تحقیق نشان داده است که ریسک سیستماتیک در بازارهای متعارف در اغلب اوقات بالاتر از ریسک سیستماتیک بازارهای اسلامی است. با این حال، نتایج تفاوت معنی‌داری بین ریسک سیستماتیک در بازارهای متعارف و بازارهای اسلامی در زمان بحران مالی سال ۲۰۰۸ نشان نمی‌دهد. در واقع بر اساس این نتایج بازارهای اسلامی در زمان بحران، قادر به پایین‌تر نگهداشتن ریسک سیستماتیک در مقایسه با بازارهای متعارف نبوده‌اند.

فرضیهٔ تحقیق

فرضیهٔ ۱: پیش‌بینی ریسک سیستماتیک بر اساس الگوریتم کلونی مورچگان امکان‌پذیر است.
فرضیهٔ ۲: پیش‌بینی ریسک سیستماتیک بر اساس الگوریتم لارس امکان‌پذیر است.
فرضیهٔ ۳: الگوریتم کلونی مورچگان در مقایسه با الگوریتم لارس توانایی بیش‌تری در پیش‌بینی ریسک سیستماتیک دارد.

روش پژوهش

در تحقیق حاضر برای سنجش میزان ریسک شرکت از عامل بتا استفاده شده است. در این تحقیق مقادیر بتا از گزارش‌های بورس اوراق بهادار قابل استخراج است. عامل بتا، ماهیتا مجموعه ای از ریسک‌های گوناگون است. از جمله این ریسک‌ها می توان به ریسک بازار، ریسک سیاسی، ریسک نوسانات نرخ ارز، ریسک نوسانات نرخ بهره و.. . اشاره نمود (چارلز، ۱۳۸۲). بتا شدت تغییرات بازدهی سهم مورد نظر نسبت به بازاراست.

$$Beta = \frac{cov\ (\text{بازدهی بازار ، بازدهی سهام})}{var\ (\text{بازدهی بازار})}$$

تعریف عملیاتی و نحوهٔ برآورد و اندازه‌گیری هر یک از متغیرهای مستقل نیز شرح جدول ۱ است:

جدول ۱: متغیرهای تحقیق

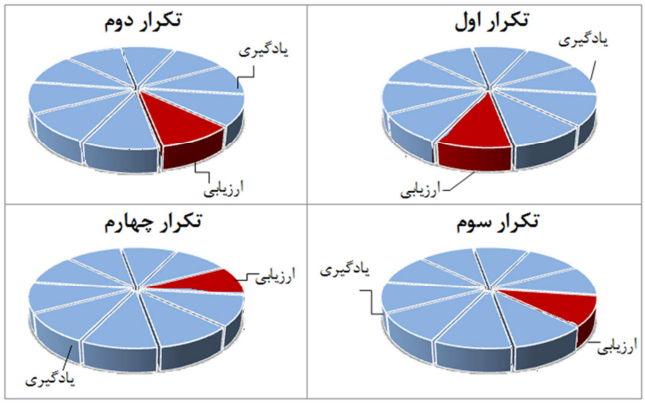
نام متغیر	نحوهٔ محاسبه	منبع
کیفیت افشا	امتیاز افشا گزارش شده توسط بورس اوراق بهادار	دستگیر و بزاززاده (۱۳۸۵)
بازده سهام	تغییر در ارزش سهام انتهای دوره نسبت به ابتدای دوره پس از تعدیل از بابت افزایش سرمایه تقسیم بر ارزش سهام در ابتدای دوره به عنوان بازده در نظر گرفته شده است	اسکندری و حسینی (۱۳۸۱) و جهانخانی و یزدانی (۱۳۷۴)
نسبت اهرمی	نسبت ارزش دفتری بدهی بلندمدت به مجموع ارزش دفتری بدهی بلندمدت و ارزش بازار حقوق صاحبان سهام تعریف می‌شود. ارزش بازار حقوق صاحبان سهام برابر با حاصل ضرب تعداد سهام منتشر شده شرکت در قیمت پایان روز معامله سهام است.	سعیدی و رامشه (۱۳۹۰)
سود عملیاتی	نسبت سود عملیاتی به فروش خالص	بال و براون (۱۹۶۸)
تغییرپذیری سود عملیاتی	انحراف معیار درصد تغییرات سود عملیاتی شرکت است.	سعیدی و رامشه (۱۳۹۰)
نرخ بهرهٔ بدون ریسک	نرخ بهره سپرده‌های بانک‌های دولتی در پایان سال	گالای و مازولیس (۱۹۷۶)
تغییرات سود	تفاوت سود در انتهای دوره نسبت به ابتدای دوره	گالای و مازولیس (۱۹۷۶)
اندازهٔ شرکت	لگاریتم طبیعی ارزش دارایی‌ها	بریمبل (۲۰۰۳) و احمدپور و نمازی (۱۳۷۷)
نرخ رشد سود عملیاتی	متوسط درصد تغییرات سود عملیاتی شرکت را نشان می‌دهد.	سعیدی و رامشه (۱۳۹۰)
نسبت پرداخت سود	سود تقسیمی هر سهم تقسیم بر سود هر سهم	بریمبل (۲۰۰۳)
ارزش جاری شرکت	ارزش بازار هر برگ سهم ضربدر تعداد سهام در انتهای دوره	گالای و مازولیس (۱۹۷۶)
نسبت جاری	نسبت دارایی جاری به بدهی جاری	بریمبل (۲۰۰۳)
اهرم مالی	نسبت بدهی به دارایی	بریمبل (۲۰۰۳) و احمدپور و نمازی (۱۳۷۷)
واریانس بازده شرکت	پراکندگی بازده سهام	گالای و مازولیس (۱۹۷۶)
نسبت پوشش بهره	این نسبت از طریق تقسیم سود قبل از بهره و مالیات به هزینه بهره شرکت به دست می‌آید.	بریمبل (۲۰۰۳)
نسبت قیمت به سود	ارزش بازار هر سهم به سود هر سهم	یحیی‌زاده‌فر و حجابی (۱۳۸۹)
سود خالص	نسبت سود خالص به فروش خالص	بال و براون (۱۹۶۹)
رشد دارایی	تغییر در دارایی‌های انتهای دوره نسبت به ابتدای دوره تقسیم بر دارایی‌های ابتدای دوره	بیور و همکاران (۱۹۷۰)
بازده دارایی	نسبت سود خالص به دارایی‌ها	نمازی و شیرزاده (۱۳۸۴)
بازده حقوق صاحبان سهام	نسبت سود خالص به حقوق صاحبان سهام	بیور و همکاران (۱۹۷۰)

جامعهٔ آماری این تحقیق شامل تمامی شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران هستند که در دورهٔ مورد بررسی تغییر دورهٔ مالی نداشته باشند، شرکت‌های سرمایه‌گذاری، واسطه‌گری‌های مالی، بانک ولیزینگ نباشند و داده‌های مورد نظر آن‌ها در دسترس باشد. بر این اساس ۱۵۴ شرکت طی سال‌های ۱۳۸۹ تا ۱۳۹۳ (۷۷۰ سال- شرکت) به عنوان نمونه آماری انتخاب شده است.در این تحقیق جهت پیش‌بینی ریسک سیستماتیک از روش‌های الگوریتم کلونی مورچه‌ها و الگوریتم لارس استفاده شده است. پنج فرآیند در روش‌های پیشنهادی وجود دارد که به ترتیب عبارت است از انتخاب داده‌ها، پاکسازی داده‌ها، تقسیم داده‌ها به مجموعه‌های آموزشی و ارزیابی، فرآیند آموزش مدل و ارزیابی مدل آموزش داده شده با داده‌های ارزیابی که تاکنون توسط الگوریتم‌ها مشاهده نشده است. در مرحلهٔ اول داده‌های تحقیق انتخاب گردیده و در جدول ۱ ارائه شده اند. مرحلهٔ دوم پاکسازی داده‌ها است. در این تحقیق جهت اجرای بخش دوم از رگرسیون گام به گام در انتخاب متغیرهای ورودی استفاده می‌گردد. انتخاب ورودی‌ها یکی از مهم‌ترین اهداف در پیش پردازش داده‌ها است. این مساله شامل فرآیند تعیین ورودی‌های مرتبط و حذف صفاتی که زائد بوده و اطلاعات اندکی فراهم می‌کنند، می‌باشد. انجام فرآیند انتخاب ورودی‌ها قبل از به کارگیری یک الگوریتم یادگیری مزایای فراوانی دارد. با حذف تعداد زیادی از ورودی‌های نامربوط، روش‌های یادگیری آموزشی، هزینهٔ محاسباتی و زمان کم‌تری را متحمل می‌شوند. همچنین مدل به دست آمده ساده‌تر می‌شود که غالباً تفسیر آن راحت‌تر شده و در عمل مفیدتر و سودمندتر می‌باشد. همچنین مدل‌های ساده هنگامی که برای پیش‌بینی به کار می‌روند، دارای کلیت و عمومیت بهتری می‌باشند. بنابراین مدلی که دارای ورودی‌های کم‌تری است مزیت‌های بیش‌تری داشته و حتی دقت بالاتری را ارائه می‌دهد. یکی از روش‌هایی که برای این منظور به کار می‌رود روش رگرسیون گام به گام می‌باشد. در این روش مجموعه‌ای از ورودی‌های معنادار برای تعیین متغیر وابسته مدل انتخاب می‌شوند. جهت استفاده از رگرسیون ابتدا باید پیش‌فرض‌هایی را مد نظر قرار داد. در هر مدل رگرسیون خطی شرایطی داریم که با فرض درستی آن‌ها، نتایج رگرسیون معتبر هستند و در غیر این صورت باید مدل دیگری را جایگزین ساخت. در این خصوص، نمودار باقیمانده‌ها در مقابل پیش‌بینی‌ها که بیان‌گر وجود همگنی واریانس و نمودار احتمال نرمال بودن باقیمانده‌های رگرسیون و باقیمانده‌های استاندارد شده بیانگر نرمال بودن داده‌ها است. عدم هم‌خطی و استقلال خطاها دو بخش دیگر مورد بررسی هستند. در صورت وجود هم‌خطی، با وجود آنکه مدل خوب به نظر می‌رسد ولی دارای متغیرهای مستقل معنی‌داری نمی‌باشد و این متغیرها بر یکدیگر تأثیر می‌گذارند. لذا پیش فرض زیربنایی بعدی آزمون هم‌خطی برای رگر سیون چندگانه می‌باشد. یکی

دیگر از پیش فرض‌هایی که در رگرسیون مدنظر قرار می‌گیرد، استقلال خطاها از یکدیگر است. برای رسیدن به این مهم از آزمون دوربین- اتسون می‌توان استفاده می‌شود. مرحلهٔ سوم تقسیم داده‌ها می‌باشد. یکی از معیارهایی که برای ارزیابی یک تخمینگر مورد استفاده قرار می‌گیرد نرخ خطا[۴] است که دارای انواع مختلفی است، به طور کلی نمی‌توان با مقایسهٔ خطای محاسبه شده روی داده‌های یادگیری، قضاوت مناسبی در خصوص توانایی‌های الگوریتم‌ها انجام داد. معمولاً نرخ خطا روی داده‌های یادگیری کم‌تر از نرخ خطا روی داده‌هایی است که در فرآیند یادگیری دیده نشده‌اند. با این استدلال، نمی‌توان از خطای یادگیری برای مقایسهٔ دو الگوریتم استفاده نمود. دلیل این است که برای مدل‌های پیچیده‌تر، تخمین‌گرهایی که معمولاً دارای پارامترهای بیش‌تری هستند، دارای مرز پیچیده‌تری هستند. این مرز پیچیده باعث کاهش خطا بر روی داده‌های یادگیری در مقایسه با مدل‌های ساده‌تر می‌شود. بنابراین علاوه بر مجموعه داده‌های یادگیری، مجموعه‌ای از داده‌ها برای ارزیابی[۱] مورد نیاز است. از داده‌های آموزش برای یادگیری مدل و از داده‌های ارزیابی به منظور محاسبهٔ نرخ خطای الگوریتم روی داده‌هایی که تاکنون مشاهده نکرده است، استفاده می‌شود. البته برای اینکه ارزیابی مناسب باشد تعداد یک اجرا الگوریتم کفایت نمی‌کند. معمولاً الگوریتم‌ها تمایل دارند که نرخ خطای تخمینی خود را به نرخ خطای واقعی نزدیک کنند و این امر با اجرای بارها و بارها فرآیند یادگیری و ارزیابی امکان‌پذیر است. بنابراین زمانی که یک مجموعه داده در اختیار گذاشته می‌شود، بایستی بخشی از آن را برای ارزیابی نهایی کنار گذاشت و از بقیه برای یادگیری استفاده کرد و مجدداً دو مجموعه‌ها را تغییر داده و دوباره مدل را ارزیابی کرد. یکی از روش‌های معمول برای این منظور روش اعتبارسنجی ده‌گانه[۶] نام دارد. در این روش مجموعه داده‌ها به K قسمت مساوی، به صورت تصادفی تقسیم میگردد. R_X زوج مجموعهٔ $X_1,X_2,...X_d$ به صورت تصادفی استخراج می‌شود که در آن X_i متغیرهای مستقل و Y_i متغیر وابسته نمونه i ام است. در اجرای اول قسمت اول از K قسمت به منظور ارزیابی، K-1 قسمت باقیمانده برای یادگیری استفاده می‌شود. در اجرای دوم قسمت دوم از K قسمت به منظور ارزیابی، K-1 قسمت باقیمانده برای یادگیری استفاده می‌شود.Y مرتبه الگوریتم به همین روال اجرا می‌گردد. مجموعه داده‌های یادگیری و ارزیابی باید به اندازهٔ کافی بزرگ باشند تا خطای تخمینی، به مقدار واقعی نزدیک‌تر باشد. در عین حال داده‌های یادگیری و ارزیابی با داده‌های یادگیری و ارزیابی سایر تکرارها، باید کم‌ترین همپوشانی را داشته باشند تا به این وسیله تمام داده‌ها در فرآیند یادگیری و ارزیابی دخالت داده شوند. در این روش دو نکته دیده می‌شود. نکتهٔ اول این که نسبت مجموعه ارزیابی به یادگیری کوچک است.

1. Evaluation

همچنین هر چقدر مقدار N (تعداد کل نمونه‌های مجموعه داده‌ها) افزایش یابد می‌توان مقدار پارامتر K را کاهش داد و اگر مقدار N کوچک باشد، باید مقدار K را آنقدر بزرگ در نظر گرفت که تعداد نمونه‌های لازم برای عمل یادگیری فراهم باشد. چنانچه مقدار K برابر N در نظرگرفته این روش به روش خارجی تبدیل می‌شود. در شکل زیر چهار تکرار اول انتخاب مجموعهٔ داده‌های یادگیری و ارزیابی روش روش اعتبارسنجی ده‌گانه شان داده شده است.



شکل ۱: روش اعتبارسنجی ده‌گانه

در هر بار تکرار یک نرخ خطا برای داده‌های یادگیری و ارزیابی محاسبه می‌گردد و در نهایت میانگین نرخ‌های خطای به دست آمده به عنوان نرخ خطا داده‌های یادگیری و داده‌های ارزیابی انتساب داده می‌شود. برای ارزیابی مدل‌های پیش‌بینی از معیار ارزیابی با نام میانگین قدرمطلق خطا[۷] (MSE) استفاده شده است که با استفاده از مدل (۱) محاسبه می‌گردند.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - d_i)^2$$ مدل ۱

که در آن d_i ریسک سیستماتیک شرکت iام، y_i ریسک سیستماتیک پیش‌بینی شده توسط الگوریتم هوشمند، n تعداد داده‌های مجموعه مورد بررسی (آموزش، اعتبارسنجی، ارزیابی) است. $y_i - d_i$ میزان خطای پیش‌بینی برای شرکت iام نشان می‌دهد. هر چه MSE نزدیک‌تر به صفر باشند پیش‌بینی الگوریتم‌ها به واقعیت نزدیک‌تر است.

مرحلهٔ چهارم، فرآیند آموزش و ارزیابی مدل‌ها در الگوریتم‌های تحقیق است.

پس از تقسیم نمونه‌ها به دو دسته داده‌های یادگیری و ارزیابی، با استفاده از داده‌های آموزشی مدل آموزش ایجاد می‌شود. برای حل مسئله ابتدا به معرفی مدل آن پرداخته می‌شود. رابطهٔ زیر تابعی است که الگوریتم تحقیق سعی در یافتن ضرایب $b_i, i=1,\ldots,m$ خواهد داشت.

مدل (۲) $$z = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m$$

که در آن b_0 عرض از مبدا و $b_i, i=1,\ldots,m$ وزن‌های هر کدام از متغیرها (x_i) (ضرایب تخمینگر) است. x_i مقدار متغیر مستقل iام است. b_iها توسط الگوریتم با استفاده از داده‌های آموزشی پیدا می‌گردند که اصطلاحا به آن آموزش مدل گفته می‌شود و سپس با داده‌های ارزیابی، مدل را ارزیابی می‌گردد. یعنی پس از محاسبهٔ b_iها، داده‌های ارزیابی به رابطهٔ بالا وارد شده و مقدار MSE محاسبه می‌گردد.

الگوریتم کلونی مورچگان

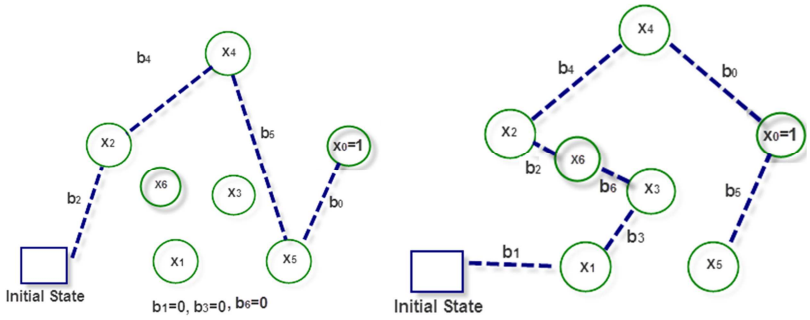
برای حل مسئله ابتدا به معرفی مدل آن پرداخته می‌شود. مدل (۳) تابعی است که الگوریتم بهینه‌سازی کلونی مورچگان سعی در یافتن ضرایب $b_{ij} = 1,...m$ خواهد داشت وقتی که $m=7$ باشد.

مدل (۳) $$z = b_0 + b_1x_1 + b_2x_2 + \Lambda + b_mx_m$$

که در آن b_0 عرض از مبدا و $b_i,\ i = 1,\cdots,m$ وزن‌های هر کدام از ۶ متغیر (x_i) (ضرایب تخمینگر) است. x_i مقدار متغیر مستقل iام است. b_iها توسط الگوریتم کلونی مورچگان با استفاده از داده‌های آموزشی پیدا می‌گردند که اصطلاحا به آن آموزش مدل گفته می‌شود و سپس با داده‌های ارزیابی، مدل را ارزیابی می‌گردد. یعنی پس از محاسبهٔ b_iها، داده‌های ارزیابی به مدل (۲) وارد شده و مقدار MSE محاسبه می‌گردد.

در ادامه به روند یافتن پارامترهای مدل (۳) با استفاده از الگوریتم و داده‌های آموزشی پرداخته می‌شود. هنگامی که مورچه‌ها به سوی منابع غذایی یا برعکس از منابع غذایی به سوی لانه حرکت می‌کنند ماده‌ای بنام فرومن روی زمین ترشح می‌کنند. مورچه‌ها می توانند فرومن را بچشند و وقتی می‌خواهند راه خود را انتخاب کنند، احتمالاً راهی را انتخاب می‌کنند که دارای غلظت فرومن زیادتری است. هرچه غلظت فرومن بیش‌تر باشد علاقه مورچه به طی این مسیر بیش‌تر می‌شود. فرومن در اثر گذشت زمان تبخیر می‌شود و در نتیجه در مسیرهای که زیاد طی نشده‌اند، فرومن کم‌تری انباشته می‌شود. با گذشت زمان کوتاهترین مسیر با حرکت مورچه‌ها به دست می‌آید، این مسیر تقویت شده و مسیرهای دیگر تضعیف می‌شوند تا همهٔ مورچه‌ها از یک

مسیر کوتاه رفت و آمد کنند. مسئلهٔ یافتن پارامترها به صورت مناسب می تواند در یک مساله فرمول‌بندی شود. در گراف الگوریتم گره‌ها متغیرهای مستقل را نشان می‌دهند و یال‌های بین آن‌ها نشان دهنده انتخاب متغیرهای مستقل توسط مورچه و پارامتر مربوط (b_i) به متغیر مستقل هستند. گراف توسط یک مورچه برای پیدا کردن یک زیرمجموعه بهینه از متغیرهای مستقل پیمایش می‌شود، تا جایی که تعداد مینیممی از گره‌ها ملاقات شوند. پیمایش یک مورچه زمانی متوقف می‌شود که به یک معیار توقف مشخص برسد. شکل ۴ نشان دهندهٔ مسیر انتخاب شده دو مورچه است. متغیرهای مستقل توسط مورچه و پارامتر مربوط به متغیر مستقل (میزان فرمون) توسط مورچه را نشان می‌دهد. برای سازگار کردن مدل – با مدل (۳)– مورچه‌ها ابتدا از یک گره مجازی حرکت می‌کنند. شش گره که پنج تای آن‌ها بیانگر متغیرهای مستقل و یکی بیانگر عرض از مبدا ($x_0 = 1$) است که مقدار آن همواره یک است در نظر گرفته می‌شود. حال با این نوع بیان مسئله می‌توان آن را با الگوریتم حل نمود.



شکل ۲: دو نمونه از حرکت یک مورچه

تمایل اکتشافی حرکت مورچه‌ها و میزان فرومن یال‌ها با همدیگر قانون احتمال انتقال را تشکیل می دهند که در مدل (۴) نشان داده شده است:

$$\begin{cases} p_{i,j}^{k}(t) = \dfrac{[\tau_{i,j}(t)]^{\alpha}[\eta_{i,j}(t)]^{\beta}}{\sum_{l \in J_i^k}[\tau_{i,j}(t)]^{\alpha}[\eta_{i,j}(t)]^{\beta}}, & j \in J_i^k \\ 0 & Otherwise \end{cases} \qquad \text{مدل (۴)}$$

که در آن $\eta_{i,j}$ یک معیار اکتشافی، $\tau_{i,j}$ میزان فرومن بین گره iام و jام، k شمارهٔ مورچه است و t زمان (تکرار) است. α و β به ترتیب وزن فرومن و معیار اکتشافی هستند که در این پروژه ۱/۰ در نظر گرفته شده است. J_i^k همسایگان گره iام برای مورچه kام را نشان

می‌دهد که در اینجا نشان دهندهٔ متغیرهای مستقلی است که هنوز انتخاب نشده‌اند. ابتدا مورچه‌ها در حالت اولیه قرار می‌گیرند. در هر گام از پیمایش، مورچهٔ k ام قانون احتمال انتقال را به کار می‌برد. احتمال این که مورچه kام در متغیر مستقل iام (گره x_iام) باشد، متغیر مستقل j ام (گره x_jام) را در تکرار tام انتخاب کرده باشد از مدل (۵) محاسبه می‌گردد.

بروز شدن فرومن بر اساس رابطه زیر انجام می‌شود.

$$\tau_{i,j}(t) = \rho.\tau_{i,j} + \sum_{k=1}^{n} \Delta\tau_{i,j}^{k}(t) \qquad \text{مدل (۵)}$$

که در آن n تعداد مورچه‌ها، $0 \prec \rho \prec 1$ میزان تبخیر فرومن را نشان می‌دهد و پارامتر ρ برای جلوگیری از تراکم بیش از حد فرومن به کار برده می‌شود در این پروژه ۰/۲ در نظر گرفته شد و الگوریتم را قادر می‌سازد که تصمیمات اشتباهی که قبلاً گرفته شده است، فراموش کند. اگر یک یال به وسیلهٔ مورچه‌ها انتخاب نشود غلظت فرومن آن به طور تدریجی کاهش پیدا میکند. $\Delta\tau_{i,j}^{k}(t)$ مقدار فرومنی است که مورچه kام بر روی یال‌هایی که ملاقات کرده است، در زمان t اضافه می‌کند و از از مدل (۶) به دست می‌آید.

$$\Delta\tau_{i,j}^{k}(t) = \begin{cases} \dfrac{Q}{L_k(t)}, & if\ \ the\ edge(i,j)\ is\ chosen\ by\ ant_k \\ 0, & Otherwise \end{cases} \qquad \text{مدل (۶)}$$

که در آن Q یک ثابت برای به روزرسانی فرومن است که معمولاً یک در نظر گرفته می‌شود و $L_k(t)$ میزان خطای تخمینگر MSE است. در هنگام محاسبهٔ خطا چون همواره میزان فرمون‌ها مثبت است و ضرایب b_iها می‌توانند منفی نیز باشند از فرمون‌ها مقدار ثابت ۳۰ کم گردید که با سعی و خطا به دست آمد. در اینجا مجموعهٔ داده‌ها به دو دستهٔ داده‌های آموزشی و ارزیابی تقسیم شدند. داده‌های آموزشی برای یادگیری (به دست اوردن پارامترهای b_i رابطهٔ مدل ۳) با استفاده از الگوریتم استفاده می‌گردد. با استفاده از $\tau_{i,j}^{k}(t)$ مقدار فرومن ترشح شده بین گره‌ها برای مورچه kام در زمان t، $L_k(t)$ برای داده‌های آموزشی محاسبه می‌گردد. لازم به ذکر است که $\tau_{i,j}^{k}(t) = b_j$ برای مورچه kام در لحظهٔ t است (bها در مدل ۳) هستند.

الگوریتم لارس

الگوریتم لارس یک الگوریتم برای رگرسیون خطی است و در آن ابتدا باید با استفاده از فرآیند آموزش ضرایب مدل (۷) را محاسبه نماید.

$$\hat{y} = b_1x_1 + b_2x_2 + \cdots + b_mx_m\ , m = 1, \ldots, 8 \qquad \text{مدل (۷)}$$

که در آن m شماره متغیرهای مستقل، x_m مقدار متغیرها، b_m ضریب اهمیت هر متغیر که باید محاسبه گردد و $\hat{y}$ مقدار پیش‌بینی ریسک سیستماتیک است. به منظور پیدا کردن b_mها فرض کنید نمونه‌های یادگیری در ماتریس $f_{ij} = -f_{ji}$ که نرمال است قرار داشته باشد که x_t یک ماتریس $n \times m$ است که n تعداد نمونه‌ها و x_t تعداد متغیرهای مستقل داده‌ها باشد. Y بردار n×۱ نرخ بازه هر کدام از نمونه‌ها را نشان دهد. x_{ij} نشان دهندهٔ متغیر مستقل iام نمونه jام و x_i بردار متغیر مستقل iام و d تعداد متغیر مستقل را نشان می‌دهد. فرض کنید r_{xj} مقدار همبستگی بین ویژگی X_j و Y است که از مدل (۸) به دست می‌آید.

$$r_{x_j} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)\, s_x\, s_y}$$

مدل (۸)

که در آن $\bar{x}$ و $\bar{y}$ بیانگر میانگین و n تعداد نمونه‌ها است و انحراف از معیار s_z از مدل (۹) محاسبه می‌گردد.

$$s_z = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (z_i - \bar{z})^2}$$

مدل (۹)

ماتریس R_x مقدار همبستگی متغیرهای $X_1, X_2, \dots, X_d$ را در خود دارد

$$(R_x = \{r_{x1}, r_{x2}, \dots, r_{xd}\})$$

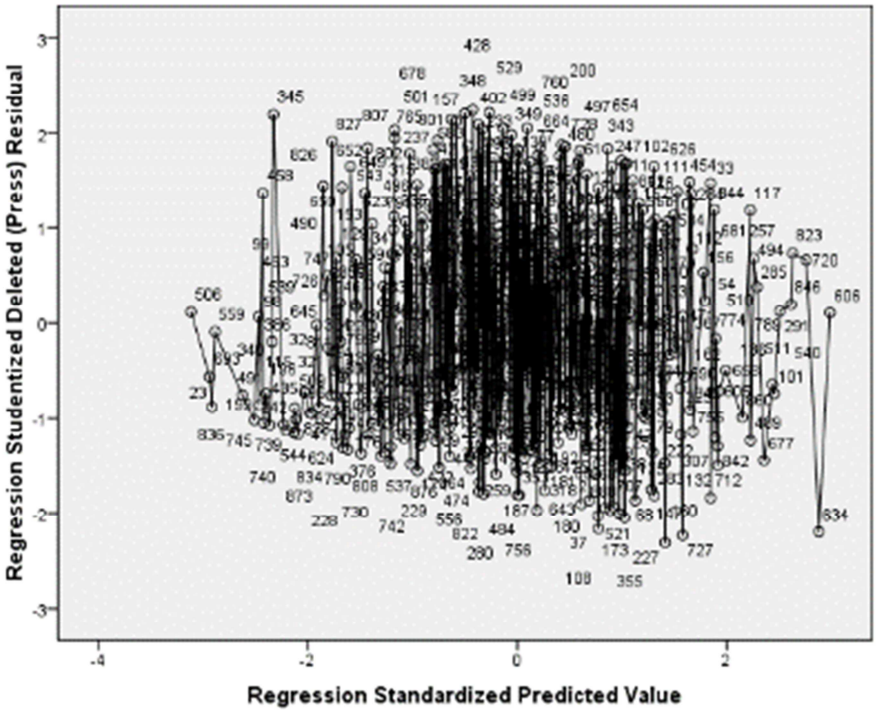
مدل (۱۰)

بنابراین f که دارای بیش‌ترین میزان قدر مطلق همبستگی با Y را دارد به عنوان اولین متغیر انتخاب می‌گردد. ویژگی m_1 که دارای بیش‌ترین میزان همبستگی است را به عنوان متغیر اول به مجموعه بیش‌ترین همبستگی اضافه می‌کند. سپس با توجه به مجموعه بیش‌ترین همبستگی به دنبال متغیر دوم می‌گردد که دارای بیش‌ترین همبستگی است. در مرحلهٔ آخر مدل با توجه به طی کردن مراحل قبل مورد ارزیابی قرار خواهد گرفت و نتایج ارائه می‌گردد. ارزیابی مدل، آخرین مرحله است.

تجزیه و تحلیل آماری

مرحلهٔ اول تجزیه و تحلیل آماری انتخاب داده‌ها است. در این مرحله داده‌های مورد نیاز که شامل ریسک سیستماتیک و متغیرهای ورودی لیست شده در جدول ۱ است جمع‌آوری می‌گردد. مرحلهٔ دوم پاکسازی داده‌ها است. در این تحقیق جهت اجرای بخش دوم از رگرسیون گام به گام در انتخاب متغیرهای ورودی استفاده شده است. مطابق بخش روش تحقیق، در این بخش

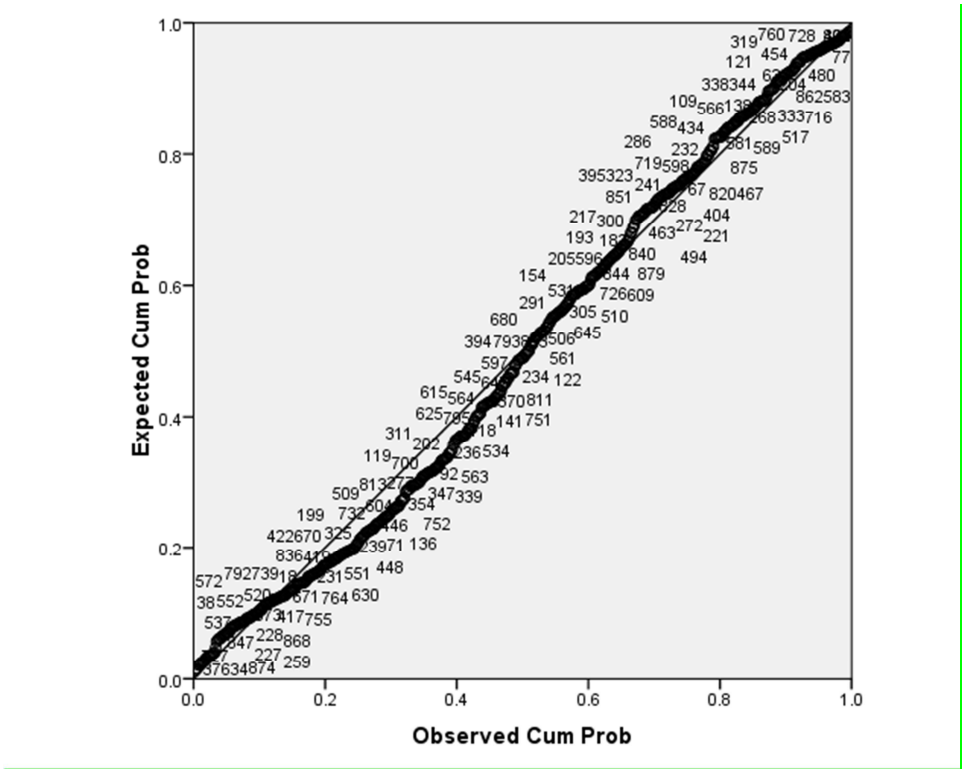
فرض‌های زیربنایی رگرسیون کنترل می‌شود. در این خصوص، نمودار باقیمانده‌ها در مقابل پیش بینی‌ها که بیان‌گر وجود همگنی واریانس است مورد بررسی قرار گرفته و همگنی واریانس تأیید می‌گردد. برای بررسی فرض نرمال بودن توزیع متغیر پاسخ، کافی است نرمال بودن توزیع باقیمانده‌ها مورد بررسی قرار گیرد چون اگر توزیع باقیمانده‌ها نرمال باشد توزیع متغیر پاسخ نیز نرمال خواهد بود. در نمودار احتمال نرمال بودن باقیمانده‌های رگرسیون و باقیمانده‌های استاندارد شده بشرح زیر مقادیر تابع توزیع تجربی نمونه بر روی نیمساز ربع اول قرار دارد که حاکی از نرمال بودن داده است.



شکل ۳: باقیمانده‌ها در مقابل پیش‌بینی‌ها

برای آزمون هم‌خطی، عامل تورم واریانس (vif) می‌بایست کم‌تر از ۱۰ باشد تا عدم وجود هم‌خطی تأیید گردد. بر اساس نتایج مکتسبه در جدول ۲، عامل تورم واریانس متغیرهای تحقیق کم‌تر از ده (vif<۱۰) بوده است. لذا با توجه به نوسان عامل تورم واریانس، همخطی چندگانه تهدیدی برای مدل نمی‌باشد (اسکندری و حسینی، ۱۳۸۱). برای بررسی استقلال خطاها از آزمون دوربین-واتسون می‌توان استفاده کرد. آمارهٔ این آزمون در جدول ۲ محاسبه شده و مقدار

آن برای فرضیه می‌بایست بین ۱ تا ۳ باشد. در واقع مقدار آمارهٔ دوربین– واتسن همیشه بین ۰ تا ۴ خواهد بود و در بهترین حالت باید مقدار آن نزدیک به ۲ باشد (دوربین و واتسن، ۱۹۵۰). در تحقیق حاضر آمارهٔ دوربین واتسن ۱/۸۹ بوده است. در نتیجه پیش فرض نهایی رگرسیون نیز مبنی بر استقلال خطاها تأیید می‌گردد. در نهایت از میان متغیرهای مستقل جدول ۱ با استفاده از آزمون گام به گام تعداد ۷ متغیر که مقدار احتمال آن‌ها کم‌تر از ۵ درصد بوده و لیست آن‌ها در جدول ۲ مشاهده می‌گردد انتخاب گردیدند.



شکل ٤: نرمال بودن باقیمانده‌های رگرسیون

طبق نتایج جدول فوق ۷ متغیر کیفیت افشا، بازده سهام، تغییرات سود، نرخ رشد سود عملیاتی، نسبت جاری، اهرم مالی و بازده دارایی جهت ورود به مدل داده کاوی در سطح خطای ۵ درصد تأیید گردیده است. سایر متغیرها نیز به دلیل وجود خطای بالاتر از ۵ درصد وارد مدل داده کاوی نمی‌گردند. نتایج حاصل از اعمال مابقی مراحل و همهٔ برازش‌ها به صورت جدول ۳ است:

جدول ۲: لیست متغیرهای مستقل انتخابی

نام متغیر	مقدار احتمال	آزمون VIF	نتیجه آزمون
کیفیت افشا	۰/۰۰۹	۱/۵۸۱	پذیرش (ورود به مدل)
بازده سهام	۰/۰۰۰	۱/۷۸۱	پذیرش (ورود به مدل)
نسبت اهرمی	۰/۱۲۲	۱/۸۷۱	رد متغیر
سود عملیاتی	۰/۶۶۳	۱/۹۸۱	رد متغیر
تغییرپـــذیری ســـود عملیـــاتی	۰/۲۱۹	۱/۰۲۵	رد متغیر
نرخ بهره بدون ریسک	۰/۴۱۰	۱/۶۲۱	رد متغیر
تغییرات سود	۰/۰۴۱	۱/۸۹۳	پذیرش (ورود به مدل)
اندازه شرکت	۰/۱۹۵	۱/۷۸۱	رد متغیر
نرخ رشد ســود عملیـاتی	۰/۰۲۱	۱/۵۶۲	پذیرش (ورود به مدل)
نسبت پرداخت سود	۰/۳۰۲	۱/۹۲۱	رد متغیر
ارزش جاری شرکت	۰/۰۷۱	۱/۹۸۱	رد متغیر
نسبت جاری	۰/۰۰۳	۱/۶۵۲	پذیرش (ورود به مدل)
اهرم مالی	۰/۰۴۹	۱/۷۷۴	پذیرش (ورود به مدل)
واریانس بازده شرکت	۰/۰۶۲	۱/۹۲۱	رد متغیر
نسبت پوشش بهره	۰/۲۵۱	۱/۰۵۱	رد متغیر
نسبت قیمت به سود	۰/۷۶۲	۱/۰۹۵	رد متغیر
سود خالص	۰/۴۲۲	۱/۹۵۲	رد متغیر
رشد دارایی	۰/۲۰۳	۱/۹۸۸	رد متغیر
بازده دارایی	۰/۰۰۰	۱/۰۲۲	پذیرش (ورود به مدل)
نسبت بازده حقوق صاحبان سهام	۰/۱۲۲	۱/۰۹۲	رد متغیر
آمارهٔ دوربین واتسن: ۱/۸۹			

بر اساس نتایج کسب شده در جدول ۳ الگوریتم‌های کلونی مورچه‌ها و لارس توانایی بالایی (بیش از ۹۸٪) جهت پیش‌بینی ریسک سیستماتیک دارند. در واقع کلونی مورچه‌ها با دقت ۹۸/۷۵٪ (خطای ۱/۲۵٪) و الگوریتم لارس با دقت ۹۸/۴۴٪ (خطای ۱/۵۶٪) توانسته اند ریسک سیستماتیک را پیش‌بینی نمایند. از این رو فرضیه‌های تحقیق مبنی بر امکان‌پذیر بودن پیش‌بینی ریسک سیستماتیک به وسیلهٔ الگوریتم کلونی مورچه‌ها و الگوریتم لارس تأیید می‌گردند. ضمن آن که فرضیهٔ سوم تحقیق مبنی بر توانایی بالاتر الگوریتم مورچه‌ها نسبت به الگوریتم لارس در پیش‌بینی ریسک سیستماتیک هم تأیید می‌گردد.

جدول ۳: نتایج تحقیق

Fold	الگوریتم کلونی مورچه‌ها		الگوریتم لارس	
	خطای داده یادگیری	خطای داده ارزیابی	خطای داده یادگیری	خطای داده ارزیابی
۱	۱/۰۱۲	۱/۰۲۵	۲/۰۱۲	۱/۸۱۲
۲	۱/۲۵۰	۱/۱۲۳	۱/۰۲۵	۱/۱۱۲
۳	۱/۰۳۶	۱/۰۸۸	۱/۸۹۲	۱/۸۱۲
۴	۱/۹۲۸	۱/۸۹۲	۲/۰۰۲	۲/۰۶۱
۵	۱/۲۰۳	۱/۶۲۳	۱/۹۸۵	۱/۵۱۹
۶	۱/۰۲۳	۱/۰۲۵	۱/۵۱۲	۱/۶۶۲
۷	۰/۹۸۹	۰/۹۱۲	۲/۴۱۲	۱/۹۱۸
۸	۱/۰۲۱	۱/۶۱۲	۱/۴۹۵	۱/۰۱۲
۹	۱/۵۱۵	۱/۰۲۳	۱/۶۱۲	۱/۴۱۱
۱۰	۱/۶۲۳	۱/۲۰۳	۱/۸۱۱	۱/۳۱۶
میانگین	۱/۲۶۵	۱/۲۵۲	۱/۷۷۵	۱/۵۶۳

نتیجه‌گیری

هدف این مطالعه ارزیابی این موضوع بود که آیا می‌توان ریسک سیستماتیک را بر اساس روش‌های مبتنی بر یادگیری ماشین کشف کرد. در این تحقیق که از اطلاعات مالی ۱۵۴ شرکت بین سال‌های ۱۳۸۹ الی ۱۳۹۳ (۷۷۰ سال- شرکت) استفاده شده است نشان می‌دهد که همبستگی بین متغیرهای کیفیت افشا، بازده سهام، تغییرات سود، نرخ رشد سود عملیاتی، نسبت جاری، اهرم مالی و بازده دارایی با ریسک سیستماتیک وجود دارد. همچنین در این تحقیق با استفاده از الگوریتم کلونی مورچگان و الگوریتم لارس اقدام به پیش‌بینی ریسک سیستماتیک شده است که نتیجهٔ پژوهش نشان می‌دهد هر دو الگوریتم توانایی بالایی (بیش از ۹۸٪) جهت پیش‌بینی ریسک سیستماتیک دارند. با توجه به دقت بالاتر از ۹۸ درصد الگوریتم کلونی مورچه‌ها و الگوریتم لارس به سرمایه‌گذاران و سایر استفاده‌کنندگان از صورت‌های مالی پیشنهاد می‌گردد از متغیرهای کیفیت افشا، بازده سهام، تغییرات سود، نرخ رشد سود عملیاتی، نسبت جاری، اهرم مالی و بازده دارایی و الگوریتم‌های پژوهش جهت پیش‌بینی ریسک سیستماتیک در شرکت‌ها استفاده نمایند. در خصوص متغیرهای به کارگرفته شده نتایج تحقیق حاضر با نتایج تحقیق دستگیر و بزاززاده (۱۳۸۵)، اسکندری و حسینی (۱۳۸۱)، جهانخانی و یزدانی (۱۳۷۴)، گالای و مازولیس (۱۹۷۶)، سعیدی و رامشه (۱۳۹۰)، بریمل (۲۰۰۳)، احمدپور و نمازی (۱۳۷۷) و نمازی و شیرزاده (۱۳۸۴) هم‌خوانی دارد. همچنین برای تحقیق‌های آتی پیشنهاد می‌گردد:

با توجه به نتایج تحقیق حاضر، مدلی ترکیبی از الگوریتم‌های کلونی مورچه‌ها و الگوریتم لارس ایجاد تا در افزایش دقت پیش‌بینی مؤثر باشد.

پیشنهاد می‌شود با استفاده از سایر تکنیک‌های گروه هوش مصنوعی از قبیل کلونی زنبورها نیز مدلی برای پیش‌بینی ریسک سیستماتیک ارائه و نتایج آن با نتایج این تحقیق مقایسه شود.

منابع

احمدپورکاسکری، ا. و نمازی، م. (۱۳۷۷). "تأثیر اهرم عملیاتی مالی و اندازهٔ شرکت روی ریسک سیستماتیک". مجله *مدرس علوم انسانی*، ۲ (۶)، ۷۴-۱۰۱.

اسکندری، آ. و حسینی، ی. (۱۳۸۱). "بررسی وضعیت ریسک سیستماتیک شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران". *مجله برنامه و بودجه*، ۷ (۵)، ۲۹-۵۴.

بهزادی، ع. (۱۳۹۱). "پیش‌بینی ریسک سیستماتیک شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران با استفاده از شبکه‌های عصبی مصنوعی". پایان‌نامهٔ کارشناسی‌ارشد حسابداری، بهشهر، دانشگاه پیام نور.

پیری، پ.، حیدری، ح. و رئوف، س. (۱۳۹۲). "رابطهٔ بین ریسک سیستماتیک و ارزش افزوده اقتصادی در ایران". *فصلنامه پژوهش‌ها و سیاست‌های اقتصادی*، ۲۱ (۶۶)، ۱۶۹-۱۸۶.

جهانخانی، ع. و یزدانی، ن. (۱۳۷۴). "بررسی تأثیر نوع صنعت، اندازه، ریسک تجاری و درجه اهرم عملیاتی شرکت‌ها بر میزان به کارگیری اهرم مالی در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران"، *فصلنامه مطالعات مدیریت*، ۱۷ (۱۸)، ۱۶۹-۱۸۶.

جونز، چ. (۱۳۸۲). "مدیریت سرمایه‌گذاری". ترجمهٔ رضا تهرانی و عسگر نوربخش، تهران، انتشارات نگاه دانش.

حیدرپور، ف. و نایب، م. (۱۳۸۸). "بررسی سودمندی ارزش افزوده اقتصادی در ارزیابی ریسک بازار شرکت‌ها در ایران". *مجله حسابداری مدیریت*، ۲ (۳)، ۳۹-۵۵.

حیدرپور، ف. و خواجه‌محمود، ز. (۱۳۹۳). "رابطهٔ بین ویژگی‌های پیش‌بینی سود هر سهم توسط مدیریت بر ریسک و ارزش شرکت با هدف آینده‌نگری در تصمیم‌گیری". *فصلنامه علمی پژوهشی دانش مالی تحلیل اوراق بهادار*، ۷ (۲۲)، ۲۵-۴۶.

دستگیر، م. و بزاززاده، ح. (۱۳۸۵). "تأثیر افشا بر ریسک سیستماتیک". *پژوهشنامه اقتصادی*، ۱ (۳)، ۲۳۵-۲۵۰.

زلقی، ح.، بیات، م. و دانش‌عسگری، ت. (۱۳۹۳). "بررسی تأثیر پیش‌بینی سود از جانب مدیریت بر ریسک غیر سیستماتیک". *راهبرد مدیریت مالی*، ۲ (۲)، ۱۲۱-۱۳۶.

سعیدی، ع. و رامشه، م. (۱۳۹۰). "عوامل تعیین کنندهٔ ریسک سیستماتیک سهام در بورس اوراق بهادار تهران". *مجله پژوهش‌های حسابداری مالی*، ۷ (۱)، ۲۲-۴۳.

صالح‌نژاد،ح. و وقفی، ح. (۱۳۹۵).”تأثیر پیش‌بینی سود توسط مدیریت بر ریسک و ارزش شرکت“. *فصلنامه راهبرد مدیریت مالی*، ۴ (۱) ، ۱۰۳-۱۲۴.

قائمی، م. و رحیم‌پور، م. (۱۳۸۹). ”اعلان سودهای فصلی و نقدشوندگی سهام“. *مجله پژوهش‌های حسابداری مالی*، ۲ (۴)، ۱۴۵-۱۸۵.

مشایخ، ب. و مطمئن، م. (۱۳۹۱). ”ریسک سیستماتیک و محافظه‌کاری مشروط“. *تحقیقات مالی*، ۱ (۱۵)، ۱۰۹-۱۲۸.

مشکی، م. و موعاصی‌ربانی، ع. (۱۳۹۰). ”بررسی رابطهٔ بین خطای پیش‌بینی سود مدیریت و بازده غیر عادی و ریسک سیستماتیک شرکت‌های پذیرفته شده در بورس اوراق بهادار“. *بررسی‌های حسابداری و حسابرسی*، ۱۸ (۶۶)، ۵۳-۶۸.

ملایی، م. (۱۳۹۰). ”محتوای اطلاعاتی معیارهای ریسک حسابداری در پیش‌بینی ریسک سیستماتیک“. *فصلنامه علمی پژوهشی حسابداری مالی*، ۳ (۹)، ۱۵۵-۱۷۷.

نمازی، م. و شیرزاده، ج. (۱۳۸۴). ”بررسی رابطهٔ ساختار سرمایه با سودآوری شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران (تأکید بر نوع صنعت)“. *بررسی‌های حسابداری و حسابرسی*، ۱۲ (۴۲)، ۲۱-۴۱.

یحیی‌زاده، م. و حجابی، ر. (۱۳۸۹). ”رابطهٔ بین نسبت قیمت به سود با ریسک سیستماتیک و بازده سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران“. *راهبردهای بازرگانی*، ۱ (۴۱)، ۲۵-۳۸.

Abdelghany, E. (2005). “Disclosure of Market Risk or Accounting Measures of Risk: an Empirical Study”. *The Managerial Auditing*, 25 (3): 867-875.

Acharya, V. and Steffen, S. (2012). Analyzing Systemic Risk of the European Banking Sector. Handbook on Systemic Risk, J.P. Fouque and J. Langsam, eds., Cambridge University Press, Forthcoming. Available at SSRN: http://ssrn. com/abstract=2150616.

Alpaydin, E. (2010). Introduction to machine learning, 2'nd ed.: Cambridge, Mass. MIT Press.

Ashbaugh-Skaife, H., Collins, D.W., Kinney, R. and Lafond, R. (2009). “The Effect of SOX Internal Control Deficiencies on Firm Risk and Cost of Equity”. *Journal of Accounting Research*, 47 (1): 1-43.

Ball, R. and Brown, P. (1968). “An Empirical Evaluation of Accounting Income Numbers”. *Journal of Accounting Research*, 6 (2): 159-178.

Beaver, W., Kettler, H.P. and Scholes, M. (1970). “The Association between Market Determined and Acccounting Determined Risk Measures, *Accounting Review*, 45 (4): 654-682.

Bougheas, S. and Kirman, P. (April 14, 2014). Complex Financial Networks and Systemic Risk: A Review, CESifo Working Paper Series No. 4756. Available at SSRN: http://ssrn. com/abstract=2436826

Brimble, M. (2003). The Relevance of Accounting Information for Valuation and Risk. www4. gu. edu. au:8080/adt–root /uploads.

Chan, K., Chan, L., Jegadeesh, N. and Lakonishok, J. (2006). "Earnings Quality and Stock Returns". *Journal of Business*, Forthcoming, 79 (3): 1041-1057.

Easley, D. and O'Hara, M. (2004). "Information and the Cost of Capital". *Journal of Finance*, 59 (4): 1553-1583.

Ferreira, M. and Laux, P. (2007). "Corporate Governance, Idiosyncratic Risk and Information Flow". *The Journal of Finance*, 62 (5): 951-989.

Francis, J., Nanda, D.J. and Olsson, P. (2008). "Voluntary Disclosure, Earnings Quality, and Cost of Capital". *Journal of Accounting Research*, 46 (1): 53-100.

Galai, D. and Masulis, R. (1976). "The Option Pricing Model and the Risk Factor of Stock". *The Journal of Financial Economics*, 3 (4): 53-81.

George, J., Jiang, D. and Xu Tong, Y. (2009). "The Information Content of Idiosyncratic Volatility". *Journal of Financial and Quantitative Analysis*, 44 (11): 1-28.

Hirst, E., Koonce, L. and Venkataraman, S. (2008). "Management Earnings Forecasts: a Review and Framework". *Accounting Horizon*, 22 (4): 1055-1087.

Philippon, T. and Richard, M. (2010). "Measuring Systemic Risk". *International Journal of Business and Management*, 4 (3): 22-32.

Rajgopal, S. and Mohan, V. (2011). "Financial Reporting Quality and Idiosyncratic Return Volatility". *Journal of Accounting and Economics*, 51 (1): 1-20.

Rogers, J.L., Skinner, D.J. and Buskirk, A.V. (2009). "Earnings Guidance and Market Uncertainty". *Journal of Accounting and Economics*, Forthcoming. 10 (11): 35-43.

Salehi, M., Moradi, M. and Molaei, S. (2015). "Forecasting systematic risk by Least Angel Regression, AdaBoost and Kernel Ridge Regression". *Modern Applied Science*; 9 (11): 135-143.

Sensoy, A. (2016). "Systematic Risk in Conventional and Islamic Equity Markets". *International Review of Finance*, 16 (3): 457-466.

Weiss, G., Neumann, S. and Bostandzic, D. (2013). "Systemic Risk and Bank Consolidation: International Evidence". *International Review of Finance*, 17 (4): 57-66.