



بکارگیری تکنیک "تحلیل مولفه‌های اصلی" در داده‌کاهی متغیرهای مؤثر بر بازده سهام

اعظم مهتدی^۱

رضوان حجازی^۲

سیدعلی حسینی^۳

منصور مؤمنی^۴

تاریخ پذیرش: ۱۳۹۶/۱۲/۰۷

تاریخ دریافت: ۱۳۹۶/۱۰/۰۴

چکیده

این مقاله با هدف مواجهه با پدیده همخطی به موضوع داده‌کاهی متغیرهای مؤثر بر بازده سهام می‌پردازد. تمرکز این مقاله بر روی روش "تحلیل مولفه‌های اصلی" است. ابتدا با مطالعه عمیق ادبیات حرفه در داخل و خارج از کشور، ۴۷ متغیر به عنوان متغیرهای مؤثر بر پیش‌بینی بازده سهام، شناسایی شدند. متغیرهای مذکور برای ۶۸ شرکت واجد شرایط، استخراج و سپس روش "تحلیل مولفه‌های اصلی" بر روی آن‌ها اجرا شد. نتایج تحقیق نشان داد می‌توان با استفاده از این تکنیک از حجم داده‌های مربوط به پیش‌بینی سهام کم کرد و همخطی بین متغیرها مذکور را بدون آنکه نیاز به حذف برخی از آن‌ها باشد از بین برد. به طور خاص پس از اعمال این روش ۲۷ متغیر مؤثر بر بازده سهام، به ۵ مؤلفه مستقل تبدیل شدند تا ضمن کم شدن از حجم داده‌ها، همخطی بین متغیرها نیز اصلاح شود.

واژه‌های کلیدی: همخطی، تحلیل مولفه‌های اصلی، بازده سهام، متغیرهای مؤثر بر بازده.

۱- دانشجوی دکتری حسابداری دانشگاه الزهراء (س)، دانشکده علوم اجتماعی و اقتصادی، تهران و عضو هیأت علمی دانشگاه پیام نور بوشهر،

دانشکده اقتصاد، حسابداری و مدیریت، بوشهر، ایران، (مسئول مکاتبات) a_mohtadi@yahoo.com

۲- استاد گروه حسابداری، دانشگاه الزهراء (س)، دانشکده علوم اجتماعی و اقتصادی، تهران، ایران، (نویسنده اصلی)

۳- استادیار گروه حسابداری، دانشگاه الزهراء (س)، دانشکده علوم اجتماعی و اقتصادی، تهران، ایران.

۴- استاد گروه حسابداری، دانشکده مدیریت دانشگاه تهران، تهران، ایران.

۱- مقدمه

پیش‌بینی بازده سهام از گذشته تا به امروز یکی از موضوعات مورد توجه محققین بازار سرمایه بوده است. محققین در بافت‌ها و زمینه‌های اقتصادی، اجتماعی و فرهنگی مختلف تلاش کرده‌اند بازده سهام را با الگوها و تکنیک‌های مختلفی پیش‌بینی و آزمون کنند. (دستگیر و همکاران، ۱۳۹۱، ۴۵). بسط مدل‌های پیش‌بینی اولیه مانند مدل‌های او و پنمن^۱ (۱۹۸۹) و فاما و فرنچ^۲ (۱۹۹۳) توسط محققان بعدی و همچنین توجه و تاکید محققان بازده، بر مفاهیم تحلیل بنیادی^۳ و تئوری آربیتراژ^۴ باعث معرفی طیف وسیعی از متغیرها به عنوان متغیرهای مؤثر بر بازده سهام شده است. از آنجایی که بکارگیری تمامی متغیرهای معرفی شده، در مدل‌های تحقیقاتی امکان پذیر نیست (به علت کم شدن بیش از حد درجه آزادی) محققان مجبور به انتخاب از بین متغیرهای مذکور و کم کردن از حجم داده‌ها هستند. داده‌کاهی^۵ متغیرهای مؤثر بر بازده سهام به روش‌های کم کردن از تعداد متغیرهای مورد استفاده در تحقیقات بازده سهام، بدون مخدوش شدن اعتبار آزمون‌ها اشاره دارد. در این مقاله به روش "تحلیل مولفه‌های اصلی" به عنوان یکی از اصلی‌ترین روش‌های داده‌کاهی پرداخته شده است و ضمن معرفی مفهوم این روش، مناسب بودن کاربرد آن در بازار سرمایه ایران آزمون و نتایج بکارگیری آن گزارش شده است.

۲- چارچوب نظری

جدا از پیچیدگی‌های اجرایی، حداقل به دو دلیل نمی‌توان از همه متغیرهای معرفی شده به عنوان متغیرهای مؤثر بر بازده سهام در تحقیقات رگرسیونی بهره برد. دلیل اول همان‌طوری که در مقدمه نیز اشاره شد کاهش بیش از حد درجه آزادی است که انجام آزمون‌ها را ناممکن می‌سازد و دلیل دوم همخطی بین برخی از متغیرهاست. از طرف دیگر حذف بدون منطق متغیرهای مهم منجر به پدیده "متغیر حذف شده" و لذا ایجاد رگرسیون کاذب و شناسایی روابط نادرست می‌شود. (افلاطونی، ۱۴۱، ۱۳۹۲) لذا برای کاستن از حجم داده‌ها باید دقت کافی مبذول شود. محققان آمار، علوم کامپیوتر و ریاضیات کاربردی در طی سال‌ها تحقیق و بررسی مجموعه‌ای از روش‌های کاستن از حجم داده‌های خطی و غیرخطی را ارائه کرده‌اند. به طور خاص دو روش برای کاستن از ابعاد داده وجود دارد (ژانگ و آنکه، ۲۰۱۷، ۱۲۷) یکی انتخاب مربوطترین متغیرها از مجموعه داده‌های اولیه است که معمولاً به آن انتخاب خصیصه^۶ می‌گویند و دیگری تولید گروه کوچک‌تری از داده با استفاده از متغیرهای جدید، به گونه‌ای است که این متغیرهای جدید، هر یک ترکیبی از متغیرهای اولیه باشند.

اکثر محققان داخلی بازده برای داده‌کاهی متغیرهای خود و مواجهه با پدیده هم‌خطی از روش اول استفاده کرده و تلاش کرده‌اند مربوط‌ترین متغیرها را که معمولاً با معیار همبستگی مشخص می‌شوند برای تحلیل‌ها انتخاب کنند. به عنوان مثال نمازی و رستمی (۱۳۸۵، ۱۱۱) ضمن سنجش رابطه بازده و نسبت‌های مالی در هر گروه نسبت مالی نسبت‌هایی را که دارای رابطه درونی کمتری با هم بودند انتخاب کردند و یا ستایش و کاظمی (۱۳۹۵) با استفاده از دو تکنیک ضریب همبستگی و ریلیف، ابتدا متغیرها را رتبه‌بندی کرده و سپس ۷ متغیر را به عنوان مدل نهایی ارائه کردند و یا همت‌فر و همکاران (۱۳۹۰) سه متغیر را از مجموعه متغیرهای مؤثر بر بازده سهام برای انجام تحلیل‌ها انتخاب کردند. به طور کلی محققان داخلی بیشتر بر ادبیات موضوع و قضاوت فردی برای انتخاب متغیرهای مناسب تکیه کرده‌اند.

دسته دوم روش‌های پیش‌گفته شامل مجموعه‌ای متنوعی از روش‌های آماری هستند. سوززانو و همکاران^۸ (۲۰۱۴، ۳۱) مجموعه روش‌های موجود در این زمینه را طبقه‌بندی کرده و مفاهیم ریاضی پشتوانه آن‌ها را شرح دادند. آن‌ها اعلام کردند که از بین تکنیک‌های کاستن از ابعاد داده، روش "تحلیل مؤلفه اصلی"^۹ و نسخه‌های مختلف آن نسبت به سایر روش‌های کاستن از ابعاد داده ساده‌تر و قابل درک‌تر هستند. "تحلیل مؤلفه‌های اصلی" قدیمی‌ترین و شناخته‌شده‌ترین روش استخراج مهم‌ترین متغیرها از بین مجموعه داده‌هایی است که دارای ابعاد زیادی هستند. این روش به کار پیرسون^{۱۰} (۱۹۰۱) برمی‌گردد و بر اساس ایده تعریف یک سیستم جدید از داده‌های خام قرار دارد که می‌تواند در قالب متغیرهای بسیار کمتری بدون از دست دادن اطلاعات زیاد ارائه شود. در واقع "تجزیه و تحلیل مؤلفه‌های اصلی" نوعی از تجزیه و تحلیل آماری است که می‌توان با استفاده از آن تعداد کمتری از متغیرها را از بین متغیرهای اولیه انتخاب کرد. ویژگی اصلی این روش تمرکز بر متغیرهایی با همبستگی بالا است به طوری که اگر متغیرها دارای همبستگی بالایی نباشند نمی‌توان از این روش استفاده کرد.

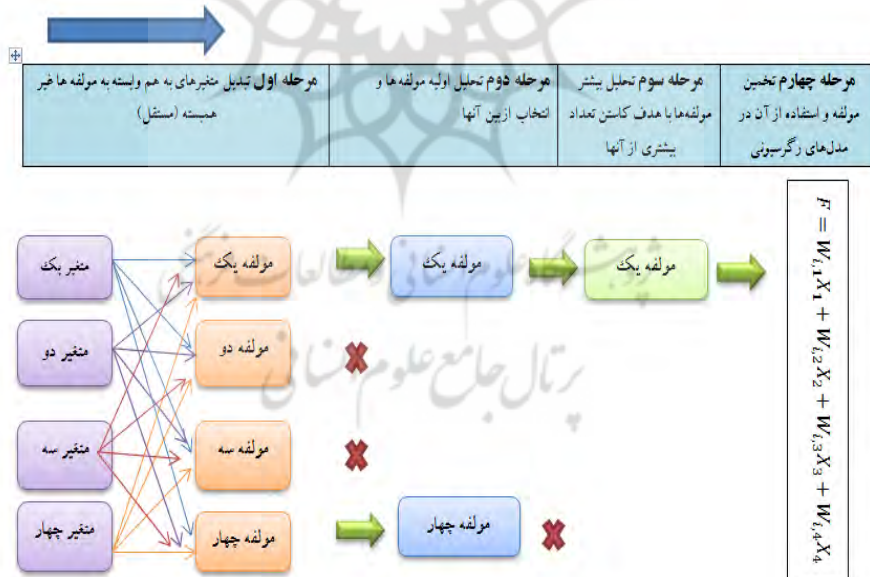
۲-۱- روش "تحلیل مولفه‌های اصلی"

روش تحلیل مولفه‌های اصلی راهی برای تعیین الگوی موجود در داده‌ها و بیان داده‌ها به گونه‌ایست که تفاوت‌ها و شباهت‌های آن‌ها برجسته شود. در جایی که پیدا کردن الگوی موجود در داده‌ها در زمانی که آن‌ها دارای ابعاد مختلفی هستند به علت عدم توانایی در استفاده از ابزارهای ترسیمی دشوار است، روش "تحلیل مولفه‌های اصلی" ابزاری قوی برای تحلیل داده‌ها است. این روش دارای کاربردهای مختلفی در علوم گوناگون از روانشناسی تا کامپیوتر، اقتصاد، مدیریت و حسابداری است. محدث (۱۳۸۹، ۶) مهم‌ترین کاربردهای روش مذکور را کاهش حجم داده‌ها،

ساختن شاخص‌ها و تحلیل داده‌هایی با ابعاد چندگانه می‌داند. به طور خلاصه در این روش متغیرهای همبسته موجود به یک (یا چند) مجموعه از مولفه‌های غیرهمبسته خلاصه می‌شوند که هر یک از آن‌ها ترکیب خطی از متغیرهای اصلی بوده و به آن‌ها مؤلفه اصلی می‌گویند. این مولفه‌ها از بردارهای ویژه ماتریس کواریانس یا ماتریس همبستگی متغیرهای اصلی به دست می‌آیند، به عبارت بهتر در این کاربرد از "تحلیل مولفه‌های اصلی"، به دنبال کم کردن یک مجموعه k عضوی به تعداد m مؤلفه هستیم به گونه ای که این m مؤلفه بیشترین میزان واریانس k متغیر را توضیح دهند. به عبارت دیگر هدف، کم کردن تعداد متغیرها به m مؤلفه مهم‌تر است. هر عامل (مؤلفه) از ترکیب خطی و موزون k متغیر تخمین می‌خورد. از این رو مؤلفه F_i^{th} عبارت است از: (معادله ۱)

$$F_i = W_{i,1}X_1 + W_{i,2}X_2 + \dots + W_{i,k}X_k \quad (1)$$

مراحل اجرای روش "تحلیل مولفه‌های اصلی" به شکل خلاصه در شکل یک ارائه شده است. در این شکل فرض می‌شود ۴ متغیر هممخت جهت تحلیل اولیه در دست است که از این بین، در نهایت یک مؤلفه (نه لزوماً) استخراج می‌شود. در ادامه هر مرحله با تفصیل بیشتری شرح داده شده است.



شکل ۱- مراحل روش "تحلیل مؤلفه اصلی"

منبع: یافته‌های پژوهشگر

۲-۱-۱- تشخیص مناسب بودن داده‌ها برای انجام روش

قدم اول در اجرای روش "تحلیل مؤلفه اصلی" شناسایی گروه‌های همبسته از بین متغیرها است. برای این منظور از دو ماتریس همبستگی یا کوواریانس می‌توان استفاده کرد. در جایی مانند تحلیل متغیرهای مؤثر بر بازده سهام که تعداد متغیرها زیاد است احتمالاً بین برخی از آن‌ها شاهد پدیده همخطی هستیم. لذا لازم است گروه‌هایی متشکل از متغیرهای همبسته ایجاد و هرکدام از گروه‌ها جداگانه مورد بررسی و تحلیل قرار گیرند زیرا همه متغیرها با یکدیگر همخط نیستند. جهت حصول اطمینان از این که اعضای گروه، همبستگی کافی برای انجام تحلیل‌های مؤلفه اصلی را دارند؟ از آزمون‌های کُرَوِیت بارتلت^{۱۱} و معیار کیسر، میسر و اولکین^{۱۲} استفاده می‌شود. در آزمون کُرَوِیت بارتلت فرض صفر این است که نمونه مورد استفاده به صورت رندوم انتخاب شده و ماتریس همبستگی به دست آمده ماتریس واحد^{۱۳} است^{۱۴} (عدم وجود همبستگی). در مورد معیار کیسر، میسر و اولکین (معادله ۲)، همبستگی نسبی یک متغیر با مجموعه متغیرها مورد توجه قرار می‌گیرد. این معیار عبارت است از نسبت مجموع مربع ضریب همبستگی ساده بین متغیر X_i و سایر X های موجود در مجموعه داده تقسیم بر همان مجموع به اضافه مربع ضریب همبستگی نسبی بین X_i و سایر X . (توجه کنیم که r می‌تواند به عنوان واریانس لحاظ گردد) یعنی:

$$MSA = \frac{\sum r_{ij}^2}{\sum r_{ij}^2 + \sum pr_{ij}^2} \quad (2)$$

معمولاً میزان قابل قبول این متغیر بالای ۰/۵ و در صورتی که بالای ۰/۸ باشد داده‌ها برای تحلیل عالی تلقی می‌شوند. (بیومونت^{۱۵}، ۲۰۱۲، ۷)

۲-۱-۲- تغییر متغیرها به مؤلفه با استفاده از محاسبه مقادیر و بردارهای ویژه^{۱۶} ماتریس‌های همبستگی

معمولاً با تشکیل k معادله و k مجهول می‌توان k مؤلفه مورد نظر را از بین داده‌ها استخراج کرد. روش کار به این صورت است که واریانس موجود در ماتریس همبستگی تبدیل به k مقدار ویژه می‌شود. هر مقدار ویژه نشان دهنده سهم مؤلفه مورد نظر از واریانس کل است و هر یک از مؤلفه‌ها ترکیب خطی k متغیر است. اولین جزء بزرگ‌ترین مقدار (درصد) واریانس را توضیح می‌دهد و دومین جزء بزرگ‌ترین درصد از میزان واریانس باقی‌مانده پس از کسر واریانس مربوط به مؤلفه اول را نشان می‌دهد و در نهایت کل واریانس توسط مؤلفه‌ها توضیح داده خواهد شد و به این ترتیب

اهمیت مولفه‌ها مشخص می‌شود. نکته مهم این است که این مولفه‌ها را می‌توان ابعاد یک فضای k بعدی تصور کرد که در آن هر بعدی بر بعد دیگر عمود است.

مقادیر ویژه (λ) از معادله ۳ به دست می‌آیند. در معادله ۳، I ماتریس واحد و A ماتریس ضرائب است. (شیخ الاسلامی و محمودآبادی، ۱۳۹۱، ۳۲۹)

$$\det(A - \lambda I) = 0 \quad (3)$$

۲-۱-۳- انتخاب مؤلفه‌های مناسب با استفاده از معیارهای موجود

از آنجایی که هدف کم کردن از تعداد مولفه‌ها است باید از بین مولفه‌های شناسایی شده در مرحله قبل مناسب‌ترین مولفه‌ها انتخاب شوند. معیار انتخاب، میزان توضیح واریانس کل توسط هر مؤلفه است. این معیار اصولاً در قالب مفهوم مقادیر ویژه بردارهای همبستگی بیان می‌شود. یک قاعده سرانگشتی این است که تنها مولفه‌هایی باقی بمانند که مقادیر ویژه آن‌ها یک یا بیشتر از آن است. (محدث، ۱۳۸۹، ۵) راه دیگر برای کم کردن از تعداد متغیرها این است که از آزمون اسکری^{۱۷} استفاده کنیم. این نمودار دیدی بصری برای تصمیم‌گیری در مورد نقطه‌ای که از آن به بعد واریانس توسط اعضا به میزان زیادی توضیح داده نمی‌شود ارائه می‌کند.^{۱۸}

۲-۱-۴- چرخش محورها^{۱۹} و محاسبه معیارهای انتخاب بعد از چرخش با هدف کم کردن بیشتر مؤلفه‌ها

می‌توان از نتایجی که تا این مرحله به دست آمده است در تحلیل‌ها استفاده کرد ولی به هر حال انجام تحلیل‌های بیشتر با هدف حذف بیشتر مولفه‌های غیرمهم و ساده‌سازی بیشتر، توصیه شده است (بیومونت، ۲۰۱۲، ۱۷-۱۸). این مرحله با تکیه بر مفهوم چرخش محورها^{۲۰} صورت می‌گیرد. به طور خلاصه در این مرحله محورها چرخیده و واریانس توضیح داده شده مجدداً برای مولفه‌های باقی مانده محاسبه می‌شود. این واریانس مساوی جمع مربع ضرایب متغیرها بر اساس مؤلفه‌هاست^{۲۱} و می‌توان مولفه‌هایی را که واریانس توضیح داده شده توسط آن‌ها بعد از چرخش، کمتر از یک می‌شود را حذف کرد. (ونش^{۲۲}، ۲۰۱۲، ۱۴)

۲-۱-۵- محاسبه ماتریس ضرائب امتیاز عناصر^{۲۳}

این ماتریس بر پایه بردارهای ویژه متغیرها و با استفاده از معادله ۴ محاسبه می‌شود. به ازای هر مقدار ویژه محاسبه شده یک بردار ویژه محاسبه می‌شود این بردارها در واقع ضرائب متغیرها در

معادلات مربوط به مولفه‌ها هستند. بعد از محاسبه مقدار ویژه، با معادله ۴، بردار مربوطه محاسبه می‌شود. در این معادله، x بردار ویژه است. سایر عناصر در معادله ۳ معرفی شده‌اند. توجه شود که این معادله زمانی جواب دارد که مقدار ویژه جواب داشته باشد.

$$(AX = \lambda X) \rightarrow x(A - \lambda I) = 0 \quad (4)$$

در این مرحله مناسب است که برای هر مؤلفه اسم مناسبی انتخاب شود تا نشان دهنده ماهیت متغیرهای موجود در مؤلفه باشد.

۲-۱-۶- تخمین مؤلفه‌ها و استفاده از آن‌ها در تحقیقات بازده سهام

در نهایت با توجه به ماتریس ضرائب امتیاز عناصر، مولفه‌های باقی مانده تخمین می‌خورند و در مدل‌های رگرسیونی به جای متغیرهای همخط مورد استفاده قرار می‌گیرند (معادله ۱ را ببینید)

۳- پیشینه پژوهش و بیان سؤالات تحقیق

هارگریوز و مانی^{۲۴} (۲۰۱۳) از روش "تحلیل مولفه‌های اصلی" برای داده‌های متغیرهای مؤثر بر بازده سهام با هدف پیش‌بینی بازده استفاده کردند. تحقیق آن‌ها در مورد شرکت‌های فعال در صنعت درمان استرالیا انجام شد. آن‌ها مجموعه‌ای متشکل از ۲۲ متغیر را مورد تحلیل قرار دادند و در نهایت با ارائه دو مؤلفه متغیرهای مؤثر بر بازده سهام را محدود کردند.

وانگ و چویی^{۲۵} (۲۰۱۳) از روش "تحلیل مولفه‌های اصلی" در خلاصه سازی عناصر یک شاخص ترکیبی متشکل از دو شاخص اصلی بازار سهام کره و هنگ‌کنگ استفاده کردند. آن‌ها با استفاده از دو مؤلفه، ابتدا عناصر شاخص مذکور را خلاصه کرده سپس تحلیل‌های خود را بر روی این شاخص ارائه و تلاش کردند جهت حرکت شاخص را پیش‌بینی کنند.

واندر ماتن و همکاران^{۲۶} (۲۰۰۹) روش "تحلیل مولفه اصلی" را با ۱۲ روش کاستن از ابعاد داده غیرخطی مشهور مقایسه کردند. آن‌ها این روش‌ها را بر روی دو مجموعه داده ساختگی و واقعی اعمال کردند. نتایج نشان داد که هرچند تکنیک‌های غیرخطی در برخی از داده‌های ساختگی، عملکرد خوبی نشان داده بودند ولی هیچ یک از آن‌ها نتوانستند در ارتباط با داده‌های واقعی، کارکردی بهتر از روش "تحلیل مولفه اصلی" داشته باشند.

ژانگ و آنکه (۲۰۱۷، ۱۲۷) با انتقاد از رویه‌های جاری پیش‌بینی بازده روزانه که مبتنی بر تکنیک‌های معمول داده‌کاوی هستند اعلام کردند این تکنیک‌ها اساساً هرگاه که تعداد متغیرها زیاد باشد ناکارا یا ناقص هستند. آن‌ها سه تکنیک "تحلیل مولفه‌های اصلی"، تحلیل مولفه‌های

اصلی مبتنی بر منطق فازی^{۲۷} و تحلیل مؤلفه‌های اساسی مبتنی بر هسته‌های مرکزی^{۲۸} را در پیش‌بینی بازده روزانه شرکت‌های "اس. اند. پی ۵۰۰"^{۲۹} آزمون و مقایسه کردند. نتایج آن‌ها نشان داد طبقه بندی داده‌ها با استفاده از روش "تحلیل مؤلفه اصلی" دقت طبقه بندی بالاتری به نسبت دو معیار دارد.

در ایران بکارگیری روش‌های کاستن از حجم داده‌ها در زمینه پیش‌بینی بازده، در بازار سرمایه مورد توجه اندکی قرار گرفته است. (ستایش و کاظمی، ۱۳۹۵، ۳). محققان اکثراً ترجیح داده‌اند با توجه به مبانی تئوریک برخی از متغیرها را انتخاب و در بافت بازار سرمایه آزمون کنند. با این وجود در سال‌های اخیر تلاش‌هایی در این زمینه دیده می‌شود. ولی‌زاده (۱۳۹۳) در حالی که در پایان نامه تحصیلی خود به دنبال ارائه مدلی برای پیش‌بینی بازده سهام بود برای تعیین اعتبار اولیه داده‌های خود از تحلیل عاملی استفاده کرد. هر چند وی تنها تحلیل عاملی را برای تأیید ساختار اولیه داده‌های خود مورد استفاده قرار داد و در ادامه مدل خود را با تکنیک رگرسیون گام به گام و بدون توجه به عوامل شناسایی شده ارائه کرد با این وجود کار وی یکی از اولین کارهای انجام شده با هدف کاستن از حجم داده‌ها است. نژاد فهیم و همکاران (۱۳۹۴) در مقاله‌ای ترویجی از تکنیک موسوم به لاسو برای تحلیل قدرت پیش‌بینی متغیرهای مؤثر بر بازده استفاده کردند. دوره و تعداد متغیر مورد بررسی این محققان بسیار اندک بود و هدف آن‌ها رتبه بندی قدرت متغیرها بود. آن‌ها در مجموع بازده دارایی‌ها را متغیر مهمی در پیش‌بینی بازده، معرفی کردند. در نهایت اخیراً ستایش و کاظم نژاد (۱۳۹۵) در بخشی از پژوهش خود که به مقایسه روش‌های انتخاب متغیرهای مؤثر بر پیش‌بینی بازده، اختصاص داشت تلاش کردند کارایی بکارگیری دو روش انتخاب متغیر بهینه شامل روش‌های همبستگی و ریلیف را آزمون کنند. آن‌ها ابتدا ۵۲ متغیر مؤثر بر بازده را شناسایی و با استفاده از این روش‌ها، متغیرها را رتبه‌بندی و از بین آن‌ها موثرترین متغیرها را انتخاب کردند. نتایج آن‌ها نشان داد استفاده از این دو روش نسبت به عدم استفاده از آن‌ها نتایج معنادارتری در پی دارد.

با توجه به خلاء تحقیقات مناسب در خصوص امکان و نحوه بکارگیری و نتایج تکنیک "تحلیل مولفه‌های اصلی" در کشور، در این پژوهش تلاش شده است تا قابلیت کاربرد این تکنیک در خصوص متغیرهای مؤثر بر بازده سهام آزمون شده و نتایج بکارگیری آن جهت استفاده محققان آتی در تحقیقات بازده ارائه شود. به طور خاص این پژوهش به دنبال پاسخ‌گویی به این سؤالات است که آیا با توجه به معیارهای آماری مناسب شامل آزمون کرویت بارتلت و معیار کیسر، میر و اوکلین، می‌توان روش "تحلیل مولفه‌های اصلی" را در بازار سرمایه ایران بکاربرد؟ چه متغیرهایی

باید به مؤلفه تبدیل شوند؟ مولفہ‌های موجود در مجموعہ نہایی متغیرها چه تعداد هستند؟ و معادلہ تخمین زنندہ آن‌ها چیست؟

۴- روش تحقیق

۴-۱- جامعہ و نمونہ

تحقیق حاضر از بعد ہدف کاربرد و از بعد روش زیرمجموعہ روش‌های تحلیل ہمبستگی است و از الگوی "تحلیل مؤلفہ اصلی" جهت تعیین ارتباط و انتخاب از بین متغیرهای مؤثر بر بازہ سهام استفادہ می‌کند. جامعہ تحقیق، شرکت‌های پذیرفتہ شدہ در بورس اوراق بہادار تہران هستند و در این تحقیق نمونہ گیری صورت نگرفتہ است و تنها شرکت‌هایی کہ حائز شرایط زیر هستند از جامعہ مورد بررسی حذف شدہ‌اند. شرایط زیر معیارهایی است کہ عموماً در تحقیقات مربوط بہ بازہ سهام (از ہر نوعی) رعایت می‌شوند.

۱) شرکت‌های سرمایہ‌گذاری و بانک‌ها بہ دلیل شرایط خاص و تفاوت‌های ساختاری با سایر شرکت‌های بورسی (مثلاً عدم امکان محاسبہ نرخ رشد) از جامعہ حذف شدند.

۲) شرکت‌هایی کہ پایان سال مالی آن‌ها زمانی غیر از بیست و نہم اسفند ماہ باشد و همچنین شرکت‌هایی کہ در طی سال‌های مورد بررسی تغییر سال مالی داشتہ‌اند از بررسی‌ها حذف شدند.

۳) الگوی پیش‌بینی بازہ در مورد شرکت‌های زیان دہ احتمالاً با الگوی پیش‌بینی در مورد شرکت‌های سوددہ متفاوت است. لذا این نوع شرکت‌ها نیز از جامعہ مورد بررسی حذف شدہ‌اند.

۴) شرکت‌هایی کہ بہ دلایلی مانند عدم معاملہ در طول سال، تعطیلی موقت، خروج از بورس و ... دادہ‌های مناسبی برای تحلیل نداشتند نیز از بررسی‌ها حذف شدند.

با اعمال شرایط زیر مجموعاً ۶۸ شرکت شناسایی و برای دورہ ۱۴ سالہ ۸۱-۹۴ مورد تحلیل قرار گرفتند.

۴-۲ - متغیرهای مورد بررسی

پس از بررسی ادبیات حرفہ مجموعاً ۴۷ متغیر مؤثر بر بازہ سهام شناسایی و انتخاب شدند. این متغیرها حیطہ وسیعی از متغیرهای شرکتی، بازار، حاکمیت شرکتی، اقتصادی و ... را در بر می‌گیرد. متغیرها، کد مربوط بہ آن‌ها و همچنین آمار توصیفی آن‌ها در جدول ۱ ارائه شدہ است.^{۳۰}

تعاریف عملیاتی مورد استفاده در جدول ۲ موجود است. به توصیه افلاطونی (۱۳۹۲، ۳۲) داده‌های پرت در سطح ۵٪ حدود بالا و پایین به نزدیک‌ترین عدد ویرایش شده‌اند.

جدول ۱- متغیرهای اولیه و آمار توصیفی مربوط به آن‌ها

کد	نام متغیر	میانگین	انحراف معیار
۱	ارزش بازار شرکت	1,030,944	1,601,838
۲	اقدام تعهدی	845,356	1,109,341
۳	بازده دارایی‌ها	۱۵/۹۵	۱۰/۸۸
۴	بازده سرمایه در گردش	۴۷/۳۲	۱۷۲/۰۵
۵	بازدهی سرمایه	۴۰/۶۶	۲۲/۲۴
۶	بتا	۰/۳۹	۰/۹۳
۷	تغییرات وجه نقد	6,307	25,701
۸	تغییرات ارزش جاری	۱۸۰,۲۷۷	۶۱۳,۷۸۳
۹	جمع کل دارایی‌ها	1,033,381	1,456,669
۱۰	حاشیه سود خالص	۲۲/۱۹	۱۷/۳۷
۱۱	حاشیه سود عملیاتی	۲۵/۵۲	۱۵/۴۷
۱۲	درصد تغییر شاخص کالا و خدمات مصرفی	۰/۱۹	۰/۰۸
۱۳	درصد تغییر قیمت سکه	۰/۳۰	۰/۲۳
۱۴	درصد تغییر قیمت نفت	۰/۱۵	۰/۱۷
۱۵	درصد تقسیم سود	۶۹/۱۳	۲۵/۵۷
۱۶	درصد سهامدار نهادی	۷۶/۷۰	۱۷/۳۰
۱۷	درصد سهامداران غیر نهادی	۲۳/۳۰	۱۷/۳۰
۱۸	مالکیت عمومی غیر دولتی	۰/۱۱	۰/۳۱
۱۹	رشد دارایی	۰/۲۱	۰/۱۹
۲۰	رشد درآمد	۰/۲۴	۰/۲۷
۲۱	سرمایه در گردش خالص	۹۹,۱۲۱	۲۵۶,۱۳۷
۲۲	سود (زیان) عملیاتی	۱۹۲,۸۷۸	۲۷۶,۷۷۰
۲۳	سود ویژه پس از کسر مالیات	۱۶۷,۰۰۹	۲۶۹,۴۰۰
۲۴	سود ویژه قبل از کسر مالیات	۱۹۲,۵۱۷	۳۰۴,۹۴۷
۲۵	سود به سود ناویژه	۶۲/۴۵	۲۹/۴۸
۲۶	سود تقسیمی هر سهم	۸۶۰/۰۶	۸۴۵/۷۸
۲۷	سود هر سهم	۱۱۱۸/۵۸	۹۳۳/۱۱
۲۸	سود هر سهم به قیمت ابتدای سال	۰/۲۰	۰/۱۲

کد	نام متغیر	میانگین	انحراف معیار
۲۹	عضویت در گروه	۰/۷۴	۰/۴۴
۳۰	قیمت تقسیم بر فروش سرانہ سهام	۱/۷۳	۱/۴۸
۳۱	گردش مجموع دارایی‌ها	۰/۸۱	۰/۳۰
۳۲	مالکیت خانوادگی	۰/۱۵	۰/۳۶
۳۳	مومنتیوم	۰/۶۸	۰/۴۷
۳۴	نرخ بهره نسبت به کل بدهی	۰/۰۵	۰/۰۳
۳۵	نرخ مؤثر مالیاتی	۰/۰۳	۰/۰۳
۳۶	نسبت آنی	۰/۷۹	۰/۳۴
۳۷	نسبت ارزش دفتری به بازار	۰/۵۷	۰/۳۸
۳۸	نسبت بدهی	۰/۶۲	۰/۱۴
۳۹	نسبت بدهی به ارزش ویژه	۲/۰۱	۱/۲۰
۴۰	نسبت جاری	۱/۲۹	۰/۴۶
۴۱	نسبت دارایی‌های جاری	۰/۶۶	۰/۱۷
۴۲	نسبت سود تقسیمی به قیمت	۰/۱۳	۰/۰۷
۴۳	نسبت قیمت بازار سهام تقسیم بر سود هر سهم عادی	۶/۹۷	۳/۸۸
۴۴	نسبت سود هر سهم تقسیم بر قیمت هر سهم عادی	۰/۱۸	۰/۰۸
۴۵	نقد شوندگی	۰/۰۴	۰/۰۲
۴۶	وجوه نقد عملیاتی	۱۳۲۹۳۶	۲۱۱۶۹۲
۴۷	حجم پول (نقدینگی جامعہ)	۰/۱۹	۰/۳۴

منبع: یافته‌های پژوهشگر

جدول ۲- تعاریف عملیاتی متغیرها

نام متغیر	تعریف عملیاتی
ارزش بازار شرکت	عبارت است از تعداد سهام در پایان سال مالی ضرب در قیمت سهام در آخر سال
اقدام تعهدی	خالص دارایی عملیاتی نماینده اقدام تعهدی است و از کسر بدهی عملیاتی از دارایی عملیاتی به دست می‌آید. دارایی عملیاتی، کل دارایی‌ها منهای وجوه نقد و سرمایه گذاری‌ها و بدهی عملیاتی، کل بدهی‌ها منهای بدهی جاری است.
بازده دارایی‌ها	عبارت است از سود خالص تقسیم بر جمع دارایی‌های ثابت شرکت‌های بورسی.
بازده سرمایه در گردش	عبارت است از سود تقسیم بر سرمایه در گردش شرکت‌های بورسی
بازدهی سرمایه	عبارت است از سود خالص تقسیم بر جمع حقوق صاحبان سهام

نام متغیر	تعریف عملیاتی
بتا	معیار ریسک سیستماتیک شرکت‌ها و عبارت است از کواریانس بازده بازار و شرکت تقسیم بر واریانس بازده بازار
تغییرات وجه نقد	عبارت است از تغییر میزان وجوه نقد و معادل‌های آن در طی سال مورد بررسی
تغییرات ارزش جاری	عبارت است از ارزش بازار شرکت در انتهای سال منهای ارزش بازار همان شرکت در ابتدای سال
جمع کل دارایی‌ها	جمع دارایی شرکت‌های بورسی است
حاشیه سود خالص	عبارت است از سود خالص تقسیم بر فروش شرکت‌های بورسی
حاشیه سود عملیاتی	عبارت است از سود عملیاتی تقسیم بر فروش شرکت‌های بورسی
درصد تغییر شاخص کالا و خدمات مصرفی	عبارت است از مابه التفاوت شاخص بهای کالا و خدمات مصرفی سال جاری و سال قبل تقسیم بر مقدار شاخص در سال قبل
درصد تغییر قیمت سکه	عبارت است از مابه التفاوت متوسط قیمت سکه در سال جاری و سال قبل تقسیم بر متوسط قیمت سکه در سال قبل
درصد تغییر قیمت نفت	عبارت است از مابه التفاوت متوسط قیمت نفت در سال جاری و سال قبل تقسیم بر متوسط قیمت نفت در سال قبل
درصد تقسیم سود	عبارت است از سود تقسیمی تقسیم بر سود کل شرکت
درصد سهامدار نهادی	متغیری که نشان دهنده میزان مالکیت شرکت‌های سرمایه‌گذاری، هلدینگ، بیمه، صندوق بازنشستگی، صندوق سرمایه‌گذاری، تأمین سرمایه، مؤسسات مالی، بانک‌ها و افراد حقیقی و حقوقی بیش از ۵٪ سهام است. (ولی زاده ۱۳۹۳)
درصد سهامداران غیر نهادی	عبارت است از یک منهای درصد مالکیت سهام‌داران نهادی
مالکیت عمومی غیر دولتی	متغیری دامی است و در صورتی که شرکت تحت کنترل (مالکیت ۵۰ درصدی) نهادهای عمومی غیردولتی باشد عدد یک و در غیر این صورت عدد صفر به خود می‌گیرد
رشد دارایی	تفاوت دارایی‌ها در طی سال تقسیم بر دارایی‌های ابتدای سال
رشد درآمد	تفاوت فروش سال جاری و سال قبل تقسیم بر فروش سال قبل
سرمایه در گردش خالص	جمع دارایی جاری منهای جمع بدهی جاری است.
سود (زیان) عملیاتی	عبارت است از فروش منهای بهای تمام شده کالای فروش رفته و هزینه‌های عملیاتی.
سود ویژه پس از کسر مالیات	سود پس از کسر مالیات شرکت‌های بورسی است.
سود ویژه قبل از کسر مالیات	سود قبل از کسر مالیات شرکت‌های بورسی است
سود به سود ناویژه	عبارت است از سود خالص تقسیم بر سود ناویژه شرکت‌های بورسی
سود تقسیمی هر	عبارت است از سود تقسیمی سهام تقسیم بر تعداد سهام عادی

نام متغیر	تعریف عملیاتی
سہم	
سود ہر سہم	عبارت است از سود خالص شرکت‌های بورسی تقسیم بر تعداد سهام عادی منتشرہ آن شرکت‌ها.
سود ہر سہم بہ قیمت ابتدای سال	عبارت است از سود ہر سہم شرکت تقسیم بر قیمت سہام در ابتدای سال
عضویت در گروہ	معیاری موهومی است کہ اگر شرکتی در گروہ تجاری عضویت داشتہ و جزئی از منابع صورت‌های مالی تلفیقی باشد عدد ۱ و در غیر این صورت عدد صفر خواهد داشت.
نسبت قیمت بہ فروش سرانہ	عبارت است از قیمت یک سہم تقسیم بر فروش سرانہ سہام
گردش مجموع دارایی‌ها	عبارت است از فروش تقسیم بر کل دارایی‌های شرکت
مالکیت خانوادگی	متغیری دامی است و در صورتی کہ شرکت تحت مالکیت یک خانوادہ (۲۰٪) باشد عدد یک و در غیر این صورت عدد صفر بہ خود می‌گیرد.
مومنتیوم	متغیر موهومی و نشان دہندہ شتاب در سود است بہ این معنی کہ اگر سود شرکت نسبت بہ سال قبل افزایش یافته است ۱ و در غیر این صورت صفر است
نرخ بہرہ نسبت بہ کل بدهی	ہزینہ بہرہ تقسیم بر کل بدهی‌ها است.
نرخ مؤثر مالیاتی	عبارت است از ہزینہ مالیات شرکت‌های بورسی تقسیم بر سود آن‌ها.
نسبت آنی	عبارت است از دارایی جاری منہای موجودی کالا و سرمایہ گذاری‌ها تقسیم بر بدهی جاری
نسبت ارزش دفتری بہ بازار	عبارت است از تقسیم ارزش دفتری سہام شرکت‌های بورسی در پایان سال تقسیم بر ارزش بازار آن‌ها.
نسبت بدهی (اہرم)	عبارت است از جمع بدهی تقسیم بر جمع دارایی شرکت‌ها
نسبت بدهی بہ ارزش ویژه	عبارت است از جمع بدهی‌ها تقسیم بر جمع حقوق صاحبان سہام شرکت‌های بورسی
نسبت جاری	عبارت است از دارایی جاری بہ بدهی جاری شرکت‌های بورسی
نسبت دارایی‌های جاری	عبارت است از دارایی جاری تقسیم بر کل دارایی‌های شرکت‌های بورسی
نسبت سود تقسیمی بہ قیمت	عبارت است از میزان سود تقسیمی ہر سہم شرکت تقسیم بر قیمت سہام در پایان سال.
نسبت قیمت بازار سہام تقسیم بر سود ہر سہم عادی	نسبتی است کہ عبارت است از قیمت ہر سہم در پایان سال تقسیم بر سود ہر سہم شرکت در همان سال.
نسبت سود ہر سہم تقسیم بر قیمت ہر	نسبتی است کہ عبارت است از سود ہر سہم در پایان سال تقسیم بر قیمت ہر سہم شرکت در همان زمان.

نام متغیر	تعریف عملیاتی
سهام عادی	
نقد شونگی	تعداد روزهایی که سهام شرکت حداقل یکبار معامله شده تقسیم بر کل روزهای معاملاتی بورس (ولی‌زاده ۱۳۹۳)
وجوه نقد عملیاتی	عبارت است از جریان‌های نقدی عملیاتی شرکت
درصد تغییر در حجم پول (نقدینگی جامعه)	عبارت است از مابه‌التفاوت حجم پول جامعه در سال جاری و سال قبل تقسیم بر حجم پول در سال قبل

۳-۴- روش اجرای پژوهش

در این پژوهش ۵ مرحله معرفی شده در خصوص روش "تحلیل مؤلفه اصلی" (بخش ۲ را ببینید) گام به گام بر روی متغیرهای انتخابی اجرا و مولفه‌های مورد نظر استخراج خواهند شد.

۵- نتایج پژوهش

۵-۱- تشکیل گروه‌های همبسته سهام و آزمون مناسب بودن آن‌ها برای "تحلیل مؤلفه اصلی"

متغیرهای تحقیق ۴۷ متغیر (شامل ۴۳ متغیر نسبی و ۴ متغیر اسمی) هستند. با توجه به این‌که ارائه جدولی با ابعاد ۴۷*۴۷ اصولاً امکان پذیر نیست لذا ابتدا این جدول در رایانه محاسبه و با بررسی همبستگی بین متغیرها، متغیرهایی که دارای همخطی بالایی بودند و امکان ترکیب آن‌ها بدون مخدوش شدن دو معیار کرویت بارتلت و "کیسر، میپر و اولکین"^{۳۱} وجود داشت، شناسایی و گروه‌بندی شدند. ضریب همبستگی داده‌هایی که حداقل یکی از آن‌ها دارای مقیاس اسمی باشد ضریب همبستگی کندال (افلاطونی، ۱۳۹۵، ۲۶) است. در این پژوهش در موارد مقتضی از این ضریب در کنار ضریب همبستگی پیرسون استفاده شده است.

جداول ۳ تا ۵ نشان دهنده گروه‌های همبسته تشکیل شده از متغیرها هستند. در این جدول‌ها ضریب همبستگی متغیرها ارائه شده است. برای بدست آوردن نام متغیرها می‌توان کد ارائه شده برای هر متغیر را در جدول یک یافت. نتایج نشان می‌دهد در هر سه گروه فرض صفر معیار بارتلت رد شده و لذا گروه‌ها همبستگی مناسب برای انجام تحلیل‌ها را دارند. همچنین معیار کیسر، میپر و اولکین نشان می‌دهد هر سه گروه برای تحلیل‌ها مناسب هستند. بنا بر این معیار داده‌های گروه دوم برای انجام تحلیل‌ها عالی و داده‌های گروه سوم و اول به ترتیب بسیار خوب و خوب تلقی می‌شوند و می‌توان تحلیل‌ها را انجام داد.

جدول ۵- همبستگی متغیرها گروه سوم

کد متغیر	3	5	10	11	15	25	26	27	30	35	37
3	۱/۰۰	۰/۸۰	۰/۷۴	۰/۷۴	۰/۳۵	۰/۵۷	۰/۷۲	۰/۷۴	۰/۵۶	۰/۵۸	۰/۴۹
5	۰/۸۰	۱/۰۰	۰/۵۷	۰/۶۰	۰/۳۳	۰/۴۸	۰/۷۵	۰/۸۰	۰/۴۴	۰/۵۰	۰/۶۵
10	۰/۷۴	۰/۵۷	۱/۰۰	۰/۸۸	۰/۳۲	۰/۷۶	۰/۵۴	۰/۵۵	۰/۷۴	۰/۴۲	۰/۳۳
11	۰/۷۴	۰/۶۰	۰/۸۸	۱/۰۰	۰/۳۷	۰/۵۲	۰/۵۷	۰/۵۷	۰/۶۹	۰/۵۹	۰/۳۹
15	۰/۳۵	۰/۳۳	۰/۳۲	۰/۳۷	۱/۰۰	۰/۲۱	۰/۵۱	۰/۳۴	۰/۲۳	۰/۲۶	۰/۲۳
25	۰/۵۷	۰/۴۸	۰/۷۶	۰/۵۲	۰/۲۱	۱/۰۰	۰/۴۲	۰/۴۵	۰/۵۰	۰/۲۶	۰/۲۳
26	۰/۷۲	۰/۷۵	۰/۳۲	۰/۳۷	۰/۵۱	۰/۴۲	۱/۰۰	۰/۹۶	۰/۴۵	۰/۵۱	۰/۵۱
27	۰/۷۴	۰/۸۰	۰/۵۵	۰/۵۷	۰/۳۴	۰/۴۵	۰/۹۶	۱/۰۰	۰/۴۴	۰/۵۳	۰/۵۳
30	۰/۵۶	۰/۴۴	۰/۷۴	۰/۶۹	۰/۲۳	۰/۵۰	۰/۴۵	۰/۴۴	۱/۰۰	۰/۳۶	۰/۳۶
35	۰/۵۸	۰/۵۰	۰/۴۲	۰/۵۹	۰/۲۶	۰/۲۶	۰/۵۱	۰/۵۳	۰/۳۶	۱/۰۰	۰/۳۸
37	۰/۴۹	۰/۶۵	۰/۳۳	۰/۳۹	۰/۲۳	۰/۲۳	۰/۵۱	۰/۵۳	۰/۳۶	۰/۳۸	۱/۰۰
	Bartlett's test		۰/۰۰		MSA		۰/۷۸				

منبع: یافته‌های پژوهشگر

۵-۲- انتخاب مؤلفه‌های مناسب با استفاده از معیارهای موجود

جدول‌های (۶،۷،۸) نشان دهنده واریانس توضیح داده شده توسط مولفه‌ها است. در سمت راست این جدول‌ها، اطلاعات مربوط به میزان واریانس توضیح داده شده توسط هر مؤلفه ارائه شده است. برای سهولت استفاده درصد واریانس مربوط به هر مؤلفه و همچنین درصد واریانس به صورت انباشته نیز ارائه شده است. همچنین در سمت چپ این جدول‌ها مولفه‌هایی که از بین کل مولفه‌ها دارای مقدار واریانس توضیح داده شده بالای یک هستند به عنوان مولفه‌های منتخب ارائه مجدد شده‌اند.

بنابر نتایج ارائه شده در گروه اول تنها یک مؤلفه دارای واریانس توضیح داده شده بالای یک بوده و انتخاب می‌شود. با توجه به تعداد نسبتاً کم متغیرهای این دسته، این نتیجه قابل پیش‌بینی بود. در دسته‌های دوم و سوم که تعداد متغیرهای همخط بیشتر هستند دو مؤلفه دارای مقدار ویژه بزرگ‌تر از یک بوده و به عنوان مؤلفه اصلی شناخته شده‌اند. این موارد در نمودارهای اسکری (اشکال ۲،۳،۴) نیز دیده می‌شود.

در زمان گروه‌بندی متغیرها، دو متغیر با همبستگی منفی کامل شناسایی شدند. (درصد سهامدار نهادی و غیر نهادی). از آنجایی که این دو متغیر اساساً یک متغیر هستند که با دو نگاه

مورد توجہ قرار می‌گیرند یکی از این متغیرها حذف شد. به هر حال نتیجه تشکیل گروه همبسته‌ای متشکل از این دو متغیر شامل معیار بارتلت (۰,۵)، تعداد مؤلفه انتخابی، یک مؤلفه و ضریب هر مؤلفه در معادله رگرسیون، ۵۰ درصد (میانگین دو متغیر) است. البته باید توجہ داشت حذف متغیرهای با همبستگی کامل در زمان استفاده از روش "تحلیل مولفه‌های اصلی" در برخی از تحقیقات دیده می‌شود (به عنوان مثال هارگریوز و مانی، ۲۰۱۵، ۱۸۵)

جدول ۶- ترکیب واریانس توضیح داده شده توسط مولفه‌ها در گروه اول

مؤلفه	مقادیر ویژه			مؤلفه استخراج شده با معیار واریانس توضیح داده شده بزرگ‌تر از یک		
	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده	% از کل واریانس توضیح داده شده (انباشته)	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده	% از کل واریانس توضیح داده شده (انباشته)
1	۲/۹۲۲	۷۳/۰۶۲	۷۳/۰۶۲	۲/۹۲۲	۷۳/۰۶۲	۷۳/۰۶۲
2	۰/۵۶۹	۱۴/۲۱۵	۸۷/۲۷۷			
3	۰/۳۴۸	۸/۶۹۲	۹۵/۹۶۹			
4	۰/۱۶۱	۴/۰۳۱	۱۰۰/۰۰۰			

روش استخراج: تحلیل مولفه‌های اصلی

منبع: یافته‌های پژوهشگر

جدول ۷- ترکیب واریانس توضیح داده شده توسط مولفه‌ها در گروه دوم

مؤلفه	مقادیر ویژه			مؤلفه استخراج شده با معیار واریانس توضیح داده شده بزرگ‌تر از یک		
	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده	% از کل واریانس توضیح داده شده (انباشته)	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده	% از کل واریانس توضیح داده شده (انباشته)
1	۶/۷۰۲	۵۵/۸۵۴	۵۵/۸۵۴	۶/۷۰۲	۵۵/۸۵۴	۵۵/۸۵۴
2	۳/۱۱۲	۲۵/۹۳۷	۸۱/۷۹۱	۳/۱۱۲	۲۵/۹۳۷	۸۱/۷۹۱
3	۰/۹۰۶	۷/۵۵۰	۸۹/۳۴۱			
4	۰/۳۷۹	۳/۱۵۶	۹۲/۴۹۷			
5	۰/۲۶۳	۲/۱۸۹	۹۴/۶۸۶			
6	۰/۲۵۲	۲/۰۹۹	۹۶/۷۸۵			
7	۰/۱۴۶	۱/۲۱۵	۹۸/۰۰۱			

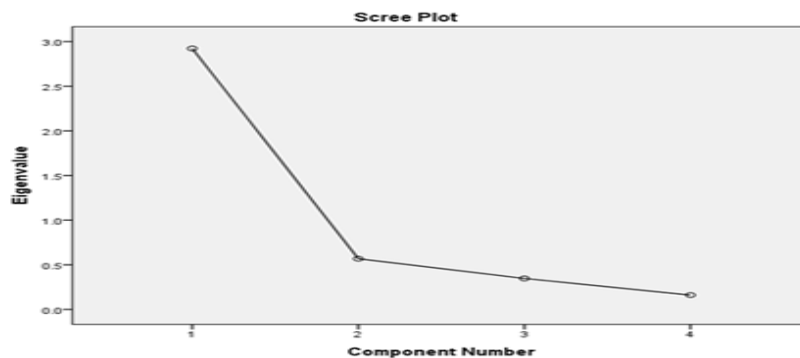
مؤلفه	مقادیر ویژه					
	مؤلفه استخراج شده با معیار واریانس توضیح داده شده بزرگ‌تر از یک			مؤلفه استخراج شده با معیار واریانس توضیح داده شده (انباشته)		
	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده
8	۰/۰۹۲	۰/۷۶۷	۹۸/۷۶۷	۰/۷۶۷		
9	۰/۰۸۰	۰/۶۶۸	۹۹/۴۳۵	۰/۶۶۸		
10	۰/۰۵۳	۰/۴۴۴	۹۹/۸۷۹	۰/۴۴۴		
11	۰/۰۱۳	۰/۱۰۶	۹۹/۹۸۵	۰/۱۰۶		
12	۰/۰۰۲	۰/۰۱۵	۱۰۰/۰۰۰	۰/۰۱۵		

روش استخراج: تحلیل مولفه‌های اصلی
منبع: یافته‌های پژوهشگر

جدول ۸- ترکیب واریانس توضیح داده شده توسط مولفه‌ها در گروه سوم

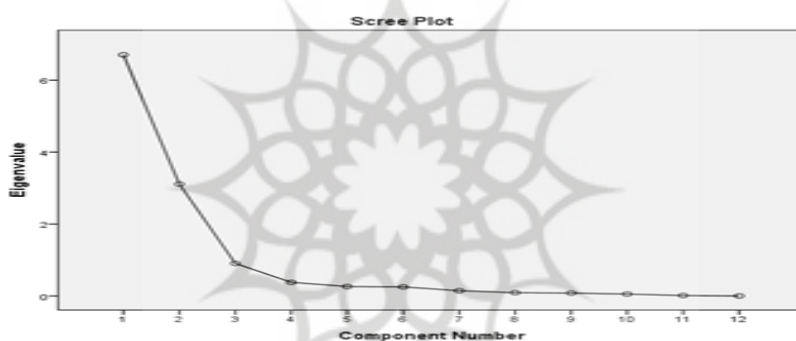
مؤلفه	مقادیر ویژه					
	مؤلفه استخراج شده با معیار واریانس توضیح داده شده بزرگ‌تر از یک			مؤلفه استخراج شده با معیار واریانس توضیح داده شده (انباشته)		
	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده	واریانس توضیح داده شده مؤلفه	% از کل واریانس توضیح داده شده
1	۶/۳۱۷	۵۷/۴۲۹	۵۷/۴۲۹	۵۷/۴۲۹	۶/۳۱۷	۵۷/۴۲۹
2	۱/۲۷۸	۱۱/۶۲۱	۶۹/۰۴۹	۱۱/۶۲۱	۱/۲۷۸	۶۹/۰۴۹
3	۰/۸۵۷	۷/۷۸۹	۷۶/۸۳۸	۷/۷۸۹		
4	۰/۷۱۳	۶/۴۸۵	۸۳/۳۲۴	۶/۴۸۵		
5	۰/۶۱۸	۵/۶۱۷	۸۸/۹۴۰	۵/۶۱۷		
6	۰/۴۵۵	۴/۱۳۲	۹۳/۰۷۲	۴/۱۳۲		
7	۰/۳۲۱	۲/۹۱۶	۹۵/۹۸۸	۲/۹۱۶		
8	۰/۲۳۳	۲/۱۱۵	۹۸/۱۰۴	۲/۱۱۵		
9	۰/۱۴۷	۱/۳۳۲	۹۹/۴۳۶	۱/۳۳۲		
10	۰/۰۴۱	۰/۳۷۷	۹۹/۸۱۳	۰/۳۷۷		
11	۰/۰۲۱	۰۰۰/۱۸۷	۱۰۰/۰۰۰	۰۰۰/۱۸۷		

روش استخراج: تحلیل مولفه‌های اصلی
منبع: یافته‌های پژوهشگر



شکل ۲- نمودار اسکری گروه اول

منبع: یافته‌های پژوهشگر



شکل ۳- نمودار اسکری گروه دوم

منبع: یافته‌های پژوهشگر



شکل ۴- نمودار اسکری گروه سوم

منبع: یافته‌های پژوهشگر

۵-۳- چرخش محورها و محاسبه معیارهای انتخاب بعد از چرخش با هدف کم کردن بیشتر مؤلفه‌ها

همانطور که قبلاً بحث شد می‌توان با تشکیل بردارهای ویژه و چرخش محورهای مربوط به آن تلاش کرد روابط قوی‌تری در خصوص ارتباط برخی از مولفه‌ها و متغیرها به دست آورد. این موضوع در روش "تحلیل مولفه‌های اصلی" برای بررسی این‌که آیا می‌توان مؤلفه دیگری را نیز حذف کرد یا خیر توصیه شده است (ونش، ۲۰۱۲، ۱۴). در مورد گروه اول با توجه به اینکه تنها یک مؤلفه شناسایی شده است این مرحله لازم نیست. حتی می‌توان در مورد گروه دوم و سوم نیز می‌توان از آن صرف نظر کرد چرا که تا این مرحله ۴۷ متغیر که ۲۷ مورد آن دارای همخطی هستند به ۱۹ متغیر (دو متغیر با همبستگی ۱- تنها یک متغیر حساب شده‌اند) و ۵ مؤلفه تبدیل شدند که دارای همخطی قابل ملاحظه‌ای نیستند. با این وجود این مرحله برای گروه‌های دوم و سوم مجدداً ارائه می‌گردد. جداول ۹ و ۱۰ نتایج مربوط به واریانس توضیح داده شده قبل و بعد از چرخش محورها را به صورت مقایسه‌ای نشان می‌دهد. برای چرخش از روش چرخش ورمکس^{۳۲} استفاده شده است. ونش (۲۰۱۲، ۱۴) این نوع از چرخش را متداول‌ترین نوع چرخش محورهای همبستگی می‌داند. نتایج مربوط به چرخش محورها نشان می‌دهد که هیچ‌یک از مولفه‌ها قابل حذف نیستند چرا که بعد از چرخش نیز معیار "جمع مربع ضرایب متغیرها" یا به عبارت دیگر میزان "واریانس توضیح داده شده" توسط آن‌ها کمتر از یک نشده است. لذا در نهایت ۵ مؤلفه شناسایی شده جهت ادامه تحلیل‌ها مورد استفاده قرار می‌گیرند.

جدول ۹- مقایسه ترکیب واریانس توضیح داده شده توسط مولفه‌های انتخابی گروه دوم قبل و بعد از چرخش محورها

مؤلفه	واریانس توضیح داده شده قبل از چرخش		واریانس توضیح داده شده بعد از چرخش	
	واریانس توضیح داده شده توسط مؤلفه	% از کل واریانس توضیح داده شده (انباشته)	واریانس توضیح داده شده توسط مؤلفه	% از کل واریانس توضیح داده شده (انباشته)
1	۶/۷۰۲	۵۵/۸۵۴	۶/۵۲۷	۵۴/۳۹۴
2	۳/۱۱۲	۲۵/۹۳۷	۳/۲۸۸	۲۷/۳۹۷

روش استخراج: تحلیل مولفه‌های اصلی

منبع: یافته‌های پژوهشگر

جدول ۱۰- مقایسه ترکیب واریانس توضیح داده شده توسط مولفه‌های انتخابی گروه سوم قبل و بعد از چرخش محورها

مؤلفه	واریانس توضیح داده شده قبل از چرخش			واریانس توضیح داده شده بعد از چرخش		
	واریانس توضیح داده شده (مؤلفه)	% از کل واریانس	% از کل واریانس	واریانس توضیح داده شده (مؤلفه)	% از کل واریانس	% از کل واریانس
1	۶/۳۱۷	۵۷/۴۲۹	۵۷/۴۲۹	۳/۷۳۷	۳۷/۳۷۰	۳۷/۳۷۰
2	۱/۲۷۸	۱۱/۶۲۱	۶۹/۰۴۹	۳/۴۴۴	۳۴/۴۳۸	۷۱/۸۰۸

روش استخراج: تحلیل مولفه‌های اصلی

منبع: یافته‌های پژوهشگر

۵-۴- محاسبه ماتریس ضرائب امتیاز عناصر

میزان تأثیر هر متغیر در تخمین مؤلفه مربوط به آن به تفکیک گروه‌ها در جداول ۱۱، ۱۲ و ۱۳ ارائه شده است. در این جداول تنها به ذکر کد متغیرها و ضریب آن در معادله رگرسیون اکتفا شده و اطلاعات مربوط به نام متغیرها و معادله کامل رگرسیونی در جدول ۱۴ ارائه شده است. همچنین معمولاً در این مرحله مولفه‌های نهایی نام‌گذاری می‌شوند تا درک ساده‌تری از آن‌ها برای خواننده ایجاد شود. نام‌گذاری بر اساس ماهیت متغیرهای تشکیل‌دهنده مؤلفه انجام می‌شود. در جدول ۱۴ نام انتخابی مولفه‌ها نیز ارائه شده است.

جدول ۱۱- ماتریس ضریب امتیاز عناصر

گروه اول	متغیر کد
مؤلفه اول	
۰/256	۲۸
۰/286	۴۲
-۰/302	۴۳
۰/321	۴۴

روش استخراج: تحلیل مولفه‌های اصلی

منبع: یافته‌های پژوهشگر

جدول ۱۲- ماتریس ضریب امتیاز عناصر

گروه دوم		متغیر کد
مؤلفه اول	مؤلفه اول	
۰/032	۰/140	۱
۰/096	۰/134	۲
۰/090	۰/136	۹
-۰/112	۰/104	۲۱
۰/042	۰/144	۲۲
۰/022	۰/145	۲۳
۰/021	۰/145	۲۴
-۰/251	۰/039	۳۶
۰/274	-۰/035	۳۸
۰/259	-۰/030	۳۹
-۰/283	۰/031	۴۰
۰/052	۰/122	۴۶

روش استخراج: تحلیل مولفه‌های اصلی

منبع: یافته‌های پژوهشگر

جدول ۱۳- ماتریس ضریب امتیاز عناصر

گروه سوم		متغیر کد
مؤلفه اول	مؤلفه اول	
-۰/003	۰/142	۳
-۰/215	۰/134	۵
۰/384	۰/133	۱۰
۰/233	۰/134	۱۱
-۰/159	۰/074	۱۵
۰/369	۰/104	۲۵
-۰/277	۰/134	۲۶
-۰/264	۰/135	۲۷
۰/335	۰/111	۳۰
-۰/136	۰/103	۳۵
310۰/	096۰/-	۳۷

روش استخراج: تحلیل مولفه‌های اصلی

منبع: یافته‌های پژوهشگر

جدول ۱۴- اطلاعات مؤلفہ‌های انتخاب شدہ نہایی

گروه	نام مؤلفہ	کد و نام و نماد متغیر	مدل نہایی(اعداد داخل مربع کد متغیر ہستند)
اول	سود سرانہ	سود ہر سہم بہ قیمت ابتدای سال (۲۸)، نسبت سود تقسیمی بہ قیمت (۴۲) نسبت قیمت بازار سہام، تقسیم بر سود ہر سہم عادی (۴۳)، نسبت سود ہر سہم تقسیم بر قیمت یک برگ سہم عادی (۴۴)	$Pps = 0.25 \begin{bmatrix} 28 \end{bmatrix} + 0.28 \begin{bmatrix} 42 \end{bmatrix} - 0.30 \begin{bmatrix} 43 \end{bmatrix} + 0.32 \begin{bmatrix} 44 \end{bmatrix}$
دوم	ارقام شرکتی ۱	ارزش بازار شرکت در آخر سال (۱)، اقلام تعہدی (۲)، جمع کل دارایی‌ها (۹)، سرمایہ در گردش خالص (۲۱)، سود (زیان) عملیاتی (۲۲)، سود (زیان) ویزہ پس از کسر مالیات (۲۳)، سود (زیان) ویزہ قبل از کسر مالیات (۲۴)، نسبت آنی (۳۶)، جح سبت بدھی (۳۸)، نسبت بدھی بہ ارزش ویزہ (۳۹)، نسبت جاری (۴۰)، وجوہ نقد عملیاتی (۴۶)	$CN_1 = 0.14 \begin{bmatrix} 1 \end{bmatrix} + 0.13 \begin{bmatrix} 2 \end{bmatrix} + 0.13 \begin{bmatrix} 9 \end{bmatrix} + 0.10 \begin{bmatrix} 21 \end{bmatrix} + 0.14 \begin{bmatrix} 22 \end{bmatrix} + 0.14 \begin{bmatrix} 23 \end{bmatrix} + 0.14 \begin{bmatrix} 24 \end{bmatrix} + 0.03 \begin{bmatrix} 36 \end{bmatrix} - 0.03 \begin{bmatrix} 38 \end{bmatrix} - 0.03 \begin{bmatrix} 39 \end{bmatrix} + 0.03 \begin{bmatrix} 40 \end{bmatrix} + 0.12 \begin{bmatrix} 46 \end{bmatrix}$
دوم	ارقام شرکتی ۲	ارزش بازار شرکت در آخر سال (۱)، اقلام تعہدی (۲)، جمع کل دارایی‌ها (۹)، سرمایہ در گردش خالص (۲۱)، سود (زیان) عملیاتی (۲۲)، سود (زیان) ویزہ پس از کسر مالیات (۲۳)، سود (زیان) ویزہ قبل از کسر مالیات (۲۴)، نسبت آنی (۳۶)، نسبت بدھی (۳۸)، نسبت بدھی بہ ارزش ویزہ (۳۹)، نسبت جاری (۴۰)، وجوہ نقد عملیاتی (۴۶)	$CN_2 = 0.03 \begin{bmatrix} 1 \end{bmatrix} + 0.09 \begin{bmatrix} 2 \end{bmatrix} + 0.09 \begin{bmatrix} 9 \end{bmatrix} - 0.11 \begin{bmatrix} 21 \end{bmatrix} + 0.04 \begin{bmatrix} 22 \end{bmatrix} + 0.02 \begin{bmatrix} 23 \end{bmatrix} + 0.02 \begin{bmatrix} 24 \end{bmatrix} - 0.25 \begin{bmatrix} 36 \end{bmatrix} + 0.27 \begin{bmatrix} 38 \end{bmatrix} + 0.25 \begin{bmatrix} 39 \end{bmatrix} - 0.28 \begin{bmatrix} 40 \end{bmatrix} + 0.05 \begin{bmatrix} 46 \end{bmatrix}$
سوم	بازدھی سود و سرمایہ ۱	بازدہ دارایی‌ها (۳)، بازدہ سرمایہ (۵)، حاشیہ سود خالص (۱۰)، حاشیہ سود عملیاتی (۱۱)، درصد تقسیم سود (۱۵)، سود بہ سود ناویژہ (۲۵)، سود تقسیمی ہر سہم (۲۶)، سود ہر سہم (۲۷)، قیمت سہام تقسیم بر فروش سرانہ ہر سہم (۳۰)، نرخ مؤثر مالیاتی (۳۵)، نسبت ارزش دفتری بہ بازار (۳۷)	$CPR_1 = 0.14 \begin{bmatrix} 3 \end{bmatrix} + 0.13 \begin{bmatrix} 5 \end{bmatrix} + 0.13 \begin{bmatrix} 10 \end{bmatrix} + 0.13 \begin{bmatrix} 11 \end{bmatrix} + 0.07 \begin{bmatrix} 15 \end{bmatrix} + 0.10 \begin{bmatrix} 25 \end{bmatrix} + 0.13 \begin{bmatrix} 26 \end{bmatrix} + 0.13 \begin{bmatrix} 27 \end{bmatrix} + 0.11 \begin{bmatrix} 30 \end{bmatrix} + 0.10 \begin{bmatrix} 35 \end{bmatrix} - 0.09 \begin{bmatrix} 37 \end{bmatrix} +$
سوم	بازدھی سود و سرمایہ ۲	بازدہ دارایی‌ها (۳)، بازدہ سرمایہ (۵)، حاشیہ سود خالص (۱۰)، حاشیہ سود عملیاتی (۱۱)، درصد تقسیم سود (۱۵)، سود بہ سود ناویژہ (۲۵)، سود تقسیمی ہر سہم (۲۶)، سود ہر سہم (۲۷)، قیمت سہام تقسیم بر فروش سرانہ ہر سہم (۳۰)، نرخ مؤثر مالیاتی (۳۵)، نسبت ارزش دفتری بہ بازار (۳۷)	$CPR_2 = -0.003 \begin{bmatrix} 3 \end{bmatrix} - 0.21 \begin{bmatrix} 5 \end{bmatrix} + 0.38 \begin{bmatrix} 10 \end{bmatrix} + 0.23 \begin{bmatrix} 11 \end{bmatrix} - 0.15 \begin{bmatrix} 15 \end{bmatrix} + 0.36 \begin{bmatrix} 25 \end{bmatrix} - 0.27 \begin{bmatrix} 26 \end{bmatrix} - 0.26 \begin{bmatrix} 27 \end{bmatrix} + 0.33 \begin{bmatrix} 30 \end{bmatrix} - 0.13 \begin{bmatrix} 35 \end{bmatrix} + 0.31 \begin{bmatrix} 37 \end{bmatrix} +$

منبع: یافته‌های پژوهشگر

۵-۵- تخمین مؤلفه‌ها و استفاده از آن‌ها در تحقیقات بازده سهام

به عنوان آخرین قدم با توجه به معادله‌های پیش‌بین معرفی شده در مرحله ۴ (جدول ۱۴) می‌توان هر مؤلفه را تخمین زد و همراه با سایر متغیرها در تحلیل‌های بازده سهام مورد استفاده قرار داد.

۶- بحث و نتیجه گیری

این پژوهش در حیطه پژوهش‌های داده‌کاهی بوده و بکارگیری روش "تحلیل مولفه‌های اصلی" در بازار سرمایه ایران را در خصوص متغیرهای مؤثر بر بازده سهام آزمون می‌کند. برای این منظور ابتدا متغیرهای مؤثر بر بازده سهام شناسایی شدند (۴۷ متغیر) و سپس برای انجام تحلیل‌ها گروه‌هایی متشکل از متغیرهای به هم وابسته تشکیل شدند (سه گروه متشکل از ۲۷ متغیر). پس از آزمون مناسب بودن گروه‌بندی‌ها، با محاسبه بردارها و مقادیر ویژه، مولفه‌های اصلی هر گروه از متغیرهای هم‌مختص تعیین و در نهایت معادله رگرسیونی هر مؤلفه اصلی با توجه به "ماتریس ضرائب امتیاز عناصر" ارائه شد. نتیجه این تحقیق نشان داد که می‌توان از این تکنیک در داده‌کاهی متغیرهای مؤثر بر بازده سهام در بازار سرمایه ایران استفاده کرد. معیارهای کرویت بارتلت و "کیسر-میر و اولکین" نشان دادند می‌توان حداقل سه گروه متغیر بهم وابسته با وضعیت مناسب برای انجام تحلیل‌ها در بازار سرمایه ایران تشکیل و با تبدیل ۲۷ متغیر وابسته به ۵ مؤلفه مستقل ضمن کم کردن از حجم متغیرهای مؤثر بر بازده سهام، پدیده هم‌خطی بین متغیرها را بدون حذف متغیرها از بین برد.

داده‌کاهی سهام در تحقیقات داخلی، اصولاً مبتنی بر رویکردهای قضاوتی بوده و توجه چندانی بر روش‌های معتبر داده‌کاهی همانند تحلیل عاملی و یا "تحلیل مولفه‌های اصلی" نشده است. از این رو مقایسه نتایج در داخل کشور امکان پذیر نیست. با این وجود در سطح بین‌الملل محققان مختلفی مانند ژانگ و آنکه (۲۰۱۷) و هارگریوز و مانی (۲۰۱۵)، سورزانو و همکاران (۲۰۱۴)، وانگ و همکاران (۲۰۱۴)، وانگ و چویی (۲۰۱۳)، ام بلدوگو^{۳۳} (۲۰۱۲) و لورینت^{۳۴} (۱۹۹۷) قابلیت کاربرد روش "تحلیل مولفه‌های اصلی" را در خصوص تحقیقات بازده در کشورهای مختلف تأیید کرده‌اند. تحقیق حاضر از بعد نتیجه با این تحقیقات همخوانی دارد. با توجه به نتایج به دست آمده از تحقیق پیشنهادت زیر برای محققان آتی ارائه می‌شود:

- ۱) آزمون بکارگیری سایر روش‌های داده‌کاهی مانند تحلیل عاملی، تحلیل مؤلفه اصلی مبتنی بر منطق فازی و... در بافت بازار سرمایه ایران

۲) مقایسه روش‌های مختلف داده‌کاهی با هدف تعیین بهترین روش کاستن از ابعاد داده‌ها در بازار سرمایه ایران

و در نهایت کاربرد رویکرد "تحلیل مؤلفه اصلی" نقاط ضعفی نیز دارد که محققان باید آن‌ها را مدنظر قرار دهند. بزرگ‌ترین مشکل رویکرد مذکور، دشواری تفسیر مولفه‌های استخراج شده است. با توجه به ترکیبی بودن این مولفه‌ها، استخراج معنا از آن‌ها دشوار است. همچنین این روش با توجه به اینکه متغیرها را حذف نمی‌کند و تنها نحوه به کارگیری آن‌ها را تغییر می‌دهد می‌تواند زمان‌بر باشد و در نهایت باید توجه داشت نتایج به دست آمده تنها برای دوره زمانی ذکر شده معتبر می‌باشد و برای کاربرد در سایر بازه‌های زمانی محاسبات باید دوباره تکرار شوند.



فهرست منابع

- ۱) برزگری خانقاه، جمال و زهرا جمالی، (۱۳۹۵)، "پیش‌بینی بازده سهام با استفاده از نسبت‌های مالی"، کنکاشی در پژوهش‌های اخیر، پژوهش‌های حسابداری، ۶(۱)، صص ۷۱-۹۲
- ۲) افلاطونی، عباس، (۱۳۹۵)، "تحلیل آماری در پژوهش‌های مالی و حسابداری با نرم افزار STATA"، انتشارات ترمه، چاپ اول
- ۳) ستایش، محمد حسین و مصطفی کاظم نژاد، (۱۳۹۵)، "سودمندی رگرسیون‌های تجمیعی و روش انتخاب متغیر پیش‌بین بهینه در پیش‌بینی بازده سهام"، پژوهش‌های حسابداری مالی و حسابرسی، ۸(۳۲)، صص ۱-۲۸
- ۴) دستگیر، محسن، ندا تاجی و رحمان ساعدی، (۱۳۹۱)، "رابطه بین متغیرهای حسابداری با بازده سهام با استفاده از مدل بازده ژانگ"، پژوهشنامه حسابداری مالی و حسابرسی، سال ۴، شماره ۱۳، صص ۴۳-۶۴.
- ۵) نژادفهییم، سیدرضا، نرگس سهرابی و هادی موقری، (۱۳۹۴)، "پیش‌بینی بازده سهام با استفاده از روش انقباضی LASSO"، مطالعات حسابداری و حسابرسی، ۴(۱۳)، صص ۴۰-۵۲
- ۶) شیخ الاسلامی، عبدالرضا، فاطمه باقری خلیلی و عباس محمودآبادی، (۱۳۹۱)، "کاهش متغیرهای ورودی در فرآیند مدل سازی تصادفات آزادراه‌ها با استفاده از روش تجزیه و تحلیل مؤلفه اصلی"، مهندسی حمل و نقل، ۳(۴)، صص ۳۲۵-۳۳۸
- ۷) محدث، فاطمه، (۱۳۸۹)، "روش تحلیل مولفه‌های اساسی و بررسی عوامل: مطالعه موردی استخراج شاخص قیمت دارایی‌ها و بررسی اثر آن بر تورم"، کتابخانه بانک مرکزی، آدرس دستیابی: www.cbi.ir، دسترسی ۹۵/۱۱
- ۸) نمازی، محمد و نورالدین رستمی، (۱۳۸۵)، "بررسی رابطه بین نسبت‌های مالی و نرخ بازده سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران"، بررسی‌های حسابداری و حسابرسی، ۴۴، صص ۱۰۵-۱۲۷
- ۹) ولی زاده لاریجانی، اعظم، (۱۳۹۳)، "تبیین مدلی برای پیش‌بینی بازده سهام"، پایان نامه دکتری حسابداری دانشگاه الزهراء(س).
- ۱۰) همت‌فر، محمود، سیدعلی اکبر حسینی، فرهاد شاه ویسی و یوسف نجفی، (۱۳۹۰)، "روابط خطی و غیرخطی بین متغیرهای حسابداری و بازده سهام شرکت‌های صنعت خودرو و ساخت قطعات"، پژوهشنامه حسابداری مالی و حسابرسی، سال ۳، شماره ۱۲، صص ۱۳۷-۱۵
- 11) Back, A. D., & Weigend, A. S, (1997), "A First Application of Independent Component Analysis to Extracting Structure From Stock Returns", International Journal of Neural Systems, 8(04), PP. 473-484.

- 12) Beaumont, R, (2012), "An Introduction to Principal Component Analysis & Factor Analysis Using SPSS 19 and R", Virtual Classroom, ONLINE AT: [Www.Floppybunny.Org/Robin/Web/Virtualclassroom/Stats/Statistics2/Pca1.Pdf](http://www.floppybunny.org/robin/web/virtualclassroom/stats/statistics2/pca1.pdf) (03/2015)
- 13) Fama, E. F., & French, K. R, (1993), "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, 33(1), PP. 3-56.
- 14) Fama, E. F., & French, K. R, (2015), "A Five Factor Asset Pricing Model", *Journal of Financial Economics*, 116 (1), PP. 1-22.
- 15) Hargreaves, C. A., & Mani, C. K, (2015), "The Selection of Winning Stocks Using Principal Component Analysis", *American Journal of Marketing Research*, 1(3), PP. 183-188.
- 16) Loretan, M, (1997), "Generating Market Risk Scenarios Using Principal Components Analysis: Methodological and Practical Considerations", *Federal Reserve Board*, PP. 23-53.
- 17) Mbeledogu, N.N., Odom., Umeh, M.N, (2012), "Stock Feature Extraction Using Principle Component Analysis", *International Conference on Computer Technology and Science*. IACSIT Press, Singapore DOI: 10.7763/IPCSIT.2012.V47.44
- 18) Ou, J. A., & Penman, S. H, (1989), "Financial Statement Analysis and the Prediction of Stock Returns", *Journal of Accounting and Economics*, 11(4), PP. 295-329.
- 19) Pearson, K, (1901), "On Lines And Planes of Closest Fit to Systems of Points in Space", *Philosophical Magazine*, 2 (6), PP. 559-572.
- 20) Sorzano, C. O. S., Vargas, J., & Montano, A. P, (2014), "A Survey of Dimensionality Reduction Techniques", *Arxiv Preprint Arxiv:1403.2877*.
- 21) Utans, J., Holt, W. T., & Refenes, A. N, (1997), "Principal Components Analysis for Modeling Multi-Currency Portfolios", *Decision Technologies for Financial Engineering (Proceedings of the Fourth International Conference on Neural Networks in the Capital Markets, NNCM-96)*, World Scientific, Singapore, PP. 359-368.
- 22) Van Der Maaten, L., Postma, E., & Van Den Herik, J, (2009), "Dimensionality Reduction: A Comparative", *Journal of Machine Learning Research*, 10 (1-41), PP. 66-71.
- 23) Wuensch, K. L, (2012), "Principle Components Analysis", Online at: [Http://Core.Ecu.Edu/Psyc/Wuensch/MV/FA.PCA.Doc T\(03/2015\)](http://Core.Ecu.Edu/Psyc/Wuensch/MV/FA.PCA.Doc.T(03/2015))
- 24) Wang, Y., & Choi, I. C, (2013), "Market Index and Stock Price Direction Prediction Using Machine Learning Techniques: An Empirical Study on the KOSPI and HIS", *Arxiv Preprint Arxiv*, PP. 1309-7119.
- 25) Wang, Z., Sun, Y., & Li, P, (2014), "Functional Principal Components Analysis of Shanghai Stock Exchange 50 Index", *Discrete Dynamics in Nature and Society*.
- 26) Zhong, X., & Enke, D, (2017), "Forecasting Daily Stock Market Return Using Dimensionality Reduction", *Expert Systems with Applications*, 67, PP. 126-139.

یادداشت‌ها

- ¹ Ou and Penman
² Fama and French

^۳ در تحلیل بنیادی بازده سهام تابعی از شرایط کلان اقتصادی، وضعیت صنعت و شرایط خاص شرکت است. (بزرگری و جمالی ۱۹۹۵، به نقل از شعری، ۱۳۳۳).

- ⁴ Arbitrage Pricing Theory (APT)
⁵ Data Reduction
⁶ Zhong and Enke
⁷ Feature selection
⁸ Sorzano
⁹ Principle Component Analysis
¹⁰ Pearson
¹¹ Bartlett's test of Sphericity
¹² Kaiser-Meyer-Olkin Measure Of Sampling Adequacy Or (MSA measure)
¹³ Identity Matrix

^{۱۴} آزمون دیگری که در این زمینه می‌تواند مورد استفاده قرار گیرد Ole Eyeball Test است.

- ¹⁵ Beaumont
¹⁶ Eigen values and Eigen vectors
¹⁷ Scree Test

^{۱۸} دو روش دیگر برای انتخاب مؤلفه‌ها عبارت‌اند از آزمونهای مولزی و آزمون نقشه ویلسر (Velicer's Map Test).

- ¹⁹ Vectors Rotation

^{۲۰} برای آشنایی بیشتر با مفاهیم چرخش محورها به منابع معرفی شده مراجعه کنید.

- ²¹ Sum Of Squared Loadings (SSL) Across Variables
²² Wuensch
²³ factor score coefficient matrix
²⁴ Hargreaves and Mani
²⁵ Wang and Choi
²⁶ Van Der Maaten
²⁷ Fuzzy Robust Principal Component Analysis
²⁸ Kernel Principal Component Analysis
²⁵ S&P 500

^{۳۰} همان‌گونه که در متن مقاله نیز ارائه شد متغیرهای پژوهش حیطه متنوعی از متغیرها را در بر می‌گیرند و هر دسته از مبنای ارائه خاصی پیروی می‌کنند. در این جدول اعداد بزرگ برای سادگی در قالب میلیون ریال ارائه شده‌اند. نرخ‌های رشد و نسبت‌ها درصد هستند. متغیرهای موهومی در قالب عددی بین صفر تا یک ارائه شده‌اند و اعداد کوچک مانند سود هر سهم و سود تقسیمی هر سهم ریال هستند. مطالعه شیوه محاسبات متغیرها که در جدول دو ارائه شده است به خواننده دید مناسبی در خصوص مبنای اعداد در جدول یک می‌دهد.

- ³¹ MSA
³² Varmax
³³ Mbeledogu
³⁴ Loretan