

## METHODS

# A survival model for wilful default prediction–Bayesian approach

**Arvind Shrivastava and Nitin Kumar\***

Reserve Bank of India, India

**\*Correspondence:**Nitin Kumar,  
nitin\_005us@yahoo.com**Received:** 26 April 2022; **Accepted:** 20 May 2022; **Published:** 30 May 2022

This study develops an insolvency model to predict the possible wilful non-payment of debt obligations that turn into bad assets. This article reveals that financially weak firms have been in deep financial distress somewhere between 2 and 3 years prior to their declaration as a wilful defaulter by the initial credit institution and its reporting on the same to the credit information companies. The Cox proportional hazards model (PHM) has been employed, which is a well-known and profusely applied approach not just in medical science but also in forecasting firm bankruptcy. This widely recognized model has been utilized to estimate the effects of different covariates influencing the times-to-event data. The application of Bayesian methods has the benefit in dealing with censored data in small sample over frequentists' approach. Herein, Bayesian survival framework is applied incorporating normal priors that generally perform better than the traditional likelihood estimation to forecast wilful default. Subsequently, Markov chain Monte Carlo (MCMC) sampling enables to provide the Bayesian estimator. In the Bayesian structure, the survival model is used with the help of hazard function. The gamma distribution is selected as the prior for the standard hazard equation in Cox PHM. To solve for posterior distribution, the Metropolis Hastings scheme is followed, by avoiding solving complicated equations with OpenBugs platform.

**Keywords:** predictions, wilful default, Bayesian approach, survival model, hazard rate

## 1. Introduction

The of the financial system is a crucial ingredient of the economic stability, in view of the special function performed by the finance in the economy. Although the weaknesses in the economic activity does affect the strength of the financial institutions (FIs), extraneous factors accompanying the economic weaknesses may exacerbate the problems of the FIs and such factors may be beyond the remedy by the normal measures available at the disposal of financial sector. Default is one such malice which affects the financial system, and default through the credit market is one that can be more easily camouflaged and thus is more prevalent. Bankruptcies resulting from wilful non-payment of dues, in which debtors have the ability to pay but do not have the motivation to do so, are not limited to financially unhealthy businesses. Therefore, development of models that predict those borrower accounts, which show signs of incipient

wilful default, will be helpful to the banks. The present study attempts to identify those variables as well as a robust methodology, which will help discern the wilful defaulters at an early stage.

As per the Reserve Bank of India (1, 2) norms, a default may be considered wilful when a borrower has reneged in fulfilling its repayment commitments to the creditor and if any of the following occurrences is recognized:

- (a) The debtor had the ability to make the repayments.
- (b) It had not utilized the finance for the purpose lent but has diverted the same for other purposes.
- (c) The borrower diverted the funds elsewhere and did not allocate it for the stated objective.
- (d) The debtor had sold or withdrawn the movable security or immovable asset that was kept as collateral for obtaining the credit without any information to creditor.

These defaults trigger penal action on part of the lenders and also filing of court cases for recovery; information on such defaults leave the confines of bankers' files and get into public domain. At the same time, the financial statements also reflect the results of the business operations and help in discerning the successful companies from those in default. The literature is rich in identifying corporate distress and probability of defaults, but most of such studies have focused on default or bankruptcy *per se*. These studies do not differentiate between (1) the default that has occurred in the course of business operations, due to either inherent or external factors, and (2) the default on account of diversion of funds or misuse of finance.

The present article is an extension of the work by Karthik et al. (3), by using an alternative Bayesian semiparametric proportional hazards model (PHM) approach and arrived at a more robust method of identifying the possible wilful defaulters. In this modeling exercise, we illustrate how the default prediction can be made easy and also show how the model performs. The arrangement of remaining article is narrated as follows. The relevance of introducing covariates in the prediction modeling is described in section "Data set and financial variables." Subsequently, Bayesian framework in Cox system of equations is developed, which is chosen as the reference hazard model. A demonstration of real data application of Bayesian survival model for wilful default prediction employing OpenBugs package is depicted in section "Analysis of the independent variables." Finally, a summary of the findings is provided, encapsulating the pertinent covariates and scope of further research.

The prevalence of intentional default has been emphasized in several research when the debtor states lower monetary worth in contrast to its actual economic status or when the abnormality in clearing its dues is influenced by observing such occurrences in general, also known as negative demonstration effect. According to a poll conducted by Ernst and Young (4), utilization of funds for undisclosed activities through deceptive techniques by entrepreneurs is one of the major reasons for bad asset quality situation in India. Distress prediction models go back to 1960s with studies by Beaver (5) and Altman (6), who have used financial ratios, in the area of liquidity, liability, and expenses. As the literature relating to defaults in general and more specific wilful or strategic defaults in particular is discussed in Karthik et al. (3), the same is not repeated in this study.

## 1.1. Survival methodology

Since long back in 17th century, the study of survival approach has drawn ample attention of researchers, leading to beginning of disciplines like actuarial science and demography. Survival data, as the phrase implies, are associated with lifetimes or, more broadly, focus on waiting period since the occurrence of an event like deterioration

of financial health of a firm, start of job of an individual, or birth of a child to the final incidence of interest like business insolvency, superannuation, or death. In 1950s, a key development happened in this research area. The seminar work of Kaplan and Meier (7) in refinement of survival theory is an important landmark due to which it is commonly used as reference by researchers. Their work provided new insights in the underpinnings of survival literature and hints on further work in this area. The impact of independent variates was initially deliberated by Cox (8) with the advent of PHM, which is a widely cited work in this theory. In demography, medical sciences, particularly in clinical trials, a phenomenon that is generally observed is that certain subjects do not die even on the completion of experiment. This leads to censoring of sample at the right side. To create standard presumptions, almost the survival that a person who is censored is at the same hazard of consequent disappointment as those who are still lively and uncensored. In the hazard structure, the peril set at any time point where the event of interest has not occurred for the units may be acceptable for the whole lot for the entire period. Such case of the censoring process is known as non-informative. The main objective in hazard theory is to understand the relationship of elapsed interval of event with exogenous variables. The estimation of the survival time distribution is the main task of this analysis. The Cox (8) formulation of PHM included period elapsed till death for the firm or person who has faced that incidence to establish the data for hazard foundation. In case the framework emanates from a known distribution, then the formulation of regression analysis follows standard hazard structure. Although, such a parametric model is troublesome to apply without earlier information of the form of the survival equations. Generating PHM from classical perspective, Kalbfleisch and Prentice (9) research comprehensively the Cox (8, 10) partial likelihood approach to estimate the unknown parameters of the hazard rate. Ibrahim et al. (11) elaborate more expansively the progress in the field of Bayesian survival. The partially parametric Bayesian tactics to hazard framework using distribution like Dirichlet, gamma, and beta employing baseline formulation and period elapsed as independent variables are illustrated. Bayesian approach was not common in survival analysis due to complex mathematical formulations of the posterior distribution under censoring that is extremely difficult to obtain directly. The investigation and headway in Gibbs testing numerical calculations, which permit getting tests from the back dissemination, has persuaded the utilization of Bayesian strategies in survival examination. The Bayesian modeling of one or multiple stages or progressive is presently conceivable utilizing reenactment strategies. The Gibbs sampler is one of the strategies of Markov chain Monte Carlo (MCMC) testing calculation which is broadly acknowledged and tested. Markov chain Monte Carlo (MCMC) procedures have presently ended up most well-known device in Bayesian

examination. For the observational workout, this chapter uses the openly accessible OpenBugs computer program for complex measurable models utilizing MCMC methods.

The significance of this thought at this point is exceptionally significant for credit hazard administration of the banking sector in India. Nowadays, the mounting terrible credits have changed the whole hazard scene of the credit field. The initial analysis has uncovered that the full value of suit-filed wilful default is 12.40% of the overall net bad resources as at the conclusion of March 2016. The significant increase within the quantum of bad asset quality has not as it were influenced the budgetary well-being of these entities but has led to severe consequences for the overall nation also. With this foundation, forecast of the wilful defaulters, i.e., a borrower who has the ability to clear off their debt, however, they renege, is of vital significance. These non-payers have essentially not utilized the funds for which it was designated for or maybe redirected or siphoned off the reserves. Consequently, there is a greater need to identify such borrowers beforehand and prepare a list detailing their names and their credit history. Thus, the aim of this research is to construct structure to foresee the likelihood of wilful default in Indian enterprises by employing combination of their publicly available operational variables and classifying them with that of the operationally viable entities. This would relieve the credit default by regressing data on the probability of wilful defaults, enable location of such defaults at the beginning, and reduce the possible harm done related to it. Considering the corporate writings like Giroud et al. (12), exogenous factors to survey the budgetary capacity of firms have been utilized. This is a process to bunch the defaulters into monetarily compelled (non-strategic) evaders and fiscally controlled (strategic) avoiders. In any case, there is exceptionally scant work that employs information from private advances for investigating or anticipating key defaults. Asimakopoulos et al. (13) have deduced that a few debtors may discover it financially more appealing not to repay their advances or renegotiate the advances on better terms and conditions for utilization of surplus amount for other pursuits or thrift. This leads to digression of the amount for unintended objectives. Moreover, the likely factors affecting the behavior of Greek companies by exploring the likelihood of wilful non-repayment to an array of companies' features like age, earnings ratios, liquidity, and collateral are performed. This research establishes that there exists direct impact of credit outstanding and macroeconomic insecurity on wilful default whereas an inverse connection is exhibited with security amount. Jayadev (14) performs empirical analysis that indicates the importance of monetary parameters in forecasting the bankruptcy of firms. Ernst and Young (4) co-authored an analytical study that focused on private-sector creditors that have faulted on more than one occasion as the essential measure behind defaulting on bank credits but intermittent-free reviews on borrowers have

uncovered redirection of stores or wilful default driving to stretch circumstances.

## 1.2. Proportional hazard model of Cox

To classify the default or success of any entity, PHM conceptualized by Cox is a widely prevalent theory gauging the risk and survival. In the event that arbitrary variable  $T$  shows the time to death of a firm as being irregular, variable  $T$  is ceaseless and having an affiliation with survival or risk time  $t$  of a firm taking after certain fundamental conveyance work which indicates the likelihood of survival of a specific firm fair sometime recently it comes up short. The equation of the distribution is as follows:

$$F(t) = P(T < t) = \int f(s)ds \quad (1)$$

The survivor function  $S(t)$  is defined as the probability that a firm will survive longer or up to time  $t$  and is given by,

$$S(t) = P(T \geq t) = 1 - F(t) \quad (2)$$

Accordingly, the hazard function is defined as the rate of death of a firm. Now,  $P(t \leq T < t + \Delta t | T \geq t)$  is the probability of death during the interval  $t$  and  $t + \Delta t$  for a random parameter whose hazard function  $\lambda(t)$  is obtained as follows:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (3)$$

Generally, there are four fundamental concepts in survival analysis: length, censoring, risk rate, and survival work. The point of commencement of the death process to the time an event occurred or the end of the period under study whichever first occur is known as duration. The latter, called observation, is right censored. The instantaneous failure rate at time  $t$  given that the individual is still alive at the time just before  $t$  is defined as hazard rate. In any case, the survival equation portrays the likelihood of surviving the person after the desired time. Bayesian theory application provides us,

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)} \quad (4)$$

$$\lambda(t) = \frac{f(t)}{S(t)}$$

Here  $f(t)$  is the probability density function of a random variable  $t$ . The survival function  $S(t)$  can be formulated as defined hereunder.

$$S(t) = \exp\{-H(t)\} \quad (5)$$

Now, the cumulative hazard function is given as  $H(t) = \int_0^t h(u) du$ . This is associated with survival equation as represented follows:

$$H(t) = -\log\{S(t)\} \quad (6)$$

Survival data induction can decide how a few of the informative factors decide the shape of the risk curve and how much is the probably value of hazard rate for a specific company. The Cox corresponding risk depiction could be a semiparametric survival examination strategy that investigation of tallying prepares information, i.e., counting survival/failure occasion information, is ordinarily based on the displaying of the concentrated. Classifying time into many small intervals, say, with interval length equal to  $\Delta t$ , where  $\Delta t$  is infinitesimal. Let  $T$  is absolutely continuous, one looks at those who have survived up to sometime  $t$  and considers the calculating the probability of an event happening over some finite time interval  $(t, t + \Delta t)$ . Now, the baseline hazard rate  $\lambda_0(t)$ , which is common to all individual, is defined as the following limit:

$$\lambda_0(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (7)$$

$\lambda_0(t)$  is the instantaneous or sudden occurrence of the event rate of an individual or business, given that the firm has still survived until time  $t$ . In particular,  $\lambda_0(t) \Delta t$  is the approximate probability of failure of the individual in the time interval  $(t, t + \Delta t)$ , given that it survives up to time  $t$ . The hazard rate is quite difficult to estimate based on an arbitrary function of time. The baseline cumulative hazard rate defined as,

$$\Lambda_0(t) = \int_0^t \lambda_0(s) ds \quad (8)$$

$\Lambda_0(t)$ , as per Nelson (1969), is a visual methodology to attain engineering data in the form of the survival distribution in reliability analysis. Altshuler (1970) and Aalen (1972) have also mentioned this estimator albeit independently. The literature about advancement within the zone of therapeutic science especially clinical trials within the early 1950s and 1960s led to the much more consideration by analysts and a major breakthrough in this heading was that the Cox relative risks show distributed in 1972 (8). The relapse investigation of survival information was conceivable (8). Particularly, the Cox structure demonstrates the perils (also called the chance work or concentrated work) for a subject  $i$  with covariates  $Z_i = Z_{i1}, Z_{i2}, \dots, Z_{ip}$  as follows:

$$\lambda(t|Z) = \lambda(t|Z = 0) \exp(\beta'Z) = \lambda_0(t) \exp(\beta'Z) \quad (9)$$

Therefore, the hazard rate is the product of the unknown standard hazard rate  $\lambda_0(t)$ , which is the distribution-free portion and the exponential function of the unknown

regression coefficient  $\beta = \sum_j \beta_j Z_{ij}$ , i.e., the distributional function. Here in the Cox PHM, it is assumed that both  $\beta$  and  $Z$  are constant over time  $t$ . Therefore,  $\beta$  is  $p$ -dimensional vector of regression coefficients.  $\int_0^T \lambda_0(t) dt < \infty$  where  $T$  denotes the time point where the study is terminated or closed. The hazard function  $\lambda(t|Z)$  for an individual with covariate vector equal to zero is  $\lambda_0(t)$ , which is the hazard function in the absence of covariates; that is why it is known as typical hazard function, essentially means that at the previous an individual or firm has to die or fail ruled by standard hazard function. Here a vital supposition is that comparative risks are unchanging over the period of study. Let us assume that there exists a single covariate. In such a condition, default is defined as,

$$\frac{\lambda(t|Z_1+1)}{\lambda(t|Z_1)} = \exp(\beta) \quad (10)$$

The relative hazard is invariant of initial point as only the standard intensity depicts association with time. In case of more than one explanatory variable in the equation, the outcome remains the same in case both the entities are compared using same set of variables. This implies that the chance of default of a company is dependent on the operational characteristics of a typical firm. The acknowledgment of Cox procedure is an assortment of survival issues and with covariates is presently an incredible accomplishment in the academic sphere. Cox (8) derived an alternate form for null likelihood that was subsequently utilized as partial likelihood in 1975 to estimate  $\beta$ . The primary objective of Cox formulation is the derivation of intensity that involves enumerating counting process information and hazard rate. Andersen and Gill (1982) examined in detail the numeration procedure and martingale constructs including the proofs for a large sample estimator in such a formulation.

### 1.3. Bayesian methodology

Assuming  $n$  firms for investigations, for firm  $i$ ,  $i = 1, 2, \dots, n$ ,  $I_i(t)$  is a prominent approach for counting process for the available exogenous vector  $Z_i = Z_{i1}, Z_{i2}, \dots, Z_{ip}$  and  $Y_{it}$  is described as the risk measure, i.e., the set of subjects still at risk at the time,  $T_i$ , of death for company  $i$  (i.e., alive and uncensored at time point just before time  $t$ ). It is assumed that  $N_i(t)$  is the count random variable of number of events occurred that resulted in the interval  $[0, t]$ . This procedure remains stable and nearly nil for survival and is recorded as unity for death. The random variate defined as  $\{N_i(t), t = 0\}$  is termed as the counting parameter.

Suppose that companies were followed to be fail or censored in a period under study. In this regard, the sample is given as  $\{D = N_i(t), Y_i(t), Z_i; i = 1, 2, \dots, n\}$  and accordingly, the unidentified regression terms  $\beta$  and the introductory hazard  $\lambda_0(t)$  to be derived. Post-censoring of

non-informative priors, the likelihood is found and obtained as follows:

$$L_i(D | \beta, \Lambda_i(t)) = \exp \left\{ \int_{t=0}^t I_i(t) dt \right\} \prod_{t=0}^{dN_i(t)} [I_i(t)]^{dN_i(t)}; \quad i = 1, 2, \dots, n \quad (11)$$

Subsequently, joint likelihood of the available information domain  $D$  is found as follows:

$$L(D | \beta, \Lambda(t)) = \prod_{i=1}^n L_i(D | \beta, \Lambda_i(t)) \quad (12)$$

Now,  $dN(t)$  represents a trivial increase of  $N(t)$  during the period  $(t, t+dt)$ .  $N(t)$  and  $dN(t)$  are assumed to be unity if the conglomerates face insolvency during the  $(0, t)$  and  $(t, t+dt)$ , respectively, and 0 otherwise. For truncation wherein researcher is unaware about the characteristics of data, the procedure is taken to be distributed as Poisson. Accordingly, the number of events happened, i.e.,  $dN(t)$  is taken to be following Poisson function with mean  $I(t) dt$ . So,

$$dN(t) \sim \text{Poisson}(I(t) dt) \quad (13)$$

Let the sample available is  $D$ , in Bayesian approach the primary concern relates to getting the functional characteristics of the posterior functional formulation. The consolidated posterior equation of the Cox PHM by applying Bayes hypothesis is obtained as follows:

$$P(\beta, \Lambda_0(t)|D) \propto L(D | \beta, \Lambda_0(t)) P(\Lambda_0(t)) P(\beta) \quad (14)$$

It may be noted that  $P(\beta, \Lambda_0(t)|D)$  and  $L(D | \beta, \Lambda_0(t))$  symbolize the expression for the final equation formulation of  $\beta, \Lambda_0(t)$ . The information set  $D$  and likelihood expression are stated as above.  $P(\Lambda_0(t))$  and  $P(\beta)$  portray the equations for anterior distribution and standard survival of the regression analysis, respectively.

## 2. Data set and financial variables

The present study uses the data submitted by lenders to the credit information companies, where the bank / FI has filed legal suit against the borrowers. The information is obtained from TransUnion CIBIL Limited, (CIBIL) [formerly, Credit Information Bureau (India) Limited]. This information set is compiled based on the reporting carried out by financial entities and is available freely online. Public and private limited companies, which are part of these data, and where data on their financial statements are also available from a publicly available source, namely, Prowess database, were taken as sample. Along with defaulting companies, data on non-defaulting companies have been taken from Prowess database. In all, about data of 368 companies (144 wilful defaulters and 224 non-defaulters) over the span of 15 years starting from 2002 to 2016 have been found suitable for the study after cleaning.

## 2.1. Subsampling strategy

For the purpose of estimating a model's ability to predict the default and non-default firms, the whole selected data set has been divided into two parts: training and testing samples. The selection of companies in training data set is fully based on randomization using Oracle software to overcome the effect of subjective biasness. The diverse system estimation procedure and the workout will be carried out on training information set and the testing information set will be utilized to find the execution of the analysis independently. The trial information set is roughly 25% of the estimate of the starting information set with comparable extents of healthy and fizzled businesses. Out of this sample, 290 companies [110 WDC (Wilfully defaulted companies) and 180 NDC (non-defaulted companies) are selected as training sample and 78 (34 WDC and 44 NDC) as testing sample]. Recognition and announcement of a debtor as "wilful" by a financial entity involves efforts, i.e., around 1–2 years. The reporting of such event is done with a lag. For arranging the most excellent observational tactics, it is imperative to recognize the factors by employing the relationship investigation. If high relationship is recognized, we prioritize the foremost commonly utilized and best performing proportions within the writings. After carrying out preliminary empirical analysis, the following variables were identified as candidates for indicating the wilful default: (1) firm age; (2) return on equity (ROE); (3) return on asset (ROA); (4) sales-to-capital ratio; (5) leverage, i.e., debt to equity; (6) cash flows of group enterprises; (7) financial and investment activities net cash flows; (8) investment in subsidiary companies; and (9) lending to subsidiary companies. As desired by the survival setup, firms categorized as deliberate non-payer are set as "1" and healthy firms as "0" along with the period since the failure has occurred is compiled.

## 2.2. Discussion on indicators

**Age of companies:** Higher the age of companies, higher the chance to survive in competitive market.

**Debt equity ratio:** Debt/equity ( $D/E$ ) ratio is calculated by dividing a company's total debt by its stockholders' equity. Debt ratio is used to measure a company's financial leverage. The  $D/E$  ratio indicates how much debt a company is using to finance its assets relative to the value of shareholders' equity.

**Sales to capital:** Sales to paid-up capital indicate how best the capital is utilized for revenue generation.

**Return on assets:** ROA is a financial ratio that shows the percentage of profit a company earns in relation to its overall resources. It is commonly defined as net income divided by total assets. Net income is derived from the income statement of the company and is the profit after taxes.

**TABLE 1** | Summary statistics of selected financial indicators for NDC (“0”) and WDC (“1”) companies: training sample.

(WDC = 1), (NDC = 0)		Group statistics			
		N	Mean	Std. deviation	Std. error mean
Company age	0	4,226	1.572	0.224	0.003
	1	110	1.428	0.155	0.015
Debt-to-equity ratio	0	4,226	1.052	2.43	0.037
	1	110	4.667	8.898	0.848
Retained earning asset ratio	0	4,226	0.422	0.216	0.003
	1	110	0.087	0.445	0.042
Sales to paid-up capital ratio	0	4,226	47.232	60.839	0.936
	1	110	11.499	17.285	1.648
Return on asset	0	4,226	6.65	8.247	0.127
	1	110	-18.011	19.882	1.896
Directors_renumeration_Employee salary ratio	0	4,226	0.048	0.143	0.002
	1	110	0.052	0.086	0.008
Net cash flow from operating activity to asset ratio	0	4,226	0.075	0.106	0.002
	1	110	-0.01	0.105	0.010
Net cash flow from financing activity to asset ratio	0	4,226	0.008	0.133	0.002
	1	110	0.024	0.108	0.010
Cash_inflow_outflow_due_to_loans_to_subs_or_group_companies_ASSET_Ratio	0	4,226	-0.002	0.019	0.000
	1	110	0.001	0.01	0.001
Net_cash_inflow_or_outflow_from_investment_activities_SALES_Ratio	0	4,226	-0.107	0.361	0.006
	1	110	-0.434	2.611	0.249
Cash_inflow_or_outflow_due_to_loans_to_subs_or_group_companies_SALES_Ratio	0	4,226	-0.003	0.053	0.001
	1	110	0.008	0.106	0.010
Cash_outflow_due_to_repayment_of_total_borrowings_SALES_RATIO	0	4,226	-0.09	0.292	0.004
	1	110	-0.106	0.329	0.031
Loans to subsidiary GP of companies to asset ratio	0	4,226	0.003	0.017	0.000
	1	110	0	0.002	0.000
Investment in subsidiary GP of companies to total investment	0	4,226	0.049	0.136	0.002
	1	110	0.055	0.133	0.013

**TABLE 2** | Cox hazard model result: training sample.

Covariates	Coeffi.	SE	P-Value
Company age	-3.645	0.251	0.000
Debt-to-equity ratio	0.053	0.008	0.000
Sales to paid-up capital ratio	-0.012	0.003	0.000
Return on asset	-0.051	0.003	0.000
Directors_renumeration_Employee salary ratio	0.892	0.078	0.000
Net cash flow from operating activity to asset ratio	-1.299	0.194	0.000
Net_cash_inflow_or_outflow_from_investment_activities_SALES_Ratio	-0.128	0.037	0.001
Loans to subsidiary GP of companies to asset ratio	-7.470	7.654	0.330
Investment in subsidiary GP of companies to total investment	1.424	0.641	0.030

**Directors’ remuneration as compensation to employees:** Directors’ remuneration as compensation to employees is the payment or compensation received for their services or employment and includes salary, bonuses, and any other economic benefits they receive during employment. However, this ratio indicates willingness of the

management to share the revenue earned from the business with the employees.

**Loans to subsidiary or group companies to asset ratio:** This indicates the amount of loan extended by the parent company to its subsidiaries or group companies. Higher loans of this nature in a distressed company indicate

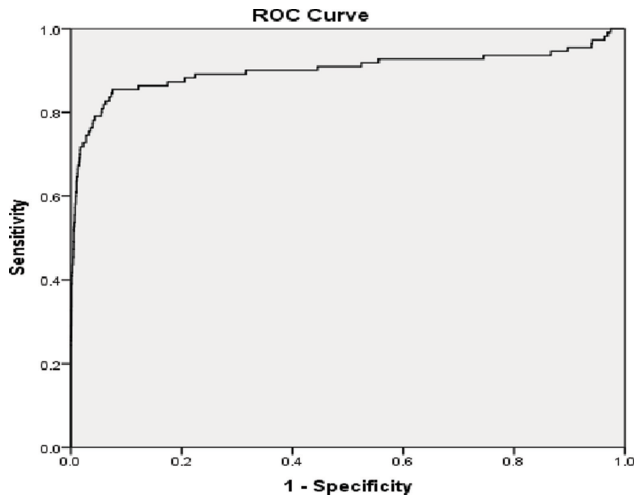


FIGURE 1 |

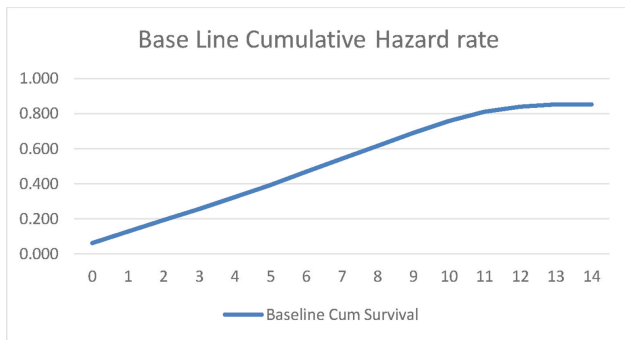


FIGURE 2 | Plot of baseline cumulative hazard rate w.r.t. time: training sample.

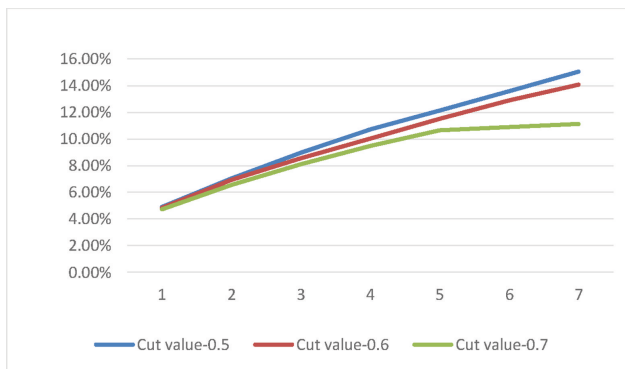


FIGURE 3 | Plot of type I error w.r.t. time: training sample.

misutilization of the available funds. It could also indicate that the funds were not utilized for the core activity of the business.

**Investment in subsidiary and group of companies to total investment ratio:** This indicates the amount of investment extended by the parent company to its subsidiaries or group companies. Higher the investment to subsidiary companies as compare to other healthy companies and if the company is itself

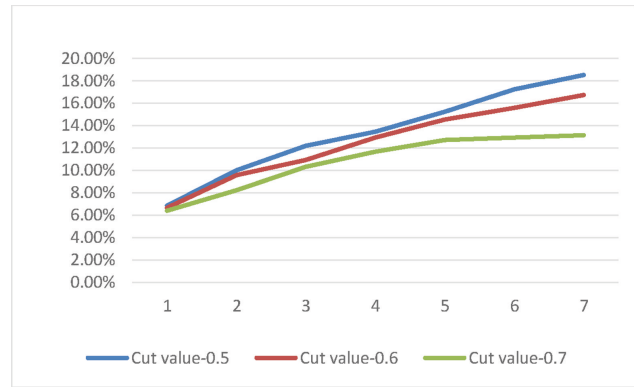


FIGURE 4 | Plot of type I error w.r.t. to time: testing sample.

in distress situations indicates the diversion and misutilization of funds.

**Cash-to-asset ratio:** This ratio is expressed as a percentage and equals net cash flows from operating activities as a share of gross assets.

**Financing activities ratio:** Cash stream from financing exercises accounts for outside exercises that permit a firm to raise capital, reimbursing financial specialists, including or changing credits, or issuing more stock. Cash stream from financing exercises appears speculators the company’s money-related quality. This, when compared to the resources of the firm, demonstrates the proportion of the accounts to it add up to resources claimed by the firm.

**Investment cash ratio:** Cash stream from contributing exercises is a thing on money stream explanation that reports the total alter in a company’s cash position coming about from any picks up (or losses) from ventures within the monetary markets and working backups and changes coming about from sums went through on ventures in capital resources such as plant and hardware. This stream when compared to the deals or turnover of the company demonstrates how well money contributed had created pay to the company.

### 2.3. Explanatory variables choosing process

This area depicts the budgetary proportions considered by the show for the disappointment expectation. Attending to the broad writing audit of existing writing on corporate default models, we identify the foremost common budgetary proportions that are inspected in several sorts of parametric and machine learning methods. As a case, the logit and danger models counting those by Altman (6), Ohlson (15), Zmijewski (16), Shumway (17), and Campbell et al. (18). Since most of the models have been tried for the non-Indian information set, we moreover attempted to work out and propose our claim budgetary proportions, which, in larger part, relate to use and action proportions to look at in case

**TABLE 3** | Classification table: training sample.

Classification table <sup>a,b</sup>					
	Observed	Predicted			
		NDC and WDC		Percentage correct	
		0	1		
Step 0	NDC and WDC	0	4,186	40	99.05
		1	21	89	80.91
Overall percentage					98.62

<sup>a</sup>Constant is included in the model.

<sup>b</sup>The cutoff value is 0.500.

NDC, non-defaulted companies; WDC, wilfully defaulted companies.

**TABLE 4** | Classification table: testing sample.

Classification table <sup>a</sup>					
	Observed	Predicted			
		NDC and WDC		Percentage correct	
		0	1		
Step 1	NDC and WDC	0	897	20	97.81
		1	8	26	76.47
Overall percentage					97.05

<sup>a</sup>The cutoff value is 0.500.

NDC, non-defaulted companies; WDC, wilfully defaulted companies.

any of these could be critical indicators other than those well built up within the past writing.

The outline insights of all the chosen covariates are given in **Table 1**. The exception choices strategy has moreover been presented to extricate out the exception firms to smoothen our information set. First, our hunch is the look of best observational show utilizing the relationship investigation. In the event that tall relationship is identified, we prioritize the foremost commonly utilized and best performing proportions within the writing. Second, the choice of factors entering our models is made by looking at the noteworthiness of the budgetary proportions. The score utilized for forecast comprises as it were of noteworthy or hardly critical factors chosen by stepwise estimation beneath the centrality level of 10%. The ultimate choice of factors is advance subordinate upon the execution premise.

## 2.4. Summary snapshots of chosen parameters

As demonstrated in ponders found beneath the survival relapse rate, the test has been isolated into two bunches: WDC and NDC. The graphic measurement given in this ponder has taken “mean” as the “measure of central tendency” and “standard deviation” as the degree for “variation.” The tables appear the graphic insights, i.e.,

cruel and standard deviation. The outline insights of chosen money-related markers for all companies have been appeared within **Table 1**. The summarized frame of mean and standard deviation of both the categories of the companies has been given as “group statistics.”

The clear measurements when examined appear that on certain factors there is a significant contrast within the scores between the two bunches taken in our test. As anticipated, the average of the productivity marker is less than the defaulted companies as compared to effective one. Moreover, as is cleared from **Table 1**, in regard of all other factors chosen for the test, there is noteworthy distinction within the values of both average and standard deviation between the two bunches of companies.

## 3. Analysis of the independent variables

The yield of Cox corresponding risk analysis is given in **Table 2** based on training sample. The ordinary distribution for relapse coefficients and gamma distribution for the standard failure rate have been utilized to calculate the back conveyance of the system parameter. A total of 15,000 emphases were performed and 2,000 beginning tests have been disposed of as burning, with each five tests taken for induction to evacuate the pollution of introductory esteem choice to begin the Markov chain. We found the relationship of the parameters based on the part thickness, follow, and autocorrelation plot. The receiver operating characteristic (ROC) curve is then plotted to demonstrate precision. The range beneath the ROC curve assesses the execution of the system. The vertical pivot indicates the extent of the WDCs that has been accurately classified beneath the suitable set by the curve. This can be called affectability.  $X$  pivot indicates specificity, which is the rate of erroneous classification (**Figure 1**). Thus, superior system is one where the ROC curve twists more toward the upper cleared out hand corner of the chart. The range beneath the curve is the precision proportion of the ROC curve. The risk and survival rate of the curve have also been plotted to evaluate the rate of default and victory. Based on the survival likelihood and diverse cutoff esteem (0.05, 0.06, and 0.07), the default and well-off companies have been arranged. If the survival probability is 0.05 or higher, the company is classified as victory. Otherwise default for other companies. The distribution of healthy and failure companies for training and testing data is presented in **Tables 3, 4**, respectively. Using the training information set, we have checked the status of the companies and compared with the calculated status to plan a two-way possibility table of victory and default companies. In this displaying workout, our fundamental point is to diminish the Sort I blunder Pr (company is defaulting but classified as victory), as higher the sort one blunder higher the chance to misfortune of



**TABLE 5** | *N* period ahead predictions based on training sample (%).

Cutoff value/ <i>N</i> -year ahead	1	2	3	4	5	6	7
<b>Training sample: type I error</b>							
0.5	4.89%	7.04%	8.95%	10.70%	12.12%	13.58%	15.04%
0.6	4.83%	6.91%	8.55%	10.03%	11.52%	12.89%	14.06%
0.7	4.72%	6.56%	8.08%	9.46%	10.63%	10.87%	11.12%
<b>Training sample: correct prediction</b>							
0.5	95.11%	92.96%	91.05%	89.30%	87.88%	86.42%	84.96%
0.6	95.17%	93.09%	91.45%	89.97%	88.48%	87.11%	85.94%
0.7	95.28%	93.44%	91.92%	90.54%	89.37%	89.13%	88.88%

**TABLE 6** | *N* period ahead predictions based on testing sample (%).

Cutoff value/ <i>N</i> -year ahead	1	2	3	4	5	6	7
<b>Type I error</b>							
<b>Testing sample: type I error</b>							
0.5	6.83%	9.99%	12.20%	13.46%	15.25%	17.25%	18.51%
0.6	6.65%	9.57%	10.94%	12.93%	14.51%	15.56%	16.72%
0.7	6.39%	8.20%	10.30%	11.67%	12.72%	12.93%	13.13%
<b>Training sample: correct prediction</b>							
0.5	93.17%	90.01%	87.80%	86.54%	84.75%	82.75%	81.49%
0.6	93.35%	90.43%	89.06%	87.07%	85.49%	84.44%	83.28%
0.7	93.61%	91.80%	89.70%	88.33%	87.28%	87.07%	86.87%

cash by financial specialist. In any case, higher the Sort II blunder, Pr (company is effective but classified as default) is issue of as it were the opportunity taken a toll. Sort I error at *T* period (1–7 years) has been plotted for each cutoff values. The evaluated parameters of error obtained through calculated training sample are used on testing sample to judge the productivity of out-of-test information. In this way, the analysis is done and results for both training and testing samples are arranged in **Tables 5, 6**, respectively.

Organization and forecast capacity of the factors on training and testing samples is shown in **Tables 7, 8**. The model accurately classified 80.9% default companies within the training sample normal premise, as its prescient capacity diminishes normal premise by 76.5% on testing sample. Over the time, Sort I blunder increments from 4.89 to 15.04% for training sample expecting the cutoff esteem 0.5 and 6.83 to 18.51% on testing sample. Then as the cutoff values increase from 0.5 to 0.7, the error decreases, thus indicating increase in prescient capacity for both training and testing samples. Receiver operating characteristic (ROC) curve recommends that the system execution is a better way as the range beneath the curve is around 85.36%.

The factors like ROAs depict the profitability and effectiveness with which the concern earns income based on its available resources. It may be mentioned that the corporate health is contingent on its income generation. There is an inverse association between the earnings and

chances of renegeing on loan shown by a minus sign. The finding from this unique forecasting tactics with an extra set of factors on pre-occupation and siphoning of reserves has led emphatically and considerably to the execution of the budgetary default forecast demonstration for domestic concerns. Moreover, it illustrates the fact about the financing stream factors in that they are better creators of intentional evaders than the accounting ratios. As anticipated, the analysis shows direct association of habitual evaders of highly indebted firms and reverse link for units with robust profit, liquid resources, and productivity indicators. Such findings are in consonance with prevailing studies in this field. Such outcome depicts that firms' monetary indicators considered within the technical framework have a strong distinguishing and foreseeable ability which the analytical tool might offer assistance to enhance the prediction for default chance by effectively utilizing data on the likelihood of premeditated evasion.

## 4. Conclusion

The present research has developed an insolvency framework under the survival framework to forecast the chance of deliberate non-payment of dues by debtors and categorize that along with healthy firms. This article reveals that financially weak firms have been in deep financial distress

somewhere between 2 and 3 years prior to their declaration as a wilful defaulter by the initial financial entity and its reporting on the same to the credit information companies. In addition, the article explores experimentally the importance of major budgetary factors that seem to offer assistance to moderate the risk due to default by progressive analysis of available information set regarding survival likelihood of wilful defaulters. This exposition deliberates extensively on all the probable factors that may impinge on the financial health of a company. This exercise shows the indicators those are found to provide most useful inputs regarding firms performing deliberate non-repayment and ascertain their forecasting efficacy. The outcome of this analysis with presentation of three cash-related factors on redirection and digression of credit to firm and its associates is a vital reason for robust findings related to insolvency forecast for Indian private sector. The assessment of forecasting efficacy of insolvency of such selected indicators has been performed for a variety of time points, yielding satisfactory results. In many scenarios, the factors are found to be technically relevant to proper direction and quantities as per the prior knowledge. Overall, the system is able to perform with a high accuracy level of 81% for proper prediction of failed corporates. Moreover, the healthy enterprises are categorized with 99% level of precision based on the available information sample. It is found that as one moves farther in time, the forecasting correctness reduces and type I error rises.

## Acknowledgments

We were grateful to Ms. Lakshmi Karthik for useful comments in an earlier draft. All the error(s) and omission(s) are solely the responsibility of the authors, if any. The authors can be contacted at [arvind\\_all\\_d\\_kr@yahoo.com](mailto:arvind_all_d_kr@yahoo.com) and [nitin\\_005us@yahoo.com](mailto:nitin_005us@yahoo.com), respectively.

## References

- Reserve Bank of India. *DBOD No. DL (W).BC. 110/20.16.003/2001-02 Dated May 30, 2002*. Mumbai: Reserve Bank of India (2002).
- Reserve Bank of India. *DBOD No. DL.BC. 111/20.16.001/2001-02 Dated June 4, 2002*. Mumbai: Reserve Bank of India (2002).
- Karthik L, Shrivastava A, Subrmanayam M, Joshi AR. Prediction of wilful defaults: an empirical study from Indian corporates. *Int J Intell Technol Appl Stat.* (2017).
- Ernst and Young. "Unmasking India's NPA issues – Can the Banking Sector Overcome This Phase?" *A Survey Report Conducted and Published by EY's Fraud Investigation & Dispute Services. E&Y.* (2015). Available online at: [www.ey.com/in](http://www.ey.com/in) (accessed Jan 04, 2017).
- Beaver WH. Financial ratios as predictors of failure. *J Account Res.* (1966) 4:71–111.
- Altman EI. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J Financ.* (1968) 23:589–609.
- Kaplan EL, Meier P. Nonparametric estimator from incomplete observations. *J Am Stat Assoc.* (1958) 53:457–81.
- Cox DR. Regression models and life-tables. *J R Stat Soc Ser B.* (1972) 34:187–220.
- Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data; Wiley Series in Probability and Statistics.* (1980).
- Cox DR. Partial likelihood. *Biometrika.* (1975) 62:269–76.
- Ibrahim JG, Chen M, Sinha D. *Bayesian Survival Analysis, Springer Series in Statistics.* (2001).
- Giroud X, Mueller HM, Stomper A, Westerkamp A. Snow and leverage. *Rev Financ Stud.* (2012) 3:680–710.
- Asimakopoulos I, Avramidis PK, Malliaropoulos D, Travlos NG. "Moral Hazard and Strategic Default: Evidence From Greek Corporate Loans," *Bank of Greece, Economic Analysis and Research Department – Special Studies Division.* (2016). Available online at: [www.bankofgreece.gr](http://www.bankofgreece.gr) (accessed Feb 26, 2017).
- Jayadev M. Predictive power of financial risk factors: an empirical analysis of default companies. *Vikalpa.* (2006) 31:45–56.
- Ohlson JA. Financial ratios and the probabilistic prediction of bankruptcy. *J Account Res.* (1980) 18:109–31.
- Zmijewski ME. Methodological issues related to the estimation of financial distress prediction models. *J Account Res.* (1984) 22:59–82.
- Shumway T. Forecasting bankruptcy more accurately: a simple hazard model. *J Bus.* (2001) 74:101–24.
- Campbell JY, Hilscher J, Szilagyi J. In search of distress risk. *J Financ.* (2008) 63:2899–939.
- Aghion BA. On the design of a credit agreement with peer monitoring. *J Econ Dev.* (1999) 60:79–104.
- Altman EI, Marco G, Varetto F. Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (The Italian Experience). *J Bank Financ.* (1994) 18:505–29.
- Bardhan S, Mukherjee V. Willful default in developing country banking system: a theoretical exercise. *J Econ Dev.* (2013) 38:101–21.
- Goel P, Pathak D. Factors affecting the repayment performance of borrowers in district central co-operative banks in Punjab. *Asia Pac J Manage Entrepr Res.* (2014) 3:47–84.
- Mishra R. Institutional credit and rural development: a case study of Dasarathpur block of Jajpur district, Odisha. *J Commerc Manage Thought.* (2014) 5:625–34.
- Reserve Bank of India. *Circular IECD.No.PMD.25/ C.446 (PL)-89/90 Dated April 5, 1990*. Mumbai: Reserve Bank of India (1990).
- Sanjeev GM. Bankers' perceptions on causes of bad loans in bank. *J Manage Res.* (2007) 7:40–6.
- Sheppard JP. The dilemma of matched pairs and diversified firms in bankruptcy prediction models. *Mid Atlantic J Bus.* (1994) 30:9–25.
- Gupta V. An empirical analysis of default risk for listed companies in India: a comparison of two prediction models. *Int J Bus Manage.* (2014) 9:223–34.
- Altman EI. Commercial bank lending: process, credit scoring, and costs of errors in lending. *J Financ Quant Anal.* (1980) 15:813–32.
- Goldman L, Cook EF, Brand DA, Lee TH, Rouan GW, Weisberg MC, et al. A computer protocol to predict myocardial information in emergency department patients with chest pain. *New Engl J Med.* (1982) 307:588–97.
- Goldman L, Weinberg M, Olshen R, Cook EF, Sargent RK, Lamas GA, et al. Hazard plotting for incomplete failure data. *J Qual Technol.* (1969) 1:27–52.
- Stone M. Firm financial stress and pension plan continuation/replacement decisions. *J Account Public Policy.* (1991) 3:175–206.
- Gepp A, Kumar K. Predicting financial distress: a comparison of survival analysis and decision tree techniques. *Procedia Comput Sci.* (2015) 54: 396–404.

## Appendix

Kernel density plot, random draws plot, and autocorrelation plot of random numbers generated from posterior distribution of coefficients.

