A Comparative Study on Existing Techniques for Variable Reduction Including Factor Analysis, Principal Component Analysis, Correlation-Based Techniques, and Relief in Predicting the Risk of Stock Price Crash in Tehran Stock Exchange

Journal of Industrial Strategic Management
Year, 2022,
Volume 7 (Issue 2),
Pages: 85 - 102.

Hassan Mohammadi[a], Alireza Zarei Soudani[b]*

a.Ph.D. Candidate, Department of Accounting, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran.
b.Assistant Professor, Department of Accounting, Flavarjan Branch, Islamic Azad University, Flavarjan, Iran.

Article history:
Received: 21/05/2022
Received in revised: 02/07/2022
Accepted: 01/08/2022

Abstract
In order to determine how well component extraction techniques (principal component analysis and factor analysis) and variable (feature) reduction techniques (correlation-based and relief techniques) perform in identifying the likelihood of future stock price crashes, the current study looked into these techniques' performance and effectiveness. To do this, a sample of 80 companies listed on Tehran Stock Exchange between 2006 and 2017 was chosen, and 17 often used major characteristics influencing the probability of a stock price drop were found by studying the literature. The criteria for evaluating the efficacy of the procedures under consideration included the mean absolute magnitude percentage error, root mean square error, and coefficient of determination. In comparison to the use of all key explanatory variables, the results showed that variable reduction and component extraction strategies work significantly better and are more effective at predicting the likelihood of future stock price crashes.

Keywords: Stock price crash risk, variable reduction technique, component extraction technique, linear prediction method.

1. Introduction

The information provided by the managers to the capital market is always an important source of information. Since, the managers are responsible for collecting and providing information, they will be able to involve their personal judgments in estimates and affect the market and stock prices by disclosing information. The internal information management of a corporation, according to many academics, including Chen et al. (2001), may be the cause of fluctuations in stock price. If information is brought into the market at random and the process of disseminating information is methodical, regardless of whether it is good or negative, i.e., if the managers swiftly reveal all information, then stock returns will be symmetrically distributed. This implies that the typical positive return on good news should be equivalent to the typical negative return on bad news (Kothari et al., 2009). The managers, meanwhile, are always driven to keep bad news and information from investors so they can invest more money in the company. By adding a period expense as an asset, for instance, they could lower costs and increase profit reported in the financial statements. So, the business unit looks better than reality and the individuals outside the business unit are more motivated to invest in it (Ball, 2009). Economic unit managers frequently simultaneously release positive news and cover up negative news. These varied reasons for disclosing are the result of several variables, including compensation contracts and career concerns (Ball et al., 2013). Stock price crash risk is crucial for the investors to consider because it greatly affects their welfare (Robin & Zhang, 2015). An increase in ambiguity increases the probability of a stock price crash. Additionally, there is a significant link between the danger of stock price crashes and profit management (Hutton et al., 2009). Because the investors often disagree with stocks that have elevated stock price crash risk and find it challenging to diversify stock price crash risk, stock price crash risk plays a significant role in stock portfolio management with respect to risk management. Companies are heavily rewarded for lowering the danger of stock price crashes.

According to Hutton et al. (2009), managers are driven to conceal some losses to maintain their positions, since there

* Corresponding author Email address: zareisoudani@finance@gmail.com

is a lack of openness in financial reporting (profit management). While the manager is present in the organization, this process—hiding real loss—continues. His departure adds a significant amount of hidden losses to the market, lowering the stock price. Investors are also unable to recognize initiatives that are losing money due to a lack of transparency. Due to inability of investors to differentiate between profit-making and loss-making projects in the early stages, loss-making projects are continued and make more losses over the time. The negative returns of these projects are accumulated in the company over the time, and whenever their information is disclosed, the stock prices fall sharply. Unfavorable issue of information asymmetry affects frequently the stock markets, the results in a discrepancy between a stock's intrinsic value and the value estimated by investors. Information asymmetry influences investors' poor economic judgement in this way. As a result, the primary goal of this study is to investigate the variables that influence the likelihood of future stock price crashes so that the findings can assist managers, investors, and all other market players in making wise investment choices. Regarding their effects on information asymmetry and firm value, these factors will alter the likelihood of a stock market crash. The current study examined the effects of corporate governance mechanisms, audit quality, income quality, earnings management, product market competition, accounting conservatism, and executive overconfidence on the risk of a future stock price crash at Tehran Stock Exchange in light of the impact of these factors on the risk of a stock price crash and the significance of risk at the national level. It should be emphasized that no research has been done to date to estimate the likelihood of a stock price fall utilizing variable reduction techniques. Additionally, the main objective and focus of earlier studies on the danger of stock price crashes was to suggest suitable and accurate prediction models, but suitable methods for choosing the predictor variables were given less consideration. In the current study, it was attempted to provide empirical evidence about the performance of linear methods and predictor variable selection methods in predicting stock price crash risk for the companies listed on Tehran Stock Exchange, taking into account the significance of predicting stock price crash risk on the Stock Exchange. In this regard, the effectiveness of linear methods such as principal component analysis, factor analysis, correlation-based methods, and relief procedures was compared to the other models without variable reduction, and additionally, the efficiency of the principal component analysis, correlation-based factor analysis, and relief procedures in choosing the best. The following key questions were created by the primary goal of determining how well variable reduction and component extraction methodologies predict the probability of stock price crashes for the companies listed on Tehran Stock Exchange:

✓ Are the optimal explanatory variable selection methods effective in modeling stock price crash risk?

✓ Is there any difference between the effectiveness of different optimal variable selection methods when modeling stock price crash risk?

These topics are discussed in the sections as follows: The theoretical underpinnings and literature review on the variables that influence the risk of a stock price falling and the well-known effective variables are included in the second section. The third portion comprises the technique, explanatory variables, and statistical sampling. The fourth section contains the results of the hypothesis testing. The fifth section concludes by discussing the arguments, limitations, and suggestions for further research.

2. Review of Literature

Factors affecting the risk of stock price crash

Corporate governance mechanism

The managers may egoistically pursue short-term strategies that are subpar for stakeholders with long-term goals (Bebchuk & Stole, 1993). Managers may be motivated to share common trends and engage in pricey projects in the market, according to Baker et al. (2003). This general hypothesis, based on previous evidence indicating over investment in fixed assets, supports inflationary performance during the period. In companies with profit bubble (inflation), very high rents and investments, as well as discretionary accruals are positively correlated with company's investment. Generally, when the real growth rate is determined, these undesirable investment policies will outline strong short-term investment opportunities, leading to a shortage of capital and a fall in stock prices (Basu, 1998).

Moreover, the managers can manipulate financial statements and engage in the well-known earnings management in support of pretending to have strong growth opportunities (Panayiotis et al., 2013). According to Kothari et al. (2009), the managers have financial incentives to keep bad information from the market in order to advance their careers and the financial well-being of their companies. Similar to this, it is suggested by Ball (2009) that managers may be driven to withhold unpleasant news for non-financial reasons such as creating governance and upholding self-esteem. Nevertheless, holding onto terrible news for a long time is still intolerable (Kothari et al., 2009). At times, unexpectedly disclosed information enters the market, causing stock prices to decline (Hutton et al., 2009). Information asymmetry between managers and stakeholders may be the cause of these opportunistic managerial actions. In fact, various financial scandals represent that such representation problem can have severe consequences.

According to the literature, the organizations with effective corporate governance will lessen these representational issues by bridging the knowledge gap between executives and stakeholders. (Karamanou & Vafeas, 2005) demonstrate that voluntary disclosure of financial information is related to the effectiveness of the board of directors and audit committees. Furthermore, corporate governance standards limit poor investment choices. Finally, while some research indicates relatively mixed

results or linkages, others show that corporate governance reduces earnings management strategies.

Executive overconfidence

Managers that are overconfident often overestimate the cash flows from their own investment projects and their own performance (Malmendier et al., 2011). Therefore, they frequently regard ongoing projects with a negative net present value as value-creating enterprises. Furthermore, a CEO who is overconfident and believes in carefully choosing investment plans may underestimate their power to influence outcomes and the likelihood of failure (Malmendier & Tate, 2005). CEO arrogance results in the long-term continuation of projects with a negative net present value. A stock market meltdown will eventually result from these projects' cumulatively poor performance (Bon Kim & Zhang, 2013). The presentation of financial facts to the stock market is also impacted by executive overconfidence. Executives may have a tendency to openly share negative feedback on projects because they believe that impatient investors who are prone to short-term news will force executives to give seemingly positive information on projects with a temporary negative net present value. This is because executive overconfidence causes them to mistakenly consider the projects with negative net present value as the ones with positive net present value. Such leaders may even use positive accruals and voluntary disclosure to convey to the stock market their upbeat views of the company's long-term outlook. As a result, CEO arrogance can also result in the hoarding of bad news, which eventually causes a stock price drop (Bon-Kim et al., 2016).

Quality of financial reporting

Asymmetry in firm stock information is decreased by the quality of financial reporting. The intrinsic worth of business stock diverges from the value placed on it by market investors when information asymmetry increases. As a result, investors will make poor financial judgments as the actual value of the shares differs from the predicted value. When one party effectively communicates with the other while possessing more information than the other, information asymmetry arises. On the other hand, as reporting quality improves, the transparency and quality of information disclosure at the level of the information environment will be better too. Increased information transparency reduces uncertainty about the company stock and fluctuations in stock returns, and ultimately leads to stock price crash. Moreover, increased quality of financial reporting reduces the risk of moral hazard and improper selection through reducing information asymmetry caused by representation problems, and reduces risk and financing cost for the company by reducing management monitoring costs for stakeholders (Verdi, 2006).

Audit quality

For the following reasons, high-quality auditors can lessen unfavorable news hoarding: First of all, competent auditors are more likely to uncover negative news quickly and raise the caliber of financial accounts because of their capacity. Second, high-quality auditors are more motivated to ensure prompt disclosure of negative news and prevent managers from hoarding information for later use for the reasons like maintaining credibility and limiting responsibility. This argument suggests that it is reasonable to believe that high audit quality will result in prompt disclosure of negative news and, as a result, a decreased chance of future stock price crashes. Unaware investors' reactions to price movements, particularly when they are unfavorable, might cause further price swings and, as a result, raise the danger of declining stock prices.

In general, a high-quality audit can mitigate these consequences by lessening investor conflict. The discussion that came before it demonstrates a negative relationship between the risk of a stock price drop and high-quality audits (Robin & Zhang, 2015).

Product market competition

Economists believe that product market exclusivity increases representation problems. Due to the more hazy information environment and resulting information asymmetry between executives and investors, businesses operating in monopolistic markets and industries suffer. Investors won't be able to quickly evaluate a company's performance in such situations.

Additionally, it is assumed that the rivalry in the product market is a great mechanism for allocating resources in the best way possible and that it has disciplinary effects on managers' behavior and inefficiency. Therefore, product market competition, as an external mechanism can effectively prevent such undesirable managerial decisions by regulating investment, preventing earnings management and improving information environment (Panayiotis et al., 2013). These results imply that product market competition serves as an external mechanism that effectively reduces managerial opportunistic behavior, so they must be connected to stock price crashes.

Earnings management

Earnings management tends to mislead the users of financial statements or deviate from the contractual results depending on accounting profit. Managers can wait until year-end and use discretionary accruals to manage reported earnings to reach the desired level, but this may increase the risk of more earnings being manipulated than the current accruals due to the limitation of discretion over discretionary accruals by generally accepted accounting principles (Barton & Simko, 2002).

In the lack of full financial reporting transparency, managers are given the opportunity to hide negative information inside the company to maintain their professional career and credibility. Hence, this negative information is accumulated in the company. When the accumulated negative information reaches its peak, keeping it for a long time becomes impossible and costly. As a result, the market receives a tremendous amount of unfavorable information leading stock prices to plummet.

Quality of earnings

If the quality of the information permits the evaluation of past performance, is effective in measuring profitability, and is predictive of future activities, investors can rely on earning information to estimate their expected returns. Reported earnings help users evaluate performance and measure the profitability of a company. As a result, the investors are very interested in the quality of earnings as a dimension of earnings information, in addition to the fact that reported earnings are significant to investors and influence their decisions. Better earnings lead to better reporting, and better financial reporting leads to greater transparency and better information disclosure. Reduced ambiguity regarding the firm stock and stock return swings ultimately result in a stock price crash because of increased information clarity. Moreover, increased quality of financial reporting reduces the risk of moral hazard and improper selection through reducing information asymmetry caused by representation problems, and reduces risk and financing cost for the company by reducing management monitoring costs for stakeholders (Verdi, 2006).

Accounting conservatism

According to (LaFond & Watts, 2008), conservatism, as a strategic mechanism, limits the motivations and abilities of management to accelerate good news disclosures and delayed bad news announcements. They believe that conservatism accelerates bad news disclosures as loss compared to good news as profit by requiring asymmetric verification capability to identify profits and losses. This reduces managers' incentives to withhold negative news, while hastening the disclosure of good news. As a result, bad news will reach the market sooner than ambiguous good news (Chen et al., 2001). Faster financial statement bad news identification and less knowledge asymmetry are two benefits of accounting practice. To represent a larger risk to the readers of financial statements, any news that could have a negative impact on the company should be recorded more quickly than news that could have a favorable impact. Conservatism is the term used in literature to describe the quicker detection of negative news, which typically depends on news. It is claimed that "conditional conservatism" lowers the probability of stock market crashes and information asymmetry (Bon Kim & Zhang, 2013).

Literature review and hypotheses

Corporate governance standards and the danger of declining stock prices were examined by Panayiotis et al. (2013) in their paper, "Corporate Governance and Firm-Specific Stock Price Crashes." According to the findings, there is a connection between institutional ownership, the proportion of managers owning adaptive stocks, and the likelihood of a stock price drop; because of the demand for immediate results that institutions and managers with institutional ownership place on management stocks. Additionally, the likelihood of a stock market crash is closely connected with financial statement transparency. A negative association between the probability of a stock price crash and the proportion of independent managers on the audit committee was also demonstrated. (Robin & Zhang, 2015) examined the relationship between stock price collapse risk and auditor industry experience. They found that there is a substantial negative link between the usage of high-quality auditors and the probability of future stock price crashes because of information intermediaries and the function of corporate governance, using a large sample from the United States between 1990 and 2009. They also recall how competent auditors can help investors by directly lowering the likelihood of future stock price crashes. Additionally, they demonstrated a significant inverse relationship between the probability of future stock price crashes and continual auditor selection. It was also demonstrated that the relationship between the likelihood of a future stock market crash and financial information transparency, accounting conservatism, and tax avoidance is influenced by the expertise of the auditing sector.

Francis et al. (2014) investigated the impact of abnormal actual operations and real earnings management on the risk of a stock price crash and came to the conclusion that the degree to which a company deviates from industry norms in its actual operations directly affects the likelihood of a future stock price crash. Additionally, they demonstrated how genuine earnings management causes a large decline in stock values the following year. The association between executive overconfidence and the probability of a stock price drop was examined by Bon-Kim et al. (2016). They discovered that companies with overconfident CEOs have a higher chance of a stock price drop than other companies, using data from 1993 to 2010. They ultimately came to the conclusion that CEO overconfidence had less of an effect on the chance of a stock price drop in companies with more cautious accounting standards The effects of several accounting and auditing features on the enhancement of return reporting in predicting the danger of stock price crash were examined by (Dimitrios et al., 2014) in a study titled "Accounting conservative quality of accounting information and crash risk of stock prices." Their findings suggested a link between conditional and unconditional conservatism and the likelihood of a future stock market meltdown. Additionally, they demonstrated how conditional conservatism and the likelihood of a future stock market catastrophe are related to the degree of unconditional conservatism. On the other hand, they demonstrated that auditing features do not appear to assist in predicting the risk of a stock price crash, although there is a strong association between the risk of delayed earnings transparency and a stock price crash, according to the literature.

The connection between institutional investors and stock price crashes was investigated by Callen and Fang (2013). They found a negative link between institutional owners and impending stock price crashes after examining two institutional investors' contrasting viewpoints on monitoring and expropriation. This is because institutional investors, who keep an eye on management's activities, prevent the accumulation of unfavorable news about the firm, which is one of the main causes of stock price crashes. In order to determine whether corporate

governance systems can foresee a company's propensity to experience a stock price collapse, (Andreou et al., 2016) looked into the relationship between corporate governance and stock price crashes. The findings indicated that the company's ownership structure, accounting discrepancies, board structure, and prevention of repeat crashes are all being monitored. These relationships, which are sometimes asymmetric, aim to increase and reduce monitoring actions, and are stronger for companies with bigger representation problems.

Does conservative accounting lessen the likelihood of a stock price crash? Was the title of a study by (Bon Kim & Zhang, 2013) that looked into the connection between the two? According to the findings, conservatism lessens managers' incentives to inflate success and conceal negative information, which lowers the probability of a stock price crash. They demonstrated that conservatism is better able to reduce the risk of future stock price crashes in the case of information asymmetry by using research and development costs, product market conditions, and stakeholder composition as the variables affecting managers' and investors' informational asymmetry.

(Real earnings management and accrual-based earnings management were both examined by Cohen & Zarowin (2010). They discovered that during seasonal stock supply, managers frequently engage in real earnings management and that reduced performance is primarily caused by real earnings management rather than accrual-based earnings management because the manipulation of real activities produces real economic outcomes. (Hong & Stein, 2003) stated that heterogeneous investor behavior is a reason for accelerating stock price crash. Their findings represented that if there is a big difference in the behavior of investors, then fanatic investors will quickly sale their stocks when hearing unfavorable bad news and try to exit the market as soon as possible, which ultimately leads to a fall in stock prices. They also argued that total accumulated news during stock market crash is more likely to become pronounced, resulting in a sharp negative stock returns adjustment.

The association between managers' inflated self-esteem and the choice of financing policy in the firms listed on the Tehran Stock Exchange was examined by (Chavoshi et al., 2015). Cash flow investment and the study's findings demonstrated that there is no connection between executive arrogance and financial choices. They also demonstrated a strong correlation between financial choices and growth prospects, profitability, firm size, and distress risk. (Ahmadpour et al., 2014) looked into how a company's qualities affected the likelihood of a stock price drop. Their findings showed a substantial negative correlation between the likelihood of a stock price fall and the variables return on assets, firm size, market-to-book value ratio, stockholder's equity, and Tobin's Q ratio. Additionally, their findings showed that there is no connection between the probability of a stock price drop and the factors return on equity and financial leverage. The relationship between corporate governance practices and the accuracy of financial reporting was examined by Mehrani et al. (2015). According to their findings, there is a significant positive correlation between independent board members, institutional ownership, and financial reporting quality. As a result, as corporate governance variables increase, income quality indices rise as well, leading to higher financial reporting quality.

The impact of audit quality on future stock returns for the companies listed on the Tehran Stock Exchange was investigated by Ahmadi and Jamali (2014). According to their findings, the size of the audit firm and the expertise of the auditor are highly connected with future stock returns, but the tenure of the auditor and those returns are significantly correlated negatively. The association between conditional conservatism in financial reporting and the possibility of a future stock price crash was examined by Foroughi and Mirzaei (2012). According to their findings, conditional conservatism and the likelihood of a future stock market drop are inversely correlated. This study also demonstrated that conditional conservatism is more effective at lowering the likelihood of a future stock price crash when there is a knowledge asymmetry between managers and investors. Using a sample of 90 companies from 2001 to 2008, Rahnama Rudposhti & Moradi (2005) investigated the impact of accounting conservatism on the risk of stock price crash in Tehran Stock Exchange. Their research revealed a sizable inverse link between the analyzed period's accounting conservatism and stock price collapse. In other words, evidence points to a reduction in the probability of stock price crashes due to accounting conservatism. Additionally, in businesses with high levels of information asymmetry, there is no statistically significant link between conservatism and stock price crashes, and information asymmetry did not enhance the effect of conservatism on lowering the probability of stock price crashes.

The connection between institutional investors and stock return volatility was examined by Fakhari and Taheri (2009). 121 businesses listed on Tehran Stock Exchange for the fiscal year 2008 were included in this descriptive correlational analysis, which was based on cross-sectional data. The findings demonstrated that institutional investors lessen information asymmetry and enhance managers' ability to monitor performance. Additionally, when these stockholders' ownership percentages rise, stock return volatility also falls. The relationship between the accuracy of financial reporting and investment effectiveness was explored by Modares and Hasarzadeh (2008).

The findings of their investigation revealed that higher-quality financial reporting resulted in greater investment efficiency, in addition to the negative and positive correlations between the quality of financial reporting and the level of investment efficiency. The quality of financial reporting can increase investment efficiency by lowering overinvestment because there is a large negative association between investment level and that indicator.

Considering the literature and theoretical foundations, the following hypotheses are developed for experimental analysis:

Main hypothesis: Compared to primary predictor variables, selected or extracted optimal predictor variables greatly improve stock price crash risk prediction.
First sub-hypothesis: optimal predictor variables derived from component analysis considerably outperform primary predictor variables in predicting stock price crash risk.
Second sub hypothesis: Optimal predictor factors derived from principal component analysis greatly outperform primary predictor variables in predicting the likelihood of a stock price fall.
Third hypothesis: Compared to primary predictor factors, optimal predictor variables chosen using the correlation-based method much better forecast the likelihood of a stock price drop.
Fourth sub-hypothesis: In comparison to primary predictor factors, optimal predictor variables chosen via the relief technique much better predict the likelihood of a stock price drop.

3. Material and methods

Statistical population and sample

In the statistical population, there were businesses that were listed on Tehran Stock Exchange between the years 2001-2006. All businesses satisfying the following criteria were included in the sample and all others excluded using the purposive sampling (systematic removal) method:
In order to have comparative data, fiscal year should be ended 20 march.

In order to have homogeneous data, they should be manufacturing companies.Their securities trading should not be suspended for more than three months.Information about the selected variables should be available.
As a result, the statistical sample was chosen from 80 companies registered on the Tehran Stock Exchange.

Data collection

Data was gathered for the current investigation using both laboratory and field techniques. From books, periodicals, and specialized Persian and Latin websites, theoretical underpinnings were gathered. The necessary financial information was gathered via the Tadbir Pardaz and Rahavard Novin software, the financial statements of the companies, and the official website of the Tehran Stock Exchange. Approximately 80 primary predictor factors were found by studying the literature in the first step (including about 250 Persian and English papers). Iranian Research Institute for Information Science and Technology (IranDoc), Islamic World Science Citation Center (ISC), Institute for Humanities and Cultural Studies (Comprehensive Portal of Human Sciences), Noormags, Scientific Information Database (SID), Magiran, Iranian Research Institute for Information Science and Technology (IranDoc), and Iranian Research Institute for Information Science and Technology (IranDoc) as well as theses indexed in IranDoc were the main Persian texts chosen. Additionally, English texts from these and other publications found in Scientific Direct, Springer, JSTOR, and ProQuest were chosen.

Among the identified variables, 20 most commonly used variables were selected to determine or predict the stock returns. The data required was available on the Securities and Exchange Organization database, or Tadbir Pardaz and Rahavard Novin software.

Variable reduction and component extraction techniques

In this study, we used to correlation-base and relief techniques for variable reduction and principal component and factor analyses for component (factor) extraction. The reason for using variable selection and extraction methods is to obtain comparative experimental results. The aforementioned techniques are variable selection techniques for prediction problems in which primary variables are chosen without change, whereas variables are adjusted in variable extraction techniques. Additionally, the aforementioned techniques are applied to prediction issues (including continuous dependent variables), as opposed to some variable selection techniques that are exclusively applied to classification issues (with nominal dependent variables, such as bankruptcy prediction). To reduce independent or explanatory factors in the stock price crash risk problem, these techniques are rarely utilized. Therefore, it is interesting to evaluate how well these methods work at decreasing variables to those of variable extraction methods, particularly factor analysis, which is utilized in most investigations.
Variable extraction methods (factor analysis and principal components analysis) are commonly used in many accounting and financial studies, e.g. (Avellaneda & Lee, 2010) and (Rahnama Rudposhti & Moradi, 2005), to compare the performance of these variable reduction methods with variable selection methods. Moreover, factor analysis is more commonly used to extract descriptive variables of stock price crash risk. Therefore, the other reason for using principal components analysis is to compare its merit with factor analysis.

Relief method

The selection of predictor variables using the relief technique is based on distance. In this method, the Euclidean distance between instances is used to calculate the weight of each variable, which represents the correlation between that variable and the class. Each variable's weight corresponds to how well the intended predictor variable can distinguish between the groups. In this method, a variable is given more weight if it has the same value for instances of the same class and a different value for instances of distinct classes. A randomly chosen instance from the training data set is chosen by the relief method, and its Euclidean distance from both its nearest neighbor in the same class and its nearest neighbor in a different class is then measured. The weights of each variable are then updated using these distances. The algorithm then chooses the variables whose weights are greater than a user-defined threshold (Ahmed & Duellman, 2012).
Relief is a commonly used filter variable selection strategy that essentially arranges variables in descending order and

works to improve classification accuracy. The importance of each characteristic in identifying distinct neighboring instances is graded. Relief looks for the two closest neighbors of each training instance: the nearest hit, which belongs to the same class, and the nearest miss, which belongs to a different class. By dividing the total of nearest hit distances for each training instance by the sum of nearest miss distances for each training instance, the rank assigned to each variable is determined (Ahmed & Duellman, 2012).

Correlation-based approach

A variable is typically considered appropriate if it correlates with the dependent variable (class, in classification) but not with any other, relevant, or additional explanatory factors. The definition can be stated as follows if the correlation between two variables is used as an appropriateness measure: Suitable variables have low correlations with other explanatory variables and high correlations with the dependent variable (class). To put it another way, a variable is considered suitable for prediction if the correlation between it and the dependent variable (class) is high enough to be relevant in predicting the dependent variable (class) and the correlation between it and other explanatory variables does not reach a certain level, making it impossible for it to be predicted by other relevant variables (classification). Finding an appropriate measure of the correlation between variables in this situation, as well as a rational process for choosing appropriate variables based on this measure, is the main challenge in variable selection (Leuz et al., 2003).
The correlation-based approach computes correlations between explanatory features and between explanatory features and dependent characteristics before searching for the feature subset space. The subset with the highest merit (according to Equation 1) is used to reduce the dimension of primary training and experimental data sets (McNichols, 2002):

$$\text{Merit}_s = \frac{k\,\overline{r_{cf}}}{\sqrt{k + k\,(k-1)\,\overline{r_{ff}}}} \qquad (1)$$

Where, Merits denotes exploratory figure of merit for subset S consisting of K features $\overline{r_{cf}}$ is the average correlation between independent and dependent features, $\overline{r_{ff}}$ is the average correlation between independent features. This equation calculates Pearson correlation coefficient in which all features are standardized. The denominator is the redundancy between the independent characteristics, while the numerator is the predictive quality of the collection of independent features.

Principal component analysis

The primary goal of principal component analysis is to reduce the number of correlated features in a set of data while preserving as much variability as possible. This reduction is accomplished by switching to a new set of characteristics (primary components) that are not correlated and are structured in such a way that the few features that were left over at the beginning preserve a significant level of variability over the entire primary set. By calculating eigenvalues and eigenvectors, one can find a linear combination of principal components with maximum variance. The first principal component takes into account as much of the data's variability as it can, and each following component takes into account as much of the variance that is left. As a result, defining and computing primary components is straightforward. A main component analysis is a generalized factor analysis. Principal component analysis and factor analysis differ in that the latter takes into consideration the total of the general and unique variances (specific variance plus error variance) in the data set, whereas factor analysis simply analyzes the general variance (Tsai, 2009). Consider the first principal component. The score of variable i on the first component, C_{i1}, use the weights w_{11}‘... ‘w_{p1} in the form of a linear combination shown by Equation (2) (Khan& Watts, 2009):

$$C_{i1} = y_{i1}w_{11} + y_{i2}w_{22} + \cdots + y_{ip}w_{p1} \qquad (2)$$

This linear combination is chosen in such a way that the sum of the squares C_1 is maximized and $w_{11}^2 + \ldots + w_{p1}^2$ is equal to 1. The second principal component is another linear combination of y_i shown by Equation (3) (Khan& Watts, 2009):

$$C_{i2} = y_{i1}w_{12} + y_{i2}w_{22} + \cdots + y_{ip}w_{p2} \qquad (3)$$

Where, C_2 is the maximum variance only if the correlation between C1 and C2 is equal to 0 and $w_{11}^2 + \ldots + w_{p1}^2$ is equal to 1. Elimination of collinearity due to the large number of effective variables in model is the main advantage of this method. Despite its advantages, it cannot be ignored that PCA is only used to simplify data and reduce its dimensionality, and interpretability of the variables extracted with this algorithm might be challenging in various sciences (Khan& Watts, 2009).

Factor analysis

Data summarization is the primary goal of factor analysis, a broad term for certain multivariate statistical techniques. This strategy analyzes the intercorrelation between a vast number of variables before classifying and identifying them as a few general factors. All variables are taken into account concurrently, and each is treated as a dependent variable in factor analysis. Because factor analysis is an interdependence technique in which all variables are viewed as dependent, it is a multivariate method that does not take independent variables into account. Instead, it attempts to condense a large number of variables into a small number of factors (Kalantari, 2006). In factor analysis, two or more independent variables change separately or interact with each other leading a change in the dependent variable. If there are two independent variables in a problem, the linear model is as follows:

$$y = a_0 + A + B + AB + e \qquad (4)$$

variable; a_0 is a common value (e.g. total average); A denotes the effect of an independent variable; B is the effect of another independent variable; AB represents the simultaneous effect of both dependent variables or the interaction effect; and e denotes the error. In addition to the

effect of a variable, A, and error, e, in one-way analysis of variance, now we face with the effect of a second factor, B, and the effect of a third factor, i.e. the simultaneous effect of A and B on the dependent variable y. In factor analysis, there is no theoretical limitation for the number of independent variables. Equation (5) shows a combination of three variables:

$$y=a_0 + A + B + AB + AC + BC + ABC + e \quad (5)$$

In this equation, there are three independent variables (A, B and C,), interactions between them (AB, AC and BC), and their simultaneous interaction (ABC) (Gunny, 2010). Factor analysis usually proceeds in four steps (Gunny, 2010): (a) Computing the correlation matrix for all variables; (b) Factor extraction; (c) Selection and rotation of factors to simplify factor structure; d) Interpretation of the results.

Data analysis and hypothesis testing

A linear regression analysis was used to predict stock returns after the variables and ideal factors were chosen using any of the variable selection and extraction procedures previously discussed. It is important to note that, in this study, the risk of a stock price drop was predicted using company data from the prior year. We examine the root mean square error, mean absolute magnitude percentage error, and coefficient of determination derived from various prediction algorithms in order to assess the efficacy of various prediction methods. These are the most commonly used performance evaluation measures in prediction problems. (Table 1) shows their formulas (Panayiotis et al., 2015):

Table1. The measures used to evaluate prediction performance

Measure	Formula
Root mean square error (RMSE)	$\sqrt{\frac{\sum_{P=1}^{P}(dp - zp)^2}{P}}$
Coefficient of determination (R^2)	$1 - \frac{\sum_{P=1}^{P}(dp - zp)^2}{\sum_{P=1}^{P}(dp - \bar{d}p)^2}$
Mean absolute magnitude percentage error (MAPE)	$\frac{100}{P} \times \sum_{P=1}^{P}\left\lvert\frac{dp - zp}{dp}\right\rvert$

Z_p: Predicted value; d_p: Actual value; $\bar{d}$: Mean value

Source: (Panayiotis et al., 2015)

Higher coefficient of determination and lower MAPE and RMSE will indicate better prediction performance. Despite the existence of other performance evaluation measures, since they can easily be calculated from the three above-mentioned measures, they are not discussed here. For instance, normalized mean square error (NMSE) and coefficient of determination are complementary, and root mean square error (RMSE) is the second root of mean square error (MSE) (Panayiotis et al., 2015). Additionally, the measurements (mean absolute magnitude percentage error, root mean square error, and coefficient of determination) obtained from variable selection methods were compared to one another and to the measures obtained in the case of no explanatory variable selection in each linear and nonlinear method in order to assess the performance of various optimal variable selection and extraction techniques. When no explanatory variable is chosen, all explanatory variables are used to make predictions (before reducing the number of variables). Weka software version 7.3 was used to do explanatory variable reduction and stock return prediction, with the exception of component analysis, which was done through SPSS software. Analysis of variance (or nonparametric Kruskal-Wallis test in the case of not satisfied parametric assumptions) and paired t-test (or Wilcoxon nonparametric test in the case of not satisfied parametric assumptions) were used to test the main and sub-hypotheses, respectively, based on 100 accuracies resulted from 10 repetitions of 10-fold cross-validation for each prediction method in SPSS software V.21.0.

Modelling

According to the literature and theoretical framework, a multivariate regression model was used in this study. In this regard, we selected the model used in (Panayiotis et al.,2013) and (Francis et al., 2008) to test the main and sub-hypotheses:

$$CRASH_{i,t} = NCSKEW_{i,t} = \alpha_0 + \beta_1 insOwn_{i,t-} + \beta_2 CentOwn_{i,t-1} + \beta_3 ManOwn_{i,t-1} + \beta_4 BrdIndep_{i,t-1} + \beta_5 BrdDobl_{i,t-1} + \beta_6 OPAQUE_{i,t-1} + \beta_7 HHI_{i,t-1} + \beta_8 DUALITY_{i,t-1} + \beta_9 STD_{i,t-1} + \beta_{10} RET_{i,t-1} + \beta_{11} Size_{i,t-1} + \beta_{12} ROE_{i,t-1} + \beta_{13} MTB_{i,t-1} + \beta_{14} Leverage_{i,t-1} + \beta_{15} CRASH_{i,t-1} + \varepsilon_{i,t} \quad (6)$$

Variables

Dependent variable

Following earlier studies (Chen et al., 2001; Hutton et al., 2009), the dependent variable listed below was utilized in the current study to calculate the likelihood that each company's stock price will crash:

$$r_{j.r} = \alpha_j + \beta_{1j} r_{m.r-1} + \beta_{2j} r_{i.r-1} + \beta_{3j} r_{m.r} + \beta_{4j} r_{i.r} + \beta_{5j} r_{m.r+1} + \beta_{6j} r_{i.r+1} + \varepsilon_{jr} \quad (7)$$

Following earlier studies (Chen et al., 2001; Hutton et al., 2009), the dependent variable listed below was utilized in the current study to calculate the likelihood that each company's stock price will crash:

$$r_{j.r} = \alpha_j + \beta_{1j} r_{m.r-1} + \beta_{2j} r_{i.r-1} + \beta_{3j} r_{m.r} + \beta_{4j} r_{i.r} + \beta_{5j} r_{m.r+1} + \beta_{6j} r_{i.r+1} + \varepsilon_{jr} \quad (7)$$

Where, $r_{j.r}$ is the return on stock j in week r, $r_{m.r}$ is the market stock return in week r, and $r_{i,r}$ is the industry stock return based on industry classification in the statistical sample. The residual return from the previous equation plus the log of one is the definition of the firm-specific weekly return (W):

Table2. Definition of primary explanatory variables

Variable	Symbol	Definition
Institutional ownership	*InsOwn*	Stockholders' percentage of ownership (such as banks, insurance and investment institutions)
Ownership concentration	*CentOwn*	Total ownership percentage of three major stakeholders
Managerial ownership	*ManOwn*	Managerial ownership percentage
Opacity of financial information	*OPAQUE*	See Section 3-6-2-1
Board independence	*BrdIndep*	Ownership percentage of non-bound board members divided by the entire board
CEO duality	*BrdDobl*	Indicator variable; it is set equal to 1 if CEO is chairman of the board, otherwise it is set equal to 0.
CEO overconfidence	*OverconfidentCEO*	See Section 3-6-2-2
Quality of financial reporting	*FRQ*	Quality of accruals (Francis et al., 2008) and (McNichols, 2002)
Audit firm size	*BIG*	Indicator variable; it is set equal to 1 if audit organization is the company's audit firm; otherwise it is set equal to 0.
Auditor industry expertise	*SPECIALIST*	Market share approach (Etemadi, 2009)
Auditor tenure	*TENURE*	The length of the auditor-client relationship
Real earnings management	*REARNINGS*	See Sections 3-6-2-3, 3-6-2-4 and 3-6-2-5
Accrual-based earnings management	*DA*	Modified Jones Model (Dechow et al., 1995)
Product market competition	*COMPET*	Includes Herfindahl-Hirschman index, Tobin's Q, Lerner Index LI, and modified Lerner Index LIIA
Conditional conservatism	*CC_SCORE*	(Khan & Watts, 2009)
Unconditional conservatism	*UC_SCORE*	(Givoly & Hayn, 2000) model
Income quality	*IQUALITY*	Includes three parameters of income persistence (Francis et al., 2014), income predictability (Francis et al., 2008), and IEQ (Leuz et al., 2003)

*In order to avoid inconvenience, detailed description is eliminated.

Table3. The optimal variables selected with correlation-based and Relies methods

Correlation-based method		Relief method	
No.	Symbol	No.	Symbol
1	ManOwn	1	ManOwn
2	InsOwn	2	FRQ
3	CentOwn	3	SPECIALIST
4	FRQ	4	TENURE
5	DA	5	DA
6	OPAQUE	6	OPAQUE

$$W_{j.r} = \ln(1 + \varepsilon_{jr}) \quad (8)$$

Where, $W_{j.r}$ is the Firm-Specific Weekly Return for company j in week r, and ε_{jr} is the residual from Equation (7). NCSKEW is a measure of stock price crash risk which is defined as negative skewness coefficient of the firm-specific weekly returns in a fiscal year. In other words, for company j in year t we have:

$$NCSKEW_{j.t} = -\left[n(n-1)^{3/2} \sum_{j.r} W^3\right] / (n-1)(n-2)(\sum_{j.r} W^2)^{3/2} \quad (9)$$

Where, n is the number of weekly stock return observations for company j in the fiscal year t. Higher NCSKEW corresponds to a higher stock price crash risk.

Independent variables

Table 2 shows 17 primary explanatory variables selected as explained in Section 3.3. In order to compare the results with the correlation-based method, we used six high ranked (more suitable) variables as presented in Table 3. It should be mentioned that 15 and 10 factors were extracted from factor analysis and principal component analysis, respectively, which respectively represent about 65% and 52% of variability or variance of the primary variables. In other words, all the primary variables can be summarized in 15 and 10 general factors maintaining about 56% and 52% of the initial information, respectively.

In general, the present study tried to use explanatory relative variables, because previous studies show that

relative valuation models better perform than absolute value models (Back et al., 1996). It can be due to their more comparative ability in relative valuation, which is a common problem with absolute values, as well as higher mathematical accuracy of financial statements (Williamson, 1984); (Back et al., 1996).

Opacity of financial information (OPAQUE)

In this study, the modified Jones model (2002) was used and discretionary accruals were measured cross-sectionally by year and were used as the financial information quality index. This variable is defined as follows:

$$OPAQUE_{i.t} = \text{Abs}(DACC_{i.t-1}) + Abs(DACC_{i.t-2}) + Abs(DACC_{i.t-3}) \quad (10)$$

Where, $\text{OPAQUE}_{\text{i.t}}$ denotes opacity of financial information for company i in the fiscal year t, Abs is absolute value, $\text{DACC}_{\text{i.t}-1}$ represents discretionary accruals for company i at the end of fiscal year t-1, $\text{DACC}_{\text{i.t}-2}$ is discretionary accruals for company i at the end of fiscal year t-2, and $\text{DACC}_{\text{i.t}-3}$ is discretionary accruals for company i at the end of fiscal year t-3.

Executive overconfidence (Overconfident CEO)

The following five factors were used to measure executive overconfidence, in such a way that if at least three of them are equal to one, then the final factor is equal to one too, otherwise it is equal to zero (Bon-Kim et al, 2016):

Overinvestment($OverInv_{i.t}$): Overinvestment, according to Ahmed and Duellman (2012), is the amount of additional investment in assets resulting from a regression of total asset growth on sales growth conducted by industry year (Equation 11). If the residual from the excess investment regression is positive (i.e., the corporation overinvested), we set Overinvested equal to one; otherwise, it is set equal to zero (Bon-Kim et al., 2016).

$$\text{ASSET.GR}_{\text{i.t}} = \beta_0 + \beta_1 \text{SALES.GR}_{\text{i.t}} + \varepsilon_t \quad (11)$$

$\text{ASSET.GR}_{\text{i.t}}$ is asset growth for company i in year t:

$$\left(\frac{ASSETS_t}{ASSETS_{t-1}}\right) - 1$$

$\text{SALES.GR}_{\text{i.t}}$ is sales growth for company i in year t:

$$\left(\frac{\text{SALES}_{\text{t}}}{\text{SALES}_{\text{t}-1}}\right) - 1$$

Net cash flow: It is a dummy variable that is set to one and denotes management overconfidence if the net cash flow for company I in year t is higher than the median net cash flow of the companies in the same industry in the same year. If not, it is set to zero (Bon-Kim et al., 2016).

Debt to equity ratio: This dummy variable is set to one and indicates management overconfidence if firm I's debt to equity ratio in year t exceeds the median debt to equity ratio of companies in the same industry in that year; otherwise, it is set to zero. The proportion of total debt (long-term debt plus short-term debt) to the market value of equity is known as the debt-to-equity ratio (Bon-Kim et al., 2016).

Dividend policy (DIVYLD): It is a dummy variable which is set equal to one if the company did not pay cash dividend, otherwise it is set equal to zero. Return on equity is equal to zero (Bon-Kim et al, 2016).

Capital expenditure ratio ($CAPEX_{I.t}$): According to (Malmendier et al., 2011). and (Malmendier & Tate, 2005), this is a dummy variable which is set equal to one and indicates managerial overconfidence if this ratio for company i in year t is greater than the median capital expenditure ratio of the corresponding industry companies in the same year, otherwise it is set equal to zero. Capital expenditure ratio in year t is calculated by dividing capital expenditures to total assets of company i at the end of year t-1 (Equation 12).

$$\left(\frac{\text{C}}{A}\right)_t = \frac{C.E_t}{TA_{t-1}} \quad (12)$$

Capital expenditure ($C.E_t$): It is the money a company spends to buy, maintain, or improve its productive assets, such as vehicles, commercial and manufacturing buildings, equipment etc. in year t. According to (Baker et al., 2003), it is the difference between the net book value of fixed assets at the beginning and the end of the fiscal year plus betterment. TA_{t-1} denotes total assets at the end of year t-1.

Real earnings management (manipulation of real activities) through abnormal discretionary expenses

According to (Francis et al., 2008), we used (Gunny, 2010) model (Equation 13) to measure abnormal discretionary expenses:

$$\frac{\text{DISX}_{\text{i.t}}}{\text{TA}_{\text{i.t}-1}} = \alpha_0 \left(\frac{1}{\text{TA}_{\text{i.t}-1}}\right) + \beta_1 \left(\frac{\text{S}_{\text{i.t}}}{\text{TA}_{\text{i.t}-1}}\right) + \beta_2 \text{MV}_{\text{i.t}} + \beta_3 \text{Q}_{\text{i.t}} + \beta_4 \left(\frac{\text{INT}_{\text{i.t}}}{\text{TA}_{\text{i.t}-1}}\right) + \beta_5 \frac{\Delta \text{S}_{\text{i.t}}}{\text{TA}_{\text{i.t}-1}} + \beta_6 \left(\frac{\Delta \text{S}_{\text{i.t}}}{\text{TA}_{\text{i.t}-1}} * \text{DD}_{\text{i.t}}\right) + \varepsilon_{\text{it}} \quad (13)$$

Where, $\text{DISX}_{\text{i,t}}$ is discretionary expenses of company i in year t (including administrative, general, sales, research and development, and advertising expenses); $\text{TA}_{\text{i.t}-1}$ is the total assets of company i in year t; $\text{S}_{\text{i.t}}$ denotes sales of company i in year t; $\text{DD}_{\text{i.t}}$ is a virtual variable which is set equal to one if the current year's sales are lower than last year's sales, otherwise it is set equal to zero; $\text{MV}_{\text{i.t}}$ is the log of market value (a measure for company size); INT is internal budget; $\Delta \text{S}_{\text{i.t}}$ is the sales change for company I in year t; $\text{Q}_{\text{i.t}}$: Tobin's Q is the ratio of profit margin to expenses per investment unit and is calculated using (Gunny, 2010) model (Equation 14):

$$\text{Qi} = \frac{\text{MTB+TDEBT}}{\text{TASSET}} \quad (14)$$

Where, Q_i is Tobin's Q ratio, MTB denotes total market value, TDEB is the book value of total debt, and TASSET represents the book value of total assets.

Real earnings management through abnormal production expenses

According to (Francis et al., 2008), we used (Gunny, 2010) model (Equation 13) to measure abnormal production expenses:

$$\frac{PROD_{i.t}}{TA_{i.t-1}} = \alpha_0\left(\frac{1}{TA_{i.t-1}}\right) + +\beta_1 MV_{i.t} + \beta_2 Q_{i.t} + \beta_3\left(\frac{S_{i.t}}{TA_{i.t-1}}\right) + \beta_4\left(\frac{\Delta S_{i.t}}{TA_{i.t-1}} + \beta_5\left(\frac{\Delta S_{t-1}}{TA_{i.t-1}}\right) + \varepsilon_{it}\right) \quad (15)$$

Where, $PROD_{i.t}$ is production expenses for company I in year t (which is equal to cost of products sold plus inventory change), and $\Delta S_{i.t-1}$ denotes sales change for company i in year t-1.

Real earnings management through abnormal cash flow from operating activities

Roychowdhury model (2006) was used to measure abnormal cash flow from operating activities:

$$\frac{CFO_{i.t}}{TA_{i.t-1}} = \alpha_0\left(\frac{1}{TA_{i.t-1}}\right) + \beta_1 MV_{i.t} + \beta_2 Q_{i.t} + \alpha\beta_3\left(\frac{S_{i.t}}{TA_{i.t-1}}\right) + \beta_4\left(\frac{\Delta S_{i.t}}{TA_{i.t-1}}\right) + \varepsilon \quad (16)$$

Where, $CFO_{i.t}$ is the cash flow from operating activities of company i in year t.

Control variables

Following the previous studies, six control variables were used in the present study as shown in Table 4.

Average firm-specific monthly return (RET)

Equations (17), (18) and (19) were used to measure the average firm-specific monthly return for each company in the fiscal year.

$$STD_{i.t} = \sum_{\theta=1}^{\theta=52}\left(\frac{\left(r_{i.\theta} - Ret_{r_{i.\theta}}\right)^2}{N}\right)^{\frac{1}{2}} \quad (17)$$

$$Ret_{r_{i.\theta}} = \sum_{\theta=1}^{\theta=52}\frac{r_{i.t}}{N} \quad (18)$$

$$r_{r_{i.\theta}} = \frac{p_{r_{i.\theta}} - p_{i.\theta-1}) + \text{Cash Benefit} + \text{Stock Award} + \text{Priority}}{r_{i.\theta-1}} \quad (19)$$

Where, $r_{i.\theta}$ is stock return for company j in month θ, $Ret_{i.j}$ is the average monthly return for company j in month t, $p_{i.t}$ is is the price at the end of the month, and $p_{i.\theta-1}$ is the price at the beginning of the month.

4. Results

Descriptive statistics for the study variables

The descriptive statistics for the study variables are presented in Table 5, which also includes descriptive parameters for each variable. These parameters largely consist of measures of dispersion, such as standard deviation, as well as measures of central tendency, such as maximum, minimum, mean, and median. The most significant indicator of central tendency, "mean," depicts the equilibrium point and the geographic center of a distribution. Displaying the center of the data is a useful measurement. For instance, when the data is centered, the mean financial leverage (LEV) is equal to 0.622. Another indicator of central tendency that depicts population status is the median.

As can be seen, the median of company size (SIZE) is equal to 5.44, which indicates that half of the data lies below and above this value. In general, measures of dispersion are used as measures of spread of data about the mean, or the spread in data set. Standard deviation is one of the most important measures of dispersion. The standard deviation is 3.82 and 0.022 for NCSKEW and STD, respectively. This indicates that maximum and minimum dispersion are related to NCSKEW and INTANG, respectively. The mean market-to-book value ratio (MTB) suggests that on average only about 2% of company's market value of equity is reported in statements. This indicates tendency of these companies to conservative accounting. Inflation and not reporting some intangible assets and intellectual property in financial statements is one of the reasons for this low value. Moreover, as shown in Table 5, given that the variables of CEO duality, executive overconfidence, audit firm size, and auditor industry expertise are indicator variables (merely accepting zero and one) they do not have significant descriptive statistics and are not included in (Table 5). In addition, the mean continuous auditor selection (about 6 years) indicates long-term cooperation between companies listed in Tehran Stock Exchange and audit firms. The mean managerial and institutional ownerships in Tehran Stock Exchange were about 75% and 67%, respectively, indicating a high percentage of ownership composition in Stock Exchange. On the other hand, the mean ownership concentration was about 70% indicating low ownership dispersion. Among the control variables, the mean financial leverage was about 62% indicating a high percentage of debt in the capital structure of these companies. The mean return on equity (ROE) indicates that on average, these companies earned about 37% return on 100 RL net assets. The mean financial leverage (LEV) shows that, on average, about 62% of the companies' assets financing is done by debt generation or borrowing. This suggests that companies tend to be financing by debt generation, so they are leveraged. It should be noted that in the present study, all outliers were removed from data set at 1% level to avoid their impact on the results.

Reliability of variables

In the present study, we used the unit root test proposed by Levin, Lin and Chu to test reliability of variables, because in the lack of reliability, for both time series and combined data sets, pseudo regression will occur. The test results shown in (Table 6) indicate reliability of all variables. It should be mentioned that dummy variables were not

Table4. Definition of control variables

Variable	Symbol	Definition
Standard deviation of monthly stock returns	*STD*	Standard deviation of firm-specific monthly return in the fiscal year
Mean of monthly stock returns	*RET*	See Section 3-5-3-1
Company size	*SIZE*	Log of total market value of equity
Return on equity	*ROE*	The ratio of net income to total equity
Market-to-book value	*MTB*	The ratio of market value to book value
Financial leverage	*LEV*	The ratio of total debt to total assets

Table5. Descriptive statistics for the study variables

Variable	Mean	Median	Minimum	Maximum	Standard deviation
NCSKEW	-0.541	-3.528	-4.243	4.243	3.798
InsOwn	0.755	0.84	0.007	0.991	0.256
CentOwn	0.768	0.809	0.064	0.983	0.183
ManOwn	0.678	0.705	0.013	0.959	0.192
OPAQUE	0.389	0.378	0.009	0.741	0.158
BrdIndep	0.590	0.601	0.221	0.691	0.858
FRQ	0.298	0.463	0.077	0.914	0.501
TENURE	5.244	0	1	11	2.769
DISX	0.066	0.062	0.014	0.112	0.119
PROD	0.316	0.275	0.054	0.552	0. 385
CFO	-2.031	2.312	-5.114	4.867	1.117
DA	0.338	0.582	-0.001	1.452	0.205
Herfindahl Index	0.118	0.468	0.001	1	0.256
Tobin's Q	0.876	3.328	-0.328	7.236	0.787
LI	0.198	0.438	-1.374	0.867	0.198
LI_{IA}	0.112	0.349	-1.159	0.856	0.195
CC_SCORE	-0.669	-0.613	-5.148	8.509	0.965
UC_SCORE	0.097	0.165	-3.306	1.827	0.412
STD	0.063	0.059	0.008	0.143	0.022
RET	-0.046	-0.092	-0.324	1.363	0.185
SIZE	5.561	5.458	4.06	7.605	0.778
ROE	0.369	0.368	0.722	0.974	0.266
MTB	2.353	1.733	0.243	9.886	1.814
LEV	0.622	0.645	0.096	0.971	0.175

* Source: Research findings

included in reliability analysis. As can be seen, variables were reliable at 99% confidence level. Therefore, parameters can be estimated without worrying about their pseudonymity. In other words, it can be said that the parameters of mean, variance, and covariance were constant over years, so these companies did not experience structural changes.

Hypothesis testing

The mean values of metrics, such as root-mean-square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R2), are shown in Table 7 for the evaluation of stock price crash risk using all variables in the linear regression method, 17 explanatory variables chosen using correlation-based and relief methods, and the variables extracted with principal component and factor analysis methods.

It is important to note that the median values of these metrics were determined using 10 iterations of 10-fold cross-validation, yielding 100 accurate predictions for each measure.

To test the main and sub-hypotheses, we used ANOVA (or nonparametric Kruskal-Wallis test in the case of not satisfied parametric assumptions) to compare the mean coefficient of determination values obtained from prediction using variables selected with the considered four variable selection and extraction methods (correlation-based, relief, principal component analysis and factor analysis), with the value obtained from linear regression method using all variables.

To compare the performance metrics of the linear regression method when using all variables with the case of using variables selected with the correlation-based, relief, principal component, and factor analysis methods, we used the paired t-test (or the Wilcoxon nonparametric test in the case of unsatisfied parametric assumptions). According to (Table 7), it can be inferred that variable selection and extraction methods have a significant positive effect on the performance of the prediction method if the performance of the prediction method when using selected or extracted variables is better than using all variables and the difference is statistically significant.

Table6. Reliability test results

Variable	Levin, Lin and Chu statistic	Confidence level
NCSKEW	-6960.00	0.000
InsOwn	-581.509	0.000
CentOwn	-220.847	0.000
ManOwn	-2215.47	0.000
OPAQUE	-20.6139	0.000
BrdIndep	-8.916	0.000
FRQ	-66.0077	0.000
TENURE	-25.0386	0.000
DISX	-1962.35	0.000
PROD	-3.27495	0.0005
CFO	-15.463	0.000
DA	-19.547	0.000
Herfindahl Index	-23.104	0.000
Tobin's Q	-56.236	0.000
LI	-75.189	0.000
LI_{IA}	-42.695	0.000
CC_SCORE	-43.316	0.000
UC_SCORE	-24.412	0.000
STD	-26.558	0.000
RET	-22.562	0.000
SIZE	-21.980	0.000
ROE	-111.890	0.000
MTB	-16.235	0.000
LEV	-19.560	0.000

* Source: Research finding

Table 7. The mean performance of measures obtained from linear regression method

Basis	Performance measure	LR(Linear regression)
All variables	RMSE	8.411
	MAPE	0.422
	R^2	0.056
Variables selected with correlation-based method	RMSE	52.502
	MAPE	0.304
	R^2	0.401
Variables selected with relief method	RMSE	56.407
	MAPE	0.203
	R^2	0.195
Variables extracted with principal component analysis	RMSE	56.563
	MAPE	0.257
	R^2	0.103
Variables extracted with factor analysis	RMSE	59.451
	MAPE	0.506
	R^2	0.174

Main hypothesis testing

The main hypothesis aimed to evaluate utility or superiority of using variable reduction methods in comparison with the use of all variables. In this case, we compared the mean coefficient of determination for prediction using all variables to the prediction using selected and extracted variables. According to the results obtained from Kolmogorov-Smirnov and Levene tests to verify normality of data distribution and homogeneity of variances, ANOVA test was used for this comparison.

According to (Table 7) and (Table 8), the mean coefficient of determination for the linear regression using all variables, variables selected with correlation-based and relief methods, and variables extracted with principal component analysis and factor analysis were 0.056, 0.0401, 0.055, 0.103, and 0.174, respectively.

F statistic and the significance level obtained from analysis of variance (3.541 and <0.01) also indicate a significant difference between the use of all variables and the use of selected or extracted variables with variable reduction methods.

Therefore, it can be said that the main hypothesis, i.e. there is a significant difference between the performance of stock price crash risk prediction using

Table8. The results of t-test using linear regression

	All	Corr	R	PCA	FA
All	-	-	-	-	-
Corr	2.562 (<0.001)	-	-	-	-
R	2.105 (<0.001)	2.402 (<0.001)	-	-	-
PCA	1.805 (<0.001)	2.205 (<0.001)	2.109 (<0.001)	-	-
FA	2.240 (<0.001)	2.287 (0.176)	1.603 (<0.001)	2.108 (<0.001)	-

selected or extracted variables and the performance of prediction using all primary variables, is confirmed.

The first sub-hypothesis testing

The average coefficients of determination for predictions made using all variables and the variables recovered using factor analysis were 0.056 and 0.054, respectively, according to (Table 7). The use of primary variables and the use of variables extracted via principal component analysis differ significantly, as indicated by the t statistic and the significance level provided in (Table 8) (2.240 and <0.001, respectively). This offers enough proof to support the first sub-hypothesis.

The second sub-hypothesis testing

The mean coefficients of determination for prediction utilizing all primary factors and the variables recovered using principal component analysis were 0.103 and 0.056 respectively, according to (Table 7). The significance level and t statistic were 1.805 and <0.001, respectively, according to (Table 8). According to these metrics using the variables extracted by principal component analysis as opposed to all primary variables results in a noticeable difference. Thus, the second sub-hypothesis is verified as principal component analysis performed better by using linear regression than utilizing all primary variables.

The third sub-hypothesis testing

The mean coefficients of determination for prediction were 0.056 and 0.401, respectively, by using all primary variables and the variables selected with correlation-based method. According to Table 8, t statistic and the significance level were 2.562 and <0.001, respectively. As a result, it can be said that, in line with the third sub-hypothesis, the use of variables selected with correlation-based method is superior to the use of all primary explanatory variables.

The fourth sub-hypothesis testing

The mean coefficients of determination for prediction using all primary variables and the variables selected with relief method were 0.056 and 0.055, respectively. According to Table 8, t statistic and the significance level were 2.105 and <0.001, respectively. Thus, it can be said that there is a significant difference between using all primary variables and using variables chosen using the relief technique, and that the variables chosen using the relief technique are better at predicting the risk of a stock price drop. The fourth sub-hypothesis is therefore supported.

5. Discussion and Conclusion

The purpose of the current study was to examine the performance and efficacy of variable reduction and extraction approaches in identifying the possibility of future stock price crashes. 80 businesses listed on the Tehran Stock Exchange between 2006 and 2017 were utilized as our sample. 17 extensively utilized primary variables were identified based on the literature. To assess the likelihood of a future stock price crash, we employed evaluation metrics such as the mean absolute magnitude percentage error, root mean square error, and coefficient of determination. With the usage of all key explanatory variables, these metrics were utilized to compare the performance of variable reduction, variable selection, and variable extraction techniques.

The findings showed a substantial difference between the principal explanatory variables and the variables chosen using variable reduction methods (correlation-based and relief approaches) or variable extraction methods (main component analysis and factor analysis). In order to anticipate the likelihood of a future stock price crash, the use of selected or extracted variables performs better than the use of all key explanatory variables.

Combination of some industries and ignorance of some other industries due to statistical constraints, excluding some of the factors affecting stock price crash risk proposed in the literature due to impossibility of computation, and excluding some companies due to lack of access to some required variables in some years were the most important limitations of the present study. Stakeholders, as target audience, can be aware of the effects of variable reduction and component extraction methods on prediction of stock price crash risk and take them into account when deciding to invest in the stock market.

In a study titled "Business Governance and the Danger of Dropping Stock Prices," Panayiotis et al. (2013) examined the connection between corporate governance standards and the risk of declining stock prices. According to them, there is a link between institutional ownership, the proportion of managers who buy comparable equities, and the probability of stock price declines. The belief that institutions and managers acquire shares on behalf of

institutions put management under pressure for immediate results. Furthermore, the likelihood of declining stock prices is directly related to the lack of transparency in financial information. Additionally, they demonstrated that there is an adverse correlation between the proportion of independent directors on the audit committee, the sector's expertise, and the likelihood of declining stock prices (Panayiotis et al., 2013).

Rubin and Zang looked into the expertise of the auditing sector as well as the danger of declining stock prices. They discovered that high-quality auditors had a detrimental and significant link with using a large sample from the year 1990 to 2009 in United States of America. Future stock price declines are possible. It serves as a reminder that competent auditors can immediately lower the danger of declining stock prices to the advantage of investors. Additionally, they demonstrated that the continuity of auditor selection has a negative and significant relationship with the risk of future falls in stock prices. They also demonstrated that the expertise of the auditing industry has an impact on the relationship between the risk of future stock price drops and the transparency of financial information, accounting conservatism, and tax avoidance (Robin & Zhang, 2015). Real earnings management has been shown to significantly increase stock price declines the following year (Francis et al., 2008). According to Bon-Kim et al., (2016), organizations with more conservative accounting standards are less likely to collapse as a result of CEO overconfidence.

In a study titled "Conservative Accounting Quality of Accounting Information and the Risk of Stock Price Fall" (Dimitrios et al., 2014), the researchers looked into the impact of various accounting and auditing features to enhance return reporting in predicting the risk of stock price fall and found a distinction between conditional and unconditional conservatism and the risk of stock price fall. In the future, there will be a bad connection. Additionally, they demonstrated how the degree of unconditional conservatism influences the link between conditional conservatism and the possibility of future stock price declines. However, they demonstrated that audit characteristics do not appear to be associated with the capacity to anticipate the risk of dropping stock prices, although the literature indicates a strong correlation between the delay in income transparency and the risk of falling stock prices.

Callen & Fang (2013) looked into the connection between institutional investors and the decline in the stock price of the company. This study contrasts the monitoring viewpoint with the expropriation viewpoint held by institutional investors. According to the study's findings, there is substantial evidence that institutional owners and future stock price declines are inversely related. This is because institutional investors prevent the company's bad news from building up by keeping an eye on the management (one of the most important reasons for the fall in stock prices).

In their study, Mehrani et al. (2015) looked into how corporate governance practices and the accuracy of financial reporting relate to one another. As a result, by increasing the number of corporate governance variables, the indicators of the quality of profit are improved, among other things. Their findings show that the variables of board member independence and ultimate ownership have a positive and significant relationship with the quality of financial reporting. Financial reporting becomes more accurate.

It is possible to provide a model to predict the factors affecting the stock price crash risk, and respectively social responsibility in the first priority, business strategies in the second priority, macroeconomic variables in the third priority, managerial ability in the fourth priority, political communication in the fifth priority, financial variables in the sixth priority, and information asymmetry in the seventh priority affect the stock price crash risk (Valizadeh et al., 2021).

Firms with influential CEOs are more likely to restate their financial statements, have a lower proportion of negative to positive earnings forecasts, and have a lower ratio of negative to positive phrases in their financial statements. The link between powerful CEOs and increased crash risk is strongest in organizations where CEO wealth is more sensitive to stock prices and where CEOs have poorer general abilities. External monitoring techniques mitigate but do not eradicate the link between powerful founding CEOs and increased crash risk (Al Mamun et al., 2020).
Identifying factors affecting the risk of falling stock prices and preventing untimely market falls Shares are of special importance, because it will directly affect the wealth of investors (Asadi Asadabad et al., 2020).

Research Limitations

Considering the combination of some industries and ignoring some industries due to statistical limitations, excluding some of the factors affecting the risk of falling stock prices raised in the background of the research due to the impossibility of calculating them and removing Due to some companies not having access to some of the required research variables in some years in the current research, it is suggested to conduct more extensive studies in this field at the country level.

Recommendations

Considering that the presence of institutional investors increases the monitoring of managers' performance, it reduces the information asymmetry and finally, with the increase in the ownership percentage of this group of shareholders, the volatility of stock returns is reduced. According to the results of the present study, it is necessary for investors as the main audience of this research, the stock market should be aware of the impact of methods of reducing variables and extracting components to predict the risk of falling stock prices and take them into consideration when making decisions about their investment.

References

Ahmadi R. & Jamali K. (2014), The Effect of Audit Quality on Future Stock Returns of Listed Companies in Tehran Stock Exchange. Accounting and Auditing Review, 20(4): 1-20.

Ahmadpour A. & Zare Bahmanmiri MJ. & Heidari Rostami KA. (2014), The effect of company characteristics on stock price crash risk (Companies listed in Tehran Stock Exchange). Quarterly Journal of Securities and Exchange, 28(1): 26-29.

Ahmed A. & Duellman S. (2012), Managerial overconfidence and accounting conservatism. Journal of Accounting Research, 51(1): 1-30.

Al Mamun M. & Balachandran B. & Duong H. (2020), Powerful CEOs Stock Price Crash Risk. Journal of Corporate Finance, 62(1): 101582-101596.

Andreou P. & Antoniou C. & Horton J. & Louca C. (2016). Corporate governance and firm-specific stock price crashes. European Financial Management, 22(5): 916-956.

Ahmadi R. & Jamali K. (2014), The Effect of Audit Quality on Future Stock Returns of Listed Companies in Tehran Stock Exchange. Accounting and Auditing Review, 20(4): 1-20.

Ahmadpour A. & Zare Bahmanmiri MJ. & Heidari Rostami KA. (2014), The effect of company characteristics on stock price crash risk (Companies listed in Tehran Stock Exchange). Quarterly Journal of Securities and Exchange, 28(1): 26-29.

Ahmed A. & Duellman S. (2012), Managerial overconfidence and accounting conservatism. Journal of Accounting Research, 51(1): 1-30.

Al Mamun M. & Balachandran B. & Duong H. (2020), Powerful CEOs Stock Price Crash Risk. Journal of Corporate Finance, 62(1): 101582-101596.

Andreou P. & Antoniou C. & Horton J. & Louca C. (2016). Corporate governance and firm-specific stock price crashes. European Financial Management, 22(5): 916-956.

Asadi Asadabad R. & Aghajani Pourghazi F. & Simkhah M. (2020), The effect of economic policy uncertainty on stock price crash risk and the power of executive management. Sixth National Conference on Applied Research in Management. Accounting and Healthy Economy in the Bank, Tehran: Stock Exchange and Insurance.

Avellaneda M. & Lee JH. (2010), Statistical arbitrage in the US equities market. Quantitative Finance, 10(7): 761-782.

Back B. & Laitinen T. & Sere K. (1996), Neural Networks and Genetic Algorithms for Bankruptcy Predictions. Expert Systems with Applications, 11(4): 407-413.

Baker M. & Ruback RS. & Wurgler J, (2003), Behavioral Corporate Finance: A Survey. Unpublished working paper. Harvard Business School and NBER, Harvard Business SchoolNYU Stern School, of Business and NBER.

Ball R. (2009), Market and Political/Regulatory Perspectives on the Recent Accounting Scandals. Journal of Accounting Research, 47(2): 277-323.

Ball R. & Kothari SP. & Nikolaev V. (2013), Econometrics of the Basu Asymmetric Timeliness Coefficient and Accounting Conservatism. Journal of Accounting Research, Wiley Blackwell, 51(5): 1071-1097.

Barton J. & Simko PJ. (2002), The balance sheet as an earnings management constraint. The Accounting Review, 13(4): 1-27.

Basu S. (1998), The conservatism principle and the asymmetric timeliness of earnings. Journal of Accounting and Economics, 24(2): 6(1): 15-27.

Bebchuk LA. & Stole L. (1993), Do short-term managerial objectives lead to under- or overinvestment in long-term projects? Journal of Finance, 48(1): 719-729.

Bon-Kim J. & Wang Z. & Zhang L. (2016), CEO Overconfidence and Stock Price Crash Risk. Contemporary Accounting Research, 33(4): 1720-1749.

Bon Kim J. & Zhang, L. (2013). Financial Reporting Opacity and Expected Crash Risk: Evidence from Implied Volatility Smirks. Contemporary Accounting Research, 31(3): 851-875.

Callen JL. & Fang X. (2013), Institutional Investors and Crashes: Monitoring or Expropriation? Journal of Banking and Finance, 16(2): 56-78.

Chavoshi K. & Rastegar M. & Mirzaei M. (2015), The relationship between managers' excessive self-esteem and financing policy selection in the companies listed in Tehran Stock Exchange. Quarterly Journal of Stock Exchange, 3(12): 42-25.

Chen J. & Hong H. & Stein J. (2001), Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. Journal of Financial Economics, 61(1): 345-381.

Cohen DA. & Zarowin P. (2010), Accrual-based and real earnings management activities around seasoned equity offerings. Journal of Accounting and Economics, 50(1): 2-19.

Dechow PM. & Sloan RG. & Sweeney AP. (1995), Detecting earnings management. The Accounting Review, 70(2): 193-225.

Dimitrios V. & Kousenidis Anestis C. & Ladas Christos I. & Negakis G. (2014), Accounting conservatism quality of accounting information and crash risk of stock prices. The JournalofEconomicAsymmetries, 11(1): 120-137.

Etemadi D. (2009), Asymmetric Timeliness Coefficient and the Accounting Conservatism. Journal of Accounting Research, 51(1): 1071-1097.

Fakhari H. & Taheri E. (2009), The relationship between institutional investors and stock return volatility in Tehran Stock Exchange. Journal of Accounting and Auditing Studies, 55(1): 98-85.

Foroughi D. & Mirzaei M. (2012), The effect of conditional accounting conservatism on stock price crash risk in Tehran Stock Exchange. Journal of Accounting Progress of Shiraz University, 63(1): 117-127.

Francis B. & Pandit S. & Li L. (2014), Abnormal real operations, real earnings management, and subsequent crashes in stock prices. Contemporary Accounting Research, 25(1): 109-147.

Francis JAH. & Huang S. & Rajgopal A. & Zang Y. (2008), CEO reputation and earnings quality. Contemporary Accounting Research, 25(1): 109-147.

Givoly D. & Hayn C. (2000), The changing time-series properties of earnings, cash flows and accruals: has financial reporting become more conservative? Journal of Accounting and Economics, 29(1): 287-320.

Gunny KA. (2010), The relation between earnings management using real activities manipulation and future performance: Evidence from Meeting Earnings Benchmarks. Contemp Acc Res, 27(3): 855-888.

Hong H. & Stein JC. (2003), Differences of opinion, short-sales constraints, and market crashes. The Review of Financial Studies, 16(1): 487-525.

Hutton A. & Marcus A. & Tehranian H. (2009), Opaque financial reports, R2, and crash risk. Journal of Financial Economics, 94(1): 67-86.

Hutton A. & Marcus A. & Tehranian H. (2009), Opaque financial reports, R2, and crash risk. Journal of Financial Economics, 94(1): 67-86.

Karamanou I. & Vafeas N. (2005), The Association between Corporate Boards, Audit Committees, and Management Earnings Forecasts: An Empirical Analysis. Journal of Accounting Research, 43(1): 453-486.

Khan M. & Watts RL. (2009), Estimation and empirical properties of a firm-year measure of accounting conservatism. Journal of Accounting and Economics, 48(1): 132-150.

Kothari SP. & Shu S. & Wysocki PD. (2009), Do Managers Withhold Bad News? Journal of Accounting Research, 47(1): 241-276.

LaFond R. & Watts R. (2008), The information role of conservatism. The Accounting Review, 83(1): 447-478.

Malmendier U. & Tate G. & Yan JON. (2011), Overconfidence and early-life experiences: The effect of managerial traits on corporate financial policies. Journal of Finance, 66(1): 1687-1733.

Leuz C. & Nanda D. & Wysocki PD. (2003), Earnings management and investor protection: an international comparison. Journal of financial economics, 69(3): 505-527.

Malmendier U. & Tate G. (2005), CEO overconfidence and corporate investment. Journal of Finance, 60(1): 2661-2700.

McNichols MF. (2002), Discussion of the quality of accruals and earnings: The role of accrual estimation errors. The Accounting Review, 77(1): 61-69.

Mehrani K. & Beig Boshrooye S. & Shahidi Z. (2015), The relationship between corporate governance mechanisms and quality of financial reporting. Journal of Accounting Research, 12(1): 19-1.

Modares A. & Hasarzadeh R. (2008), Quality of financial reporting and investment performance. Quarterly Journal of Stock Exchange, 2(1): 85-116.

Panayiotis CA. & Constantinos A. & Joanne H. & Christodoulos L. (2015), Corporate Governance and Firm-Specific Stock Price Crashes. European Financial Management, 22(5): 1-41.

Panayiotis CA. & Constantinos A. & Joanne H. & Christodoulos L. (2013), Corporate Governance and Firm-Specific Stock Price Crashes.

Rahnama Rudposhti F. & Moradi MR. (2005), Factor analysis of the arbitrage pricing mechanism in Tehran Stock Exchange. Quarterly Journal of Financial Research, 19(1): 65-95.

Robin A. & Zhang H. (2015), Do Industry-Specialist Auditors Influence Stock Price Crash Risk? Journal of Practice and Theory, 34(3): 47-79.

Robin A. & Zhang H. (2015), Do Industry-Specialist Auditors Influence Stock Price Crash Risk? The Accounting Review, 83(6): 1571-1603.

Valizadeh F. & Mohamadzadeh A. & Seighali M. & Torabian M. (2021), Presentation the Model for Predicting the Factors Affecting Stock Price Crash Risk in the Tehran Stock Exchange. Journal of Financial Management Perspective, 11(33): 217-243.

Verdi R. (2006), Financial Reporting Quality and Investment Efficiency. The Accounting Review, 86(4): 1255-1288.

Williamson RW. (1984), Evidence on the Selective Reporting of Financial Ratios. The Accounting Review, 59(2): 296-298.