

مطالعه‌ای در خصوص انواع روش‌های رگرسیونی در پژوهش‌های حسابداری و مالی

سیداحمد میرزائی^۱، پرویز سعیدی^۲

^۱ گروه مهندسی مالی، واحد علی‌آبادکتول، دانشگاه آزاد اسلامی، علی‌آبادکتول، ایران (نویسنده مسئول)

^۲ گروه حسابداری و مالی، واحد علی‌آبادکتول، دانشگاه آزاد اسلامی، علی‌آبادکتول، ایران

چکیده

استفاده از داده‌ها به منظور کشف رابطه بین آن‌ها اساس داده‌کاوی است. یکی از ابزارهای سنجش رابطه، مدل‌سازی و پیش‌بینی استفاده از ابزار آماری رگرسیون است. تکنیک‌های رگرسیون یکی از محبوب‌ترین تکنیک‌های آماری است که برای داده‌کاوی، مدل‌سازی و پیش‌بینی مورد استفاده قرار می‌گیرد. امروزه به منظور تحلیل و کشف مدل روی داده‌های کلان، روش‌های مختلف رگرسیون توسعه یافته است. استفاده از تحلیل رگرسیون در علوم مختلف مانند علوم انسانی، علوم زیستی، فیزیک و شیمی و سایر حوزه‌ها کاربرد بسیاری دارد. اغلب پژوهش‌گران و تحلیل‌گران که با مدل‌های رگرسیونی سروکار دارند، از دو یا سه روش عمومی مانند رگرسیون خطی، رگرسیون چندکی و رگرسیون لجستیک استفاده می‌کنند، در حالیکه تکنیک‌های رگرسیونی بسیار گسترده می‌باشند و انواع مختلفی از روش‌های رگرسیون به منظور تحلیل داده‌های کیفی و کمی طراحی شده است. هر تحلیلگر باید بداند که بسته به نوع داده و توزیع از کدام شکل رگرسیون استفاده کند. در این پژوهش ۱۶ نوع رگرسیون معرفی و بررسی شده است و به ارائه راهکارهایی به منظور انتخاب صحیح مدل رگرسیون و جلوگیری از مشکلات رایج در این حوزه پرداخته‌ایم.

واژه‌های کلیدی: مدل‌سازی، روش‌های پیش‌بینی، روش‌های رگرسیونی، رگرسیون خطی، ضریب همبستگی

۱. مقدمه

در سال ۱۸۷۷ فرانسیس گالتون در مقاله‌ای که درباره بازگشت به میانگین منتشر کرده بود، اظهار داشت که متوسط قد پسران دارای پدران قدبلند، کمتر از قد پدرانشان می‌باشد. به این ترتیب گالتون پدیده بازگشت به طرف میانگین را در داده‌هایش مورد تأکید قرار داد. برای گالتون رگرسیون مفهومی زیست‌شناختی داشت، اما کارهای او توسط کارل پیرسون برای مفاهیم آماری توسعه داده شد. گرچه گالتون برای تأکید بر پدیده «بازگشت به سمت مقدار متوسط» از تحلیل رگرسیون استفاده کرد، اما به‌رحال امروزه واژه تحلیل رگرسیون جهت اشاره به مطالعات مربوط به روابط بین متغیرها به کار برده می‌شود. به عبارت دیگر، تعبیر جدید رگرسیون کاملاً متفاوت از حالت قبل است و به طور کلی تحلیل‌های رگرسیون به مطالعه وابستگی یک متغیر (متغیر وابسته) به یک یا چند متغیر دیگر (متغیر مستقل) می‌پردازد که با تخمین یا پیش‌بینی مقدار متوسط یا میانگین مقادیر متغیر نوع اول در حالتی که مقادیر متغیر نوع دوم معلوم یا معین باشند صورت می‌پذیرد.

تکنیک‌های رگرسیون یکی از محبوب‌ترین تکنیک‌های آماری است که برای مدل‌سازی، پیش‌بینی و داده‌کاوی مورد استفاده قرار می‌گیرد و در علوم مختلف مانند علوم انسانی، علوم زیستی، فیزیک و شیمی و سایر حوزه‌ها کاربرد بسیاری دارد (عریضی، گل‌پرور، ۱۳۸۸). به‌صورتی که بخش قابل ملاحظه‌ای از تحقیقات تجربی با استفاده از مدل‌های رگرسیونی صورت می‌گیرد (بدری، ۱۳۹۳).

تحلیل رگرسیونی به عنوان روش کاربردی برای تحلیل داده‌ها، در بسیاری از پژوهش‌های حسابداری مورد استفاده قرار گرفته است (مومنی، فعال قیومی، ۱۳۸۸). شناخت دقیق انواع تحلیل‌های رگرسیونی و استفاده مناسب و به‌جا از آن‌ها می‌تواند نتایج مفیدتر و اثربخش‌تری را به همراه داشته باشد (مومنی، فعال قیومی، ۱۳۸۸). اغلب پژوهشگران و تحلیل‌گران که با مدل‌های رگرسیونی سروکار دارند، از دو یا سه روش عمومی رگرسیون مانند رگرسیون خطی (بت‌شکن، فرهادی، ۱۳۹۵؛ امام وردی و دیگران، ۱۳۹۸؛ صمدی، ۱۳۸۶؛ مومنی، فعال قیومی، ۱۳۸۸) رگرسیون چندکی (خانیان و دیگران، ۱۳۹۶؛ فروشانی و دیگران، ۱۴۰۰؛ کمیجانی و دیگران، ۱۳۹۵؛ بت‌شکن، فرهادی، ۱۳۹۵؛ دانش جعفری و دیگران، ۱۳۹۵؛ بارانی، علیزاده، ۱۳۹۷) و رگرسیون لجستیک (اعتمادی، زلّی، ۱۳۹۲؛ دانشور مفرد و دیگران، ۱۳۹۳؛ اکبرلو و دیگران، ۱۳۹۹؛ نوظهور، کاویانی، ۱۳۹۹) استفاده می‌کنند، در حالیکه تکنیک‌های رگرسیونی بسیار گسترده می‌باشند و انواع مختلفی از روش‌های رگرسیونی به‌منظور تحلیل داده‌های کیفی و کمی طراحی شده است. هر تحلیلگر باید بداند که با توجه به نوع و توزیع داده از کدام شکل رگرسیون استفاده کند.

ادامه این مقاله به صورت زیر ساختار بندی شده است. در بخش ۲ به معرفی برخی از روش‌های رگرسیونی پرداخته و به‌صورت خلاصه هریک از این روش‌ها را توضیح داده‌ایم. در بخش ۳ مروری بر مطالعات رگرسیونی در حوزه حسابداری و مالی انجام گرفته است. در بخش ۴ نیز به بیان نکات، توصیه و پیشنهادهایی به‌منظور انتخاب صحیح مدل رگرسیون و استخراج نتایج قابل‌اتکا و اجتناب از اشتباهات رایج در این حوزه پیشنهاد شده است. نهایتاً در بخش آخر به بیان نتیجه‌گیری پرداخته‌ایم.

۲. معرفی برخی از روش‌های رگرسیونی

اغلب کسانی که با مدل‌های رگرسیونی سروکار دارند، از دو یا سه شیوه عمومی استفاده می‌کنند. در حالیکه تکنیک‌های رگرسیونی بسیار گسترده‌تر هستند و به‌منظور تحلیل‌های مختلف روی داده‌های کیفی و کمی به شکل‌های متفاوتی به کار می‌روند. در جدول ۱ تعداد ۱۶ روش از تکنیک‌های رگرسیونی فهرست شده است.

Dependent Variable

Independent Variable

جدول ۱- برخی از روش های رگرسیون

نام فارسی	نام انگلیسی	ردیف
رگرسیون خطی	Linear Regression	۱
رگرسیون لجستیک	Logistic Regression	۲
رگرسیون چندکی	Quantile Regression	۳
رگرسیون ستیغی	Ridge Regression	۴
رگرسیون لاسو	Lasso Regression	۵
رگرسیون شبکه الاستیک	Elastic Net Regression	۶
رگرسیون مؤلفه های اصلی	Principle Components Regression (PCR)	۷
رگرسیون کمترین مربعات جزئی	Partial Least Squares (PLS)	۸
رگرسیون بردار پشتیبان	Support Vector Regression (SVR)	۹
رگرسیون ترتیبی	Ordinal Regression	۱۰
رگرسیون پواسون	Poisson Regression	۱۱
رگرسیون شبه پواسون	Quasi Poisson Regression	۱۲
رگرسیون کاکس	Cox Regression	۱۳
رگرسیون توبیت	Tobit Regression	۱۴
رگرسیون دو جمله ای منفی	Negative Binomial Regression	۱۵
رگرسیون چند جمله ای	Polynomial Regression	۱۶

هر یک از روش های رگرسیونی، پیش فرض های مخصوص خود را دارد که بر حسب ویژگی و مشخصات متغیرهای مستقل و متغیر وابسته تعیین می شوند. حال سوالی که مطرح می شود این است که چگونه مدل رگرسیون صحیح را انتخاب نماییم؟

قبل از اجرای مدل رگرسیونی باید به وسیله رسم نمودار یا محاسبه ضریب همبستگی و نظایر آن، وجود وابستگی بین متغیرها توصیفی با متغیر وابسته را مورد بررسی قرار دهیم و در صورتی که نتایج حاصل از این ابزارها، نشانگر وجود رابطه (رابطه خطی یا رابطه غیرخطی) بین این دو دسته متغیر باشد، به دنبال ایجاد مدل رگرسیونی باشیم. پس از محاسبات و برآورد پارامترهای مدل باید فرضیه هایی که مدل رگرسیونی براساس آن استوار شده است نیز آزمون شود. بیشتر این گونه آزمون ها براساس باقی مانده های مدل صورت می گیرد در نتیجه باید آن ها را پس از پیدا کردن مدل و محاسبه باقی مانده ها مورد سنجش قرار داد. این آزمون ها در فهرست زیر معرفی و توضیح داده شده اند.

بررسی نقاط پرت : از آنجایی که وجود نقاط دور افتاده یا پرت، باعث می شود که برآورد پارامترهای مدل رگرسیونی به درستی صورت نگیرد، قبل از اجرای محاسبات مربوط به مدل رگرسیونی باید از عدم چنین داده های اطمینان حاصل پیدا کنیم و اگر به این گونه مشاهدات برخوردیم، آن ها را از مدل خارج کرده، سپس دوباره مدل سازی را انجام دهیم. نقاط پرت در اینجا به مشاهداتی اشاره دارد که باقی مانده های خیلی بزرگی نسبت به بقیه نقاط دارند.

Outlier

هم خطی و هم خطی چندگانه: اگر متغیرهای توصیفی نسبت به هم وابستگی داشته باشند، پارامترهای مدل، واریانس بزرگی پیدا کرده، در نتیجه قابل اعتماد نیستند. در این حالت باید مدل رگرسیونی را از جهت وجود مشکل هم خطی یا هم خطی چندگانه مورد بررسی قرار داد.

ناهم واریانسی: باقی مانده های حاصل از مدل رگرسیونی باید دارای واریانس ثابت باشند. معمولاً این موضوع به متغیر وابسته نیز بر می گردد. اگر با تغییر مقادیر متغیرهای توصیفی، واریانس متغیر وابسته تغییر کند، با مشکل ناهم واریانس مواجه هستیم. وزن دهی به مشاهدات یا تبدیل روی متغیر وابسته، روش هایی برای ثابت کردن واریانس متغیر پاسخ محسوب می شوند.

بیش برآزش و کم برآزش: به کارگیری متغیرهای توصیفی زیاد در کاهش میزان خطای مدل رگرسیونی مؤثر است ولی در این صورت ممکن است مدل دچار مشکل بیش برآزش شود. به این معنی که مدل قادر به پیش بینی دقیق مقادیر متغیر پاسخ برای مشاهداتی است براساس آن مدل ساخته شده ولی برای داده های جدید، توانایی پیش بینی مناسب را ندارد. عکس این حالت که کم برآزش نامیده می شود، ناکافی بودن متغیرهای توصیفی در مدل است، به طوری که واریانس مقادیر پیش بینی بسیار زیاد خواهد بود. به این ترتیب باید از مدل رگرسیونی استفاده کرد که تعداد متغیرهای آن مناسب بوده و چنین مسائلی را به وجود نیاورد.

۲-۱ رگرسیون خطی

یکی از ساده ترین تکنیک های رگرسیون، رگرسیون خطی است که در آن متغیر وابسته مقادیر عددی و پیوسته دارند. در این حالت رابطه بین متغیر وابسته و مستقل، یک رابطه خطی برحسب پارامترهای مدل است (مومنی، فعال قیومی، ۱۳۸۸). زمانی که فقط یک متغیر مستقل وجود داشته باشد، مدل رگرسیونی خطی را ساده می نامند و اگر بیش از یک متغیر مستقل وجود داشته باشد، رگرسیون را چندگانه می گویند. فرم مدل رگرسیون خطی ساده به صورت رابطه (۱) است.

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

رابطه (۱) معادله یک خط است که جمله خطا یا همان ε به آن اضافه شده است. از آنجایی که پیش بینی رابطه بین متغیر وابسته و مستقل به شکل دقیق نیست، جمله خطا را یک «متغیر تصادفی» با میانگین صفر در نظر می گیرند. پارامترهای مدل رگرسیون خطی رابطه (۱) عرض از مبدا (β_0) و شیب خط (β_1) است. باید توجه داشت که منظور از رابطه خطی در مدل رگرسیون، وجود رابطه خطی بین ضرایب است نه بین متغیرهای مستقل. برای مثال مدل $Y = \beta_0 + \beta_1 X^2 + \varepsilon$ نیز یک مدل خطی می باشد.

در رگرسیون خطی سعی می شود، به کمک معادله خطی که توسط روش رگرسیون معرفی می شود، برآورد مقدار متغیر وابسته به ازای مقدارهای مختلف متغیر مستقل توسط خط رگرسیون به دست آید. به منظور برآورد پارامترهای مناسب برای مدل، کوشش می شود براساس داده های موجود، مدلی انتخاب می شود که کمترین خطا را داشته باشد. روش های مختلفی برای تعریف خطا و حداقل کردن خطا وجود دارد. معیاری که در مدل رگرسیون خطی به کار می رود، کمینه کردن مجموع مربعات

^۴Colinearity

^۵Multicolinearity

^۶Heteroscedasticity

^۷Overfitting

^۸Underfitting

خطا است. یکی از روش های مرسوم برای تخمین پارامترهای مدل رگرسیونی شماره (۱) روش حداقل مربعات معمولی است که به اختصار روش OLS نیز گفته می شود. استفاده از این روش دارای یک سری مفروضاتی می باشد که در ادامه معرفی شده است:

- میانگین خطاها صفر باشد.
- بین خطاها همبستگی وجود نداشته باشد.
- واریانس خطا ثابت باشد.
- بین متغیرهای مستقل همبستگی وجود نداشته باشد (دارای هم خطی نباشد).
- توزیع خطاها باید دارای توزیع نرمال باشد.

برآورد پارامترهای مدل رگرسیون خطی ساده به صورت زیر خواهد بود.

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad (2)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3)$$

در تجزیه و تحلیل رگرسیون معیارهایی برای سنجش قابلیت اتکا و دقیق بودن رگرسیون وجود دارد. منظور از دقیق بودن اشاره به صحت برآورد در رگرسیون است، درحالی که منظور از قابلیت اتکا دلالت بر استمرار واقعی بودن ارتباطات موجود بین متغیرها در رگرسیون دارد. به بیان دیگر آیا از مدل رگرسیون برای برآورد در آینده همچنان می توان استفاده نمود. این معیارها در ارزیابی مفید بودن رگرسیون به کار می رود. ضریب تعیین، ارزش t- و ارزش p- برای اندازه گیری قابلیت اتکا رگرسیون مورد استفاده قرار می گیرد. درحالی که، خطای معیار برای اندازه گیری دقیق بودن، یا صحت رگرسیون است. در این بخش به معرفی این چهار معیار کلیدی می پردازیم:

ضریب تعیین : که به آن ضریب تشخیص نیز گفته می شود، معمولاً با R^2 نشان داده می شود، عددی است بین صفر و یک که اغلب به عنوان معیاری برای توصیف قدرت رگرسیون یعنی توان پیش بینی درجه تغییرات متغیر وابسته در اثر تغییرات متغیرهای مستقل مورد استفاده قرار می گیرد.

خطای معیار (استاندارد) : معیاری برای اندازه گیری دامنه پراکندگی مشاهدات واقعی^۱ در اطراف خط رگرسیون، و تبع آن، معیاری برای پی بردن به میزان صحت برآورد از طریق رگرسیون است. از این معیار می توان برای تعیین یک دامنه در اطراف برآورد رگرسیون که در آن ارزش واقعی و ناشناخته برآورد می تواند به طور معقول مورد انتظار باشد، استفاده می شود.

^۱Sum of Square Error

^۲Ordinary Least Squares

^۳Reliability

^۴Precision

^۵Coefficient of Determination

^۶Standard Error

ارزش- t ^۱: معیار آماری قابلیت اتکا برای متغیر مستقل است. قابلیت اتکا درجه ارتباطی ارزشمند و با ثبات در طولانی مدت بین متغیر مستقل با یک متغیر وابسته را نشان می‌دهد. از این رو یک ارزش- t کوچک، علامت نبود یا کمبود ارتباط آماری بین متغیرهای مستقل و وابسته است (به طور کلی ارزش- t باید همواره بزرگتر از ۲ باشد). به این دلیل متغیری با ارزش- t پایین را به خاطر ساده‌سازی مدل و این که منجر به برآورد صحیح‌تر شود باید از رگرسیون حذف کرد.

ارزش- p ^۱: معیار اندازه‌گیری ریسک برای متغیر مستقلی است که از احتمال کم داشتن ارتباط با یک متغیر وابسته برخوردار بوده و هیچ ارتباط آماری قابل توجهی بین آن دو مشاهده نمی‌شود. بر این اساس اندازه کوچک ارزش- p (ریسک کم) مناسب است. اغلب، در عمل از ارزش- p معادل ۰,۵ یا کمتر به عنوان راهنما استفاده می‌شود.

۲-۲ رگرسیون لجستیک

یکی از روش‌های دسته‌بندی در مبحث آموزش نظارت شده رگرسیون لجستیک است. تحلیل رگرسیونی لجستیک بسیار متداول بوده که در آن احتمال یک پیامد دو مقوله‌ای (دو ارزشی) به تعدادی متغیر پیش بینی کننده بالقوه ارتباط داده می‌شود (اعتمادی، زلفی، ۱۳۹۲).

از رگرسیون لجستیک برای تحلیل رابطه بین متغیرها بخصوص در زمینه‌های پزشکی، روانشناسی و علوم اجتماعی بسیار کمک گرفته می‌شود. در رگرسیون لجستیک، متغیر وابسته، به صورت دودویی است. به این معنی که مقادیر آن به دو طبقه صفر و یک دسته‌بندی شده‌اند. البته زمانی که از رگرسیون چند جمله‌ای لجستیک استفاده می‌کنید، ممکن است تعداد سطوح متغیر طبقه‌ای بیشتر از دو باشد. واضح است که در این مدل رگرسیونی، خطاها، دارای توزیع نرمال نیستند و متغیر وابسته دارای توزیع دو یا چند جمله‌ای است در نتیجه نمی‌توان از مدل رگرسیون ساده یا خطی استفاده کرد.

۲-۳ رگرسیون چندکی

در مواقعی که توزیع خطا غیر نرمال می‌باشد و یا مدل رگرسیون نسبت به داده‌های پرت در حالت واریانس ناهمسانی حساس می‌باشد، رگرسیون چندکی جهت عبور از محدودیت‌های اشاره‌شده برای رگرسیون خطی معرفی شد. این روش به تدریج به رویکردی جامع برای تجزیه و تحلیل آماری ارتباط میان دو یا چند متغیر در مدل‌های خطی و غیرخطی مبدل شد. به کارگیری رگرسیون چندکی امکان استفاده از شاخص‌های پراکندگی و مرکزی مختلف به‌ویژه در دنباله‌های ابتدایی و انتهایی برای تحلیل جامع‌تر از روابط میان متغیرها را امکان‌پذیر می‌کند و حتی توانایی استخراج روابط ضعیف میان متغیرها در قسمت‌های مختلف توزیع را دارد، این در حالی است که رگرسیون خطی امکان دخالت متغیرهای مستقل را تنها در مرکز ثقل داده‌ها (کوانتایل ۰,۵) فراهم می‌کند (کميجانی و دیگران، ۱۳۹۵).

رگرسیون چندکی را می‌توان جایگزینی برای مدل رگرسیون خطی در نظر گرفت که نسبت به مشکلات حاصل از نقاط پرت، چولگی زیاد و همچنین ناهم‌واریانس مقاوم است. همان‌طور که می‌دانید میانگین باوجود داده‌های پرت، چولگی و ناهم‌واریانس در متغیر وابسته، شاخص مناسبی به‌عنوان معیار تمرکز نیست. بنابراین شیوه رگرسیون خطی، نمی‌تواند تغییرات چنین متغیری را به‌خوبی توصیف و مدل مناسبی برای نمایش رابطه بین متغیر مستقل و وابسته ایجاد کند. استفاده از چندک‌ها که نسبت به شرایط گفته شده، مقاوم‌تر هستند، می‌تواند مدل رگرسیون کامل‌تر و دقیق‌تری را ارائه کند. در

^۱t-value

^۱p-value

^۱Binary

رگرسیون چندکی به جای برآورد میانگین متغیر وابسته، از برآورد چندک‌های آن (مانند صدک، دهک یا چارک) به شرط متغیر مستقل، کمک گرفته می‌شود. در رگرسیون چندکی، باید متغیر وابسته از نوع عددی و از نوع مقیاس با مقادیر پیوسته باشد تا امکان محاسبه چندک‌ها وجود داشته باشد.

۲-۴ رگرسیون ستیغی

در مباحث مربوط به رگرسیون چندگانه، تعیین تعداد متغیرهای مستقلی که باید در مدل به کار گرفته شوند، یک مشکل محسوب می‌شود. با افزایش تعداد متغیرها، بیش برآزش رخ داده و با کاهش آن‌ها نیز ممکن است با مسئله کم برآزش مواجه شویم. در صورتی که مدل رگرسیونی دچار بیش برآزش شود، خطای آن برای برآورد مقادیر جدید متغیر وابسته زیاد خواهد بود در حالیکه وجود متغیرهای کمتر از حد لازم در مدل (کم برآزش) واریانس مدل را افزایش می‌دهد. بنابراین با افزایش تعداد متغیرها مشکل هم خطی و بیش برآزش ظاهر شده و با کاهش آن‌ها، واریانس مدل افزایش خواهد یافت. یکی از روش‌های غلبه بر این مسائل در رگرسیون چندگانه، استفاده از مدل رگرسیون ستیغی است. از آنجایی که در زمانی که متغیرهای مدل، زیاد و یا هم خطی چندگانه وجود داشته باشد، واریانس برآوردگرها متورم شده و به شکل قله (ستیغ) در می‌آید، از همین روی، این روش رگرسیونی که بر این مشکل غلبه می‌کند، رگرسیون ستیغی نام‌گذاری شده است.

در مسئله رگرسیون خطی، از یک تابع خطا استفاده شده و سعی بر آن است که مجموع مربعات خطا را کمینه کنند. در رگرسیون ستیغی به کمک ترکیب تابع مجموع مربعات خطا و مقدار جریمه مرتبط با تعداد پارامترها، تابع جدیدی ساخته می‌شود که برای برآورد پارامترهای مدل رگرسیونی به کار می‌رود. حالت کلی‌تر این روش را به نام قاعده‌سازی تیکانوف می‌شناسند.

۲-۵ رگرسیون لاسو

با استفاده از رگرسیون لاسو سعی می‌کنیم، روش مناسب برای مدل‌سازی متغیر پاسخ براساس کمترین و البته مناسب‌ترین تعداد متغیرهای مستقل را ارائه دهیم. در مدل رگرسیون خطی فرض بر این است که π مشاهده از نتایج (متغیر وابسته) و p متغیر مستقل (یا ویژگی) در اختیار ما قرار گرفته است. پیش‌بینی متغیر پاسخ براساس داده‌های مشاهده شده، هدف اصلی در رگرسیون خطی است. همچنین عمل پیش‌بینی ممکن است برای داده‌هایی که در هنگام مدل‌سازی حضور نداشته‌اند (داده‌های آزمون) به کار گرفته شود. در بیشتر موارد، برآورد برای پارامترهای این مدل صفر نخواهد بود و هیچ تغییری بدون اثر در مدل به دست نمی‌آید. پس همه متغیرها در مدل نقش داشته و برای پیش‌بینی مقادیرهای متغیر پاسخ به کار گرفته می‌شوند. در رگرسیون لاسو (عملگر گزینش و انقباض کمترین قدر مطلق) که توسط رابرت تیب‌شیرانی (۱۹۹۶) معرفی شد، روش اول^۱ یعنی کاهش بعد متغیرها به کاررفته و از کمینه‌سازی مجموع مربعات تغییر یافته استفاده می‌شود. به این ترتیب با استفاده از یک تابع جریمه روی جمع قدر مطلق ضرایب مدل رگرسیونی، تعداد پارامترها کنترل می‌شود. در این حالت، مجموع مربعات خطای رگرسیونی لاسو به صورت رابطه (۴) نوشته می‌شود:

^۱Penalty

^۱Tikhonov Regularization

^۱Least Absolute Shrinkage and Selection Operator

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_j |\beta_j| \quad (4)$$

در این رابطه λ پارامتر تنظیم کننده است، به این معنی که اگر مقدارش برابر با صفر باشد، مدل به رگرسیون عادی تبدیل شده و همه متغیرها در آن حضور خواهند داشت و اگر مقدار آن افزایش یابد تعداد متغیرهای مستقل در مدل کاهش خواهند یافت. بنابراین با انتخاب بی نهایت برای λ عملاً هیچی متغیری در مدل وجود ندارد. تعیین مقدار برای این پارامتر معمولاً توسط روش اعتبارسنجی متقابل انجام می شود.

۲-۶ رگرسیون شبکه الاستیک

رگرسیون شبکه الاستیک، با ترکیب رگرسیون لاسو و رگرسیون ستیجی، بر معایب آن‌ها غلبه کرده و جایگزین مطمئن برای آن‌ها است. به این ترتیب اگر با مدلی مواجه هستید که متغیرهای توصیفی آن با یکدیگر همبستگی دارند، بهتر است از رگرسیون شبکه الاستیک استفاده کنید.

۲-۷ رگرسیون مؤلفه‌های اصلی

زمانی که هم خطی یا هم خطی چندگانه در مدل رگرسیونی وجود داشته باشد، بهتر است از مدل رگرسیون مؤلفه‌های اصلی که به اختصار با PCR نشان داده می شود، استفاده کنیم. رگرسیون مؤلفه‌های اصلی در دو گام اجرا می شود.

✓ گام اول: استخراج مؤلفه‌های اصلی براساس متغیرهای توصیفی

✓ گام دوم: اجرای رگرسیون براساس مؤلفه‌های ایجاد شده به عنوان متغیرهای مستقل با متغیر وابسته

به این ترتیب، مشکل هم خطی یا هم خطی چندگانه از مدل رگرسیونی خارج شده و از طرفی با توجه به استفاده از مؤلفه‌های کمتر از تعداد متغیرهای توصیفی، ابعاد یا تعداد متغیرهای به کاررفته در مدل رگرسیونی نیز کاهش می یابد.

۲-۸ رگرسیون کمترین مربعات جزئی

زمانی که بین متغیرهای توصیفی، وابستگی شدید وجود داشته باشد، به جای رگرسیون مؤلفه‌های اصلی بهتر است از رگرسیون کمترین مربعات جزئی استفاده شود. همچنین زمانی که تعداد متغیرهای توصیفی زیاد هستند و می خواهیم مؤثرترین متغیرها در مدل حضور داشته باشند، از رگرسیون کمترین مربعات جزئی استفاده می کنیم.

هم در روش رگرسیون مؤلفه‌های اصلی و هم کمترین مربعات جزئی، متغیر جدیدی به عنوان متغیر پیش گو ساخته می شود که به آن مؤلفه گفته می شود. این متغیر جدید، ترکیب خطی از متغیرهای توصیفی است. ولی تفاوت در این است که در تحلیل رگرسیون PCR، مؤلفه‌ها براساس توصیف واریانس کل متغیرهای توصیفی تولید می شوند بدون آنکه به مقادیر متغیر پاسخ توجه شود. در حالیکه در PLS با در نظر گرفتن متغیر پاسخ و متغیرهای پیشگو، مؤلفه‌ها تولید می شوند و در نهایت مدلی ایجاد می شود که با کمترین عوامل، بهترین برازش را دارد.

^۲Cross Validation

^۲Component

۹-۲ رگرسیون بردار پشتیبان

به کمک رگرسیون بردار پشتیبان که گاهی آن را با SVR نیز نشان می‌دهند، می‌توان مدل‌های خطی و غیرخطی را ایجاد و پارامترهای آن را محاسبه کرد. این کار توسط به‌کارگیری یک تابع هسته غیرخطی (مانند چندجمله‌ای) حاصل می‌شود. محاسبه پارامترهای این تابع به شکل است که خطا کمینه شود به طوری که فاصله بین صفحاتی که عمل جداسازی بین دسته‌ها را ایجاد می‌کنند، بیشینه شود.

۱۰-۲ رگرسیون ترتیبی

اگر متغیر پاسخ یا وابسته به صورت ترتیبی یا رتبه‌ای باشد، بهتر است^۴ از مدل رگرسیون ترتیبی استفاده شود. به این ترتیب اگر با استفاده از طیف لیکرت، با پاسخ‌های مربوط به پرسشنامه مواجه هستید، بهتر است برای مدل‌سازی بین این متغیر وابسته با متغیرهای مستقل، از رگرسیون ترتیبی استفاده کنید. برای مثال ممکن است براساس یک پرسش‌نامه که پاسخ‌های آن با مقیاس یا طیف لیکرت ۵ سطحی تشکیل شده بخواهید میزان رضایت ذی‌نفعان از عملکرد شرکت را بسنجید. چنین مدلی به کمک رگرسیون ترتیبی ساخته خواهد شد.

۱۱-۲ رگرسیون پواسون

زمانی که متغیر وابسته به صورت شمارشی باشد، از مدل رگرسیون پواسون استفاده می‌کنیم. متغیر وابسته در این مدل باید شرایط زیر را داشته باشد:

- ✓ متغیر وابسته (Y) دارای توزیع پواسون است.
- ✓ مقدار متغیر وابسته نباید منفی باشد (زیرا حاصل از شمارش است).
- ✓ مدل رگرسیون پواسون برای زمانی که مقادیر متغیر وابسته متعلق به مجموعه اعداد طبیعی نیستند، نباید استفاده شود.

از رگرسیون پواسون برای داده‌هایی که از نوع نرخ (درصد) هستند نیز استفاده می‌شود. در اینجا منظور از نرخ، میزان یا تعداد رخداد یک پیشامد در واحد است. معمولاً این واحد از مشاهدات را با Exposure نشان می‌دهند.

۱۲-۲ رگرسیون شبه پواسون

یکی از روش‌های رگرسیونی که می‌تواند به‌عنوان جایگزین رگرسیون دوجمله‌ای منفی به کار رود، رگرسیون شبه پواسون است. هرچند هر دو شیوه پاسخ‌های یکسانی تولید می‌کنند ولی در ماتریس کوواریانس برآوردگرها اختلافاتی وجود دارد. به طوری که واریانس مدل شبه پواسون، یک ترکیب خطی از میانگین بوده ولی واریانس مدل رگرسیون دوجمله‌ای منفی، به صورت تابعی درجه چهار از میانگین است.

^۴Kernel

^۵Ordinal

^۶Rank

^۷Likert Scale

^۸Rate

۲-۱۳ رگرسیون کاکس

از رگرسیون کاکس برای داده‌های وابسته به زمان استفاده می‌شود. متغیر وابسته به زمان، متغیری است که مقدار آن برای یک آزمودنی معین می‌تواند در طول زمان متفاوت باشد. در مقابل، متغیر مستقل از زمان متغیری است که مقدارش برای یک آزمودنی معین در طول زمان ثابت باقی می‌ماند (دانشور مفرد و دیگران، ۱۳۹۳).

مدل رگرسیون کاکس مانند مدل رگرسیون لجستیک است با این تفاوت که در رگرسیون لجستیک تعداد پیشامدها مهم است ولی در رگرسیون کاکس، طول یا بازه زمانی بین پیشامدها مورد بررسی قرار می‌گیرد.

چنین تحلیل‌هایی را به‌عنوان تحلیل بقا شناخته می‌شوند. همان‌طور که مشخص است متغیر وابسته از دو بخش تشکیل شده. در بخش اول که مقداری پیوسته است، زمان یا طول دوره رخداد یک پیشامد ثبت شده و در بخش دوم، رخداد یا عدم رخداد آن پیشامد مشخص شده است.

۲-۱۴ رگرسیون توبیت

رگرسیون توبیت که گاهی رگرسیون حساس شده و یا سانسور شده نامیده می‌شود، به منظور بررسی روابط خطی در شرایطی که در مورد متغیر وابسته یک حد بحرانی در سمت راست یا چپ مشاهده شود استفاده می‌شود. از این رو به این نوع خاص از توابع پیش بینی رگرسیون سانسور شده نیز اطلاق می‌شود. وجود موارد فراتر از حد بحرانی یا پائین‌تر از حد بحرانی در متغیر وابسته بیانگر یک مشکل جدی و اریب در معادله رگرسیون است و نیازمند استفاده از رگرسیون توبیت است. به عبارت دیگر وجود این سطح از پراکنش سبب خطای جدی در شرایط استفاده از رگرسیون خطی می‌شود.

۲-۱۵ رگرسیون دو جمله‌ای منفی

در رگرسیون دو جمله‌ای منفی، متغیر وابسته یک تعداد مشاهده شده که از توزیع دو جمله‌ای منفی پیروی می‌کند. بنابراین، مقادیر ممکن متغیر وابسته اعداد صحیح غیر منفی مانند ۰، ۱، ۲، ۳، و ... هستند. رگرسیون دو جمله‌ای منفی شکل تعمیم یافته رگرسیون پواسون است که فرض برابری واریانس با میانگین در مدل پواسون است، از بین می‌برد.

۲-۱۶ رگرسیون چند جمله‌ای

گاهی می‌توان یک مدل غیرخطی بین متغیرهای مستقل و وابسته را به صورت یک چند جمله‌ای درجه k تعیین کرد. مدل رگرسیونی چند جمله‌ای در چنین مواقعی می‌تواند رابطه بین متغیرهای مستقل و وابسته را به خوبی بیان کند.

در فضای اقلیدسی فقط یک خط راست وجود دارد که از دو نقطه می‌گذرد. همچنین برای مشخص کردن یک منحنی درجه دو، کافی است سه نقطه از آن معلوم باشد تا معادله مربوط به آن را به همراه پارامترهای معادله مشخص و شکل آن را ترسیم کنیم. هر چه تعداد نقاط بیشتر باشد، درجه منحنی یا چند جمله‌ای که از آن نقاط بگذرد، بیشتر خواهد شد. این طور به نظر می‌رسد که همیشه درجه منحنی یا چند جمله‌ای، یک واحد کمتر از تعداد نقاط است. در نتیجه اگر بخواهیم براساس ۱۰ نقطه یک منحنی برازش کنیم، به یک چندجمله‌ای نیازمندیم که درجه آن ۹ باشد. پیچیدگی در این مدل زیاد است زیرا چند جمله‌ای درجه ۹ دارای ۱۰ پارامتر است. یک رابطه یا تابع چند جمله‌ای مرتبه k به صورت زیر نوشته می‌شود.

^۲Survival Analysis

^۲Censored

$$f(x) = a. + a_1x + a_2x^2 + \dots + a_kx^k \quad (5)$$

رگرسیون چندجمله ای سعی دارد با استفاده از تعداد متغیرهای بسیار کمتر از تعداد نقطه‌ها، پارامترهای چند جمله‌ای یا رگرسیون را برآورد کند. در نتیجه پیچیدگی مدل کاهش یافته ولی در عوض خطا نیز بوجود خواهد آمد. به این ترتیب یک توازن بین بیش‌برازش و کم‌برازش در مدل برقرار خواهد شد.

به این ترتیب می‌توان مدل رگرسیون خطی زیر را برای رگرسیون چند جمله ای در نظر گرفت.

$$Y = \beta. + \beta_1X + \beta_2X^2 + \dots + \beta_kX^k + \varepsilon \quad (6)$$

بر همین اساس می‌توان رگرسیون چندجمله‌ای را به صورت رگرسیون خطی چندگانه و به شکل زیر درج کرد.

$$X_1 = X, X_2 = X^2, \dots, X_k = X^k \quad (7)$$

۳. مروری بر مطالعات رگرسیونی در حوزه حسابداری و مالی

ایده رگرسیون جعلی در اقتصادسنجی توسط گرنجر و نیوبولد (۱۹۷۴) مطرح شد. محققان نشان دادند در رگرسیون خطی ساده و تخمین پارامترهای آن به روش حداقل مربعات معمولی باعث پیدایش نتایج غیرواقعی (یا جعلی) خواهد شد. در پژوهش صمدی (۱۳۸۶) به بررسی سیر تکامل تاریخی و مفهوم رگرسیون جعلی پرداخته شده است.

مومنی و فعال قیومی (۱۳۸۸) مدل‌های رگرسیون را با توجه به داده‌های مورد بررسی، به انواع مختلفی تقسیم‌بندی کرده و بیان داشته‌اند: «این مدل‌های رگرسیون عبارت است از: رگرسیون سری زمانی، رگرسیون مقطعی، رگرسیون میانگین و رگرسیون تجمعی. مقایسه تحلیل‌های رگرسیونی با استفاده از داده‌های حسابداری محدود و به کمک ضریب تعیین، آزمون F و آزمون t انجام شده است. نتایج پژوهش نشان داده است که رگرسیون سری زمانی برای مطالعه‌های کلی و فراگیر مناسب است. البته رگرسیون تجمعی با توجه به فراهم آوردن امکان دستیابی به نتایج کلی‌تر و به دور از سوگیری، در پژوهش‌های حسابداری بیشتر مورد استفاده قرار گرفته است. رگرسیون میانگین نیز به دلیل تمرکز بر داده‌های مرکزی، در پژوهش‌های حسابداری کاربرد اندکی داشته است.» در این پژوهش محققان با توجه به داده‌های مورد بررسی و تنها با استفاده از روش رگرسیون خطی، به مقایسه و تحلیل نتایج پرداخته است و از سایر روش‌های رگرسیونی استفاده و مقایسه نشده است.

حسن زاده و کاظمی (۱۳۹۰) در پژوهشی با عنوان «مقایسه انواع مدل‌های رگرسیون در تحلیل داده‌های شمارشی» به بررسی این گونه از مدل‌های رگرسیون پرداختند. محققان بیان داشته‌اند: «مدل متداول در برازش داده‌های شمارشی پواسون است که ویژگی مهم آن برابری میانگین و واریانس می‌باشد. اما در مباحث کاربردی در صورت عدم برقراری این شرط استفاده از مدل پواسون منجر به استنباط نادرستی از پارامترهای مدل خواهد شد.» در این پژوهش محققان ضمن بررسی موضوع بیش پراکنش و آزمون‌های مربوط به آن، مدل‌های دو جمله‌ای منفی و پواسون تعمیم یافته به عنوان جایگزین‌هایی برای مدل

^۱Spurious Regression

^۲Time Series

^۳Cross Section

^۴Pooled

پواسون معرفی کرده‌اند. البته این پژوهش در حوزه علوم پزشکی انجام گرفته است و تنها به بررسی روش‌های رگرسیونی برای داده‌های شمارشی پرداخته است.

اعتمادی و زلقی (۱۳۹۲) برای تدوین مدل تحقیق، از تحلیل رگرسیون لجستیک استفاده کرده است و بیان داشته است: «تحلیل رگرسیون لجستیک بسیار متداول بوده که در آن احتمال یک پیامد دو مقوله‌ای (دو ارزشی) به تعدادی متغیر پیش‌بینی کننده بالقوه ارتباط داده می‌شود. مدل لجستیک موردنظر در این تحقیق با استفاده از نسبت‌های مالی برآورد شده است.»

دانشور مفرد و دیگران (۱۳۹۳) به بررسی عوامل موثر بر تغییرات عمده قیمت سهام، اثر عامل بازار، اندازه، ارزش، نقدشوندگی و تکانه بر افزایش‌های عمده قیمت سهام پرداخته است. محققان به منظور برآورد مدل، از رگرسیون کاکس و برای داده‌های بازگشتی استفاده کرده‌اند و بیان داشته‌اند: استفاده از رگرسیون کاکس به این دلیل انجام می‌شود که روش‌های مرسوم استفاده شده در مطالعه عوامل موثر بر رویدادهای مالی مانند رگرسیون لجستیک، وابستگی بین رویدادهای مالی را در نظر نمی‌گیرند.

کاوایی و دیگران (۱۳۹۵) در رابطه با انتخاب مدل رگرسیون‌های غیرخطی پیشنهاد داده‌اند: «شکل مدل رگرسیونی در برخی از پژوهش‌ها از حالت ارتباط خطی به غیرخطی (درجه دوم و درجه سوم) تبدیل می‌گردد. به عبارتی اگر در نتایج یک پژوهش توسط پژوهش‌گران مختلف در جوامع متفاوت، ارتباط منفی و مثبت زیادی به دست آید بهتر است پژوهش‌گر در تدوین این‌گونه پژوهش‌ها از رابطه غیرخطی استفاده نماید.» در این پژوهش فقط اشاره کلی به رگرسیون‌های خطی و غیرخطی شده است در حالیکه روابط غیرخطی حوزه‌ای بسیار گسترده می‌باشد.

در پژوهش کمیجانی و دیگران (۱۳۹۵) به بررسی اثر واکنش به قیمت‌داری‌ها در سیاست پولی بر تثبیت اقتصاد کلان مبتنی بر رویکرد کوانتایل پرداخته‌اند و بیان داشته‌اند: «به‌طور کلی رگرسیون حداقل مربعات معمولی به‌منظور بررسی ارتباط خطی دو یا چند متغیر در موضوعات مختلف به کار می‌رود. درواقع این روش می‌تواند ارتباط بین میانگین شرطی وابسته برحسب یک یا چند متغیر توضیحی را تبیین می‌کند. علی‌رغم ویژگی‌های مناسب این مدل برای تحلیل رگرسیونی، به دنبال عملکرد ضعیف آن در زمان‌هایی که توزیع خطا غیر نرمال بوده (به‌ویژه در توزیع‌هایی با دنباله‌های نامتقارن) یا حساس بودن رگرسیون نسبت به داده‌های پرت و دورافتاده در حالت واریانس ناهمسانی برای نخستین بار رگرسیون کوانتایل را جهت گذر از محدودیت‌های اشاره‌شده برای رگرسیون معمولی، معرفی نمودند. این روش به تدریج به رویکردی جامع برای تجزیه و تحلیل آماری ارتباط میان دو یا چند متغیر در مدل‌های خطی و غیرخطی مبدل شد. به‌کارگیری رگرسیون کوانتایل امکان استفاده از شاخص‌های پراکندگی و مرکزی مختلف به‌ویژه در دنباله‌های ابتدایی و انتهایی برای تحلیل جامع‌تر از روابط میان متغیرها را امکان‌پذیر می‌کند و حتی توانایی استخراج روابط ضعیف میان متغیرها در قسمت‌های مختلف توزیع را دارد. این در حالی است که رگرسیون حداقل مربعات معمولی امکان دخالت متغیرهای توضیحی را تنها در مرکز ثقل داده‌ها (کوانتایل ۰,۵) فراهم می‌کند. از دیگر ویژگی‌های رگرسیون کوانتایل می‌توان به نشان دادن چگونگی تأثیر متغیرهای مستقل بر روی مکان و مقیاس و شکل توزیع نیز اشاره کرد. درمجموع، رگرسیون کوانتایل برخلاف رگرسیون حداقل مربعات معمولی از حداقل کردن مجذور خطاها و مجموع قدر مطلق باقی‌مانده‌های موزون برای برآورد پارامتر الگو استفاده می‌کند که به آن روش حداقل قدر مطلق انحرافات گفته می‌شود.»

در پژوهش خانیان و دیگران (۱۳۹۶) به بررسی اثرات نامتقارن تورم بر کسری بودجه در ایران با رویکرد رگرسیون کوانتایل پرداخته‌اند و بیان داشته‌اند: «انگیزه اصلی به‌کارگیری رگرسیون چندکی این است که با نگاهی دقیق و جامع در ارزیابی متغیر پاسخ، مدلی ارائه شود تا امکان دخالت متغیرهای مستقل نه‌تنها در مرکز ثقل داده‌ها، بلکه در تمام قسمت‌های توزیع به‌ویژه در دنباله‌های ابتدایی و انتهایی فراهم گردد، بدون این‌که با محدودیت‌های مفروضات رگرسیون معمولی، واریانس ناهمسانی و حضور تأثیرگذار داده‌های دورافتاده در برآورد ضرایب روبه‌رو باشد. در رگرسیون چندکی برخلاف رگرسیون معمولی از حداقل

نمودن مجموع قدر مطلق باقی مانده های موزون برای برآورد پارامتر الگو استفاده می شود که به آن روش حداقل قدر مطلق انحرافات گفته می شود.»

فروشانی و دیگران (۱۴۰۰) با استفاده از روش تسلط تصادفی تخمینی و به کمک رگرسیون چندکی، به ارزیابی کارایی بازار سرمایه پرداخته شده است. محققان بیان داشته اند: با توجه به اینکه روش رگرسیون خطی نمی تواند میزان اثربخشی متغیر وابسته از متغیر توضیحی را در سطوح مختلف توزیع نشان دهد و فقط با پارامتر میانگین مرتبط هستند مدل رگرسیون چندکی را جایگزین رگرسیون خطی و روش حداقل مربعات معمولی کرده است.

شناخت دقیق انواع تحلیل های رگرسیونی و استفاده مناسب و به جا از آن ها می تواند نتایج مفیدتر و اثربخش تری را به همراه داشته باشد (مومنی، فعال قیومی، ۱۳۸۸). در مطالعات و تحقیقات انجام شده پیشین، به مقایسه و معرفی مدل های رایج رگرسیون مانند رگرسیون خطی، رگرسیون چندکی و رگرسیون لجستیک پرداخته شده است و مقایسه جامعی میان سایر انواع مدل های رگرسیون تحقیقی صورت نگرفته است. در مقاله حاضر به شکل تکمیلی و در راستای این مطالعات، به معرفی ۱۶ مدل رگرسیون در حوزه مالی پرداخته شده است و همچنین به بیان راهکارهایی به منظور انتخاب صحیح مدل رگرسیون، استخراج نتایج قابل اتکا و جلوگیری از مشکلات رایج در این حوزه پرداخته ایم.

در جدول ۲ پژوهش های صورت گرفته با استفاده از روش های رگرسیون که در ۲ سال اخیر ۱۳۹۹ و ۱۴۰۰ در دو مجله «راهبرد مدیریت مالی» و «مهندسی مالی و مدیریت اوراق بهادار» منتشر شده، فهرست شده است. همان طور که از جدول ۲ مشاهده می شود، اکثریت پژوهش های صورت گرفته با استفاده از رگرسیون خطی صورت گرفته است.

جدول ۲- خلاصه تحقیقات انجام شده با استفاده از روش های رگرسیون

ردیف	نام پژوهش گر	عنوان مقاله	مدل رگرسیون مورد استفاده	نشریه	سال چاپ
۱	حسن زلفی، کفسان منصور، عباس افلاطونی	ویژگی های جمعیت شناختی مدیران ارشد مالی و استراتژی مدیریت سود	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۲	عبداله امیری، پرویز سعیدی	بررسی رابطه بحران مالی، چرخه عمر و استراتژی های تجدید ساختار دارایی و مالی شرکت های دارای بحران مالی	رگرسیون لاجیت	راهبرد مدیریت مالی	۱۴۰۰
۳	حمید بدآغی، حسن ولیان، سیده فاطمه مرتضویان، مهدی واسعی چارمحالی	بررسی تاثیر توانایی مدیریتی مدیر عامل در ابهام شرکتی و ارزش نهایی وجه نقد شرکت های پذیرفته شده در بورس اوراق بهادار تهران	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۴	آرش پوررضایی، علی محمد کیمیگری	تخمین تابع سودآوری با استفاده از متغیرهای ساختار سرمایه در صنایع ایران	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۵	محمد رضا رستمی، حجت اله انصاری، فرزانه رحیمی	ویژگی های شرکتی توضیح دهنده نوسانات غیرسیستماتیک در بورس اوراق بهادار تهران	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۶	محمد اسماعیل فدایی	بررسی چگونگی ارتباط مدیریت جریان وجوه	رگرسیون	راهبرد	۱۴۰۰

ردیف	نام پژوهش گر	عنوان مقاله	مدل رگرسیون مورد استفاده	نشریه	سال چاپ
	نژاد، محمد اصولیان، پروانه شمس	نقد با ساختار سرمایه	کوانتایل	مدیریت مالی	
۷	مهسا کفاش پور یزدی، اکرم تفتیان، محمود معین الدین	تأثیر تقارن اطلاعاتی بر رابطه کیفیت اطلاعات داخلی و ریسک سیستماتیک با استفاده از مدل های ایستا و پویا	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۸	روح اله عرب، سیده زهرا حسینی، محمد غلامرضا پور	بررسی رابطه بین قدرت مدیرعامل و اهرم مالی شرکت	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۹	سید کاظم ابراهیمی، علی بهرامی نسب، حسین سجادی نسب	بررسی اثر روابط سیاسی بر محافظه کاری سود و زیانی و هزینه بدهی شرکت های پذیرفته شده در بورس اوراق بهادار تهران	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۱۰	شکراله خواجهوی، علی غیوری مقدم	بررسی نقش توانایی مدیران و تضاد منافع در مفید بودن چسبندگی هزینه ها: شواهدی از تصمیمات آگاهانه مدیران در پدیده چسبندگی هزینه	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۱۱	معصومه صابرمهانی، محمد رضا نیکبخت، مصطفی دلدار	آزمون اثربخشی تمرکز مشتری بر سرعت تعدیل نگهداشت وجه نقد در بورس اوراق بهادار تهران	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۱۲	سید علی واعظ، رحیم بنابی قدیم، ساناز قاسمی	تأثیر تضاد مالکانه بر رابطه بین انعطاف پذیری مالی با شدت سرمایه گذاری و تقسیم سود	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۱۳	سامان رحمانی، نوروزآباد، علی اصغر انواری رستمی، کرم خلیلی، اسفندیار محمدی	تأثیر نیاز به تامین مالی برون سازمانی بر شاخص حاکمیت شرکتی و ارزش شرکت: شواهدی از بورس اوراق بهادار تهران	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۱۴	رضا دانشور بنداری، ابوالقاسم مسیح آبادی، محمد رضا شورورزی	نقش کیفیت سود در برآورد ریسک اعتباری	رگرسیون خطی	راهبرد مدیریت مالی	۱۴۰۰
۱۵	احمد خدای پور، علیرضا رحیمی، عارف فروغی	بررسی تأثیر توانایی مدیریتی بر رابطه بین اجتناب مالیاتی و کارایی سرمایه گذاری	رگرسیون خطی	راهبرد مدیریت مالی	۱۳۹۹
۱۶	عبدالرضا شاکری، نگار خسروی پور، سیده	برآورد ریسک سیستمی نظام بانکی با استفاده از سنجه های CoVaR و MES	رگرسیون خطی و رگرسیون	راهبرد مدیریت	۱۳۹۹

ردیف	نام پژوهش گر	عنوان مقاله	مدل رگرسیون مورد استفاده	نشریه	سال چاپ
	محبوبه جعفری		چندک	مالی	
۱۷	محمد مشهدی زاده ، محسن دستگیر ، سهیل سلحشور	الگویی برای شناسایی عدم اطمینان و افزایش انعطاف در تصمیمات بودجه بندی سرمایه‌ای با رویکرد اختیار سرمایه‌گذاری (مطالعه موردی: نیروگاه برق فتوولتائیک جنوب اصفهان)	رگرسیون خطی	راهبرد مدیریت مالی	۱۳۹۹
۱۸	حسین فخاری ، مهراب نصیری	تأثیر عملکرد شرکت بر ریسک سقوط آتی قیمت سهام	رگرسیون خطی	راهبرد مدیریت مالی	۱۳۹۹
۱۹	رویا دارابی	بررسی کارایی الگوریتم های بهینه سازی کرم شب تاب و رگرسیون ماشین بردار پشتیبان جهت پیش بینی هزینه سرمایه	رگرسیون بردار پشتیبان	راهبرد مدیریت مالی	۱۳۹۹
۲۰	سمیرا زمهریرلو ، غلامرضا منصور فر ، فرزاد غیور	تبیین عوامل مؤثر بر صدور و ابطال واحدهای صندوقهای سرمایه‌گذاری مشترک - مقایسه رفتار صندوق های سرمایه‌گذاری در سهام و صندوق های سرمایه‌گذاری با درآمد ثابت	رگرسیون لجستیک	راهبرد مدیریت مالی	۱۳۹۹
۲۱	مریم دولو، علیرضا ورزیده، آرین صفری	بررسی تاثیر دیرش سهام بر ریسک کل و ریسک سیستماتیک	رگرسیون خطی	راهبرد مدیریت مالی	۱۳۹۹
۲۲	هانیه نیکو ، سید کاظم ابراهیمی ، فاطمه جلالی	گرایش احساسی سرمایه‌گذار، نوسان‌پذیری غیرسیستماتیک و قیمت‌گذاری نادرست سهام در شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران	رگرسیون خطی	راهبرد مدیریت مالی	۱۳۹۹
۲۳	قادر داداش زاده ، رضوان حجازی	ارزش انعطاف‌پذیری مالی، کارایی سرمایه‌گذاری و سرعت تعدیل سرمایه در گردش	رگرسیون خطی	راهبرد مدیریت مالی	۱۳۹۹
۲۴	سید محمود موسوی شیری ، احسان قدردان ، روژین صالحی آزاد	شواهد تجربی از نقش خوش بینی مدیریت بر بازده سهام	رگرسیون خطی	راهبرد مدیریت مالی	۱۳۹۹
۲۵	سیمین راجی زاده ، امیرحسین تائبی نقندری، حدیث زینلی	تأثیر میانجی شکنندگی سیستم بانکی بر رابطه‌ی حساسیت سهامداران کنترلی و واگرایی نرخ بهره	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۴۰۰
۲۶	علی مهرنوش، علی	ابزارهای مالی مشتقه (اختیار معامله و اختیار	رگرسیون خطی	مهندسی	۱۴۰۰

ردیف	نام پژوهش‌گر	عنوان مقاله	مدل رگرسیون مورد استفاده	نشریه	سال چاپ
	جعفری، سید حسین نسل موسوی	فروش تبعی (و هم زمانی بازده سهام ؛ شواهدی از بازار سرمایه ایران		مالی و مدیریت اوراق بهادار	
۲۷	شعبان محمدی ، هادی سعیدی ، عبدالحسین طالبی نجف آبادی ، قاسم الهی شیروان	ارائه مدلی برای پیش‌بینی سری زمانی قیمت های نوین‌دار سهام با استفاده از تحلیل طیف تکین، رگرسیون بردار پشتیبان به‌همراه بهینه‌سازی ازدحام ذرات و مقایسه آن با عملکرد مدل های تبدیل موجک، شبکه عصبی، فرآیند میانگین متحرک خودرگرسیون و رگرسیون چندجمله‌ای	رگرسیون بردار پشتیبان- رگرسیون چندجمله‌ای	مهندسی مالی و مدیریت اوراق بهادار	۱۴۰۰
۲۸	حمزه حسین پور ،احمد خدای پور، امید پور حیدری	تاثیر عوامل رفتاری مبتنی بر نظریه چشم انداز بر قدرت توضیح دهنده مدل پنج عاملی فاما و فرنچ	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۴۰۰
۲۹	سپیده راجی زاده ، امیرحسین تائبی نقندری ، حدیث زینلی	تأثیر اصطکاک مالی بر سرعت همگرایی قیمت سهام	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۴۰۰
۳۰	سعید مشتاق ، فرهاد حسین زاده لطفی ، اسمعیل فدایی نژاد	ارائه مدل پرتفوی بهینه از طریق مدل پیش بینی شاخص بازار و با وجود حافظه بلندمدت با استفاده از شبکه عصبی	رگرسیون بردار پشتیبان	مهندسی مالی و مدیریت اوراق بهادار	۱۴۰۰
۳۱	کامیار طالب نیا ، حمیدرضا وکیلی فرد	بررسی رابطه متقابل توانایی مدیریت و ارزش افزوده اقتصادی با استفاده از معادلات همزمان	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۴۰۰
۳۲	لیلا عبدالله زاده ،فرهاد حنیفی ، میر فیض فلاح	ارائه مدلی برای پیش‌بینی ریسک ریزش قیمت سهام در بورس اوراق بهادار تهران براساس مدل های هاتن و چن	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۴۰۰
۳۳	محمد یوسفیان امیری ، بابک شیرازی ، علی تاجدین ، حسین محمدیان پیشه	تبیین بازده پرتفوی انرژی با استفاده از مدل پنج عاملی فاما و فرنچ	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۴۰۰
۳۴	رضا تهرانی ، محمد	مدلسازی پویای غیرخطی عوامل موثر بر بازار	رگرسیون	مهندسی	۱۳۹۹

ردیف	نام پژوهش گر	عنوان مقاله	مدل رگرسیون مورد استفاده	نشریه	سال چاپ
	اصولیان ، سعید باجلان ، وحید عباسیون	سهام: رویکرد رگرسیون کوانتیل آستانه با اثرات ناهمسانی واریانس بیزی BQTR-GARCH	کوانتیل	مالی و مدیریت اوراق بهادار	
۳۵	محمدعلی صادقی لقمجانی ، جواد رضائی ، مهدی خلیل پور	تبیین نقش تعدیل کننده افق سرمایه گذاری بر بازده اضافی حاصل از به کارگیری استراتژی های شتاب - معکوس در نوسانات قیمت سهام	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۳۹۹
۳۶	سولماز سلامی ، عبدالمجید عبدالباقی عطاآبادی ، روح الله فرهادی	مدلسازی اثر معاملات نوپزی بر بازده حدی سهام بر مبنای رهیافت رگرسیون چندکی	رگرسیون چندکی	مهندسی مالی و مدیریت اوراق بهادار	۱۳۹۹
۳۷	سهیلا حقوقی ، محمد ابراهیم آقابابائی	بررسی اثربخشی معاملات آتی سکه طلا جهت پوشش ریسک نوسانات قیمت سهام	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۳۹۹
۳۸	سپیده عرب ، مجید زنجیردار ، حسن زارعی	نقش عوامل ریسکی اخلاص (نویز) و عمق بازار در تبیین بازده آتی سهام	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۳۹۹
۳۹	علیرضا اسلام پور، رویا دارابی	مقایسه قدرت پیش بینی الگوریتم کرم شب تاب، الگوریتم درخت تصمیم و الگوریتم رگرسیون ماشین بردار پشتیبان جهت پیش بینی ریسک سیستماتیک	رگرسیون ماشین بردار پشتیبان	مهندسی مالی و مدیریت اوراق بهادار	۱۳۹۹
۴۰	زینب آریامند ، سید عباس ابراهیمی	اثر هموارسازی سود بر کیفیت درآمدها و عملکرد مالی شرکت های پذیرفته شده در بورس اوراق بهادار تهران	رگرسیون لجستیک	مهندسی مالی و مدیریت اوراق بهادار	۱۳۹۹
۴۱	اکبر خیام بور ، سینا خریدار ، فرزین رضایی ، محمدرضا وطن پرست	ارزیابی سودمندی در تصمیم افشاء اطلاعات مولفه های ریسک	رگرسیون خطی	مهندسی مالی و مدیریت اوراق بهادار	۱۳۹۹
۴۲	علیرضا عاطفی فر ، زاده فتحی	بررسی اثربخشی شاخص های سلامت مالی به عنوان نمادهای بحران مالی بانکی با بکارگیری مدل های لاجیت چند متغیره (مطالعه موردی بانکهای پذیرفته شده در	رگرسیون لجستیک	مهندسی مالی و مدیریت اوراق بهادار	۱۳۹۹

ردیف	نام پژوهش‌گر	عنوان مقاله	مدل رگرسیون مورد استفاده	نشریه	سال چاپ
		(بورس)			

۴. نکاتی در رابطه با مدل‌های رگرسیونی

در نهایت به منظور انتخاب صحیح مدل رگرسیون و استخراج نتایج قابل‌اتکا و اجتناب از اشتباهات رایج در این حوزه در این بخش راه‌کارهای زیر پیشنهاد می‌گردد:

نکته ۱: قبل از شروع پژوهش، تحقیقات و مطالعات زیادی انجام دهید.

قبل از شروع تجزیه و تحلیل رگرسیون، باید ادبیات تحقیق را به خوبی مرور و بازبینی کنید تا درک درستی از متغیرهای مربوطه و روابط آن‌ها داشته باشید. توسعه پایگاه دانش شما در وهله اول به شما کمک می‌کند تا داده‌های صحیح را جمع‌آوری کنید و به شما امکان می‌دهد بهترین معادله رگرسیون را مشخص کنید.

نکته ۲: در صورت امکان از یک مدل ساده استفاده کنید.

به نظر می‌رسد که مسائل پیچیده باید به معادلات رگرسیونی پیچیده نیاز داشته باشند. با این حال، بررسی تحقیقات نشان می‌دهد که ساده‌سازی معمولاً مدل‌های دقیق‌تری ارائه می‌دهد. در بسیاری از موارد، سه متغیر مستقل برای مسائل پیچیده کافی است. نکته این است که با یک مدل ساده شروع کنید و تنها زمانی که واقعاً به آن نیاز است، آن را پیچیده‌تر کنید. وقتی چندین مدل با توانایی‌های پیش‌بینی قابل مقایسه وجود دارد، ساده‌ترین مدل را انتخاب کنید زیرا احتمالاً بهترین مدل است. مزیت دیگر این است که مدل‌های ساده‌تر قابل توضیح و درک آسان برای دیگران هستند.

سعی کنید متغیرهایی را که به هم مرتبط نیستند از مدل حذف کنید و فقط آنهایی را که دارای رابطه واقعی هستند نگهداری کنید تا به یک تعادل طلایی برسید.

نکته ۳: نمودارها، اطمینان و فواصل پیش‌بینی را در نتایج قرار دهید.

این نکته بر این واقعیت تمرکز دارد که نحوه ارائه نتایج خود می‌تواند بر نحوه تفسیر افراد تأثیر بگذارد. اطلاعات می‌تواند یکسان باشد، اما سبک ارائه می‌تواند واکنش‌ها و برداشت‌های متفاوتی را در پی داشته باشد.

نکته ۴: نمودارهای باقیمانده را بررسی کنید.

یکی از روش‌های مرسوم برای تخمین پارامترهای مدل رگرسیونی روش حداقل مربعات معمولی می‌باشد که استفاده از این روش دارای یک سری مفروضاتی می‌باشد. اگر مفروضات را نقض کنید، خطر تولید نتایجی را دارید که نمی‌توانید به آن‌ها اعتماد کنید. نمودارهای باقیمانده راهی سریع و آسان برای بررسی مشکلات در مدل رگرسیونی شما هستند. این نمودارها همچنین می‌توانند به شما در انجام تنظیمات کمک کنند. پس از برازش یک مدل رگرسیون، بررسی نمودارهای باقیمانده بسیار مهم است. اگر نمودارهای شما الگوهای ناخواسته‌ای را نشان می‌دهند، نمی‌توانید به ضرایب رگرسیون و سایر نتایج عددی اعتماد کنید.

نکته ۵: متغیرهایی با دامنه تغییرات بالا را بی‌مقیاس نماید.

بی مقیاس کردن متغیرها روشی برای یکنواخت کردن بازه مقادیر مربوط به متغیرهای پژوهش است. اگر واحد سنجش متغیرهای مورد مطالعه متنوع باشد با استفاده از روش های بی مقیاس کردن می توان متغیرهای عددی را بی مقیاس کرد.

امروزه در پژوهش های کمی مالی از متغیر اندازه سازمان یا اندازه شرکت به دلیل اهمیت تأثیر آن بر متغیرهای مالی توجه زیادی شده است و بیشتر پژوهش ها از اندازه شرکت به عنوان متغیرهای مستقل و کنترل استفاده نموده اند. اندازه شرکت ها یا بزرگی و کوچکی آن ها غالباً از طریق میزان دارایی های یک شرکت اندازه گیری می شود. در اکثر پژوهش های مالی صورت گرفته برای محاسبه متغیر اندازه شرکت، استفاده از لگاریتم طبیعی ارزش بازار سهام، دارایی های شرکت و یا کل فروش شرکت استفاده می شود (کاویانی و دیگران، ۱۳۹۵). استفاده از لگاریتم طبیعی برای محاسبه اندازه شرکت باعث می شود تا ضرایب احتمالی این متغیر در مدل، تحت تأثیر اثرات مقیاس های بزرگ قرار گیرد و یکی از روش های بی مقیاس کردن اعداد می باشد. سایر روش های بی مقیاس کردن متغیرها در ادامه معرفی شده است.

نرمال سازی خطی در بازه $[a, b]$ توسط **Error! Reference source not found.** محاسبه می شود (میرزائی و دیگران، ۱۴۰۰؛ دعائی و دیگران، ۲۰۲۱).

$$x_{norm} = (b - a) \left(\frac{(x - x_{min})}{(x_{max} - x_{min})} \right) + a \quad (8)$$

نرمال سازی استاندارد توسط رابطه (۸) محاسبه می شود. با استفاده از این روش نرمال سازی، میانگین متغیرهای بی مقیاس شده برابر با صفر و انحراف معیار آن ها نیز برابر یک می شود. به این روش نرمال سازی، استاندارد کردن متغیر نیز گفته می شود.

$$x_{norm} = \frac{(x - \mu)}{\sigma} \quad (9)$$

دقت نماید نرمال سازی متغیرها با نرمال سازی توزیع متغیرها متفاوت می باشد.

نکته ۶: نرمال بودن توزیع متغیرها را بررسی نمایید.

یکی از مسائل مهم در تحلیل آماری داده ها، انتخاب بین روش های پارامتری یا ناپارامتری به منظور انجام محاسبات و استنباط آماری است. هرچند روش های ناپارامتری ساده تر هستند ولی استفاده از آن ها در اکثر مواقع باعث می شود که نتایج تحلیل ها توان کمتری نسبت به روش های مشابه پارامتری داشته باشد. بنابراین اگر بتوانیم با آزمون نرمال بودن توزیع متغیر، توزیع آن ها را نرمال فرض کنیم، بهتر است که از روش های پارامتری استفاده کنیم. در روش های پارامتری فرض بر این است که متغیرها دارای توزیع نرمال هستند. اما اگر متغیرها دارای توزیع نرمال نبوده ولی شما از روش های پارامتری بر مبنای توزیع نرمال استفاده کنید، ممکن است این یافته های شما در طرح تحقیقاتی و تجزیه و تحلیل آماری داده ها، گمراه کننده یا اشتباه باشد. در نتیجه انجام آزمون نرمال بودن یا تأیید فرض نرمال بودن متغیرها یک نکته کلیدی در تصمیم گیری برای انتخاب روش برای استنباط آماری بر مبنای نمونه های تصادفی است.

بخش عمده ای از تحلیل های آماری مربوط به متغیرهای تصادفی و جوامع با توزیع نرمال است. بسیاری از پدیده هایی که در طبیعت مشاهده می شوند دارای توزیع نسبتاً نرمال می باشند. در صورتی که توزیع داده ها نرمال نباشد می توان با استفاده از رابطه یا تبدیلی، متغیرهای غیر نرمال را به نرمال تغییر شکل داد و بر اساس آن ها استنباط و روش های پارامتری را به کار گرفت. به عنوان مثال اگر توزیع متغیر دارای چولگی باشد می توان با حذف داده های پرت، توزیع را نرمال کرد. روش های مختلفی برای بررسی نرمال بودن داده ها وجود دارد. مانند ترسیم نمودار هیستوگرام و یا آزمون کولموگروف-اسمیرنوف در محیط SPSS

نکته ۷: اگر متغیر وابسته، متغیری پیوسته باشد و مدل از هم خطی بودن رنج می برد و یا تعداد متغیرهای مستقل بسیار زیادی وجود داشته باشد، می توان از رگرسیون مؤلفه های اصلی، رگرسیون کمترین مربعات جزئی، رگرسیون ستیغی، رگرسیون شبکه الاستیک استفاده کرد. اگر روی داده های شمارشی کار می کنید، بهتر است رگرسیون شبه پواسون و رگرسیون دوجمله ای منفی را امتحان کنید. برای جلوگیری از بیش برآزش، می توانیم از روش اعتبارسنجی متقاطع برای ارزیابی مدل های مورد استفاده برای پیش بینی استفاده کنیم. همچنین می توانیم از تکنیک های رگرسیون ستیغی، رگرسیون لاسو و رگرسیون شبکه الاستیک برای اصلاح مشکل بیش برآزش استفاده کنیم.

۵. نتیجه گیری

رگرسیون یک تحلیل آماری بسیار قدرتمند است و به شما امکان می دهد اثرات متغیرهای فردی را منفک کرده و درک کنید، تعاملات را مدل کنید و پیش بینی کنید. تجزیه و تحلیل رگرسیون انعطاف پذیری بالایی را ارائه می دهد. اغلب محققان از دو یا سه روش رایج رگرسیون مانند رگرسیون خطی، رگرسیون چندکی و رگرسیون لجستیک استفاده می کنند، در حالیکه تکنیک های رگرسیونی بسیار گسترده می باشند و انواع مختلفی از روش های رگرسیون به منظور تحلیل داده های کیفی و کمی طراحی شده است. انتخاب صحیح مدل رگرسیون کار پیچیده و مشکلی می باشد. در صورتی که مدل انتخاب شده به درستی انتخاب شود باعث استخراج نتایج صحیح و قابل اتکا خواهد شد. در این پژوهش به معرفی و بررسی این ابزار در حوزه مالی پرداخته و به ارائه راهکارهایی به منظور انتخاب صحیح مدل رگرسیون و جلوگیری از مشکلات رایج در این حوزه پرداختیم.

منابع و مراجع

۱. ابراهیمی، سید کاظم، بهرامی نسب، علی، سجادی نسب، حسین. (۱۴۰۰). بررسی اثر روابط سیاسی بر محافظه کاری سود و زبانی و هزینه بدهی شرکت های پذیرفته شده در بورس اوراق بهادار تهران. راهبرد مدیریت مالی، ۹(۲)، ۸۸-۱۰۷. doi: 10.22051/jfm.2018.21372.1734
۲. اسلام پور، علیرضا، دارابی، رویا. (۱۳۹۹). مقایسه قدرت پیش بینی الگوریتم کرم شب تاب، الگوریتم درخت تصمیم و الگوریتم رگرسیون ماشین بردار پشتیبان جهت پیش بینی ریسک سیستماتیک. مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۳)، ۱-۲۹.
۳. اعتمادی، حسین، زلفی، حسن. (۱۳۹۲). کاربرد رگرسیون لجستیک در شناسایی گزارشگری مالی متقلبانه. دانش حسابرسی، ۱۳(۵۱)، ۰-۰. <https://www.sid.ir/fa/journal/ViewPaper.aspx?id=253168>
۴. امام وردی، قدرت اله و اصغری، لیلا و نقی زاده هنجنی، سمانه، (۱۳۹۸). برآورد رگرسیون خطی در شرایط غیر نرمال بودن جملات خطا: برآورد مدل مناسب بازده سهام صنعت بانکداری، <https://civilica.com/doc/1243217>
۵. امیری، عبدالله، سعیدی، پرویز. (۱۴۰۰). بررسی رابطه بحران مالی، چرخه عمر و استراتژی های تجدید ساختار دارایی و مالی شرکت های دارای بحران مالی. راهبرد مدیریت مالی، ۹(۴)، ۱۷۵-۱۹۰. doi: 10.22051/jfm.2020.25221.2019
۶. آریامند، زینب، ابراهیمی، سید عباس. (۱۳۹۹). اثر هموارسازی سود بر کیفیت درآمدها و عملکرد مالی شرکت های پذیرفته شده در بورس اوراق بهادار تهران. مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۲)، ۱-۲۲.
۷. بارانی، عنایت، گل علیزاده، موسی. (۱۳۹۷). استفاده از رگرسیون آمیخته چندکی برای تحلیل هزینه و درآمد خانوارهای ایرانی. مجله ی بررسی های آمار رسمی ایران، ۲۹ (۲): ۱۸۵-۲۰۵. URL: <http://ijoss.srtc.ac.ir/article-1-278-fa.html>

۸. بت شکن، محمدهاشم، فرهادی، روح الله. (۱۳۹۵). بده و بستان ریسک و بازده: شواهدی از مدل رگرسیون چارکی. بورس اوراق بهادار، ۹(۳۳)، ۴۷-۷۰. <https://www.sid.ir/fa/journal/ViewPaper.aspx?id=349909>
۹. بداغی، حمید، ولیان، حسن، مرتضویان، سیده فاطمه، واسعی چارمحالی، مهدی. (۱۴۰۰). بررسی تاثیر توانایی مدیریتی مدیر عامل در ابهام شرکتی و ارزش نهایی وجه نقد شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. راهبرد مدیریت مالی، ۹(۴)، ۱۹۱-۲۱۶. doi: 10.22051/jfm.2019.21610.1752
۱۰. بدری، احمد. (۱۳۹۳). واکاوی و نقد روش شناسی پژوهش‌های تجربی مالی و حسابداری در ایران. پژوهش‌های تجربی حسابداری، ۴(۲)، ۹۹-۱۲۷. doi: 10.22051/jera.2014.620
۱۱. برادران خانیان، زینب، اصغری‌پور، حسین، پناهی، حسین، کازرونی، سید علیرضا. (۱۳۹۶). اثرات نامتقارن تورم بر کسری بودجه در ایران: رویکرد رگرسیون کوانتایل. فصلنامه علمی نظریه های کاربردی اقتصاد، ۴(۳)، ۱۶۹-۱۹۴.
۱۲. پوررضایی، آرش، کیمیاگری، علی محمد. (۱۴۰۰). تخمین تابع سودآوری با استفاده از متغیرهای ساختار سرمایه در صنایع ایران. راهبرد مدیریت مالی، ۹(۴)، ۴۳-۶۴. doi: 10.22051/jfm.2021.۳۳۹۳۹.۲۴۶۱
۱۳. پیمانی فروشانی، مسلم، امیری، میثم، شیرازی، فاطمه. (۱۴۰۰). کاربرد تسلط تصادفی مبتنی بر رگرسیون چندک در تحلیل فرصت‌های آربیتراژی، کارایی بازار و ترجیحات سرمایه‌گذاران. تحقیقات مالی، ۲۳(۲)، ۱۷۲-۱۹۵. doi: ۱۰.۲۲۰۵۹/frj.2021.301157.1007014
۱۴. تهرانی، رضا، اصولیان، محمد، باجلان، سعید، عباسیون، وحید. (۱۳۹۹). مدل‌سازی پویای غیرخطی عوامل موثر بر بازار سهام: رویکرد رگرسیون کوانتیل آستانه با اثرات ناهمسانی واریانس بیزی (BQTR-GARCH). فصلنامه علمی و مدیریتی اوراق بهادار، ۱۱(۴۵)، ۳۴۳-۳۷۲.
۱۵. حسن‌زاده فاطمه، کاظمی ایرج. مقایسه‌ی انواع مدل‌های رگرسیون در تحلیل داده‌های شمارشی (۱۳۹۰). مجله‌ی بررسی‌های آمار رسمی ایران، ۲۲(۱): ۳۲-۱۷. URL: <http://ijoss.srtc.ac.ir/article-۱۷-۳۲>; ۱-۶۳-fa.html
۱۶. حسین پور، حمزه، خدای پور، احمد، پور حیدری، امید. (۱۴۰۰). تاثیر عوامل رفتاری مبتنی بر نظریه چشم انداز بر قدرت توضیح دهنده‌ی مدل پنج عاملی فاما و فرنچ. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۹)، ۲۱۱-۲۲۹.
۱۷. حقوقی، سهیلا، آقابابائی، محمد ابراهیم. (۱۳۹۹). بررسی اثربخشی معاملات آتی سکه طلا جهت پوشش ریسک نوسانات قیمت سهام. مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۳)، ۱۳۱-۱۵۰.
۱۸. خدای پور، احمد، رحیمی، علیرضا، فروغی، عارف. (۱۳۹۹). بررسی تأثیر توانایی مدیریتی بر رابطه بین اجتناب مالیاتی و کارایی سرمایه‌گذاری. راهبرد مدیریت مالی، ۸(۴)، ۱۱۱-۱۳۴. doi: 10.22051/jfm.2020.27194.2147
۱۹. خواجوی، شکراله، غیوری مقدم، علی. (۱۴۰۰). بررسی نقش توانایی مدیران و تضاد منافع در مفید بودن چسبندگی هزینه‌ها: شواهدی از تصمیمات آگاهانه مدیران در پدیده چسبندگی هزینه. راهبرد مدیریت مالی، ۹(۲)، ۱-۲۰. doi: ۱۰.۲۲۰۵۱/jfm.2020.28808.2240
۲۰. خیام بور، اکبر، خریدار، سینا، رضایی، فرزین، وطن پرست، محمدرضا. (۱۳۹۹). ارزیابی سودمندی در تصمیم افشاء اطلاعات مولفه های ریسک. مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۲)، ۴۱۹-۴۴۵.
۲۱. داداش زاده، قادر، حجازی، رضوان. (۱۳۹۹). مقاله پژوهشی: ارزش انعطاف‌پذیری مالی، کارایی سرمایه‌گذاری و سرعت تعدیل سرمایه در گردش. راهبرد مدیریت مالی، ۸(۱)، ۱۷۷-۱۹۶. doi: 10.22051/jfm.2018.19359.1615
۲۲. دارابی، رویا. (۱۳۹۹). مقاله پژوهشی: بررسی کارایی الگوریتم های بهینه سازی کرم شب تاب و رگرسیون ماشین بردار پشتیبان جهت پیش‌بینی هزینه سرمایه. راهبرد مدیریت مالی، ۸(۲)، ۱۱۱-۱۳۳. doi: ۱۰.۲۲۰۵۱/jfm.2019.23780.1911

۲۳. دانش جعفری، داوود، بت شکن، محمد-هاشم، پاشازاده، حامد (۱۳۹۵). رتبه بندی بانک ها از نظر مقاومت در برابر ریسک سیستمیک در راستای نظام مالی مقاومتی (روش رگرسیون کوانتایل و همبستگی شرطی پویا). مطالعات راهبردی بسیج، ۱۹(۷۲)، ۷۹-۹۹. <https://www.sid.ir/fa/journal/ViewPaper.aspx?id=358411>
۲۴. دانشور بنداری، رضا، مسیح آبادی، ابوالقاسم، شورورزی، محمد رضا. (۱۴۰۰). نقش کیفیت سود در برآورد ریسک اعتباری. راهبرد مدیریت مالی، ۹(۱)، ۲۰۳-۲۲۴. doi: 10.22051/jfm.2019.23338.1869
۲۵. دانشور مفرد، مریم، حجازی، رضوان، موسوی، میر حسین. (۱۳۹۳). بررسی اثر بازار، نقدشوندگی و تکانه بر افزایش عمده قیمت سهام با استفاده از روش رگرسیون کاکس. راهبرد مدیریت مالی، ۲(۲)، ۱-۱۹. doi: 10.22051/jfm.2015.980
۲۶. دولو، مریم، ورزیده، علیرضا، صفری، آرین. (۱۳۹۹). مقاله پژوهشی: بررسی تاثیر دیرش سهام بر ریسک کل و ریسک سیستماتیک. راهبرد مدیریت مالی، ۸(۲)، ۹۳-۱۰۹. doi: 10.22051/jfm.2020.23721.1905
۲۷. راجی زاده، سپیده، تائبی نقندری، امیرحسین، زینلی، حدیث. (۱۴۰۰). تأثیر اصطکاک مالی بر سرعت همگرایی قیمت سهام. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۷)، ۴۳-۶۹.
۲۸. راجی زاده، سیمین، تائبی نقندری، امیرحسین، زینلی، حدیث. (۱۴۰۰). تأثیر میانجی شکنندگی سیستم بانکی بر رابطه حساسیت سهامداران کنترلی و واگرایی نرخ بهره. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۹)، ۸۰-۱۰۳.
۲۹. رحمانی نوروزآباد، سامان، انواری رستمی، علی اصغر، خلیلی، کرم، محمدی، اسفندیار. (۱۴۰۰). تاثیر نیاز به تامین مالی برون سازمانی بر شاخص حاکمیت شرکتی و ارزش شرکت: شواهدی از بورس اوراق بهادار تهران. راهبرد مدیریت مالی، ۹(۱)، ۲۴-۱. doi: 10.22051/jfm.2020.26819.2126
۳۰. رستمی، محمد رضا، انصاری، حجت اله، رحیمی، فرزانه. (۱۴۰۰). ویژگی های شرکتی توضیح دهنده نوسانات غیرسیستماتیک در بورس اوراق بهادار تهران. راهبرد مدیریت مالی، ۹(۴)، ۲۱-۴۲. doi: 10.22051/jfm.2021.35498.2526
۳۱. زلفی، حسن، منصوری، کفسان، افلاطونی، عباس. (۱۴۰۰). ویژگی های جمعیت شناختی مدیران ارشد مالی و استراتژی مدیریت سود. راهبرد مدیریت مالی، ۹(۴)، ۶۵-۸۶. doi: 10.22051/jfm.2021.34355.2477
۳۲. زمهریرلو، سمیرا، منصور فر، غلامرضا، غیور، فرزاد. (۱۳۹۹). مقاله پژوهشی: تبیین عوامل مؤثر بر صدور و ابطال واحدهای صندوقهای سرمایه گذاری مشترک - مقایسه رفتار صندوقهای سرمایه گذاری در سهام و صندوقهای سرمایه گذاری با درآمد ثابت. راهبرد مدیریت مالی، ۸(۲)، ۶۱-۷۰. doi: 10.22051/jfm.2019.22515.1808
۳۳. سلامی، سولماز، عبدالباقی عطاآبادی، عبدالمجید، فرهادی، روح الله. (۱۳۹۹). مدل سازی اثر معاملات نویزی بر بازده حدی سهام بر مبنای رهیافت رگرسیون چندکی. مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۴)، ۱۸۸-۲۰۶.
۳۴. شاکری، عبدالرضا، خسروی پور، نگار، جعفری، سیده محبوبه. (۱۳۹۹). برآورد ریسک سیستمی نظام بانکی با استفاده از سنجح های MES و CoVaR. راهبرد مدیریت مالی، ۸(۴)، ۲۳۵-۲۵۶. doi: 10.22051/jfm.2019.27534.2170
۳۵. صابرمهانی، معصومه، نیکبخت، محمد رضا، دلدار، مصطفی. (۱۴۰۰). آزمون اثربخشی تمرکز مشتری بر سرعت تعدیل نگهداشت وجه نقد در بورس اوراق بهادار تهران. راهبرد مدیریت مالی، ۹(۱)، ۵۵-۷۸. doi: 10.22051/jfm.2019.23264.1864
۳۶. صادقی لقمجانی، محمدعلی، رضانی، جواد، خلیل پور، مهدی. (۱۳۹۹). تبیین نقش تعدیل کننده افق سرمایه گذاری بر بازده اضافی حاصل از به کارگیری استراتژی های شتاب - معکوس در نوسانات قیمت سهام. مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۴)، ۱۱۴-۱۳۲.
۳۷. صمدی، علی-حسین. (۱۳۸۶). رگرسیون جعلی: مفهوم و نتایج. پژوهشهای اقتصادی ایران، ۹(۳۲)، ۱۱۱-۱۳۶. <https://www.sid.ir/fa/journal/ViewPaper.aspx?id=74666>
۳۸. طالب نیا، کامیار، وکیلی فرد، حمیدرضا. (۱۴۰۰). بررسی رابطه متقابل توانایی مدیریت و ارزش افزوده اقتصادی با استفاده از معادلات همزمان. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۷)، ۳۱۴-۳۳۰.

۳۹. عاطفی فر، علیرضا، فتحی، زاداله. (۱۳۹۹). بررسی اثربخشی شاخص های سلامت مالی به عنوان نمادهای بحران مالی بانکی با بکارگیری مدل های لاجیت چند متغیره (مطالعه موردی بانکهای پذیرفته شده در بورس). مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۲)، ۳۳۳-۳۶۱.
۴۰. عبدالله زاده، لیلا، حنیفی، فرهاد، فلاح، میر فیض. (۱۴۰۰). ارائه مدلی برای پیش بینی ریسک ریزش قیمت سهام در بورس اوراق بهادار تهران براساس مدل های هاتن و چن. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۶)، ۱۷۰-۱۴۶.
۴۱. عرب، روح اله، حسینی، سیده زهرا، غلامرضا پور، محمد. (۱۴۰۰). بررسی رابطه بین قدرت مدیرعامل و اهرم مالی شرکت. راهبرد مدیریت مالی، ۹(۳)، ۱۲۳-۱۴۰. doi: 10.22051/jfm.2020.26457.2103
۴۲. عرب، سپیده، زنجیردار، مجید، زارعی، حسن. (۱۳۹۹). نقش عوامل ریسکی اخلاقی (نویز) و عمق بازار در تبیین بازده آتی سهام. مهندسی مالی و مدیریت اوراق بهادار، ۱۱(۴۳)، ۲۹۱-۳۱۲.
۴۳. عربی، حمیدرضا، گل پرور، محسن. (۱۳۸۸). سبک رهبری و رضایت شغلی: مقایسه رگرسون گام به گام با رگرسون ریج. روانشناسی تحولی: روانشناسان ایرانی، ۶(۲۱)، ۲۵-۳۴.
۴۴. علی اکبرلو، علیرضا، منصورفر، غلامرضا، غیور، فرزاد. (۱۳۹۹). مقایسه معیارهای تشخیص شرکت های درمانده مالی با استفاده از رگرسون لجستیک و روش های هوش مصنوعی. نشریه چشم انداز مدیریت مالی، ۱۰(۲۹)، ۱۴۷-۱۶۶. doi: 10.29147/jfmp.10.29.147
۴۵. فخاری، حسین، نصیری، مهرباب. (۱۳۹۹). مقاله پژوهشی: تأثیر عملکرد شرکت بر ریسک سقوط آتی قیمت سهام. راهبرد مدیریت مالی، ۸(۳)، ۴۳-۶۲. doi: 10.22051/jfm.2019.25489.2037
۴۶. فدایی نژاد، محمد اسماعیل، اصولیان، محمد، شمس، پروانه. (۱۴۰۰). بررسی چگونگی ارتباط مدیریت جریان وجوه نقد با ساختار سرمایه. راهبرد مدیریت مالی، ۹(۳)، ۶۱-۷۸. doi: 10.22051/jfm.2021.33007.2415
۴۷. کاویان، میثم و سعیدی، پرویز و منتظری، زینب، ۱۳۹۵، مرور و نقدی بر برخی از چالش های موجود در تدوین و ارائه پژوهش های تجربی مالی و حسابداری ایران، <https://civilica.com/doc/541589>
۴۸. کفاش پور یزدی، مهسا، تفتیان، اکرم، معین الدین، محمود. (۱۴۰۰). تأثیر تقارن اطلاعاتی بر رابطه کیفیت اطلاعات داخلی و ریسک سیستماتیک با استفاده از مدل های ایستا و پویا. راهبرد مدیریت مالی، ۹(۳)، ۱۴۱-۱۶۰. doi: 10.22051/jfm.2020.28848.2243
۴۹. کمیجانی، اکبر، رحمانی، تیمور، سیفی، طاهره، مهرآرا، محسن. (۱۳۹۵). بررسی اثر واکنش به قیمت دارایی ها در سیاست پولی بر تثبیت اقتصاد کلان مبتنی بر رویکرد کوانتایل (مطالعه بین کشوری). فصلنامه علمی پژوهشی راهبرد اقتصادی، ۵(۱۷)، -.
۵۰. محمدی، شعبان، سعیدی، هادی، طالبی نجف آبادی، عبدالحسین، الهی شیروان، قاسم. (۱۴۰۰). ارائه مدلی برای پیش بینی سری زمانی قیمت های نوپرداز سهام با استفاده از تحلیل طیف تکین، رگرسون بردار پشتیبان به همراه بهینه سازی ازدحام ذرات و مقایسه آن با عملکرد مدل های تبدیل موجک، شبکه عصبی، فرآیند میانگین متحرک خودرگرسون و رگرسون چندجمله ای. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۹)، ۳۹۹-۴۲۱.
۵۱. مشتاق، سعید، حسین زاده لطفی، فرهاد، فدایی نژاد، اسمعیل. (۱۴۰۰). ارائه مدل پرتفوی بهینه از طریق مدل پیش بینی شاخص بازار و با وجود حافظه بلندمدت با استفاده از شبکه عصبی. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۷)، ۴۵۰-۴۶۹.
۵۲. مشهدی زاده، محمد، دستگیر، محسن، سلحشور، سهیل. (۱۳۹۹). الگویی برای شناسایی عدم اطمینان و افزایش انعطاف در تصمیمات بودجه بندی سرمایه ای با رویکرد اختیار سرمایه گذاری (مطالعه موردی: نیروگاه برق فتوولتائیک جنوب اصفهان). راهبرد مدیریت مالی، ۸(۴)، ۱۵۷-۱۸۶. doi: 10.22051/jfm.2019.22744.1827

۵۳. موسوی شیرینی، سید محمود، قدردان، احسان، صالحی آزاد، روزین. (۱۳۹۹). مقاله پژوهشی: شواهد تجربی از نقش خوش بینی مدیریت بر بازده سهام. راهبرد مدیریت مالی، ۸(۱)، ۱۴۳-۱۶۲. doi: 10.22051/jfm.2019.22752.1821
۵۴. مومنی، منصور، و فعال قیومی، علی. (۱۳۸۸). مقایسه انواع تحلیل های رگرسیونی برای داده های حسابداری. بررسی های حسابداری و حسابرسی، ۱۶(۵۸)، ۱۰۳-۱۱۲. <https://www.sid.ir/fa/journal/ViewPaper.aspx?id=111469>
۵۵. مهرنوش، علی، جعفری، علی، نسل موسوی، سید حسین. (۱۴۰۰). ابزارهای مالی مشتقه (اختیار معامله و اختیار فروش تبعی) و هم زمانی بازده سهام؛ شواهدی از بازار سرمایه ایران. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۹)، ۴۴۶-۴۶۰.
۵۶. میرزائی، سیداحمد، نیکدل، زکیه، نیکدل، زهرا. (۱۴۰۰). بررسی آموزش شبکه عصبی با استفاده از الگوریتم های فراابتکاری به منظور پیش بینی شاخص کل در بورس ایران. فصلنامه تحلیل بازار سرمایه، ۱(۲)، ۱۸۸-۲۱۲.
۵۷. نوظهور، ندا، کاویانی، میثم. (۱۳۹۹). تحلیل رابطه بین نسبت های دوپونت و سیاست تقسیم سود (رهیافت تحلیل رگرسیون های توبیت و لجستیک). مدیریت نوآوری و راهبردهای عملیاتی، ۱(۲)، ۰-۰. <https://www.sid.ir/fa/journal/ViewPaper.aspx?id=542153>
۵۸. نیکو، هانیه، ابراهیمی، سید کاظم، جلالی، فاطمه. (۱۳۹۹). مقاله پژوهشی: گرایش احساسی سرمایه گذار، نوسان پذیری غیرسیستماتیک و قیمت گذاری نادرست سهام در شرکت های پذیرفته شده در بورس اوراق بهادار تهران. راهبرد مدیریت مالی، ۸(۱)، ۶۵-۸۵. doi: 10.22051/jfm.2019.24325.1952
۵۹. واعظ، سید علی، بنایی قدیم، رحیم، قاسمی، ساناز. (۱۴۰۰). تاثیر تضاد مالکانه بر رابطه بین انعطاف پذیری مالی با شدت سرمایه گذاری و تقسیم سود. راهبرد مدیریت مالی، ۹(۱)، ۱۳۳-۱۵۴. doi: 10.22051/jfm.2020.26310.2095
۶۰. یوسفیان امیری، محمد، شیرازی، بابک، تاجدین، علی، محمدیان پیشه، حسین. (۱۴۰۰). تبیین بازده پرتفوی انرژی با استفاده از مدل پنج عاملی فاما و فرنچ. مهندسی مالی و مدیریت اوراق بهادار، ۱۲(۴۶)، ۲۶۹-۲۹۳.
۱. Doaei, M., Mirzaei, S., Rafigh, M. (2021). Hybrid Multilayer Perceptron Neural Network with Grey Wolf Optimization for Predicting Stock Market Index. *Advances in Mathematical Finance and Applications*, 6(4), 883-894. doi: 10.22034/AMFA.2021.19.3474.1452
 ۲. Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, 2(2), 111-120.
 ۳. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.