بسم الله الرحمن الرحيم

وزارت علوم، تحقیقات و فناوری

دانشگاه علم و فرهنگ

دانشکده فنی و مهندسی

پایان‌نامه کارشناسی ارشد مهندسی مالی

انتخاب سبد سهام ردیابی کننده شاخص با استفاده از ابزارهای داده‌کاوی

نگارش

راحله رضاوند

استاد راهنما

دکتر فریدون رهنمای رودپشتی

خرداد ۱۳۹۳

تقدیم به انسان‌های مهربانی که:

لحظات ناب باور بودن، لذت و غرور دانستن،

جسارت خواستن، عظمت رسیدن

و تمام تجربه‌های یکتا و زیبای زندگیم مدیون حضور سبز آن‌هاست

پدرم

مادرم

با تقدیر و تشکر شایسته از

استاد فرهیخته و فرزانه جناب آقای دکتر رهنمای رودپشتی که با نکته های دلاویز و گفته های بلند ، صحیفه های سخن را علم پرور نمود و همواره راهنما و راه گشای من در اتمام واکمال پایان نامه بوده است.

چکیده

ردیابی شاخص نوعی از مدیریت سرمایه‌گذاری منفعل (غیرفعال) است که به دنبال دستیابی به عملکردی مشابه شاخص از طریق سرمایه‌گذاری در تعداد معدودی از سهم‌های تشکیل دهنده‌ی شاخص می‌باشد. این پایان‌نامه مسئله انتخاب سبد بهینه ردیابی‌کننده شاخص کل قیمت بورس تهران را دنبال می‌کند که در آن با استفاده از خوشه‌بندی و دو معیار تعداد روزهای معاملاتی سهم و مجموع قدر مطلق تغییر قیمت سهم، سهم‌های ورودی به سبد سهام ردیابی کننده شاخص انتخاب می‌شوند و جهت تعیین اوزان بهینه نیز از شبکه عصبی استفاده می‌شود. در مرحله بعد از شبکه عصبی جهت تعیین سهم های ورودی به سبد سهام و تعیین اوزان بهینه استفاده می‌شود. در نهایت نتایج حاصل از این دو روش مورد مقایسه قرار می‌گیرد. نتایج مقایسه نشان می‌دهد که روش دوم به نسبت از عملکرد بهتری برخوردار است.

واژگان کلیدی: ردیابی شاخص، سبد سهام ردیابی‌کننده شاخص، داده‌کاوی، خوشه‌بندی، شبکه عصبی، مدیریت سرمایه‌گذاری غیرفعال.

فهرست مطالب

فصل ۱- مقدمه و کلیات ۱

۱-۱- مقدمه ۱

۱-۲- اهمیت و ضرورت انجام تحقیق ۲

۱-۳- سؤالات تحقیق ۲

۱-۴- هدف تحقیق ۳

۱-۵- فرضیه‌های تحقیق ۳

۱-۶- روش تحقیق ۳

۱-۷- قلمرو تحقیق ۴

۱-۸- روشهای گرد آوری اطلاعات ۴

فصل ۲- مروری بر ادبیات موضوع و پیشینه تحقیق ۵

۲-۱- بخش اول: مفاهیم و پیشینه تحقیق مرتبط با سبد سهام ردیابی‌کننده شاخص ۵

۲-۱-۱- رودیکردهای مدیریت سبد سهام ۵

۲-۱-۲- ردیابی کامل شاخص ۸

۲-۱-۳- پیشینه تحقیق مرتبط با ردیابی شاخص ۹

۲-۲- تعریف مسئله ۱۶

فصل ۳ - روش شناسی تحقیق ۲۰

۳-۱- روش تحقیق ۲۰

۳-۲- سؤالات تحقیق ۲۱

۳-۳- فرضیه‌های تحقیق ۲۱

۳-۴- قلمرو تحقیق ۲۱

۳-۵- جامعه آماری ۲۲

۳-۶- روش‌های گردآوری اطلاعات ۲۲

۳-۷- داده‌کاوی ۲۲

۳-۷-۱- مقدمه ۲۲

۳-۷-۲- داده‌کاوی و داده‌های مالی ۲۴

۳-۷-۳- مفاهیم داده‌کاوی ۲۴
۳-۷-۳-۱- مراحل (فرآیند) داده‌کاوی ۲۴
۳-۷-۳-۲- روش‌های داده‌کاوی ۲۵
۳-۷-۴- خوشه‌بندی ۲۷
۳-۷-۴-۱- روش‌های اصلی خوشه‌بندی ۲۸
۳-۷-۴-۲- مقایسه خوشه‌بندی سلسله مراتبی و غیر سلسله مراتبی ۳۰
۳-۷-۴-۳- الگوریتم K-means ۳۰
۳-۷-۴-۴- نقاط قوت و ضعف روش k-means ۳۲
۳-۷-۵- شبکه عصبی ۳۳
۳-۷-۵-۱- شبکه های عصبی تک نورونه، تک لایه، چند لایه ۳۵
۳-۷-۵-۲- انواع شبکه های عصبی مصنوعی از نظر برگشت پذیری ۳۷
۳-۷-۵-۳- مراحل طراحی یک شبکه عصبی مصنوعی ۳۸
۳-۷-۶- معرفی ساختار شبکه‌های عصبی برای حل مسائل بهینه سازی ۴۲
۳-۷-۶-۱- بروز رسانی نورون‌ها در تکرارهای متوالی ۴۷
۳-۷-۶-۲- آماده‌سازی توابع انرژی جهت حل مسئله ۴۸
۳-۷-۷- دلایل استفاده از شبکه‌های عصبی ۵۰
فصل ۴- تجزیه و تحلیل داده‌ها ۵۱
۴-۱- نمونه آماری و روش نمونه‌گیری ۵۲
۴-۲- آماده‌سازی و پردازش داده‌ها ۵۲
۴-۳- نتایج حاصل از اجرای الگوریتم k-means ۵۳
۴-۴- تعیین اوزان سهم‌های ورودی به سبد سهام ردیابی کننده شاخص ۵۵
۴-۵- نتایج حاصل از اجرای شبکه عصبی جهت تعیین اوزان بهینه ۵۷
۴-۶- طراحی شبکه عصبی جهت تشکیل سبد سهام ردیابی کننده شاخص ۵۹
۴-۷- نتایج حاصل از اجرای الگوریتم شبکه عصبی در تشکیل سبدسهام ردیابی کننده شاخص ۶۳
فصل ۵- نتیجه‌گیری و پیشنهادات ۶۶
۵-۱- نتیجه‌گیری ۶۶
۵-۲- نتیجه‌گیری نهایی ۶۷

۵-۳- پیشنهاد برای تحقیقات آتی ۶۸
نتایج حاصل از اجرای برنامه متلب جهت تعیین اوزان بهینه سهم‌های انتخابی از روش k-means ۷۲
الف-۱: تعیین اوزان بهینه برای سهم‌های انتخابی بر اساس معیار تعداد روزهای معاملاتی سهم در دوره مورد بررسی ۷۲
الف-۲: تعیین اوزان بهینه برای سهم‌های انتخابی بر اساس معیار مجموع قدر مطلق تغییر قیمت سهم در روزهای دوره مورد بررسی ۷۴

فهرست جداول

جدول ۲-۱- خلاصه‌ای از تحقیقات صورت گرفته در زمینه ردیابی شاخص ۱۴
جدول ۴-۱- تغییر در مراکز خوشه ها در تکرارهای متوالی ۵۴
جدول ۴-۲- نتایج خوشه بندی حاصل از اجرای الگوریتم k-means..... ۵۴
جدول ۴-۳- سهم‌های انتخاب شده از هر خوشه ۵۵
جدول ۴-۴- مقدار پارامترهای شبکه عصبی تعیین اوزان بهینه..... ۵۸
جدول ۴-۵- نتایج حاصل از اجرای الگوریتم شبکه عصبی در تعیین اوزان بهینه ۵۸
جدول ۴-۶- پارامترهای شبکه عصبی تعیین شده برای انتخاب سهام..... ۶۴
جدول ۴-۷- سهم های انتخاب شده توسط الگوریتم شبکه عصبی ۶۴
جدول ۴-۸- مقایسه خطای ردیابی ۶۴

فهرست نمودارها

نمودار ۲-۱- روش‌های داده کاوی..... ۳۰

فهرست اشکال

شکل ۳-۱-گام دوم الگوریتم k-means ۳۱
شکل ۳-۲-گام دوم الگوریتم k-means ۳۱
شکل ۳-۳- گام سوم اگوریتم k-means..... ۳۲

شکل ۳-۴-گام چهارم الگوریتم k-means ۳۲
شکل ۳-۵- ساختار یک نورون عصبی مصنوعی ۳۴
شکل ۳-۶- نمونه ای از یک شبکه ی تک لایه ۳۶
شکل ۳-۷- نمونه ای از یک شبکه سه لایه ۳۶
شکل ۳-۸- نمونه ای از یک شبکه پیشخور ۳۷
شکل ۳-۹- نمونه‌ای از یک شبکه برگشتی ۳۸
شکل ۳-۱۰- یادگیری با ناظر ۴۱
شکل ۳-۱۱- یادگیری بدون ناظر ۴۲
شکل۳-۱۲- نمایش چگونگی ورودی و خروجی نورون‌ها ۴۴
شکل ۳-۱۳- معماری شبکه عصبی جفت نورونی ۴۴

فصل۱

مقدمه و کلیات

۱-۱- مقدمه

سرمایه‌گذاران همواره در حال تصمیم‌گیری برای افزایش مطلوبیت انتظاری خود در سطح معینی از ریسک هستند. در گذشته این عمل با توجه به تجربیات و تخصص فردی انجام می‌گرفت و امروزه پیشرفت علم در زمینه مدیریت مالی و انجام مطالعات فراوان، سرمایه‌گذاران را در راستای اتخاذ تصمیم بهینه با بکارگیری مدل‌های مختلف یاری می نماید.

سبد سرمایه‌گذاری به ترکیبی از دارایی‌های واقعی و مالی اطلاق می‌شود که توسط افراد جهت سرمایه‌گذاری انتخاب می‌شود. از جمله دارای‌های مالی سهام می‌باشد. هدف از تشکیل سبد سهام، تقسیم‌کردن ریسک سرمایه‌گذاری بین چند سهم است؛ بدین ترتیب، سود یک سهم می‌تواند ضرر سهم دیگر را جبران کند.

تلاش برای دستیابی به عملکردی مشابه شاخص از طریق سرمایه‌گذاری در تعداد معدودی از اقلام تشکیل‌دهندهٔ آن را ردیابی شاخص می نامند. ردیابی شاخص یکی از راهبردهای کم‌ریسک تخصیص

منابع است که به اعتقاد بسیاری از صاحب‌نظران، بازدهی بالاتر را نسبت به رویکرد فعال در یک افق زمانی بلندمدت ایجاد می‌کند (شارپ، ۱۹۹۱). مزایای منحصر به فرد این روش در کاهش ریسک و کنترل هزینه منجر به رشد بی‌سابقه آن شده و بی‌شک پایه و اساس راهبردهای سرمایه‌گذاری در آینده خواهد بود. بحران مالی و عدم اطمینان از بازگشت سرمایه‌گذاری‌ها لزوم بهره‌گیری از یک راهبرد کم‌ریسک با بازدهی مناسب را بیش از پیش نمایان می‌سازد (حنیفی، ۱۳۸۸).

۱-۲- اهمیت و ضرورت انجام تحقیق

ایجاد سبد ردیابی کننده شاخص به شیوه ردیابی کامل بسیار هزینه‌بر می‌باشد و تعدیل و تغییر در این سبد سهام مشکل است. بنابراین مدیران به دنبال ایجاد سبد سهام ردیابی شامل تعداد محدودی از سهام می‌باشند. در حقیقت ایجاد چنین سبد سهامی که با عنوان مسئله ردیابی شاخص شناخته می‌شود شامل دو زیر مسئله می‌باشد: اول انتخاب زیر مجموعه‌ای از سهام‌ها و دوم اختصاص اوزان مناسب به گونه‌ای که خطای ردیابی کمینه شود و بیشترین انطباق با شاخص صورت گیرد. این مدل در گروه مسائل NP complete قرار می‌گیرد و راه حل تحلیلی مشخصی ندارد. فضای جواب چنین مسائلی بسیار بزرگ بوده و به همین دلیل اغلب محققان برای دستیابی به جواب‌های بهینه از روش‌های ابتکاری استفاده می‌نمایند (جریسون، ۲۰۰۸). به عنوان مثال در ردیابی شاخص S&P 500 با ۲۰۰ سهم از ۵۰۰ سهم، تعداد سبدهای قابل بررسی برابر با $500!/(200! * 300!) \approx 5{,}05 *$ 10^{144} می‌باشد و به منظور دستیابی به سبد سهام با کمترین خطای ردیابی ناگزیر به حل تمامی سبدهای سهام هستیم (رازقی، ۱۳۹۰). برای رفع این مشکل در این پژوهش از روش خوشه‌بندی و شبکه عصبی استفاده می‌شود. این روش‌ها تا کنون در تشکیل سبد سهام ردیاب شاخص بورس تهران به کار برده نشده است.

۱-۳- سؤالات تحقیق

چگونه می توان سبد سهام ردیابی کننده شاخص را انتخاب کرد؟

آیا روش داده کاوی برای انتخاب سبد سهام ردیابی کننده شاخص مناسب است؟

۱-۴- هدف تحقیق

هدف این تحقیق انتخاب سبد بهینه ردیابی‌کننده شاخص کل قیمت تهران با استفاده از زیر مجموعه‌ای از سهام تشکیل‌دهنده آن توسط داده‌کاوی می‌باشد.

۱-۵- فرضیه‌های تحقیق

انتخاب سبد سهام ردیابی کننده شاخص کل قیمت تهران با استفاده از داده‌کاوی برای سرمایه گذاری مناسب می باشد.

۱-۶- روش تحقیق

تحقیق فوق از نوع مدلسازی و کاربردی می‌باشد. این تحقیق چون در جستجوی دستیابی به یک هدف علمی است و اطلاعات سودمندی درباره واقعیات ارائه می دهد از نوع کاربردی است و چون برای تشکیل سبد ردیابی کننده شاخص از داده‌کاوی استفاده می‌کند از نوع مدلسازی است.در این پایان‌نامه ابتدا با استفاده از روش خوشه‌بندی تمامی سهم‌های موجود در شاخص سهام خوشه‌بندی شده و پس از تعیین خوشه‌ها جهت انتخاب یک سهم از هر خوشه دو معیار تعریف می‌شود و در نهایت سهم‌های ورودی به سبد سهام ردیابی کننده انتخاب شده و سپس جهت تعیین اوزان بهینه این سهم‌ها از یک شبکه عصبی با متغیرهای پیوسته استفاده می‌کنیم. در مرحله بعد جهت تشکیل سبد سهام از یک شبکه عصبی با دو متغیر پیوسته و گسسته برای یافتن بهترین سهام جهت ورود به سبد سهام منتخب با هدف کاهش خطای ردیابی استفاده می‌کنیم. پس از اینکه بهترین سهم‌های ممکن جهت ورود به سبد سهام را تعیین نمودیم باید وزن هر یک از آن‌ها را نیز مشخص کنیم که برای انجام این کار از شبکه عصبی که دارای متغیرهای پیوسته است استفاده می‌نماییم. در آخر جهت مقایسه نتایج این دو روش از معیار خطای ردیابی استفاده می‌نماییم.

۱-۷- قلمرو تحقیق

قلمرو زمانی تحقیق از تاریخ ۱۳۸۴/۰۸/۰۶ تا تاریخ ۱۳۸۵/۰۵/۱۰ می‌باشد. در این دوره زمانی شاخص کمترین تغییر را داشته است.

قلمرو مکانی تحقیق بازار سهام ایران می‌باشد.

۱-۸- روشهای گرد آوری اطلاعات

روش جمع آوری اطلاعات به صورت کتابخانه ای و اسنادی است. جهت گردآوری اطلاعات در زمینه مبانی نظری و ادبیات تحقیق، از بخش کتابخانه ای از طریق مراجعه به کتب، مقالات و منابع موجود در کتابخانه، جستجو در شبکه جهانی اینترنت و همچنین مطالعه و بررسی پایان نامه‌ها و رساله‌های کارشناسی ارشد و دکتری استفاده، و مطالب مرتبط با موضوع تحقیق جمع‌آوری شده است. در بخش اسنادی، ابتدا داده‌های مرتبط با موضوع پژوهش یعنی قیمت روزانه سهام شرکت‌های مختلف در بورس تهران، از پایگاه اطلاعاتی سایت مدیریت پژوهش توسعه و مطالعات اسلامی اخذ شده و سپس داده‌ها به منظور آزمونهای مورد نظر آماده‌سازی می‌شوند.

فصل ۲

مروری بر ادبیات موضوع و پیشینه تحقیق

در این فصل ابتدا به بیان خلاصه‌ای از مفاهیم و موضوعات مرتبط با انواع روش‌های مدیریت سبد سهام می‌پردازیم و سپس تحقیقات و پژوهش‌هایی که در این زمینه صورت گرفته است را مرور می‌نماییم. در بخش دوم از این فصل به معرفی مدل ریاضی که برای انتخاب سبد بهینه سهام ردیابی‌کننده شاخص طراحی شده است می‌پردازیم.

۲-۱- بخش اول: مفاهیم و پیشینه تحقیق مرتبط با سبد سهام ردیابی‌کننده شاخص

۲-۱-۱- رودیکردهای مدیریت سبد سهام

مدیریت سبد دارایی‌ها دربرگیرنده‌ی فعالیتهای متفاوت مربوط به سرمایه‌گذاری در دارایی‌ها می‌باشد. فرآیند مدیریت سبد سهام شامل: برنامه‌ریزی ، سرپرستی ، زمانبندی ، منطق‌گرائی و محافظه‌کاری در انتخاب اوراق بهادار متفاوت برای دستیابی به اهداف سرمایه‌گذاری است. این فعالیت فرآیندی در جهت انتخاب مجموعه متنوعی از دارایی‌ها برای به حداکثر رساندن بازده مورد انتظار با توجه به میزان ریسک‌پذیری یا ریسک‌گریزی فرد سرمایه‌گذار می‌باشد. در واقع هر سرمایه‌گذاری که

در بازار اقدام به سرمایه‌گذاری می‌کند از سرمایه‌گذاری صورت گرفته مقدار مشخصی بازده (سود) انتظار دارد و در مقابل مقدار مشخصی هم ریسک (عدم دستیابی به بازده مورد نظر) می‌پذیرد. در مورد هر تصمیمی که در خصوص سرمایه‌گذاری در هر بازار صورت می‌گیرد، پیوند بسیار تنگاتنگی بین ریسک و بازده هر یک از سرمایه‌گذاری‌ها وجود دارد.

مدیریت سبد سهام، مطالعه‌ی تمامی ابعاد سهام از جمله ترکیب سهم‌ها، وزن هریک از سهام موجود در سبد و زمان بهینه جهت تغییر ترکیب سبد می‌باشد. علاوه بر این، مدیریت سبد سهام باعث کاهش ریسک و افزایش بازده می‌گردد.

بطور کلی دو راهبردی که برای مدیریت دارایی‌ها و دستیابی به بازده و ریسک مورد انتظار سرمایه گذاران بکار گرفته می‌شود عبارتند از:

مدیریت فعال[1] سبد سهام: دیدگاه مدیران فعال سبد سهام این است که بازارهای مالی به طور کامل کارا نمی‌باشند بنابراین با بهره‌گیری از تجربه و دانش خود در انتخاب اوراق بهادار و یا زمانبندی مناسب تصمیمات خرید / فروش قادر به ایجاد ارزش هستند (بیزلی،۲۰۰۳).

مدیریت انفعالی[2] (غیر فعال) سبد سهام: این دسته از مدیران بر این باورند که بازارها کارا هستند و همچنین از انعطاف‌پذیری کمتری برخوردارند بنابراین ملزم به در نظرگرفتن محدودیت‌هایی برای تضمین حداقل سطحی از بازده هستند. یکی از معمول‌ترین این محدودیت‌ها دستیابی به عملکردی مشابه با شاخص مالی مبنا است.

دو روش ذکر شده جزء روش‌های خالص می‌باشند اما روش سومی نیز وجود دارد که می‌توان از آن به عنوان روش ترکیبی یاد کرد، بدین صورت که سرمایه‌گذار قسمتی از سرمایه خود را به روش فعال و مابقی را به صورت غیر فعال سرمایه‌گذاری می‌کند.

هر یک از رویکردهای مدیریت سبد سهام دارای مزایا و معایبی می‌باشند که در ادامه به آن‌ها پرداخته خواهد شد:

[1] Active Management
[2] Passive Management

مدیریت فعال دارای هزینه‌های ثابت بالایی شامل هزینه‌های مدیریتی که به تیم مدیریت پرداخته می‌شود و همچنین هزینه‌های معاملاتی که به سبب معامله سهم‌های مختلف ایجاد می‌گردد می‌باشد، در عوض این مدیران در صورتی که احساس کنند بازار با خطر سقوط مواجه است دست به تغییرات آنی می‌زنند. در مقابل مدیریت غیرفعال هزینه‌های ثابت و معاملاتی کمتری دارد اما در شرایطی که شاخص سقوط کند بازده حاصل از سبد سهام نیز کاهش پیدا می‌کند.

در مدیریت فعال یک سرمایه‌گذار در معرض هم ریسک شرکت و هم ریسک بازار قرار دارد در حالی که در مدیریت غیرفعال سرمایه‌گذار در معرض تنها ریسک بازار قرار دارد.

در مدیریت فعال، کسب بازده با توجه به مهارت در تجزیه و تحلیل و همچنین تجربه مدیران می‌تواند متفاوت باشد و حتی به حداکثر میزان خود برسد در حالی که در مدیریت غیر فعال تنها به میزان مشخصی از بازده (بازدهی برابر با بازده شاخص مورد ردیابی) می‌توان دست یافت.

اگرچه هر یک از روش‌های مدیریت فعال و انفعالی دارای مزایا و معایب مخصوص به خود می‌باشند اما امروزه استفاده از مدیریت انفعالی بسیار رواج پیدا کرده است. بسیاری از متخصصان معتقدند که رویکرد انفعالی بازدهی بالاتر را در یک بازه بلندمدت نسبت به رویکرد فعال ایجاد می‌کند .

صاحب نظران، دیدگاه‌های مختلفی را در تأیید رویکرد سرمایه‌گذاری شاخص محور ارائه کرده‌اند که به یکی از آنها اشاره می‌کنیم: چارلز الیس در مقاله‌ای با عنوان « بازی بازنده‌ها » نشان داد که ۸۵ ٪ مدیران فعال نتوانسته‌اند بازدهی بالاتر از شاخص S&P500 را در یک بازهٔ زمانی ۱۰ ساله بدست آورده و بر آن غلبه کنند. الیس بیان داشت که سرمایه‌گذاری در بازار سهام یک بازی با برایند صفر است، زیرا تمام سرمایه‌گذاران در مجموع، بازدهٔ بازار را بدست خواهند آورد. معادل یک بازنده در بازار، بایستی یک برنده وجود داشته باشد. به جای کوشش برای غلبه بر بازار، سرمایه‌گذاران بایستی بازتابی از بازار را با حداقل هزینه از طریق یک پورتفوی ردیابی کنندهٔ شاخص بدست آورند. نکته قابل توجه آن است که سرمایه‌گذاران به طور میانگین، بازده بازار منهای هزینه‌های معاملاتی را دریافت خواهند

کرد و هرچه فعال‌تر باشند، با هزینه‌های معاملاتی، تأثیرات بازار و هزینه‌های مالیاتی بیشتر مواجه خواهند بود (حنیفی، ۱۳۸۸).

۲-۱-۲- ردیابی کامل[3] شاخص

ردیابی شاخص همانطور که بیان شد عبارت است از ایجاد یک پورتفوی سرمایه‌گذاری به منظور دستیابی به عملکردی مشابه با شاخص مبنا. برای اینکه سبد سهام عملکرد شاخص را عینا تکرار کند و به بازدهی برابر با شاخص مورد نظر دست یابد می‌بایست تمامی سهام‌هایی که در شاخص موجود می‌باشد را دقیقا به همان نسبت موجود دارا باشد اما این کار مشکلاتی را در پی دارد که برخی از آن‌ها عبارتند از:

۱. اگر تعداد سهم‌های موجود در شاخص زیاد باشد جهت ردیابی کامل از تمامی سهام موجود با نسبت بسیار کمی خرید صورت می‌گیرد. این کار باعث افزایش بیش از اندازه هزینه‌های معاملاتی می‌گردد، زیرا به ازای خرید از هر سهم می‌بایست هزینه معاملاتی پرداخت گردد. از طرف دیگر این کار در عمل نامطلوب است و در برخی از بازارها به دلیل وجود محدودیت‌هایی امکان‌پذیر نمی‌باشد.
۲. ترکیب سهم‌های شاخص همیشه ثابت نیست و در طول زمان تغییر می‌کند. این تغییرات ممکن است به دلایل مختلفی رخ دهد، ممکن است شرکتی به این حد از رشد رسیده باشد که بتواند وارد شاخص شود و ورود این شرکت به شاخص باعث خروج سهام شرکت دیگری از شاخص گردد. از دلایل دیگر تغییر ترکیب شاخص ادغام چند شرکت با یکدیگر می‌باشد. به هر حال برای اینکه شاخص به طور کامل ردیابی گردد تمامی این تغییرات باید در سبد سهام منعکس گردد و این عمل بسیار مشکل است. با این شرایط خرید تمامی سهم‌های موجود در شاخص منطقی به نظر نمی‌رسد.

[3] Perfect Tracking

بنابراین با توجه به معایبی که برای ردیابی کامل ذکر گردید مدیران ترجیح می‌دهند سبد سهامی تشکیل دهند که تعداد محدودی از سهام‌ها در آن وجود داشته باشد.

۲-۱-۳- پیشینه تحقیق مرتبط با ردیابی شاخص

تحقیقات انجام شده در این زمینه بر اساس مدلسازی به سه دسته تقسیم‌بندی می‌شود: مدل‌های مارکویتزی ، مدل‌های فاکتوری ، سایر مدل‌ها (مدل‌های ابتکاری و فرا ابتکاری).

در مدل‌های مارکویتزی با استفاده از مدلی شبیه مدل میانگین – واریانس مارکویتز وزن‌های بهینه سهام‌ها به دست می‌آید. تابع هدف این مدل‌ها کمینه‌سازی خطای ردیابی می‌باشد. این مدل اولین بار توسط هاجز (۱۹۷۶) مورد استفاده قرار گرفت. رول (۱۹۹۲) مدلی با استفاده از ترکیب مدل مارکویتز و مدل عاملی با اضافه کردن محدودیت برای β برای حل مسأله مورد استفاده قرار داد. فرانگز (۱۹۹۲) از مدل مارکویتز در یک شبیه‌سازی برای بیش از ۵۰ سال و ۵ سهام مسأله را حل کرد. رودر (۱۹۹۸) هزینه مبالات را نیز در مدل خود منظور کرد. وانگ (۱۹۹۹) علاوه بر منظور کردن هزینه مبادلات در تابع هدف، تابع هدف را بر مبنای چند شاخص ارائه نمود.

در مدل‌های فاکتوری با این مفهوم که بازده سهام با فاکتورهای اقتصادی ارتباط دارد مسأله ردیابی شاخص حل می‌شود، به این صورت وزن‌های سهام از طریق حل مدل کمینه‌سازی واریانس و با این محدودیت که β پورتفوی ردیاب برابر یک باشد ، به دست می آید.

رود (۱۹۸۰) یک مدل تک فاکتوری را برای ردیابی شاخص S&P500 و با در نظر گرفتن هزینه مبادلات مورد استفاده قرار داد . هاجن و بیکر (۱۹۹۰) و کوریلی و مارسلینو (۲۰۰۶) مدل تک فاکتوری را به صورت چند فاکتوری گسترش دادند. در روش مورد استفاده آن‌ها با توجه به تبعیت شاخص و سبد سهام ردیاب از یک ساختار فاکتوری، مدل به این صورت بود که بعد از مشخص کردن فاکتورها ، سهمی که بیشترین همبستگی را با فاکتورها دارد به پورتفوی اضافه می‌شود.

مید و سالکین (۱۹۸۹) خطای ردیابی را به گونه‌ای تقریب زدند که بتوان مسئله را با استفاده از برنامه‌ریزی کوادراتیک حل کرد ولی با اجرا بر روی داده‌های واقعی بازار سهام توکیو دریافتند که این مدل کارایی بالایی ندارد و در سال ۱۹۹۰ با فرض پیروی بازده سهام وشاخص از فرآیند ARCH (واریانس ناهمسانی شرطی اتورگرسیو) و با در نظر گرفتن ضمنی هزینه مبادلات در تابع هدف و نه به شکل محدودیت، مدل خود را گسترش دادند و همچنین طبقه‌بندی مختلفی نیز برای صنایع در نظر گرفتند. معایب این روش ناکارامدی روش و عدم تاثیر طبقه‌بندی بر بهبود نتایج می‌باشد.

رادلف و ولتر (۱۹۹۹) با در نظر گرفتن اینکه خطای ردیابی در عمل به صورت انحراف خطی بازده پورتفوی ردیاب از شاخص در نظر گرفته می‌شود و این با دیدگاه آکادمیک که خطای ردیابی را به صورت غیرخطی تعریف می‌کنند در تناقض است ، چهار تعریف خطی متفاوت برای خطای ردیابی بر اساس اختلاف بازده پورتفوی ردیاب و شاخص ارائه نمودند : مدل MinMax که ماکزیمم قدر مطلق اختلاف بازده پورتفوی ردیاب و شاخص را مینیمم می کند، مدل Downside MinMax که بیشترین انحراف منفی را مینیمم می‌کند ، مدل Min Absolute Deviation (MAD) که مجموع قدر مطلق خطا را مینیمم می کند و Min Absolute Downside deviations (MADD) که مجموع انحرافات منفی را مینیمم می کند. و با اجرای این مدل های خطی دریافتند که هر کدام از این مدل ها جواب های متفاوتی ارائه می دهند .مزیت این مدل سادگی مسئله و دسترسی به جوابهای بهینه و معایب آن عدم توانایی مقایسه نتایج و بالا رفتن زمان حل با افزایش متغیرها می‌باشد.

شاپ کات (۱۹۹۲) در مقاله‌ی خود از الگوریتم ژنتیک استفاده کرده است. تابع هدف مورد استفاده او همان تابع هدف مانند مدل ساده مارکویتز می‌باشد . که به صورت زیر می‌باشد:

$$f(X) = (X - H)^T G(X - H)$$

X ماتریسی است که عناصر آن نسبتی از سرمایه که در هر یک از سهام‌ها سرمایه‌گذاری می‌شود را نشان می‌دهد و H نشان‌دهنده نسبت سهام‌ها در شاخص می‌باشد و G ماتریس کوواریانس واریانس می‌باشد. در این مقاله مجموعه‌ی کلی شامل ۱۰۰ سهام و زیر مجموعه‌ها شامل ۲۰ سهام مربوط به

شاخص FTSE-100 در نظر گرفته شده است. از مزایای این روش کارایی بالا و توانایی حل مسائل با تعداد زیاد متغیر می‌باشد. این روش دارای معایبی که عبارتند از: در نظر نگرفتن هزینه‌های عملیاتی و عدم توجه به بازده مازاد می‌باشد.

گیلی و کلزی (۲۰۰۲) از یک الگوریتم ابتکاری بر پایه آستانه پذیرش استفاده نمودند و الگوریتم جوابی را مورد پذیرش قرار می‌داد که از حد آستانه داده شده بدتر نباشد. این کار از گیر افتادن در بهینه‌های محلی جلوگیری می‌نمود. در مدل آن‌ها هزینه‌های معامله نیز مورد محاسبه قرار می‌گرفت اما به بازده مازاد به عنوان هدف توجهی نمی‌شد. بیزلی و مید (۲۰۰۳) یک روش ابتکاری با استفاده از الگوریتم ژنتیک ارائه دادند، آن‌ها هزینه‌های مبادلات را به صورت صریح وارد مدل کرده و برای هزینه کل مبادلات محدودیت قرار دادند. آن‌ها مدل خود را بر روی پنج دسته داده Hang Seng (Hong Kong) و DAX 100 (Germany) و FTSE 100 (UK) و S&P 100 (USA) وNikkei 225(Japan) اجرا کردند. از مزیتهای این مدل می‌توان به در نظر گرفتن هزینه‌های عملیاتی و بازده مازاد و همچنین استفاده از روش‌های ابتکاری به منظور محدود کردن فضای جستجو اشاره کرد و از جمله معایب آن می توان زمان حل بالا با بالا رفتن تعداد سهم‌های وارد شونده در شاخص را نام برد.

اوکی و اکمن (۲۰۰۳) معادلات خود را بر مبنای مدل بیزلی قرار داده و با تجمیع محدودیت‌ها یک مدل عدد صحیح غیرخطی ایجاد نمودند. نتایج آن‌ها در دقت و صحت تفاوت چندانی با مدل قبل (بیزلی ۲۰۰۳) نداشت ولی با وجود آن سرعت دستیابی به جواب بیشتر بود.

کلمن و همکاران (۲۰۰۴) با محدود ساختن تعداد سهام تشکیل دهنده پورتفو یک تابع هدف موزون از خطای ردیابی و تعداد سهم‌ها ایجاد نمودند و با مینیمم‌سازی آن با استفاده از یک روش برنامه ریزی کوادراتیک جوابهای بهینه را محاسبه نمودند.

جریسون و برگ (۲۰۰۵) روش الگوریتم ژنتیک را روی داده‌های Dutch AEX و انتخاب زیر مجموعه های ۱۰ عضوی از مجموعه ۲۵ سهم و با تابع هدف کمینه‌سازی خطای ردیابی اجرا کردند و با دقت