

دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی صنایع

استفاده از داده کاوی در انتخاب برخط سبد سرمایه گذاری با در نظر گرفتن
اطلاعات جانبی

مهدی خزائی

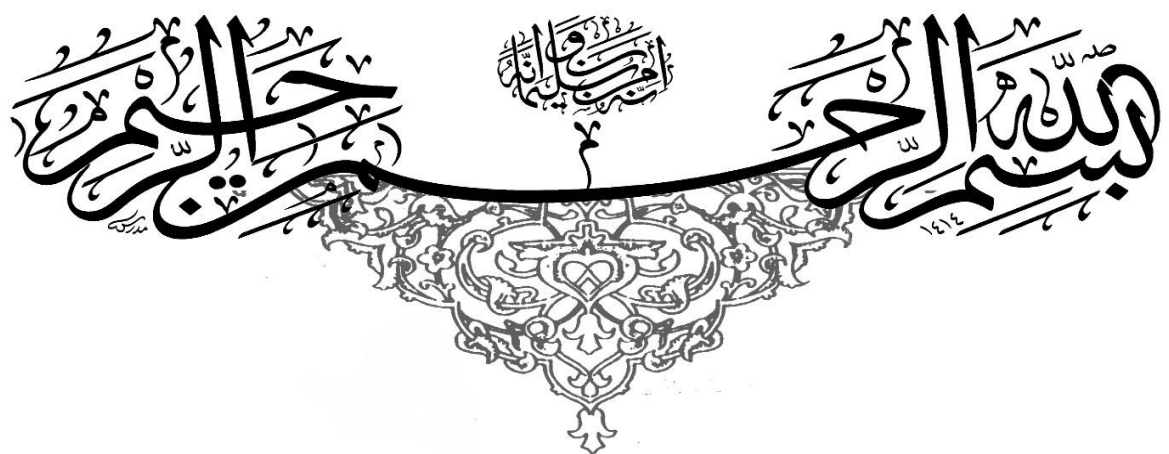
استاد راهنما:

دکتر امیر عباس نجفی

پایان نامه کارشناسی ارشد

رشته مهندسی صنایع گرایش مهندسی مالی

بهمن ۱۴۰۰



پدر بزرگوار و مادر عزیزم

ماحصل آموخته‌هایم را تقدیم می‌کنم به شما که مهر آسمانی‌تان آرام بخش آلام زمینی ام است

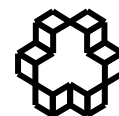
تقدیم به استوارترین تکیه‌گاهم، دستان پر مهر پدرم

و تقدیم به زیباترین نگاه زندگیم، چشمان نازنین مادرم

امروز هستی ام به امید شماست و فردا کلید باغ بهشتم رضای شماست

با سپاس فراوان از

استاد بزرگوار و فرزانه، جناب آقای دکتر امیرعباس نجفی که مراد این مسیر راهنمایی
فرمودند و با تشکر از دوست بسیار خوبم آقای فرهنگ رایانی که در این مسیر همراهی ام نمودند.



تأسیس ۱۳۰۷
دانشگاه صنعتی خواجه نصیرالدین طوسی

تأییدیه هیأت داوران

شماره:

تاریخ:

هیأت داوران پس از مطالعه پایان نامه و شرکت در جلسه دفاع تهیه شده تحت عنوان:
استفاده از داده کاوی در انتخاب برخط سبد سرمایه گذاری با در نظر گرفتن اطلاعات جانبی
توسط آقای/ خانم مهدی خزائی، صحت و کفایت تحقیق انجام شده را برای اخذ درجه کارشناسی ارشد در
رشته مهندسی صنایع گرایش مهندسی مالی در تاریخ ۱۴۰۰/۱۱/۲۷ مورد تأیید قرار می دهند.

امضاء

جناب آقای دکتر امیرعباس نجفی

۱- استاد راهنما

امضاء

جناب آقای/ سرکار خانم

۲- استاد مشاور

امضاء

جناب آقای دکتر حسین محسنی

۳- ممتحن داخلی

امضاء

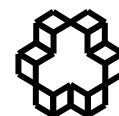
جناب آقای دکتر حمیدرضا آرین

۴- ممتحن خارجی

امضاء

جناب آقای دکتر حامد سلمان زاده

۵- معاونت آموزشی و
تحصیلات تکمیلی
دانشکده



تأسیس ۱۳۰۷
دانشگاه صنعتی خواجه نصیرالدین طوسی

اظہارنامہ دانشجو

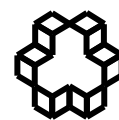
شماره:

تاریخ:

اینجانب مهدی خزائی دانشجوی کارشناسی ارشد رشته مهندسی صنایع گرایش مهندسی مالی دانشکده مهندسی صنایع دانشگاه صنعتی خواجه نصیرالدین طوسی گواهی می‌نمایم که تحقیقات ارائه شده در پایان نامه با عنوان استفاده از داده‌کاوی در انتخاب برخط سبدسرمایه گذاری با در نظر گرفتن اطلاعات جانبی با راهنمایی استاد محترم جناب آقای دکتر امیر عباس نجفی، توسط شخص اینجانب انجام شده و صحت و اصالت مطالب نگارش شده در این پایان نامه مورد تأیید می‌باشد، و در مورد استفاده از کار دیگر محققان به مرجع مورد استفاده اشاره شده است. بعلاوه گواهی می‌نمایم که مطالب مندرج در پایان نامه تاکنون برای دریافت هیچ نوع مدرک یا امتیازی توسط اینجانب یا فرد دیگری در هیچ جا ارائه نشده است و در تدوین متن پایان نامه چارچوب (فرمت) مصوب دانشکده مهندسی صنایع را بطور کامل رعایت کرده‌ام. چنانچه در هر زمان خلاف آنچه گواهی نموده‌ام مشاهده گردد خود را از آثار حقیقی و حقوقی ناشی از دریافت مدرک کارشناسی ارشد محروم می‌دانم و هیچگونه ادعائی نخواهم داشت.

امضاء دانشجو:

تاریخ:



تأسیس ۱۳۰۷

دانشگاه صنعتی خواجه نصیرالدین طوسی

حق طبع و نشر و مالکیت

شماره:

تاریخ:

۱- حق چاپ و تکثیر این پایان نامه متعلق به نویسنده و استاد راهنمای آن می باشد. هر گونه کپی برداری بصورت کل پایان نامه یا بخشی از آن تنها با موافقت نویسنده یا استاد راهنما یا کتابخانه دانشکده مهندسی صنایع دانشگاه صنعتی خواجه نصیرالدین طوسی مجاز می باشد. ضمناً متن این صفحه نیز باید در نسخه تکثیر شده وجود داشته باشد.

۲- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی خواجه نصیرالدین طوسی می باشد و بدون اجازه کتبی دانشگاه به شخص ثالث قابل واگذاری نیست. همچنین استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مراجع مجاز نمی باشد.

* توجه: این فرم می بایست پس از تکمیل، در نسخ تکثیر شده قرار داده شود.

چکیده

امروزه در بازارهای مالی، با توجه به حجم و سرعت بالای معاملات، نیاز به افزایش سرعت در تحلیل‌ها و تصمیم‌گیری‌ها غیرقابل چشم‌پوشی می‌باشد. یکی از شیوه‌های نوین که مبتنی بر رایانه می‌باشد، معاملات الگوریتمی می‌باشد و از جمله تکنیک‌های معاملات الگوریتمی می‌توان به انتخاب برخط سبد سرمایه‌گذاری اشاره کرد که می‌تواند حجم زیادی از داده‌های مربوط به قیمت و بازده سهام را به عنوان ورودی دریافت کرده، بر روی آنها عملیات پردازش انجام داده و سرمایه را به تعدادی معینی از سهام اختصاص دهد که در هر دوره معاملاتی سبد سهام را بر طبق الگوریتم از پیش تعیین شده باز توازن کند. این الگوریتم‌ها بدون توجه به ریسک صرفاً به دنبال بیشینه‌سازی سود می‌باشند و این عدم توجه به ریسک از جمله نقاط ضعف موجود در ادبیات انتخاب برخط سبد سهام محسوب می‌شود. این پژوهش دو الگوریتم انتخاب سبد سرمایه‌گذاری برخط بر اساس اصل تطابق با الگو ارائه می‌دهد. این دو روش با الهام از الگوریتم HRP و استفاده از مدل‌ها IVP و برابری خطر و اطلاعات جانبی سعی در کنترل ریسک انتخاب سبد سرمایه‌گذاری برخط را دارند. روش پیشنهادی شامل سه مرحله کلی است: (۱) اطلاعات جانبی (۲) الگوریتم HRP (۳) انتخاب برخط سبد سرمایه‌گذاری. در مرحله اول زیر مجموعه‌ای از سهام بر اساس اطلاعات جانبی از مجموعه سهام موجود انتخاب می‌شوند. در مرحله دوم نمادهای انتخاب شده در مرحله اول توسط الگوریتم HRP خوشه‌بندی شده و به هر کدام از خوشه‌ها بر اساس ریسک آن‌ها یک وزن تخصیص داده می‌شود. در مرحله سوم برای هر کدام از خوشه‌های مرحله دوم، ابتدا با استفاده از الگوریتم‌های خوشه‌بندی شامل k-means، k-medoids، خوشه‌بندی طیفی و سلسله مراتبی پنجره‌های زمانی مشابه پنجره زمانی اخیر کشف شده و انتخاب نمونه صورت می‌پذیرد. پس از یافتن پنجره‌های زمانی مشابه و پیش‌بینی رفتار بازار در روز بعد، از تابع بهینه‌سازی به همراه هزینه معاملاتی برای تشکیل پرتفو استفاده می‌شود. نتایج نشان می‌دهد که الگوریتم‌های ارائه شده در این تحقیق، عملکرد بهتری نسبت به الگوریتم‌های ارائه شده در ادبیات این حوزه دارند.

واژگان کلیدی: HRP، انتخاب برخط سبد سرمایه‌گذاری، تطابق با الگو، معاملات الگوریتمی، اطلاعات جانبی

فهرست مطالب

فصل اول: مقدمه و کلیات تحقیق	۱
۱-۱- مقدمه	۲
۱-۲- هدف از پژوهش	۳
۱-۳- توضیح موضوع پژوهش	۳
۱-۴- توجیه، انگیزه و علت انتخاب موضوع	۵
۱-۵- اهمیت موضوع	۵
۱-۶- مرور کلی بر ادبیات موضوع	۶
۱-۷- جنبه‌های جدید بودن موضوع	۸
۱-۸- کاربردهای موضوع پژوهش	۸
۱-۹- کاربران نتایج پژوهش	۸
۱-۱۰- جمع‌بندی	۹
فصل دوم: مرور ادبیات و پیشینه تحقیق	۱۰
۱-۲- مقدمه	۱۱
۲-۲- داده‌کاوی	۱۲
۲-۲-۱- روش‌های آماری	۱۳
۲-۲-۲- خوشه بندی	۱۳
۲-۲-۳- درخت تصمیم	۱۳
۲-۲-۴- قوانین انجمنی	۱۴
۲-۲-۵- شبکه‌های عصبی	۱۴
۲-۳- انتخاب بر خط سبد سرمایه‌گذاری	۱۴
۲-۳-۱- الگو برداری	۱۹

- ۱۹..... ۲-۳-۱-۱- استراتژی متوازن سازی ثابت پرتفو
- ۲۱..... ۲-۳-۱-۲- استراتژی خرید و نگهداری
- ۲۲..... ۲-۳-۲- الگوریتم های تبعیت از برنده
- ۲۲..... ۲-۳-۲-۱- الگوریتم تبعیت از رهبر
- ۲۴..... ۲-۳-۲-۲- الگوریتم پرتفوی یونیورسال
- ۲۶..... ۲-۳-۳-۲- الگوریتم گرادیان نمایی
- ۲۸..... ۲-۳-۴- الگوریتم برخط گام نیوتن
- ۳۰..... ۲-۳-۳- الگوریتم های تبعیت از بازنده
- ۳۱..... ۲-۳-۱- الگوریتم ضد همبستگی
- ۳۲..... ۲-۳-۲- الگوریتم بازگشت به میانگین منفعلانه-تهاجمی
- ۳۵..... ۲-۳-۳- الگوریتم ترکیبی منفعلانه-تهاجمی
- ۳۵..... ۲-۳-۴- الگوریتم بازگشت به میانگین اطمینان وزنی
- ۳۸..... ۲-۳-۵- الگوریتم بازگشت به میانگین متوسط متحرک وزن دار
- ۳۸..... ۲-۳-۶- الگوریتم بازگشت به میانگین متحرک برخط
- ۳۹..... ۲-۳-۷- الگوریتم بازگشت به میانه استوار
- ۴۱..... ۲-۳-۴- الگوریتم تطابق با الگو
- ۴۶..... ۲-۳-۱-۴- تکنیک های بهینه سازی پرتفو
- ۴۸..... ۲-۳-۵- الگوریتم های فرا یادگیری
- ۴۹..... ۲-۳-۱-۵- الگوریتم های جمع کننده
- ۵۰..... ۲-۳-۲-۵- یونیورسال شدن سریع
- ۵۱..... ۲-۳-۳-۵- الگوریتم گرادیان برخط
- ۵۲..... ۲-۳-۴-۵- الگوریتم تبعیت از تاریخ راهنما
- ۵۲..... ۲-۳-۵-۵- الگوریتم گرادیان برخط
- ۵۲..... ۲-۳-۶- دیگر الگوریتم ها

۵۳.....	۶-۲- الگوریتم HRP
۵۵.....	فصل سوم: استفاده از اطلاعات جانبی در انتخاب برخط سبد سرمایه‌گذاری با روش تطابق با الگو
۵۶.....	۳-۱- مقدمه
۵۸.....	۳-۲-۱- نفرین مارکویتز
۵۹.....	۳-۲-۲- مدل‌های درختی
۶۱.....	۳-۲-۳- گام‌های الگوریتم HRP
۶۲.....	۳-۲-۳-۱- خوشه بندی درختی
۶۳.....	۳-۲-۳-۲- شبه قطری سازی ماتریس
۶۴.....	۳-۲-۳-۳- تخصیص دوبخشی بازگشتی
۶۵.....	۳-۳- اطلاعات جانبی
۶۵.....	۴-۳- وزن دهی بر اساس معکوس واریانس
۶۶.....	۵-۳- پرتفوی برابری ریسک
۶۷.....	۶-۳- خوشه بندی
۶۷.....	۳-۶-۱- خوشه بندی طیفی
۶۸.....	۳-۶-۲- خوشه بندی سلسله مراتبی
۶۹.....	۳-۶-۲-۱- محاسبه ماتریس linkage
۷۰.....	۳-۶-۳- خوشه بندی به روش k-means
۷۲.....	۳-۶-۴- خوشه بندی به روش k-medoids
۷۳.....	۷-۳- انتخاب برخط سبد سرمایه‌گذاری به روش تطابق با الگو با کمک اطلاعات جانبی
۸۸.....	۶-۳- جمع بندی
۸۹.....	فصل چهارم: پیاده‌سازی و نتایج پژوهش
۹۰.....	۴-۱- مقدمه
۹۰.....	۴-۲- داده‌های ورودی پژوهش

۳-۴- معیارهای ارزیابی مدل‌ها	۹۱
۴-۴- نتایج مدل‌های پیاده‌سازی شده	۹۳
۴-۵- مقایسه نتایج کلی	۹۹
۴-۶- تاپسیس	۱۰۰
۴-۷- جمع‌بندی	۱۰۴
فصل پنجم: نتیجه‌گیری و پیشنهادات برای تحقیقات آتی	۱۰۶
۵-۱- مقدمه	۱۰۷
۵-۲- خلاصه‌ای از پژوهش و جمع‌بندی	۱۰۷
۵-۳- پیشنهادات برای تحقیقات آتی	۱۰۸
پیوست‌ها	۱۱۰
فهرست منابع	۱۱۱

فهرست جداول

جدول ۱-۲- فرم ریاضی مسئله CRP با M دارایی	۲۰
جدول ۲-۱- خلاصه‌ای از ادبیات موجود در زمینه الگوریتم‌های تبعیت از بازنده	۴۱
جدول ۲-۲ انتخاب نمونه $C(x_1^t, \omega)$	۴۳
جدول ۲-۳ خلاصه‌ای از ادبیات موجود در زمینه الگوریتم‌های تطابق با الگو	۴۸
جدول ۴-۱ معرفی معیارهای ارزیابی معیارهای ارزیابی	۹۲
جدول ۴-۲ الگوریتم‌های BNN	۹۳
جدول ۴-۳ الگوریتم CORN	۹۴
جدول ۴-۴ الگوریتم KMNLOG	۹۵
جدول ۴-۵ الگوریتم HRCLOG	۹۶
جدول ۴-۶ الگوریتم SPCLOG	۹۷
جدول ۴-۷ الگوریتم KMDLOG	۹۸
جدول ۴-۸ رتبه‌بندی الگوریتم‌های معرفی شده بر اساس روش تاپسیس	۱۰۱
جدول ۴-۹ مقایسه الگوریتم‌های معرفی شده بر اساس تمامی معیارها	۱۰۲

فهرست نمودار

نمودار ۱-۳- مقادیر ویژه ماتریس همبستگی برای پرتفوی های مختلف، دپرادو (۲۰۱۸).....	۵۸
نمودار ۲-۳- دندوگرام نمادهای پرتفو	۷۰
نمودار ۱-۴- بازدهی الگوریتم های BNN	۹۳
نمودار ۲-۴- بازدهی الگوریتم CORN	۹۴
نمودار ۳-۴- بازدهی الگوریتم KMNLOG	۹۶
نمودار ۴-۴- بازدهی الگوریتم HRCLOG	۹۷
نمودار ۵-۴- بازدهی الگوریتم SPCLOG	۹۷
نمودار ۶-۴- بازدهی الگوریتم KMDLOG	۹۸
نمودار ۷-۴- بازدهی تمامی الگوریتم ها	۹۹

فهرست اشکال

- شکل ۳-۱- نمونه ای از گراف کامل برای ۵۰ دارایی، دپرادو (۲۰۱۸) ۶۰
- شکل ۳-۲- ساختار درختی برای دارایی‌ها، دپرادو (۲۰۱۸) ۶۱
- شکل ۳-۳- نقشه حرارت ماتریس همبستگی و ماتریس همبستگی شبه قطری ۶۳
- شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه‌گذاری به روش تطابق با الگو با کمک اطلاعات جانبی ۷۹

فصل اول

مقدمه و کلیات تحقیق

۱-۱- مقدمه

در این پژوهش الگوریتم‌هایی در زمینه انتخاب برخط سبد سرمایه‌گذاری ارائه می‌شود که با استفاده از اطلاعات جانبی و با الهام از الگوریتم HRP^1 و استفاده از مدل‌های IVP^2 و مدل برابری خطر^۳ سعی در کنترل ریسک آن‌ها دارد که در ادامه به تشریح و تحلیل الگوریتم‌های مذکور پرداخته می‌شود.

در این فصل، اهداف و کلیات پژوهش ذکر خواهد شد و در ادامه به توجیه انتخاب موضوع انتخاب برخط سبد سرمایه‌گذاری و کاربردهای آن خواهیم پرداخت.

¹ Hierarchical Risk Parity

² Inverse-variance portfolio

³ Risk Parity

۱-۲- هدف از پژوهش

همانطور که می‌دانیم، در بازارهای مالی سرمایه‌گذاران به دنبال افزایش ثروت خود از طریق سرمایه‌گذاری می‌باشند. این سرمایه‌گذاران اعم از صندوق‌های مشترک، صندوق‌های شاخصی و سرمایه‌گذاران خرد، برای کنترل ریسک و بازده خود طبق نظریه مدرن پورتفو^۴ که توسط هری مارکوویتز^۵ (۱۹۵۲) مطرح شد، اقدام به تشکیل سبد سرمایه‌گذاری می‌کنند که این مسئله را مسئله انتخاب سبد سرمایه‌گذاری^۶ می‌نامند. سرمایه‌گذاران معمولاً دارایی‌های ارزشمند را از طریق تحلیل‌های تکنیکی^۷ و بنیادی^۸ با استفاده از ابزارهای متنوع تجزیه و تحلیل و شناسایی می‌کنند اما دلایلی مانند، نیاز به سرعت بالا در معاملات، به حداقل رساندن خطای انسانی در تصمیم‌گیری، تنوع بالای ابزارهای مالی و غیره سبب می‌شود که استفاده دستی از روش‌های مذکور خواسته‌های سرمایه‌گذاران را برآورده نسازد. به همین دلیل تکنیک‌های معاملات الگوریتمی^۹ مطرح می‌شوند که با استفاده از تکنولوژی سعی در برآورده کردن نیاز سرمایه‌گذاران در امر سرمایه‌گذاری می‌نمایند. انتخاب برخط سبد^{۱۰} از جمله این الگوریتم‌ها می‌باشد که در کنار الگوریتم‌های کنترل ریسک و اطلاعات جانبی می‌تواند ما را به این هدف نزدیک کند.

۱-۳- توضیح موضوع پژوهش

سرمایه‌گذاران در بازارهای مالی دستیابی به سود از طریق کشف قیمت ذاتی سهام و مقایسه آن با قیمت بازاری آن دارایی و یا کشف نقاط تغییر قیمت و کمترین و بیشترین قیمت سهم در گذشته می‌باشند. از جمله

^۴ Modern Portfolio Theory

^۵ Markowitz

^۶ Portfolio Selection Problem (PSP)

^۷ Technical

^۸ Fundamental

^۹ Algorithmic Trading Techniques

^{۱۰} Online Portfolio Selection (OLPS)

مهم‌ترین روش‌های مورد استفاده در این تحلیل‌های دستی تحلیل‌های بنیادی و تکنیکی می‌باشد. در تحلیل بنیادی، تحلیل‌گر بر اساس صورت‌های مالی و با محاسبه نسبت‌های مالی، ارزش ذاتی سهم را محاسبه کرده و آن را با ارزش بازاری^{۱۱} سهم مقایسه می‌کند و بر اساس نتایج به‌دست‌آمده در خصوص خرید یا فروش سهم تصمیم‌گیری می‌کند. در تحلیل تکنیکی، به‌جای محاسبه ارزش ذاتی سهم، عملکرد گذشته هر سهم معیاری کافی برای پیش‌بینی عملکرد آینده آن قلمداد می‌شود. به همین منظور تحلیل‌گران تکنیکی از نمودارها، اندیکاتورها^{۱۲} و اسیلاتورها^{۱۳} برای تشخیص الگوها کمک می‌گیرند تا بتوانند هرچه بهتر قیمت‌های آینده را پیش‌بینی کنند.

روش‌های مذکور دستی بوده و به همین علت نمی‌تواند یکی از مهمترین نیازهای سرمایه‌گذاران در بازارهای مالی امروزی یعنی سرعت بالا در انجام تحلیل‌ها و معاملات را برآورده سازد. برای دستیابی به این خواسته، تکنیک‌های معاملات الگوریتمی مطرح شد که یکی از مهمترین این تکنیک‌ها انتخاب بر خط سبد سهام می‌باشد. این تکنیک به دنبال تشکیل یک سبد سرمایه‌گذاری از طریق تخصیص بهینه سرمایه بین چند سهم به منظور دستیابی به بیشترین سود در افق زمانی مورد نظر می‌باشد و سبد سرمایه‌گذاری در ابتدای هر دوره به روز رسانی می‌شود. یادگیری ماشینی^{۱۴} از جمله دیدگاه‌های موجود در انتخاب بر خط سبد سرمایه‌گذاری است. یادگیری ماشینی برنامه‌نویسی برای بهینه‌سازی یک عملکرد با استفاده از داده‌ها و تجربیات گذشته بوده و به دنبال راهی برای ایجاد برنامه‌های است که عملکرد را به‌صورت خودکار و با توجه به تجربیات ارتقا دهد. با استفاده از یادگیری ماشینی، می‌توان روابط میان پایگاه داده‌های گذشته را تشخیص داد و تصمیمات هوشمندانه‌ای برای داده‌های جدید گرفت. لازم به ذکر است یکی از نقاط ضعف الگوریتم‌های انتخاب بر خط سبد سرمایه‌گذاری نادیده گرفتن ریسک سرمایه‌گذاری است که ما با استفاده از الگوریتم HRP سعی در کنترل ریسک در کنار بیشینه کردن بازده را داریم.

¹¹ Market Value

¹² Indicator

¹³ Oscillator

¹⁴ Machine Learning

۱-۴- توجیه، انگیزه و علت انتخاب موضوع

در بازارهای مالی امروزی با گسترش ابزارهای مالی و تنوع زیاد در این حوزه و همچنین بالا بودن سرعت و حجم معاملات، منطقی به نظر می‌رسد که به جای استفاده از روش‌های سنتی و دستی به دنبال استفاده از روش‌های نوین که عمدتاً مبتنی بر رایانه هستند، باشیم. از جمله‌ی این روش‌ها می‌توان به داد و ستد الگوریتمی اشاره کرد. همچنین از مهمترین تکنیک‌های معاملات الگوریتمی، انتخاب بر خط سبد سهام می‌باشد، که با کمک تکنولوژی سعی در حل مسائل مربوط به انتخاب سبد سرمایه گذاری دارد. بنابراین نیاز به الگوریتم‌هایی احساس می‌شود که بتوانند با سرعت بالا سبد سرمایه‌گذاری در دوره‌های مختلف را با هدف حداکثر سازی بازده یا ثروت در انتهای افق سرمایه‌گذاری تشکیل دهند.

۱-۵- اهمیت موضوع

امروزه صنعت سرمایه‌گذاری عمدتاً با چالش‌ها و مشکلات گوناگونی روبه‌رو می‌باشد و فعالان این صنعت به دنبال راه‌حلی برای مواجهه و برطرف کردن این مشکلات می‌باشند. با توجه به پیشرفت فناوری و تکنولوژی، یادگیری ماشینی و داده کاوی از جمله راه‌حل‌های بالقوه در این زمینه‌ها می‌باشند که به سرمایه گذاران و فعالان بازار در رفع این چالش‌ها کمک می‌کنند. از جمله این چالش‌ها می‌توان به تنوع بسیار زیاد در ابزارهای مالی و ابزارهای مشتقه مالی مانند معاوضه^{۱۵}، تاثیر احساسات انسانی در تصمیم‌گیری‌ها متعدد و سریع، خطاهای رفتاری انسانی و معاملات با تنوع بالا می‌باشد. همچنین در سال‌های گذشته با پیشرفت بازارهای مالی، افزایش تعداد شرکت کنندگان در بازار و افزایش تعداد شرکت‌های موجود در بازار و ابزارهای مالی بسیار زیاد، حجم داده‌ها و اطلاعات بسیار زیاد شده است. با این شرایط می‌توان گفت که امروزه تحلیل و بررسی این داده‌ها با سرعت لازم و همچنین تصمیم‌گیری سریع و بدون اشتباه و به دور از خطاهای رفتاری

¹⁵ Swap

انسانی از عهده روش‌های سنتی خارج می‌باشد. به همین علت روی آوردن به روش‌های نوین مانند معاملات الگوریتمی برای محاسبات، تحلیل‌ها و تصمیم‌گیری‌های لازم غیر قابل اجتناب می‌باشد.

۱-۶- مرور کلی بر ادبیات موضوع

در ادبیات مدل‌های انتخاب سبد سرمایه‌گذاری، دو اصل تئوریک عمده وجود دارد: مارکوییتز (۱۹۵۲) و (۱۹۵۹) تئوری اول را معرفی نمود و بهبود داد. تئوری دوم را کلی^{۱۶} (۱۹۵۶) معرفی کرد که به تئوری رشد سرمایه^{۱۷} مشهور می‌باشد و برای حداکثر کردن امید ریاضی لگاریتم بازده سرمایه‌گذاری است که در پورتفوهای چند دوره‌ای کاربرد دارد. مارکوییتز قبل از آن که دکترای خود را در سال ۱۹۵۴ کسب کند اولین مدل بهینه‌سازی پورتفو یعنی الگوریتم خط بحرانی^{۱۸} (CLA) را ارائه کرد به طوری که این الگوریتم به کمک بهینه‌سازی درجه دوم بر اوزان دارایی‌ها، پرتفوی بهینه‌ای با بیشینه نسبت شارپ به ما می‌دهد. کوهن و تاکر (۲۰۱۴) اثبات کردند که همانند تمامی مدل‌های بهینه‌سازی درجه دوم، در صورت برقرار بودن شرایط کاروش-کوهن-تاکر^{۱۹} این الگوریتم بهینه‌سازی بعد از تعداد محدودی تکرار به جواب می‌رسد که این ویژگی یکی از مهم‌ترین مزیت‌های این مدل است. توضیحات بیشتر و یک نمونه به کارگیری این مدل در مقاله بیلی و دپرادو^{۲۰} (۲۰۱۳) یافت می‌شود. در ادامه دپرادو در سال ۲۰۱۶ مدل HRP را ارائه داد.

عمدتاً تکنیک‌های انتخاب برخط سبد سرمایه‌گذاری را به چهار دسته الگوریتم‌های تبعیت از برنده، تبعیت از بازنده، تطابق با الگو و الگوریتم‌های فرا یادگیری تقسیم‌بندی می‌کنند.

¹⁶ Kelly

¹⁷ Capital Growth Theory

¹⁸ Critical Line Algorithm

¹⁹ Karush-Kuhn-Tucker

²⁰ MARCOS LÓPEZ DE PRADO

تبعیت از برنده، مهم‌ترین اصل در پژوهش‌های انتخاب برخط سبد سرمایه‌گذاری می‌باشد از نمونه‌های آن می‌توان به الگوریتم متوازن‌سازی ثابت متوالی، الگوریتم پرتفوی یونیورسال، الگوریتم گرادیان نمایی و الگوریتم برخط گام نیوتن را نام برد.

دسته بعدی الگوریتم‌های تبعیت از بازنده می‌باشد که برودین^{۲۱} (۲۰۰۴) تحت عنوان الگوریتم ضد همبستگی آن را معرفی نمود. لی^{۲۲} و همکاران (۲۰۱۲) الگوریتم بازگشت به میانگین منفعلانه-تهاجمی معرفی نمود. گائو و ژانگ^{۲۳} (۲۰۱۳) استراتژی ترکیبی منفعلانه-تهاجمی، لی و همکاران (۲۰۱۳) الگوریتم بازگشت به میانگین اطمینان وزنی، گائو و ژانگ (۲۰۱۳) الگوریتم بازگشت به میانگین متوسط متحرک وزن‌دار، لی و همکاران (۲۰۱۲) الگوریتم بازگشت به میانگین متحرک برخط و هانگ و همکاران (۲۰۱۳) الگوریتم بازگشت به میانه استوار معرفی نمودند. همچنین پژوهش ولیدی و نجفی (۱۳۹۸) از دیگر ادبیات موجود در این زمینه است. یک استراتژی که از هر دوی سهم‌های برنده و بازنده استفاده می‌کند الگوریتم تطابق با الگو نام دارد. پژوهش‌های گیورفی و شافر^{۲۴} (۲۰۰۳)، اتوساک و وایدا^{۲۵} (۲۰۰۷)، گیورفی و وایدا (۲۰۰۸)، عبدی و نجفی (۱۳۹۶)، از ادبیات موجود در این زمینه است.

الگوریتم‌های فرا یادگیری که مرتبط با یادگیری اکسپرت‌ها می‌باشد از دیگر الگوریتم‌های انتخاب آنلاین سبد سهام می‌باشد. الگوریتم‌های جمع‌کننده^{۲۶}، یونیورسال شدن سریع^{۲۷}، الگوریتم گرادیان برخط^{۲۸} و الگوریتم تبعیت از تاریخ راهنما^{۲۹} از نمونه‌های آن می‌باشد.

²¹ Borodin

²² Li

²³ Gao and Zhang

²⁴ Györfi and Schäfer

²⁵ Ottucsák and Vajda

²⁶ Aggregating Algorithms

²⁷ Fast Universalization

²⁸ Online Gradient Update (OGU)

²⁹ Follow-the-Leading History (FLH)

۱-۷- جنبه‌های جدید بودن موضوع

در حال حاضر پژوهش‌های گوناگونی در زمینه انتخاب برخط سبد سرمایه‌گذاری انجام شده است اما از آنجایی که الگوریتم‌های ارائه‌شده از کارایی خوبی برخوردار نیستند، پژوهشگران سعی در بهبود الگوریتم‌های پیشین دارند. برای اینکار می‌توان با ادغام تئوری‌های مختلف اقتصادی و تکنیک‌های یادگیری برخط، الگوریتم‌هایی را در جهت انتخاب بهتر سبد سهام به‌صورت خودکار ارائه داد. در ایران دو پژوهش توسط عبدی (۱۳۹۶) و ولیدی (۱۳۹۷) انجام شده است. بنابراین پژوهش حاضر جزو جدیدترین تحقیقات می‌باشد.

۱-۸- کاربردهای موضوع پژوهش

همانطور که در قسمت‌های قبل اشاره شد، سرمایه‌گذاران در بازارهای مالی به دنبال بیشینه کردن بازده خود از سرمایه‌گذاری می‌باشند. برای رسیدن به این هدف آن‌ها باید تصمیمات درست و به دور از احساسات را در مورد خرید و فروش دارایی‌شان و همچنین زمان درست این معاملات را اتخاذ کنند. همچنین با توجه به سرعت بالا در بازارهای امروزه، در این کار باید از سرعت مناسبی برخوردار باشند. الگوریتم‌های معاملاتی که برای انتخاب برخط سبد سهام به کار می‌رود با استفاده از داده‌های مناسب که از تکنیک‌های نوین داده کاوی به دست می‌آید این کار را برای ما انجام می‌دهند.

۱-۹- کاربران نتایج پژوهش

انتخاب برخط سبد سهام نقش اساسی در گستره‌ای از کاربردهای مالی نظیر مدیریت خودکار سرمایه، مدیریت صندوق‌های پوشش ریسک^{۳۰} و غیره ایفا می‌کند. درانتخاب برخط سبد سرمایه‌گذاری، سرمایه‌گذار به دنبال

³⁰ Hedge Funds

آن است تا فرآیند انتخاب سبد یا تخصیص سرمایه خود بین دارایی‌ها را به‌منظور بیشینه‌سازی ثروت خود در بلندمدت خودکارسازی کند. با توجه به موارد فوق، می‌توان مخاطبین اصلی این پژوهش را شرکت‌های مشاوره سرمایه‌گذاری، صندوق‌های سرمایه‌گذاری در اوراق با درآمد ثابت و صندوق‌های بازنشستگی کشوری دانست. همچنین افراد زیر می‌توانند از نتایج حاصل از این پژوهش استفاده کنند:

- دانشجویان یا فارغ‌التحصیلان رشته‌های مهندسی مالی، ریاضیات مالی و مدیریت مالی
- سرمایه‌گذاران در اوراق با درآمد ثابت
- تحلیل‌گران و مشاوران اقتصادی
- شرکت‌های سبدگردان
- شرکت‌های مشاور سرمایه‌گذاری
- اساتید و دانش‌پژوهان

۱-۱۰- جمع‌بندی

ابزارهای مالی متنوع، خطاهای رفتاری انسان و معاملات با تناوب بالا، سه چالش اساسی در بازارهای مالی امروزی هستند که باعث احساس نیاز به روش‌ها و الگوریتم‌های جدیدی می‌شود که دارای سرعت بالایی بوده و حجم زیادی از داده‌ها را در مدت زمان کوتاه پردازش کنند. همین مسئله سبب پیدایش معاملات الگوریتمی در بازارهای مالی شد که یکی از شاخه‌های آن را می‌توان انتخاب برخط سبد سرمایه‌گذاری دانست. در دهه‌های گذشته الگوریتم‌های متعددی در زمینه انتخاب برخط سبد سرمایه‌گذاری ارائه شده است که هر کدام دارای مزایا و البته معایب خاص خود بوده است. این مسئله سبب ارائه الگوریتم‌های جدیدتری شده است تا بتوانند معایب الگوریتم‌های گذشته را پوشش دهند

..

فصل دوم

مرور ادبیات و پیشینه تحقیق

۲-۱- مقدمه

مسئله انتخاب سبد، مسئله‌ای است که در آن ما به دنبال آن هستیم تا مقدار معینی از ثروت خود را در افق زمانی مشخصی سرمایه‌گذاری کنیم. در حل این مسئله سرمایه‌گذار تصمیم می‌گیرد که در هر دوره چه نسبتی از ثروت خود را به دارایی‌های مختلف اختصاص دهد به گونه‌ای که در انتهای دوره بیشترین عایدی را کسب نماید. این مسئله بسته به تعداد دوره‌ها به دو دسته تقسیم می‌شود: تک دوره‌ای و چند دوره‌ای. در انتخاب سبد تک دوره‌ای نیاز به تنها یک تصمیم در طول افق سرمایه‌گذاری است اما انتخاب سبد چند دوره‌ای نیازمند تصمیم‌گیری‌های متعدد در فواصل زمانی مشخص در افق زمانی پیش رو است و می‌توان آن را یک مسئله برخط لحاظ کرد. برای انتخاب بر خط سبد سرمایه‌گذاری، با استراتژی‌های تطابق با الگو دو مرحله اساسی وجود دارد، ابتدا انتخاب نمونه و سپس تشکیل سبد سرمایه‌گذاری با استفاده از داده‌های نمونه، برای انتخاب نمونه می‌توان از تکنیک‌های داده‌کاوی^۱ استفاده کرد و سپس با استفاده از داده‌های خروجی در این مرحله سبد سرمایه‌گذاری خود را تشکیل داده و آن را با استفاده از تکنیک‌های انتخاب بر خط سبد سرمایه‌گذاری به‌روز رسانی کرد. اما یکی از نقاط ضعف انتخاب بر خط سبد سرمایه‌گذاری در نظر نگرفتن ریسک می‌باشد که ما در اینجا با کمک گرفتن از الگوریتم HRP سعی در کنترل ریسک سبدهای سرمایه‌گذاری کرده‌ایم. در این فصل ابتدا ما داده کاوی، که ابزار گام نخست ما می‌باشد را توضیح خواهیم داد و سپس مفاهیم و الگوریتم‌های انتخاب بر خط سبد سرمایه‌گذاری،

^۱ Data Mining

که گام دوم می باشد را مورد بررسی قرار می دهیم. و در انتها به معرفی الگوریتم HRP خواهیم پرداخت که به وسیله ی آن ریسک سرمایه گذاری را کنترل خواهیم کرد.

۲-۲- داده کاوی

همانطور که می دانیم قدم اول برای انتخاب برخط سبد سرمایه، انتخاب نمونه می باشد. از راه های انتخاب نمونه می توان به روش های گوناگون داده کاوی اشاره کرد. در اینجا ما ابتدا داده کاوی را تعریف کرده و سپس انواع آن را مورد بررسی قرار می دهیم.

همانطور که می دانیم امروزه با پیشرفت و گسترش روز افزون تکنولوژی، سرعت تولید و انباشت داده های حاصل از پردازش، تبادلات و ارتباطات در فضای کسب و کار بسیار زیاد می باشد. این حجم بسیار بالای اطلاعات و داده ها باعث شده است که روش های سنتی و مرسوم تحلیل داده ها، مانند آمار کارایی خود را از دست بدهند و برای استخراج اطلاعات مورد نیاز از داده ها نیاز به حفاری در آنها باشد و داده کاوی فرآیندی است که با استفاده از تکنولوژی های پیشرفته این امکان را به ما می دهد.

تعاریف متفاوتی از داده کاوی وجود دارد، اما به طور کلی می توان گفت: استخراج دانش از مجموعه ای بزرگ از داده ها را داده کاوی گویند. داده کاوی با بهره گیری از ابزارهای پیشرفته تجزیه و تحلیل داده ها، به دنبال کشف الگوها و روابط معتبر بین متغیرها از پایگاه های عظیم داده ها می باشد به نحوی که این الگوها و روابط تا کنون ناشناخته بوده اند و در ادامه می توان این الگوها را دسته بندی و مدل سازی کرد. تکنیک های داده کاوی میتوان به چند دسته کلی زیر تقسیم بندی کرد که در ادامه به معرفی آنها می پردازیم:

۱- روش‌های آماری ۲- خوشه بندی^۲ ۳- درخت تصمیم^۳ ۴- قوانین انجمنی^۴ ۵- شبکه‌های عصبی^۵

۲-۱-۲- روش‌های آماری

روش‌های آماری از روش‌های پرکاربرد در تحلیل داده‌ها محسوب می‌شود. چنانچه تخمین و پیشگویی را یکی از وظایف داده‌کاوی بدانیم، تحلیل‌های آماری بیش از یک قرن است که داده‌کاوی را اجرا می‌کنند. روش‌های آماری بر روی داده‌ها اعمال شده و به منظور کشف موضوعات و ساختن الگوهای پیشگویانه استفاده می‌شود و از تکنیک‌های آن می‌توان به رگرسیون و دسته بندی به روش بیز اشاره کرد.

۲-۲-۲- خوشه بندی

فرآیند گروه‌بندی مجموعه‌ای از داده‌ها و قرار دادن آنها در طبقاتی از نمونه‌های مشابه خوشه‌بندی نام دارد. یک خوشه مجموعه‌ای از داده‌هاست که نسبت به دیگر داده‌های همان خوشه شبیه بوده ولی متفاوت از نمونه‌های دیگر خوشه‌ها هستند. تفاوت خوشه‌بندی و دسته‌بندی در این است که خوشه‌بندی، خوشه‌ها را تعریف می‌کند و داده‌ها را درون آنها می‌ریزد در حالی که در دسته‌بندی داده‌ها در دسته‌های از پیش تعیین شده قرار می‌گیرند. از تکنیک‌های آن می‌توان به روش‌های k-means ، c-means و خوشه‌بندی سلسله مراتبی اشاره کرد.

۲-۲-۳- درخت تصمیم

درخت تصمیم از ابزارهای قدرتمند و مشهور در داده‌کاوی بوده و در واقع ساختمان داده‌ای به منظور تقسیم مجموعه‌ای بزرگ از رکوردها به مجموعه‌های کوچکتر می‌باشد. این کار با استفاده از مجموعه‌ای از سوالات و قوانین تصمیم‌گیری بسیار ساده صورت می‌پذیرد. فرآیند مهم در داده‌کاوی برای درخت تصمیم، استفاده از روش‌های یادگیری ماشین به منظور آموزش درخت است و به آن یادگیری درخت

² Clustering

³ Decision Tree

⁴ Association Rules

⁵ Neural Networks

تصمیم گویند. یک مدل درخت تصمیم از مجموعه‌ای از قوانین برای تقسیم یک جمعیت ناهمگن و وسیع به گروه‌های کوچکتر و همگن‌تر بر اساس یک متغیر هدف خاص تشکیل شده است.

۲-۲-۴- قوانین انجمنی

یکی دیگر از روش‌های مهم در داده کاوی قوانین انجمنی می‌باشد که به دنبال یافتن میزان فراوانی ارتباط بین صفات در مجموعه‌ای از داده‌هاست، در اینجا ما به دنبال کشف الگوهای مکرر هستیم. در واقع در این روش وابستگی‌های مهم میان اقلام موجود در یک پایگاه داده را مشخص می‌کنیم به نحوی که حضور برخی از داده‌ها بر حضور بعضی داده دیگر از داده‌ها در همان پایگاه داده دلالت دارند.

۲-۲-۵- شبکه‌های عصبی

به زبان ساده شبکه‌های عصبی مجموعه‌ای از سیستم‌ها و روش‌های محاسباتی نوین برای یادگیری ماشین و نمایش دانش به دست آمده از این سیستم‌ها می‌باشد. دانش استخراج شده از این محاسبات را برای پیش بینی مورد استفاده قرار می‌دهند. شبکه عصبی مصنوعی الهام گرفته شده از فعالیت‌های بیولوژیکی مغز و شبکه عصبی انسان می‌باشد و در نتیجه این توانایی را دارند که از طریق تحلیل نمونه‌های مختلف به یادگیری بپردازند.

۲-۳- انتخاب بر خط سبد سرمایه‌گذاری

پس از مشخص کردن زیر مجموعه‌ای از سهام از روش‌های داده کاوی، در مرحله بعد ما به دنبال تشکیل سبد سرمایه گذاری خود و به روز رسانی آن هستیم. انتخاب سبد سرمایه گذاری را می‌توان به طور کلی به دو دسته‌ی الگوریتم‌های آنالین و آفلاین، مطرح شده توسط کارپ^۶ (۱۹۹۲) و آلبرز^۷ (۲۰۰۶)، تقسیم

^۶ Carp

^۷ Albers

نمود. از الگوریتم‌های آفلاین انتخاب سبد سرمایه گذاری می‌توان به مسئله‌ی انتخاب سبد سرمایه گذاری، مسئله تبدیل، استراتژی خرید و نگهداری، مسئله نیم انتخاب سبد سرمایه گذاری و استراتژی نیم انتخاب سبد سرمایه گذاری اشاره نمود که می‌توان با استفاده از نرم افزار هایی مانند گمز مسئله ریاضی این روش‌ها را حل کرده و تابع هدف آن را بهینه نمود و نسبت‌های بهینه دارایی‌های سبد مورد نظر خود را محاسبه کرد. اما همانطور که پیش از این نیز ذکر شد، امروزه به علت نیاز به سرعت و دقت بالا در تصمیم گیری‌های مالی ما را به استفاده از تکنولوژی‌های پیشرفته در انتخاب سبد سرمایه‌گذاری وا می‌دارد. الگوریتم‌های آنلاین از جمله این روش‌ها می‌باشد و همچنین از مهمترین روش‌های الگوریتم‌های آنلاین می‌توان به انتخاب بر خط سبد سرمایه گذاری اشاره نمود. چهار دسته اصلی الگوریتم‌های انتخاب بر خط سبد سرمایه گذاری را می‌توان الگوریتم‌های تبعیت از برنده^۸، الگوریتم‌های تبعیت از بازنده^۹، الگوریتم‌های تطابق با الگو^{۱۰} و الگوریتم‌های فرا یادگیری^{۱۱} نامید. با این حال در برخی از مقالات روش‌های دیگری نیز ارائه شده است که شاید نتوان آنها را در هیچ یک از دسته‌های بالا گنجانند. با این وجود ما در این فصل این چهار اصل مهم و زیر شاخه‌های آنها را توضیح داده و سعی می‌کنیم معرفی مناسبی از آنها داشته باشیم.

به علت اینکه مسئله انتخاب سبد سرمایه گذاری یک مسئله ریاضی می‌باشد، ابتدا متغیرهای موجود در این مدل‌ها را تعریف کرده و برخی عبارات، مفاهیم و مفروضات آنها را توضیح خواهیم داد و سپس الگوریتم‌های انتخاب بر خط سبد سرمایه‌گذاری را مورد بررسی قرار داده و مدل‌های ریاضی هر یک از آنها را معرفی خواهیم کرد.

مفروضات مسئله: در بیشتر الگوریتم‌های انتخاب بر خط سبد سرمایه گذاری مفروضات ساده کننده زیر برای ساده کردن مدل‌ها در نظر گرفته می‌شوند. این فرضیات عبارتند از:

^۸ Follow-the-Winner Algorithms

^۹ Follow-the-Loser Algorithms

^{۱۰} Pattern-matching Algorithms

^{۱۱} Meta-algorithms

- لحاظ نکردن هزینه‌های معاملاتی

- در نظر نگرفتن اثر استراتژی انتخاب شده بر روی بازار

- لحاظ نکردن نقدشوندگی^{۱۲} سهام

متوازن سازی^{۱۳}: تغییر وزن ثروت سرمایه‌گذاری شده از یک دارایی به سایر دارایی‌های سبد سرمایه‌گذاری را گویند.

اکسپرت^{۱۴}: اکسپرت یا خبره را یک سرمایه‌گذار فرضی را گویند که با استراتژی مختص به خود در بازار حضور دارد و ثروت اولیه آن $W_0 = 1$ می‌باشد.

پنجره زمانی: غالباً الگوریتم‌ها، از اطلاعات یک دوره برای بدست آوردن سبد سرمایه‌گذاری دوره بعد استفاده می‌کنند به این معنی که اطلاعات جدیدتر نسبت به اطلاعات قدیمی‌تر دارای ارزش بیشتری می‌باشد. اما استفاده از اطلاعات چند دوره قبل احتمالاً نتایج بهتری به همراه دارد. مثلاً می‌توان، اطلاعات مربوط به دوره‌های زمانی $t - \omega + 1$ تا t دوره گذشته را برای بدست آوردن پرتفوی دوره بعد به کار برد. که ω اندازه پنجره زمانی^{۱۵} می‌باشد و باید به وسیله سرمایه‌گذار تعیین شود.

اطلاعات مرتبه اول و دوم: بازده دوره نگهداری^{۱۶} با رابطه زیر بدست می‌آید:

$$x_t = \sum_{i=1}^m x_{it} b_{it} = \frac{W_t}{W_{t-1}} \quad (۱-۲)$$

¹² Liquidity

¹³ Rebalancing

¹⁴ Expert

¹⁵ Time Window

¹⁶ Holding Period Return

افزایش یا کاهش ثروت در یک دوره معاملاتی را شامل می‌شود. برخی از الگوریتم‌هایی که در ادامه به آنها پرداخته می‌شود اطلاعات مرتبه اول و دوم را به کار می‌برند که با مفهوم بازده دوره نگهداری به صورت مستقیم رابطه دارد.

اطلاعات مرتبه اول از رابطه زیر بدست می‌آید:

$$\theta_t^i = \frac{\partial \ln x_t}{\partial b_{it}} = \frac{x_{it}}{x_t} \quad (2-2)$$

که نشان دهنده تغییرات قیمتی دارایی i م در رابطه با بازده دوره نگهداری سبد سرمایه حال حاضر b_{it} می‌باشد. اگر θ_t^i بزرگ‌تر (کوچک‌تر) از یک باشد، نشان می‌دهد که دارایی i م دارای عملکرد بهتری (بدتری) نسبت به پرتفوی فعلی b_{it} می‌باشد.

اطلاعات مرتبه دوم از رابطه زیر بدست می‌آید:

$$\theta_t^{ij} = \frac{\partial^2 \ln x_t}{\partial b_{it} \partial b_{jt}} = -\frac{x_{it} x_{jt}}{x_t^2} \quad (3-2)$$

مقادیر θ_t^{ij} کمتر (بیشتر) از یک نشان می‌دهد پرتفوی با وزن یکسان متشکل از دارایی‌های i و j دارای عملکرد بهتر (بدتر)ی از پرتفوی فعلی b_{it} است. هلمبلد^{۱۷} (۱۹۹۸) از اطلاعات مرتبه اول استفاده کرده و می‌توان استفاده از اطلاعات مرتبه اول و دوم را در پژوهش آگاروال^{۱۸} و همکاران (۲۰۰۶) دید.

پارامترها و متغیرهای مورد استفاده در فرمول‌بندی مسائل عبارتند از:

t : شمارنده دوره زمانی

i : شمارنده دارایی

m : تعداد دارایی‌ها

¹⁷ Helmbold

¹⁸ Agarwal

T : تعداد دوره‌های زمانی

p_{it} : قیمت دارایی i م در زمان t

Q_{it}^j : نرخ تبدیل ثروت از دارایی i م به دارایی j م در زمان t

x_{it} : بردار نسبت قیمتی دارایی i م در زمان t

W_t : ثروت در زمان t

Y_i^{init} : تعداد اولیه دارایی i م در زمان $t = 0$

Y_{it} : تعداد سهام دارایی i م که پس از متوازن‌سازی در زمان t بدست می‌آید

y_{it} : تعداد سهام دارایی i م که توسط یک معامله در زمان t خریداری یا فروخته می‌شود.

روابط موجود بین متغیرهای معرفی‌شده شامل روابط زیر می‌باشد.

$$Q_{it}^j = \frac{p_{it}}{p_{jt}} \quad (۴-۲)$$

$$x_{it} = \frac{p_{it}}{p_{i,t-1}} \quad (۵-۲)$$

$$Y_{it} = Y_{i,t-1} + y_{it} \quad (۶-۲)$$

$$W_t = \sum_{i=1}^m p_{it} Y_{it} \quad (۷-۲)$$

$$W_T = \sum_{i=1}^m p_{iT} (Y_i^{init} + \sum_{t=1}^T y_{it}) \quad (۸-۲)$$

۲-۳-۱- الگو برداری

۲-۳-۱-۱- استراتژی متوازن سازی ثابت پرتفو^{۱۹}

استراتژی متوازن سازی ثابت پرتفو یکی از مهمترین استراتژی‌های الگو برداری است که در ابتدای هر دوره، پرتفو را به پرتفوی ثابت b متوازن سازی می‌کنیم. در واقع استراتژی انتخاب پرتفو به صورت $b_1^T = \{b, b, \dots, b\}$ می‌باشد. و ثروت سرمایه‌گذاری شده در دارایی i ام در ابتدای همه دوره‌ها از رابطه زیر محاسبه می‌شود:

$$b_{i,t} = \frac{Y_{i,t-1} \times p_{i,t-1}}{W_{t-1}} \quad (۹-۲)$$

که در آن شرط $\sum_{i=1}^m b_{i,t} = 1$ برقرار است. نسبت $b_{i,t}$ نیز در زمان $t - 1$ محاسبه می‌گردد.

همچنین ثروت نهایی در پایان دوره از رابطه زیر بدست می‌آید:

$$W_T(CRP(b)) = \prod_{t=1}^T b_t^\top x_t \quad (۱۰-۲)$$

(\top علامت ترانپوز ماتریس است.)

انواع استراتژی متوازن سازی ثابت :

(۱) استراتژی CRP بهینه آفلاین^{۲۰}: دارای پرتفو زیر است:

$$b^* = \arg \max W_T(CRP(b)) = \arg \max \prod_{t=1}^T b_t^\top x_t \quad (۱۱-۲)$$

استراتژی CRP که از این پرتفو استفاده کند، استراتژی بهترین متوازن سازی ثابت^{۲۱} پرتفو است و ثروت نهایی در پایان دوره سرمایه‌گذاری طبق فرمول زیر بدست می‌آید:

^{۱۹} Constant Rebalanced Portfolio

^{۲۰} Offline CRP Strategy

^{۲۱} Best Constant Rebalanced Portfolio (BCRP)

$$W_T(BCRP) = \max W_T(CRP(b)) = W_T(CRP(b^*)) \quad (۱۲-۲)$$

۲) استراتژی متوازن سازی ثابت یکنواخت^{۲۲}: در این استراتژی $b = (\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m})$ در همه دوره ها برقرار است. به این معنی که در همه دوره ها باید وزن سهام موجود در سبد با هم مساوی هستند.

۳) استراتژی متوازن سازی نیم ثابت^{۲۳}: در این استراتژی، در زمان های خاصی می بایست عملیات متوازن سازی انجام گیرد. می توان فرمول زیر را جهت متوازن سازی پرتفو به کار برد:

$$\dot{b}_{i,t} = 1_{t-1} b_{i,t} + (1 - 1_{t-1}) \dot{b}_{i,t-1} \quad (۱۳-۲)$$

و اگر در دوره t متوازن سازی مجاز باشد، یعنی $1_{t-1} = 1$ ، از فرمول

$$b_{i,t} = \frac{Y_{i,t-1} \times p_{i,t-1}}{W_{t-1}} \quad (۱۴-۲)$$

در غیر این صورت، یعنی $1_{t-1} = 0$ ، رابطه (۲-۱۳) بکار برده می شود.

جدول ۲-۱- فرم ریاضی مسئله CRP با m دارایی

مقادیر داده شده:

$$p_{i0}, \dots, p_{iT} \quad , i = 1, \dots, m$$

$$Y_i^{init} \quad , i = 1, \dots, m$$

پیدا کن:

$$y_{i0}, \dots, y_{iT} \quad , i = 1, \dots, m$$

تابع هدف:

²² Uniform Constant Rebalanced Portfolio (UCRP)

²³ Semi-Constant Rebalanced Portfolio

$$\max W_T = \sum_{i=1}^m p_{iT} (Y_i^{init} + \sum_{t=0}^T y_{it})$$

محدودیت‌ها:

$$(I) \sum_{t=0}^{\tau} y_{it} \geq -Y_i^{init}, \tau = 0, \dots, T, i = 1, \dots, m$$

$$(II) \sum_{i=1}^m p_{it} y_{it} = 0, t = 0, \dots, T$$

$$(III) \frac{(Y_i^{init} + \sum_{t=0}^{\tau} y_{it}) p_{i\tau}}{W_{\tau}} = \frac{(Y_i^{init} + \sum_{t=0}^{\tau-1} y_{it}) p_{i,\tau-1}}{W_{\tau-1}}, \tau = 1, \dots, T, i = 1, \dots, m-1$$

$$(IV) y_{it} \in R, t = 0, \dots, T, i = 1, \dots, m$$

۲-۳-۱-۲- استراتژی خرید و نگهداری

بر طبق این استراتژی سرمایه گذار در ابتدای دوره پرتفوی خود را تشکیل داده و آن را تا انتهای دوره مد نظر نگهداری می‌کنند. پرتفو در پایان دوره t از روابط زیر بدست می‌آید:

$$b_{i,t} = \frac{Y_{i,t-1} \times p_{i,t-1}}{W_{t-1}} \quad (۱۵-۲)$$

یا

$$\frac{b_t \odot x_t}{b_t^{\top} x_t} \quad (۱۶-۲)$$

در عبارت فوق، \odot ضرب مولفه در مولفه دو بردار است.

در این استراتژی ارزش پرتفو در انتهای دوره برابر با میانگین موزون بازدهی هر دارایی بر اساس پرتفوی اولیه است از طریق فرمول زیر محاسبه می‌شود:

$$W_T(BAH(b_1)) = b_1^{\top} \cdot (\odot_{t=1}^T x_t) \quad (۱۷-۲)$$

۲-۳-۲- الگوریتم‌های تبعیت از برنده

در دنیای مالی، پیروی از روند یک استراتژی معروف است که فرض می‌کند سهم‌های دارای عملکرد بهتر در گذشته احتمالاً در آینده نیز عملکرد بهتری را از خود نشان می‌دهند به این ایده مومنتوم گفته می‌شود. می‌توان گفت تبعیت از برنده، یکی از مهم‌ترین اصل‌ها در تحقیقات انتخاب برخط سبد سرمایه‌گذاری است. در این الگوریتم‌ها، وزن‌ها بیشتر به سهامی اختصاص می‌یابد که در گذشته عملکرد بهتری از خود به جای گذاشته‌اند. الگوریتم‌های این دسته استراتژی بازار و استراتژی بهترین سهام دنبال نمی‌کنند و بیشتر استراتژی BCRP را با هدف رسیدن به یونیورسال بودن^{۲۴} دنبال می‌کنند. از جمله زیر مجموعه‌های آن می‌توان به الگوریتم‌های زیر اشاره کرد:

۲-۳-۱- الگوریتم تبعیت از رهبر

به این دلیل که اطلاعاتی در ابتدای دوره اول موجود نیست، سرمایه‌گذار پرتفوی $b_1 = (\frac{1}{m}, \dots, \frac{1}{m})$ را برمی‌گزیند و در دوره‌های بعد، به دنبال بهترین نسبت‌ها برای b^* است درحالی‌که استراتژی تبعیت از رهبر را برای دوره‌های $1, \dots, t$ را انتخاب نموده است. پرتفوی بهینه b^* از رابطه زیر بدست می‌آید:

$$b_{t+1}^{SCR} = b_t^* = \arg \max \sum_{\tau=1}^t \log(b \cdot x_{\tau}) \quad (۱۸-۲)$$

این الگوریتم، که الگوریتم متوازن‌سازی ثابت متوالی نیز نامیده می‌شود توسط گایورونسکی و استلا^{۲۵} (۲۰۰۰) ارائه شد و از اکسپرت‌های CRP استفاده می‌کند با این تفاوت که SCR همواره از اکسپرت‌های BCRP تا دوره t استفاده می‌کند. در این روش، در هر دوره از BCRP دوره‌های گذشته به‌عنوان رهبر استفاده شده و رهبر نهایی، BCRP تمام دوره‌ها است. اوردنتلیچ (۱۹۹۶) استراتژی BCRP دوره‌های گذشته و پرتفوی یکنواخت را تلفیق کرد و فرمول به‌روزرسانی زیر را ارائه داد:

²⁴ Universality

²⁵ Gaivoronski and Stella

$$b_{t+1} = \frac{t}{t+1} b_t^* + \frac{1}{t+1} \frac{1}{m} \mathbf{1} \quad (۱۹-۲)$$

که در آن $\mathbf{1}$ بردار سطری واحد است.

در ادامه الگوریتم متوازن سازی ثابت پیاپی موزون پرتفو^{۲۶} برای بازارهای ایستا^{۲۷} توسط گایورونسکی و استلا (۲۰۰۰) ارائه شد که در آن، پرتفوی دوره بعد شامل ترکیب محدب SCRCP و پرتفوی دوره قبل می باشد.

$$b_{t+1} = (1 - \gamma) b_t^* + \gamma b_t \quad (۲۰-۲)$$

سپس این دو برای بازارهای غیر ایستا، استراتژی متوازن سازی متغیر پرتفو^{۲۸} را ارائه دادند که در این استراتژی برای مشخص کردن پرتفوی دوره آینده از BCRP پنجره زمانی اخیر با طول پنجره زمانی ω به صورت رابطه زیر استفاده شده است:

$$b_{t+1} = \arg \max \sum_{\tau=t-\omega+1}^t \log(b \cdot x_\tau) \quad (۲۱-۲)$$

همچنین برای الگوریتم VRP الگوریتم های تکمیلی متوازن سازی متغیر پیاپی پرتفوی^{۲۹} و متوازن سازی متغیر پیاپی موزون پرتفو^{۳۰} ارائه شد.

این الگوریتم، که الگوریتم متوازن سازی ثابت متوالی نیز نامیده می شود توسط گایورونسکی و استلا^{۳۱} (۲۰۰۰) ارائه شد و از اکسپرت های CRP استفاده می کند با این تفاوت که SCRCP همواره از اکسپرت های BCRP تا دوره t استفاده می کند. در این روش، در هر دوره از BCRP دوره های گذشته به عنوان رهبر استفاده شده و

²⁶ Weighted Successive Constant Rebalanced Portfolio (WSCRCP)

²⁷ Static Market

²⁸ Variable Rebalanced Portfolio (VRP)

²⁹ Successive Variable Rebalanced Portfolio (SVRP)

³⁰ Weighted Successive Variable Rebalanced Portfolio (WSVRP)

³¹ Gaivoronski and Stella

رهبر نهایی، BCRP تمام دوره‌ها است. اوردنتلیج (۱۹۹۶) استراتژی BCRP دوره‌های گذشته و پرتفوی یکنواخت را تلفیق کرد.

در ادامه الگوریتم متوازن‌سازی ثابت پیاپی موزون پرتفو برای بازارهای ایستا توسط گایورونسکی و استلا (۲۰۰۰) ارائه شد.

۲-۳-۲- الگوریتم پرتفوی یونیورسال

در این الگوریتم، تخصیص ثروت به اکسپرت‌های یک گروه خاص می‌باشد. به طور مثال $\omega = 1, \dots, \theta$ اکسپرت CRP داریم که هرکدام از آن‌ها پرتفوی CRP خاص خود را دارد. برای تعیین نسبت ثروت سرمایه‌گذاری شده در دارایی t م، ثروت هر اکسپرت در زمان t و نسبت دارایی t م در پرتفو در نظر گرفته خواهد شد. در این الگوریتم اگر یک دارایی نسبت به سایر دارایی‌ها بازدهی بیشتر باشد، اکسپرت-هایی که دارای مقدار بیشتری از این دارایی باشند، باید به طور متوسط، نسبت به سایر اکسپرت‌ها دارای ثروت دوره‌ای بزرگ‌تری باشند. اولین بار این الگوریتم توسط کاور (۱۹۹۱) ارائه شد.

$$S_n(\mathbf{b}) = \prod_{i=1}^n \mathbf{b}^1 \mathbf{x}_i \quad (22-2)$$

جایی که سرمایه اولیه برابر با یک می‌باشد. $S_0(\mathbf{b}) = 1$

$$S_n^* = \max_b S_n(\mathbf{b}) \quad (23-2)$$

$$\hat{\mathbf{b}}_{k+1} = \frac{\int \mathbf{b} S_k(\mathbf{b}) d\mathbf{b}}{\int S_k(\mathbf{b}) d\mathbf{b}} \quad (24-2)$$

$$\hat{\mathbf{b}}_1 = \left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m} \right) \quad (25-2)$$

$$S_k(\mathbf{b}) = \prod_{i=1}^k \mathbf{b}^t \mathbf{x}_i \quad (26-2)$$

$$B = \left\{ b \in \mathbb{R}^m : b_i \geq 0, \sum_{i=1}^m b_i = 1 \right\} \quad (27-2)$$

ثروت نهایی حاصل از پرتفوی یونیورسال توسط فرمول زیر به دست می‌آید:

$$\hat{S}_n = \prod_{k=1}^n \hat{\mathbf{b}}_k^t \mathbf{x}_k \quad (28-2)$$

این الگوریتم به پرتفوی یونیورسال موزون به μ ^{۳۲} توسط کاور و اوردنتلیچ^{۳۳} (۱۹۹۶) بهبود داده شد که در آن μ به توزیع داده شده بر روی فضای پرتفو اشاره می‌کند. در این الگوریتم، سرمایه اولیه با نسبت $d\mu$ بین اکسپرت‌ها تقسیم و هر اکسپرت یک استراتژی CRP مخصوص خود را دارد. سپس در پایان دوره دارایی هر کدام از طریق فرمول $W_T(b)d\mu(b)$ بدست خواهد آمد. در نهایت بازده تمام آنها جمع شده و به صورت رابطه زیر خواهد شد:

$$W_T(UP) = \int_{\Delta_m} W_T(b) d\mu(b) \quad (29-2)$$

در هر دوره، نسبت‌های سرمایه‌گذاری، از میانگین وزنی استراتژی دوره قبل از طریق رابطه زیر بدست می‌آید:

$$b_{t+1} = \frac{\int_{\Delta_m} b W_t(b) d\mu(b)}{\int_{\Delta_m} W_t(b) d\mu(b)} \quad (30-2)$$

در فضای گسسته پرتفوی دوره بعد از رابطه زیر بدست می‌آید:

³² μ -weighted Universal Portfolio

³³ Cover and Ordentlich

$$b_{i,t+1} = \frac{\sum_{\omega=1}^{\theta} b_i^{\omega} W_t(b^{\omega})}{\sum_{\omega=1}^{\theta} W_t(b^{\omega})} \quad (31-2)$$

اگر توزیع $d\mu$ به صورت یکنواخت باشد، همان الگوریتم کلاسیک پرتفوی یونیورسال بدست می آید:

اکسپرت هایی که دارای مقدار بیشتری از این دارایی باشند، باید به طور متوسط، نسبت به سایر اکسپرت ها دارای ثروت دوره ای بزرگتری باشند. این الگوریتم که توسط کاور (۱۹۹۱) ارائه شد. کاور و اوردنتلیچ (۱۹۹۶) این الگوریتم را به پرتفوی یونیورسال موزون به μ بهبود دادند. در این الگوریتم، سرمایه اولیه با نسبت $d\mu$ بین اکسپرت ها تقسیم و هر اکسپرت یک استراتژی CRP مخصوص خود را دارد. کاور و اوردنتلیچ (۱۹۹۶) اطلاعات جانبی مانند نظر خبرگان و اطلاعات بنیادین را نیز وارد مدل کردند. بلوم و کالای (۱۹۹۹) مدل را با هزینه معاملاتی ارتقا دادند. کالای و ومپالا (۲۰۰۲) روشی را ارائه دادند که سریع تر به جواب می رسد و مدت زمان محاسبات آن کاهش می یابد. کراس و بارون (۲۰۰۳) و آکوگلو و همکاران (۲۰۰۵)، حالت تعمیم یافته ای از الگوریتم اولیه را ارائه دادند به این صورت که هر اکسپرت می تواند استراتژی های گوناگونی را به کار گیرد و لزومی به استفاده از استراتژی CRP نیست. کوزات و سینگر (۲۰۱۱)، نیمه CRP را به عنوان استراتژی هر اکسپرت به جای CRP ارائه کردند.

۲-۳-۲-۳ الگوریتم گرادیان نمایی

الگوریتم گرادیان نمایی بر اطلاعات مرتبه اول متکی بوده بطوریکه که پرتفوی b_{t+1} بر مبنای پرتفوی b_t و اطلاعات مرتبه اول محاسبه خواهد شد. در این الگوریتم اگر نسبت قیمتی دارایی i م در دوره t بزرگتر از نسبت تغییر ثروت در دوره فعلی باشد، آنگاه نسبت دارایی i م در پرتفوی دوره $t + 1$ باید افزایش یابد که مطابق با اصل تبعیت از برنده می باشد. این الگوریتم را هلمبلد و همکاران (۱۹۹۸) ارائه نمودند و فرمول بهینه سازی آن به صورت رابطه زیر می باشد:

$$b_{t+1} = \arg \max \eta \log b \cdot x_t - R(b, b_t) \quad (32-2)$$

که در آن η نرخ یادگیری^{۳۴} و $R(b, b_t)$ عبارت تنظیم^{۳۵} است که از طریق رابطه زیر محاسبه شده است:

$$R(b, b_t) = \sum_{i=1}^m b_i \log \frac{b_i}{b_{it}} \quad (۳۳-۲)$$

فرمول گرادیان نمایی محدب می‌باشد. چون که محاسبه تابع غیر خطی لگاریتمی کار دشواری است، لذا می‌توان از بسط تیلور مرتبه اول^{۳۶} تابع لگاریتم به صورت رابطه زیر بهره گرفت:

$$\log b \cdot x_t \approx \log(b_t \cdot x_t) + \frac{x_t}{b_i \cdot x_t} (b - b_t) \quad (۳۴-۲)$$

در نتیجه جزء غیرخطی لگاریتم به خطی تبدیل شده و حل مسئله بهینه‌سازی ساده می‌شود. پس از حل، سبد سرمایه EG به شکل فرمول زیر است:

$$b_{i,t+1}^{EG} = \frac{b_{i,t} \exp(\eta \theta_t^i)}{\sum_{i=1}^m b_{it} \exp(\eta \theta_t^i)} \quad (۳۵-۲)$$

که در آن θ_t^i اطلاعات مرتبه اول است.

در این اینجا، می‌توان برای دوره اول b_1 از هر تخصیصی بهره برد. اما اگر $\eta = 0$ باشد، آنگاه همان پرتفوی b_1 که در ابتدا در نظر گرفته شده، برای دوره‌های آتی نیز بکار گرفته می‌شود و عملاً استراتژی EG تبدیل به الگوریتم متوازن سازی ثابت یکنواخت می‌شود. و اگر $\eta < 0$ باشد، آنگاه الگوریتم EG یک الگوریتم تبعیت از بازنده خواهد بود.

³⁴ Learning Rate

³⁵ Regularization Term

³⁶ First-order Taylor Expansion

یانگ و همکاران (۲۰۱۹) اطلاعات جانبی را وارد مدل کردند و نشان دادند که در این حالت الگوریتم نتیجه بهتری می‌دهد. آنها نام این الگوریتم را EGS گذاشتند.

۲-۳-۴- الگوریتم برخط گام نیوتن

این الگوریتم علاوه بر اطلاعات مرتبه اول، اطلاعات مرتبه دوم را نیز در نظر می‌گیرد. ONS از پارامترهای β و δ استفاده کرده که در ادامه آنها را توضیح می‌دهیم. به ازای هر دوره ماتریس:

$$A_t = \begin{pmatrix} 1 - \sum_{\tau=1}^t \theta_{\tau}^{11} & \dots & 0 - \sum_{\tau=1}^t \theta_{\tau}^{1m} \\ \vdots & \ddots & \vdots \\ 0 - \sum_{\tau=1}^t \theta_{\tau}^{m1} & \dots & 1 - \sum_{\tau=1}^t \theta_{\tau}^{mm} \end{pmatrix} \quad (36-2)$$

را در نظر بگیرید که در آن عنصر مربوط به سطر i و ستون j آن توسط a_t^{ij} نمایش داده شده است. توجه داشته باشید $a_t^{ij} = 1 - \sum_{\tau=1}^t \theta_{\tau}^{ij}$ بر روی عناصر قطری ماتریس و $a_t^{ij} = 0 - \sum_{\tau=1}^t \theta_{\tau}^{ij}$ بر روی عناصر غیر قطری ماتریس وجود دارد. A_t^{-1} را ماتریس معکوس A_t در نظر گرفته به نحوی که عنصر مربوط به سطر i و ستون j آن به صورت \bar{a}_t^{ij} قابل نمایش است. در این مرحله از محاسبات از اطلاعات مرتبه دوم استفاده می‌شود. تغییرات قیمت هر دارایی را در مقایسه با تغییر ثروت در نظر بگیرید. بردار o_t اطلاعات مرتبه اول و دوم را ترکیب می‌کند که در آن

$$o_t = \begin{pmatrix} \delta(1 + \frac{1}{\beta}) \sum_{j=1}^m \bar{a}_t^{1j} \sum_{\tau=1}^t \theta_{\tau}^j \\ \vdots \\ \delta(1 + \frac{1}{\beta}) \sum_{j=1}^m \bar{a}_t^{mj} \sum_{\tau=1}^t \theta_{\tau}^j \end{pmatrix} \quad (37-2)$$

و یک عنصر o_t به صورت o_{it} به ازای $i = 1, \dots, m$ قابل نمایش است. پرتفوی دوره $t + 1$ از رابطه زیر بدست می‌آید:

$$b_{t+1}^{ONS} = \arg \min (o_t - b)^T A_t (o_t - b) \quad (38-2)$$

$$b_{t+1}^{ONS} = \arg \min \left(\sum_{i=1}^m a_t^{ii} (o_{it} - b_i) + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^m a_t^{ij} (o_{it} - b_i)(o_{jt} - b_j) \right) \quad (39-2)$$

برای محاسبه آنلاین الگوریتم ONS از گام‌های زیر استفاده کنید:

$$a_t^{ij} = a_{t-1}^{ij} - \theta_t^{ij} \quad (40-2)$$

که در آن برای عناصر قطری $a_0^{ij} = 0$ و برای عناصر غیر قطری $a_0^{ij} \neq 0$. اطلاعات مرتبه اول توسط رابطه زیر بدست می‌آید:

$$o'_{it} = \left(1 + \frac{1}{\beta}\right) \theta_t^i + o'_{i,t-1} \quad (41-2)$$

که در آن به ازای $i = 1, \dots, m$ داریم: $o'_{i0} = 0$

سپس اطلاعات مرتبه اول و دوم توسط رابطه زیر به ازای $i = 1, \dots, m$ ترکیب می‌شود:

$$o_{it} = \delta \sum_{j=1}^m \bar{a}_t^{ij} o'_{jt} \quad (42-2)$$

الگوریتم توسعه داده‌شده الگوریتم برخط نیوتن، می‌گوید که عملکرد کلی نمی‌تواند بدتر از هیچ‌یک از ترکیبات مختلف اکسپرت‌ها باشد. که به آن الگوریتم به‌روزرسانی برخط نیوتن^{۳۷} گویند. این الگوریتم توسط داس و بانرجی^{۳۸} (۲۰۱۱) توسعه داده شده است.

فانفانگ و همکاران^{۳۹} (۲۰۱۸) در مقاله خود مطرح می‌کنند که ONS خطرات نزولی را در نظر نمی‌گیرد، و منجر به زیان زیادی سرمایه‌گذاری در بعضی از بازارها می‌شود. برای غلبه بر این محدودیت، در این مقاله یک

³⁷ Online Newton Update (ONU)

³⁸ Das and Banerjee

³⁹ Fanfang et.al

روش انتخاب جدید نمونه، یعنی ONS با اطلاعات جانبی ONS-SI که شامل ONS با اطلاعات جانبی حاصل از بازار است، برای کاهش ریسک سرمایه گذاری پیشنهاد می شود.

هوانگ^{۴۰} (۲۰۱۸) اظهار می کند که ONS ناپایدار است و ممکن است تغییرات غیرضروری را در تخمین نهایی ایجاد کند. او در مقاله خود با استفاده از برآورد کننده های پیش بینی کننده ترکیبی از پدیده بازگشت استفاده می کند.

۲-۳-۳- الگوریتم های تبعیت از بازنده

استراتژی BCRP در صورتی که توزیع بازار مستقل و i.i.d باشد بهینه است. اما فرض مذکور در دنیای واقعی وجود نداشته و بنابراین ممکن است موجب عملکرد نامطلوب الگوریتم های تبعیت از برنده گردد. برخلاف ردیابی برنده ها که در الگوریتم های تبعیت از برنده وجود دارد، در واقع الگوریتم های تبعیت از بازنده ثروت را از دارایی های با عملکرد خوب در گذشته به دارایی های با عملکرد بد در گذشته جابجا می کنند بوند و تالر (۱۹۸۵) به این فرض، بازگشت به میانگین گفته می شود. طبق این اصل، دارایی های با عملکرد خوب (بد)، در دوره های بعدی دارای عملکرد بد (خوب) خواهند بود. بر طبق این اصل باور سرمایه گذاران این است که سهام با عملکرد بد می تواند از نوسانات بازار بهره مند شود لونبرگر^{۴۱} (۱۹۹۸). حال مهم ترین الگوریتم هایی که در این زمینه صورت گرفته است را مورد بررسی قرار خواهیم داد.

طبق اصل بازگشت به میانگین، دارایی های با عملکرد خوب (بد) در گذشته، در دوره های آتی دارای عملکرد بد (خوب) خواهند بود. الگوریتم های تبعیت از بازنده از این اصل پیروی می کنند. از الگوریتم های آن می توان به کار برودین (۲۰۰۴) تحت عنوان الگوریتم ضد همبستگی اشاره کرد.

⁴⁰ Huang

⁴¹ Luenberger

۲-۳-۱- الگوریتم ضد همبستگی

این الگوریتم توسط برودین و همکاران (۲۰۰۴) ارائه شده است، فرض می‌شود که بازار از اصل بازگشت به میانگین تبعیت خواهد کرد. برای استفاده از این اصل در مدل، داده‌ها باید خودهمبستگی^{۴۲} منفی و همبستگی متقاطع^{۴۳} مثبت با تاخیر داشته و فرض دیگر آن این است که تمام سهم‌هایی که در بازار هستند، در طولانی مدت دارای نرخ رشد نمایی کاملاً مشابه هستند. نسبت‌های قیمتی لگاریتمی در دو پنجره زمانی به شرح زیر بکار گرفته می‌شود:

$$y_1 = \log(x_{t-2\omega+1}^{t-\omega}) \quad , \quad y_2 = \log(x_{t-\omega+1}^t) \quad (۴۳-۲)$$

$$M_{cov}(i, j) = \frac{1}{\omega-1} (y_{1,i} - \bar{y}_1)^\top (y_{2,j} - \bar{y}_2) \quad (۴۴-۲)$$

$$M_{cor}(i, j) = \begin{cases} \frac{M_{cov}(i, j)}{\sigma_1(i) \times \sigma_2(j)} & \sigma_1(i), \sigma_2(j) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (۴۵-۲)$$

و سپس ماتریس همبستگی با توجه به فرمول فوق به دست می‌آید. $M_{cov}(i, j)$ ماتریس کوواریانس و $M_{cor}(i, j)$ ماتریس همبستگی متقاطع دو دارایی i و j می‌باشد. با توجه به اصل بازگشت به میانگین، این الگوریتم وزن را از دارایی‌های که بیشتر رشد داشته‌اند به دارایی‌های با رشد قیمت کمتر منتقل کرده و مقدار آن از فرمول زیر بدست می‌آید:

$$C_{-}(i, j) = M_{cor}(i, j) - [\min\{0, M_{cor}(i, +\min\{0, M_{cor}(j, j)\})\}] \quad (۴۶-۲)$$

و پرتفوی دوره بعد b_{t+1} از طریق رابطه زیر محاسبه می‌شود:

$$b_{i,t+1} = b_{i,t} \left(1 - \sum_{j=1}^m \frac{C_{i,j}}{\sum_{j'=1}^m C_{i,j'}}\right) + \sum_{j=1}^m b_{j,t} \left(\frac{C_{i,j}}{\sum_{j'=1}^m C_{i,j'}}\right) \quad (۴۷-۲)$$

^{۴۲} Autocorrelation

^{۴۳} Cross-correlation

این الگوریتم ابتکاری است و تضمین تئوریک ندارد و به همین علت اگر خاصیت بازگشت به میانگین را توسط الگوریتم‌های یادگیری اجرا کنیم، می‌توان آن را با اطمینان بیشتری قبول کرد.

چانگ^{۴۴} (۲۰۱۹) در هر الگوریتم Anticor اصلی هر دو حرکت و معکوس را بررسی می‌کند و یک الگوریتم انتخاب جدید پرتفو آنلاین را به نام الگوریتم ضد همبستگی Kalman Momentum W-KACM پیشنهاد می‌کند که می‌تواند به طور کامل از ویژگی قیمت نوسانات قیمت بهره‌بردار.

WKFRP، بر خلاف نسبی قیمت نسبی که تنها میزان قیمت را از یک دوره به دوره بعدی منتقل می‌کند، اندازه‌گیری می‌کند که چقدر قیمت از ارزش روند ذاتی متفاوت است.

در این الگوریتم، بجای اینکه هیچ فرضی در مورد توزیع بازار وجود نداشته باشد (مانند الگوریتم UP ارائه‌شده توسط کاور) فرض می‌شود بازار از اصل بازگشت به میانگین پیروی می‌کند. دیگر فرضی که وجود دارد این است که تمام سهم‌هایی که در بازار وجود دارند، در دراز مدت از نقطه نظر نرخ رشد نمایی کاملاً مشابه عمل می‌کنند.

۲-۳-۳-۲- الگوریتم بازگشت به میانگین منفعلانه-تهاجمی

این الگوریتم که توسط لی و همکاران (۲۰۱۲) معرفی شد، از اصل بازگشت به میانگین استفاده کرده و از یادگیری برخط منفعلانه-تهاجمی که توسط کرامر^{۴۵} (۲۰۰۶) ارائه شده است، بهره‌می‌برد، و همچنین به‌خوبی می‌تواند موازنه‌ای را بین بازده و ریسک برقرار کرده و برای نمایش مفهوم بازگشت به میانگین از تابع زیان به‌صورت زیر استفاده می‌کند:

$$l_{\varepsilon}(b, x_t) = \begin{cases} 0 & b \cdot x_t \leq \varepsilon \\ b \cdot x_t - \varepsilon & \text{otherwise} \end{cases} \quad (2-48)$$

و $\varepsilon \geq 0$ پارامتر حساسیت^{۴۶} بوده که آستانه بازگشت به میانگین را کنترل خواهد کرد. چون نوسان بردار نسبت قیمتی x_t حول عدد یک است، بنابراین $\varepsilon \leq 1$ قرار می‌دهیم تا بتوانیم دارایی‌های با عملکرد

⁴⁴ Chang

⁴⁵ Crammer

⁴⁶ Sensitivity Parameter

ضعیف را خریداری کنیم. اگر بازده انتظاری بر اساس آخرین قیمت‌ها بیشتر از پارامتر حساسیت باشد، زیان به صورت خطی افزایش خواهد یافت. در غیر این صورت زیان برابر با صفر می‌شود.

در این الگوریتم از یادگیری برخط منفعلانه-تهاجمی استفاده می‌شود. این تکنیک، تحت شرایطی که در ادامه آورده شده، نیازمند این است که پرتفو به صورت منفعلانه حفظ و یا به صورت فعالانه به روز رسانی گردد. در نهایت مسئله بهینه‌سازی به صورت رابطه زیر می‌باشد:

$$b_{t+1} = \arg \min \frac{1}{2} \|b - b_t\|^2 \text{ s.t. } l_\varepsilon(b, x_t) = 0 \quad (49-2)$$

در فرمول مذکور، هدف پیدا کردن پرتفو از طریق مینیمم‌سازی انحراف از آخرین پرتفو b_t است در صورتی که شرط صفر بودن تابع زیان برقرار گردد. در هر مرحله از متوازن‌سازی، اگر تابع زیان فوق برابر با صفر باشد، آنگاه آخرین پرتفو حفظ خواهد شد یعنی $b_{t+1} = b_t$. و اگر تابع زیان غیر صفر باشد، الگوریتم به صورت تهاجمی پرتفوی قدیم را به پرتفوی جدید به روز رسانی کرده تا محدودیت صفر بودن تابع زیان حاصل گردد. اگر در بازار مالی نوسانات شدید وجود داشته باشد، ممکن است نتیجه مسئله بهینه‌سازی فوق، راه‌حل‌های اشتباه برای سرمایه‌گذار گردد. برای مثال، وجود یک نسبت قیمتی پرت ممکن است به یکباره پرتفو را در یک جهت اشتباه به روز رسانی نماید. برای حل این مشکل، دو نوع دیگر از PAMR مطرح گردید که موازنه‌ای را بین منفعلانه و یا تهاجمی رفتار کردن الگوریتم ایجاد می‌نماید. در یکی از این الگوریتم‌ها، مسئله بهینه‌سازی به صورت رابطه زیر است:

$$b_{t+1} = \arg \min \left\{ \frac{1}{2} \|b - b_t\|^2 + C\mathfrak{L} \right\} \text{ s.t. } l_\varepsilon(b, x_t) \leq \mathfrak{L} \text{ and } \mathfrak{L} \geq 0 \quad (50-2)$$

$\mathfrak{L} \geq 0$ متغیر کمبود^{۴۷} و $C \geq 0$ پارامتری برای کنترل تاثیر متغیر کمبود در تابع هدف می‌باشد.

همچنین مسئله بهینه‌سازی به صورت رابطه زیر بر طبق این الگوریتم نشان داده می‌شود:

⁴⁷ Slack Variable

$$b_{t+1} = \arg \min \left\{ \frac{1}{2} \|b - b_t\|^2 + C\beta^2 \right\} \text{ s.t. } l_\varepsilon(b, x_t) \leq \beta \quad (51-2)$$

چیزی که در مورد دو نوع اخیر این الگوریتم وجود دارد، این است که این دو می‌توانند موازنه‌ای بین بازده و ریسک سرمایه‌گذاری ایجاد نماید. بنا بر گفته نویسندگان این پژوهش، این دو نوع PAMR توانسته‌اند عملکرد خوبی را در سنجه‌های بازده تعدیل‌شده به ریسک از خود نشان دهند.

پس از حل مسئله بهینه‌سازی فوق و با فرض عدم فروش استقراضی ($b \geq 0$)، نتایج زیر حاصل می‌شود:

$$b = b_t - \tau_t(x_t - \bar{x}_t \mathbf{1}) \quad (52-2)$$

که در آن:

$$\tau_t = \begin{cases} \max \left\{ 0, \frac{b_t \cdot x_t - \varepsilon}{\|x_t - \bar{x}_t \mathbf{1}\|^2} \right\} & \text{(PAMR)} \\ \max \left\{ 0, \min \left\{ C, \frac{b_t \cdot x_t - \varepsilon}{\|x_t - \bar{x}_t \mathbf{1}\|^2} \right\} \right\} & \text{(PAMR - 1)} \\ \max \left\{ 0, \frac{b_t \cdot x_t - \varepsilon}{\|x_t - \bar{x}_t \mathbf{1}\|^2 + \frac{1}{2C}} \right\} & \text{(PAMR - 2)} \end{cases} \quad (53-2)$$

پس از قرار دادن τ_t در رابطه مذکور، باید نرمال‌سازی را انجام شود تا پرتفوی دوره بدست آید. طبق (54-2)، $\tau_t \geq 0$ و \bar{x}_t بازده بازار است. پس عبارت $x_t - \bar{x}_t \mathbf{1}$ نشان‌دهنده بازده اضافی سهام نسبت به بازده بازار در دوره t می‌باشد. همچنین علامت منفی نشان دهنده اصل بازگشت به میانگین است.

عبارت مشترک در هر سه نوع PAMR، $\frac{l_\varepsilon(b, x_t)}{\|x_t - \bar{x}_t \mathbf{1}\|^2}$ است. با فرض ثابت در نظر گرفتن سایر متغیرها، چنانچه بازدهی دارای مقدار زیاد (کم) باشد، منجر به ایجاد مقداری بزرگ (کوچک) برای τ_t خواهد شد که باعث انتقال وزن بیشتر (کمتر) ثروت از دارایی‌های با عملکرد بهتر به دارایی‌های با عملکرد بدتر می‌گردد.

و اگر C که نشان دهنده ریسک است، بزرگ (کوچک) باشد، سبب می‌شود τ_t کوچک (بزرگ) شود. پس انتقال وزن کمتر (بیشتر) از دارایی‌های با عملکرد بهتر به دارایی‌های با عملکرد بدتر انجام می‌شود. پس همواره $0 \leq \tau_t \leq C$ خواهد بود که میزان وزن انتقالی را کنترل کرده و مانع از انتقال اوزان شدید می‌شود.

الگوریتم بازگشت به میانگین منفعلانه-تهاجمی که توسط لی و همکاران (۲۰۱۲) ارائه شد از این الگوریتم برای نشان دادن مفهوم بازگشت به میانگین از تابع زیان استفاده می‌کند.

۲-۳-۳-۲ الگوریتم ترکیبی منفعلانه-تهاجمی

این الگوریتم ارائه شده به وسیله گائو و ژانگ (۲۰۱۳)، برای کنترل وزن انتقالی از پرتفو یک دوره به پرتفو دوره بعد از یک تابع زیان چند ضابطه‌ای^{۴۸} طبق رابطه زیر استفاده می‌شود:

$$l_{\varepsilon}(b, x_t) = \begin{cases} 0 & b \cdot x_t \geq 1 + \frac{\varepsilon}{2} \text{ or } b \cdot x_t \leq 1 - \frac{\varepsilon}{2} \\ b \cdot x_t - \varepsilon & 1 - \frac{\varepsilon}{2} < b \cdot x_t \leq 1 \\ b \cdot x_t - \frac{\varepsilon}{3} & 1 < b \cdot x_t \leq 1 + \frac{\varepsilon}{3} \\ b \cdot x_t - \frac{\varepsilon}{6} & 1 + \frac{\varepsilon}{3} < b \cdot x_t < 1 + \frac{\varepsilon}{2} \end{cases} \quad (۵۴-۲)$$

و بهینه‌سازی در الگوریتم PACS مانند الگوریتم PAMR می‌باشد. در ضابطه فوق اگر از بالا به پایین بیاییم، به معنای بازدهی بیشتر می‌باشد که باعث بزرگتر شدن تابع زیان و در نتیجه بیشتر شدن مقدار τ_t خواهد شد. همین امر باعث انتقال وزن بیشتری از پرتفو یک دوره به پرتفو دوره بعد می‌گردد. دلیل استفاده از عبارت «ترکیبی» در عنوان این الگوریتم، به جمله اول در تابع زیان مرتبط است به طوریکه اگر $b \cdot x_t \geq 1 + \frac{\varepsilon}{2}$ باشد، تابع زیان دارای مقدار صفر می‌شود و در نتیجه پرتفو دوره بعد نیز بدون تغییر نسبت به پرتفو فعلی می‌باشد. (اصل مومنوم).

در الگوریتم ترکیبی منفعلانه-تهاجمی که توسط گائو و ژانگ (۲۰۱۳) ارائه شد، از یک تابع زیان چند ضابطه‌ای به منظور کنترل وزن انتقالی از پرتفو یک دوره به پرتفو دوره بعد استفاده می‌شود.

۲-۳-۳-۲ الگوریتم بازگشت به میانگین اطمینان وزنی

این الگوریتم توسعه‌ای از الگوریتم PAMR می‌باشد و توسط لی و همکاران (۲۰۱۳) ارائه گردیده که از تکنیک یادگیری اطمینان وزنی استفاده می‌کند. در این الگوریتم، حفظ توزیع گاوسی^{۴۹} و یا به صورت سلسله مراتبی

⁴⁸ Piecewise Loss Function

⁴⁹ Gaussian Distribution

به روز رسانی کردن توزیع مطابق با یادگیری منفعلانه-تهاجمی می باشد. الگوریتم CWMR بردار پرتفو را به صورت یک توزیع گاوسی مدل سازی می کند و به صورت سلسله مراتبی آن را مطابق با اصل بازگشت به میانگین به روز رسانی می نماید. برای توضیح بیشتر میتوان چنین عنوان کرد:

b را به عنوان یک توزیع گاوسی با میانگین $\mu \in R^m$ و ماتریس کوواریانس قطری $\Sigma \in R^{m \times m}$ در نظر بگیرید. در Σ عناصر قطری غیر صفر و عناصر غیر قطری صفر هستند.

در ابتدای دوره t ، در نظر می گیریم $b \sim N(\mu, \Sigma)$. پس از مشخص شدن x_t ، ثروت با عامل $b^\top x_t$ افزایش پیدا می کند. چنانچه بازدهی به صورت $D = b^\top x_t$ تعریف گردد، در نتیجه:

$$D \sim N(c, x_t^\top \Sigma x_t) \quad (55-2)$$

مطابق با اصل بازگشت به میانگین احتمال یک پرتفو سودآور b با توجه به آستانه بازگشت به میانگین ε طبق رابطه زیر می باشد:

$$Pr[D \leq \varepsilon] = Pr[b^\top x_t \leq \varepsilon] \quad (56-2)$$

اما الگوریتم به دنبال آن است تا اطمینان حاصل کند که احتمال کم بودن بازده پرتفو از مقدار مشخص ε از مقدار معین θ بیشتر باشد. $\theta \in [0,1]$ پارامتر سطح اطمینان^{۵۰} است. یعنی:

$$Pr[b^\top x_t \leq \varepsilon] \geq \theta \quad (57-2)$$

این رابطه بر اساس اصل بازگشت به میانگین می باشد.

در ادامه مطابق تکنیک منفعلانه-تهاجمی، الگوریتم CWMR نزدیک ترین توزیع را به توزیع فعلی مطابق با اصل واگرایی کولبک-لیبلر^{۵۱} بر می گزیند.

⁵⁰ Confidence Level

⁵¹ Kullback and Leibler Divergence

در الگوریتم فوق مسئله بهینه‌سازی به صورت رابطه زیر می‌باشد:

$$\begin{aligned}
 (\mu_{t+1}, \Sigma_{t+1}) &= \arg \min D_{KL}\{N(\mu, \Sigma) \| N(\mu_t, \Sigma_t)\} \\
 s. t. \quad Pr[b^\top x_t \leq \varepsilon] &\geq \theta \\
 \mu &\in \Delta_m
 \end{aligned}
 \tag{۵۸-۲}$$

و می‌توان نوشت:

$$\begin{aligned}
 &D_{KL}\{N(\mu, \Sigma) \| N(\mu_t, \Sigma_t)\} \\
 &= \frac{1}{2} \left(\log \left(\frac{\det \Sigma_t}{\det \Sigma} \right) + Tr(\Sigma_t^{-1} \Sigma) + (\mu_t - \mu)^\top \Sigma_t^{-1} (\mu_t - \mu) - d \right)
 \end{aligned}
 \tag{۵۹-۲}$$

و

$$Pr[D \leq \varepsilon] = pr \left[\frac{D - \mu_D}{\sigma_D} \leq \frac{\varepsilon - \mu_D}{\sigma_D} \right]
 \tag{۶۰-۲}$$

متغیر تصادفی دارای توزیع نرمال است و بنابراین احتمال آن برابر است با $\Phi\left(\frac{\varepsilon - \mu_D}{\sigma_D}\right)$ که Φ تابع توزیع تجمعی نرمال است. بنابراین می‌توان محدودیت را به صورت $\Phi^{-1}(\theta) \geq \frac{\varepsilon - \mu_D}{\sigma_D}$ بازنویسی کرد. عبارت فوق را به صورت رابطه می‌توان بیان کرد:

$$\varepsilon - \mu^\top x_t \geq z \sqrt{x_t^\top \Sigma x_t}
 \tag{۶۱-۲}$$

که در آن $z = \Phi^{-1}(\theta)$ و در ادامه داریم:

$$(\mu_{t+1}, \Sigma_{t+1}) = \frac{1}{2} \left(\log \left(\frac{\det \Sigma_t}{\det \Sigma} \right) + Tr(\Sigma_t^{-1} \Sigma) + (\mu_t - \mu)^\top \Sigma_t^{-1} (\mu_t - \mu) \right)
 \tag{۶۲-۲}$$

$$s. t. \quad \varepsilon - \mu^\top x_t \geq z \sqrt{x_t^\top \Sigma x_t}$$

$$\mu^T \mathbf{1} = 1, \mu \geq 0$$

لی و همکاران (۲۰۱۳) الگوریتم بازگشت به میانگین اطمینان وزنی را به عنوان توسعه‌ای از الگوریتم PAMR ارائه کردند که از تکنیک یادگیری اطمینان وزنی ارائه شده توسط کرامر و همکاران (۲۰۰۸) و دریدزه و همکاران (۲۰۰۸) استفاده می‌کند. ایده اصلی تکنیک یادگیری اطمینان وزنی، حفظ توزیع گاوسی و یا به صورت سلسله مراتبی به روز رسانی کردن توزیع مطابق با یادگیری منفعلانه-تهاجمی است.

۲-۳-۵- الگوریتم بازگشت به میانگین متوسط متحرک وزن دار

هدف این الگوریتم، استفاده از نوسانات قیمت‌های سهام در طی پنجره زمانی مشخص و نه فقط آخرین روز معاملاتی است و توسط گائو و ژانگ (۲۰۱۳) ارائه گردیده است. می‌توان گفت الگوریتم‌های PAMR و CWMR استراتژی‌های انتخاب سبد تک دوره‌ای می‌باشند و همین مسئله سبب کاهش عملکرد آن‌ها می‌گردد. این الگوریتم پیش‌بینی بردار نسبت قیمتی را برای دوره مطابق رابطه زیر بدست می‌آورد:

$$\tilde{x}_{t+1} = \sum_{i=1}^w \omega_i x_{t-i+1} \quad (2-63)$$

که در آن $(\omega_1, \dots, \omega_w)$ بردار وزن‌ها است. ادامه حل مانند PAMR می‌باشد.

۲-۳-۶- الگوریتم بازگشت به میانگین متحرک برخط

الگوریتم‌های موجود از بازگشت به میانگین تک دوره‌ای استفاده می‌کنند که در برخی موارد موجب عملکرد ضعیف الگوریتم می‌شود. برای حل این مشکل لی و همکاران (۲۰۱۵) از بازگشت به میانگین چند دوره‌ای در کنار تکنیک‌های یادگیری برخط استفاده کردند که آن را بازگشت به میانگین متحرک برخط گویند.

اگر \tilde{x}_{t+1}^i ، i مین پیش‌بینی سرمایه‌گذار از بردار نسبت قیمتی دوره $t+1$ و p_i احتمال هر یک از این پیش‌بینی‌ها باشد، در نتیجه تابع هدف لگاریتمی به صورت رابطه می‌شود:

$$b_{t+1} = \arg \max \sum_{i=1}^k p_i \log(b \cdot \tilde{x}_{t+1}^i) \quad (۶۴-۲)$$

در الگوریتم‌های PAMR و CWMR:

$$\tilde{x}_{t+1}^1 = \frac{1}{x_t} \text{ with } p_1 = 100\%$$

در نتیجه:

$$\tilde{x}_{t+1} = \frac{1}{x_t} \rightarrow \frac{\tilde{p}_{t+1}}{p_t} = \frac{p_{t-1}}{p_t} \rightarrow \tilde{p}_{t+1} = p_{t-1}$$

اما الگوریتم OLMAR به جای $\tilde{p}_{t+1} = p_{t-1}$ فرض می‌کند بردار نسبت قیمتی در دوره $t + 1$ به میانگین متحرک نسبت‌های قیمتی باز خواهد گشت یعنی $\tilde{p}_{t+1} = MA_t$ که MA_t میانگر میانگین متحرک تا انتهای دوره t است. در آنالیز سری‌های زمانی، از میانگین متحرک برای هموارسازی نوسانات قیمتی در کوتاه‌مدت استفاده می‌گردد.

گائو و ژانگ (۲۰۱۳) الگوریتم بازگشت به میانگین متوسط متحرک وزن‌دار را ارائه کردند که ایده اصلی آن، بهره بردن از نوسانات قیمت‌های سهام در طی پنجره زمانی مشخص و نه فقط آخرین روز معاملاتی است. در واقع الگوریتم‌های PAMR و CWMR استراتژی‌های انتخاب سبد تک دوره‌ای هستند که همین مسئله سبب کاهش عملکرد آن‌ها می‌شود. طبق بیان نویسندگان این پژوهش، این الگوریتم‌ها نمی‌توانند به صورت کامل اثر نوسانات قیمتی در بازارهای سهام را پوشش دهند.

۲-۳-۷- الگوریتم بازگشت به میانه استوار

این الگوریتم که توسط هوانگ و همکاران (۲۰۱۳) ارائه شد از تخمین‌زننده استوار میانه L_1 ^{۵۲} جهت پیش‌بینی قیمت دوره بعد استفاده خواهد شد. در واقع، پیش‌بینی قیمت در روش‌های ایجادشده تاکنون دارای دو مشکل اساسی می‌باشد: اول اینکه فرض تک دوره‌ای بودن در روش‌های PAMR و CWMR در دنیای واقعی نتایج

⁵² L1-median Estimator

قابل قبولی ندارد. و دوم اینکه در همه الگوریتم‌ها، وجود نویزها و داده‌های پرت عملکرد الگوریتم را تحت تأثیر قرار خواهد داد. اما در الگوریتم RMR بجای آنکه $\tilde{p}_{t+1} = p_{t-1}$ و یا $\tilde{p}_{t+1} = MA_t(\omega)$ باشد، از تخمین‌زننده استوار میانه L_1 در انتهای دوره t استفاده می‌شود یعنی $\tilde{p}_{t+1} = L_1 med_{t+1}(\omega) = \mu$ که در آن ω اندازه پنجره زمانی و μ مقدار تخمین‌زننده استوار میانه L_1 می‌باشد که از حل مسئله بهینه‌سازی طبق رابطه زیر بدست می‌آید:

$$\mu = \arg \min \sum_{i=0}^{\omega-1} \|p_{t-i} - \mu\| \quad (۶۵-۲)$$

در نتیجه μ نقطه‌ای با کمترین فواصل اقلیدسی^{۵۳} از ω نقطه قیمتی داده شده است. پس از حل مسئله بهینه‌سازی فوق، می‌توان \tilde{x}_{t+1} را طبق رابطه زیر بدست آورد:

$$\tilde{x}_{t+1} = \frac{\mu}{p_t} \quad (۶۶-۲)$$

در الگوریتم RMR مشابه سایر الگوریتم‌های موجود، از تکنیک یادگیری منفعلانه-تهاجمی استفاده خواهد شد.

پس می‌توان بیان کرد که الگوریتم RMR، نه تنها از بازگشت به میانگین چند دوره‌ای استفاده می‌کند بلکه به نحوی با استفاده از تخمین‌زننده استوار میانه L_1 به دنبال حذف داده‌های پرت می‌باشد.

هانگ و همکاران (۲۰۱۳) الگوریتم بازگشت به میانه استوار که در آن از تخمین‌زننده استوار میانه L_1 جهت پیش‌بینی قیمت دوره بعد استفاده می‌شود. در واقع، پیش‌بینی قیمت در روش‌های ایجادشده تاکنون دارای دو مشکل اساسی است: ۱) فرض تک دوره‌ای بودن در روش‌های PAMR و CWMR در دنیای واقعی نتایج قابل قبولی در بر نخواهد داشت. ۲) در همه الگوریتم‌ها، وجود نویزها و داده‌های پرت عملکرد الگوریتم را تحت تأثیر قرار می‌دهد. این دو مشکل دلیلی برای ایجاد این الگوریتم شد. در الگوریتم RMR بجای آنکه $\tilde{p}_{t+1} = p_{t-1}$ و یا $\tilde{p}_{t+1} = MA_t(\omega)$ باشد، از تخمین‌زننده استوار میانه L_1 در انتهای دوره t استفاده می‌شود.

⁵³ Euclidean Distance

جدول ۲-۲- خلاصه‌ای از ادبیات موجود در زمینه الگوریتم‌های تبعیت از بازنده

الگوریتم‌های تبعیت از بازنده									
ردیف	نویسندگان	سال انتشار	نام الگوریتم	تابع زیان			بازگشت به میانگین		تکنیک یادگیری
				تک ضابطه‌ای	چند ضابطه‌ای	عدم استفاده	تک دوره-ای	چند دوره-ای	
۱	برودین و همکاران	۲۰۰۴	Anticor			✓		✓	اطمینان وزنی
۲	لی و همکاران	۲۰۱۲	PAMR	✓			✓		منفعلا نه-تهاجمی
۳	گائو و ژانگ	۲۰۱۳	PACS		✓		✓		
۴	لی و همکاران	۲۰۱۳	CWMR	✓			✓		✓
۵	گائو و ژانگ	۲۰۱۳	WMAMR			✓		✓	✓
۶	لی و همکاران	۲۰۱۲	OLMAR			✓		✓	✓
۷	هانگ و همکاران	۲۰۱۳	RMR			✓		✓	✓
۸	زیجین	۲۰۱۶	WMAAR		✓			✓	✓
۹	چانگ چو	۲۰۱۹			✓				✓
۱۰	ولیدی و نجفی	۲۰۲۰				✓		✓	✓

۲-۳-۴- الگوریتم تطابق با الگو

دسته‌ای از استراتژی‌ها وجود دارند که از هر دوی استراتژی‌های تبعیت از برنده و تبعیت از بازنده استفاده می‌کنند که آنها را استراتژی‌های تطابق با الگو می‌نامند. این الگوریتم دربرگیرنده استراتژی‌های سرمایه‌گذاری پیوسته غیر پارامتریک است که رشد بهینه سرمایه را تحت کمترین فرضیات تضمین می‌نماید. فرضیات این مدل، شامل ایستایی^{۵۴} و ارگودیک^{۵۵} بودن سری زمانی داده‌های مالی می‌باشد. این الگوریتم‌ها به صورت تجربی، از الگوهای تکراری در طی زمان‌های گذشته استفاده می‌کنند و دارای عملکرد تجربی قابل قبولی نیز می‌باشند. از مشکلات مهم در این استراتژی، شناسایی الگوهای تکراری بوده که منجر به ایجاد پژوهش‌های متعددی در

⁵⁴ Stationary

⁵⁵ Ergodic

این زمینه شده است. طبق نظر گیورفی و همکاران (۲۰۰۶) الگوریتم‌های تطابق با الگو شامل دو گام انتخاب نمونه و بهینه‌سازی پرتفو است. انتخاب نمونه می‌تواند بر اساس همبستگی^{۵۶}، نزدیک‌ترین همسایه^{۵۷}، هیستوگرام^{۵۸} و کرنل^{۵۹} باشد.

گام دوم در الگوریتم‌های تطابق با الگو انتخاب پرتفوی بهینه بر اساس مجموعه G_t است. دو رویکرد عمده در انتخاب پرتفو وجود دارد؛ تئوری رشد سرمایه کلی (۱۹۵۶) و پرتفوی میانگین-واریانس مارکوویتز (۱۹۵۲).

گیورفی و همکاران (۲۰۰۶) روشی را برای انتخاب پرتفو بر اساس رویکرد تئوری رشد سرمایه معرفی کردند که در واقع لگاریتم بهینه است. گیورفی و همکاران (۲۰۰۷) برای حل مشکل محاسباتی روش بالا، روش نیمه لگاریتم بهینه را ارائه دادند که در تابع مطلوبیت از تقریب لگاریتم استفاده می‌شود وایدا (۲۰۰۶) تحلیل نظری این روش را ارائه کرد و یونیورسال بودن آن را اثبات نمود. گیورفی و همکاران (۲۰۰۷) روش فوق را با در نظر گرفتن تابع توزیع یکنواخت تطبیق دادند. اتوساک و وایدا (۲۰۰۷) تابع مطلوبیت بر مبنای رویکرد مارکوویتز ارائه دادند که بیشتر به استراتژی نیمه لگاریتم بهینه عمومیت بخشید. ایده اصلی این رویکرد، مشابه رویکرد مارکوویتز، موازنه بین میانگین (بازدهی) و واریانس (ریسک) است. گیورفی و وایدا (۲۰۰۸)، تابع مطلوبیتی را که دربرگیرنده فاکتور هزینه‌های معاملاتی است، ارائه کردند و آن را GV-Type نامیدند. لی و همکاران (۲۰۱۱) الگوریتمی را ارائه کردند که از تکنیک انتخاب نمونه همبستگی و تکنیک بهینه‌سازی پرتفو لگاریتم بهینه استفاده کردند. آن‌ها این الگوریتم را CORN نامیدند. مطالعات تجربی نشان داده است که این الگوریتم در مقایسه با سایر الگوریتم‌ها که از پرتفو لگاریتم بهینه و سایر تکنیک‌های انتخاب نمونه استفاده کردند، دارای عملکرد بهتری است. اورموس و اوربان (۲۰۱۳) عملکرد استراتژی‌های لگاریتم بهینه را تحت شرایط هزینه معاملاتی به صورت تجربی تحلیل کردند.

⁵⁶ Correlation-driven

⁵⁷ Nearest Neighbor-based

⁵⁸ Histogram-based

⁵⁹ Kernel-based

عبدی (۱۳۹۶) در مقاله خود برای انتخاب نمونه از تکنیک خوشه‌بندی طیفی که یکی از تکنیک‌های نوین در مبحث داده‌کاوی است، استفاده کرد. وانگ و همکاران (۲۰۱۸) نوعی از استراتژی تطابق با الگو را ارائه دادند که ریسک را در بهینه‌سازی پرتفوی خود در نظر می‌گیرد. آن‌ها دو نوع استراتژی RA-CORN-K و RA-CORN(C)-K را ارائه دادند که توسعه‌ای از استراتژی CORN ارائه شده توسط لی و همکاران (۲۰۱۱) است. نتایج تجربی بر روی الگوریتم‌های فوق نشان داده است که در معیارهای ارزیابی شامل ثروت نهایی، شارپ و حداکثر افت سرمایه پیشرفت‌های چشمگیری حاصل شده است.

تکنیک‌های انتخاب نمونه

ایده کلی این گام، انتخاب نمونه‌های مشابه از مجموعه بردارهای نسبت قیمتی است که این امر با مقایسه دو پنجره زمانی گذشته بازار محقق می‌شود. در این گام، پنجره زمانی ω و همچنین دنباله تاریخی بازار x_1^t را به‌عنوان ورودی به الگوریتم می‌دهیم. الگوریتم نیز اندیس‌هایی را که در آن دو پنجره زمانی $x_{t-\omega+1}^t$ و $x_{t-\omega}^{t-1}$ مشابه باشند به‌عنوان خروجی در اختیار کاربر قرار می‌دهد. در شکل ۲-۶، الگوریتم انتخاب نمونه آورده شده است.

جدول ۲-۳ انتخاب نمونه $(C(x_1^t, \omega))$

ورودی: X_1^t : دنباله تاریخی بازار و ω : اندازه پنجره زمانی

خروجی: C_t : مجموعه اندیس‌های با نسبت‌های قیمتی مشابه

قرار بده: $C_t = \emptyset$

اگر $t \leq \omega$ آنگاه

از ابتدا شروع کن؛

پایان

به ازای $i = \omega + 1, \omega + 2, \dots, t$

اگر $X_{i-\omega}^{i-1}$ مشابه با $X_{t-\omega+1}^t$ است آنگاه

$$C_t = C_t \cup i$$

پایان

پایان

بر اساس هیستوگرام^{۶۰}

در این روش که توسط گیورفی و همکاران (۲۰۰۶) ارائه شد، تعدادی پارتیشن^{۶۱} به صورت گسسته تعریف می شود و در صورتی که پنجره زمانی اخیر ($X_{t-\omega+1}^t$) و پنجره زمانی تاریخی ($X_{i-\omega}^{i-1}$) در یک پارتیشن قرار گیرند، آن دو پنجره زمانی مشابه قلمداد می شوند.

فضای عددی مورد مطالعه به d پارتیشن گسسته مساوی تقسیم شده است و پارتیشن $p = A_j, j = 1, 2, \dots, d$ تعریف می شود. همچنین تابع $G(x) = j$ بیانگر شماره پارتیشنی است که x در آن قرار گرفته است. در صورتی که شرط زیر برقرار باشد، آنگاه دو پنجره زمانی مشابه هستند:

$$C_H(x_1^t, \omega) = \{\omega < i < t + 1 : G_L(X_{t-\omega+1}^t) = G_L(X_{i-\omega}^{i-1})\} \quad (۶۷-۲)$$

بر اساس کرنل^{۶۲}

در این روش که نیز توسط گیورفی و همکاران (۲۰۰۶) ارائه شد، فاصله اقلیدسی دو پنجره زمانی محاسبه می شود و اگر از مقدار $\frac{c}{l}$ کمتر باشد دو پنجره زمانی مشابه قلمداد می شوند. در این تکنیک، داریم:

⁶⁰ Histogram-based

⁶¹ Partition

⁶² Kernel-based

$$C_K(x_1^t, \omega) = \left\{ \omega < i < t + 1 : \left\| (X_{t-\omega+1}^t) - (X_{i-\omega}^{i-1}) \right\| \frac{c}{l} \right\} \quad (68-2)$$

که در آن، c و l آستانه‌هایی برای کنترل تعداد نمونه‌های مشابه هستند.

بر اساس نزدیک‌ترین همسایه^{۶۳}

این روش که توسط گیورفی و همکاران (۲۰۰۷) ارائه شد، به دنبال نسبت‌های قیمتی است که جزء k پنجره زمانی نزدیک‌تر به پنجره اخیر قرار گیرد. معیار نزدیکی در این روش، فاصله اقلیدسی است. و داریم:

$$C_N(x_1^t, \omega) = \left\{ \omega < i < t + 1 : (X_{i-\omega}^{i-1}) \text{ is among the } k \text{ NNs of } (X_{t-\omega+1}^t) \right\} \quad (69-2)$$

که در آن، k یک پارامتر آستانه^{۶۴} است.

بر اساس همبستگی^{۶۵}

در این روش، معیار تشابه ضریب همبستگی^{۶۶} پنجره زمانی اخیر و پنجره زمانی تاریخی است. طبق رابطه زیر:

$$C_C(x_1^t, \omega) = \left\{ \omega < i < t + 1 : \frac{\text{cov}(X_{i-\omega}^{i-1}, X_{t-\omega+1}^t)}{\text{Std}(X_{i-\omega}^{i-1}) \text{Std}(X_{t-\omega+1}^t)} \geq \rho \right\} \quad (70-2)$$

که در آن ρ ضریب همبستگی بوده و پارامتر آستانه محسوب می‌شود. بدین‌صورت که اگر همبستگی بین دو پنجره زمانی از مقدار ρ بیشتر شود، آن دو پنجره مشابه قلمداد شوند.

اما گام دوم که بهینه‌سازی پرتفو می‌باشد که در ادامه به آن می‌پردازیم. در این گام یک پرتفو بهینه بر اساس شباهت با مجموعه به‌دست‌آمده در گام اول را بر اساس تابع هدف از رابطه زیر بدست می‌آید:

$$b_{t+1} = \arg \max U(b; C_t) \quad (71-2)$$

⁶³ Nearest Neighbor-based

⁶⁴ Threshold Parameter

⁶⁵ Correlation-driven

⁶⁶ Correlation Coefficient

که در آن $U(\cdot)$ یک تابع مطلوبیت^{۶۷} خاص مانند تابع مطلوبیت لگاریتمی می‌باشد. در صورتی که هیچ شباهتی با مجموعه وجود نداشته باشد، پرتفوی یکنواخت به عنوان پرتفوی بهینه انتخاب می‌شود.

۲-۳-۴-۱- تکنیک‌های بهینه‌سازی پرتفو

در اینجا انتخاب پرتفوی بهینه بر اساس مجموعه C_t بوده که به طور کلی دو رویکرد انتخاب پرتفو وجود دارد؛ تئوری رشد سرمایه کلی^{۶۸} (۱۹۵۶) و پرتفوی میانگین-واریانس مارکوییتز^{۶۹} (۱۹۵۲).

لگاریتم بهینه^{۷۰} روشی است که بر اساس رویکرد تئوری رشد سرمایه معرفی شده است و تابع مطلوبیت در این روش به صورت زیر است:

$$U_L(b, C_t) = E\{\log b \cdot x | x_i, i \in C_t\} = \sum_{i \in C_t} P_i \log b \cdot x_i \quad (۷۲-۲)$$

که P_i بیانگر احتمالی است که به هر x_i و $i \in C_t$ نسبت داده می‌شود و اگر از توزیع احتمال یکنواخت استفاده شود، رابطه فوق مطابق با رابطه زیر است:

$$U_L(b, C_t) = \sum_{i \in C_t} \log b \cdot x_i \quad (۷۳-۲)$$

ماکزیمم‌سازی تابع مطلوبیت فوق منجر به ایجاد پرتفوی BCRP در سراسر نسبت‌های قیمتی مشابه می‌گردد.

پس از آن روش نیمه لگاریتم بهینه مطرح شد که تابع مطلوبیت آن به صورت رابطه زیر است:

$$U_S(b, C_t) = E\{f(b \cdot x) | x_i, i \in C_t\} = \sum_{i \in C_t} P_i f(b \cdot x_i) \quad (۷۴-۲)$$

که در آن $f(\cdot)$ بیان‌کننده بسط تیلور مرتبه دوم $\log z$ به ازای $z = 1$ است.

$$f(z) = z - 1 - \frac{1}{2}(z - 1)^2 \quad (۷۵-۲)$$

⁶⁷ Utility Function

⁶⁸ Kelly's Capital Growth Portfolio

⁶⁹ Markowitz's Mean Variance Portfolio

⁷⁰ Log-Optimal

همچنین می‌توان روش فوق را با در نظر گرفتن تابع توزیع یکنواخت تطبیق داد که بر اساس آن:

$$U_S(b, C_t) = \sum_{i \in C_t} f(b \cdot x_i) \quad (۷۶-۲)$$

سپس تابع مطلوبیت بر مبنای رویکرد مارکوییتز ارائه شد که بیشتر به استراتژی نیمه لگاریتم بهینه عمومیت بخشید. ایده اصلی این رویکرد، مشابه رویکرد مارکوییتز، موازنه بین میانگین (بازدهی) و واریانس (ریسک) است. تابع مطلوبیت این رویکرد به صورت رابطه (۶۹-۲) عبارتست از:

$$U_M(b, C_t) = E\{b \cdot x | x_i, i \in C_t\} - \lambda Var\{b \cdot x | x_i, i \in C_t\} \quad (۷۷-۲)$$

$$= E\{b \cdot x | x_i, i \in C_t\} - \lambda E\{(b \cdot x)^2 | x_i, i \in C_t\} + \lambda (E\{b \cdot x | x_i, i \in C_t\})^2$$

که در آن λ پارامتر موازنه^{۷۱} می‌باشد.

⁷¹ Trade-off Parameter

جدول ۲-۳- خلاصه‌ای از ادبیات موجود در زمینه الگوریتم‌های تطابق با الگو

الگوریتم‌های تطابق با الگو												
ردیف	نویسندگان	سال انتشار	تکنیک انتخاب نمونه							تکنیک بهینه‌سازی		
			هیستوگرام	کرنل	نزدیک‌ترین همسایه	همبستگی	خوشه-بندی	لگاریتم بهینه	نیمه لگاریتم بهینه	مارکویتز	GV	یونیورسال
۱	گیورفی و شافر	۲۰۰۳	✓					✓				✓
۲	گیورفی و همکاران	۲۰۰۶		✓				✓				✓
۳	گیورفی و همکاران	۲۰۰۸			✓			✓				✓
۴	لی و همکاران	۲۰۱۱				✓		✓				✓
۵	گیورفی و همکاران	۲۰۰۷		✓					✓			✓
۶	اتوساک و وایدا	۲۰۰۷		✓						✓		
۷	گیورفی و وایدا	۲۰۰۸		✓							✓	✓
۸	عبدی و نجفی	۱۳۹۶					✓	✓				
۹	وانگ و همکاران	۲۰۱۸				✓		✓				✓
۹	خدمتی و آذین	۲۰۲۰					✓	✓				

۲-۳-۵- الگوریتم‌های فرا یادگیری

از دیگر الگوریتم‌ها در حوزه انتخاب آنلاین پرتفو، الگوریتم‌های فرا یادگیری هستند که مرتبط با یادگیری اکسپرت‌ها می‌باشد. به‌طور کلی در الگوریتم فرا یادگیری، تعدادی اکسپرت در نظر گرفته می‌شوند که

استراتژی‌های آن‌ها می‌تواند از دسته استراتژی‌های مشابه یا متفاوت باشد. هر اکسپرت یک پرتفو برای دوره بعد ارائه می‌دهد و در نهایت الگوریتم فرا یادگیری این پرتفوها را ترکیب کرده تا پرتفوی نهایی برای دوره آینده را ارائه دهد. سیستم کلی توصیف‌شده می‌تواند بهترین عملکرد را در بین اکسپرت‌ها داشته باشد و این مسئله نیز برای پرتفوهایی که یونیورسال نیستند، می‌تواند مناسب باشد. الگوریتم‌های فرا یادگیری مشابه الگوریتم UP در رویکرد تبعیت از برنده هستند با این تفاوت که این الگوریتم‌ها حیطه وسیع‌تری از دسته‌های اکسپرت‌ها را شامل می‌شوند و CRP که در الگوریتم UP استفاده می‌شود، می‌تواند یک مورد خاص از آن باشد. بنابراین در این نوع الگوریتم‌ها، الزامی وجود ندارد که اکسپرت‌ها صرفاً از یک دسته استراتژی خاص نظیر CRP در انتخاب سبد استفاده کنند.

و همچنین الگوریتم‌های فرا یادگیری می‌توانند عملکرد نهایی را با در نظر گرفتن همه اکسپرت‌ها هموار کنند مخصوصاً زمانی که اکسپرت‌ها نسبت به پارامترها و یا محیط‌هایی حساسیت داشته باشند. از سوی دیگر با ترکیب استراتژی‌های یونیورسال و الگوریتم‌های ابتکاری مانند الگوریتم ضد همبستگی، می‌توان مشخصه یونیورسال بودن را برای کل الگوریتم فرا یادگیری ایجاد نمود. همچنین می‌توان الگوریتم فرا یادگیری را در همه الگوریتم‌های موجود ترکیب کرد تا بتوان گستره وسیع‌تری از کاربردها را ایجاد نمود و الگوریتم‌های جدیدی را ساخت.

۲-۳-۵-۱- الگوریتم‌های جمع‌کننده

وووک و واتکینز^{۷۲} (۱۹۹۸) الگوریتم جمع‌کننده را در مبحث انتخاب برخط سبد سهام ارائه کردند. الگوریتم‌های جمع‌کننده متعددی ارائه شده‌اند تا مجموعه متفاوتی از اکسپرت‌ها را شامل شوند. اکسپرت‌ها در این دسته‌بندی، به‌جای این‌که متعلق به اکسپرت‌های پیچیده از دسته‌های مختلف باشند، متعلق به یک دسته خاص هستند. در این الگوریتم می‌بایست مجموعه مشخصی از اکسپرت‌ها را تعیین کرده و ثروت را به‌صورت

⁷² Vovk and Watkins

سلسله مراتبی به اکسپرت‌های مختلف تخصیص داده تا عملکرد مطلوبی حاصل گردد که بدتر از هیچ ترکیب ثابتی از اکسپرت‌ها نباشد. فرمول به‌روز رسانی پرتفو در این الگوریتم به‌صورت رابطه زیر می‌باشد:

$$b_{t+1} = \frac{\int_{\Delta m} b \prod_{i=1}^{t-1} (b \cdot x_t)^\eta P_0(db)}{\int_{\Delta m} \prod_{i=1}^{t-1} (b \cdot x_t)^\eta P_0(db)} \quad (78-2)$$

به‌عنوان یک مورد خاص، الگوریتم UP ارائه‌شده توسط کاور مطابق با الگوریتم جمع‌کننده و با توزیع اولیه یکنواخت و $\eta = 1$ است.

اکسپرت‌ها در این دسته‌بندی، به‌جای این‌که متعلق به اکسپرت‌های پیچیده از دسته‌های مختلف باشند، متعلق به یک دسته خاص هستند. سینگر (۱۹۹۷) پرتفوی تعویضی (به‌اختصار، SP) را ارائه کرد که در میان مجموعه‌ای از استراتژی‌هایی که رژیم‌های مختلف را اختیار می‌کنند سوئیچ می‌کند. وووک و واتکینز (۱۹۹۸) الگوریتم جمع‌کننده را در مبحث انتخاب برخط سبد سهام ارائه کردند که در آن الگوریتم UP یک مورد خاص محسوب می‌شود. کوزات و سینگر (۲۰۰۷) SP را توسعه داده و یک استراتژی کسر ثابت چند ضابطه‌ای ارائه دادند که دوره‌های زمانی را به بخش‌های مختلف تقسیم‌بندی کرده و بین این بخش‌ها نیز حرکت می‌کند.

لوینا و شافر (۲۰۰۸) استراتژی گام تصادفی گاوسی را ارائه کردند که در آن بین استراتژی‌های مختلف مطابق با توزیع گاوسی سوئیچ می‌کند. کوزات و سینگر (۲۰۰۸) مدل قبلی خود را توسعه داده و آن را با در نظر گرفتن هزینه‌های معاملاتی حل کردند. کوزات و همکاران (۲۰۰۸) استراتژی انتخاب سبد یونیورسال چند ضابطه‌ای را ارائه کردند که در آن از مفهوم درخت تصمیم استفاده شده است. کوزات و سینگر (۲۰۰۹) مدل خود را به مسئله‌ای که در آن نیاز به تصمیم‌گیری سلسله‌مراتبی است توسعه دادند.

۲-۳-۵-۲- یونیورسال شدن سریع

آکوگلو و همکاران (۲۰۰۵) استراتژی UP کاور را توسعه دادند و استراتژی یونیورسال شدن سریع را ارائه کردند. در استراتژی یونیورسال شدن سریع به‌جای استفاده از استراتژی CRP، از استراتژی‌های دیگری مانند استراتژی خرید تک سهم و استراتژی‌هایی که سرمایه را به کل بازار اختصاص می‌دهند، استفاده

می‌شود. ایده اصلی در این استراتژی، تقسیم مساوی سرمایه بین اکسپرت‌ها بوده و هر اکسپرت‌ها استراتژی خاص خود را دنبال کرده و طبق آن عمل می‌کنند و در انتها ثروت هر کدام جمع می‌شود تا ثروت نهایی کل الگوریتم به دست آید. به روز رسانی استراتژی یونیورسال شدن سریع مانند استراتژی UP است و نرخ رشد دارایی حاصل شده مجانباً به سمت ترکیب محدب ثابت و بهینه اکسپرت‌ها میل می‌نماید. در مواردی که در آن همه اکسپرت‌ها CRP هستند، استراتژی معادل با استراتژی UP می‌شود. در این استراتژی پرتفو به صورت رابطه زیر محاسبه می‌شود:

$$b_t = \frac{\int_W S_t(w) R_t(w) d\mu(w)}{\int_W R_t(w) d\mu(w)} \quad (2-79)$$

که در آن، $R_0(w) = 1$.

علاوه بر یونیورسال شدن در فضای پیوسته، ترکیباتی از استراتژی‌های متنوع خرید و نگهداری گسسته نیز در بسیاری از این الگوریتم‌ها استفاده شده است.

این استراتژی به جای استفاده از استراتژی CRP، از استراتژی‌های دیگری مانند استراتژی خرید تک سهم و استراتژی‌هایی که سرمایه را به کل بازار تخصیص می‌دهند، استفاده می‌کند. ایده اصلی در این استراتژی، تقسیم مساوی سرمایه بین اکسپرت‌ها است.

آگوگلو و همکاران (۲۰۰۵) استراتژی UP کاور را توسعه دادند و استراتژی یونیورسال شدن سریع را ارائه کردند. بوردین و همکاران (۲۰۰۴) از استراتژی خرید و نگهداری برای ترکیب اکسپرت‌های ضد همبستگی و با توجه به پنجره‌های زمانی و با پارامترهای مشخص استفاده کردند. علاوه بر این، تمام الگوریتم‌های تطابق با الگو نیز از استراتژی BAH برای ترکیب اکسپرت‌های موجود استفاده کردند.

۲-۳-۵-۳- الگوریتم گرادیان برخط

الگوریتم به روز رسانی گرادیان برخط، بر اساس الگوریتم‌های گرادیان نمایی توسط داس و بانرجی (۲۰۱۱) ارائه شده است. همچنین به روز رسانی آن مانند الگوریتم EG می‌باشد. از نظر تئوری، الگوریتم دارای رشد نرخ

مشابه با ترکیب محدب و بهینه خبره‌ها می‌باشد. اگر هر یک از اکسپرت‌ها یونیورسال باشند، آنگاه کل الگوریتم نیز دارای این خاصیت می‌باشد. این مسئله سبب می‌شود الگوریتم فرا یادگیری بتواند الگوریتم ابتکاری و الگوریتم یونیورسال را ادغام کرده و بتواند از عملکرد خوب هر دو بهره‌مند شود

۲-۳-۵-۴- الگوریتم تبعیت از تاریخ راهنما^{۷۳}

این الگوریتم برای محیط‌های متغیر توسط هازان و سشادری^{۷۴} (۲۰۰۹) ارائه شده است. در اینجا ما از ترکیب چند الگوریتم یونیورسال استفاده می‌کنیم و همچنین این الگوریتم‌ها در دوره‌های مختلف ثابت نیستند و همچنین به‌صورت تجربی نشان داده شده است که عملکرد بهتری نسبت به ONS دارد.

۲-۳-۵-۵- الگوریتم گرادیان برخط

داس و بانرجی (۲۰۱۱) الگوریتم به‌روز رسانی گرادیان برخط را بر اساس الگوریتم‌های گرادیان نمایی ارائه دادند.

۲-۳-۶- دیگر الگوریتم‌ها

ژینجو و همکاران (۲۰۱۹) الگوریتم جمع آوری ضعیف WAACS را ارائه دادند. آنها می‌گویند چون بازارهای مالی به سرعت در حال تغییر هستند سرمایه‌گذاران باید به صورت پویا موقعیت دارایی خود را با توجه به اطلاعات مختلف بازار تنظیم کنند. این الگوریتم تمام استراتژی‌های CRP را با توجه به بازده تجمعی آنها در نظر می‌گیرد و همچنین اطلاعات جانبی را در مدل خود لحاظ می‌کند.

⁷³ Follow The Leading History

⁷⁴ Hazan and Seshadhri

۲-۶- الگوریتم HRP

مارکویتز قبل از آنکه دکترای خود را در سال ۱۹۵۴ کسب کند اولین مدل بهینه‌سازی پورتفو یعنی الگوریتم خط بحرانی^{۷۵} (CLA) را ارائه کرد به طوری که این الگوریتم به کمک بهینه‌سازی درجه دوم بر اوزان دارایی‌ها، پرتفوی بهینه‌ای با بیشینه نسبت شارپ به ما می‌دهد. کوهن و تاکر^{۷۶} (۲۰۱۴) اثبات کردند که همانند تمامی مدل‌های بهینه‌سازی درجه دوم، در صورت برقرار بودن شرایط کاروش-کوهن-تاکر^{۷۶} این الگوریتم بهینه‌سازی بعد از تعداد محدودی تکرار به جواب می‌رسد که این ویژگی یکی از مهم‌ترین مزیت‌های این مدل است. توضیحات بیشتر و یک نمونه به کارگیری این مدل در مقاله بیلی و دپرادو^{۷۷} (۲۰۱۳) یافت می‌شود.

بر خلاف نبوغی که در توسعه مدل بهینه‌سازی مارکویتز به کار رفته، به کارگیری آن در عمل با مشکلاتی مواجه است که از مهم‌ترین این مشکلات می‌توان به عدم ثبات اوزان در تشکیل دادن پرتفو اشاره کرد به طوری که همانطور که در مقاله میچاد^{۷۸} (۲۰۰۱) به آن اشاره شده با کوچک‌ترین تغییری در پیش‌بینی بازدهی‌های آینده، مدل مارکویتز یک پرتفوی کاملاً متفاوت با پورتفوی قبل به ما تحویل می‌دهد. با توجه به اینکه پیش‌بینی بازدهی دارایی‌ها با یک سطح اطمینان قابل قبول تقریباً ناممکن می‌باشد اکثر مدیران سبد به استفاده از ماتریس کواریانس^{۷۸} و مدل‌های ریسک‌پایه^{۷۹} روی آورده‌اند که از مهم‌ترین آنها می‌توان به مدل تعادل ریسک^{۸۰} اشاره کرد. که چاوز^{۸۱} (۲۰۱۱) استراتژی‌ای کاملاً بر مبنای کمینه کردن ریسک بدون استفاده از الگوریتم‌های بهینه‌سازی خطی و درجه دوم توسعه داده است. روش‌های دیگری مانند بلک لیترمین^{۸۱} (1991) هم در طول سال‌ها توسعه داده شده‌اند که سعی در بهبود نواقص مدل مارکویتز داشته‌اند اما هیچ کدام نتوانسته‌اند نتایجی به خوبی نتایج الگوریتم CLA داشته باشند.

⁷⁵ Critical Line Algorithm

⁷⁶ Karush-Kuhn-Tucker

⁷⁷ Marcos Lopez De Prado

⁷⁸ Covariance matrix

⁷⁹ Risk based

⁸⁰ Risk parity

⁸¹ Black-Litterman

ابتدا به نظر می‌رسید با استفاده نکردن از بازدهی‌های پیش‌بینی شده مشکل عدم ثبات اوزان دارایی‌ها برطرف شده است اما با توجه به اینکه مدل‌های بهینه‌سازی درجه دوم به معکوس پذیری ماتریس کوواریانس احتیاج دارند، یا به عبارت دیگر تمامی مقادیر ویژه ماتریس کوواریانس مثبت باشند، در صورتی که ماتریس کوواریانس عدد شرطی بالایی داشته باشد خطای دیگری بر محاسبات آنان تحمیل شد که دپرادو (۲۰۱۶) از آن به عنوان نفرین مارکویتز نام می‌برد.

فصل سوم

استفاده از اطلاعات جانبی در انتخاب برخط سبد سرمایه‌گذاری با
روش تطابق با الگو

۳-۱- مقدمه

روش شناسی تحقیق در این فصل مورد بررسی قرار می‌گیرد. در ابتدا مفاهیم HRP و نحوه استفاده از اطلاعات جانبی را مورد بررسی قرار می‌دهیم و سپس به مبحث تطابق با الگو و روش‌های خوشه‌بندی خواهیم پرداخت. در ادامه چندین الگوریتم که از ترکیب اطلاعات جانبی، الگوریتم و روش‌های خوشه‌بندی به دست آمده است را معرفی خواهیم کرد. الگوریتم‌های موجود، علاوه بر بازده سبد سرمایه‌گذاری به کنترل ریسک سبد نیز توجه کرده‌اند و سعی در تشکیل پرتفوی متنوع و با بازده بالا دارند.

۳-۲- الگوریتم HRP

تشکیل دادن پرتفو یکی از تکرارشونده‌ترین مفاهیم در سرمایه‌گذاری است. روزانه مدیران سرمایه‌گذاری باید پرتفوهایی تشکیل دهند که منعکس کننده دیدگاه‌ها و پیش‌بینی‌های خود در مورد ریسک‌ها و بازدهی‌های همه‌ی دارایی‌های آن پرتفو باشد. مارکویتز قبل از آنکه دکترای خود را در سال ۱۹۵۴ کسب کند اولین مدل بهینه‌سازی پرتفو یعنی الگوریتم خط بحرانی^۱ (CLA) را ارائه کرد به طوریکه این الگوریتم به کمک بهینه‌سازی درجه دوم بر اوزان دارایی‌ها، پرتفوی بهینه‌ای با بیشینه نسبت شارپ به ما می‌دهد. کوهن و تاکر (۲۰۱۴) اثبات کردند که همانند تمامی مدل‌های بهینه‌سازی درجه دوم، در صورت برقرار بودن شرایط کاروش-کاهن-

^۱ Critical Line Algorithm

تا کر^۲ این الگوریتم بهینه‌سازی بعد از تعداد محدودی تکرار به جواب می‌رسد که این ویژگی یکی از مهم‌ترین مزیت‌های این مدل محسوب می‌شود. توضیحات بیشتر و یک نمونه به کارگیری این مدل در مقاله بیلی و دپرادو^۳ (۲۰۱۳) یافت می‌شود.

بر خلاف نبوغی که در توسعه مدل بهینه‌سازی ماکویتز به کار رفته، به کارگیری آن در عمل با مشکلاتی مواجه است که از مهم‌ترین این مشکلات می‌توان به عدم ثبات اوزان در تشکیل دادن پرتفو اشاره کرد به طوری که همانطور که در مقاله میچاد (۲۰۰۱) به آن اشاره شده با کوچک‌ترین تغییری در پیش‌بینی بازدهی‌های آینده، مدل مارکویتز یک پرتفوی کاملاً متفاوت با پورتفوی قبل به ما تحویل می‌دهد. با توجه به اینکه پیش‌بینی بازدهی دارایی‌ها با یک سطح اطمینان قابل قبول تقریباً ناممکن می‌باشد اکثر مدیران سبد به استفاده از ماتریس کواریانس^۴ و مدل‌های ریسک‌پایه^۵ روی آورده‌اند که از مهم‌ترین آنها می‌توان به مدل تعادل ریسک^۶ اشاره کرد. چاوز (۲۰۱۱) این استراتژی که کاملاً بر مبنای کمینه کردن ریسک، بدون استفاده از الگوریتم‌های بهینه‌سازی خطی و درجه دوم می‌باشد را توسعه داده است. روش‌های دیگری مانند بلک لیترمین^۷ (۱۹۹۱) هم در طول سال‌ها توسعه داده شده‌اند که سعی در بهبود نواقص مدل مارکویتز داشته‌اند اما هیچ کدام نتوانسته‌اند نتایجی به خوبی نتایج الگوریتم CLA داشته باشند.

ابتدا به نظر می‌رسد با استفاده نکردن از بازدهی‌های پیش‌بینی شده مشکل عدم ثبات اوزان دارایی‌ها برطرف شده است اما با توجه به اینکه مدل‌های بهینه‌سازی درجه دوم به معکوس پذیری ماتریس کواریانس احتیاج دارند، یا به عبارت دیگر تمامی مقادیر ویژه ماتریس کواریانس مثبت باشند، در صورتی که ماتریس کواریانس عدد شرطی بالایی داشته باشد خطای دیگری بر محاسبات آنان تحمیل شد که دپرادو (۲۰۱۶) از آن به عنوان نفرین مارکویتز نام می‌برد.

² Karush-Kuhn-Tucker

³ MARCOS LÓPEZ DE PRADO

⁴ Covariance matrix

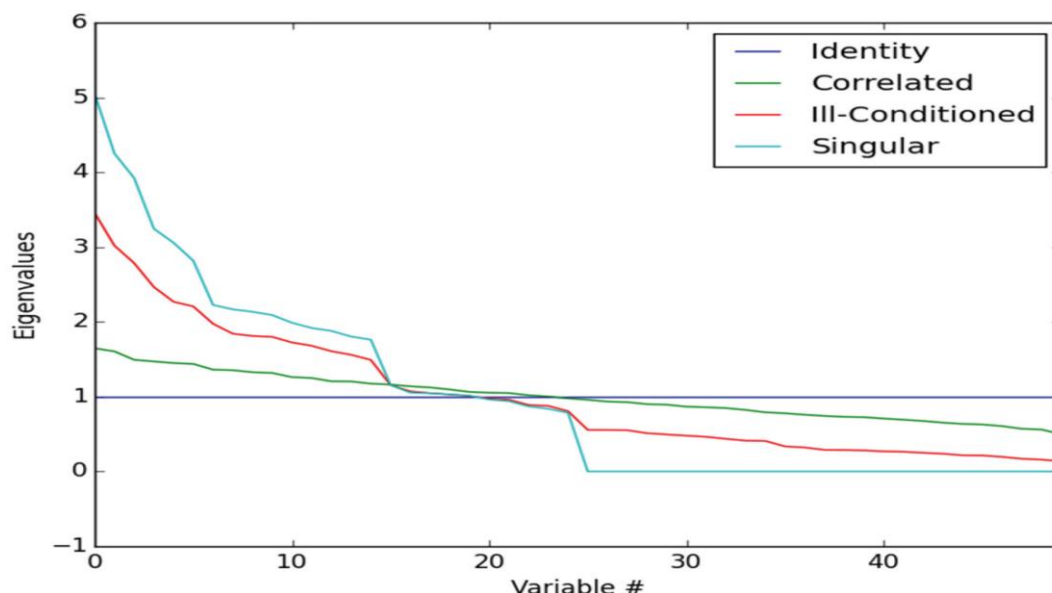
⁵ Risk based

⁶ Risk parity

⁷ Black-Litterman

۳-۲-۱- نفرین مارکویتز

به طور کلی عدد شرطی یک ماتریس، حاصل تقسیم قدر مطلق بزرگ‌ترین مقدار ویژه یک ماتریس بر کوچک‌ترین مقدار ویژه آن است. در نمودار ۱ نمودارهای مقادیر ویژه چندین ماتریس همبستگی^۸ از بزرگ به کوچک مرتب شده است به طوری که عدد شرطی این ماتریس‌ها اولین عدد نشان داده شده در نمودار، تقسیم بر آخرین عدد آن است. کوچک‌ترین عدد شرطی متعلق به ماتریس همانی است زیرا خودش معکوس خودش می‌باشد. فرض کنید چند دارایی با یکدیگر همبستگی بسیار کمی داشته باشند، حال با اضافه کردن دارایی‌هایی که با همدیگر همبستگی بیشتری دارند عدد شرطی ماتریس بیشتر می‌شود تا جایی که با کوچک‌ترین تغییر در درایه‌های ماتریس همبستگی یا ماتریس کوواریانس، معکوس آن به کلی تغییر خواهد کرد و باعث عدم ثبات در اوزان دارایی‌ها می‌شود. پس می‌توان گفت زمانی که همبستگی دارایی‌ها ما بیشتر می‌شود نیاز ما به متنوع سازی پورتفو هم بالاتر می‌رود تا علاوه بر متنوع کردن ریسک‌هایی که در بازارهای مختلف تحمل می‌کنیم، تاثیر خطاهای برآوردمان را در بازار های مختلف را کاهش دهیم.



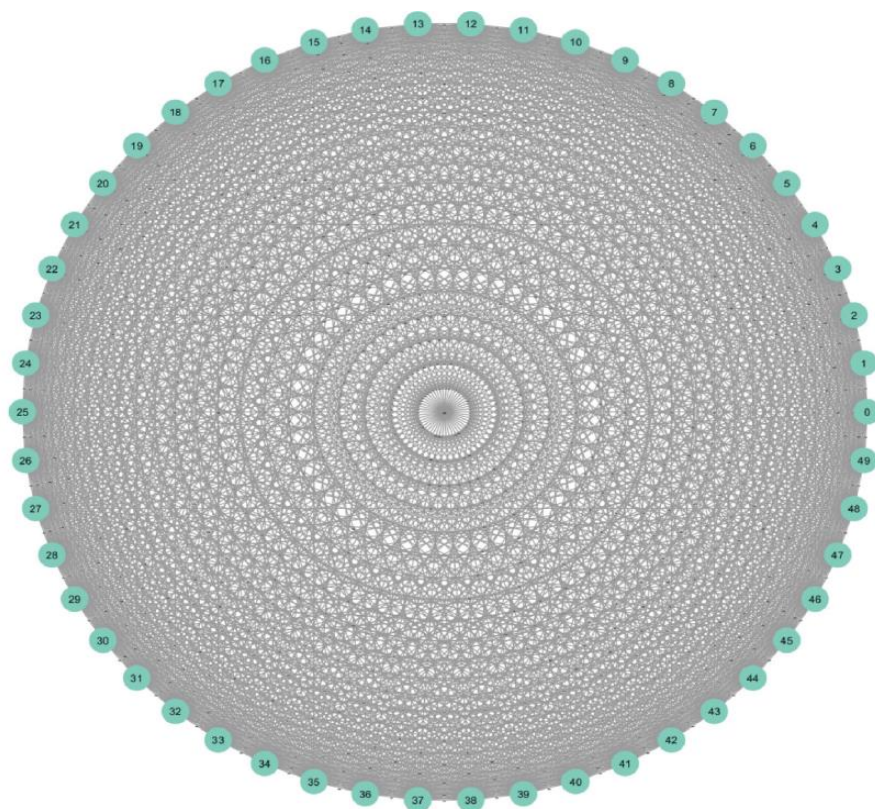
نمودار ۳-۱- مقادیر ویژه ماتریس همبستگی برای پرتفوی های مختلف، دپرادو (۲۰۱۸)

^۸ Correlation matrix

اما اضافه کردن اندازه ماتریس کوواریانس در مدل‌های بهینه‌سازی درجه دوم شرایط را بدتر می‌کند زیرا برای محاسبه ماتریس کوواریانس با N دارایی حداقل به $\frac{1}{2} N(N-1)$ داده تاریخی IID احتیاج داریم. به عنوان مثال برای محاسبه یک ماتریس کوواریانس برای ۵۰ دارایی مختلف تقریباً به ۵ سال داده تاریخی از دارایی‌ها احتیاج داریم که در دنیای سرمایه‌گذاری داده‌های مربوط به ۵ سال قبل نسبت به شرایط فعلی دارایی‌ها اطلاعات دقیقی را ارائه نداده و از ارزش کمتری برخوردارند.

۳-۲-۲- مدل‌های درختی

از جمله ضعف‌های مربوط به مدل‌های کلاسیک بهینه‌سازی پورتفو، عدم ثبات در اوزان دارایی‌هاست. نگرانی‌های ناشی از این بی‌ثباتی در سال‌های اخیر مورد توجه بسیاری قرار گرفته است که کلم و همکاران (۲۰۱۴) با دقت آن‌ها را مستند کرده‌اند. اکثر الگوریتم‌های جایگزین سعی در معرفی روشی دارند تا به تغییرات عددی در معکوس ماتریس کوواریانس ثبات ببخشند. در حقیقت ماتریس همبستگی یک ابزار از جبر خطی است که کسینوس زوایای بین هر دو بردار را در فضای بردار ایجاد شده توسط سری بازدهی را اندازه‌گیری می‌کند. یکی از دلایل بی‌ثباتی بهینه‌سازهای درجه دوم این است که فضای بردار به صورت یک گراف کامل یا کاملاً متصل، مدل شده است، جایی که هر گره یک نامزد احتمالی برای جایگزینی دیگری است. از نظر الگوریتمی، معکوس کردن ماتریس به معنی ارزیابی همبستگی‌های جزئی در گراف کامل است.

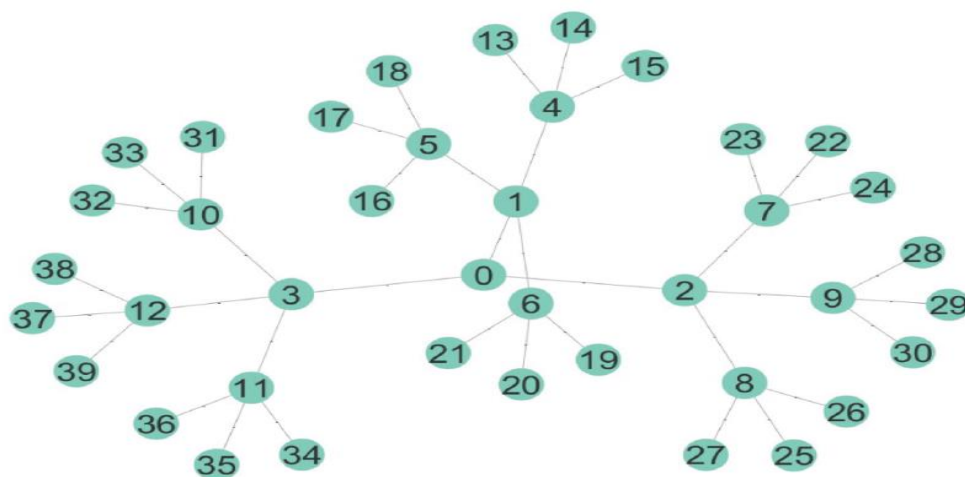


شکل ۳-۱- نمونه ای از گراف کامل برای ۵۰ دارایی، دپرادو (۲۰۱۸)

به عنوان مثال همانطور که در شکل ۳-۲ دیده می‌شود، گراف بالا روابط دلالت شده توسط ماتریس کوواریانس 50×50 را تجسم می‌کند، که شامل ۵۰ گره و ۱۲۲۵ لبه است. این ساختار نشان می‌دهد که برای محاسبه روابط نمونه کوچکی با ۵۰ نماد، مطابق ساختار بالا خطاهای تخمین به مراتب زیاد می‌شود و در نهایت منجر به راه‌حل‌های نادرست می‌شود. برای پرهیز از بزرگ شدن خطاهای تخمین محاسبه معکوس ماتریس کوواریانس به طور شهودی می‌توان گفت با کنار گذاشتن لبه‌های غیر ضروری می‌توانیم تعداد خطاهای کوچک تخمین را به اندازه مطلوبی کاهش دهیم.

فرض کنید که یک سرمایه‌گذار می‌خواهد مجموعه‌ای متنوع از اوراق بهادار شامل صدها سهام، اوراق قرضه، صندوق‌های پوشش ریسک، املاک و مستغلات، قراردادهای خصوصی و غیره بسازد. به نظر می‌رسد برخی از سرمایه‌گذاری‌ها جایگزین یکدیگر و سایر سرمایه‌گذاری‌ها مکمل یکدیگر هستند به عنوان مثال، سهام را می‌توان بر اساس مواردی همچون نقدینگی، اندازه، صنعت و منطقه جغرافیایی گروه‌بندی کرد که در آن

سهام‌ها در یک گروه معین با یکدیگر رقابت می‌کنند اما نمی‌توانند به عنوان مثال جایگزین اوراق قرضه یا حتی سهام‌هایی از گروه دیگر شوند. با این حال در یک ماتریس همبستگی، همانند شکل ۳-۲، همه سرمایه‌گذاری‌ها جایگزین‌های بالقوه یکدیگر هستند. به عبارت دیگر، ماتریس همبستگی فاقد مفهوم سلسله مراتب هستند. این فقدان ساختار سلسله مراتبی اجازه می‌دهد وزن‌های دارایی‌ها آزادانه به روش‌های ناخواسته تغییر کنند، که دلیل اصلی بی‌ثباتی CLA است. شکل ۳-۳ یک ساختار سلسله مراتبی معروف به درخت را تجسم می‌کند. ساختار درختی دو ویژگی مطلوب را معرفی می‌کند: اول آنکه فقط $N - 1$ لبه برای اتصال N گره‌ها دارد، بنابراین وزن‌ها فقط بین هم‌مرتبه‌ها در سطوح مختلف سلسله مراتبی تعادل می‌یابند. و دوم آنکه وزن‌ها از بالا به پایین توزیع می‌شود که این الگوریتم دقیقاً مانند عملکرد مدیران دارایی‌ها در وزن دهی به عناصر سبدشان است. در نتیجه می‌توان گفت، بنا به دلایل ذکر شده، ساختارهای سلسله مراتبی ساختارهای بهتری از جهت طراحی برای پایدار بودن اوزان دارایی‌ها و برای معنادار بودن آنها است.



شکل ۳-۲- ساختار درختی برای دارایی‌ها، دپرادو (۲۰۱۸)

۳-۲-۳- گام‌های الگوریتم HRP

ابتدا به معرفی رویکرد متعادل سازی ریسک سلسله مراتبی می‌پردازیم. الگوریتم HRP برای برطرف کردن سه ایراد عمده بی‌ثباتی، تمرکز و عملکرد ضعیف در مدل‌های بهینه‌سازی درجه دوم و به طور خاص مدل

مارکوویتز، توسعه داده شده است. HRP از تکنیک های ریاضیات مدرن، نظریه نمودار و یادگیری ماشین، استفاده می کند تا با اطلاعات موجود در ماتریس کوواریانس، پرتفویی با دارایی های متنوع ارایه دهد. با این حال، برخلاف بهینه سازی های درجه دوم، HRP نیازی به قابلیت معکوس پذیری ماتریس کوواریانس ندارد که این امر به HRP اجازه می دهد تا متنوع ترین پورتفو را برای دارایی ها با ماتریس کوواریانس که دترمینان آن صفر است محاسبه کند که تاکنون برای مدل های بهینه سازی ناممکن بوده است. همچنین شبیه سازی مونت کارلو نشان می دهد که الگوریتم HRP در عمل، خارج از نمونه^۹، پورتفوهای کم ریسکی به خوبی مدل CLA ایجاد کرده است درحالی که نقاط ضعفی که پیش تر ذکر شد را ندارد. بهترین روش برای اجرای الگوریتم HRP استفاده کردن از روش های یادگیری ماشین است که بعدتر به آن خواهیم پرداخت.

۳-۲-۱- خوشه بندی درختی

در این مرحله دارایی های سرمایه گذاری در یک ساختار درختی طبقه بندی می شوند به طوری که هر چه دارایی های ما شباهت بیشتری داشته باشند، که در این جا معیار شباهت همبستگی بالاتر نمادها می باشد، در شاخه های پایین تر درخت قرار می گیرند. برای فهم بیشتر این ساختار درختی، به نحوه ی تقسیم بندی تقسیم دارایی ها توسط یک مدیر دارایی توجه کنید. یک مدیر سبد ابتدا با استفاده از داده های کمی و کیفی دارایی ها را بر حسب ناحیه جغرافیایی تقسیم بندی کرده، سپس هر ناحیه جغرافیایی را به گروه های مختلف صنایع تقسیم بندی کرده و هر گروه صنایع به سهم های تشکیل دهنده آن تقسیم بندی می کند. هدف اصلی ما از خوشه بندی درختی تولید کردن ساختاری مشابه با نحوه فکر کردن این مدیر سبد در دنیای واقعی است. برای خوشه بندی درختی از متریک روبهرو استفاده خواهیم کرد:

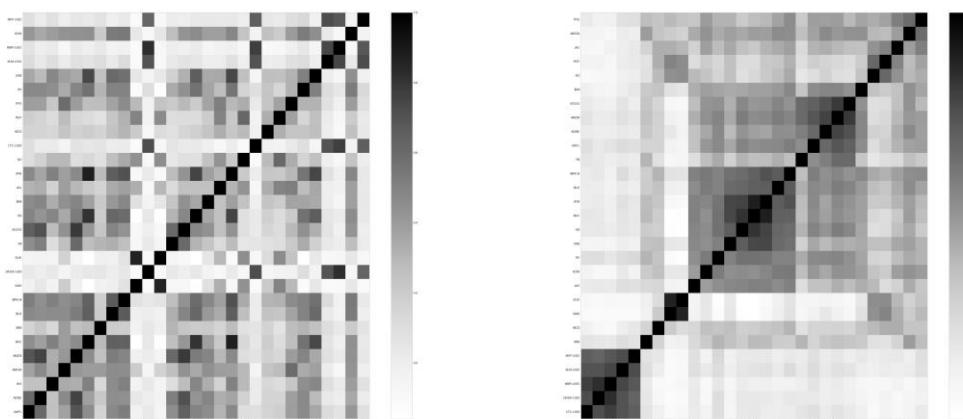
$$d[i,j] = d_{i,j} = \sqrt{\frac{1}{2}(1 - \rho_{i,j})} \quad (۳-۱)$$

^۹ Out of sample

برای محاسبه ماتریس فواصل از فرمول بالا استفاده می‌کنیم که $\rho_{i,j}$ ضریب همبستگی میان دو دارایی i, j است. سپس به کمک ماتریس فاصله‌ها ماتریس linkage را محاسبه می‌کنیم. دندوگرام حاصل از خوشه‌بندی سلسله مراتبی ماتریس فواصل به کمک ماتریس linkage ساختار درختی مطلوب را به ما می‌دهد.

۲-۳-۲-۳_شبه قطری سازی ماتریس

در این مرحله تنها با جابجایی سطرها و ستون‌های ماتریس کوواریانس، و همچنین ماتریس همبستگی، سعی به شبه قطری سازی این ماتریس داریم به طوری که بزرگ‌ترین درایه‌های این ماتریس در مسیر قطر ماتریس قرار گیرد. به طور کلی برای انجام این کار باید به ساختار درختی توجه کنیم و دارایی‌های مشابه، عضو یک شاخه، را کنار یکدیگر قرار دهیم. ما می‌دانیم که هر سطر ماتریس linkage دو شاخه را به یکدیگر متصل می‌کند پس تنها کاری که باید انجام دهیم جابجایی $\mathcal{Y}_{N-1,1}, \mathcal{Y}_{N-1,2}$ با عناصر، خوشه‌های تشکیل دهنده آن به صورت تودرتو است، آنقدر این جابجایی را ادامه می‌دهیم تا هیچ خوشه‌ای با دو دارایی باقی نماند. شکل ۱ ماتریس همبستگی برای ده دارایی شبه قطری شده همان ماتریس را در کنار هم نشان می‌دهد.



شکل ۳-۴_ نقشه حرارت ماتریس همبستگی و ماتریس همبستگی شبه قطری

۳-۳-۲-۳_ تخصیص دوبخشی بازگشتی

کار کردن با ماتریس همبستگی شبه قطری شده مزیت‌های خیلی بیشتری نسبت به کار کردن با ماتریس همبستگی عادی دارد که از مهم‌ترین آن می‌توان بهینه بودن نحوه محاسبه معکوس ماتریس را برشمرد. ما به دو شکل می‌توانیم از این مزیت‌ها استفاده کنیم :

I. الگوریتم با قراردادن تمامی دارایی‌ها، به ترتیب قطر ماتریس همبستگی شبه قطری در لیست L آغاز می‌شود. در ابتدا برای همه دارایی وزن ۱ را تخصیص می‌دهیم

$$II. \quad \forall L_i \in L: |L_i| = 1 \text{ در صورتی که برای هر عضو}$$

III. سپس به شرطی که $|L_i| > 1$ برای هر $\forall L_i \in L$ (یعنی خوشه انتخابی بیش از یک عضو داشته باشد).

(a) هر خوشه L_i را از وسط به دو زیر مجموعه $L_i^1 \cup L_i^2 = L_i$ تقسیم بندی می‌کنیم به طوری که $|L_i^1| = \lfloor 0.5 * |L_i| \rfloor$. ترتیب اعضای L_i حفظ شده باشد.

(b) سپس واریانس هر زیر خوشه $L_i^j, j = 1, 2$ محاسبه می‌شود به طوری که ابتدا وزن اولیه $w_i^j = \frac{\text{diag}(v_i^j)^{-1}}{\text{trace}(\text{diag}(v_i^j)^{-1})}$ به دارایی‌های عضو آن خوشه داده می‌شود و سپس وزن هر زیر خوشه با رابطه $\tilde{v}_i^j = w_i^{jT} v_i^j w_i^j$ محاسبه می‌شود.

(c) برای هر زیر خوشه ضریب تعدیل کننده الف با رابطه $\alpha_i = 1 - \frac{\tilde{v}_i^1}{\tilde{v}_i^1 + \tilde{v}_i^2}$ تعیین می‌شود. سپس وزن دارایی‌های مربوط به زیر خوشه یک با ضریب α_i و اوزان دارایی‌های زیر خوشه ۲ با ضریب $1 - \alpha_i$ تعدیل می‌شوند.

IV. بازشت به مرحله II

۳-۳- اطلاعات جانبی

در الگوریتم‌های توسعه داده شده در این پژوهش از نسبت P/E گذشته نگر به عنوان اطلاعات جانبی استفاده شده است. برای تعدیل وزن دهی بر مبنای Pattern Matching از امتیاز بر مبنای اطلاعات جانبی PE سهام ها استفاده شد. به طوری که از تاریخ شروع الگوریتم و هر سه ماه یکبار بر مبنای داده فصلی EPS سایت Yahoo Finance، PEtm هر سهام محاسبه شد. سپس میانگین متحرک آن در ۴ دوره سه ماهه که در مجموع یک سال می‌شود، در هر تاریخ محاسبه شد. در نهایت در هر بار تشکیل سبد سهامی جدید بر مبنای اطلاعات جانبی سهام ها بر اساس $\frac{PE}{MA(PE,4)}$ از کوچک به بزرگ مرتب شده و ۱۵ سهام برتر انتخاب می‌شود.

۳-۴- وزن دهی بر اساس معکوس واریانس^{۱۰}

در روش IVP یا وزن دهی بر اساس معکوس واریانس، واریانس هر دارایی به عنوان معیار ریسک آن در نظر گرفته می‌شود و در این روش از همبستگی میان بازدهی‌ها صرف نظر می‌شود. در این روش ابتدا واریانس بازدهی‌های هر دارایی در یک بازه زمانی محاسبه می‌شود سپس اوزان دارایی‌های هر سبد بر اساس معکوس ریسک آن دارایی تعیین می‌شود. برای این کار ابتدا اوزان همه دارایی برابر با معکوس واریانس بازدهی آن تعیین شده و در نهایت با تقسیم بر مجموع اوزان اولیه، مجموع اوزان سبد به یک نرمال می‌شود. هدف از اعمال روش IVP، تشکیل سبدی با اهرم ۱ و کمینه کردن ریسک در میانگین وزنی است.

¹⁰ Inverse variance portfolio

۳-۵- پرتفوی برابری ریسک^{۱۱}

شناخته شده ترین روش Risk parity، توزیع ریسک بر مبنای سهم ریسک برابر برای تمامی دارایی‌ها می‌باشد. در این روش بر خلاف IVP، ما به دنبال نوسانات یکسان نیستیم بلکه می‌خواهیم در انتهای تشکیل شدن سبد، اعانه هر دارایی به نوسانات سبد یکسان باشد. برای این کار ابتدا باید سهم هر دارایی از ریسک سبد را تعریف کنیم. سبدی با دارایی‌های $X = (x_1, x_2, \dots, x_n)$ را در نظر بگیرید به طوری که w_i وزن متعلق به دارایی x_i در سبد باشد. w بردار اوزان تمامی دارایی‌ها و Σ ماتریس کوواریانس برای بازدهی‌های تمامی دارایی‌های سبد باشد. آنگاه نوسانت سبد X به صورت انحراف معیار متغیر تصادفی $w^t X$ به صورت زیر محاسبه می‌شود.

$$\sigma(w) = \sqrt{w' \Sigma w} \quad (۲-۳)$$

از آنجا که تابع $\sigma(w)$ بر حسب w ، همگن از مرتبه اول است، بر اساس نظریه اوایلر برای توابع همگن داریم:

$$\sigma(w) = \sum_{i=1}^n \sigma_i(w) \quad (۳-۳)$$

$$\sigma_i(w) = w_i \partial_{w_i} \sigma(w) = \frac{w_i (\Sigma w)_i}{\sqrt{w' \Sigma w}} \quad (۴-۳)$$

همانطور که در تساوی بالا دیده می‌شود، می‌توان $\sigma_i(w)$ را سهم ریسک دارایی w_i از تمامی سبد تعریف کرد.

آنگاه برای اینکه سهم ریسک همه دارایی‌ها برابر است می‌توان گفت $\sigma_i(w) = \frac{\sigma(w)}{n}$ پس اوزان دارایی‌ها با

حل کردن معادله

$$w_i = \frac{\sigma(w)^2}{(\Sigma w)_i n} \quad (۵-۳)$$

یا حل مسئله کمینه سازی زیر

^{۱۱} Risk parity portfolio

$$\operatorname{argmin}_w \sum_{i=1}^n \left(w_i - \frac{\sigma(w)^2}{(\Sigma w)_i n} \right)^2 \quad (6-3)$$

به دست می‌آید.

۳-۶- خوشه بندی

یک از تکنیک‌های بسیار پرکاربرد در علوم داده و فرآیندهای یادگیری ماشین خوشه‌بندی می‌باشد. فرآیند گروه‌بندی مجموعه‌ای از داده‌ها و قرار دادن آن‌ها در گروه‌هایی از نمونه‌های مشابه خوشه‌بندی نام دارد. یک خوشه مجموعه‌ای از داده‌هاست که نسبت به دیگر داده‌های همان خوشه شبیه بوده ولی متفاوت از نمونه‌های دیگر خوشه‌ها هستند. خوشه بندی یک تکنیک یادگیری بدون نظارت است. تفاوت خوشه‌بندی و دسته‌بندی در این است که خوشه‌بندی، خوشه‌ها را تعریف می‌کند و داده‌ها را درون آنها می‌ریزد در حالی که در دسته‌بندی داده‌ها در دسته‌های از پیش تعیین شده قرار می‌گیرند. مزیت اصلی خوشه بندی این است که الگوها و ساختارها را می‌توان مستقیماً از مجموعه داده های بسیار بزرگ با کم یا هیچ یک از آنها یافت به همین دلیل الگوریتم های خوشه بندی را می‌توان در تمامی حوزه ها استفاده کرد.

از تکنیک‌های آن می‌توان به روش‌های k-means, k-medoids خوشه‌بندی طیفی و خوشه‌بندی سلسله مراتبی اشاره کرد.

۳-۶-۱- خوشه بندی طیفی

خوشه‌بندی طیفی یک تکنیک تجزیه و تحلیل داده می‌باشد که مجموعه‌ی داده‌های چند بعدی پیچیده را به خوشه‌هایی از داده‌های مشابه با ابعاد کمتر کاهش می‌دهد. هدف این است که طیف کاملی از نقاط داده را بر

اساس شباهت آن‌ها در چندین گروه دسته‌بندی کنیم. این روش داده‌های مشابه را، بدون توجه به ویژگی‌ها، حول نقاط مشترک گروه‌بندی می‌کند.

۳-۶-۲- خوشه بندی سلسله مراتبی

خوشه‌بندی سلسله مراتبی یک روش ساده اما اثبات شده برای تجزیه و تحلیل داده‌ها با ساخت خوشه‌هایی از داده‌هایی با الگوهای مشابه است. این کار با گروه‌بندی مکرر داده‌هایی انجام می‌شود که با یکدیگر همبستگی بالایی دارند و با یک دندروگرام می‌تواند نمایش داده شود. شاخه‌های دندروگرام نشان‌دهنده شباهت‌های بین داده‌ها هستند به نحوی که هر چه شاخه کوتاه‌تر باشد، شباهت بیشتری در الگوی بین داده وجود دارد. خوشه‌بندی سلسله مراتبی یکی از روش‌های محبوب در خوشه‌بندی بوده که در علوم مالی نیز می‌تواند کاربردهای فراوانی داشته باشد.

الگوریتم‌های خوشه بندی سلسله مراتبی تجمعی می‌توانند در معنای الگوریتمی به عنوان حریص شناخته شود. به طوری که دنباله ای از مراحل برگشت ناپذیر برای پیدا کردن خوشه های مورد نظر انجام می‌شود. فرض کنید یک جفت از خوشه‌ها، احتمال زیاد خوشه هایی با یک عضو، در یکی از مراحل خوشه‌بندی تجمعی با یکدیگر ادغام شده باشند. آن موقع ممکن است زمانی که الگوریتم را با n خوشه و $n - 1$ پارتیشن بندی (n خوشه) آغاز می‌کنیم در نهایت به یک خوشه n عضوی یا یک دندوگرام با درخت باینری یا یک مجموعه نیمه مرتب (POSET) برسیم.

طیف گسترده ای از الگوریتم‌های خوشه بندی سلسله مراتبی تجمعی در طول زمان ارایه شده. چنین الگوریتم‌های سلسله مراتبی را می‌توان به راحتی به دو گروه از روش‌ها تقسیم کرد. گروه اول روش‌های پیوندی است. روش‌های تکی، کامل، وزن دار و بدون وزن از روش‌های مبتنی بر پیوند linkage هستند. خوشه بندی به روش‌های پیوند را می‌توان با یک گراف نشان داد که اصطلاحاً آنرا دندوگرام می‌نامیم. گروه دوم خوشه بندی سلسله مراتبی روش‌هایی هستند که در آنها مراکز خوشه‌ها به وسیله میانگین یا میانگین وزنی اعضای هر خوشه مشخص می‌شود. سه روش معمول در این روش می‌توان تنها از متریک‌های عدم شباهت دو خوشه

استفاده کرد یا می‌توان موقعیت مرکز خوشه‌ها را در این روش خوشه‌بندی دخیل کرد که روش ترکیبی، تمامی مدل‌های خوشه‌بندی سلسله‌مراتبی را پوشش می‌دهد که به طور خاص می‌توان به روش خوشه‌بندی به روز رسانی فورمول لنس و ویلیامز اشاره کرد.

به طور کلی عدم شباهت دو خوشه را می‌توان از فورمول زیر محاسبه کرد. فرض کنید عناصر i, j ، $i \cup j$ تجمیع شده و در خوشه $i \cup j$ قرار گیرند، آنگاه به طور کلی معیار عدم شباهت خوشه جدید با عنصر تصادفی k در مجموعه داده‌ها به شکل زیر است:

$$d(i \cup j, k) = \alpha_i d(i, k) + \alpha_j d(j, k) + \beta d(i, j) + \gamma |d(i, k) - d(j, k)| \quad (7-3)$$

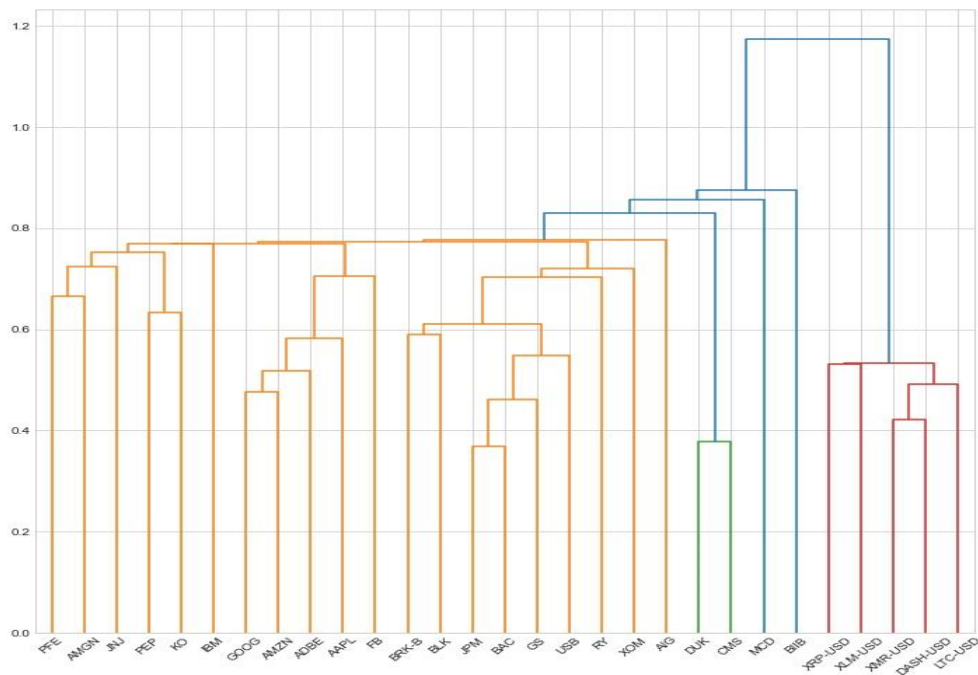
۳-۲-۱- محاسبه ماتریس linkage

این مرحله به ما اجازه می‌دهد تا ماتریس پیوند را به عنوان ماتریس $(N-1) \times 4$ با ساختار تعریف کنیم.

$Y = \{(y_{m,1}, y_{m,2}, y_{m,3}, y_{m,4})\}_{m=1,2,\dots,N-1}$ که عناصر $y_{m,1}, y_{m,2}$ عناصر ما در خوشه‌بندی، $y_{m,3}$ فاصله اقلیدسی میان دو بردار ستونی ماتریس شباهت هست که از رابطه زیر به دست می‌آید و $y_{m,4}$ تعداد عناصر موجود در کلاستر m است.

$$\tilde{d}_{i,j} = \sqrt{\sum_{n=1}^N (d_{n,i} - d_{n,j})^2} \quad (8-3)$$

برای محاسبه ماتریس پیوند ما از تابع از پیش نوشته شده پایتون در کتابخانه scipy استفاده می‌کنیم. برای استفاده خاص ما از خوشه‌بندی سلسله‌مراتبی، ماتریس پیوند به متد single محاسبه می‌شود یعنی فاصله فاصله میان دو خوشه منینم فاصله میان عناصر دو خوشه در نظر گرفته می‌شود. به طور کلی می‌توان گفت ستون سوم ماتریس پیوند در خوشه‌بندی طیفی فواصلی است که در دندوگرام مربوط به خوشه‌بندی سلسله‌مراتبی ترسیم می‌شود.



نمودار ۳-۲_ دندوگرام نمادهای پرتفو

تصویر بالا یک دندوگرام از بیست و نه دارایی پرتفو در یکی از مراحل الگوریتم توسعه داده شده می‌باشد که از بازار نیویورک آمریکا انتخاب شده‌اند را نشان می‌دهد. با توجه به حداکثر فاصله قابل تحمل میان خوشه‌های مختلف، می‌توان خطی به موازات محور افقی کشید و با توجه به تقاطع آن با دندوگرام خوشه‌ها را تعیین کرد. برای مثال در شکل بالا اگر از فاصله ۰.۸ خطی به موازات محور افقی رسم شود آنگاه، نماد MCD, BIB هر کدام به تنهایی یک خوشه، نمادهای بازار رمزارزها که با خطوط قرمز نشان داده شده است یک خوشه، نمادهای DUK, CMS که با رنگ سبز مشخص می‌باشد یک خوشه و در نهایت دیگر دارایی‌ها در یک خوشه دیگر، طبقه بندی می‌شوند.

۳-۶-۳ خوشه بندی به روش k-means

در الگوریتم k-means مجموعه داده‌ها به تعداد خوشه‌هایی که از پیش تعیین شده‌اند تقسیم خواهند شد. ایده اصلی در این الگوریتم تعریف K مرکز برای هر یک از خوشه‌ها است. پس از آن هر مشاهده در مجموعه

داده به نزدیکترین مرکز خوشه اختصاص داده می‌شود. فاصله اقلیدسی و فاصله همینگ از مشهورترین معیارهای محاسبه فاصله مشاهدات می‌باشند که با توجه به ویژگی‌های داده می‌توانند به کار گرفته شوند. الگوریتم K-Means یکی از ساده ترین نوع الگوریتم های یادگیری بدون نظارت است که اکثر مسایل خوشه بندی را به خوبی حل می کند. این روش یک راه ساده و آسان را برای طبقه بندی یک مجموعه داده پیشینی معین از طریق تعداد معینی از خوشه ها (فرض کنید k خوشه) را دنبال می کند. ایده اصلی این است که ابتدا k مرکز را برای هر خوشه تعریف می کنیم. این مرکز ها باید به روشی خلاقانه طوری قرار بگیرند که بهترین و دقیق ترین جواب ممکن را فراهم کنند زیرا نتیجه این الگوریتم تنها از موقعیت مرکزها تبعیت می کند. سپس هر داده در مجموعه داده هایمان را با یکی از این مرکز ها متناظر می کنیم به طوری که حداقل فاصله (اقلیدسی) را از آن مرکز داشته باشد. زمانی که این گام را برای تمامی داده ها انجام دادیم، اینبار باید دوباره K مرکز جدید برای داده هایمان محاسبه کنیم که به طوری که وضعیت بهتری را نسبت به حالت قبل ارائه دهند به طوری که دوباره همانند گام قبلی همه داده ها را به یک مرکز متناظر کرده و دوباره تابع هدف را برای آن محاسبه می کنیم. همانطور که شرح داده شد بین گام های دوم و سوم الگوریتم یک حلقه ایجاد شده که تکرار کردن (iteration) موقعیت مرکز ها، هر بار جواب جدیدی (بهتری) تولید می شود تا معمولاً این حلقه تا زمانی که به جواب بهینه برسیم ادامه پیدا می کند. تابع هدف در این الگوریتم به نوعی مجموع مربعات خطا را نمایش می دهد به طوری که برای n دیتا و k خوشه با مرکز های C_1, C_2, \dots, C_k داریم:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (9-3)$$

در نهایت این فرایند باعث ایجاد جدایی اشیاء از یکدیگر می شوند و گروه هایی را به وجود می آورند که یک متریک خاص، در این جا مجموع مربعات خطا را بهینه می کند. اگرچه می توان ثابت کرد که این رویه همیشه خاتمه می یابد، اما الگوریتم K-Means لزوماً جواب بهینه مربوط به تابع هدف را پیدا نمی کند. الگوریتم نیز به طور قابل توجهی حساس به مراکز خوشه ای اولیه که به طور تصادفی انتخاب شده اند، است. برای کاهش دادن این خطا می توان الگوریتم را چند بار اجرا کرد و پاسخ های مشابه را معتبر دانست. در الگوریتم k-means مرکز خوشه ها نماینگر تمامی ویژگی های آن خوشه هستند.

۳-۶-۴- خوشه بندی به روش k-medoids

ایده این الگوریتم، حل برخی مشکلات الگوریتم K-Means می باشد، که در آن بجای مینیمم^{۱۲} کردن مجموع مجذور فاصله اقلیدسی بین مشاهدات، که عموماً به عنوان تابع هدف در الگوریتم K-Means به کار برده می شود، مجموع تفاوت های فواصل دو به دو مشاهدات را مینیمم می کند. و برای به دست آوردن مراکز جدید در حلقه های مدل از میانه اعضای هر خوشه بهره می برد.

از آنجایی که یک داده با ارزش بسیار زیاد می تواند توزیع بخش داده ها را تحت تاثیر قرار دهد، از مهمترین مشکلات الگوریتم Kmeans، می توان به حساسیت بالای آن به داده های پرت اشاره کرد زیرا در صورتی که این داده پرت با چند داده دیگر در یک خوشه قرار گیرند، داده پرت به تنهایی مرکز آن خوشه را به اندازه قابل توجهی جا به جا می کند. روش سنتی برای حل این مشکل استفاده از میانه به جای میانگین است زیرا با وجود داده های پرت، میانه تقریب بهتری برای داده ها نسبت به میانگین است. بنابراین روش پارتیشن بندی خوشه بندی kmeans همچنان در این روش جواب می دهد به طوری که ابتدا به صورت کاملاً تصادفی k دیتا را به عنوان مرکز خوشه ها فرض می کنیم. سپس بقیه داده ها را براساس نزدیک بودن فاصله (اقلیدسی) به این مرکز ها نسبت می دهیم. (به طور کلی در گام های بعد نیز میانه هر خوشه مرکز آن خوشه قلمداد می شود). سپس یک داده که مرکز هیچ کدام از خوشه ها نباشد را انتخاب کرده O_{random} و تابع جابجایی^{۱۳} S را برای آن نقطه و مرکز هر خوشه O_j محاسبه می کنیم. اگر $S > 0$ بود آنگاه O_j را با O_{random} عوض کرده و در غیر این صورت الگوریتم را انقدر ادامه می دهیم که شاهد هیچ تغییری نباشیم.

¹² Minimum

¹³ Swap

۳-۷- انتخاب برخط سبد سرمایه‌گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی

گام ۱: دریافت دیتای ورودی ماتریس قیمت P نمادها

در این مرحله ماتریس قیمت به کد معرفی می‌شود. در ستون این ماتریس نام دارایی‌ها و در سطرها این ماتریس قیمت نمادها قرار دارد. قیمت پایانی یک دارایی به عنوان قیمت در آن تاریخ در نظر گرفته شده و این دیتا از سایت یاهو فاینانس^{۱۴} گرفته شده است. برای این کار ابتدا باید تاریخ نهایی الگوریتم و تعداد روزهایی که می‌خواهیم الگوریتم در آن روزها اجرا شود را تعیین کنیم که در اینجا مجموع داده‌های تست^{۱۵} و ترین^{۱۶} را ۱.۵ سال در نظر گرفته‌ایم. در صورت استفاده از اطلاعات جانبی باید تعداد و اندازه زمانی پنجره‌های زمانی نگهداری سبدها بر اساس دیتای جانبی نیز مشخص شود. سپس پنجره زمانی‌ای که می‌خواهیم بر اساس آن تصمیم‌گیری کنیم را تعیین می‌کنیم. برای مثال ۱.۵ سال گذشته را می‌توانیم به ۶ فصل تقسیم کنیم. می‌توانیم برای هر نوع خوشه‌بندی یک پنجره زمانی تعیین کنیم در این‌جا برای خوشه‌بندی طیفی، پنجره زمانی دو ساله و برای خوشه‌بندی سلسله مراتبی پنجره زمانی یک ساله در نظر گرفته شده است. در این صورت به اندازه بیشترین پنجره زمانی تعیین شده به داده‌های تست و ترین می‌افزاییم و دیتا را در آن تاریخ‌ها از yahooofinance دانلود می‌کنیم. سپس داده‌ها را به دو قسمت تست و ترین تقسیم‌بندی می‌کنیم.

به طور کلی الگوریتم تابع سه متغیر اصلی پنجره زمانی خوشه‌بندی سلسله مراتبی، پنجره زمانی خوشه‌بندی طیفی و اندازه زمانی زیر ماتریس‌ها برای خوشه‌بندی طیفی می‌باشد. برای به‌دست آوردن این متغیرها ابتدا الگوریتم را برای حالت‌بندی‌های مختلف بر روی داده‌های ترین اجرا می‌کنیم. برای پنجره زمانی خوشه‌بندی سلسله مراتبی سه ماه، شش ماه و یک سال، و برای پنجره زمانی خوشه‌بندی طیفی شش ماه، یک سال و دو سال، و برای اندازه زمانی زیر ماتریس‌ها از ۲ تا ده روز استفاده کرده‌ایم. معیار تصمیم‌گیری ما برای پنجره

¹⁴ yahooofinance

¹⁵ Test

¹⁶ Train

زمانی خوشه‌بندی‌ها، معیار نیکویی برازش خوشه‌بندی یعنی سیلوئت^{۱۷} و برای اندازه زمانی زیر ماتریس‌ها، معیار تصمیم‌گیری ما بازده سالیانه می‌باشد. در نهایت با مقادیر ثابت ۱ سال برای خوشه‌بندی سلسله مراتبی، ۲ سال برای خوشه‌بندی طیفی و ۲ روز برای اندازه زمانی زیر ماتریس‌ها الگوریتم تست را آغاز می‌کنیم.

نمادهای استفاده شده در پژوهش که شامل سهام و رمز ارزها می‌شوند، در انتهای پژوهش پیوست شده است.

گام ۲: به کمک ماتریس قیمت‌ها، ماتریس نسبت‌های قیمتی را به دست می‌آوریم به طوری که $R_{i,t} = \frac{P_{i,t}}{P_{i,t-1}}$

گام ۳: در صورت استفاده از اطلاعات جانبی، ابتدا برای تمامی سهام‌ها باید در تاریخ‌های باز توازن^{۱۸}، نسبت

$\frac{PE}{MA(PE,4)}$ آن‌ها را محاسبه کرده و در هر تاریخ، تمامی سهام را بر این اساس از کوچک به بزرگ مرتب

می‌کنیم و در نهایت ۱۵ سهام برتر را به عنوان سهام‌های موجود در آن دوره نگهداری در نظر بگیریم. گام‌های بعدی مربوط به الگوریتم تست خواهد بود.

گام ۴: ابتدا ماتریس اوزان را تعریف می‌کنیم به طوری که سطرها نمایان‌گر دارایی‌ها و ستون‌ها، اوزان دارایی‌ها

در هر لحظه را نشان می‌دهند. در صورت عدم استفاده از اطلاعات جانبی ماتریس اوزان تنها در اجرای

بهینه‌سازی الگوریتم تغییر می‌کند. اما در صورت استفاده از اطلاعات جانبی، ابتدا بررسی می‌کنیم که با توجه

به زمان‌های متعادل سازی اوزان، در چه پنجره زمانی قرار داریم. در ابتدای هر پنجره زمانی نام دارایی‌هایمان

را از تمامی رمز ارزها و سهام‌های انتخاب شده در پنجره زمانی متناظر گام ۳ قرار می‌دهیم و شمارنده الگوریتم

را برابر با صفر قرار می‌دهیم.

گام ۵: خوشه‌بندی سلسله مراتبی

ابتدا پنجره زمانی داده مورد نیاز خوشه‌بندی را با توجه به پنجره زمانی تعیین شده در گام ۱ جدا می‌کنیم و

ماتریس بازدهی مورد نیاز خوشه‌بندی سلسله مراتبی را تشکیل می‌دهیم. سپس به کمک ماتریس بازدهی

مورد نیاز خوشه‌بندی سلسله مراتبی، ماتریس کوواریانس و ماتریس Correlation را تشکیل می‌دهیم. سپس

¹⁷ Silhouette

¹⁸ Rebalancing

ماتریس فاصله ها را از رابطه $dist = \sqrt{\frac{1-corr}{2}}$ محاسبه می کنیم. در نهایت خوشه بندی سلسله مراتبی را آغاز می کنیم.

گام ۵.۱: به ازای $k = 2, 3, \dots, [\sqrt{n}]$ تعداد خوشه، ماتریس فاصله ها را خوشه بندی می کنیم. برای ارزیابی نیکویی برازش این خوشه بندی از سیلوئت استفاده می کنیم به طوری که در هر بار خوشه بندی سیلوئت را محاسبه کرده و ذخیره می کنیم.

گام ۵.۲: تعداد خوشه ها را برابر با تعداد خوشه های متناظر با مقدار ماکسیمم سیلوئت در گام ۵.۱ در نظر گرفته و دوباره خوشه بندی را انجام می دهیم سپس خوشه ها را بر اساس فاصله ی آن ها از دیگر خوشه ها مرتب می کنیم. (خوشه های L_1, L_2, \dots, L_J)

گام ۵.۳: به طور کلی اگر V_j ماتریس کوواریانس دارایی های مربوط به خوشه l_j ($j \leq J$) باشد، اوزان اولیه $w_j = \frac{diag(v_j)^{-1}}{trace(diag(v_j)^{-1})}$ را به دارایی های درون خوشه l_j اختصاص می دهیم. آنگاه واریانس این خوشه با اوزان فعلی $\widehat{Var}(l_j) = w_j^T V_j w_j$ برآورد می شود. در نهایت وزن خوشه $\frac{\widehat{Var}(l_j)^{-1}}{\sum_1^J \widehat{Var}(l_j)^{-1}}$ را به خوشه l_j اختصاص می دهیم.

گام ۶: خوشه بندی طیفی

گام ۶.۱: ابتدا دارایی ها را بر اساس خوشه بندی انجام شده در گام ۵.۲ تفکیک می کنیم. سپس برای هر خوشه، خوشه بندی طیفی را انجام می دهیم.

گام ۶.۲: همانند گام ۵.۱ مطابق با پنجره زمانی تعیین شده N در گام ۱، ماتریس بازدهی مورد نیاز برای خوشه بندی طیفی را تشکیل می دهیم به طوری که سطرهای این ماتریس بازدهی دارایی ها و ستون های این ماتریس اختلاف زمانی مربوط به آن تاریخ و تاریخ تشکیل ماتریس است (به روز) به طوری که ستون مربوط

به تاریخ تشکیل ماتریس است، ستون 0 ، روز قبل آن ستون 1 و برای بقیه روزها به همین ترتیب ادامه می‌دهیم.

سپس با توجه به اندازه زیر ماتریس‌ها که در گام 1 تعیین شده، زیرماتریس‌های ، ماتریس بازدهی مورد نیاز برای خوشه‌بندی طیفی را محاسبه می‌کنیم به طوری که روز های $1, 2, 3, \dots, tw$ در زیر ماتریس اول، روز های $2, 3, 4, \dots, tw + 1$ در زیر ماتریس دوم و بقیه روزها به طور مشابه در زیر ماتریس‌های متناظر قرار می‌گیرند و در انتها $N - tw + 1$ زیر ماتریس محاسبه می‌شود. در نهایت با سطری سازی هر زیر ماتریس، یک بردار را با آن متناظر می‌سازیم.

گام ۶.۳: در این مرحله ابتدا زیر ماتریس مربوط به تاریخ تشکیل ماتریس بازدهی مورد نیاز برای خوشه‌بندی طیفی یعنی مربوط به روز های $0, 1, \dots, tw - 1$ و بردار متناظر آن را محاسبه می‌کنیم و آن را بردار پایه می‌نامیم. سپس مجموعه بردارهای محاسبه شده در گام ۶.۲ و بردار پایه را به روش طیفی خوشه‌بندی می‌کنیم. برای خوشه‌بندی همانند گام ۵.۱ به ازای تعداد خوشه‌های $k = 2, 3, \dots, [\sqrt{N - tw + 1}]$ خوشه‌بندی را انجام می‌دهیم سپس تعداد خوشه‌ها را برابر با تعداد خوشه‌های متناظر با بیشترین مقدار سیلوئت در نظر گرفته و دوباره خوشه‌بندی را با تعداد خوشه‌های بهینه انجام می‌دهیم. سپس بردارهایی که با بردار پایه در یک خوشه قرار نمی‌گیرند را حذف می‌کنیم. در نهایت برای بردارهای منتخب، همبستگی آن‌ها با بردار پایه را بعد از نرمال سازی در بردار ستونی W و روز بعد از زیر ماتریس‌های متناظر با هر بردار را در ماتریس C قرار می‌دهیم. که در فاز بهینه سازی تنها از بردار W و ماتریس C استفاده خواهد شد.

گام ۷. بهینه سازی

آخرین گام از الگوریتم مذکور، تشکیل پرتفو بهینه برای روز آینده می‌باشد. این گام با استفاده از ماتریس C و بردار W که در مرحله‌ی قبل به دست آمده است صورت می‌پذیرد.

در این گام با دو حالت مواجه خواهیم بود، حالت اول زمانی رخ می‌دهد که ماتریس C شامل هیچ روز مشابهی با نمونه اخیر نباشد و به عبارت دیگر این ماتریس تهی باشد و حالت دوم بیانگر زمانی است که ماتریس C

تهی نبوده و یا به بیان دیگر روز و یا روزهای مشابه با نمونه اخیر وجود داشته باشد. حال به بررسی این دو حالت مذکور می‌پردازیم.

حالت (۱) ماتریس C فاقد روز مشابه بوده و تهی است: بر طبق موارد مذکور بالا، در گام سوم زمانی که هیچ روز مشابهی با روز اخیر یافت نشده باشد ماتریس C تهی بوده و در نتیجه با دو حالت مواجه خواهیم شد. حالت اول: این حالت زمانی رخ می‌دهد که در روز اول اجرای الگوریتم قرار داریم و ماتریس C نیز تهی باشد. در این حالت پرتفوی یکنواخت به عنوان پرتفوی بهینه انتخاب خواهد شد.

حالت دوم: این حالت زمانی رخ می‌دهد که ما در روز اول اجرای الگوریتم نباشیم و ماتریس C نیز تهی باشد. در این صورت هیچ تغییری در پرتفوی روز گذشته صورت نمی‌پذیرد و در واقع هیچ گونه معامله‌ای انجام نخواهد شد. در نتیجه پرتفوی بهینه روز آینده همان پرتفو تشکیل شده در روز قبل می‌باشد.

پرتفو تعدیل شده : با تغییر قیمت سهم‌های مختلف در طول روز ، اوزان آنها در پرتفو دچار تغییر شده و با اوزان اولیه متفاوت خواهد بود. این اوزان جدید سبب ایجاد یک پرتفوی جدید خواهند شد که به آن پرتفو تعدیل شده می‌گویند. در واقع به واسطه‌ی تغییر قیمت سهام و بدون انجام معامله‌ای یک پرتفو جدید به وجود خواهد آمد. رابطه پرتفوی تعدیل شده مطابق فرمول زیر است:

$$B_{adj}(t) = \frac{B^*(t) \odot x(t)}{B^*(t) \cdot x(t)} \quad (۱۰-۳)$$

صورت کسر فوق ضرب مولفه به مولفه‌ی دو بردار است که از آن یک بردار به دست می‌آید. مخرج عبارت بالا نیز ضرب داخلی دو بردار بوده که حاصل آن یک عدد خواهد شد و در نهایت حاصل ضرب یک عدد در بردار، یک بردار خواهد بود.

حالت (۲) ماتریس C تهی نباشد:

این حالت زمانی رخ خواهد داد که روزهای مشابه روز اخیر وجود داشته باشد و در این صورت پرتفوی بهینه روز آینده با بهینه سازی مدل زیر به دست خواهد آمد:

$$\begin{aligned} \max \quad & Z = E\{\log(b \odot (1 - TC)). x \mid x_i, i \in C(x_1^t)\} = \\ & \log((b \odot (1 - TC)). C) \cdot W = \sum_{i \in C(x_1^t)} w_i \log(b \odot (1 - TC)). ci \\ \text{s.t.} \quad & \sum bi = 1 \\ & l \leq bi \leq u \quad \forall i \end{aligned} \quad (11-3)$$

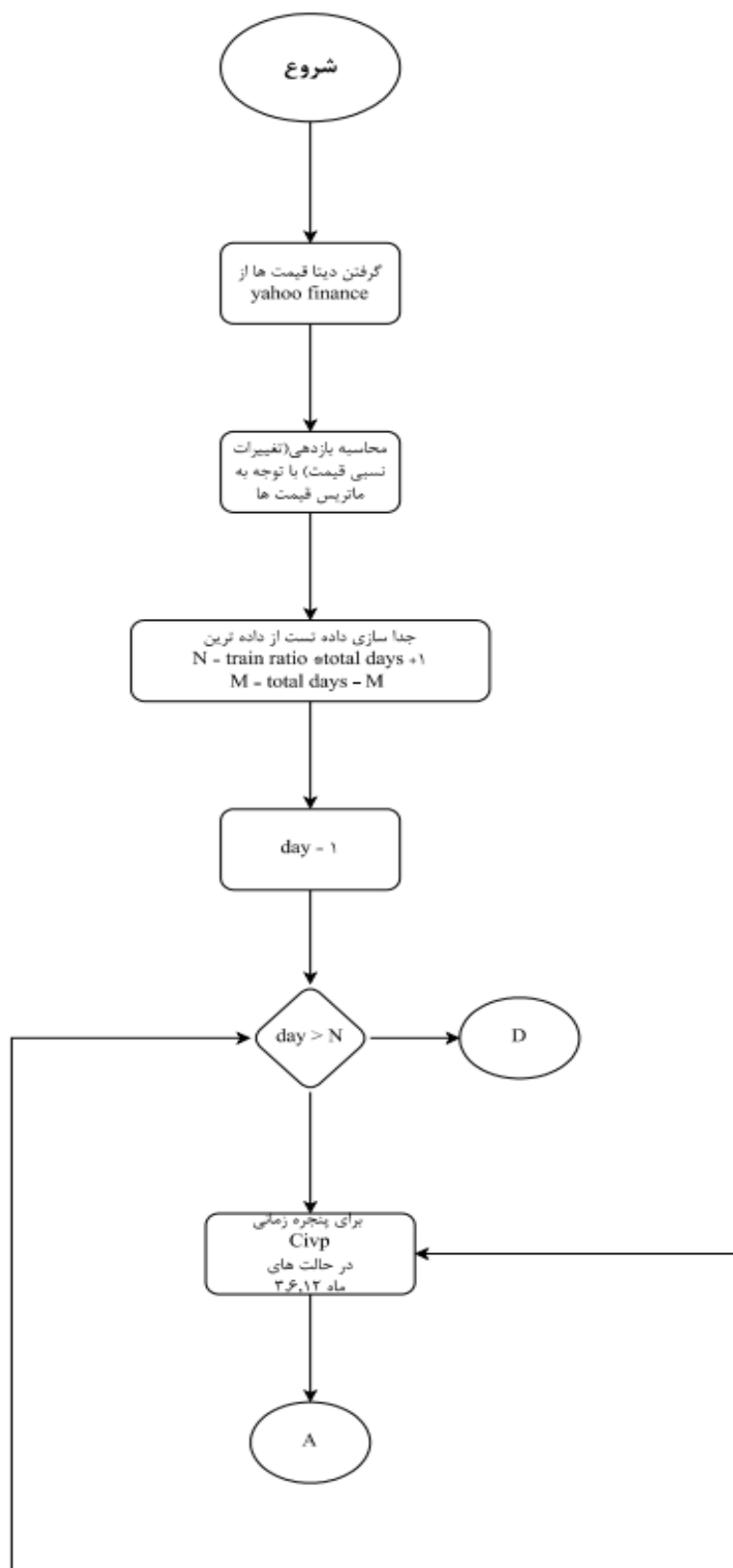
ورودی تابع هدف لگاریتم بهینه بالا، ماتریس C و بردار W می‌باشند که از مراحل قبل به دست آمده‌اند و به ترتیب ماتریس روزهای مشابه و بردار وزن روزهای مشابه می‌باشند. بردار TC نیز هزینه معاملاتی بوده که به وسیله فرمول زیر محاسبه خواهد شد:

$$TC = |\gamma| \odot \tau \quad (12-3)$$

$$\gamma = B_{adj}(t - 1) - b \quad (13-3)$$

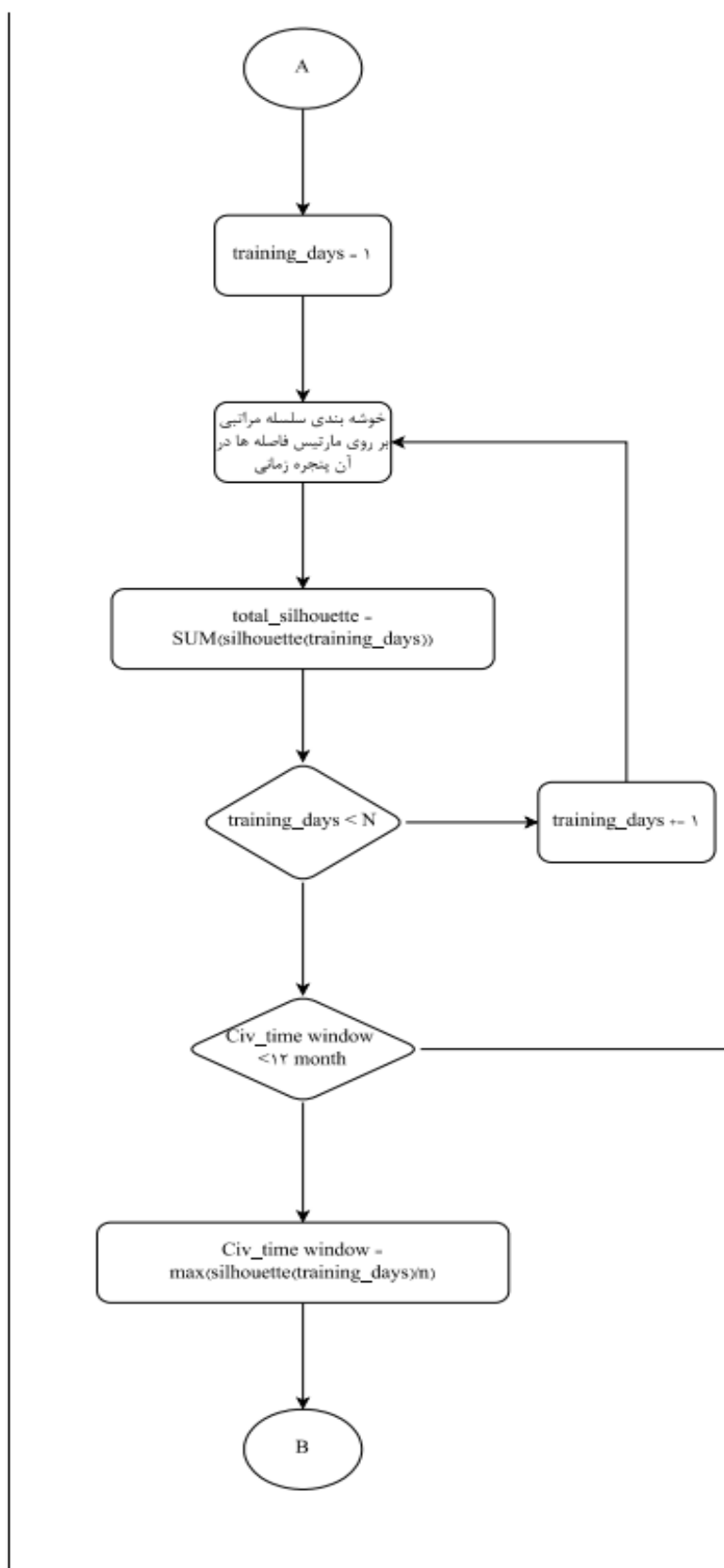
$$\tau = \begin{cases} TC_{Buy} & \gamma < 0 \\ TC_{Sell} & \gamma > 0 \\ 0 & \gamma = 0 \end{cases} \quad (14-3)$$

به این دلیل که کل سرمایه در پرتفو مورد استفاده قرارگیرد از محدودیت اول استفاده شده و محدودیت دوم حد بالا و حد پایین سرمایه‌گذاری را متناسب با نظر سرمایه‌گذار مشخص خواهد کرد و با بهینه‌سازی مدل فوق مقدار b^* به دست خواهد آمد که برابر با پرتفوی بهینه روز آینده می‌شود. فلوچارت الگوریتم فوق در شکل (۲-۳) نمایش داده شده است:



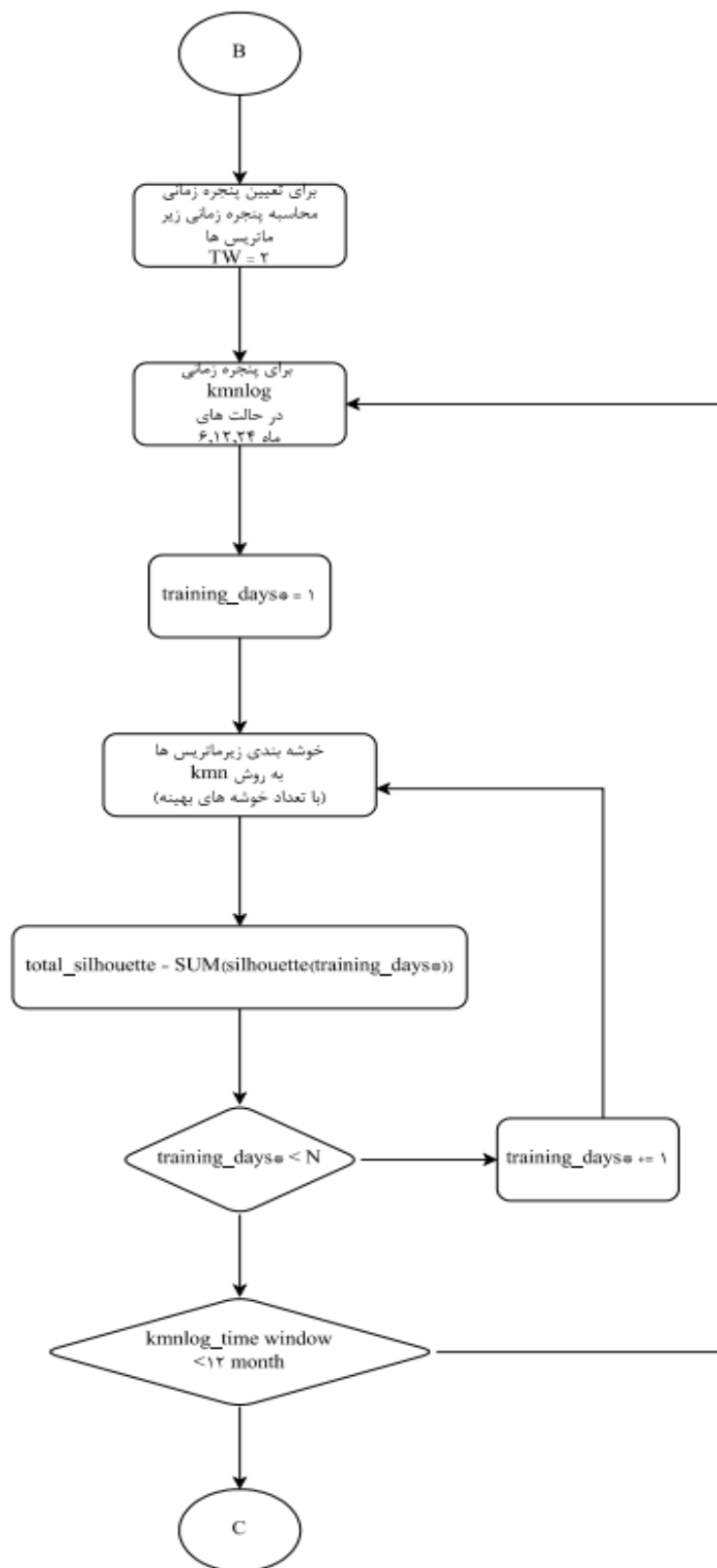
شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی



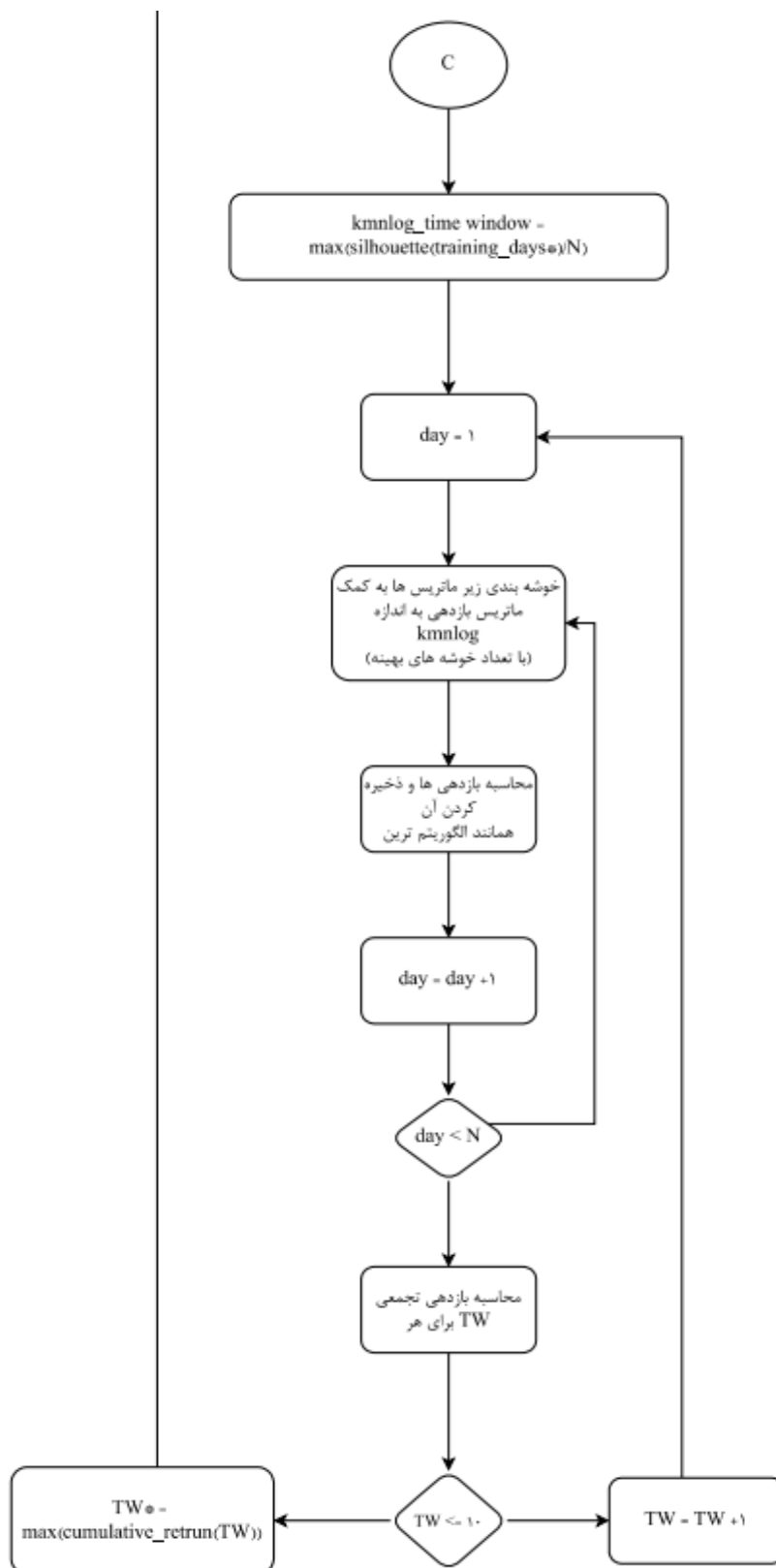
شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی (ادامه)



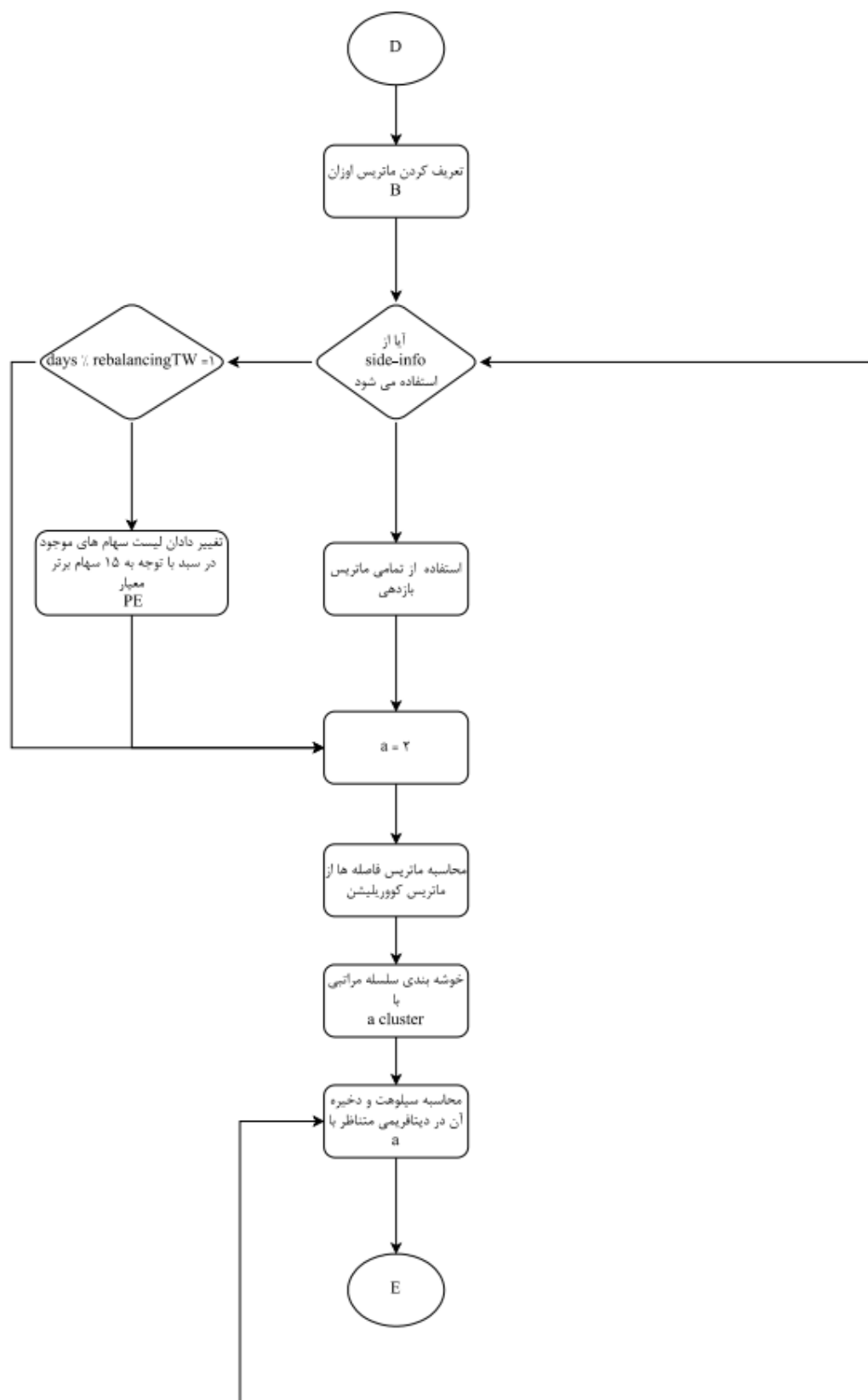
شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی(ادامه)



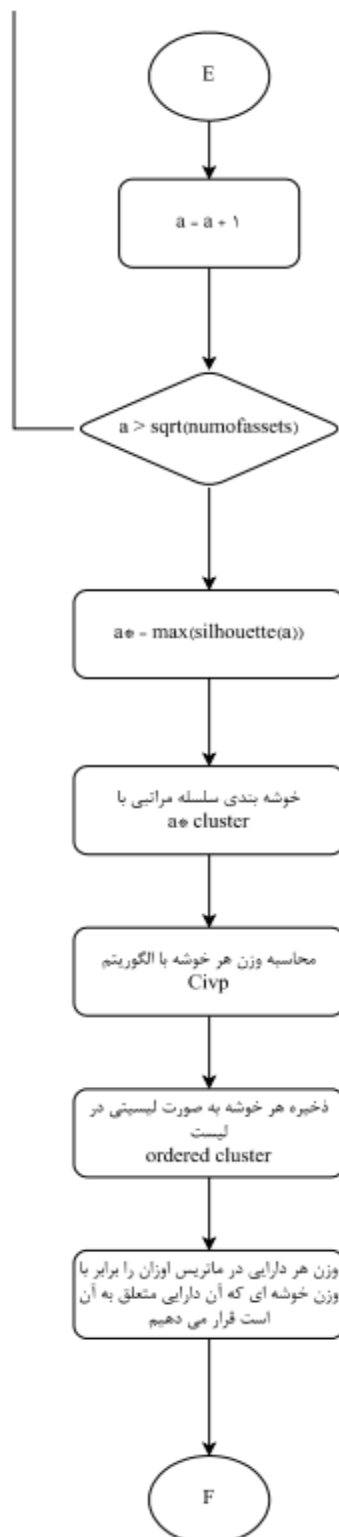
شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه گذاری به روش تطابق با الگو با کمک اطلاعات

جانبی (ادامه)



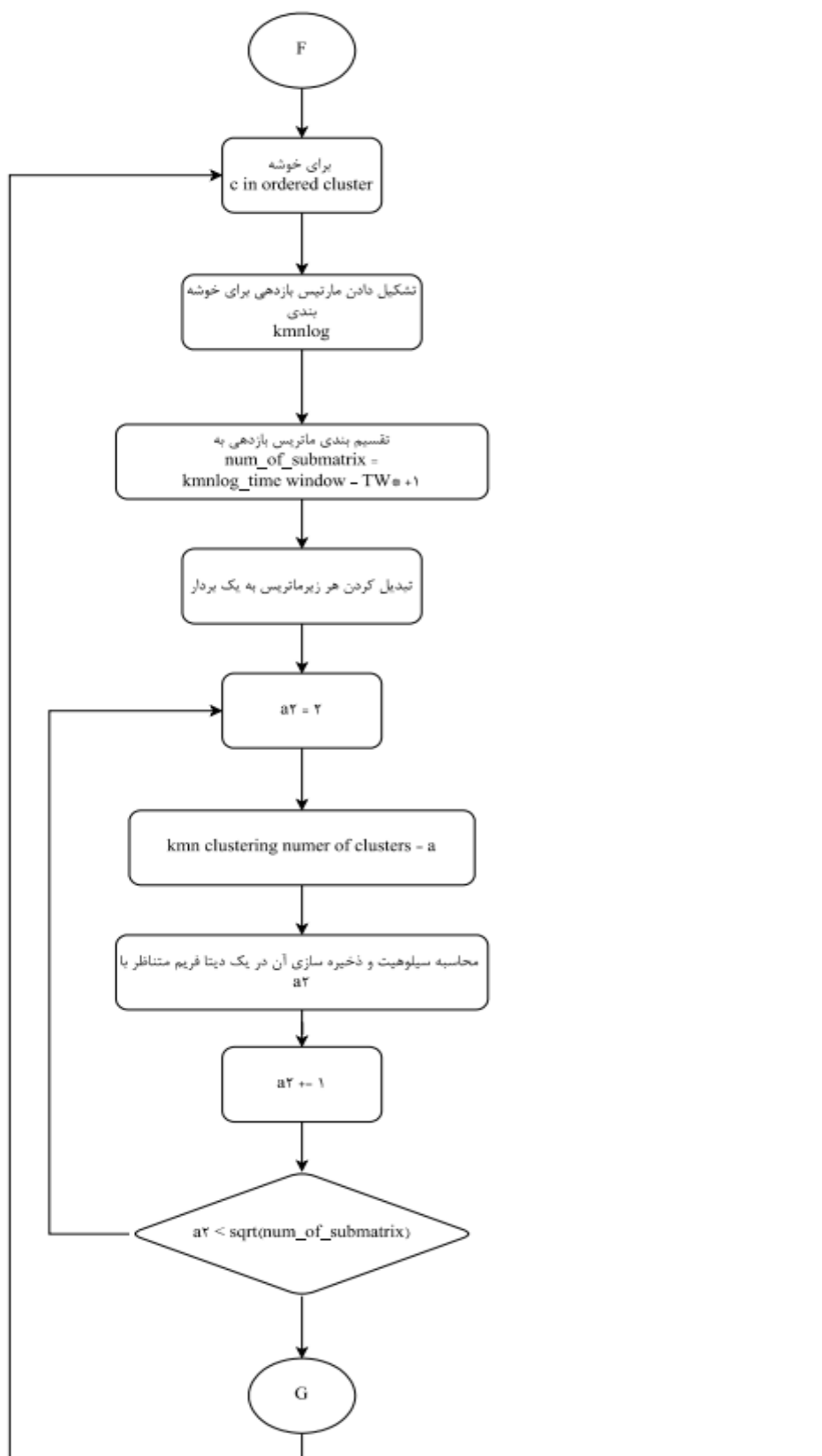
شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی(ادامه)



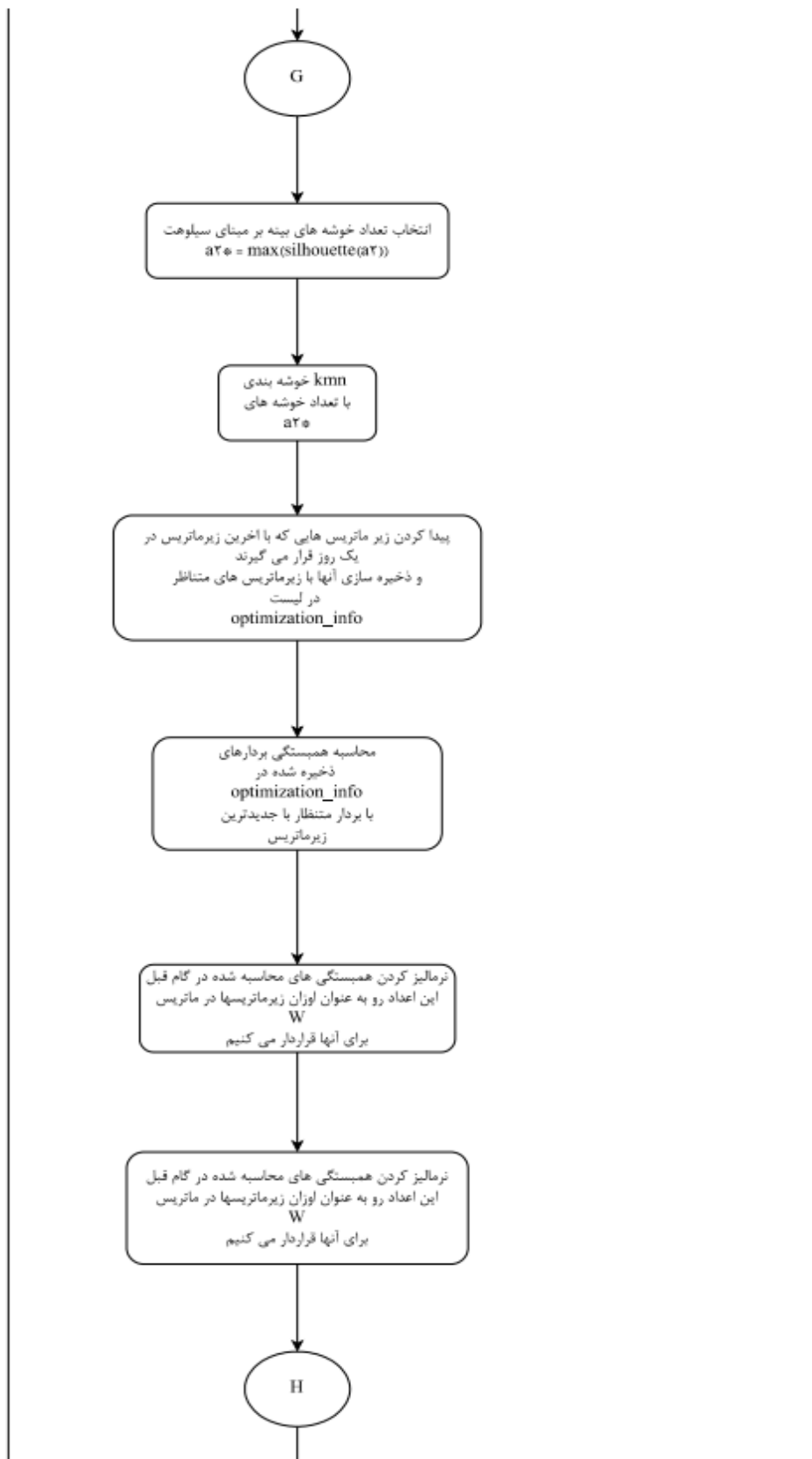
شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه‌گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی(ادامه)



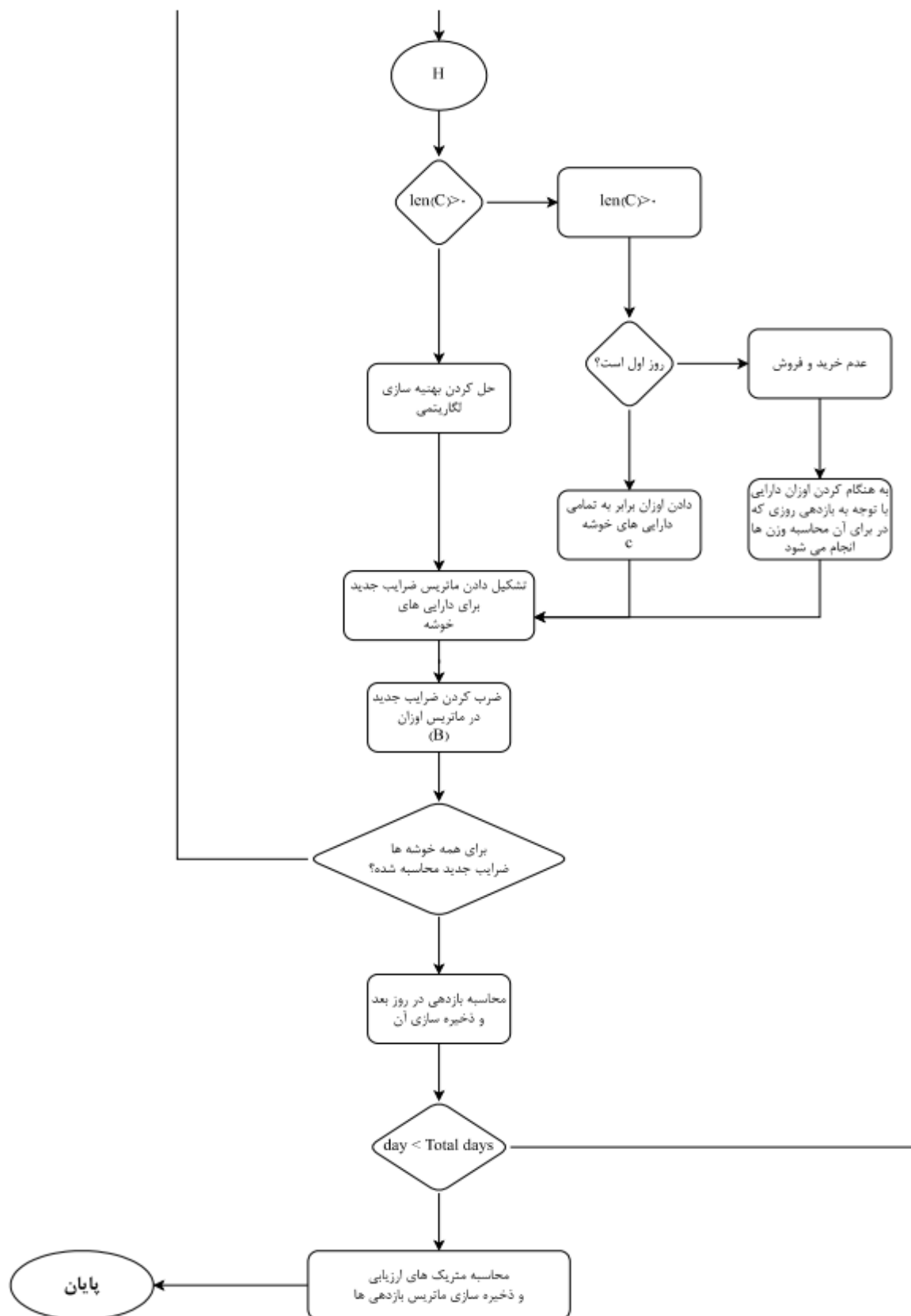
شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه‌گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی(ادامه)



شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی(ادامه)



شکل ۳-۴- فلوچارت الگوریتم انتخاب برخط سبد سرمایه‌گذاری به روش تطابق با الگو با کمک

اطلاعات جانبی(ادامه)

۳-۶- جمع بندی

در فصل جاری متدولوژی پژوهش به طور کامل تشریح شده است. الگوریتم‌های معرفی شده در این فصل به دو دسته کلی تقسیم بندی می‌شوند. دسته اول الگوریتم‌هایی هستند که از اطلاعات جانبی به منظور انتخاب سهام اولیه برای مراحل بعد استفاده می‌کند. دسته‌ی بعدی الگوریتم‌ها از اطلاعات جانبی استفاده نمی‌کند و بنابراین از تمامی نمادهای انتخاب شده برای پژوهش به منظور تشکیل پرتفو بهره می‌برد. از جمله نوآوری‌های این پژوهش نسبت به مدل‌های پیشین انتخاب بر خط سبد سرمایه‌گذاری استفاده از معیارهای کنترل ریسک به منظور انتخاب زیر مجموعه‌ی مناسب از نمادهای بازار به جهت استفاده در روش تطابق با الگوی انتخاب بر خط سبد سهام می‌باشد. نتایج تمامی این الگوریتم‌ها در فصل آتی مورد بررسی قرار گرفته است.

فصل چهارم

پیاده‌سازی مدل و نتایج پژوهش

۴-۱- مقدمه

در این فصل نتایج مبتنی بر مدل‌های توسعه داده شده که در فصل پیشین معرفی شده بودند مورد بررسی قرار می‌گیرند. در ابتدا به داده‌های استفاده شده در پژوهش پرداخته می‌شود و در ادامه نتایج و معیارهای بدست آمده از پیاده سازی هر مدل مورد بررسی قرار گرفته و در نهایت تمامی مدل‌های پیاده‌سازی شده براساس معیارهای معرفی شده مورد مقایسه قرار خواهند گرفت.

۴-۲- داده‌های ورودی پژوهش

داده‌های مورد استفاده در این پژوهش شامل داده‌های قیمتی روزانه ۲۴ شرکت بزرگ بازار نیویورک و ۵ نماد بزرگ بازار رمزارزها می‌باشند. شرکت‌های انتخاب شده از بازار نیویورک دارای ویژگی‌هایی از جمله ارزش معاملات بالا و ارزش بازار بالا و از صنایع مختلف بازار می‌باشند. نمادهای انتخاب شده از بازار رمزارزها دارای ویژگی‌هایی از جمله ارزش بازار بالاتر و همچنین سابقه طولانی‌تر نسبت به دیگر نمادها می‌باشند. اسامی این شرکت‌ها در پیوست ۱ آورده شده‌اند.

بازه زمانی مورد استفاده در پژوهش از ماه می سال ۲۰۱۷ تا ماه دسامبر سال ۲۰۲۰ می باشد که به طور مساوی به دو بخش داده های مدل و داده های تست تقسیم شده اند.

۴-۳- معیارهای ارزیابی مدل ها

به منظور بررسی و مقایسه نتایج الگوریتم های معرفی شده در این پژوهش با یکدیگر و همچنین مقایسه این مدل ها با الگوریتم های پیشین موجود در ادبیات انتخاب برخط سبد سرمایه گذاری و روش تطابق با الگو از ۶ معیار ارزیابی استفاده شده است. معیارهای مذکور به سه دسته کلی تقسیم می شوند: دسته اول معیارهایی هستند که صرفاً بازدهی را مد نظر قرار می دهند که از جمله آن ها می توان به بازدهی تجمعی نهایی و یا درصد بازدهی سالیانه اشاره کرد. دسته دوم معیارهایی هستند که تنها ریسک را مورد توجه قرار می دهند که از جمله آن ها می توان به انحراف استاندارد سالیانه و یا بیشترین نزول اشاره کرد. دسته سوم الگوریتم هایی هستند که بازده تعدیل شده با ریسک را محاسبه می کنند و از نماینده های این دسته از معیارهای ارزیابی، نسبت شارپ سالیانه و نسبت کالمار می باشند.

جدول ۱-۴_ معرفی معیارهای ارزیابی

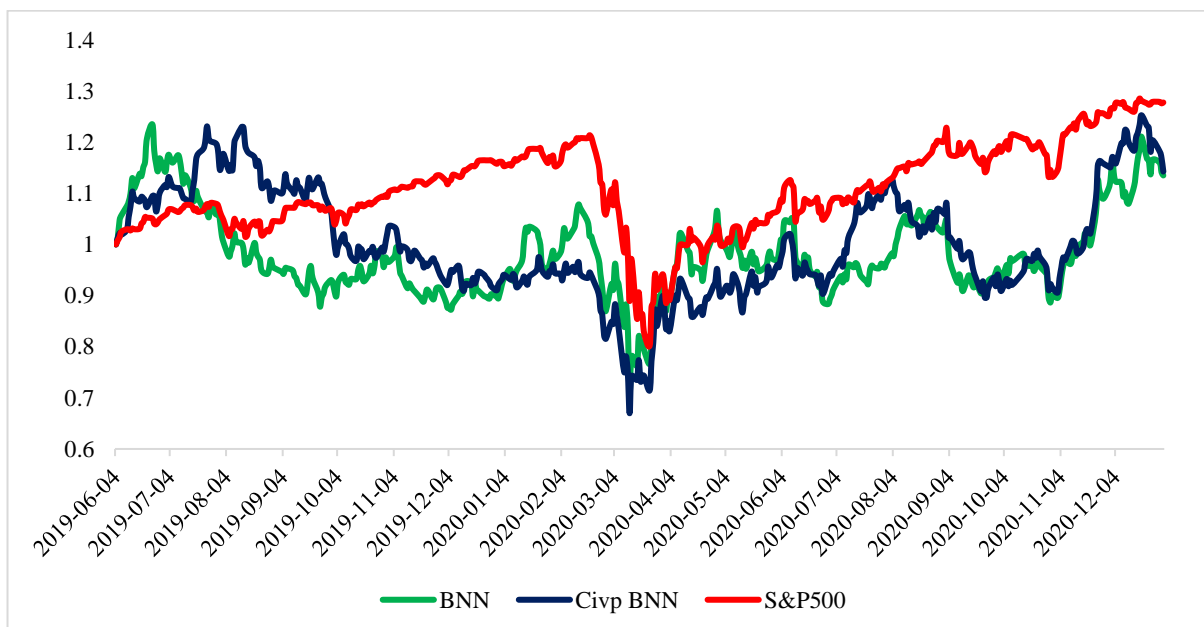
نام متریک	توضیحات	فرمول
بازدهی تجمعی نهایی (S_n)	بازده انباشته کل تغییر در قیمت سرمایه گذاری در یک زمان معین است .	$R = \prod (1 + r_i)$
بازدهی سالیانه به درصد (APY)	بازده درصدی سالانه، نرخ واقعی بازدهی است که در یک سرمایه گذاری به دست می آید، با در نظر گرفتن تأثیر بهره مرکب. برخلاف سود ساده، بهره مرکب به صورت دوره ای محاسبه می شود و مبلغ بلافاصله به مانده اضافه می شود.	$\left(1 + \frac{R}{N}\right)^N - 1$ N: تعداد سال ها
انحراف استاندارد سالیانه (SY)	انحراف استاندارد سالانه، انحراف استاندارد ضرب در جذر تعداد دوره های یک سال است.	$std(r_1, r_2, \dots, r_n)$
نسبت شارپ سالیانه	نسبت شارپ توسط ویلیام اف. شارپ عملکرد یک سرمایه گذاری را پس از کسر نرخ بازده بدون ریسک و تقسیم بر انحراف استاندارد بازده اضافی اندازه گیری می کند. نرخ بدون ریسک، نرخ بازده یک سرمایه گذاری نسبتاً مطمئن، مانند اوراق قرضه دولتی ایالات متحده است. انحراف استاندارد یک شاخص ریسک است که نوسانات قیمت را از قیمت متوسط اندازه گیری می کند. نسبت شارپ ابزار مفیدی برای سرمایه گذاران است که می خواهند بازده های تعدیل شده با ریسک پرتفوی یا دارایی های مشابه را مقایسه کنید. پورتفولیوی با نسبت شارپ بالاتر باید پس از تعدیل ریسک، عملکرد نسبتاً بهتری داشته باشد.	$\frac{\bar{r} - R_f}{\sigma_r \sqrt{N}}$ N: تعداد روزها
بیشترین نزول (MDD)	حداکثر کاهش (MDD) حداکثر زیان مشاهده شده از یک قله به پایین یک سبد، قبل از رسیدن به یک اوج جدید است. حداکثر افت، نشانگر ریسک نزولی در یک دوره زمانی مشخص است. این شاخص معمولاً بر حسب درصد بیان می شود.	$MDD = MAX \left(\frac{through - peak}{peak} \right)$
نسبت کالمار	نسبت Calmar معیاری برای عملکرد صندوق های سرمایه گذاری مانند صندوق های تأمین و مشاوران معاملات کالا (CTAs) است. این تابعی از میانگین نرخ بازده مرکب اضافی سالانه صندوق در مقابل حداکثر نزول آن است. هرچه نسبت Calmar بالاتر باشد، بر اساس ریسک تعدیل شده در بازه زمانی معین، که معمولاً ۳۶ ماه تعیین می شود، بهتر عمل می کند.	$calmar = \frac{R_p - R_f}{MDD}$

۴-۴- نتایج مدل‌های پیاده‌سازی شده

الگوریتم‌های معرفی شده در فصل سوم با استفاده از داده‌های معرفی شده در بخش‌های قبلی این فصل، پیاده سازی شده است و نتایج آن‌ها در قالب نمودارها و جداول زیر مورد بررسی قرار گرفته است.

جدول ۲-۴_الگوریتم BNN

Rank	Alg	Total_ Cumulative_return	Annual_ STD	APY	Annual_ Sharpe	MDD	Calmar
1	S&P500	1.28	0.28	1.23	4.33	0.34	3.63
2	Civp BNN	1.14	0.36	1.05	2.90	0.46	2.30
3	BNN	1.14	0.37	1.04	2.79	0.43	2.43

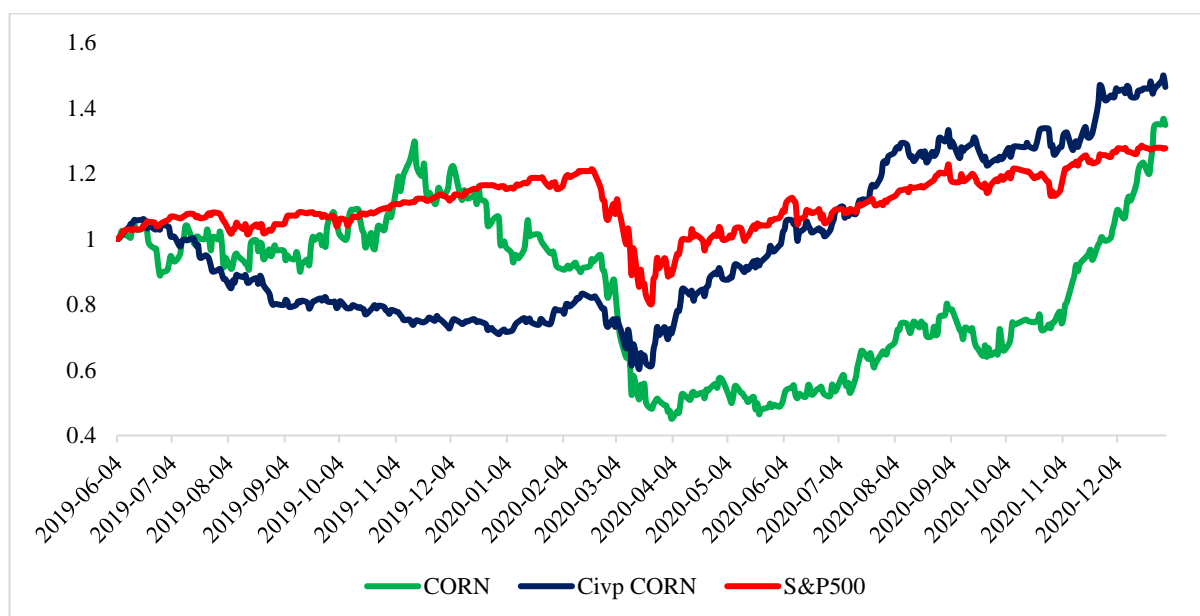


نمودار ۱-۴_ بازدهی الگوریتم BNN

همانطور که در جدول بالا مشخص است روش خوشه‌بندی و وزن‌دهی بر اساس معکوس واریانس می‌تواند از نظر معیارهای ارزیابی معرفی شده در قسمت (۳-۴)، مدل BNN را بهبود ببخشد.

جدول ۳-۴_الگوریتم CORN

Rank	Alg	Total_ Cumulative_return	Annual_ STD	APY	Annual_ Sharpe	MDD	Calmar
	Civp CORN	1.46	0.34	1.51	4.38	0.43	3.50
	S&P500	1.28	0.28	1.23	4.33	0.34	3.63
	CORN	1.35	0.55	1.33	2.44	0.65	2.04

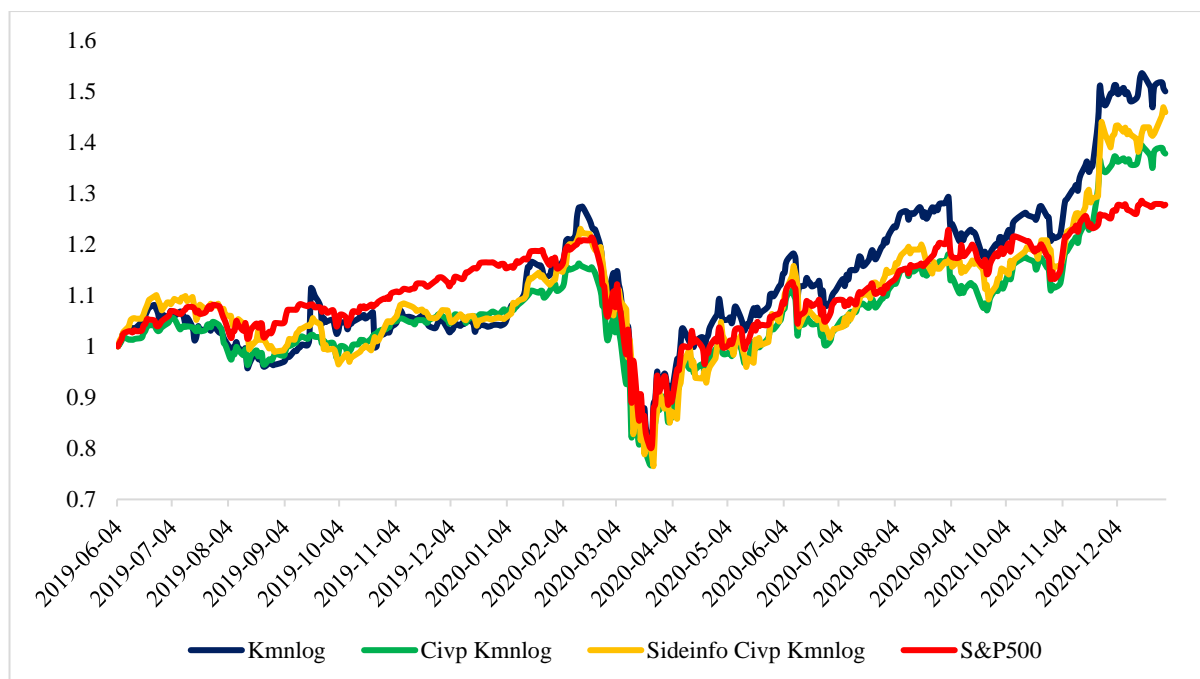


نمودار ۲-۴_بازدهی الگوریتم CORN

همانطور که در جدول بالا مشخص است روش خوشه‌بندی و وزن‌دهی بر اساس معکوس واریانس می‌تواند از نظر معیارهای ارزیابی معرفی شده در قسمت (۳-۴)، مدل CORN را بهبود ببخشد.

جدول ٤-٤_الگوریتم KMNLOG

Rank	Alg	Total_ Cumulative_return	Annual_ STD	APY	Annual_ Sharpe	MDD	Calmar
1	Kmnlog	1.50	0.33	1.57	4.78	0.36	4.36
2	Civp Kmnlog	1.38	0.29	1.38	4.72	0.34	4.06
3	Sideinfo Civp Kmnlog	1.46	0.32	1.50	4.72	0.38	3.97
4	S&P500	1.28	0.28	1.23	4.33	0.34	3.63

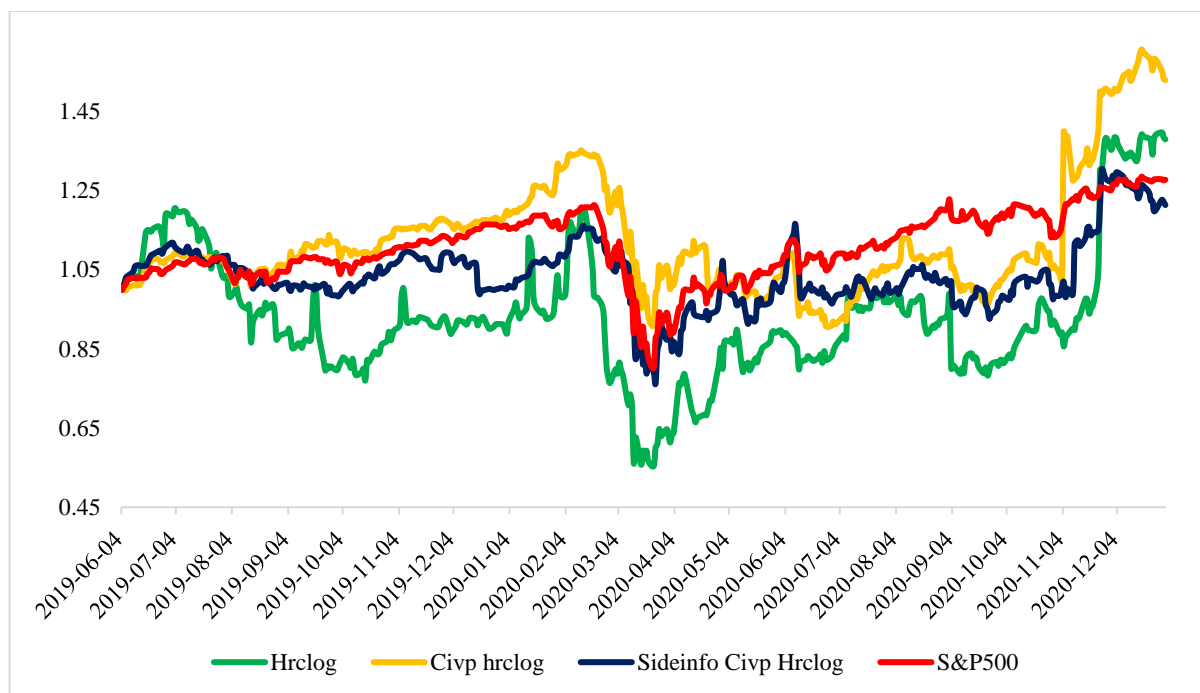


نمودار ۳-۴ بازدهی الگوریتم Kmnlog

در جدول بالا نشان داده شده است روش خوشه‌بندی و وزن‌دهی بر اساس معکوس واریانس می‌تواند از لحاظ کنترل ریسک، مدل Kmnlog را بهبود ببخشد و همانطور که در ابتدای فصل گفته شد نوآوری پیشنهاد شده در این مدل می‌تواند به کنترل ریسک سبد سهام کمک کند که با توجه داده‌ها این موضوع مشخص می‌باشد.

جدول ۵-۴ الگوریتم HRCLOG

Rank	Alg	Total_ Cumulative_return	Annual_ STD	APY	Annual_ Sharpe	MDD	Calmar
	Cvp hrclog	1.53	0.44	1.61	3.64	0.33	4.91
	S&P500	1.28	0.28	1.23	4.33	0.34	3.63
	Sideinfo Cvp Hrclog	1.21	0.36	1.14	3.18	0.34	3.32
	Hrclog	1.38	0.62	1.38	2.22	0.54	2.55

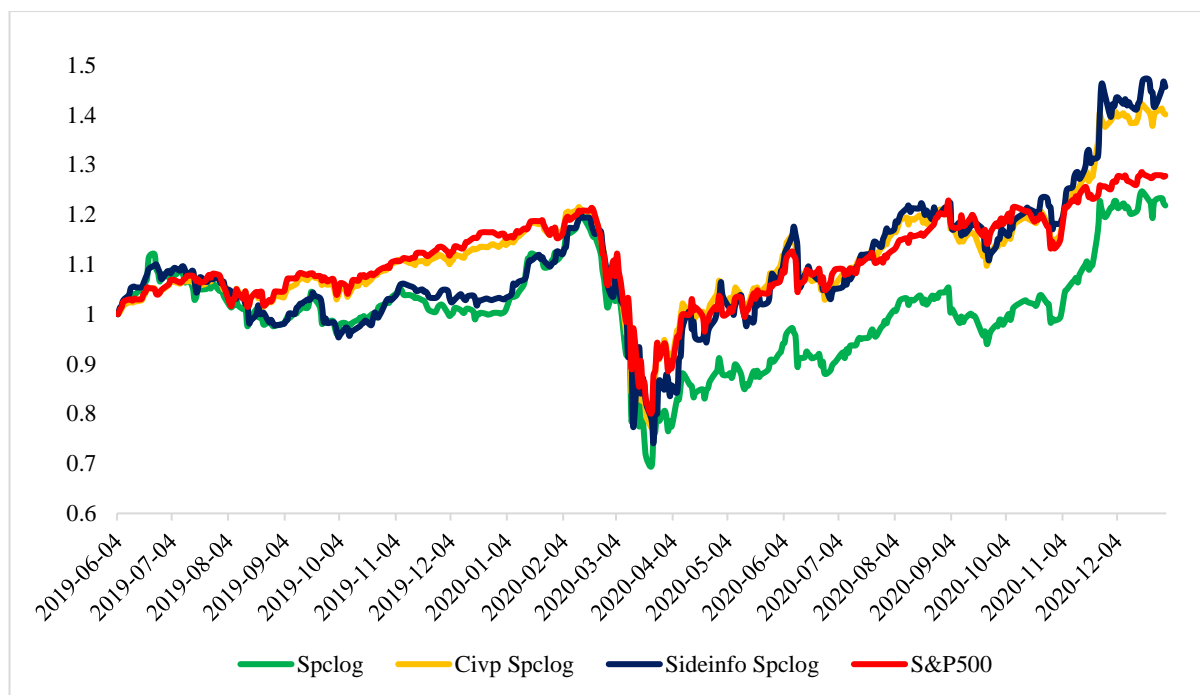


نمودار ۴-۴_ بازدهی الگوریتم Hrclog

همانطور که در جدول بالا مشخص است روش خوشه‌بندی و وزن‌دهی بر اساس معکوس واریانس می‌تواند از نظر معیارهای معرفی شده در بخش‌های قبلی، مدل Hrclog را بهبود ببخشد و همانطور که در ابتدای فصل گفته شد نوآوری پیشنهاد شده در این مدل می‌تواند به کنترل ریسک سبد سهام کمک کند که با توجه داده‌های به دست آمده این موضوع به خوبی نشان داده شده است.

جدول ۴-۶_ الگوریتم SPCLOG

Rank	Alg	Total_ Cumulative_ return	Annual_ STD	APY	Annual_ Sharpe	MDD	Calmar
	Cvp Spclog	1.40	0.29	1.41	4.87	0.37	3.87
	S&P500	1.28	0.28	1.23	4.33	0.34	3.63
	Sideinfo Spclog	1.46	0.35	1.50	4.22	0.38	3.90
	Spclog	1.22	0.32	1.15	3.61	0.42	2.75

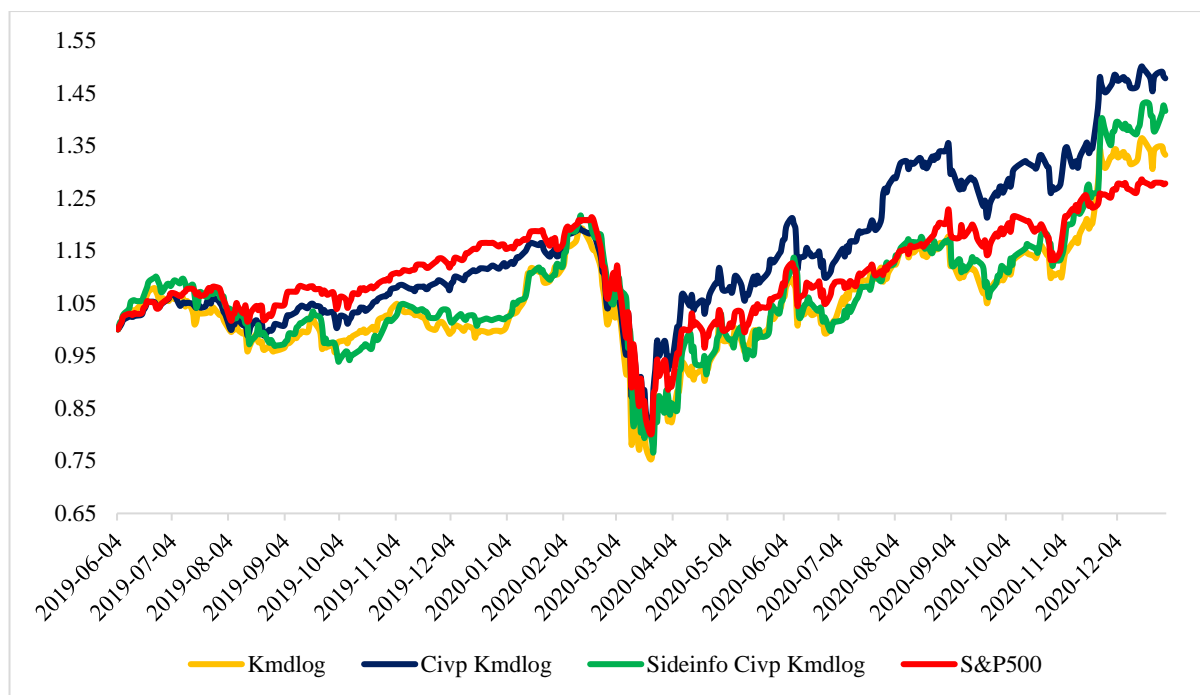


نمودار ۴_۵_ بازدهی الگوریتم Spclog

همانطور که در جدول بالا مشخص است روش خوشه‌بندی و وزن‌دهی بر اساس معکوس واریانس می‌تواند از نظر معیارهای معرفی شده در بخش‌های قبلی، مدل Spclog را بهبود ببخشد و همانطور که در ابتدای فصل گفته شد نوآوری پیشنهاد شده در این مدل می‌تواند به کنترل ریسک سبد سهام کمک کند که با توجه داده‌های به دست آمده این موضوع به خوبی نشان داده شده است.

جدول ۴_۷_ الگوریتم KMDLOG

Rank	Alg	Total_ Cumulative_return	Annual_ STD	APY	Annual_ Sharpe	MDD	Calmar
1	Cvp Kmdlog	1.48	0.30	1.53	5.09	0.32	4.81
2	Sideinfo Cvp Kmdlog	1.42	0.32	1.43	4.45	0.37	3.88
3	S&P500	1.28	0.28	1.23	4.33	0.34	3.63
4	Kmdlog	1.33	0.32	1.31	4.15	0.37	3.57



نمودار ۴-۶_ بازدهی الگوریتم Kmdlog

همانطور که در جدول بالا مشخص است روش خوشه‌بندی و وزن‌دهی بر اساس معکوس واریانس می‌تواند از نظر معیارهای معرفی شده در بخش‌های قبلی، مدل Kmdlog را بهبود ببخشد و همانطور که در ابتدای فصل گفته شد نوآوری پیشنهاد شده در این مدل می‌تواند به کنترل ریسک سبد سهام کمک کند که با توجه داده‌های به دست آمده این موضوع به خوبی نشان داده شده است.

۴-۵- مقایسه نتایج کلی

در این بخش نتایج بدست آمده با استفاده از معیارهای معرفی شده در بخش‌های قبلی این فصل مورد مقایسه قرار گرفته‌اند و نتایج نهایی آن‌ها در قالب جداول زیر می‌باشد. یکی از روش‌های که برای تصمیم‌گیری چند معیاره استفاده می‌شود، روش تاپسیس^{۱۹} می‌باشد که ما در این پژوهش با کمک روش مذکور اقدام به رتبه‌بندی مدل‌ها و الگوریتم‌های مورد استفاده کرده‌ایم.

۴_۶_ تاپسیس

تاپسیس از جمله روش‌های تصمیم‌گیری چند هدفه معیاره (MADM) می‌باشد که به گزینه‌های مورد قضاوت را با استفاده از معیارهای از پیش تعیین شده رتبه‌بندی می‌کند. در این روش از دو مفهوم حل ایده‌آل و شباهت به حل ایده‌آل استفاده می‌شود. حل ایده‌آل، آن حلی است که از هر نظر از سایر گزینه‌ها بهتر باشد که معمولاً در عمل وجود نداشته و سعی بر آن است که به آن دست پیدا کنیم.

برای اندازه‌گیری شباهت یک گزینه به حل ایده‌آل و ضد ایده‌آل، فاصله آن گزینه از حل ایده‌آل و ضد ایده‌آل محاسبه می‌شود. سپس گزینه‌ها بر اساس نسبت فاصله از حل ضد ایده‌آل به مجموع فاصله از حل ایده‌آل و ضد ایده‌آل ارزیابی و رتبه‌بندی خواهند شد. واژه TOPSIS از حروف اول عبارت Technique for Order of Preference by Similarity to Ideal Solution گرفته شده است.

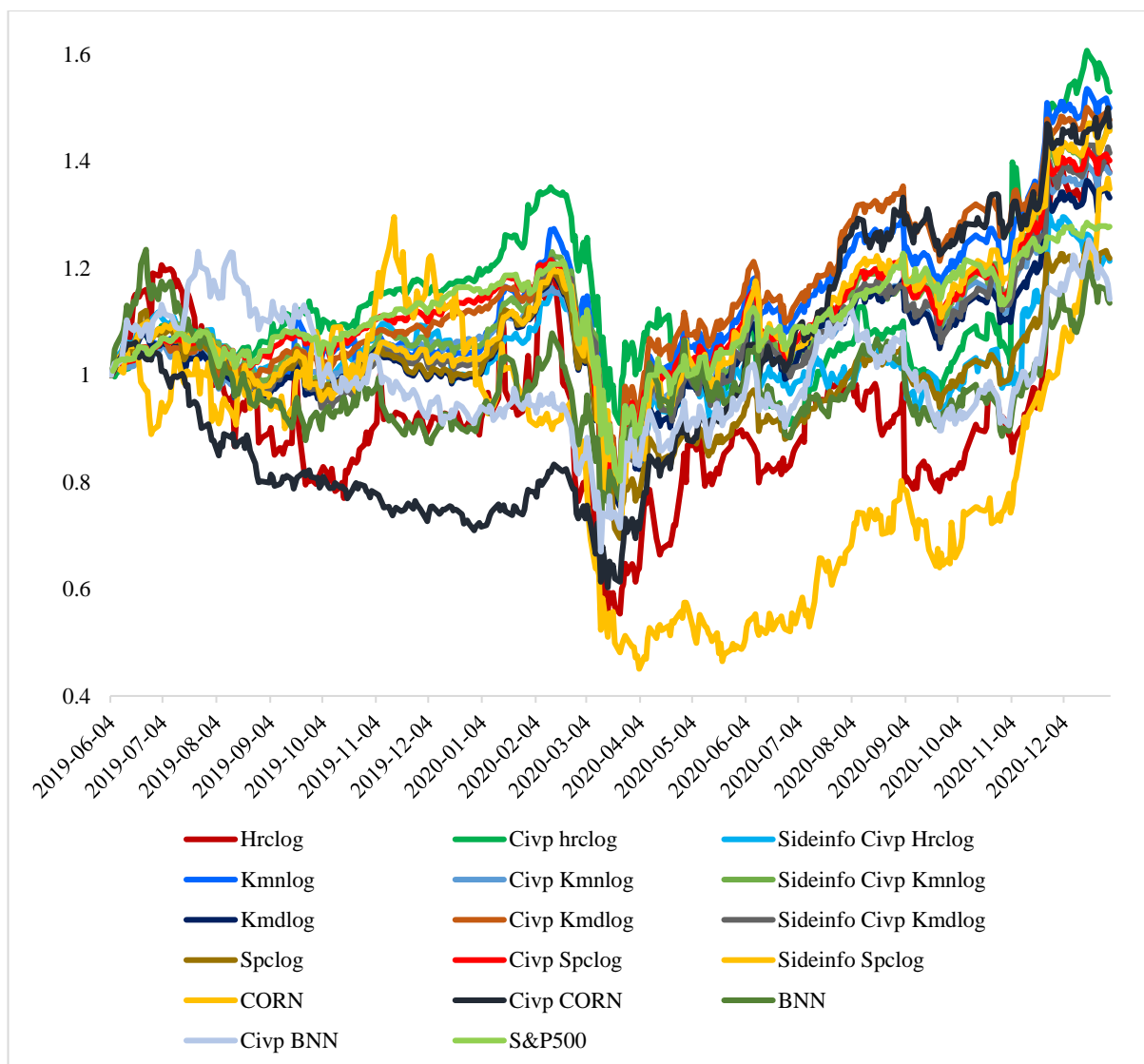
جدول ۴-۸. رتبه‌بندی الگوریتم‌های معرفی شده بر اساس روش تاپسیس

Rank	Alg	Total_ Cumulative_return	Annual_ STD	APY	Annual_ Sharpe	MDD	Calmar	TOPSIS
1	Civp Kmdlog	1.48	0.30	1.53	5.09	0.32	4.81	0.95
2	Kmnlog	1.50	0.33	1.57	4.78	0.36	4.36	0.86
3	Civp Kmnlog	1.38	0.29	1.38	4.72	0.34	4.06	0.82
4	Sideinfo Civp Kmnlog	1.46	0.32	1.50	4.72	0.38	3.97	0.80
5	Civp Spclog	1.40	0.29	1.41	4.87	0.37	3.87	0.80
6	Sideinfo Civp Kmdlog	1.42	0.32	1.43	4.45	0.37	3.88	0.77
7	Sideinfo Spclog	1.46	0.35	1.50	4.22	0.38	3.90	0.74
8	S&P500	1.28	0.28	1.23	4.33	0.34	3.63	0.73
9	Civp hrclog	1.53	0.44	1.61	3.64	0.33	4.91	0.71
10	Kmdlog	1.33	0.32	1.31	4.15	0.37	3.57	0.71
11	Civp CORN	1.46	0.34	1.51	4.38	0.43	3.50	0.69
12	Sideinfo Civp Hrclog	1.21	0.36	1.14	3.18	0.34	3.32	0.58
13	Spclog	1.22	0.32	1.15	3.61	0.42	2.75	0.56
14	BNN	1.14	0.37	1.04	2.79	0.43	2.43	0.45
15	Civp BNN	1.14	0.36	1.05	2.90	0.46	2.30	0.44
16	Hrclog	1.38	0.62	1.38	2.22	0.54	2.55	0.23
17	CORN	1.35	0.55	1.33	2.44	0.65	2.04	0.18

همانطور که در جدول فوق مشخص است، بهترین مدل از بین مدل‌های ارائه شده الگوریتم Civp Kmdlog می‌باشد. همچنین می‌توان به خوبی این موضوع را مشاهده کرد که توانسته است الگوریتم‌های موجود در حوزه ادبیات تطابق با الگو را به صورت قابل توجهی بهبود ببخشد.

جدول ۹-۴ مقایسه الگوریتم‌های معرفی شده بر اساس تمامی معیارها

Rank	Alg	Total_ Cumulative_return	Annual_ STD	APY	Annual_ Sharpe	MDD	Calmar
1	Civp Kmdlog	1.48	0.30	1.53	5.09	0.32	4.81
2	Civp hrclog	1.53	0.44	1.61	3.64	0.33	4.91
3	Civp Kmnlog	1.38	0.29	1.38	4.72	0.34	4.06
4	Kmnlog	1.50	0.33	1.57	4.78	0.36	4.36
	Spclog	1.22	0.32	1.15	3.61	0.42	2.75
	Sideinfo Spclog	1.46	0.35	1.50	4.22	0.38	3.90
	Sideinfo Civp Kmnlog	1.46	0.32	1.50	4.72	0.38	3.97
	Sideinfo Civp Kmdlog	1.42	0.32	1.43	4.45	0.37	3.88
	Sideinfo Civp Hrclog	1.21	0.36	1.14	3.18	0.34	3.32
	S&P500	1.28	0.28	1.23	4.33	0.34	3.63
	Kmdlog	1.33	0.32	1.31	4.15	0.37	3.57
	Hrclog	1.38	0.62	1.38	2.22	0.54	2.55
	CORN	1.35	0.55	1.33	2.44	0.65	2.04
	Civp Spclog	1.40	0.29	1.41	4.87	0.37	3.87
	Civp CORN	1.46	0.34	1.51	4.38	0.43	3.50
	Civp BNN	1.14	0.36	1.05	2.90	0.46	2.30
	BNN	1.14	0.37	1.04	2.79	0.43	2.43



نمودار ۴_۷_ بازدهی تمامی الگوریتم‌ها

همانطور که در جداول و نمودارهای بالا مشاهده می‌شود، بهترین مدل از بین مدل‌های ارائه شده الگوریتم Civp Kmdlog می‌باشد. این الگوریتم در یک سال و نیم، یعنی از اواسط سال ۲۰۱۹ تا انتهای سال ۲۰۲۰ حدود ۴۸ درصد بازدهی تجمعی کسب کرده است.

۴-۷- جمع‌بندی

همانطور که در بخش‌های پیشین فصل جاری نشان داده شد، الگوریتم‌های پیشنهاد شده در این پژوهش از منظرهای گوناگون نسبت به الگوریتم‌های موجود در ادبیات انتخاب برخط سبد سرمایه‌گذاری و همچنین روش تطابق با الگو عملکرد بهتری از خود نشان داده‌اند. مهم‌ترین دانش و دستاوردی که در این پژوهش به صورت تجربی به دست آمده است، توانایی کنترل و تعیین ریسک مورد نظر، برای انتخاب برخط سبد سرمایه‌گذاری می‌باشد که یکی از چالش‌های موجود در ادبیات این حوزه می‌باشد. نتایج تحقیق حاضر نشان می‌دهد با استفاده از تکنیک‌های معرفی شده می‌تواند نمادهای ورودی به مدل را فیلتر کرده و ریسک خود را کنترل نماییم. به طور کلی انتظار داریم که در بلند مدت، مدل‌های کاهش ریسک از مدل‌های حداکثر کردن بازدهی بدون در نظر گرفتن ریسک بازدهی بیشتری تولید کنند. ابتدا باید به این نکته توجه کرد که بازه انتخاب شده، یک بازه پر نوسان در هر دو بازار رمز ارزها و بازار سهام است. به طوری که بازار سهام یک ریزش ۴۰ درصدی و یک رشد ۸۰ درصدی و بازار رمز ارزها تقریباً یک روند صعودی را حفظ می‌کند اما نوسان شدیدتری نسبت به بازار سهام دارد. یکی از مزایای استفاده کردن از روش Civr، لحاظ کردن ماتریس کوواریانس به عنوان معیاری از ریسک و تعدیل کردن وزن رمزارزهاست که در نهایت به کاهش ریسک کلی سبد در بازه زمانی تست منجر می‌شود. همچنین همزمان که نزول سهام‌ها بعد از چند ماه تثبیت می‌شود، با شناخته شدن الگوی رند نزولی توسط پترن متچینگ، وزن سهام‌ها تعدیل شده و به وزن رمزارزها اضافه می‌شود پس می‌توان نتیجه گرفت در بازارهای روند دار کنترل کاهش ارزش سبد و یا رشد ارزش سبد تابعی از قدرت کلاسترینگ مدل است. همانطور که در نتایج پیداست، در این بازه زمانی مدل Civr hrc بیشترین ارزش تجمعی و را تجربه می‌کنند اما بهترین الگوریتم در میان همه‌ی مدل‌ها الگوریتم Civr Kmdlog می‌باشد. به طور کلی می‌توان گفت Civr باعث بهتر شدن هر کدام از الگوریتم‌های تست شده در مقالات موجود در این زمینه شده است.

اضافه کردن side info در بیشتر موارد بهبود قابل توجهی برای الگوریتم‌ها در پی نداشت. شاید با اینکه سهام های انتخاب شده از نظر ارزش گذاری وضعیت بهتری داشتند و انتظار داشتیم بازار اقبال بیشتری به آنها نشان دهد اما با توجه به اینکه دقت و کیفیت خوشه‌بندی در روندهای بازار می‌تواند به صورت قابل توجهی بازدهی سبد سرمایه‌گذاری را تحت تاثیر قرار دهد در نتیجه کاهش دادن دیتای کلی و حذف کردن چندین سهام به دقت خوشه‌بندی ما صدمه می‌زند.

فصل پنجم

نتیجه‌گیری و پیشنهادات برای تحقیقات آتی

۵-۱- مقدمه

در این فصل در ابتدا خلاصه‌ای از پژوهش ارائه خواهد شد و همچنین به، علل پرداختن به این مبحث، فرآیندهای تحقیق و استراتژی‌های به کار گرفته شده در تحقیق خواهیم پرداخت و در ادامه پس از نتیجه‌گیری نهایی، پیشنهاداتی برای پژوهش‌های آتی ارائه شده است.

۵-۲- خلاصه‌ای از پژوهش و جمع‌بندی

در دنیای مالی امروز، تنوع در ابزارهای مالی، نیاز به سرعت بالا در معاملات و خطاهای انسانی از جمله چالش‌های اساسی سرمایه‌گذاری در بازارهای مالی محسوب می‌شوند. این موارد در کنار پیشرفت ابزارهای یادگیری ماشین باعث شد که سرمایه‌گذاران از روش‌ها و الگوریتم‌های جدیدی برای انتخاب سبد سرمایه‌گذاری خود استفاده کنند. انتخاب بر خط سبد سرمایه‌گذاری از جمله تکنیک‌هایی است که رای حل چالش‌های فوق به کار برده می‌شود. در انتخاب بر خط سبد سرمایه‌گذاری، پس از آنکه زیر مجموعه‌ای از شرکت‌ها توسط سرمایه‌گذاران انتخاب می‌شود، سپس الگوریتم بدون دخالت سرمایه‌گذار و به صورت خودکار یک سبد سرمایه‌گذاری را تشکیل می‌دهد که هدف آن بیشینه سازی ثروت سرمایه‌گذار در انتهای افق زمانی مورد نظر می‌باشد. در این پژوهش در فصل اول به هدف پژوهش، توضیح موضوع و توجیه انتخاب موضوع پژوهش

پرداخته می‌شود. سپس در ادامه مرور کلی بر ادبیات موضوع صورت گرفته و در نهایت به کاربردها و کاربران موضوع پژوهش پرداخته شده است. در فصل دوم ادبیات موجود در این زمینه مورد مطالعه قرار گرفته است. در این فصل در ابتدا به مفاهیم موجود در زمینه داده‌کاوی پرداخته شده و در ادامه پس از ارائه مفاهیم موجود در ادبیات انتخاب برخط سبد سرمایه‌گذاری به مرور تحقیقات صورت گرفته در این زمینه پرداخته‌ایم. در حالت کلی می‌توان ادبیات موجود در زمینه انتخاب برخط سبد سرمایه‌گذاری را به چهار دسته تبعیت از برنده، تبعیت از بازنده، اصل تطابق با الگو و الگوریتم‌های فرا ابتکاری دسته‌بندی کرد. در فصل سوم به ارائه روش و متدولوژی تحقیق پرداخته شده است. با توجه به خلاءهای موجود در ادبیات انتخاب برخط سبد سرمایه‌گذاری تمرکز پژوهش بر روی روش‌های اصل تطابق با الگو و استفاده از اطلاعات جانبی، در کنار استفاده از روش HRP برای کنترل ریسک سبد سرمایه‌گذاری می‌باشد. در این روش ما با استفاده از معیار نسبت قیمت به درآمد شرکت‌ها به عنوان اطلاعات جانبی سعی در وارد کردن اطلاعات بنیادی شرکت‌ها و نمادها را به الگوریتم خود داریم. همچنین با استفاده کردن از بخشی از مدل HRP به عنوان کنترل کننده ریسک، تلاش کردیم که بی‌توجهی به ریسک را که نقص اساسی در مدل‌های انتخاب برخط سبد سرمایه‌گذاری می‌باشد را بهبود دهیم. در فصل چهارم پس از مشخص کردن داده‌های استفاده شده در تحقیق، نتایج روش‌ها و الگوریتم‌های استفاده شده در پژوهش را ارائه داده‌ایم. نتایج به دست آمده از الگوریتم‌های پیشنهادی در این پژوهش، با الگوریتم‌های موجود در زمینه انتخاب برخط سبد سرمایه‌گذاری و الگوریتم‌های اصل تطابق با الگو، بر اساس معیارهای مختلف، از جمله معیارهای تعدیل شده با ریسک مقایسه شده است و نتایج به دست آمده نشان می‌دهد روش civp Kmdlog از دیگر الگوریتم‌ها نتایج بهتری دارد.

۵-۳- پیشنهادات برای تحقیقات آتی

با توجه به مطالعات صورت گرفته در پژوهش موجود، به نظر می‌رسد موارد زیر را می‌توان به عنوان نوآوری برای بهبود نتایج در تحقیقات آتی به کار برد:

(۱) ما در این پژوهش از معیار نسبت قیمت به درآمد به عنوان اطلاعات جانبی استفاده کرده ایم. در تحقیقات آتی می توان از دیگر پارامترهای بنیادی و تکنیکال استفاده کرد و نتایج آن را بررسی نمود.

(۲) در این پژوهش محدودیت هایی مانند عدم توانایی فروش استقراضی و یا عدم توانایی استفاده از اهرم برای معاملات در نظر گرفته شده است. در نتیجه می توان مدل های ارائه شده در این پژوهش را بدون در نظر گرفتن محدودیت های مذکور توسعه داد.

(۳) استفاده از دیگر روش ها و ابزارهای یادگیری ماشین و یادگیری عمیق در گام های مختلف الگوریتم های ارائه شده می تواند منجر به بهبود نتایج در این زمینه شود. اذا استفاده از این ابزارها برای تحقیقات پیش رو جذاب به نظر می رسد.

(۴) با توجه به مزیت ها و نواقص هر کدام از الگوریتم های موجود در این زمینه، ادغام روش های مختلف می تواند منجر به نتایج مطلوب تر شود.

پیوست‌ها

پیوست ۱:

Stocks	cryptocurrencies
USB	XMR-USD
GS	XLM-USD
BAC	DASH-USD
AIG	LTC-USD
JPM	XRP-USD
BRK-B	
RY	
BLK	
GOOG	
AAPL	
IBM	
FB	
ADBE	
AMZN	
XOM	
CMS	
DUK	
JNJ	
PFE	
AMGN	
BIIB	
PEP	
KO	
MCD	

فهرست منابع

فسنقری، مهدی وهمکاران، روش های داده کاوی در بورس، انتشارات بورس وابسته به شرکت بورس، ۱۳۹۴
عبدی، متین و نجفی، امیرعباس، انتخاب برخط سبد سرمایه گذاری با استفاده از الگوریتم های تطابق با الگو،
مجله مهندسی مالی و مدیریت اوراق بهادار شماره سی و چهارم بهار ۱۳۹۷
ولیدی، جواد. ۱۳۹۷. انتخاب بر خط سبد سرمایه گذاری براساس الگوریتم های تبعیت از بازنده، پایان نامه
کارشناسی ارشد، دانشگاه صنعتی خواجه نصیرالدین طوسی، دانشکده مهندسی صنایع.

De Prado, M.L., 2018. *Advances in financial machine learning*. John Wiley & Sons.

Li, B. and Hoi, S.C., 2014. Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, 46(3), pp.1-36.

Li, B. and Hoi, S.C.H., 2018. *Online portfolio selection: principles and algorithms*. Crc Press.

Khedmati, M. and Azin, P., 2020. An online portfolio selection algorithm using clustering approaches and considering transaction costs. *Expert Systems with Applications*, 159, p.113546.

Yang, X., Li, H., Zhang, Y. and He, J.A., 2018. Reversion strategy for online portfolio selection with transaction costs. *International Journal of Applied Decision Sciences*, 11(1), pp.79-99.

Huang, D., Yu, S., Li, B., Hoi, S.C. and Zhou, S., 2018. Combination forecasting reversion strategy for online portfolio selection. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9(5), p.58.

Koyano, S. and Ikeda, K., 2017. Online portfolio selection based on the posts of winners and losers in stock microblogs. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1-4). IEEE.

Ye, Z., Huang, K., Zhou, S. and Guan, J., 2017, November. Gaussian Weighting Reversion Strategy for Accurate On-Line Portfolio Selection. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 929-936). IEEE.

Yang, X., He, J.A., Xian, J., Lin, H. and Zhang, Y., 2019. Aggregating expert advice strategy for online portfolio selection with side information. *Soft Computing*, pp.1-15.

- Yang, F., Li, X., Yang, J. and Ye, N., 2018, July. Online Newton Step for Portfolio Selection with Side Information. In *2018 5th International Conference on Information Science and Control Engineering (ICISCE)* (pp. 869-873). IEEE.
- Li, B., Wang, J., Huang, D. and Hoi, S.C., 2018. Transaction cost optimization for online portfolio selection. *Quantitative Finance*, 18(8), pp.1411-1424.
- Agarwal, A., Hazan, E., Kale, S. and Schapire, R. E. 2006. Algorithms for portfolio management based on the newton method. In *Proceedings of International Conference on Machine Learning*, Vol. 23, pp. 9–16
- Wang, Y., Wang, D., Wang, Y. and Zhang, Y. 2018. RACORN-K: RISK-AVERSION PATTERN MATCHING-BASED PORTFOLIO SELECTION. *Quantitative Finance > Risk Management* [online]. [accessed 28 Feb 2018].
- Zhang, W., Xu, Y., Zheng, F. and Dong, Y. 2012. Optimal algorithms for online time series search and one-way trading with interrelated prices. *Journal of Combinatorial Optimization* 23(2), pp. 159–166.
- Zijin, p. 2016. On-Line Portfolio Selection Strategy Based on Weighted Moving Average Asymmetric Mean Reversion. *Management Science and Engineering*, Vol. 10, No. 1, pp. 43-48
- Akcoglu, K., Drineas, P. and Kao, M.Y. 2005. Fast universalization of investment strategies. *SIAM Journal on Computing*, 34(1), pp. 1–22.
- Li, B. and Hoi, S.C., 2014. Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, 46(3), p.35.
- Ahmad, I., Mohr, E. and Schmidt, G. 2014. Risk-adjusted on-line portfolio selection, *Operations Research Proceedings 2012*, Springer, pp. 63–69.
- Albers, S. 2006. Online algorithms. *Interactive Computation: The New Paradigm*, Springer, pp. 143–164.
- Balvers, R., Wu, Y. and Gilland, E. 2000. Mean Reversion across National Stock Markets and Parametric Contrarian Investment Strategies. *The Journal of Finance*, No. 2, pp. 745–772.
- Blum, A. 1998. On-line algorithms in machine learning. *Online Algorithms: The State of the Art*, Springer, chapter 14, pp. 306–325.
- Blum, A. and Kalai, A. 1999. Universal portfolios with and without transaction costs. *Machine Learning* 35(3), pp. 193–205.
- Borodin, A., El-Yaniv, R. and Gogan, V. 2004. Can we learn to beat the best stock. *Journal of Artificial Intelligence Research* 21, pp. 579–594.

- Cover, T. M. 1991. Universal portfolios. *Mathematical Finance* 1(1), pp. 1–29.
- Cover, T. M. and Gluss, D. H. 1986. Empirical bayes stock market portfolios. *Advances in Applied Mathematics* 7(2), pp. 170–181.
- Cover, T. M. and Ordentlich, E. 1996. Universal portfolios with side information. *IEEE Transactions on Information Theory* 42(2), pp. 348–363.
- Cover, T.M. and Ordentlich, E. 1998. Universal portfolios with short sales and margin. *In Proceedings of Annual IEEE International Symposium on Information Theory*, Cambridge, MA.
- Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S. and Singer, Y. 2006. Online passive aggressive algorithms, *Journal of Machine Learning Research* 7(3), pp. 551–585.
- Crammer, K., Dredze, M. and Pereira, F. 2008. Exact convex confidence weighted learning. *Advances in Neural Information Processing Systems* 21, pp. 345–352.
- Cross, J.E. and Barron, A. R. 2003. Efficient universal portfolios for past-dependent target classes. *Mathematical Finance*, 13(2), pp. 245–276.
- Cunado, J., Gil-Alana, L.A., Fernando Perez de Gracia. 2010. Mean reversion in stock market prices: New evidence based on bull and bear markets. *Research in International Business and Finance*, pp. 113–122.
- Györfi, L., Ottucsák, G. and Urban, A. 2012. Empirical log-optimal portfolio selections: a survey. *Machine Learning for Financial Engineering*, World Scientific, chapter 2, pp. 79–115.
- Györfi, L., Udina, F. and Walk, H. 2008. Nonparametric nearest neighbor based empirical portfolio selection strategies. *Statistics and Decisions*, 26(2), pp. 145–157.
- Györfi, L., Urban, A. and Vajda, I. 2007. Kernel-based semi-log-optimal empirical portfolio selection strategies. *International Journal of Theoretical and Applied Finance* 10(3), pp. 505–516.
- Haussler, D., Kivinen, J. and Warmuth M. K. 1995. Tight worst-case loss bounds for predicting with expert advice. *Computational Learning Theory: Second European Conference, EuroCOLT '95*, pp. 69–83.
- Helmbold, D., Schapire, R., Singer, Y. and Warmuth, M. 1997. A comparison of new and old algorithms for a mixture estimation problem. *Machine Learning*, 27(1), pp. 97–119.
- Helmbold, D., Schapire, R., Singer, Y. and Warmuth, M. 1998. On-line portfolio selection using multiplicative updates. *Mathematical Finance* 8(4), pp. 325–347.

Huang, D., Zhou, J., Li, B., Hoi, S. C. and Zhou, S. 2013. Robust median reversion strategy for on-line portfolio selection. *Proceedings of the 23th International Joint Conference on Artificial Intelligence* Vol. 23, pp. 2006–2012.

Ishijima, H. 2001. Numerical methods for universal portfolios. Working Paper.

Jegadeesh, N. 1991. Seasonality in Stock Price Mean Reversion: Evidence from the U.S. and the U.K. *The Journal of Finance*, No. 4, pp. 1427–1444.

Jensen, M. C. 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* 42(2), pp. 167–247.

Kalai, A. and Vempala, S. 2002. Efficient algorithms for universal portfolios. *Journal of Machine Learning Research*, pp. 423–440.

Karp, R. M. 1992. On-line algorithms versus off-line algorithms: how much is it worth to know the future *Proceedings of the IFIP Twelfth World Computer Congress*, pp. 416–429.

Kelly, J. L. 1956. A new interpretation of information rate. *Bell System Technical Journal* 35(4), pp. 917–926.

Kivinen, J. and Warmuth, M. K. 1994. Using experts for predicting continuous outcomes. *Computational Learning Theory: EuroCOLT '93*, pp. 109–120.

Kozat, S. S. and Singer, A. C. 2007. Universal constant rebalanced portfolios with switching. *In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, pp. 1129–1132.

Kozat, S. S. and Singer, A. C. 2008. Universal switching portfolios under transaction costs. *In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, NV, pp. 5404–5407.

Kozat, S. S. and Singer, A. C. 2009. Switching strategies for sequential decision problems with multiplicative loss with application to portfolios. *IEEE Transactions on Signal Processing*, 57(6), pp. 2192–2208.

Kozat, S. S. and Singer, A. C. 2011. Universal semiconstant rebalanced portfolios. *Mathematical Finance* 21(2), pp. 293–311.

Kozat, S. S., Singer, A. C. and Bean, A.J. 2008. Universal portfolios via context trees. *In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, NV, pp. 2093–2096.

Levina, T. and Shafer, G. 2008. Portfolio selection and online learning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 16(4), pp. 437–473.

Li, B. and Hoi, S. C. H. 2012. On-line portfolio selection with moving average reversion. *Proceedings of the 29th International Conference on Machine Learning*, pp. 1–8.

Li, B. and Hoi, S. C. H. 2014. Online portfolio selection: A survey. *ACM Computing Surveys* 46(3).

Li, B. and Hoi, S. C. H. 2015. *Online Portfolio Selection: Principles and Algorithms*. CRC Press.

Li, B., Hoi, S. C. H. and Gopalkrishnan, V. 2011. Corn: Correlation-driven nonparametric learning approach for portfolio selection. *ACM Transactions on Intelligent Systems and Technology* 2(3), pp. 1–29.

Li, B., Hoi, S. C. H., Zhao, P. and Gopalkrishnan, V. 2013. Confidence weighted mean reversion strategy for online portfolio selection. *ACM Transactions on Knowledge Discovery from Data* 7(1), pp. 1–38.

Li, B., Zhao, P., Hoi, S. C. H. and Gopalkrishnan, V. 2012. PAMR: Passive aggressive mean reversion strategy for portfolio selection. *Machine Learning* 87(2), pp. 221–258.

Liberti, L., Cafieri, S. and Tarissan, F 2009. Reformulations in mathematical programming: A computational approach. *Foundations of Computational Intelligence*, Vol. 3, Springer, pp. 153–234.

Littlestone, N. and Warmuth, M. K. 1994. The weighted majority algorithm. *Inform. and Comput.* 108, pp.212–261.

Lo, A. W. and MacKinlay, A. C. 1990. Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies* 3(3), pp. 431–467.

Magdon-Ismail, M., Atiya, A. F., Pratap, A. and Abu-Mostafa, Y. S. 2004. On the maximum drawdown of a Brownian motion. *Journal of Applied Probability* 41(1), pp. 147–161.

Malin, M. and Bornholt, G. 2013. Long-term return reversal: Evidence from international market indices. *Journal of International Financial Markets, Institutions & Money* 25, pp. 1–17.

Markowitz, H. M. 1952. Portfolio selection. *Journal of Finance* 7(1), pp. 77–91.

Mayo, H. 2010. *Investments: An Introduction*, 10 edn, Cengage Learning.

McMillan, M. G., Pinto, J. E., Pirie, W. and de Venter, G. V. 2011. *Investments - Principles of Portfolio and Equity Analysis*, Wiley.

Mitra, G., Kyriakis, T., Lucas, C. and Pirbhai, M. 2003. A review of portfolio planning: models and systems. *Advances in Portfolio Construction and Implementation*, chapter 1, pp. 2–39.

Modigliani, F. and Modigliani, L. 1997. Risk-adjusted performance. *Journal of Portfolio Management* 23(2), pp. 45–54.

Mohr, E., Ahmad, I. and Schmidt, G. 2014. Online algorithms for conversion problems: A survey. *Surveys in Operations Research and Management Science* 19(2), pp. 87–104.

Mohr, E. and Schmidt, G. 2013. How much is it worth to know the future in online conversion problems, *Discrete Applied Mathematics* 161(10), pp. 1546–1555.

Rader, D. J. 2010. *Deterministic Operations Research: Models and Methods in Linear Optimization*, Wiley.

Schmidt, G., Mohr, E. and Kersch, M. 2010. Experimental analysis of an online trading algorithm. *Electronic Notes in Discrete Mathematics* 36, pp. 519–526.

Serletis, A. and Rosenberg, A.A. 2007. Mean reversion in the US stock market. *Chaos, Solitons and Fractals* 40, pp. 2007–2015.

Sharpe, W. F. 1963. A simplified model for portfolio analysis. *Management Science* 9(2), pp. 277–293.

Sharpe, W. F. 1966. Mutual fund performance. *Journal of Business* 39(1), pp. 119–138.

Singer, Y. 1997. Switching portfolios, *International Journal of Neural Systems* 8(4), pp. 445–455.

Treynor, J. L. 1965. How to rate management of investment funds *Harvard Business Review* 43(1), pp. 63.75.

Vecer, J. 2006. Maximum drawdown and directional trading. *Risk* 19(12), pp. 88–92.

Vajda, I. 2006. Analysis of semi-log-optimal investment strategies. *Proceedings of Prague stochastic*, pp. 719-727.

Vovk, V. G. 1990. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, pp. 321–383.

Vovk, V. G. and Watkins, C. 1998. Universal portfolio selection. *Proceedings of the Annual Conference on Learning Theory*, Vol. 11, pp. 12–23.

Abstract

In today's financial markets, due to the high volume and speed of transactions, the need to increase the speed of analysis and decision-making is inevitable. One of the new methods that is computer-based is algorithmic trading and among the algorithmic trading techniques is the online selection of a portfolio that can receive large amounts of data on price and stock returns as input then Perform processing operations on them and allocate capital to a certain number of stocks that rebalance the stock portfolio according to the predetermined algorithm in each trading period. These algorithms, regardless of risk, only seek to maximize profits, and this lack of attention to risk is one of the weaknesses in the online portfolio selection literature.

This research presents two online investment portfolio selection algorithms based on the principle of compliance with the model. Inspired by the HRP algorithm and using IVP models and risk parity and side information, these two methods try to control the risk of choosing an online investing portfolio.

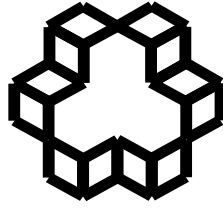
The proposed method consists of three general steps: 1) side information 2) HRP algorithm, 3) online portfolio selection.

In the first step, a subset of stocks are selected based on side information from the existing stock set.

In the second step, the selected symbols are clustered in the first step by the HRP algorithm and each cluster is assigned a weight based on their risk.

In the third stage, for each of the second stage clusters, first, clustering algorithms including k-means, k-medoids, spectral clustering and hierarchical time windows similar to the recent time window are discovered and sample selection is performed. After finding similar time windows and predicting market behavior the next day, the optimization function is used along with the transaction cost to form a portfolio. The results shows that the algorithms presented in this research perform better than the algorithms presented in the literature in this field.

Keywords: HRP, Online Portfolio Selection, Pattern Matching, Algorithmic Trading, Side Information



K. N. Toosi University of Technology

Faculty of Industrial Engineering

Using data mining for Online Portfolio Selection with Side Information

Mahdi Khazaei

:Supervisor

Dr. Amir Abbas Najafi

Master Thesis

Industrial Engineering majoring in Financial Engineering

February 2022