

پایدار سازی شاخه نگارهای مالی به عنوان روشی برای اندازه گیری تغییرات سیستمیک

(مطالعه ای در شاخص های بورس اوراق بهادار تهران)

چکیده

دندروگرام ها، اصلی ترین ابزار نگاره سازی در خوشه بندی های سلسله مراتبی هستند. بنابراین مقایسه ی دندروگرام ها، متضمن مقایسه خوشه بندی های متناظر با آن هاست. در این پژوهش، مقایسه دندروگرام شاخص های محاسبه شده در بورس اوراق بهادار تهران، به عنوان شاخصی از تغییرات سیستمیک در نظر گرفته می شود. با در نظر گرفتن شاخص های مذکور، به عنوان ارکان سیستمیک بورس و با مقایسه خوشه بندی ها و نیز محاسبه ی میزان همگرایی آن ها ، میزان پایداری موقعیت یا فاصله ی متریک شاخص ها نسبت به یک دیگر، کمی سازی می شود. به همین منظور، دندروگرام های ماهانه در قلمرو زمانی پژوهش، با ابزارهایی چون نمودار تانگلگرام و شاخص های کمی بیکر و ضریب همبستگی همسانی ها، مقایسه شده و میزان شباهت آن ها اندازه گیری شده اند. آن گاه داده های اولیه با استفاده از تحلیل مولفه های اصلی، تبدیل و ضمن حذف مولفه هایی که اهمیت کمتری دارند، خوشه بندی ها و شباهت مجددا محاسبه شدند. از مقایسه شاخه نگار های پیش از تحلیل PCA پس از آن میزان ناپایداری سیستم در قلمرو پژوهش اندازه گیری شده است.

کلیدواژه ها: خوشه بندی سلسله مراتبی، دندروگرام (شاخه نگار)، تانگلگرام (همشاخه نگار)، تحلیل مولفه های اصلی

۱- مقدمه

متعاقب بحران مالی سال ۲۰۰۸-۲۰۰۷، تلاش های بسیاری برای صورتبندی مفهوم تغییرات سیستمیک مالی صورت پذیرفته است. ریسک سیستمیک (و نه سیستماتیک) به احتمال رخدادهایی اشاره دارد که منجر به کنش-واکنش هایی در کل سیستم مالی شده و به دنبال خود زیان هایی را به فعالان این سیستم تحمیل می کنند. کمی سازی میزان تغییرات سیستمیک از جمله مهمترین دغدغه های نظریات مالی در سال های اخیر بوده است [۲۷]. در این بین، تحلیل هریک از زیرسیستم های

مالی مانند بورس اوراق بهادار یا ارزهای خارجی، بخشی از وضعیت کل سیستم را بازتاب می دهد. در این پژوهش با تمرکز بر بورس اوراق بهادار تهران، موقعیت متریک شاخص های محاسبه شده ی آن به عنوان قلمرو مطالعه انتخاب شده اند. شاخص ها محاسبه شده، اندیکاتورهای صنایع مختلف اند و دوری یا نزدیکی آن ها به یکدیگر، نمایانگر فعل و انفعالات واقعی اقتصاد در آن صنایع است. مساله اصلی آن است که آیا در طی زمان، موقعیت متریک شاخص های بورس تهران نسبت به یکدیگر، ثابت بوده و بنابراین گونه ای پایداری بر سیستم حاکم بوده است یا برعکس، تغییرات این سیستم، بر شکلی از ناپایداری دلالت دارد. بر این اساس، در بازه های مختلف زمانی، با انجام خوشه بندی سلسله مراتبی شاخص ها، دندروگرام های متناظر با آن ها تدوین و با یکدیگر مقایسه می شوند. این مقایسه به روش های مختلفی صورت می پذیرد، که در قسمت روش شناسی مطالعه، تشریح شده اند. در ادامه این نوشتار، ابتدا، ضمن تشریح مفهوم خوشه بندی سلسله مراتبی، برخی از کاربردهای متاخر آن در بازارهای مالی مرور می شود. سپس، روش های مقایسه دندروگرام ها توضیح داده شده و نتایج عددی حاصل از آن ها در بورس تهران، ارائه و تفسیر خواهد شد.

۲- مرور ادبیات

در ادبیات داده کاوی، خوشه بندی سلسله مراتبی یا همان تحلیل سلسله مراتبی خوشه ها، روشی از خوشه بندی است که به دنبال پیدا کردن سلسله مراتبی از خوشه هایی از مشاهدات، اشیاء یا اعیان مورد نظر است. نحوه جستجوی این خوشه از یکی از این دو منطق پیروی می کند. در منطق پایین به بالا یا تجمعی، هر مشاهده ای در خوشه ای تک عضوی عضویت دارد. یعنی خوشه ای با حضور خودش. در گام بعد، شبیه ترین خوشه ها، با یکدیگر ادغام می شوند، تا به خوشه هایی بزرگتر از مشاهدات منجر شود [۳۳]. معیارهای گوناگونی برای این شباهت یابی وجود دارند. بنابراین سلسله مراتب خوشه ها، از طریق ترکیب خوشه های کوچک به بزرگ حاصل می شود. در منطق دوم، ابتدا همه مشاهدات در یک خوشه بزرگ قرار گرفته و با تجزیه آن خوشه، خوشه های کوچک تر اما همگن تر شکل می گیرند. سلسله مراتب خوشه ها در این روش، منطقی بالا به پایین یا تجزیه ای دارد. هم ادغام ها (در روش اول) و هم تجزیه ها (در روش دوم) بر اساس الگوریتم های خاصی صورت می پذیرند که در داده کاوی به آن ها پرداخته می شود [۱۳، ۲۴]. نکته دیگر به انتخاب معیاری بر می گردد که شباهت هر مشاهده با مشاهده ی دیگر را نشان می دهد. متریک های مختلفی برای اندازه گیری شباهت مشاهدات یا فاصله آن ها از یکدیگر وجود دارد که از جمله آن ها می توان فاصله اقلیدسی، فاصله منتهن و فاصله ماکسیمم اشاره کرد [۲۰]. در تقسیم یا ادغام خوشه منطق پیوند آن ها با یکدیگر نیز مهم است که روش

های خاص خود را دارد. حاصل یک خوشه بندی سلسله مراتبی به صورت نگاره ای که دندروگرام یا شاخه نگار نامیده می شود، نشان داده می شود.

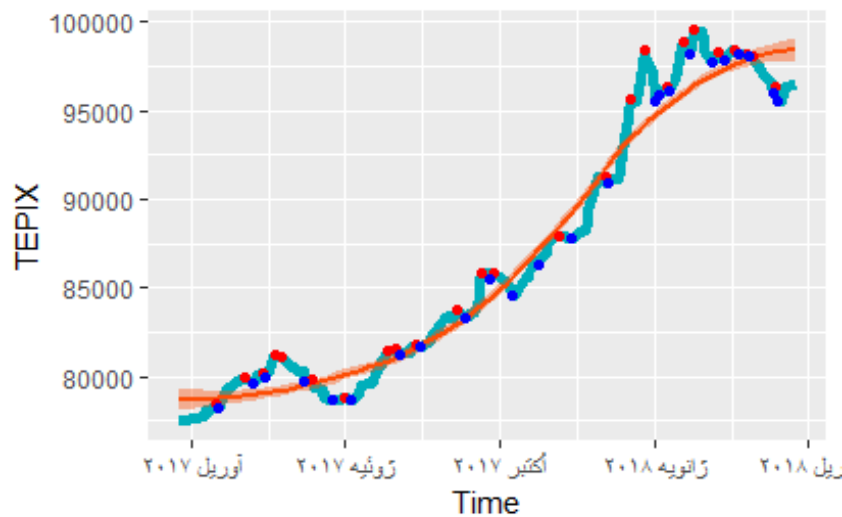
استفاده از خوشه بندی به عنوان رویکردی در تحلیل بازارهای مالی، کاربردهای روز افزونی یافته است. در این رویکرد، هر یک از سهم های موجود در بازارهای مالی به عنوان اشیاء یا مشاهداتی در نظر گرفته می شوند که می بایست ابتدا خوشه بندی شوند. خوشه بندی سهام، ممکن است بر اساس بازده آن ها، عوامل فاندamental متناظر و یا سایر مختصات مالی مهم صورت پذیرد. در بخش قابل توجهی از مطالعات، ابتدا شبکه ای از سهام تدوین می شود. منظور از شبکه، گرافی است که رئوس آن، سهم های منتخب و یال های آن، فاصله یا شباهت سهام از یکدیگر است. تحلیل این شبکه یا گراف، به روش های گوناگونی امکان پذیر است که خوشه بندی سلسله مراتبی یک از آن هاست. ویژگی های یک گراف، مانند معیارهای مرکزیت رئوس یا پیروی درجه رئوس از قانون توانی، استخراج و تحلیل درخت کمینه پوشا و سایر تحلیل های کمی تحلیل شبکه در مطالعات متعدد مالی صورت گرفته است [۱۵]. در پژوهشی سهام موجود در شاخص DAX30 با همین منطق تجزیه و تحلیل شدند [۳]. این شبکه البته در طول زمان پویایی های خود را دارد که تغییر پذیری تاریخی آن ها، موضوع برخی از پژوهش ها بوده است. دینامیسم تغییرات سیستم سهام، اطلاعات مفیدی جهت پیش بینی آن ها در زمان ها بعدی را به دست می دهد [۳۶]. تغییرات سیستم سهام در مواقع بحران نظیر بحران مالی ۲۰۰۸-۲۰۰۷ نیز در ذیل همین عنوان جای می گیرد [۳۵]. به عنوان مثال در پژوهشی متاخر که در بورس برزیل صورت گرفته است ابتدا شاخصی کمی برای حال و هوای فعالان بازار تکوین شده و سپس تاثیرات گراف سهام بورس مذکور از این شاخص کمی مطالعه شده است. شاخصی که حال و هوای فعالان بازار را نشان بدهد از تحلیل متن ها و بسامد واژگان با بار معنایی مثبت و منفی استخراج می شود که روش های تحلیل متن را به استخدام می گیرد [۱]. در مطالعه ای دیگر، که در بورس کازابلانکا، صورت گرفته است، سهام مختلف بر اساس نمای هرست خوشه بندی شده اند. نمای هرست که از سری زمانی بازده سهام محاسبه می شود، نشان دهنده تصادفی بودن قیمت یا رونددار بودن آن و به طور کلی حافظه بلند مدت سری های زمانی است. موضعی که می تواند به فرضیه بازار کارا یا شکل بدیل آن، فرضیه بازار فراکتال، مربوط باشد. این مطالعه، به چند روش، نمای هرست سهام منتخب محاسبه شده، و خوشه بندی بر اساس آن ها صورت می پذیرد [۱۸]. در برخی از مطالعات، چند بازار مالی به عنوان قلمرو پژوهش انتخاب شده اند. به عنوان مثال در مطالعه ای، ۵۲ بازار مالی مختلف از کشورهای پراکنده انتخاب و خوشه بندی شده اند. در این تحقیق حاصل خوشه بندی ها و موقعیت جغرافیایی این ۵۲

شاخص تجزیه و تحلیل شده است [۴]. در پژوهشی دیگر ۴۱ شاخص از ۴۰ کشور برای مدت ۱۲ فصل، انتخاب و میزان یکپارچگی یا عدم یکپارچگی آن ها به عنوان معیاری از روند جهانی شدن تحلیل شده است. در این پژوهش نیز دندوگرام این شاخص ها به عنوان یکی از ابزارهای منتخب بوده است [۸]. از خوشه بندی سلسله مراتبی به عنوان روشی در تکوین پرتفوی سهام نیز استفاده می شود. پس از تکوین خوشه ها، می توان پرتفوی های مرکزی و پرتفوی های پیرامونی را از یکدیگر تمیز داد. آنگاه، با استفاده از روش های بهینه سازی سید سهام در هر یک از خوشه ها، وزن بهینه ی هر یک از سهم ها در سید منتخب را تعیین کرد. مطالعه ای در بورس چین، با همین روش صورت گرفته است [۲۲]. علاوه بر محاسبات عددی، شیوه نمایش خوشه بندی ها و گراف نیز اهمیت دارد. پژوهش های مستقل اما ناکافی ای که نگاره سازی داده های مالی را دنبال کرده اند، روش های متعددی برای تدوین نگاره ها، فهم و تفسیر واقعیات مالی پیشنهاد و نقاط ضعف و قوت آن ها را بررسی کرده اند [۲۶]. رویکرد های مبتنی بر شبکه و ارتباط متقابل شاخص های مختلف در اندازه گیری و تحلیل ریسک سیستمیک نیز، پیگیری شده است. در پژوهشی ویژگی های توپولوژیک شبکه های مالی به عنوان شاخصی از ریسک سیستمیک مطرح و در بورس اروپایی استاکس محاسبه شده است [۲۱]. در این پژوهش نیز، پس از انجام خوشه بندی سلسله مراتبی در زمان های مختلف، از مقایسه این خوشه بندی ها، تغییرات سیستمیک بورس تهران، محاسبه شده است. پس از کمی کردن تغییرات سیستمی در که در قالب ضریب همبستگی شاخه نگارها، بیان شده است، استفاده از روش تجزیه مولفه های اصلی به عنوان روشی جهت پیراستن بازده ها و رسیدن به داده هایی که حاصل بازسازی مولفه های اصلی است، پیگیری شده است. این روش از تحلیل در پژوهش های پیشین مالی کاربرد های فراوانی داشته است. به عنوان مثال در پژوهشی از PCA جهت پویایی ها و تعاملات بین شاخص های بورس های گوناگون در فاصله سالها ۱۹۹۸ تا ۲۰۱۲ پرداخته شده است [۲۵]. از این روش اخیرا در تحلیل ماتریس واریانس-کواریانس حاصل از داده های با بسامد بالا نیز استفاده شده است [۱۷]. پس از بحران مالی ۲۰۱۸-۲۰۱۷ در چند پژوهش دیگر از این روش برای اندازه گیری ریسک سیستمیک در نهاد مالی و بانک های چین استفاده شده است. [۱۹، ۱۴، ۱۰] در بخش بعدی، به روش تحقیق شامل قلمرو های پژوهش، خوشه بندی سلسله مراتبی، روش مقایسه دندروگرام ها و نیز روش مولفه های اصلی، پرداخته می شود.

۳- روش تحقیق

قلمرو تحقیق و داده

داده های روزانه ی ۲۶ شاخص بورس اوراق بهادار تهران در ۲۴۱ روز کاری از سایت بورس اوراق بهادار تهران از تاریخ ۱۳۹۶/۰۱/۰۵ تا ۱۳۹۷/۰۱/۰۵ دریافت شدند. اگر $P_i(t)$ مقدار شاخص i در روز t و $P_i(t-1)$ مقدار آن در پایان روز $t-1$ را نشان دهد، آنگاه لگاریتم بازده شاخص در یک فاصله زمانی Δt با استفاده از رابطه زیر به دست می آید:



نمودار (۱). رفتار شاخص کل در قلمرو پژوهش

$$R_i(t) = \ln \frac{P_i(t)}{P_i(t-1)} \quad (۱) \text{ رابطه}$$

در بازه زمانی انتخاب شده شاخص کل بورس اوراق بهادار به صورت نمودار (۱) تغییر داشته است.

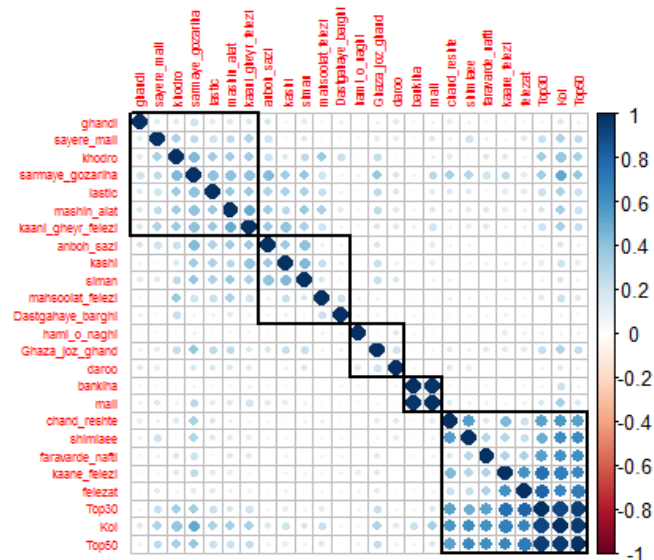
در این نمودار مشخص است که شاخص کل از عدد ۷۷۴۸ آغاز و در انتهای قلمرو زمانی پژوهش به ۹۶۴۲ می رسد. علاوه بر رفتار شاخص، روند کلی و نقاط قله و دره ی آن در نمودار مشخص است. در همین بازه، ماتریس همبستگی ۲۶ شاخص نیز محاسبه شده است که در نمودار (۲) نشان داده شده

است. عناصر قطر اصلی ماتریس همبستگی که اساس، ماتریس فاصله ی شاخص ها از یکدیگر است، یک و سایر عناصر میزان ضریب همبستگی آن ها را نشان می دهد.

بخش عمده ای از شبکه های مالی بر اساس همبستگی بازده ها به دست می آید[۳۴]. محاسبه ماتریس همبستگی اولین گام در تدوین شبکه سهام و به طور کلی خوشه بندی های مبتنی بر همبستگی است. در این پژوهش از همبستگی پیرسون برای به دست آوردن ماتریس همبستگی استفاده شده است.

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}} \quad \text{رابطه (۲)}$$

که در آن i و j نمادهای عددی از سهم ها هستند و $\langle \dots \rangle$ نشان دهنده متوسط آماری در طول دوره مورد مطالعه است. ضریب همبستگی ρ_{ij} می تواند از بازه -1 تا $+1$ را داشته باشد که در آن به ترتیب -1 بدین معناست که دو قیمت سهام همبستگی کاملاً مخالف هم دارند و $+1$ یعنی دو قیمت سهام کاملاً همبسته هستند. زمانی که $\rho_{ij} = 0$ باشد، آنگاه قیمت های دو سهم غیر همبسته هستند. ماتریس ضریب همبستگی، یک ماتریس متقارن $N \times N$ با $\rho_{ii} = 1$ در قطر اصلی است[۳۱، ۱۶].



نمودار(۲). همبستگی شاخص ها

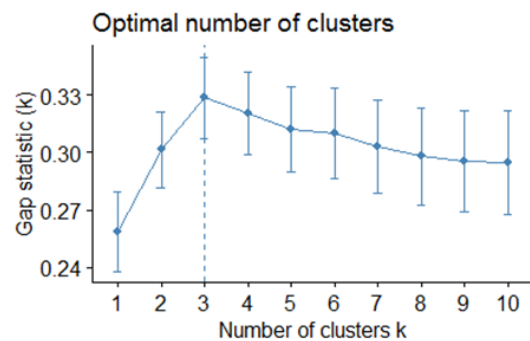
پس از محاسبه همبستگی، با استفاده از تبدیل زیر، همبستگی را به متریک فاصله تبدیل می کنند.

$$d(i, j) = \sqrt{2(1 - \rho_{ij})} \quad \text{رابطه (۳)}$$

بر اساس این متریک، بیشترین فاصله بین دو سهم، زمانی رخ می دهد که همبستگی بازده آن ها، ۱- کمترین فاصله، در همبستگی کامل ۱+ اتفاق می افتد. استفاده از این متریک به عنوان مبنایی برای خوشه بندی به دلیل خواص ریاضی ای چون متقارن بودن و نامساوی مثلثی، توجیه پذیر است [۶،۷]. در گام بعد، ماتریس فاصله ها، به عنوان ورودی روشی که جو/چ. وارد ارائه کرد، استفاده می شود. بر اساس روش وارد، تابع هدف، کمینه سازی مجموع واریانس های درون خوشه ای است. در خوشه بندی، هر یک از خوشه ها، واریانسی درون خوشه ای دارند که در حالت ایده آل، صفر است. واریانس درون خوشه ای کمتر، نشان دهنده همگنی بیشتر مشاهداتی است که در یک خوشه، جای گرفته اند. در این روش، در هر مرحله خوشه هایی با یکدیگر ادغام می شوند که ترکیب آن ها، کمترین میزان افزایش در مجموع واریانس های درون خوشه ای را به دنبال داشته باشد. به عبارت دیگر، همگن ترین خوشه ها تا حد ممکن با یکدیگر ادغام و خوشه های بزرگتر و النهایه، سلسله مراتبی از خوشه ها شکل می گیرد. بنابراین، روش وارد از جمله الگوریتم های پایین به بالا در خوشه بندی سلسله مراتبی است [۵]. در گام بعد، خوشه بندی حاصل از روش وارد، به صورت دندروگرام نمایش داده می شود. پیش از نمایش خوشه ها، می توان تعداد خوشه ها را تعیین کرد. تعیین تعداد بهینه خوشه ها به روش های گوناگونی از جمله روش البو، روش میانگین سیلووت و روش آماره شکاف است. در این

پژوهش از روش آماره شکاف که در سال ۲۰۰۱ معرفی شد، استفاده شده است [۳۰]. بر اساس داده های

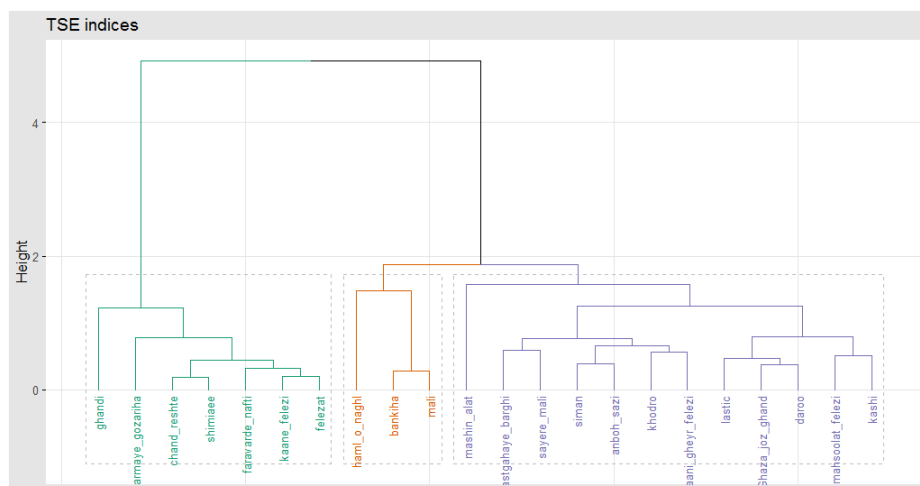
قلمرو این پژوهش، مقدار آماره شکاف محاسبه شده، در تعداد ۳ خوشه، بهینه است.



نمودار (۳). مقدار بهینه تعداد خوشه ها بر اساس آماره شکاف

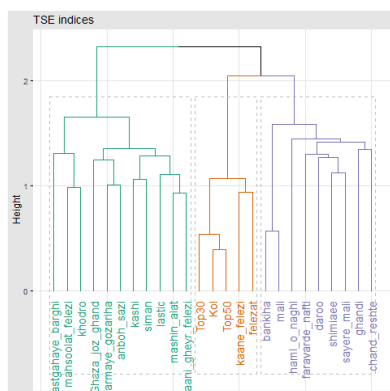
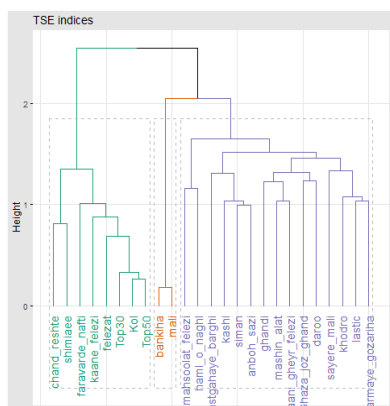
اصلی ترین ابزار نگاره سازی در خوشه بندی های سلسله مراتبی، دندروگرام یا شاخه نگار است.

در ارائه ی شاخه نگار به تعداد، خوشه ها و سلسله مراتب آن ها و نیز موقعیت سهم ها از یکدیگر توجه شده است.



نمودار(۴) شاخه نگار شاخص ها پیش از انجام PCA

بر اساس این خوشه بندی، شاخص های قندی، سرمایه گذاری، چند رشته ای، شیمیایی فراورده های نفتی، کانه فلز و فلزات در خوشه اول، بانک ها و مالی و حمل و نقل، در خوشه دوم، و سایر شاخص ها در خوشه سوم جای گرفته اند. موقعیت متریک هر یک از شاخص ها نسبت به یکدیگر در نمودار نشان داده شده است. طبیعی است که با تغییر قلمرو زمانی، خوشه بندی های متفاوتی بروز پیدا می کند. به عنوان مثال اگر قلمرو زمانی این پژوهش، به دو دوره شش ماه تقسیم شده و در هر بازه پس از خوشه بندی، شاخه نگار مربوطه تدوین شوند، می توان میزان تغییرات این دو دوره را تعیین کرد. در نمودار مشخص است که ترکیب خوشه ها در دو بازه ی مورد نظر، تا حدودی متفاوت اند و موقعیت شاخص ها، نسبت به یکدیگر، شاهد تغییراتی بوده است.



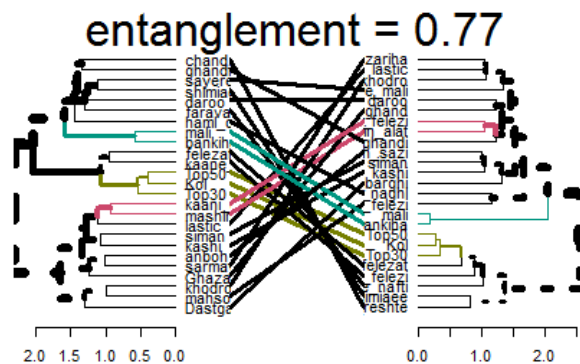
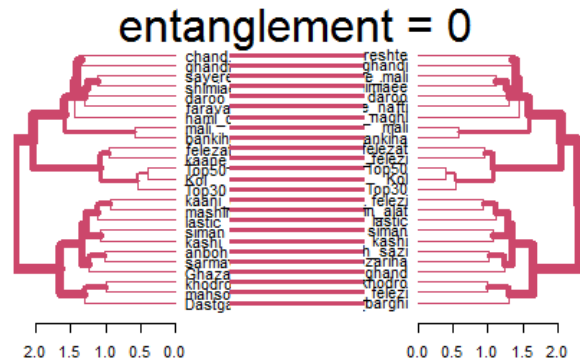
نمودار (۵). شاخه نگارهای شش ماهه

تغییرات شاخه نگارها در طول زمان می تواند نشان دهنده تغییرات سیستم باشد. مساله اصلی این پژوهش، مقایسه شاخه نگارها به دو روش است: تدوین همشاخه نگار ها و محاسبه شاخص های کمی ای که برای مقایسه ی شاخه نگار ها. نتیجه چنین مقایسه ای، رسیدن به پاسخ این پرسش است که آیا شاخه نگارها که نشان دهنده موقعیت متریک شاخص های گوناگون بورس اوراق بهادار تهران هستند، در طول زمان تغییر کرده و بنابراین دلالت بر گونه ای ناپایداری در سیستم را دارند؟ یا برعکس شاخه نگارهای مذکور در طول زمان، پایدار بوده و تغییرات چشم گیری نداشته اند. برای پاسخ به این پرسش ها در ادامه کار، ابتدا مفهوم همشاخه نگار، شاخص گامای بیکر و ضریب همبستگی همسانی توضیح داده و در قسمت بعد، نتایج محاسبه ی آن در قلمرو زمانی این پژوهش تشریح می شود.

همشاخه نگارها که میزان شباهت دو یا چند شاخه نگار را نشان می دهند، یکی از روش های مقایسه ی خوشه بندی های سلسله مراتبی به شمار می روند. در برخی از پژوهش ها، همشاخه نگارهایی

که دو یا بیش از دو خوشه بندی را مقایسه می کند، معرفی و روش تدوین آن ها ارائه شده است [۳۲].

بیشتر کاربردهای این روش در آمار زیستی و مسائلی از قبیل شباهت و خوشه بندی مشاهدات زیست شناختی بوده است [۲۳، ۲۸]. به نظر می رسد تعمیم و به کارگیری آن در مطالعات مالی، فارغ از نوآوری این پژوهش، می تواند در آشکارسازی تغییرات سیستم مالی، سهم داشته باشد. همشاخه نگار، دو شاخه نگار را به عنوان ورودی می گیرد و آن ها را مقابل یکدیگر قرار می دهد و مشاهدات مشابه، را پیدا و با خط به یکدیگر متصل می کند. در صورتی که شاخه نگارهای ورودی عین هم باشند، هیچ گونه درهم تنیدگی اتفاق نمی افتد و خطوطی که مشاهدات مشابه را به یکدیگر وصل می کنند موازی اند. اما اگر شاخه نگار ها، متفاوت باشند، برخی از این خطوط یک دیگر را قطع می کنند. هر چه میزان این درهم تنیدگی خطوط، بیشتر باشد، شاخه نگارها، تفاوت بیشتری دارند. بر همین اساس میزان درهم تنیدگی محاسبه که مقدار صفر آن، یعنی شباهت کامل شاخه نگارهای ورودی (بدون هیچ خط متقاطع) و مقدار یک آن، یعنی عدم شباهت کامل آن ها (مقاطع بودن همه خطوط) است [۱۲]. در شکل، همشاخه نگار خوشه بندی شش ماهه نخست یا خودش (به عنوان مثالی از شباهت کامل با درهم تنیدگی صفر) و با شش ماهه دوم (به عنوان مثالی از شباهت جزئی با درهم تنیدگی ۷۷٪) نشان داده شده است.



نمودار(۶). تانگلگرام (همشاخه نگار) شش ماهه اول و دوم

بیکر در سال ۱۹۷۴، روشی برای مقایسه خوشه بندی های سلسله مراتبی ارائه کرد. بر اساس روش وی، دو مشاهده (در پژوهش حاضر دو شاخص) انتخاب می شود، تعیین می کنند در خوشه بندی اول، در چه صورت آن دو شاخص هم خوشه می شوند و تعداد خوشه ها در این حالت چند است؟ حالا همین دو شاخص را در خوشه بندی دوم کنترل می کنند و باز هم تعداد خوشه ها را شمارش می کنند. برای این زوج از شاخص ها، دو عدد (تعداد خوشه ها در حالت همخوشگی اول و تعداد خوشه ها در

حالت همخوشگی دوم) به دست می آید. برای همه ی زوج ها (همه انتخاب های دوتایی از شاخص ها) این دو عدد محاسبه می شود. حال همبستگی رتبه ای (اسپیرمن) این دو ستون عدد، محاسبه می شود که عددی بین -1 و $+1$ است و مقادیر نزدیک به صفر آن عدم شباهت آماری خوشه بندی های مورد نظر است. [۲]. روش دیگر مقایسه خوشه بندی های سلسله مراتبی، استفاده از مفهوم ناهمسانی مشاهداتی است که خوشه بندی شده اند. این فاصله، ناهمسانی خوشه های مختلف از یکدیگر را اندازه گیری می کند. در خوشه بندی، هدف اولیه، افزایش این ناهمسانی بین خوشه ای و کاهش ناهمسانی درون خوشه هاست. همبستگی ناهمسانی ها، عبارت است از ضریب همبستگی دو ماتریس؛ ماتریس هایی که هریک فاصله های ناهمسانی دو خوشه بندی را نشان می دهند. تفسیر همبستگی ناهمسانی ها، نیز مشابه شاخص بیکر است [۲۹]. شاخص فاولکس-مالوز نیز، شباهت دو خوشه بندی را نشان می دهد. این شاخص، عددی بین صفر تا ۱ را می گیرد و هر چه شباهت خوشه بندی ها بیشتر باشد، به ۱ نزدیک تر می شود [۱۱].

در ادامه این نوشتار، شاخص های انتخاب شده در قلمرو زمانی تحقیق به صورت ماهانه، خوشه بندی شده و سپس این خوشه بندی ها با یکدیگر مقایسه می شوند. با توجه به اینکه رسم 12×12 همشاخه نگار ممکن نیست، بنابراین در مقایسه آن ها، به شاخص بیکر و ضریب همبستگی ناهمسانی ها اکتفا شده است. در این پژوهش از بسته ی *dendextend* از نرم افزار *R* استفاده شده است [۱۲]. پس از آن که مشخص شدن میزان ناپایداری خوشه بندی ها، با استفاده از روش تجزیه مولفه های اصلی (PCA)، سری های زمانی بازده شاخص های منتخب را به صورت ترکیبی از عناصر پایدار

سیستم تحلیل کرده و خوشه بندی نهایی بر اساس مولفه های اصلی سیستم صورت می گیرد. در انجام تجزیه مولفه های اصلی، به صورت کلی مراحل زیر طی شده است:

الف. ابتدا بازده لگاریتمی شاخص های قلمرو مطالعه ، محاسبه و به صورت ماتریس $T \times n$ که در آن n تعداد شاخص ها و T تعداد بازده ها یا عمق سری زمانی است، نشان داده می شود. (ماتریس X).

ب. ماتریس V یا همان ماتریس همبستگی بازده ها یک ماتریس $n \times n$ است که از آن با محاسبه بردار ها و مقدار های ویژه ی آن می توان به ماتریسی $n \times n$ رسید که W یا ماتریس بارعاملی نامیده می شود.

ج. با استفاده از تبدیل $P = XW$ ماتریس بازده ها به ماتریس مولفه های اصلی تبدیل می شود که یک ماتریس $T \times n$ است.

د. در آخرین گام k مولفه اصلی که بیشترین بارعاملی را دارند و ماتریس بازده ها بر اساس آن ها، به این شکل تبدیل می شوند: $X \approx P^*W^{*'}.$ ماتریس X ، همان ماتریس بازده هایی است که مولفه های اصلی آن ها را ساخته اند و بنابراین مولفه های غیر اصلی که نوفه یا تصادفی اند از آن حذف شده اند [۹]. با اجرای این روش و رسیدن به این داده های پیراسته و انجام خوشه بندی بر اساس آن ها می توان به خوشه بندی سلسله مراتبی ای دست یافت که بر اساس تحلیل مولفه های اصلی،

شکل گرفته است و بنابراین، نشان دهنده موقعیت متریک ثبت شده ی آن ها نسبت به یکدیگر است.

جدول زیر مراحل انجام کار و توابع استفاده را نشان می دهد.

۴- تحلیل نتایج

پس از تقسیم بازه زمانی به صورت ماهانه، شاخه نگار مربوط به هر ماه، تدوین شد و پس از

آن، شاخص های بیکر، همبستگی همسانی ها و شاخص FM به صورت زوجی محاسبه و نتایج حاصله

نگاره سازی شد.

جدول (۱). توابع و بسته های استفاده شده

مرحله	شرح	تابع استفاده شده	بسته استفاده شده
۱	ورود داده ها شاخص ها به R	read.csv()	base
۲	تبدیل داده ها به سری زمانی	timeDate()	timeSeries
۳	محاسبه بازده لگاریتمی شاخص ها	returns	quantmod
۴	خوشه بندی سلسله مراتبی شاخص ها	hclust()	stats
۵	تدوین همشاخه نگار بازه های شش ماهه	tanglegram()	dendextend
۶	نگاره سازی شاخه نگارها	fviz_dend()	factextra
۷	محاسبه شاخص بیکر برای خوشه بندی های ماهانه	cor_bakers_gamma()	dendextend
۸	محاسبه شاخص همبستگی ناهمسانی ها	cor_cophenetic()	dendextend
۹	نگاره سازی نتایج	ggplot()	ggplot2
۱۰	تجزیه مولفه های اصلی	princom()	stats

جدول (۲)، مقادیر شاخص بیکر را نشان می دهد. قطر اصلی ماتریس، بیانگر شباهت هر خوشه بندی به خودش یعنی یک است اما همچنانکه مشاهده می شود، خوشه بندی ها در اکثر موارد به یکدیگر شبیه نیستند و مقدار شاخص بیکر در آن ها به صفر نزدیک است. روند قابل ملاحظه ای در شاخص بیکر و ضریب همبستگی همسانی ها وجود ندارد. این واقعیت در مقایسه با حالتی فرضی که در آن همبستگی ها، نزدیک به یک و بیانگر این حقیقت باشند که پایداری قابل ملاحظه ای در سیستم وجود دارد، قابل تفسیر است. از طرف دیگر، ضریب همبستگی همسانی ها، نیز در بازه ی مورد نظر رفتاری مشابه شاخص بیکر دارد.

جدول (۲). مقادیر شاخص بیکر قبل از PCA

	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	-0.042	1										
3	0.028	-0.013	1									
4	0.036	0.013	-0.036	1								
5	-0.054	-0.057	-0.013	-0.029	1							
6	0.045	-0.029	-0.071	-0.03	-0.031	1						
7	0.14	-0.009	-0.019	0.017	-0.031	-0.002	1					
8	-0.051	-0.02	-0.016	-0.001	-0.039	-0.043	0.051	1				
9	0.116	-0.055	-0.034	-0.031	0.002	0.206	0.033	-0.042	1			
10	0.054	0.006	-0.015	-0.061	0.107	-0.017	0.037	0.046	-0.01	1		
11	-0.052	0.01	-0.044	-0.055	0.036	0.018	-0.02	0.055	0.006	0.088	1	
12	-0.034	-0.031	0.019	-0.038	-0.041	0.009	-0.009	0.146	-0.029	-0.005	0.033	1

جدول (۳) نشان می دهد که بر اساس ضریب همبستگی همسانی ها، شباهت بین خوشه بندی ها در ماه های مختلف سال ناچیز و به ندرت، در برخی از مواقع این ضریب همبستگی همسانی ها، قابل ملاحظه به نظر می رسد.

به منظور درک روشن تری از پایداری بورس اوراق بهادار تهران، ماتریس های فوق را تبدیل به ماتریس هایی بالامثلی نموده، تا مقادیر تکراری حذف شوند. روشن است که هر دو ماتریس، متقارن اند و با در نظر گرفتن ماتریس بالا مثلثی یا پایین مثلثی، اطلاعات خاصی از بین نخواهد رفت. آمار توصیفی مقادیر باقی مانده و چگالی احتمال آن ها در ادامه مورد بررسی قرار می گیرد.

جدول (۳). شاخص همبستگی همسانی ها پیش از PCA

	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	-0.058	1										
3	0.034	-0.001	1									
4	0.056	-0.025	-0.018	1								
5	-0.071	-0.064	-0.058	-0.021	1							
6	0.013	-0.005	-0.093	-0.031	-0.007	1						
7	0.065	-0.013	-0.043	0.013	-0.023	0.015	1					
8	-0.062	-0.035	0.026	0.022	-0.006	-0.042	0.007	1				
9	0.078	-0.076	-0.016	-0.033	-0.006	0.122	0.04	-0.042	1			
10	0.04	0.04	-0.024	-0.046	0.045	0.019	0.043	0.021	-0.032	1		
11	-0.032	0.01	-0.017	-0.051	-0.005	0.032	-0.008	-0.001	-0.017	0.056	1	
12	-0.043	-0.039	0.065	-0.039	-0.04	0.018	-0.011	0.172	-0.016	0.008	0.042	1

جدول (۴). آمار توصیفی شاخص های بیکر و همبستگی همسانی

	mean	sd	median	min	max	range
شاخص بیکر	0.227	0.169	0.196	-0.010	0.579	0.590
شاخص همبستگی همسانی	0.287	0.135	0.250	0.052	0.623	0.570

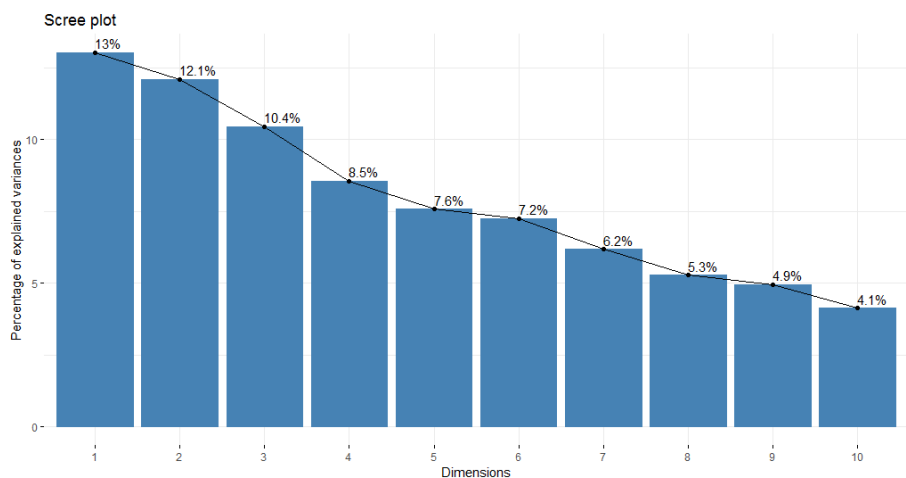
با این توصیف، میانگین شاخص بیکر و شاخص همبستگی همسانی، به ترتیب ۰,۲۲۷ و ۰,۲۸۷ است که هر دو گویای پایداری کم سیستم مورد مطالعه است. پایداری کم سیستم، نشان دهنده تغییر پذیری مداوم شاخه نگارها و خوشه بندی های سلسله مراتبی است. بنابراین نتیجه عملی، آن است که در هر زمان موقعیت متریک شاخص ها نسبت به یکدیگر تغییر می کند. با این وجود روش هایی وجود دارند که امکان اندازه گیری جزء پایدار سیستم را به دست می دهند. یکی از این روش ها، تجزیه مولفه های اصلی (PCA) است. در روش PCA، ماتریس سری های زمانی ای که مبنای خوشه بندی در یک سال بود، براساس مولفه های اصلی (حاصل از مقادیر ویژه و بردارهای ویژه ماتریس همبستگی آن ها) بازسازی می شوند. در این بازسازی براساس مبانی روش PCA که در قسمت روش تحقیق به اختصار شرح داده شد، تاثیر عوامل اصلی یا همان مولفه های با ثبات سیستم که بیشترین بار عاملی در شکل گیری نوسان کل سیستم را داشته اند در نظر گرفته می شود و از تاثیر عواملی که اهمیت کمتری دارند و به سبب ناپایداری این سیستم بوده اند، چشم پوشی می شود. در گام اول مقادیر ویژه ماتریس همبستگی بازده ها، محاسبه می شود. مقادیر ویژه ای بالای یک باشند، یک ملاک برای انتخاب تعداد ابعاد است. ملاک دیگر درصد تجمعی واریانس است. در پژوهش های مالی معمولاً تعداد ابعادی که ۷۰ درصد تغییرات را نشان دهد، به عنوان ملاک انتخاب تعداد ابعاد در نظر گرفته می شود. این تعداد، در پژوهش حاضر، ۸ بعد است.

جدول (۵). مقادیر ویژه

	مقادیر ویژه	درصد واریانس	درصد تجمعی واریانس
Dim.1	31.22892717	13.01205299	13.01205299
Dim.2	29.01820576	12.09091907	25.10297205
Dim.3	25.07628085	10.44845036	35.55142241
Dim.4	20.51604719	8.548352997	44.0997754
Dim.5	18.1528791	7.563699624	51.66347503
Dim.6	17.36242041	7.234341836	58.89781686
Dim.7	14.81127722	6.171365507	65.06918237
Dim.8	12.64148888	5.267287033	70.3364694
Dim.9	11.85150799	4.938128331	75.27459774
Dim.10	9.887668641	4.119861934	79.39445967
Dim.11	9.237442159	3.848934233	83.2433939
Dim.12	7.827616942	3.261507059	86.50490096

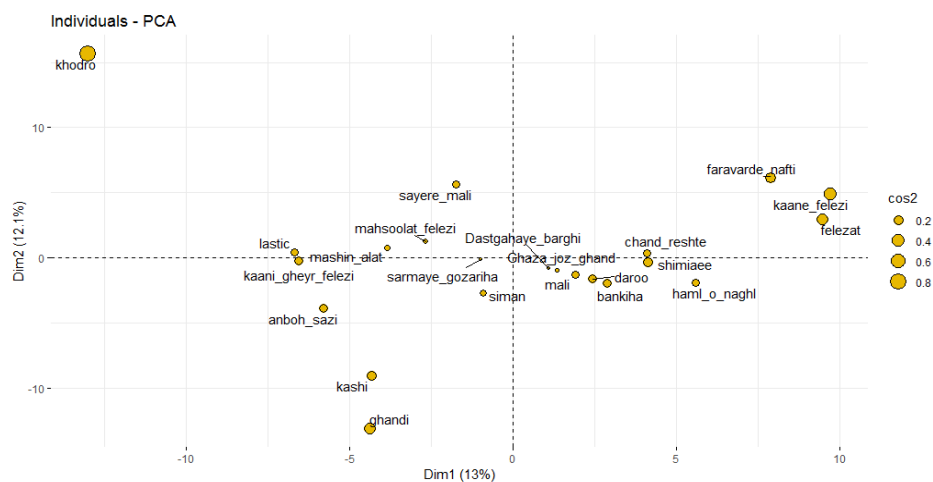
Dim.13	7.219492864	3.008122027	89.51302299
Dim.14	5.505109039	2.293795433	91.80681842
Dim.15	5.290905688	2.204544037	94.01136246
Dim.16	4.319508346	1.799795144	95.8111576
Dim.17	3.59349817	1.497290904	97.30844851
Dim.18	3.02944633	1.262269304	98.57071781
Dim.19	1.992446934	0.830186223	99.40090403
Dim.20	1.425661835	0.594025765	99.9949298
Dim.21	0.012168486	0.005070202	100

مقادیر ویژه و نقش آن ها در تشکیل مولفه های اصلی این پژوهش به صورت زیر است. همان گونه که مشخص است این نمودار در تناظر با مقادیر جدول (۵) است.



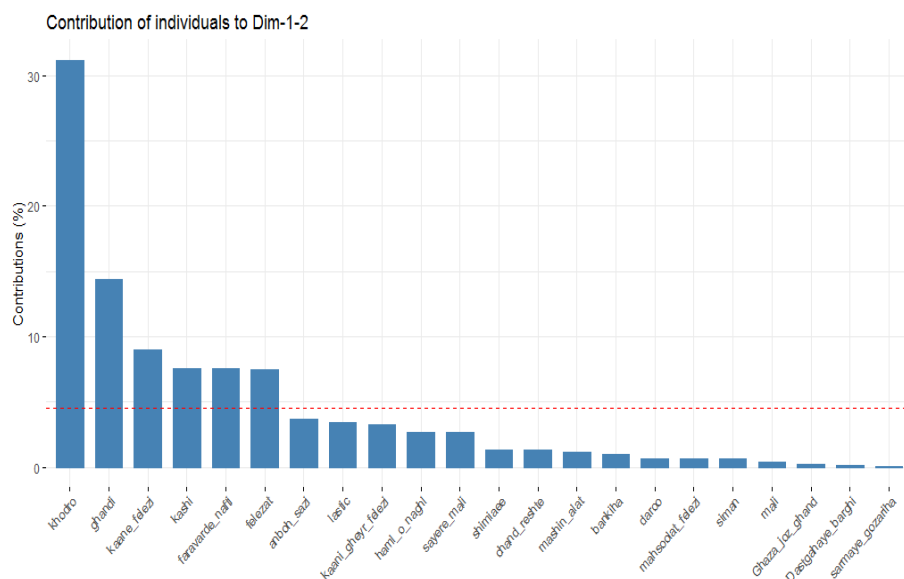
نمودار (۷). میزان اهمیت مولفه های ویژه

پس از آن که تعداد مولفه های اصلی مشخص شد، حال می توان ماتریس بارزده ها را به ماتریس بارزده های تعدیل شده بر مبنای PCA تبدیل کرد و خوشه بندی را بر اساس آن انجام داد. در نمودار (۸)، موقعیت هر یک از شاخص ها نسبت به دو عامل نخست که بیشترین اهمیت را دارند و زاویه ای که با مولفه دوم دارند، مشخص شده است.



نمودار(۸). موقعیت هر یک از شاخص ها نسبت به مولفه های اول و دوم

نمودار(۹) نقشی که هر یک از صنایع در دو مولفه ی اصلی (یکم و دوم) داشته اند را نشان می دهد. بر اساس نتایج به دست آمده، صناعی که بیشترین نقش را در مولفه های اول و دوم داشته اند و به تعبیری بیشترین تاثیر گذاری سیستمی را به خود اختصاص داده اند عبارت اند از: خودرو، قندی، کانی فلزی، کاشی، فراورده نفتی، فلزات و انبوه سازی.

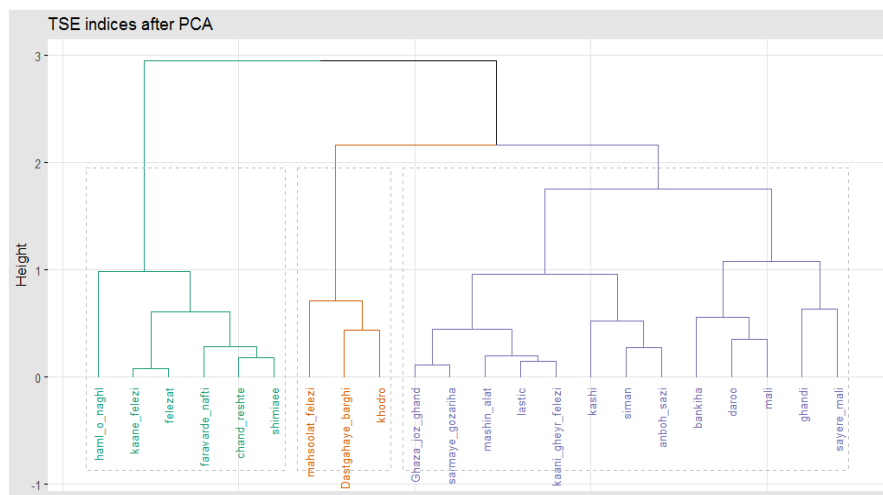


نمودار(۹). سهم هر یک از شاخص ها در مولفه های اول و دوم

پس از تولید سری های زمانی جدید بر مبنای PCA، تقسیم آن ها به صورت ماهانه، خوشه بندی سلسله مراتبی بر اساس آن ها صورت پذیرفت و شاخص بیکر برای آن ها مجددا محاسبه و در جدول گزارش شده است. همان گونه که اعداد جدول نشان می دهد، همبستگی تعدادی از خوشه بندی ها در ماه مختلف به مقادیر نزدیکی به عدد یک رسیده است و این بدان معناست که شاخه نگارهای متناظر با آن ها تقریبا مثل هم بوده یا از شباهت بالایی برخوردارند. در بسیاری از حالت ها، نیز همبستگی از مقادیری که نزدیک به صفر بودند به مقادیر فاصله دار از صفر رسیده اند. این به مفهوم آن است که سیستم پس از تحلیل PCA نسبت به حالت پیشین، پایدار تر است و بنابراین می توان نتیجه گرفت در صورتی که خوشه بندی سلسله مراتبی بر اساس PCA صورت پذیرد، نتایج حاصل از آن، از اتکاپذیری بالاتری برخوردار خواهد بود.

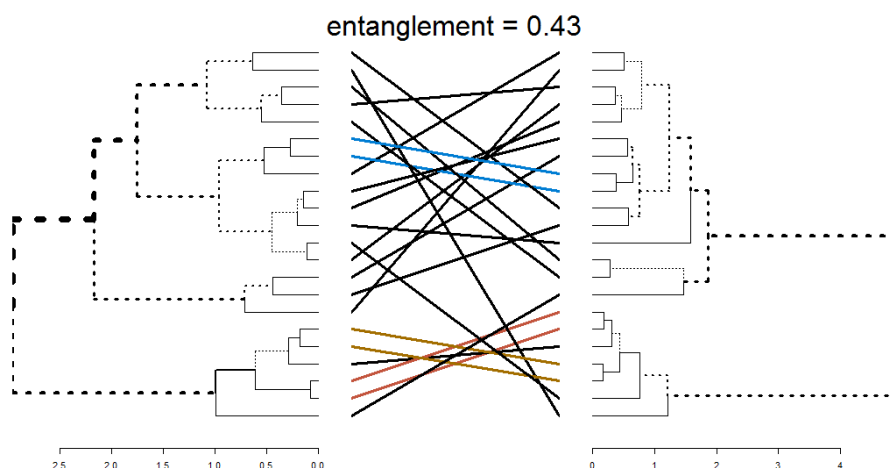
جدول (۶). شاخص بیکر بعد از PCA

	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	0.367	1										
3	0.522	0.559	1									
4	0.478	0.703	0.939	1								
5	0.501	0.663	0.844	0.899	1							
6	0.508	0.677	0.815	0.868	0.971	1						
7	0.548	0.676	0.863	0.903	0.981	0.983	1					
8	0.721	0.272	0.493	0.416	0.448	0.452	0.503	1				
9	0.35	0.239	0.328	0.399	0.519	0.481	0.464	0.328	1			
10	0.47	0.668	0.817	0.872	0.975	0.979	0.978	0.417	0.481	1		
11	0.388	0.3	0.322	0.376	0.446	0.553	0.474	0.375	0.341	0.463	1	
12	0.592	0.599	0.877	0.868	0.936	0.938	0.954	0.516	0.443	0.933	0.453	1



نمودار (۱۰). خوشه بندی سلسله مراتبی بعد از PCA

در پایان خوشه بندی پایداری که حاصل از تحلیل PCA است، ارائه می شود. در این خوشه بندی، گروه های سه گانه ای مشخص شده است که ترکیب آن با خوشه بندی اولیه متفاوت است. این تفاوت، به دلیل نوفه هایی است که سیستم را تحت تاثیر قرار می دهند و در اثر تحلیل PCA بخش زیادی از آن ها، حذف شده اند. در صورتی که این شاخه نگار و شاخه نگار قبل از PCA با یکدیگر مقایسه شوند، میزان شباهت آن ها نشان دهنده ثبات سیستم و عدم شباهت آن ها نشان دهنده ی ناپایداری آن است.



نمودار(۱۱). همشاخه نگار خوشه بندی های سلسله مراتبی بعد و قبل از PCA

میزان درهم تنیدگی این دو خوشه بندی ۴۳ درصد است، این عدد در صورتی که صفر باشد یعنی شباهت کامل خوشه ها و در صورتی که ۱۰۰ درصد باشد، یعنی عدم شباهت کامل خوشه بندی ها. بنابراین در صورتی که دو خوشه بندی (یعنی قبل از تحلیل PCA و پایداری سازی خوشه ها یا بعد از تحلیل PCA) مقایسه شود، این عدد می تواند کمیتی برای اندازه گیری پایداری سیستم باشد. هر چه این عدد به صفر نزدیک باشد، ناپایداری آن کمتر است و باثبات تر است.

۵- بحث و نتیجه گیری

از آن جا که اندازه گیری پایداری یک سیستم مالی، اولین گام در تخمین ریسک سیستمیک است، در این پژوهش با استفاده از خوشه بندی سلسله مراتبی شاخص های بورس اوراق بهادار و پایداری سازی آن با استفاده از تحلیل PCA، کمیت جدیدی برای اندازه گیری میزان پایداری سیستم معرفی شد. نتایج نشان داد که شاخه نگارهای شاخص ها در طول زمان متغیر اند و بنابراین موقعیت متریک شاخص ها

نسبت به یکدیگر مدام تغییر می کند. در صورتی که این تغییر را بتوان به مولفه های اصلی تجزیه نمود، می توان سهم مولفه های مهم و مولفه کمتر مهم را از یکدیگر تفکیک کرد. بنابراین پس از حذف تاثیر مولفه های کم اهمیت که در تحلیل PCA صورت می گیرد، می توان به خوشه بندی ای دست یافت که از ثبات بیشتری برخوردارند. البته این پایداری کامل نیست چرا که مولفه های اصلی انتخاب شده، خود دارای تغییراتی هستند. تغییرات مذکور را می توان تغییرات سیستمیک تلقی کرد. برای مقایسه خوشه بندی ها و تشخیص شباهت موجود بین آن ها از شاخص های گوناگونی چون بیکر و ضریب همبستگی همسانی ها، استفاده می شود. در این پژوهش، مشخص گردید، خوشه بندی ها، پس از تحلیل PCA از شباهت بیشتری برخوردارند و بنابراین پایداری بالاتری دارند. در نهایت خوشه بندی سلسله مراتبی حاصل از PCA با خوشه بندی اولیه مقایسه شد که میزان عدم پایداری (۴۳ درصد) سیستم در قلمرو مطالعه به دست آمد. این پژوهش برای محدوده زمانی یکساله صورت گرفته است و طبیعی است که با توسعه قلمرو پژوهش به چند سال و به ویژه ها سال هایی که تغییرات سیستمیک آشکاری در بازار بورس صورت گرفته است و نیز توسعه قلمرو آن به شاخص های صنایع گوناگون و بازارهای دیگر مالی می توان نتایج حاصل از آن را تعمیم داد. به عنوان نکته آخر، به سرمایه گذاران، تحلیلگران و سبدگردانان و نیز مدیران ریسک توصیه می شوند، در استفاده از تکنیک های خوشه بندی به عنوان ابزاری در متنوع سازی یا مدیریت ریسک به این مهم توجه داشته باشند در زمان های مختلف موقعیت متریک مشاهداتی که خوشه بندی می شوند بر اساس میزان ناپایداری سیستم تغییر می کند و بنابراین اتکاء به آن ها ممکن است قابل تعمیم به زمان های دیگر نباشد. مگر آنکه با روش هایی از جمله روش هایی که در این نوشتار به آن اشاره شد، خوشه بندی ها را پایدار و از شکل پایداری شده ی آن ها برای تصمیم گیری های مالی استفاده شود. در تحقیقات آتی، پژوهشگران می توانند سایر روش های پایداری سازی را مدنظر قرار دهند و شکاف پژوهشی موجود را از طریق بسط قلمروهای این پژوهش یا اصلاح و تکمیل روش های آن، ترمیم کنند.

۶- منابع

1. Araújo, Tanya, Eleutério, Samuel, & Louçã, Francisco. (2018). Do sentiments influence market dynamics? A reconstruction of the Brazilian stock market and its mood. *Physica A: Statistical Mechanics and its Applications*, 505, 1139-1149. doi: <https://doi.org/10.1016/j.physa.2018.04.045>
2. Baker, Frank B. (1974). Stability of Two Hierarchical Grouping Techniques Case I: Sensitivity to Data Errors. *Journal of the American Statistical Association*, 69(346), 440-445. doi: 10.1080/01621459.1974.10482971

3. Brida, J. Gabriel, & Risso, W. Adrián. (2010). Hierarchical structure of the German stock market. *Expert Systems with Applications*, 37(5), 3846-3852. doi: <https://doi.org/10.1016/j.eswa.2009.11.034>
4. Cai, Yumei, Cui, Xiaomei, Huang, Qianyun, & Sun, Jianqiang. (2017). Hierarchy, cluster, and time-stable information structure of correlations between international financial markets. *International Review of Economics & Finance*, 51, 562-573. doi: <https://doi.org/10.1016/j.iref.2017.07.024>
5. de Amorim, Renato Cordeiro. (2015). Feature relevance in ward's hierarchical clustering using the L p norm. *Journal of Classification*, 32(1), 46-62.
6. Deza, Michel-Marie, & Deza, Elena. (2006). *Dictionary of distances*: Elsevier.
7. Deza, Michel Marie, & Deza, Elena. (2009). Encyclopedia of distances *Encyclopedia of Distances* (pp. 1-583): Springer.
8. Esmalifalak, Hamidreza, Ajirlou, Ali Irannezhad, Behrouz, Sahar Pordeli, & Esmalifalak, Maryam. (2015). (Dis)integration levels across global stock markets: A multidimensional scaling and cluster analysis. *Expert Systems with Applications*, 42(22), 8393-8402.
9. Everitt, Brian S, & Dunn, Graham. (2001). Principal components analysis. *Applied multivariate data analysis*, 48-73.
10. Fang, Libing, Xiao, Binqing, Yu, Honghai, & You, Qixing. (2018). A stable systemic risk ranking in China's banking sector: Based on principal component analysis. *Physica A: Statistical Mechanics and its Applications*, 492, 1997-2009. doi: <https://doi.org/10.1016/j.physa.2017.11.115>
11. Fowlkes, E. B., & Mallows, C. L. (1983). A Method for Comparing Two Hierarchical Clusterings. *Journal of the American Statistical Association*, 78(383), 553-569. doi: 10.1080/01621459.1983.10478008
12. Galili, Tal. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22), 3718-3720. doi: 10.1093/bioinformatics/btv428
13. Goldberger, Jacob, & Tassa, Tamir. (2008). A hierarchical clustering algorithm based on the Hungarian method. *Pattern Recognition Letters*, 29(11), 1632-1638. doi: <https://doi.org/10.1016/j.patrec.2008.04.003>
14. Huang, Wei-Qiang, & Wang, Dan. (2018). Systemic importance analysis of chinese financial institutions based on volatility spillover network. *Chaos, Solitons & Fractals*, 114, 19-30.
15. Huang, Wei-Qiang, Yao, Shuang, Zhuang, Xin-Tian, & Yuan, Ying. (2017). Dynamic asset trees in the US stock market: Structure variation and market phenomena. *Chaos, Solitons & Fractals*, 94, 44-53.
16. Iori, Giulia, & Mantegna, Rosario N. (2018). Chapter 11 - Empirical Analyses of Networks in Finance. In C. Hommes & B. LeBaron (Eds.), *Handbook of Computational Economics* (Vol. 4, pp. 637-685): Elsevier.
17. Jian, Zhihong, Deng, Pingjun, & Zhu, Zhican. (2018). High-dimensional covariance forecasting based on principal component analysis of high-frequency data. *Economic Modelling*, 75, 422-431.

18. Lahmiri, Salim. (2016). Clustering of Casablanca stock market based on hurst exponent estimates. *Physica A: Statistical Mechanics and its Applications*, 456, 310-318. doi: <https://doi.org/10.1016/j.physa.2016.03.069>
19. Lee, Byung-Joo. (2018). Asian financial market integration and the role of Chinese financial market. *International Review of Economics & Finance*. doi: <https://doi.org/10.1016/j.iref.2018.10.012>
20. Li, Chun-Zhong, Xu, Zong-Ben, & Luo, Tao. (2013). A heuristic hierarchical clustering based on multiple similarity measurements. *Pattern Recognition Letters*, 34(2), 155-162. doi: <https://doi.org/10.1016/j.patrec.2012.09.025>
21. Li, Wenwei, Hommel, Ulrich, & Paterlini, Sandra. (2018). Network topology and systemic risk: Evidence from the Euro Stoxx market. *Finance Research Letters*. doi: <https://doi.org/10.1016/j.frl.2018.02.016>
22. Lu, Ya-Nan, Li, Sai-Ping, Zhong, Li-Xin, Jiang, Xiong-Fei, & Ren, Fei. (2018). A clustering-based portfolio strategy incorporating momentum effect and market trend prediction. *Chaos, Solitons & Fractals*, 117, 1-15.
23. Matsen, F. A., Billey, S. C., Kas, A., & Konvalinka, M. (2018). Tanglegrams: A Reduction Tool for Mathematical Phylogenetics. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(1), 343-349. doi: 10.1109/TCBB.2016.2613040
24. Mondal, Sakib A. (2018). An improved approximation algorithm for hierarchical clustering. *Pattern Recognition Letters*, 104, 23-28.
25. Nobi, Ashadun, & Lee, Jae Woo. (2016). State and group dynamics of world stock market by principal component analysis. *Physica A: Statistical Mechanics and its Applications*, 450, 85-94.
26. Rea, Alethea, & Rea, William. (2014). Visualization of a stock market correlation matrix. *Physica A: Statistical Mechanics and its Applications*, 400, 109-123. doi: <https://doi.org/10.1016/j.physa.2014.01.017>
27. Rodríguez-Moreno, María, & Peña, Juan Ignacio. (2013). Systemic risk measures: The simpler the better? *Journal of Banking & Finance*, 37(6), 1817-1831. doi: <https://doi.org/10.1016/j.jbankfin.2012.07.010>
28. Scornavacca, Celine, Zickmann, Franziska, & Huson, Daniel H. (2011). Tanglegrams for rooted phylogenetic trees and networks. *Bioinformatics*, 27(13), i248-i256. doi: 10.1093/bioinformatics/btr210
29. Sokal, Robert R, & Rohlf, F James. (1962). The comparison of dendrograms by objective methods. *Taxon*, 33-40.
30. Tibshirani, Robert, Walther, Guenther, & Hastie, Trevor. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423.
31. Tumminello, Michele, Lillo, Fabrizio, & Mantegna, Rosario N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1), 40-58.
32. Venkatachalam, B., Apple, J., John, K. St., & Gusfield, D. (2010). Untangling Tanglegrams: Comparing Trees by Their Drawings. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 7(4), 588-597.

33. Vinothkumar, K., & Selvan, M. P. (2014). Hierarchical Agglomerative Clustering Algorithm method for distributed generation planning. *International Journal of Electrical Power & Energy Systems*, 56, 259-269. doi: <https://doi.org/10.1016/j.ijepes.2013.11.021>
34. Wang, Yanli, Li, Huajiao, Guan, Jianhe, & Liu, Nairong. (2019). Similarities between stock price correlation networks and co-main product networks: Threshold scenarios. *Physica A: Statistical Mechanics and its Applications*, 516, 66-77. doi: <https://doi.org/10.1016/j.physa.2018.09.154>
35. Zhao, Longfeng, Li, Wei, & Cai, Xu. (2016). Structure and dynamics of stock market in times of crisis. *Physics Letters A*, 380(5), 654-666.
36. Zhao, Longfeng, Wang, Gang-Jin, Wang, Mingang, Bao, Weiqi, Li, Wei, & Stanley, H. Eugene. (2018). Stock market as temporal network. *Physica A: Statistical Mechanics and its Applications*, 506, 1104-1112.

Stabilization of Financial Dendrograms as a Method of Systemic Changes Measurement (A Study of TSE's Indices)

Abstract

Dendrograms are considered as the most important visualization technique of Hierarchical Clustering. We studied the stability of different dendrograms on the Tehran Stock Exchange and then stabilized the system of TSE indices using Principal Component Analysis (PCA). It seems that the comparison of dendrograms before and after PCA is a quantitative measure of system stability. (43% in our study of 21 different indices of TSE which are the representations of the economic sectors of Iranian financial systems. We measured the similarity of different dendrograms according to Baker's Index and the Cophenetic Correlation Index. The results show that stabilized hierarchical clusterings have got better Baker's Index and more reliable dendrogram. Therefore, It is highly recommended to investors, portfolio managers and risk hedgers to denoise and stabilize their clustering with efficient methods like PCA before financial decision making.

Keywords: Hierarchical Clustering, Dendrograms, Tanglegrams, Principal Component Analysis (PCA).

