

پیش بینی بازده سهام با استفاده از روش انقباضی LASSO

سیدرضا سیدنژادفهم^۱، نرگس سهرابی^۲، هادی موقری^۳

^۱عضو هیأت علمی گروه حسابداری دانشگاه آزاد اسلامی واحد لاهیجان،

E.mail: s.rezafahim@liau.ac.ir، شماره تماس: ۰۹۱۱۳۴۳۴۱۳۲

^۲کارشناس ارشد آمار، دانشگاه صنعتی امیرکبیر، دانشکده علوم پایه، تهران، ایران،

narges.sohrabi89@gmail.com، شماره تماس: ۰۹۳۷۰۶۰۱۷۴۶

^۳کارشناس ارشد آمار، دانشگاه تربیت مدرس، دانشکده علوم ریاضی، تهران، ایران،

hadigilan@gmail.com، شماره تماس: ۰۹۳۵۲۱۸۴۱۱۳

پیش بینی بازده سهام با استفاده از روش انقباضی LASSO

چکیده

انتخاب متغیر، یکی از مراحل مهم در مدل سازی آماری است. برای این منظور، معمولاً از روش هایی نظیر حذف پسرو استفاده می شود. از آنجایی که در این روش ها دو مرحله ی برآورد مدل و انتخاب متغیر به طور جداگانه صورت می گیرد، نتیجه ی حاصل بی-ثبات خواهد بود. به همین دلیل اخیراً گروه دیگری از روش های انتخاب متغیر به نام روش های انقباضی مطرح شده اند که در این بین، LASSO از محبوبیت ویژه ای برخوردار است. در این تحقیق با استفاده از LASSO در نرم افزار R، مهم ترین متغیرهای بنیادی حسابداری موثر بر بازدهی سهام شرکت های پذیرفته در بورس اوراق بهادار تهران در بازه ی زمانی ۱۳۹۰-۱۳۸۶، شناسایی می شوند. براساس یافته های تحقیق، بازدهی دارایی ها در تمام سال ها به عنوان یک متغیر مهم برای پیش بینی بازدهی سهام برگزیده شد. در مقابل سه متغیر سرمایه گذاری ها، حاشیه سود ناخالص و تغییرات نقدینگی در هیچ یک از سال ها، انتخاب نشدند. پیچیده ترین و ساده ترین مدل، مربوط به سال های ۱۳۸۸ و ۱۳۸۹ به ترتیب با دو و شش متغیر است.

واژگان کلیدی: انتخاب متغیر، بازدهی سهام، روش انقباضی LASSO، متغیرهای بنیادی حسابداری.

۱ مقدمه

در بسیاری از مسایل عملی، تحلیل گر با تعداد زیادی متغیر مستقل مواجه است که باید، تعداد محدودی از آن‌ها را جهت استفاده در مدل نهایی انتخاب نماید. زیرا اگر مدل، حاوی متغیرهای بی‌اهمیت باشد علاوه بر آنکه تعبیر و تفسیر چنین مدلی مشکل است، نتایج حاصل از آن با عدم قطعیت زیادی همراه خواهد بود. در چنین حالتی گفته می‌شود که مدل دچار بیش برآوردی^۱ است. برعکس، مدلی که دچار کم‌برآوردی^۲ است (یعنی متغیرهای مهم از مدل حذف شوند)، داده‌های موجود را به خوبی توصیف نمی‌کند [۲۵]. پیدا کردن مدلی که گریبانگیر مسایل بیش‌برآوردی و کم‌برآوردی نباشد در حیطه‌ی انتخاب متغیر^۳ قرار می‌گیرد [۲۳] و [۱۶]. می‌توان ثابت نمود که مدل‌های کم‌برآورد، سبب اریبی برآوردهای کمترین توان‌های دوم ضرایب رگرسیونی (OLS) و اریبی مقادیر پیش‌بینی می‌شوند. همچنین مدل‌های بیش برآورد، سبب افزایش واریانس برآوردهای کمترین توان‌های دوم ضرایب رگرسیونی و افزایش واریانس مقادیر پیش‌بینی می‌گردند [۲۷].

در مسایل مالی، یکی از مهمترین معیارهای تصمیم‌گیری برای سرمایه‌گذاران، بازدهی سهام شرکت‌ها است. بر اساس مدل‌ها و نظریه‌های موجود، بسیاری از متغیرهای حسابداری بر بازدهی سهام موثر می‌باشند. حال سوال اساسی آن است که از بین انبوه متغیرهای حسابداری، کدام یک، پیشگوی بهتری برای بازدهی سهام محسوب می‌شوند؟ در صورت وجود طیف وسیعی از متغیرهای مستقل، انتخاب متغیر جز جدانشدنی مدل‌سازی آماری است. این مساله مخصوصاً در بُعدهای بی‌نهایت^۴، اهمیت دو چندان پیدا می‌کند [۲۱]. روش‌های انتخاب متغیر کلاسیک عبارتند از: انتخاب پیشرو، حذف پسرو و گام‌به‌گام [۱۰]. از آنجایی که در این

^۱ Overfitting

^۲ Underfitting

^۳ Variable Selection

^۴ High Dimensionality

روش‌ها دو مرحله‌ی برآورد مدل و انتخاب متغیر به صورت جداگانه انجام می‌شود، لذا نتیجه‌ی حاصل، بی‌ثبات خواهد بود [۱۵]. برای غلبه بر این مشکل، روش‌های انقباضی^۱ مطرح شده‌اند که نقطه‌ی آغاز آن‌ها روش ابداعی تیب‌شیرانی با عنوان LASSO^۲ می‌باشد [۳۰]. در این تحقیق با استفاده از LASSO، مهم‌ترین متغیرهای بنیادی حسابداری موثر بر بازدهی سهام شناسایی می‌شوند. این متغیرها توسط لو و تیاگاراگان معرفتی شده‌اند [۲۲]. لازم به ذکر است، اگر چه در چند تحقیق داخل ایران نظیر ثقفی و سلیمی [۵] انتخاب متغیر انجام شد، اما در آن‌ها از روش‌های انتخاب متغیر کلاسیک استفاده گردید. در واقع تحقیق انجام شده توسط ونگ و تن جز معدود مطالعاتی است که در آن از روش انقباضی برای شناسایی متغیرهای موثر بر بازدهی سهام استفاده گردید [۳۱].

۲ مبانی نظری و پیشینه تحقیق

یکی از معیارهای اساسی برای تصمیم‌گیری در بورس اوراق بهادار، بازدهی سهام می‌باشد. بازدهی سهام خود به تنهایی دارای محتوای اطلاعاتی است و بیشتر سرمایه‌گذاران بالفعل و بالقوه در تجزیه و تحلیل مالی از آن استفاده می‌نمایند [۹]. به طور کلی در مورد قابلیت پیش‌بینی بازدهی سهام، دو فرضیه متضاد معروف به فرضیه‌های رقیب وجود دارد. یکی از این فرضیه‌ها، فرضیه قدم زدن تصادفی است که بر غیرقابل پیش‌بینی بودن بازدهی سهام تأکید دارد. فرضیه‌ی مقابل، اعتقاد دارد که بر اساس مجموعه‌ای از اطلاعات می‌توان بازده را پیش‌بینی نمود [۴]. در راستای فرضیه‌ی دوم مطالعات وسیعی انجام شده است. علاوه بر تلاش‌هایی که برای پیش‌بینی بازدهی سهام با استفاده از داده‌های تاریخی در تحلیل فنی^۳ [۲۴] انجام گرفته شناسایی

^۱ Shrinkage Methods

^۲ Least Absolute Shrinkage and Selection Operator

^۳ Technical Analysis

عوامل موثر بر آن نیز یکی دیگر از شاخه‌های اصلی در تحقیقات مالی است. این عوامل با عنوان *عوامل ریسک* شناخته می‌شوند. تحقیقات زیادی برای مدل‌سازی ارتباط عوامل ریسک با بازدهی سهام صورت گرفته است: مانند مدل قیمت‌گذاری دارایی‌های سرمایه‌ای شارپ [۲۹] مدل سه عاملی فاما و فرنچ [۱۹]. علاوه بر موارد مذکور، تجزیه و تحلیل بنیادی^۱ بسیار مورد توجه واقع شده است [۲۲]. تجزیه و تحلیل بنیادی را می‌توان به عنوان هنر استفاده از اطلاعات موجود برای انجام پیش‌بینی‌های بهتر قلمداد کرد [۱۳]. در ادامه‌ی برخی از مطالعات صورت گرفته‌ی داخلی و خارجی مرتبط با موضوع را مرور می‌کنیم.

ونگ و تن با استفاده از روش انقباضی LARS، عوامل موثر بر بازدهی سهام شرکت‌های عضو بورس آمریکا را در بازه‌ی زمانی ۱۹۹۹ تا ۲۰۰۳ شناسایی نمودند. در این تحقیق طیف وسیعی از متغیرهای حسابداری (۶۵ متغیر) مورد مطالعه قرار گرفت. نتایج تحقیق نشان می‌دهد که در سال ۲۰۰۳ تنها یک متغیر بر بازدهی سهام موثر است در حالی که در سال ۲۰۰۰ بیش از ده متغیر انتخاب شدند [۳۱]. تحقیق الچ و ترابلسی نشان می‌دهد که متغیرهای بنیادی رابطه‌ی مثبت و معناداری با عملکرد آتی دارند. به گونه‌ای که پرتفویهای موفق با امتیازات بنیادی بالا، دارای عملکرد بالاتری نسبت به پرتفویهای ناموفق هستند [۱۷]. پایتروسکی تحلیل صورت‌های مالی را برای توسعه‌ی یک استراتژی سرمایه‌گذاری برای شرکت‌های با ارزش دفتری به ارزش بازار بالا به کار گرفت. نتایج وی نشان می‌دهند در بین شرکت‌های ارزشی، بازدهی اضافی شرکت‌های با قویترین عوامل بنیادی، ۲۰ درصد بیش از بازدهی اضافی شرکت‌های دارای ضعیف‌ترین عوامل بنیادی است [۲۶]. در ایران نیز مطالعات زیادی در مورد ارتباط متغیرهای حسابداری و بازدهی سهام انجام شده است. عباسی و علیدوست به بررسی تاثیر چهار متغیر بنیادی بر بازده سبد سهام پرداختند. نتایج بیانگر آن است که نسبت‌های سود به قیمت و ارزش دفتری به قیمت، تاثیر منفی و معنی‌داری بر بازده سبد سهام دارند [۸]. عارفی و دادرس در تحقیق خود، تاثیر

یازده متغیر بنیادی حسابداری را بر بازدهی سهام بررسی نمودند. نتایج تحقیق آن‌ها نشان داد که متغیرهای موجودی کالا و تغییرات گردش دارایی‌ها رابطه‌ی مثبت و معنادار و سرمایه‌گذاری‌ها رابطه‌ی منفی و معنادار با بازده دارند [۶]. پورحیدری و بیات به بررسی سودمندی سه متغیرهای بنیادی و چهار متغیرهای مبتنی بر بازار در تعیین بازده سهام شرکتها پرداختند. نتایج حاصل از این مطالعه نشان می‌دهد که از متغیرهای بنیادی، تنها نسبت ارزش دفتری به ارزش بازار مورد انتظار رابطه مثبت و معنی‌داری با بازده سهام دارد [۲۶]. عباس‌زاده و آتشی‌گلستانی رابطه‌ی بین ده متغیر حسابداری و بازدهی سهام را قبل و پس از لازم‌الاجرا شدن استانداردهای حسابداری ایران بررسی کردند. نتایج تحقیق نشان می‌دهد که توانایی پنج متغیر جمع‌داری‌ها و تغییرات آن، سود عملیاتی و تغییرات آن و ارزش دفتری حقوق صاحبان سهام، برای پیش‌بینی بازدهی سهام در دوره‌ی زمانی بعد از استانداردهای حسابداری نسبت به قبل از آن، کاهش معناداری یافته است [۷]. ثقفی و سلیمی رابطه‌ی نه متغیر بنیادی حسابداری را با بازدهی سهام شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران بررسی نمودند. آن‌ها با استفاده از روش حذف پس‌رو، اقدام به انتخاب متغیرهای مهم کردند که در انتها، سه متغیر تغییرات سودآوری، نوع گزارش حسابرس و درصد تغییرات کل دارایی‌ها شناسایی شدند [۵]. جدول ۱ برخی از تحقیقات داخلی را که در آن‌ها به بررسی ارتباط بیش از پنج متغیر حسابداری با بازدهی سهام پرداخته شده است، معرفی می‌کند. در جدول مذکور، مطالعاتی که در آن‌ها انتخاب متغیر صورت گرفت، مشخص شده‌اند. همان‌طور که ملاحظه می‌شود تنها در برخی از مطالعات، انتخاب متغیر صورت پذیرفته است. به طوری که در مطالعات مذکور از روش‌های کلاسیک گام‌به‌گام و حذف پس‌رو استفاده گردید. در این تحقیق از روش نوین انقباضی با اهداف برآورد و انتخاب همزمان متغیرهای مهم استفاده می‌شود و این، وجه تمایز تحقیق حاضر از تحقیقات گذشته است.

جدول ۱: خلاصه‌ای از تحقیقات صورت گرفته داخلی در زمینه شناسایی عوامل موثر بر بازدهی سهام

نویسنده	متغیرهای مستقل	تعداد متغیرها	انتخاب متغیر
هاشمی و همکاران (۱۳۹۱)	ساختار سرمایه، رشد مورد انتظار، اندازه، سودآوری، روند بلند مدت معکوس، تکانه قیمت سهام، ارزش شرکت، نقدینگی	۸	-
بدری و عبدالباقی (۱۳۹۰)	بازدهی دارایی‌ها، بازدهی نقدی دارایی‌ها، تفاوت جریان نقد و سود خالص عملیاتی، تغییرپذیری عایدات، ثبات رشد، رشد حساب‌های دریافتی، حاشیه سود ناخالص	۷	-
عارفی و دادرس (۱۳۹۰)	موجودی کالا، حساب‌های دریافتی، سرمایه‌گذاری، حاشیه سود ناخالص، بازدهی دارایی‌ها، تغییرات بازدهی دارایی‌ها، جریان وجوه نقد، ارقام تعهدی، تغییرات اهرم مالی، تغییرات نقدینگی و تغییرات گردش دارایی‌ها	۱۲	-
عباس‌زاده و گلستانی (۱۳۸۹)	جمع دارایی‌ها و تغییرات آن، سود عملیاتی و تغییرات آن، ارزش دفتری حقوق صاحبان سهام و تغییرات آن، سود خالص و تغییرات آن، جریان‌های نقدی عملیاتی و تغییرات آن	۱۰	حذف پسرو
پورحیدری و عطارد (۱۳۸۷)	موجودی کالا، حساب‌های دریافتی، مخارج سرمایه‌ای، حاشیه سود ناخالص، بهره‌وری نیروی انسانی، هزینه‌های اداری و فروش، نرخ موثر مالیات، کیفیت حسابرسی و هزینه‌های مالی	۹	-
نمازی و رستمی (۱۳۸۵)	نسبت جاری، نسبت آبی، سود ناویژه به فروش، سود به فروش، گردش موجودی کالا، گردش دارایی ثابت، بازدهی دارایی، بازدهی دارایی ویژه، سود هر سهم، قیمت به درآمد	۱۰	-
قائم‌ی و طوسی (۱۳۸۵)	شاخص ریسک سیستماتیک، اندازه، ارزش دفتری به ارزش بازار، حجم مبادلات، قیمت به درآمد	۵	گام به گام
ثقفی و سلیمی (۱۳۸۴)	موجودی کالا، حساب‌های دریافتی، حاشیه سود ناخالص، هزینه‌های اداری و فروش، ذخیره مطالبات، تغییرات دارایی‌ها، نوع گزارش حسابرس، سودآوری	۸	حذف پسرو
ثقفی و شعری (۱۳۸۲)	بازدهی دارایی‌ها، بازدهی سرمایه‌گذاری‌ها، رشد فروش به جمع دارایی‌ها، رشد سود خالص به فروش، هزینه‌های مالی به فروش، رشد بازدهی حقوق صاحبان سهام، رشد بدهی به ارزش ویژه، سود نقدی به سود خالص و...	۴۲	گام به گام

۳ انتخاب متغیر و روش LASSO

تاکنون معیارها و روش‌های مختلفی برای انتخاب متغیر ارائه شده‌اند که معروف‌ترین آن‌ها معیارهای اطلاع AIC^1 و BIC^2 هستند [۱۴ و ۲۸]. علیرغم کاربرد گسترده‌ی معیارهای اطلاع، استفاده‌ی از آنها باعث بی‌ثباتی در انتخاب متغیر می‌شود [۱۵]. به این معنا که اگر مجموعه‌ی داده‌ها دستخوش تغییرات اندک شود، نتیجه‌ی انتخاب متغیر به میزان زیادی تغییر می‌کند. یکی از دلایل این امر آن است که دو مرحله‌ی انتخاب متغیر و برآورد ضرایب رگرسیونی به طور جداگانه انجام می‌شود. یک راه غلبه بر این مشکل که اخیراً بسیار مورد توجه قرار گرفته استفاده از روش انقباضی است که در این بین، روش ابداعی تیب شیرانی موسوم به LASSO از مقبولیت بیشتری برخوردار است [۳۰]. روش LASSO در حقیقت یک روش بهینه‌سازی است که با وضع کردن شرایطی بر روی ضرایب رگرسیونی، سبب خروج متغیرهای بی‌اهمیت از مدل می‌شود. مدل رگرسیون خطی زیر را در نظر بگیرید:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_d x_{id} + \varepsilon_i, \quad i = 1, \dots, n$$

که در آن n تعداد مشاهدات و d تعداد متغیرهای مستقل هستند. تیب شیرانی برای برآورد ضرایب رگرسیونی از تابع هدف زیر استفاده نمود:

$$\min_{\beta} \{(y - X\beta)'(y - X\beta)\} \quad s.t \quad \sum_{j=1}^d |\beta_j| < c$$

که در آن y بردار متغیر وابسته، X ماتریس حاوی متغیرهای مستقل و β بردار ضرایب رگرسیونی هستند. در عبارت بالا c را پارامتر نظم^۳ یا آستانه^۴ می‌نامند. با تغییر مقدار این پارامتر، تعداد متغیرهایی که اجازه‌ی ورود به مدل می‌یابند، تغییر می‌کند. به طوری که اگر مقدار آن

^۱ Akaike Information Criterion

^۲ Bayesian Information Criterion

^۳ Tuning Parameter

^۴ Threshold parameter

خیلی بزرگ (کوچک) باشد، متغیرهای بیشتری (کمتری) در مدل نهایی حضور خواهند یافت. روش LASSO پنجره‌ی جدیدی را به روی مبحث انتخاب متغیر گشوده است. به طوری که بسیاری از روش‌های انتخاب متغیر نظیر SCAD، LARS و Elastic Net به نوعی با LASSO در ارتباط هستند [۱۸، ۲۰ و ۳۲].

۴ روش تحقیق

پژوهش حاضر از نوع همبستگی است. پس از استخراج اطلاعات از نرم‌افزارهایی نظیر ره‌آورد نوین، متغیرهای تحقیق توسط نرم‌افزار Excel 2007 محاسبه شدند. هم‌اکنون امکان استفاده از روش LASSO تنها در برخی از نرم‌افزارهای آماری مهیا است. در این تحقیق از بسته ی Ice در نرم‌افزار R استفاده می‌شود. نرم‌افزار R یکی از قدرتمندترین نرم‌افزارهای آماری است که دائماً با ایجاد بسته‌های نرم‌افزاری جدید، امکان استفاده از روش‌های نوین آماری را برای کاربر فراهم می‌کند.

۴-۱ فرضیه‌های تحقیق

فرضیه اصلی:

بنابر روش LASSO، تنها برخی از متغیرهای بنیادی حسابداری بر بازدهی سهام موثر هستند.

فرضیه های فرعی:

- ۱) بنابر روش LASSO، موجودی کالا بر بازدهی سهام تاثیر دارد.
- ۲) بنابر روش LASSO، حساب‌های دریافتی بر بازدهی سهام تاثیر دارد.
- ۳) بنابر روش LASSO، سرمایه‌گذاری‌ها بر بازدهی سهام تاثیر دارد.
- ۴) بنابر روش LASSO، حاشیه‌ی سود ناخالص بر بازدهی سهام تاثیر دارد.
- ۵) بنابر روش LASSO، بازدهی دارایی‌ها بر بازدهی سهام تاثیر دارد.
- ۶) بنابر روش LASSO، تغییرات بازدهی دارایی‌ها بر بازدهی سهام تاثیر دارد.

- ۷) بنابر روش LASSO، جریان وجوه نقد بر بازدهی سهام تاثیر دارد.
- ۸) بنابر روش LASSO، ارقام تعهدی بر بازدهی سهام تاثیر دارد.
- ۹) بنابر روش LASSO، تغییرات اهرم مالی بر بازدهی سهام تاثیر دارد.
- ۱۰) بنابر روش LASSO، تغییرات نقدینگی بر بازدهی سهام تاثیر دارد.
- ۱۱) بنابر روش LASSO، تغییرات گردش دارایی‌ها بر بازدهی سهام تاثیر دارد.

۴ متغیرهای تحقیق

متغیرهای تحقیق به دو دسته مستقل و وابسته تقسیم می‌شوند. متغیر وابسته، بازدهی سهام است. سایر متغیرها به عنوان متغیر مستقل در نظر گرفته می‌شوند. این متغیرها توسط لو و تیاگارajan معرفی شده و پس از آن توسط نویسندگان مختلف مورد مطالعه و بررسی قرار گرفته‌اند [۲۲]. جدول ۲ از عارفی و دادرس نحوه‌ی محاسبه‌ی این متغیرها را نشان می‌دهد [۶].

جدول ۲: نحوه محاسبه متغیرهای تحقیق (اقتباس از عارفی و دادرس، [۵])

نماد	متغیر	نحوه محاسبه
X1	موجودی کالا	= درصد تغییرات فروش منهای درصد تغییرات موجودی کالا
X2	حساب‌های دریافتی	= درصد تغییرات فروش منهای درصد تغییرات حساب‌های دریافتی
X3	سرمایه گذاری‌ها	= درصد تغییرات سرمایه‌گذاری‌ها منهای درصد تغییرات فروش
X4	حاشیه سود ناخالص	= درصد تغییرات حاشیه سود ناخالص منهای درصد تغییرات فروش
X5	بازدهی دارایی	= سود خالص قبل از ارقام غیرعادی بر کل دارایی سال قبل
X6	تغییرات بازدهی دارایی	= تغییرات بازدهی دارایی‌ها نسبت به سال قبل
X7	جریان وجوه نقد	= (سود خالص قبل از ارقام غیرعادی منهای ارقام تعهدی) بر کل دارایی سال قبل
X8	ارقام تعهدی	= تغییرات دارایی جاری منهای تغییرات بدهی جاری منهای تغییرات وجه نقد منهای هزینه استهلاک بر کل دارایی سال قبل

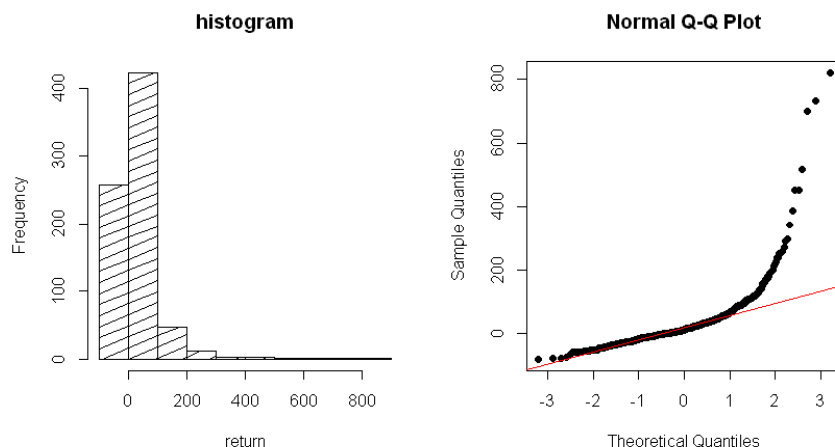
X9	تغییرات اهرم مالی	= (بدهی بلند مدت بر میانگین دارایی) منهای (بدهی بلند مدت سال قبل بر میانگین دارایی سال قبل)
X10	تغییرات نقدینگی	= (دارایی جاری بر بدهی جاری) منهای (دارایی جاری سال قبل بر بدهی جاری سال قبل)
X11	تغییرات گردش دارایی‌ها	= (فروش بر دارایی سال قبل) منهای (فروش سال قبل بر دارایی دو سال قبل)

۴ جامعه و نمونه آماری

جامعه‌ی آماری تحقیق، شامل تمام شرکت‌های عضو بورس اوراق بهادار تهران در بازه‌ی زمانی ۱۳۸۶ تا ۱۳۹۰ می‌باشد. شرکت‌هایی به عنوان نمونه آماری انتخاب شدند که (۱) سال مالی آنها منتهی به ۲۹ اسفند باشد. (۲) اطلاعات مربوط به تمام متغیرهای تحقیق در دسترس باشد. (۳) جز شرکت‌های سرمایه‌گذاری نباشند. (۴) تغییر سال مالی نداده باشند. با در نظر گرفتن شرایط مذکور در مجموع ۱۴۹ شرکت به عنوان نمونه‌ی نهایی انتخاب شدند. لازم به ذکر است که برای محاسبه‌ی متغیرهای تحقیق از اطلاعات مالی دو سال ۱۳۸۴ و ۱۳۸۵ نیز استفاده گردیده است.

۵ یافته‌های تحقیق

در این بخش با استفاده از روش انقباضی LASSO، از بین متغیرهای بنیادی حسابداری آنهایی را که بر بازده‌ی سهام تاثیرگذار هستند، شناسایی می‌کنیم. به دلیل نرمال نبودن توزیع آماری بازده سهام، از لگاریتم آن در فرایند انتخاب متغیر استفاده می‌شود. نمودارهای مستطیلی و چندک-چندک در شکل ۱، نمایی کلی از توزیع متغیر وابسته را ارائه می‌کنند.



شکل ۱: نمودار مستطیلی (سمت چپ) و چندک-چندک (سمت راست) بازدهی سهام

نتیجه‌ی حاصل از انتخاب متغیر با روش LASSO به تفکیک سال در جدول ۳ آورده شده است. این نتایج به ازای ۰/۱ برای پارامتر آستانه بدست آمده‌اند. تاثیر پارامتر نظم بر انتخاب متغیر در ادامه‌ی همین بخش بررسی می‌شود. بر اساس نتایج بدست آمده، تنها متغیری که در تمام سال‌های تحقیق به عنوان متغیر موثر بر بازدهی سهام برگزیده شده است، x_5 (بازدهی دارایی‌ها) می‌باشد. اهمیت بازدهی دارایی‌ها با نتایج بدست آمده توسط ثقفی و شعری همخوانی دارد [۵]. متغیر x_8 (اقلام تعهدی) نیز در چهار مدل (سال‌های ۱۳۸۶، ۱۳۸۷، ۱۳۸۸ و کل داده‌ها) حضور داشته است. متغیر x_{11} (تغییرات گردش دارایی‌ها) در سه مورد انتخاب گردید. در تحقیق ثقفی و شعری نیز گردش دارایی‌ها به عنوان یکی از متغیرهای مهم و موثر بر بازده معرفی گردید [۵]. هر یک از متغیرهای x_2 (حساب‌های دریافتی)، x_6 (تغییرات بازدهی دارایی‌ها)، x_7 (جریان وجوه نقد) و x_9 (تغییرات اهرم مالی) تنها در دو مدل به عنوان متغیرهای موثر بر بازدهی سهام شناسایی شده‌اند. متغیر x_1 (موجودی کالا) نیز تنها در سال ۱۳۸۹ به عنوان متغیر مهم مطرح شده است. از بین متغیرهای بنیادی مورد مطالعه، سه متغیر x_3 (سرمایه گذاری

ها)، x4 (حاشیه سود ناخالص) و x10 (تغییرات نقدینگی) در هیچ یک از سال‌های تحقیق به عنوان متغیر مهم انتخاب نشده‌اند. انتخاب نشدن حاشیه سود ناخالص، با یافته‌های ثقفی و سلیمی مطابقت می‌کند [4].

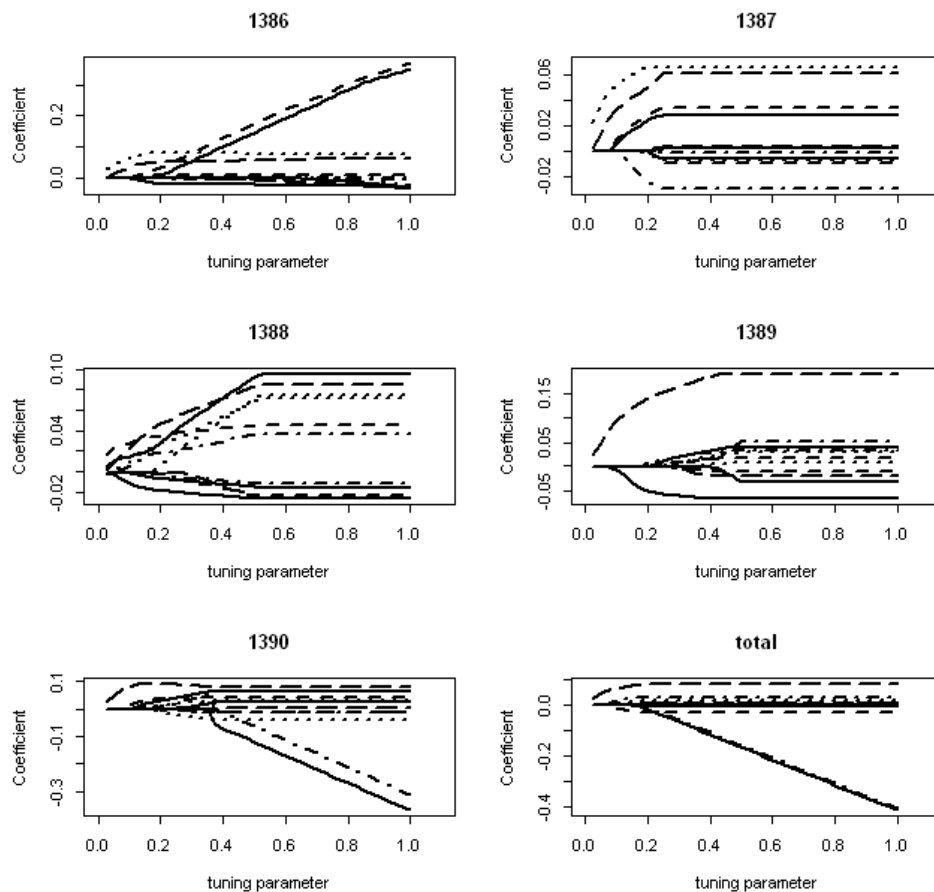
جدول ۳: نتایج انتخاب متغیر با استفاده از LASSO به تفکیک سال و در کل

متغیر مستقل	ضرایب رگرسیونی					
	کل	۱۳۹۰	۱۳۸۹	۱۳۸۸	۱۳۸۷	۱۳۸۶
ضریب ثابت	۰/۸۲۳	۰/۷۵۷	۰/۸۸۱	۰/۶۴۴	۰/۷۰۹	۰/۹۴۴
X1	-	-	-۰/۰۰۷	-	-	-
X2	-	-	-	۰/۰۲۹	۰/۰۰۹	-
X3	-	-	-	-	-	-
X4	-	-	-	-	-	-
X5	۰/۰۶۵	۰/۰۸۳	۰/۰۹۳	۰/۰۲۵	۰/۰۳۳	۰/۰۳۵
X6	-	-	-	-۰/۰۱۱	-	-۰/۰۰۰
X7	-۰/۰۱۴	۰/۰۱۶	-	-	-	-
X8	۰/۰۰۶	-	-	۰/۰۱۲	۰/۰۵۲	۰/۰۶۴
X9	۰/۰۱۵	-	-	۰/۰۰۷	-	-

-	-	-	-	-	-	X10
-	۰/۰۰۰	-	۰/۰۱۶	۰/۰۰۶	-	X11

ساده‌ترین مدل مربوط به سال ۱۳۸۹ است که تنها شامل دو متغیر بازدهی دارایی‌ها و موجودی کالا می‌باشد. پیچیده‌ترین مدل نیز متعلق به سال ۱۳۸۸ با شش متغیر است. این نتیجه را می‌توان با دهمین دوره‌ی انتخابات ریاست جمهوری در ۲۲ خرداد سال ۱۳۸۸ و حوادث پس از آن، مرتبط دانست. به طوری که حوادث مذکور سبب تغییرات اساسی در بازار سرمایه‌ی کشور شده و از کارایی متغیرهای مالی در پیش‌بینی بازدهی سهام کاسته است. ونگ و تن نیز به چنین نتایجی دست یافتند [۳۱]. آن‌ها در سال ۲۰۰۰ بیش از ۱۰ متغیر را به عنوان متغیرهای موثر بر بازدهی سهام شرکت‌های آمریکایی معرفی نمودند. در حالی که در سال ۲۰۰۳ تنها یک متغیر را انتخاب کردند. آن‌ها این نتایج را با حوادث بازدهی سپتامبر مرتبط دانستند. با توجه به نتایج بدست آمده از روش LASSO، فرضیات فرعی ۳، ۴ و ۱۰ در تمام سال‌های تحقیق رد می‌شوند. در مقابل فرضیه‌ی ۵ در تمام سال‌ها پذیرفته شده است. سایر فرضیات در برخی از سال‌ها پذیرفته و در برخی دیگر رد می‌شوند. به طور کلی، فرضیه اصلی تحقیق پذیرفته می‌شود. عملکرد روش‌های انقباضی به انتخاب مناسب پارامتر نظم بستگی دارد. برای بررسی تاثیر اندازه‌ی پارامتر نظم بر انتخاب متغیر با استفاده از روش انقباضی LASSO، به عنوان مثال برآورد ضرایب رگرسیونی در برابر چهل مقدار پارامتر نظم که به صورت $\frac{1}{40}, \frac{2}{40}, \dots, \frac{40}{40}$ در نظر گرفته شده‌اند، در شکل ۲ نشان داده شده است. مقادیر فوق، آستانه‌ی بسیار کوچک $(c = \frac{1}{40} = 0.025)$ و آستانه‌ی بسیار بزرگ $(c = \frac{40}{40} = 1)$ را در بر می‌گیرند. در شکل ۲ برآورد ضرایب رگرسیونی روی محور عمودی و مقادیر پارامتر نظم روی محور افقی نشان داده شده است. همان‌گونه که در شکل‌ها ملاحظه می‌شود، با افزایش پارامتر نظم از ۰/۰۲۵ به ۱، تعداد متغیرهای مستقل بیشتری به مدل افزوده می‌شوند. تا جایی که به ازای $c=1.0$ ، تمام

متغیرهای مستقل در مدل نهایی حضور می‌یابند. نکته‌ای که از نمودارهای شکل ۲ برداشت می‌شود، آن است که تقریباً از $c=0.2$ به بعد، به پیچیدگی مدل افزوده شده و تعداد متغیرهای مستقل بیشتری فرصت حضور در مدل را پیدا می‌کنند. بنابراین برای دستیابی به یک مدل ساده با تعداد متغیرهای مستقل کم، باید پارامتر نظم را کمتر از 0.2 در نظر گرفت. این ویژگی روش‌های انقباضی، یکی از دلایل برتری آن‌ها بر روش‌های کلاسیک انتخاب متغیر است. زیرا محقق می‌تواند، با تنظیم پارامتر نظم (بزرگ و کوچک کردن آن) پیچیدگی مدل رگرسیونی مورد نظر را به طور دلخواه کنترل نماید.



شکل ۲: نحوه تاثیر پارامتر نظم بر نتیجه انتخاب متغیر با استفاده از LASSO

۶ بحث و نتیجه گیری

انتخاب متغیر یکی از مراحل اساسی در مدل سازی آماری است. وجود متغیرهای زاید علاوه بر آنکه تفسیر مدل را مشکل می سازد، سبب افزایش عدم قطعیت پیش بینی های صورت گرفته نیز می شوند. در عین حال، عدم حضور متغیرهای مهم در مدل، سبب عدم توانایی آن در

توصیف پدیده‌ی مورد نظر می‌گردد. به منظور برقراری تعادل بین بیش‌برآوردی و کم‌برآوردی از تکنیک انتخاب متغیر استفاده می‌شود. یکی از روش‌های انتخاب متغیر که اخیراً بسیار مورد توجه قرار گرفته است، روش انقباضی می‌باشد. در این روش، دو مرحله‌ی برآورد مدل و انتخاب متغیر به طور همزمان صورت می‌گیرد. در این تحقیق، از روش انقباضی LASSO برای شناسایی مهمترین متغیرهای بنیادی موثر بر بازدهی سهام شرکت‌های عضو بورس اوراق بهادار تهران استفاده شد. یافته‌های تحقیق نشان می‌دهد، بازدهی دارایی‌ها در تمام سال‌ها به عنوان یک متغیر موثر بر بازدهی سهام مطرح است. اقلام تعهدی در درجه‌ی بعدی اهمیت قرار دارد. این متغیر در چهار مدل از شش مدل برآورد شده حضور داشته است. برخی از متغیرهای بنیادی در هیچ یک از مدل‌های برآورد شده حضور نداشتند. ساده‌ترین مدل مربوط به سال ۱۳۸۹ با دو متغیر و پیچیده‌ترین مدل نیز متعلق به سال ۱۳۸۸ با شش متغیر بود.

۷ پیشنهاد برای تحقیقات آتی

۱. تنوع روش‌های انتخاب متغیر انقباضی دائماً در حال افزایش است [۱۸، ۲۰ و ۳۲]. لذا پیشنهاد می‌شود، ارتباط بین متغیرهای بنیادی و بازدهی سهام با روش‌های جدیدتر نیز بررسی گردد.
۲. پیشنهاد می‌شود که روش انقباضی را بر روی گروه‌های دیگری از متغیرهای حسابداری نظیر نسبت‌های مالی به کار برد.
۳. در این تحقیق، ضرایب رگرسیونی تمامی متغیرهای بنیادی حسابداری در شرط بهینه-سازی، شرکت داده شدند. به این ترتیب هر یک از متغیرها ممکن است از مدل کنار گذاشته شوند. بسته‌ی نرم افزاری Ilce این امکان را برای محقق فراهم می‌کند که متغیرهای دلخواهی را در مدل نهایی بگنجانند. بنابراین پیشنهاد می‌شود، تاثیر حضور اجباری هر یک از متغیرها در مدل، بر نتیجه‌ی نهایی انتخاب متغیر بررسی شود.

منابع

۱. بدری، احمد و عبدالباقی، عبدالمجید (۱۳۹۰). "سودمندی استراتژی تجزیه و تحلیل بنیادی در کسب بازده غیرعادی"، **مجله پژوهش های حسابداری مالی**، ۹، ۳۸-۱۹.
۲. پورحیدری، امید و بیات، علی (۱۳۸۹). " بررسی سودمندی متغیرهای بنیادی و متغیرهای مبتنی بر بازار در تعیین بازده سهام"، **تحقیقات حسابداری و حسابرسی**، ۵، ۴۴۴-۴۴۴.
۳. پورحیدری، امید و حسن عطارد (۱۳۸۷). "بررسی تحلیل های بنیادی در بورس اوراق بهادار تهران"، **پژوهشنامه اقتصادی (ویژه نامه بازار سرمایه)**، ۲، ۳۹-۱۵.
۴. ثقفی، علی و سلیمی، محمدجواد (۱۳۸۴). متغیرهای بنیادی حسابداری و بازده سهام، **مجله علوم اجتماعی و انسانی دانشگاه شیراز**، ۴۳، ۷۴-۶۱.
۵. ثقفی، علی و شعری، صابر (۱۳۸۳). "نقش اطلاعات بنیادی حسابداری در پیش بینی بازده سهام"، **فصلنامه مطالعات حسابداری**، ۸، ۱۲۰-۸۷.
۶. عارفی، اصغر، دادرس، عباس (۱۳۹۰). "پیش بینی بازده سهام با استفاده از استراتژی تحلیل بنیادی"، **بررسی های حسابداری و حسابرسی**، ۶۵، ۹۸-۸۱.
۷. عباس زاده، محمدرضا و آتشی گلستانی، حجت اله. (۱۳۸۹). "متغیرهای حسابداری و پیش بینی بازده سهام (قبل و بعد از لازم الاجرا شدن استانداردهای حسابداری)"، **مجله دانش و توسعه**، ۳۳، ۲۸-۱.
۸. عباسی، ابراهیم و علیدوست اقدم، محمد (۱۳۹۱). "تاثیر متغیرهای بنیادی بر بازده سبد سهام در بورس تهران"، **تحقیقات حسابداری و حسابرسی**، ۱۶، ۴۴۴-۴۴۴.
۹. قائمی، محمدحسین و طوسی، سعید (۱۳۸۵). "بررسی عوامل موثر بر بازده سهام عادی شرکت های پذیرفته شده در بورس اوراق بهادار تهران"، **پیام مدیریت**، ۱۷،

۱۰. مونت گمری، داگلاس. سی. و پک، الیزابت (۱۳۸۲). **مقدمه‌ای بر تحلیل رگرسیون خطی**، ترجمه‌ی ابراهیم رضوی پاریزی، انتشارات دانشگاه شهید باهنر کرمان.

۱۱. نمازی، محمد، رستمی، نورالدین (۱۳۸۵). "بررسی رابطه بین نسبت های مالی و نرخ بازده سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران"، **بررسی های حسابداری و حسابرسی**، ۴۴، ۱۲۷-۱۰۵.

۱۲. هاشمی، سیدعباس، صمدی، سعید و سارکسیان، طاده (۱۳۹۱). "بررسی ارتباط متقابل عوامل موثر بر ساختار سرمایه و بازده سهام شرکت های پذیرفته شده در بورس اوراق بهادار تهران"، **فصلنامه بورس اوراق بهادار تهران**، ۱۷، ۶۷-۴۳.

۱۳. هیبتی، فرشاد و حسنی، محمد (۱۳۹۰). "بررسی مقایسه‌ای توان شاخص های اطلاعاتی حسابداری و غیرحسابداری در تبیین و پیش‌بینی شاخص های ارزش بازار سهام شرکت‌ها"، **حسابداری مدیریت**، ۹، ۸۵-۹۸.

14. Akaike, H. (1973). "Information theory an Extension of the Maximum Likelihood Principle". **Proceeding 2nd Inter Symposium on Information Theory**, 267-281, Budapest.
15. Breiman, L. (1996). "Heuristics of instability and stabilization in model selection". **Annals of Statistics**, 24, 2350-2383.
16. Burnham, K.P. and Anderson, D.R. (2002). **Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach**, Springer.
17. Elleuch J. (2009). "Fundamental Analysis Strategy and the Prediction of Stock Returns", **International Research Journal of Finance and Economics**, 30, 95-107.
18. Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). "Least angle regression". **The Annals of statistics**, 32, 407-499.
19. Fama, E. and French, K. (1993). "Common risk factors in the returns on stocks and bonds". **Journal of Financial Economics**, 33, 3-56.
20. Fan, J. and Li, R. (2001). "Variable selection via nonconcave penalized likelihood and its oracle properties". **Journal of the American Statistical Association**, 96, 1348-1360.
21. Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality:

- Feature selection in knowledge discovery. **arXiv preprint math/0602133**.
22. Lev, B. and Thiagarajan, R. (1993). "Fundamental Information Analysis", **Journal of Accounting Research**, 31, 190-215.
 23. McQuarrie, A. D. and Tsai, C. L. (1998). **Regression and time series model selection**, Singapore: World Scientific.
 24. Murphy, J. J. (1999). **Technical analysis of the financial markets: A comprehensive guide to trading methods and applications**. Prentice Hall Press.
 25. Myung, I. J. (2000). "The importance of complexity in model selection". **Journal of Mathematical Psychology**, 44, 190-204.
 26. Piotroski, J. (2000). "Value Investing: the Use of Historical Financial Statement Information To Separate Winners From Losers", **Journal of Accounting Research**, 38, 1-41
 27. Rencher, A. C. and Schaalje, G. B. (2008), **Linear Models in Statistics**, 2nd ed., John Wiley and Sons.
 28. Schwarz, G. (1978). "Estimating the Dimension of a Model", **the Annals of Statistics**, 6, 461- 464.
 29. Sharpe, W. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", **the Journal of Finance**, 19, 425 – 442.
 30. Tibshirani, R. J. (1996). "Regression Shrinkage and Selection Via the LASSO", **Journal of the Royal Statistical Society Ser B**, 58, 267- 288.
 31. Wang, Z. and Tan, S. (2009). "Identifying idiosyncratic stock return indicators from large financial factor set via least angle regression". **Expert Systems with Applications**, 36, 8350-8355.
 32. Zou, H. and Hastie, T. (2005). "Regularization and variable selection via the elastic net". **Journal of the Royal Statistical Society: Series B**, 67, 301-320.

Stock Return Prediction via LASSO Shrinkage Method

Seyed reza seyed nezhad fahim¹, Narges Sohrabi², Hadi Movaghari³

¹ Member of Faculty, Department of Accounting, Lahijan Branch, Islamic Azad University, Lahijan, Iran

² Amirkabir University, MSc. of Statistics, Faculty of Science, Iran, Tehran.

³ Tarbiat Modares University, Department of Statistics, MSc of Statistics Iran, Tehran.

Abstract

Variable selection is one of the most important steps in statistical modeling. For this purpose, typically backward elimination technique is used. Since in these methods two stage model fitting and variable selection done separately, the results is unstable. Recently, another group of variable selection methods called shrinkage methods have been proposed. Among them, LASSO Enjoys special popularity. In this article, the most important fundamental accounting variables affecting on stock return of listed companies in Tehran Stock Exchange identified by LASSO in R in the period 1386-1390. Based on the findings, the return on assets is selected in all the years. In contrast three variables investment, gross profit margin and changes in equity were not selected in any years. The Simplest and most complex models belong to 1388 and 1389, respectively with two and six variables.

Keywords: Variable Selection, Stock Return, LASSO Shrinkage Method, Fundamental Accounting Variables.