

## Research Article

# Analyzing the Association between Pattern and Returns Using Goodman–Kruskal Prediction Error Reduction Index ( $\lambda$ )

Parmod Kumar Paul <sup>1</sup>, Om Prakash Mahela <sup>2</sup> and Baseem Khan <sup>3</sup>

<sup>1</sup>Dept. of Mathematics, Indian Institute of Technology, Jodhpur, India

<sup>2</sup>Power System Planning Division, Rajasthan Rajya Vidyut Prasaran Nigam Ltd., Rajasthan, India

<sup>3</sup>Department of Electrical and Computer Engineering, Hawassa University, Awassa, Ethiopia

Correspondence should be addressed to Baseem Khan; [baseem.khan04@gmail.com](mailto:baseem.khan04@gmail.com)

Received 29 October 2021; Revised 6 December 2021; Accepted 21 December 2021; Published 15 January 2022

Academic Editor: Paulo Jorge Silveira Ferreira

Copyright © 2022 Parmod Kumar Paul et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For selecting and interpreting appropriate behaviour of proportion between buy/neutral/sell patterns and high/moderate/low returns, the prediction error reduction index is a very useful tool. It is operationally interpretable in terms of the proportional reduction in error of estimation. We first obtain the buy/sell pattern using an Optimal Band. The analysis of the association between patterns and returns is based on the Goodman–Kruskal prediction error reduction index ( $\lambda$ ). Empirical analysis suggests that the prediction of returns from patterns is more impressive or of less error as compared to the prediction of patterns from returns. We demonstrated the prediction index for Index NIFTY 50, BANK-NIFTY, and NIFTY-IT of NSE (National Stock Exchange), for the period 2010–2020.

## 1. Introduction

In the domain of the stock market, there have been several studies well formulated on the relationship between two financial variables. In this domain, the association or relation is very important. With the help of one known variable, one can predict/estimate other unknown variables. There are a number of literature studies on the association between different kinds of financial variables, like exchange rates, stock prices, returns, volatility, and many more factors. Here, we discussed some of the literature.

*1.1. Related Work.* The relationship between market sentiment index and stock rates of returns in the Brazilian market is explored in [1]. The relation between common stock returns, trading activity, and market value is explored in [2]. There are relationships between strange variables such as Quantile relationships between oil and stock return is presented in [3], which is evidence from emerging and frontier stock markets. Further, the relationship between

music sentiment and stock returns presented in [4]. Moreover, when the market drives investors crazy, the relationship between stock market returns and fatal car accidents is presented in [5]. Firm efficiency and stock returns during the COVID-19 crisis are discussed in [6]. Football sentiment and stock market returns are expounded in [7]. Apart from these, there are many more association between investor sentiments and market returns as in [8]. In [9], authors discussed the relationship between firm size and international content of earnings, while in [10], authors discussed the relationship between transaction cost and small firm effect.

*1.2. Motivation.* In the above literature, there is some gap of relation between patterns and returns. Based on some known pattern of a particular index, we predict the future returns of the index. In most back-testing processes, returns is a known variable for us, and based on this variable, we predict pattern or combination of different patterns that depend on the price series of the index. We

test our pattern by back-testing used in the current market that is combination of pattern and returns. This is the major gap of research on relation between these two variables based on different price series. In this research article, we try to fill this gap. This is our main motivation behind this study.

*1.3. Objective.* In this research paper, we discuss the association between pattern and returns with prediction error reduction index on either side of prediction, pattern from returns and returns from pattern. These two variables play a very important role in the stock market. When developing an investment strategy and selecting index or stocks for our portfolio, the association can be a very helpful tool. There are various algorithms and models available in the literature, for predicting the pattern of financial time series [11–14]. In the modern era, the most trendy pattern prediction technique is artificial intelligence. There is a lot of research analysis of pattern prediction using AI algorithm [15–19]. Each model has its own advantages and limitations associated with it and also shows an error when we execute it. For example, these models have prediction errors while predicting the buy/sell pattern. We can use appropriate preprocessing techniques which can help in reducing the prediction error significantly. We construct such a method by using the Goodman–Kruskal index [20]. Goodman–Kruskal’s lambda has been widely used in applications. Jaroszewicz et al. used the lambda for constructing a minimal classifier for cancer data [21, 22]. Taha and Hadi used it to compare the performance of a new measure of association [23].

Here, we initially constructed a two-dimension contingency table using the count of elements of pattern (buy/neutral/sell) and returns (high/moderate/low). The count of contingency tables depends on the trading/investing way or strategy. There are a number of strategies and trading styles to construct contingency. In this paper, analysis is done using Optimal Band to classify the financial data into patterns and returns; for more details see reference [24]. Some details of Optimal Band and the construction of the contingency table are as follows:

- (i) The construction of Optimal Band is based on the global and local extremums of given financial time series data.
- (ii) This gives a two-dimensional contingency table which consists of two variables, returns and pattern.
- (iii) The table can be constructed in two different ways:
  - (1) Optimal Band divides the pattern data into three categories of sell, neutral, and buy and then uses each of these categories for prediction of returns (high, moderate, and low)
  - (2) Optimal Band divides the returns data into categories of high, moderate, and low and then uses each of these categories for prediction of patterns (sell, neutrals and buy).

With the help of these tables, we find the prediction error of the data with the help of the Goodman–Kruskal index of

prediction proportion ( $\lambda$ ) [25–27]. Based on different values of  $\lambda$ , we decide which way is better: prediction of pattern from returns or prediction of returns from pattern, that is, whether to categorize the pattern data first or to categorize the returns data first.

Here, we proposed a noble method to find the perfect pattern using given returns in back-testing data based on Goodman–Kruskal  $\lambda$ . And then, we use the same pattern to find returns in the live market. We defined the different kinds of pattern and returns as in the research article by Vijay and Paul [24]. We analyze the statistical significance of  $\lambda$ , using errors defined as

$$\lambda = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1} \quad (1)$$

The remaining part of the paper is structured as follows: Section 2 contains the algorithm for construction of a contingency table and uses it to obtain the prediction error reduction index proportion ( $\lambda$ ). The methodology is demonstrated with empirical analysis and their results for Index NIFTY 50, BANK-NIFTY, and NIFTY-IT data from 2010 to 2020 in Sections 3–6. Conclusion of the work for all index data from 2010 to 2020 is provided in Section 7.

## 2. Proposed Methodology

Consider the daily close price time series  $Y_1, Y_2, \dots, Y_n$  of a stock. We define the process of construction of contingency table for two variables, returns and pattern, of financial time series. We divide the data into three categories, sell, neutral, and buy, of patterns using the classifier, Optimal Band [24]. In this section, a brief summary of the construction of an Optimal Band is given; for more details, see [24]:

Step 1: define

$$\begin{aligned} \alpha &= \text{Max}(Y_1, Y_2, \dots, Y_n); \\ \delta &= \text{Min}(Y_1, Y_2, \dots, Y_n); \\ \beta_i &= \text{Max}(Y_i, Y_{i+1}, \dots, Y_{i+5}), \quad 1 \leq i \leq n-5; \\ \gamma_i &= \text{Min}(Y_i, Y_{i+1}, \dots, Y_{i+5}), \quad 1 \leq i \leq n-5. \end{aligned} \quad (2)$$

Step 2: define the linear function as

$$\begin{aligned} f &= a * \alpha + b * \bar{\beta} + c * \bar{\gamma} + d * \delta, \\ \text{where } \bar{\beta} &= \text{mean}(\beta_1, \beta_2, \dots, \beta_{n-5}) \\ &= \frac{1}{n-5} \sum_{i=1}^{n-5} \beta_i, \\ \bar{\gamma} &= \text{mean}(\gamma_1, \gamma_2, \dots, \gamma_{n-5}) \\ &= \frac{1}{n-5} \sum_{i=1}^{n-5} \gamma_i. \end{aligned} \quad (3)$$

Step 3: the following optimization problem is now solved to estimate the parameters a, b, c, and d:

$$\underset{a,b,c,d}{\text{Max}} f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) = a * \alpha + b * \bar{\beta} + c * \bar{\gamma} + d * \delta, \quad (4)$$

such that  $f > 0$ ,  $f < \frac{(\alpha - \bar{\beta})}{2}$ , where  $a, b, c, d \in \mathbb{R}$ .

Step 4: define the bands, for  $1 \leq i \leq n - 5$ ,

$$\begin{aligned} \text{UpperBand}[UB_1] &= \beta_i + f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) \\ \text{UpperBand}[UB_2] &= \beta_i - f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) \\ \text{MiddleLayer}[ML] &= f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) \\ \text{LowerBand}[LB_1] &= \gamma_i + f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) \\ \text{LowerBand}[LB_2] &= \gamma_i - f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) \end{aligned}$$

These bands are used to divide the pattern data into three of its categories, sell, neutral, and buy, as shown in Table 1, that is,

$$\begin{aligned} Y_i \in Y^S, & \quad \text{if } UB_1 \leq Y_i < UB_2, \\ Y_i \in Y^N, & \quad \text{if } UB_2 \leq Y_i < LB_1, \\ Y_i \in Y^B, & \quad \text{if } LB_1 \leq Y_i \leq LB_2, \end{aligned} \quad (5)$$

for  $1 \leq i \leq n - 5$ .

where  $Y^S$ ,  $Y^N$ , and  $Y^B$  are, respectively, the sell, neutral, and buy categories.

Let us define new variables as follows:

$$\begin{aligned} |Y^S| & \text{ is the cardinality of the subsets of sell} \\ |Y^N| & \text{ is the cardinality of the subsets of neutral} \\ |Y^B| & \text{ is the cardinality of the subsets of buy} \end{aligned}$$

Now, we find the prediction error of single variable of pattern [22].

$$\epsilon_1 = 1 - \max \frac{(|Y^S|, |Y^N|, |Y^B|)}{\text{Total}}. \quad (6)$$

We further divide each categories of sell, neutral, and buy of patterns using the same classification technique into subcategories (high, moderate, and low) of returns [24]. We construct Table 2 which is the 2-dimensional contingency table [28].

From the table,  $Y^{SH}$  is the cell corresponding to pattern (sell) and returns (high). The error of prediction for Table 2, where column  $I = C - I$ , Maximum =  $M$ , sum =  $S$ , and Total =  $T$  is given by

$$\epsilon_2 = 1 - S \frac{(M(C - I), M(C - II), M(C - III))}{T}. \quad (7)$$

Goodman–Kruskal prediction error reduction index ( $\lambda$ ): Goodman and Kruskal introduced the idea of proportional reduction in error (PPE) of prediction [25]. The value of  $\lambda$  measures the association of nominal variables for cross tabulations. The value of  $\lambda$  depends upon the proportions of the constructed model. The value of measure of association  $\lambda$  represents the reduction of error of dependent variables (pattern or returns) for a given value of independent variables (returns or pattern). For any given data of a nominal independent variable and dependent variable, it indicates the

TABLE 1: Table of pattern counts.

Pattern	Count	Proportion
Sell	$ Y^S $	$p_1$
Neutral	$ Y^N $	$p_2$
Buy	$ Y^B $	$p_3$

TABLE 2: 2-dimensional frequency table.

Pattern	Returns		
	High	Moderate	Low
Sell	$ Y^{SH} $	$ Y^{SM} $	$ Y^{SL} $
Neutral	$ Y^{NH} $	$ Y^{NM} $	$ Y^{NL} $
Buy	$ Y^{BH} $	$ Y^{BM} $	$ Y^{BL} $

extent to which the model categories and frequencies for each value of the independent variable differs from the overall model category and frequency, denoted by  $\lambda$ . It can be calculated using the following equation:

$$\lambda = \frac{\epsilon_1 - \epsilon_2}{\epsilon_1}, \quad (8)$$

where  $\epsilon_1$  and  $\epsilon_2$  are defined in equations (6) and (7), respectively.

The range of  $\lambda$  varies between 0 (zero association) and 1 (complete association).

### 3. Experiments and Results

In this section, we implement the classification method, Optimal Band, and Goodman–Kruskal prediction error reduction index ( $\lambda$ ), using the daily returns of Index NIFTY 50 for the year 2010. We use Optimal Band to classify the data into three categories of pattern (sell, neutral, and buy). We plot the data with Optimal Band to create the three categories of pattern as shown in Figure 1. For a detailed explanation of Figure 1, please refer to reference [24]. Each of these categories of pattern is further divided into three subcategories of returns (high, moderate, and low) using Optimal Band (Table 3).

Table 4 is the table of counts of different categories of patterns constructed by using the algorithm given in Section 2.

The highest proportion corresponding to sell implies that the best prediction of new instance of Index NIFTY 50 of year 2010 data might fall into the sell category as this category consists of the largest number of items in the observed data set. In this case, we are assuming the sample proportion to be an unbiased reflection of the general population of data set. The estimated probability proportion of correct prediction is  $146/247 = 0.5911$ , and the estimated probability prediction error is

$$\epsilon_1 = 1 - 0.5911 = 0.4089. \quad (9)$$

Now, these categories of pattern are concurrently divided into three further categories of returns (high, moderate, and low).

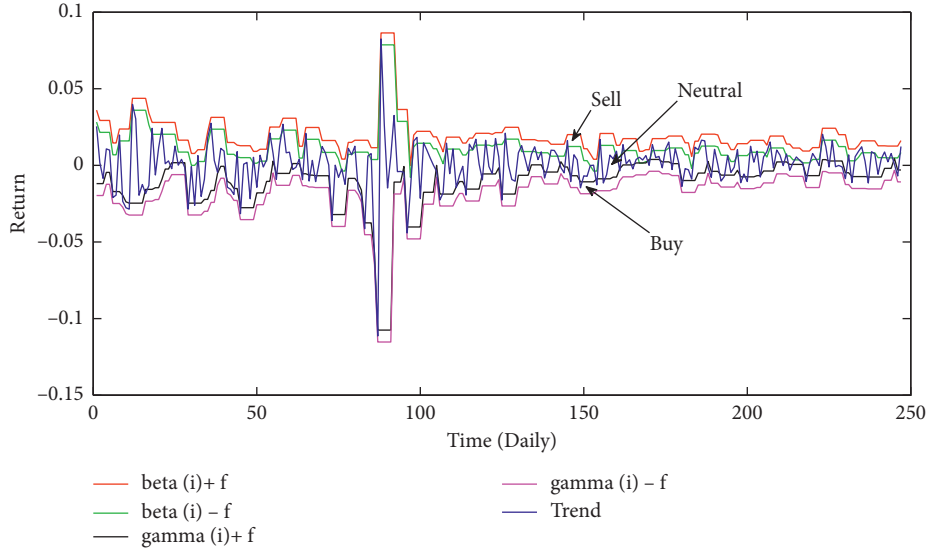


FIGURE 1: Optimal Band classification for NIFTY 50 of 2010.

TABLE 3: Two-variable classification.

Pattern	Returns			Total
	High	Moderate	Low	
Sell	2	115	29	146
Neutral	2	11	4	17
Buy	42	41	1	84
Total	68	113	66	247

TABLE 4: Single-variable classification of pattern of NIFTY 50.

Pattern	Count	Proportion = count/total
Sell	146	0.5911
Neutral	17	0.0692
Buy	84	0.3401

In this case, the prediction error is refined. Table 3 represents that the data set belongs to the sell category of pattern. The best category of returns is moderate. Similarly, if the data set belongs to neutral and buy categories, the respective best prediction of returns is moderate and high. The refined estimated probability of prediction is  $(42 + 115 + 29)/247 = 0.7530$ , and the estimated probability error is

$$\epsilon_2 = 1 - 0.7530 = 0.2470. \quad (10)$$

The probability of prediction error is  $\epsilon_1 = 0.4089$ , as the association between pattern and returns is not established. Once the association is established, the error reduces to  $\epsilon_2 = 0.2470$ . The Goodman-Kruskal prediction error index gives the measure of proportion by which the prediction error is reduced in aforementioned situations [27]. The following equation gives the value of lambda ( $\lambda_1$ ) for the case of predicting returns from pattern:

$$\lambda_1 = \frac{(0.4089 - 0.2470)}{0.4089} = 0.3960. \quad (11)$$

In equation (11), lambda is asymmetric in nature [25]. We turn things around so as to make categorical predictions of pattern from returns.

Our best bet in the absence of information about pattern would be moderate, due to the returns category with the largest number of instances (see Table 5). The initial estimated probability of error in this case would be  $1 - (155/247) = 0.3723$ . Once we factor the relationship between pattern and returns, we could refine the guesses by predicting low when data are sell category; moderate when data are neutral category; and high when data are buy category. The estimated probability of correct prediction would now be  $(43 + 104 + 36)/247 = 0.7409$  as shown in Table 6, the estimated probability of error would be  $1 - 0.7409 = 0.2591$ , and the proportionate reduction in prediction error lambda ( $\lambda_2$ ) is

$$\lambda_2 = \frac{(0.3723 - 0.2591)}{0.3723} = 0.3040. \quad (12)$$

Now, we extended our analysis for the Index NIFTY 50 from 2010 to 2020 to find the value of  $\lambda$  of returns from pattern and pattern from returns for each year as shown in Table 7. Also, we extend the analysis for other indexes BANK-NIFTY and NIFTY-IT for same period of time 2010–2020 (see Tables 8 and 9).

In Tables 7–9, the values of  $\lambda_1$  and  $\lambda_2$  represent the prediction error reduction index corresponding to Index NIFTY 50, BANK-NIFTY, and NIFTY-IT. Also, Table 10 shows the average value of  $\lambda_1$  and  $\lambda_2$  representing the average prediction error reduction index corresponding to some stocks. These tables have column  $\lambda$  value prediction error reduction index for returns from pattern and pattern from returns. The value of  $\lambda$  is more in case returns from pattern than in pattern from returns. If this factor is more, it means prediction error is going to be reduced and prediction will be more perfect. Reduction indexes minimize the error that occurs during the analysis of data.

TABLE 5: Single-variable classification of returns.

Returns	Count	Proportion = count/total
High	42	0.1700
Moderate	155	0.6275
Low	50	0.2024

TABLE 6: Two-variable classification.

Returns	Pattern			Total
	Sell	Neutral	Buy	
High	2	4	36	42
Moderate	23	104	28	155
Low	43	5	2	50
Total	68	113	66	247

TABLE 7: Prediction error reduction index of Index NIFTY 50 for the period 2010–2020.

Year	Returns from pattern			Pattern from returns		
	$\epsilon_1$	$\epsilon_2$	$\lambda_1$	$\epsilon'_1$	$\epsilon'_2$	$\lambda_2$
2010	0.4089	0.2470	0.3959	0.3723	0.2591	0.3040
2011	0.6327	0.4857	0.2323	0.4875	0.4256	0.1270
2012	0.2058	0.1646	0.2000	0.5621	0.5124	0.0884
2013	0.4204	0.3424	0.1845	0.6214	0.5746	0.0753
2014	0.4315	0.3610	0.1635	0.6321	0.5463	0.1357
2015	0.4089	0.3124	0.2360	0.5214	0.4125	0.2089
2016	0.6327	0.5124	0.1901	0.4365	0.3321	0.2392
2017	0.5648	0.3654	0.3530	0.3625	0.2547	0.2974
2018	0.4875	0.3427	0.2970	0.2154	0.1754	0.1857
2019	0.6584	0.5487	0.1666	0.3654	0.3154	0.1368
2020	0.5487	0.3258	0.4062	0.2614	0.2315	0.1144

TABLE 8: Prediction error reduction index of Index BANK-NIFTY for the period 2010–2020.

Year	Returns from pattern			Pattern from returns		
	$\epsilon_1$	$\epsilon_2$	$\lambda_1$	$\epsilon'_1$	$\epsilon'_2$	$\lambda_2$
2010	0.5264	0.3612	0.3138	0.2356	0.1745	0.2593
2011	0.6541	0.4857	0.2575	0.2457	0.1956	0.2039
2012	0.7845	0.6213	0.2080	0.3564	0.2956	0.1706
2013	0.7541	0.6214	0.1760	0.4521	0.3874	0.1431
2014	0.5413	0.4125	0.2379	0.5478	0.3214	0.4133
2015	0.4521	0.3874	0.1431	0.6321	0.5961	0.0570
2016	0.3241	0.2145	0.3382	0.5647	0.4851	0.1410
2017	0.3874	0.2654	0.3149	0.7453	0.6541	0.1221
2018	0.2341	0.1145	0.5109	0.8451	0.6321	0.2520
2019	0.2874	0.1254	0.5637	0.7412	0.6521	0.1202
2020	0.1245	0.0478	0.6161	0.6214	0.5124	0.1754

#### 4. Recession Periods

A financial crisis is any of a number of scenarios in which certain financial assets lose a significant portion of their nominal value all of a sudden. Numerous financial crises were coupled with banking panics throughout the 19th and early twentieth centuries, and many recessions corresponded with these panics, as illustrated in Figure 2. Stock

TABLE 9: Prediction error reduction index of Index NIFTY-IT for the period 2010–2020.

Year	Returns from pattern			Pattern from returns		
	$\epsilon_1$	$\epsilon_2$	$\lambda_1$	$\epsilon'_1$	$\epsilon'_2$	$\lambda_2$
2010	0.6547	0.5321	0.1873	0.6521	0.6012	0.0781
2011	0.3265	0.2856	0.1253	0.2314	0.2031	0.1223
2012	0.6214	0.5124	0.1754	0.3265	0.2854	0.1259
2013	0.8745	0.6321	0.2772	0.4152	0.3265	0.2136
2014	0.4236	0.3214	0.2413	0.3628	0.2856	0.2128
2015	0.5621	0.4521	0.1957	0.5621	0.5012	0.1083
2016	0.4127	0.3965	0.0993	0.3265	0.3215	0.0153
2017	0.2356	0.1965	0.1660	0.8562	0.7921	0.0749
2018	0.3261	0.2745	0.1582	0.6321	0.5932	0.0615
2019	0.1254	0.1154	0.0797	0.3261	0.1423	0.3973
2020	0.4126	0.3124	0.2429	0.3124	0.2654	0.1504

TABLE 10: Average prediction error reduction index of stocks for the period 2010–2020.

Stocks	Returns from pattern			Pattern from returns		
	$\epsilon_1$	$\epsilon_2$	$\lambda_1$	$\epsilon'_1$	$\epsilon'_2$	$\lambda_2$
HDFC	0.5674	0.4851	0.1450	0.6153	0.5324	0.1347
SBI	0.4365	0.3942	0.0969	0.4158	0.3654	0.1212
REL	0.6124	0.4215	0.3117	0.5621	0.5347	0.0487
ICICI	0.3265	0.2965	0.0919	0.3214	0.2729	0.1509
TCS	0.4851	0.3874	0.2014	0.5128	0.2741	0.4655
INFY	0.6314	0.3214	0.4910	0.4127	0.3941	0.0451
LUPIN	0.7131	0.5236	0.2657	0.5321	0.4752	0.1069
ITC	0.6254	0.4785	0.2349	0.6951	0.6354	0.0859
CIPLA	0.4125	0.2951	0.2846	0.4237	0.3684	0.1305
NTPC	0.3264	0.1452	0.5551	0.5312	0.4012	0.2447
UPL	0.5136	0.2345	0.5434	0.4356	0.3745	0.1403

market collapses and the bursting of other financial bubbles, currency crises, and sovereign defaults are examples of circumstances that are commonly referred to as financial crises. However, there is no agreement, and financial crises of the sort described in the following continue to occur from time to time:

- (i) Banking crisis
- (ii) Currency crisis
- (iii) Speculative bubbles and crashes
- (iv) International financial crisis

Here, we will discuss major financial crises such as the Asian Financial Crisis of 1997 (2 July 1997). This crisis arose as a result of investors fleeing emerging Asian stocks, notably Hong Kong's inflated stock market. Crashes occurred in Thailand, Indonesia, South Korea, the Philippines, and elsewhere, with the mini-crash on October 27, 1997, serving as a high point. The Dot-com bubble burst on March 10, 2000, as a result of a technological bubble burst. The financial crisis of 2007–08 is the third (16 Sep 2008). Failures of large financial institutions in the United States, primarily due to exposure of securities of packaged subprime loans and credit default swaps issued to insure these loans and their issuers, quickly devolved into a global crisis on September 16, 2008,

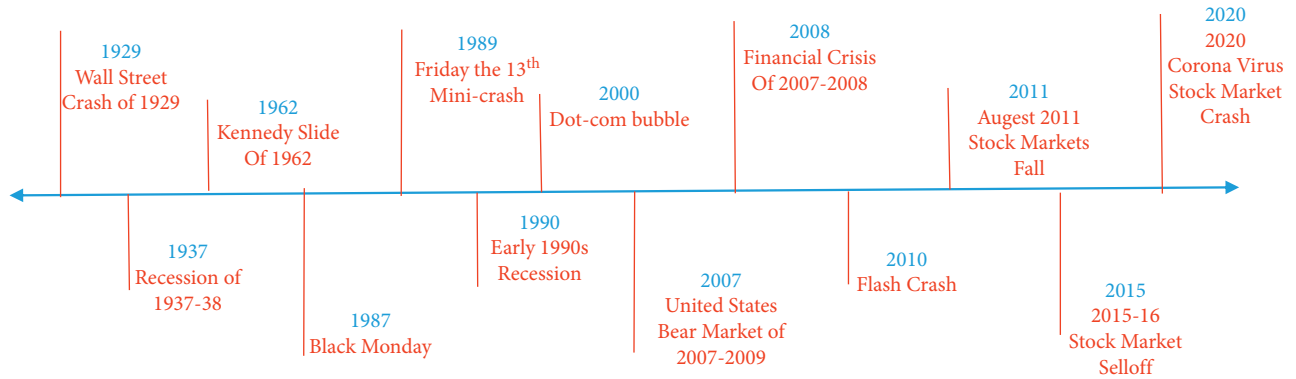


FIGURE 2: Financial crisis.

TABLE 11: Yearly average prediction error reduction index of NSE indices.

Index	NIFTY 50		NIFTY-IT		BANK-NIFTY	
	$\lambda_1$	$\lambda_2$	$\lambda_1$	$\lambda_2$	$\lambda_1$	$\lambda_2$
1997	0.3284	0.2847	0.5847	0.6164	0.3645	0.2847
2008	0.4129	0.3845	0.3456	0.2963	0.3156	0.2999
2016	0.5247	0.4985	0.5129	0.4258	0.3954	0.3387
2020	0.4782	0.4111	0.4786	0.3965	0.4258	0.3859

resulting in a number of bank failures in Europe and sharp drops in the value of equities (stocks) and commodities worldwide. The most recent stock market catastrophe happened in 2020 (24 Feb 2020). This crash was part of a worldwide recession caused by the COVID-19 pandemic.

During these financial crises mentioned above, the mechanism of selecting pattern does not vary. However, pattern selection varies, and it may be biased toward short or long patterns. In back-testing during these financial crises, the error pattern from returns is as shown in Table 11, which is much lower than the error pattern from return. Table 11 shows the yearly average prediction error index for patterns in terms of  $\lambda_1$  and  $\lambda_2$  in four financial crisis periods of 1997, 2000, 2008, and 2020 for NIFTY 50, NIFTY-IT, and BANK-NIFTY. In case of NIFTY-IT, the value of  $\lambda_2$  is higher than  $\lambda_1$  that means errors occur more in selection patterns from returns and all other patterns are selected smooth.

## 5. Comparison with Related Work

Presently, there are lots of research works on association between two or more variables. Here, we define the association between returns and patterns based on back-testing and live trading prediction of returns. In back-testing, we have returns of the data and try to find patterns with the Goodman–Kruskal prediction error index, and in live trading, we have back-tested patterns and predicted returns of future data. Most research works concentrate on prediction of future data pattern without knowing back-testing data pattern accuracy. But here, we try to recommend strong back-testing patterns using Goodman–Kruskal prediction error index.

## 6. Scalability to Economic Significance and Practical Implications

Stock markets are critical to the economy's functioning since they serve as the backbone of a contemporary nation's economic infrastructure. Companies can use stock markets to obtain funds to expand, recruit more qualified employees, and repair or replace equipment. Individuals can also invest in businesses through these platforms.

Stock exchanges provide companies the ability to raise capital to expand their businesses. When a company needs to raise money, they can sell shares of the company to the public. They accomplish this by listing their shares on a stock exchange. Annual reports help investors analyze the performance of companies listed on an exchange.

Investors can purchase shares in public offerings, and the funds collected are deployed by the firm to expand operations, acquire another company, or hire extra employees. All of this contributes to an increase in economic activity, which serves to propel the economy forward. The banking sector, the information technology industry, the pharmaceutical business, and other manufacturing industries all contribute significantly to the country's economic growth. In this study, we look at three NSE (National Stock Exchange) indices, NIFTY 50, BANK-NIFTY, and NIFTY-IT, which cover vital business stocks that are a large part of our economy's growth, as shown in Figure 3. Investors can use our research to determine the optimum pattern for investing in indexes (futures) or stocks. From 2010 to 2020, Table 12 displays the results of pattern selection based on returns in terms of average prediction error reduction index of indices. Table 12 shows that from 2010 to 2020, the error for pattern selection from returns will be decreased.

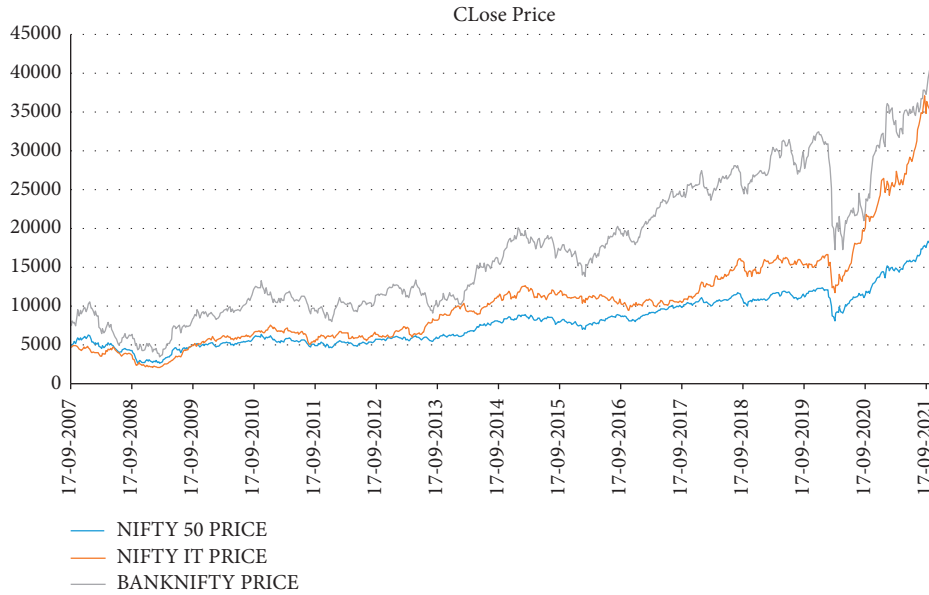


FIGURE 3: NSE indices.

TABLE 12: Average prediction error reduction index of NSE indices for the period 2010–2020.

Stocks	Returns from pattern			Pattern from returns		
	$\epsilon_1$	$\epsilon_2$	$\lambda_1$	$\epsilon'_1$	$\epsilon'_2$	$\lambda_2$
NIFTY 50	0.4875	0.4123	0.1543	0.6945	0.6345	0.0864
BANK-NIFTY	0.5689	0.3956	0.3046	0.5231	0.4482	0.1432
NIFTY-IT	0.4632	0.3348	0.2772	0.4896	0.4156	0.1511

## 7. Conclusions

Here, we conclude from the whole analysis that the prediction of association between two variables is very important, but the way you predict the association is also very important. In economic analysis, any economic factor has two ways, top to bottom and bottom to top. In a similar manner, we try to find the best way to predict the association from one known variable to an unknown variable, which has less prediction error.

In the present analysis, we find prediction analysis error index of patterns of returns from seen data or back-testing data and patterns of returns from the unseen data or future data. The reduction error index of pattern from returns is less which helps to collect better patterns based on given returns that are used in live data to predict returns.

Here, we use the classifier, Optimal Band, and the measure of association ( $\lambda$ ) to find the Goodman–Kruskal prediction error reduction index. It works effectively to find the error in prediction. We did the analysis in two ways to classify the data for association, from returns to patterns, and in a reverse way from pattern to returns, using Optimal Band. We observe that the prediction error reduction index of returns from patterns is more than that of patterns from returns using Goodman–Kruskal index ( $\lambda$ ) for all data sets. Data of Index NIFTY 50, BANK-NIFTY, NIFTY-IT, and stocks for 2010–2020 were used; if the prediction error

reduction index lambda is more, error is less. This lambda that predicts the returns from pattern is better than that which predicts the patterns from returns in all three indices. The constituent (stocks) of the indices also follow the same pattern. The prediction of returns from this pattern is better. In 2014, BANK-NIFTY had a lower prediction error reduction index, followed by NIFTY 50 in 2016 and NIFTY-IT in 2014. Also, we make good selection of patterns in different financial crises for NIFTY 50, BANK-NIFTY, and NIFTY-IT.

## Data Availability

Data will be made available on request to the corresponding author.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] C. E. Yoshinaga and F. H. F. d. Castro Junior, “The relationship between market sentiment index and stock rates of return: a panel data analysis,” *BAR-Brazilian Administration Review*, vol. 9, no. 2, pp. 189–210, 2012.
- [2] C. James and R. O. Edmister, “The relation between common stock returns trading activity and market value,” *The Journal of Finance*, vol. 38, no. 4, pp. 1075–1086, 1983.
- [3] M. Balçilar, R. Demirer, and S. Hammoudeh, “Quantile relationship between oil and stock returns: evidence from emerging and Frontier stock markets,” *Energy Policy*, vol. 134, Article ID 110931, 2019.
- [4] A. Fernandez-Perez, A. Garel, and I. Indriawan, “Music sentiment and stock returns,” *Economics Letters*, vol. 192, Article ID 109260, 2020.
- [5] C. Giulietti, M. Tonin, and M. Vlassopoulos, “When the market drives you crazy: stock market returns and fatal car

- accidents,” *Journal of Health Economics*, vol. 70, Article ID 102245, 2020.
- [6] D. Neukirchen, N. Engelhardt, M. Krause, and P. N. Posch, “Firm efficiency and stock returns during the COVID-19 crisis,” *Finance Research Letters*, vol. 102037, 2021.
- [7] Q.-T. Truong, Q.-N. Tran, W. Bakry, D. N. Nguyen, and S. Al-Mohamad, “Football sentiment and stock market returns: evidence from a Frontier market,” *Journal of Behavioral and Experimental Finance*, vol. 30, Article ID 100472, 2021.
- [8] M. D. Jørgensen and S. Gåsbygg, *Investor Sentiments and Stock Returns*, Master’s thesis, BI Norwegian Business School, Oslo, Norway, 2017.
- [9] M. J. Wohlgemuth, “The relation between firm size and the informational content of earnings,” *Quarterly Journal of Business & Economics*, vol. 27, no. 4, pp. 135–148, 1988.
- [10] P. Schultz, “Transaction costs and the small firm effect,” *Journal of Financial Economics*, vol. 12, no. 1, pp. 81–88, 1983.
- [11] V. Ingle and S. Deshmukh, “Ensemble deep learning framework for stock market data prediction (EDLF-DP),” *Global Transitions Proceedings*, vol. 2, no. 1, pp. 47–66, 2021.
- [12] X. Li, X. Chen, B. Li, T. Singh, and K. Shi, “Predictability of stock market returns: new evidence from developed and developing countries,” *Global Finance Journal*, vol. 100624, 2021.
- [13] Z. Dai, H. Zhu, and J. Kang, “New technical indicators and stock returns predictability,” *International Review of Economics & Finance*, vol. 71, pp. 127–142, 2021.
- [14] Z. Dai, X. Dong, J. Kang, and L. Hong, “Forecasting stock market returns: new technical indicators and two-step economic constraint method,” *The North American Journal of Economics and Finance*, vol. 53, Article ID 101216, 2020.
- [15] D. K. Mohanty, A. K. Parida, and S. S. Khuntia, “Financial market prediction under deep learning framework using auto encoder and kernel extreme learning machine,” *Applied Soft Computing*, vol. 99, Article ID 106898, 2021.
- [16] S. Garcia-Vega, X.-J. Zeng, and J. Keane, “Stock returns prediction using kernel adaptive filtering within a stock market interdependence approach,” *Expert Systems with Applications*, vol. 160, Article ID 113668, 2020.
- [17] D. Kumar, P. Kumar Sarangi, and R. Verma, “A systematic review of stock market prediction using machine learning and statistical techniques,” *Materials Today Proceedings*, vol. 49, no. 8, pp. 3187–3191, 2022.
- [18] C. H. Ellaji, P. Jayasri, C. Pradeepthi, and G. Sreehitha, “AI-based approaches for profitable investment and trading in stock market,” *Materials Today Proceedings*, 2020, In press.
- [19] S. Ruder, *An Overview of Multi-Task Learning in Deep Neural Networks*, Cornell University, Ithaca, NY, USA, 2017.
- [20] L. A. Goodman and W. H. Kruskal, “Measures of association for cross classifications,” *Journal of the American Statistical Association*, vol. 49, no. 268, pp. 732–764, 1954.
- [21] S. Jaroszewicz, D. A. Simovici, W. P. Kuo, and L. Ohno-Machado, “The Goodman-Kruskal coefficient and its applications in genetic diagnosis of cancer,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1095–1102, 2004.
- [22] W. J. Youden, “Index for rating diagnostic tests,” *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.
- [23] A. Taha and A. S. Hadi, “Pair-wise association measures for categorical and mixed data,” *Information Sciences*, vol. 346–347, pp. 73–89, 2016.
- [24] V. Vijay and P. K. Paul, “An optimal band for prediction of buy sell signal and forecasting of States: optimal bank for buy sell signal,” *International Journal of Applied Management Sciences and Engineering*, vol. 2, no. 2, pp. 34–54, 2015.
- [25] L. A. Goodman and W. H. Kruskal, “Measures of association for cross classifications. II: further discussion and references,” *Journal of the American Statistical Association*, vol. 54, no. 285, pp. 123–163, 1959.
- [26] L. A. Goodman and W. H. Kruskal, “Measures of association for cross classifications III: approximate sampling theory,” *Journal of the American Statistical Association*, vol. 58, no. 302, pp. 310–364, 1963.
- [27] L. A. Goodman and W. H. Kruskal, “Measures of association for cross classifications, IV: simplification of asymptotic variances,” *Journal of the American Statistical Association*, vol. 67, no. 338, pp. 415–421, 1972.
- [28] V. Vijay and P. K. Paul, “Analyzing returns and pattern of financial data using log-linear modeling,” *Computational Research*, vol. 4, no. 1, pp. 1–7, 2016.