International Journal of Data Science

ISSN online: 2053-082X - ISSN print: 2053-0811
https://www.inderscience.com/ijds

Life cycle prediction and survival model construction of digital economy enterprises integrating survival analysis

Shulei Yin

DOI: 10.1504/IJDS.2025.10071082

Article History:

Received:	20 January 2025
Last revised:	03 March 2025
Accepted:	11 March 2025
Published online:	09 May 2025

Life cycle prediction and survival model construction of digital economy enterprises integrating survival analysis

Shulei Yin

School of Economics and Management,
Qilu Normal University,
Jinan, 250000, Shandong, China
Email: rhscfxysl@163.com

Abstract: The life cycle of digital economy enterprises is affected by many complex and nonlinear factors. Traditional methods can only handle some simple linear relationships. This paper proposes an improved gradient boosting regression tree (GBRT) model to further enhance the prediction ability. Firstly, the Kaplan-Meier survival curve is used for descriptive statistics and exploratory analysis. Then, the accelerated failure time (AFT) model is used to model the enterprise life cycle. Finally, the GBRT model is used to predict the mean-square error (MSE) value, and it is compared with the MSE value of the linear regression model. The MSE value of the survival model is 0.015, much smaller than the MSE value of the linear regression model of 0.232, reflecting the superiority of the survival model.

Keywords: AFT; accelerated failure time; Kaplan-Meier survival curve; GBRT; gradient boosting regression tree; survival model; linear regression method.

Reference to this paper should be made as follows: Yin, S. (2025) 'Life cycle prediction and survival model construction of digital economy enterprises integrating survival analysis', *Int. J. Data Science*, Vol. 10, No. 6, pp.1–19.

Biographical notes: Shulei Yin received her Master degree from Broas University. Sweden. Now, she works in School of Economics and Management, Qilu University. Her research interests include economics and management.

This paper was originally accepted for a special issue on 'AI-based data application and management' guest edited by Prof. S. Tsai and Assoc. Prof. H-T Wu'.

1 Introduction

The core characteristics of the digital economy lie in the in-depth application of information technology and data-driven business model transformation, which has a profound impact on the operating environment, competition rules, and growth path of enterprises. In this context, studying the dynamic changes in the enterprise life cycle can provide theoretical support for the enterprise's strategic planning and new empirical basis for policy making. The traditional enterprise life cycle usually includes the start-up phase, growth phase, maturity phase, and decline phase. However, driven by the digital economy, the boundaries between these phases are becoming increasingly blurred. In the start-up phase, the digital economy has significantly lowered the barriers for enterprises to enter the market. The widespread popularity of technologies such as internet platforms, cloud computing, and big data has enabled start-up enterprises to enter the market quickly at a lower cost, shortening the time cycle from concept to commercialisation. At the same time, these technologies provide enterprises with innovative tools and resources to help them stand out in a fiercely competitive environment. For example, e-commerce platforms and social media give the enterprises efficient marketing tools, and crowdfunding platforms and venture capital provide enterprises with more convenient financing channels. However, start-up enterprises face a high risk of failure due to increasing market competition and rapid changes in consumer preferences. Therefore, enterprises in this phase must quickly build digital capabilities to enhance competitiveness. During the growth phase, the main characteristics of enterprise growth in the digital economy are the emergence of network effects and rapid scale expansion. Through digital platforms, enterprises can quickly attract many users and form a positive feedback mechanism. For example, an increase in the number of users further attracts more users to join (network effect), driving the rapid expansion of the enterprise. In addition, enterprises can also use big data technology to analyse user behaviour, optimise products and services, and increase market share. However, the information transparency brought about by the digital economy has also intensified market competition, and growing enterprises must continue to invest resources to maintain their competitive advantage. After entering the maturity phase, traditional enterprises often face the problem of slowing growth, but the digital economy has injected new impetus into enterprises in this phase. Mature enterprises can maintain their market competitiveness through data-driven decision-making and service optimisation. For example, enterprises can launch personalised products or services by deeply mining consumer preferences. In addition, the digital economy also encourages enterprises to build ecosystems and achieve resource integration and diversified growth through cooperation or mergers and acquisitions with other enterprises. In this phase, enterprises focus more on digital transformation and continuous innovation to extend their life cycle. Finally, during the recession phase, the digital economy allows enterprises to achieve 'reverse growth'. Many enterprises have regained market vitality through digital transformation. For example, traditional manufacturing enterprises have achieved intelligent production through the industrial internet, thereby improving efficiency and competitiveness and successfully escaping the predicament of the recession. Therefore, the digital economy can change the life cycle trajectory of an enterprise and provide enterprises with a more flexible transformation path.

The widespread application of digital technology has significantly improved the efficiency of information dissemination, accelerated the dynamic changes in the market environment, and promoted the rapid evolution and progress of the enterprise life cycle. On the one hand, start-up enterprises can use digital platforms to enter the market quickly and gain market share. On the other hand, increasingly fierce market competition has increased the risk of enterprises being eliminated, resulting in a general shortening of their life cycles. At the same time, enterprises need to continue to adapt to dynamic changes in technological innovation and consumer demand, making adjustments to the life cycle more flexible and diversified. In the context of the digital economy, data has become an essential strategic asset for enterprises. By using technologies such as big data analysis, artificial intelligence, and the Internet of Things, enterprises can more precisely predict market demand and optimise resource allocation and operational efficiency, thereby effectively extending the life cycle and enhancing market competitiveness. This data-driven transformation provides new possibilities for enterprises to achieve sustained growth in a rapidly changing environment. For example, enterprises can use customer relationship management (CRM) systems to analyse customer behaviour, develop more precise marketing strategies during the growth phase, and explore new business growth points during the maturity phase. Unlike the traditional life cycle model that emphasises internal factors of the enterprise and the external market environment, in the context of the digital economy, platform ecology, network effects, and data collaboration have become key factors affecting the life cycle of enterprises. During the growth phase, enterprises usually rely on platforms to achieve scale expansion. In the maturity phase, maintaining sustained growth through open cooperation or ecological integration is necessary. In the digital economy environment, the life cycles of enterprises in different industries, sizes, and business models also show significant differences. Technology-driven businesses tend to have shorter life cycles and exhibit explosive growth. Traditional manufacturing is expected to extend its life cycle through digital empowerment. Platform-based enterprises can achieve rapid expansion, stable operations, and sustainable growth by relying on network effects, data-driven, ecosystem construction, and continuous innovation. Service-oriented businesses rely more on customer needs and user experience and can usually achieve long-term development. In addition, there are apparent differences in the life cycle performance between small and micro enterprises and large enterprises. Small and micro enterprises have a shorter life cycle and are vulnerable to market fluctuations and funding constraints, but they are more flexible and can respond quickly to digital trends. Large enterprises have relatively longer life cycles and are more resistant to risks, but the speed and efficiency of digital transformation are often constrained by organisational structure and culture. Overall, the digital economy has given enterprises a new logic of survival and growth. Through technology application, data-driven, and ecological construction, enterprises can adapt to dynamic market changes and open up broader growth space in a complex and changing environment.

Firstly, the Kaplan-Meier survival curve is used to analyse the probability of overall enterprise survival and determine the key turning points. The accelerated failure time (AFT) model is then used to model the enterprise life cycle and capture the impact of characteristics on its life cycle. Finally, the gradient boosting regression tree (GBRT) model is used to predict the MSE value, which is then compared with the MSE value of the linear regression to understand the excellence of the survival model. The innovation of this paper is to propose an improved GBRT model that combines the AFT model with the GBRT model, which can handle complex nonlinear relationships, automatically select features and enhance data adaptability.

2 Related work

Survival analysis originated in the medical field and was mainly used to study the survival time of patients or the recurrence time of diseases. Tinguely et al. (2019) proposed a random forest destruction prediction model combined with survival analysis, which provided a new research perspective for disease diagnosis and treatment. Li et al. (2023) explored the application of whole slide images in survival analysis based on a multi-feature fusion network of self-supervised learning. This method showed great potential in cancer prognosis and pathological analysis. Huang et al. (2024) combined Gaussian random fuzzy numbers to develop a new evidence-based multimodal survival analysis model, further improving complex medical data's analysis efficiency. Aguirre Paz et al. (2024) proposed a survival analysis method based on multi-tree structure decision fusion, which showed excellence in processing high-dimensional and nonlinear data. Wang et al. (2019) combined survival analysis with machine learning algorithms through systematic research, providing new ideas for model optimisation and data processing. In and Lee (2019) explored the core concepts of survival analysis. Nagy et al. (2021) studied the survival analysis of pancreatic cancer hallmark genes, providing valuable genetic information for early cancer detection and treatment decisions. Wiegrebe et al. (2024) promoted substantial development in survival analysis by applying deep learning technology, significantly improving its ability in complex data analysis. Lanczky and Gyorffy (2021) proposed a survival analysis method suitable for medical research based on network tools, providing a convenient and efficient solution for clinical applications. Survival analysis has also been widely used in public health during the epidemic. For example, Salinas-Escudero et al. (2020) conducted a detailed analysis of the survival time of COVID-19 cases in the Mexican population to reveal the epidemic's impact on different population groups, providing an essential basis for policy making. In addition, Jing et al. (2019) developed a loss function based on regression and rank constraints to optimise survival analysis models, further improving their accuracy and stability. Lee and Lim (2019) outlined statistical methods for survival analysis using genomic data, providing systematic guidance for studying biomedical data. In recent years, the application of deep learning in survival analysis has also made significant progress. Zhao and Feng (2020) proposed a deep neural network model based on pseudo-values, effectively improving survival analysis's adaptability in complex data processing. Zhong et al. (2021) further developed a deep extended risk model, making the application of survival analysis under multivariate conditions more efficient and precise. Gyorffy (2021) conducted survival analysis on all genes and found some genes that performed best in chemotherapy for estrogen-positive breast cancer, providing a new direction for personalised medicine and precision treatment.

Different stages of the enterprise life cycle have a profound impact on financial management strategies and decisions. Habib and Hasan (2019) analysed the core role of the enterprise life cycle in accounting, finance and corporate governance, discussed the financial management strategies of enterprises at different life cycle stages, and looked forward to future research directions. Wang et al. (2020) studied the relationship between working capital management and financial performance at different stages of the life cycle and found that growth-stage enterprises are more concerned with capital expansion and market share, while mature-stage enterprises focus on capital stability and profit optimisation. Bansal (2022) focused on the impact of the life cycle on income and expense transfer, revealing the changes in financial strategies at each stage and their constraints. Singh et al. (2023) studied the relationship between the dividend strategy and life cycle of Indian enterprises and found that start-up enterprises tend to use funds for growth investment and pay less dividends; while mature-stage enterprises enhance market trust through stable dividend payments. Moshashaie and Mirzajani (2019) studied the relationship between the life cycle and equity cost of listed companies on the Tehran Stock Exchange and found that changes in the life cycle have a significant impact on equity cost, especially playing a key role in financing decisions and capital market performance.

The corporate life cycle and non-financial areas mainly focus on the changes and impacts of social responsibility, profit management, investment decisions, human resource management, etc. at different stages of the corporate life cycle. Trihermanto and Nainggolan (2020) explored the relationship between the life cycle and social responsibility, revealing the adjustments of policies and investment strategies at each stage. Thu and Khuong (2023) found through a study of Vietnamese companies that start-up and growth-stage companies pay more attention to short-term profits, while mature companies use social responsibility activities to enhance brand image and promote sustainable development. Zhao et al. (2019) analysed the relationship between social responsibility and investment from a life cycle perspective, pointing out that growth-stage companies pay more attention to capital expansion and market development, while mature companies focus on enhancing brand value and competitiveness through social responsibility activities. Uttamagana and Wirakusuma (2023) explored the cost allocation effect of social responsibility on market performance and believed that companies need to adjust their social responsibility strategies according to the life cycle stage to optimise costs and market performance. Yarahmadi (2019) pointed out that financial resources play a regulatory role in the relationship between corporate life cycle and social responsibility. Mature companies can assume more social responsibilities with stronger financial capabilities, thereby enhancing their market image. Hussain et al. (2020) pointed out from the perspective of earnings management that there are significant differences in earnings management pressures among enterprises at different stages, which is closely related to financial goals, market demand and capital structure. Michalkova et al. (2022) found through a study of European transport companies that there are significant differences in the quality of earnings at different life cycle stages, which provides an important basis for management decisions. Eulaiwi et al. (2020) pointed out that the investment committee structure and cash holding decisions of non-financial enterprises at different stages of the life cycle show significant differences, reflecting the profound impact of the life cycle on capital structure and investment strategy. Ma et al. (2020) analysed the role of venture capital in the growth stage, especially in supporting technological innovation and market expansion. Petris (2023) studied the life cycle of real estate investment trust companies, emphasising the impact of their characteristics on investment decisions. Hejazi and Salehi (2019) focused on the relationship between the enterprise life cycle and human resource investment and internal control system, indicating that as the life cycle progresses, the investment in these two aspects gradually increases to support efficient operations and sustainable development. Hong (2020) found that mature enterprises can significantly improve performance through resource integration and optimisation. Abiahu et al. (2019) studied Nigerian listed companies and found that changes in the life cycle may lead to adjustments in financial statement classification and accounting standards. Michalkova (2021) based on the Goodenough-Kinson rule showed that the profit growth of mature companies slowed down, while start-up and growth companies mainly focused on capital accumulation and market expansion, providing targeted management suggestions for companies at different stages. Ostad et al. (2022) proposed an analysis of accrual anomalies at different stages of the life cycle of companies listed on the Tehran Stock Exchange (Shang and Asif, 2023; Zhang et al., 2023).

This paper attempts to compare with some of the referenced literature, comparing research methods, contributions, limitations, etc. The details are shown in Table 1.

Table 1 Literature review and comparison of related papers

Research papers	*Research methods*	*Contribution points*	*Limitation*
Habib and Hasan (2019)	Theoretical summary and case analysis	Explore the role of life cycle in finance and corporate governance	Lack of quantitative analysis, mainly focusing on theoretical aspects
Wang et al. (2020)	Empirical analysis	Study the relationship between fund management and financial performance at different stages of the life cycle	Focusing only on financial aspects without considering other factors
Singh et al. (2023)	Regression analysis	Analyse the relationship between dividend strategy and life cycle	Only involves dividend strategy, which is relatively simple
Zhao et al. (2019)	Empirical analysis	Research on the relationship between life cycle and social responsibility	Not enough focus on survival analysis and lack of quantitative models
Hussain et al. (2020)	Earnings management perspective	Studying the impact of life cycle on earnings management	Focusing on earnings management and ignoring overall enterprise survival prediction
This paper	Survival Analysis + Machine Learning	A new prediction model combining AFT and GBST is proposed	The model is complex and the computational cost is high

3 Methods

3.1 Kaplan-Meier survival curve

The Kaplan-Meier method is well suited for dealing with survival data because it can estimate survival rates with incomplete data and does not require making assumptions about the distribution of survival times. This method is a non-parametric method used to estimate the survival probability distribution from time to event (business exit or bankruptcy). It can visualise the trend of survival probability over time and help identify key turning points in the life cycle. Firstly, data on the enterprise life cycle are collected, including each enterprise's survival time (t) and status (d). Among them, t is the time when the enterprise delists, and d represents whether the enterprise exits, with exit as 1 and non-exit as 0. The data is then sorted, from largest to smallest according to the survival time *t*. The survival probability at each time point is calculated next. Some symbols are defined. t_i represents the *i*th time point (the time when the enterprise delists). n_i represents the number of enterprises still under observation at time t_i. d_i represents the number of enterprises that experience the event at time t_i.This paper needs to calculate the estimated value of the survival function at each time point and then draw the survival curve.

The formula for calculating the conditional survival probability is:

$$p_i = 1 - \frac{d_i}{n_i} \tag{1}$$

Then, the cumulative survival probability is calculated, which is defined as:

$$S(t) = \prod_{t \le i}^{k} p_i = \prod_{t \le i}^{k} (1 - \frac{d_i}{n_i}) \tag{2}$$

where *S*(*t*) is the survival probability of the firm at time *t*, is the number of times the firm declines at a point in time, is the number of firms still at risk before a point in time, and *k* is the number of time points.

The Kaplan-Meier curve is drawn, which is step-like and decreases at each event node. Finally, the key turning point is determined, which is the interval or inflection point where the survival curve decreases significantly. In this study, Kaplan-Meier survival curves were used for exploratory analysis to help us understand the distribution characteristics of each stage of the enterprise life cycle (such as start-up, growth, and maturity) and observe the proportion of enterprises surviving at different time points.

3.2 *Accelerated failure time*

Why did this paper choose the AFT model? Because compared with the Cox proportional hazard model, the AFT model does not require the assumption of risk proportionality, but analyses the acceleration effect of time through logarithmic transformation. It is suitable for scenarios where we need to understand the influencing factors of time to events (i.e., the life cycle of an enterprise). AFT is a vital survival analysis method used to model the logarithm of survival time directly and evaluate the accelerating or delaying effect of specific features on the life cycle. Figure 1 shows the specific steps of this method.

The model is assumed first. Instead of directly modelling the risk rate, it is assumed that specific features can have a proportional effect on the life cycle, affecting its ability to accelerate or decelerate. Assuming that the survival time of the enterprise is *T*, The basic form of AFT is

$$T = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2) \cdot \in \tag{3}$$

where *T* is the life cycle of the enterprise, X_1, X_2 are the influencing factors of the enterprise life cycle, β_0, β_1, β_2 are regression coefficients, indicating the impact of each factor on the life cycle.

Then, the survival data, variable (*T*), feature (*X*), and truncating status (δ), are collected. They are cleaned and standardised. The missing values can be filled with the mean, and the formula is:

$$x_{new} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{4}$$

Data standardisation can be done using the Z-Score standardisation method, which is defined as:

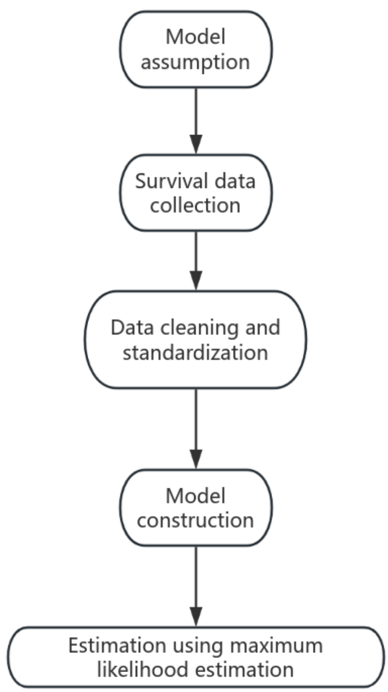
$$x_{new} = \frac{x - \mu}{\sigma} \tag{5}$$

Then, the model is built. The appropriate error term $\in$ is selected according to the distribution of *T*. The model formula is:

$$S(t|X) = S_0\left(\frac{t}{\exp(X\beta)}\right) \tag{6}$$

Among them, if the acceleration factor is greater than 1, it means that feature *X* extends the life cycle of the enterprise. If it is less than 1, it means that the life cycle of the enterprise is shortened.

Figure 1 Steps in applying the AFT model



Then the model is estimated using maximum likelihood estimation. The logarithm of the enterprise life cycle *T* follows a normal distribution, which can be defined as:

$$\ln(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \in \tag{7}$$

where $\in$ follows a normal distribution, $\in \sim N(0, \sigma^2)$

For the given data, we need to construct a joint likelihood function. There may be missing data in the data. In order to handle the missing data, we need to merge the likelihood functions of the censored samples and the uncensored samples, where the likelihood function of the uncensored samples is:

$$f(T_i \mid X_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln(T_i) - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2})^2}{2\sigma^2} \right) \tag{8}$$

The formula represents the probability density of the survival time of enterprise i's life cycle T_i under the given covariate X_i.

The likelihood function of the censored sample is:

$$S(T_i^* \mid X_i) = 1 - F(T_i^* \mid X_i) \tag{9}$$

where $F(T_i^* \mid X_i)$ is the cumulative distribution function of the lifetime.

The joint likelihood function is defined as:

$$L(\beta_0, \beta_1, \beta_2, \sigma) = \prod_{i \in D} f(T_i \mid X_i) \prod_{i \in C} S(T_i^* \mid X_i) \tag{10}$$

where D is the set of uncensored sample data, and C is the set of censored sample data.

The AFT model considers the relationship between the duration of the enterprise life cycle and various influencing factors. Through this model, we can analyse how different factors accelerate or delay the progress of the enterprise life cycle.

3.3 Gradient boosting regression tree model and linear regression

GBRT is a technique that combines the survival regression method with the gradient boosting tree model, focusing on processing survival data and improving prediction precision. The model uses a tree structure to capture complex nonlinear relationships, predicts the survival space by gradually building an additive model, and guides the model training process with the help of the log-likelihood loss function in survival regression.

Minimise the joint log-likelihood function in the AFT model, which is defined as:

$$L_{GBRT} = -\ln L(\beta_0, \beta_1, \beta_2, \sigma) \tag{11}$$

In GBRT, a new set of decision trees is constructed through multiple iterations, and each iteration minimises the loss function to improve the prediction precision. Firstly, the data is prepared, including survival time (T) and event-related feature (X). Whether an event occurs (δ) is marked, and the truncation is obtained. Then, the model is initialised using constant prediction, and the tree model is fitted using the gradient boosting algorithm. In each iteration, a new tree is constructed based on the residual of the current model. The formula is:

$$f_m(x) = f_{m-1}(x) + \eta \cdot h_m(x) \tag{12}$$

Among them, $f_m(x)$ is the prediction function of the mth epoch. η is the learning rate (usually less than 1). $h_m(x)$ is the decision tree trained in the mth epoch of iteration.

Then, the gradient is used to train each tree, ensuring that the tree minimises the loss function in each iteration. Multiple iterations are repeated, and the model is continuously adjusted with new trees until convergence. Finally, the trained model is used to predict new data to get the survival function estimate. The mean-square error (MSE) is adopted to measure precision.

This paper compares the survival regression model with the linear regression model. The prediction of the enterprise life cycle using linear regression is first made by predicting the same features mentioned above, such as enterprise scale and market competition intensity. The linear relationship is:

$$T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \in \tag{14}$$

Then, the model parameter $\beta = [\beta_0, \beta_1, \beta_2]$ is estimated by least squares, which is defined as:

$$\beta = (X^T X)^{-1} X^T Y \tag{15}$$

Finally, the precision of the model is measured by the MSE, reflecting the deviation between the model-predicted value and the true value. The smaller the value, the more precise the prediction. It is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{16}$$

GBRT gradually improves the prediction accuracy by training multiple weak predictors (i.e., decision trees) step by step, each time making corrections based on the residuals of the previous model. In this study, GBRT was used to predict the length of the enterprise life cycle and calculate the MSE value.

4 Experimental analysis

4.1 Analysis of survival probability

Data from ten groups of enterprises is collected, as shown in Table 2. The enterprises are not named by company name but by number.

Table 2 Enterprise life cycle and event indication

Company number	*Life cycle* (*year*)	*Event indication* (*d*)
1	5	1
2	2	1
3	3	1
4	10	0
5	7	1
6	8	0
7	6	1
8	4	0
9	9	1
10	2	1

The data in Table 2 shows ten companies numbered 1 to 10. Among them, the two companies numbered 2 and 10 have the same life cycle, both two years, and both exit the market. The other eight companies have different life cycles. Three of these ten companies are still alive, and the remaining seven exit the market.

The conditional survival probability and cumulative survival probability are calculated according to Formulas (1) and (2), as shown in Table 3. The survival curve is drawn according to the data in Table 2.

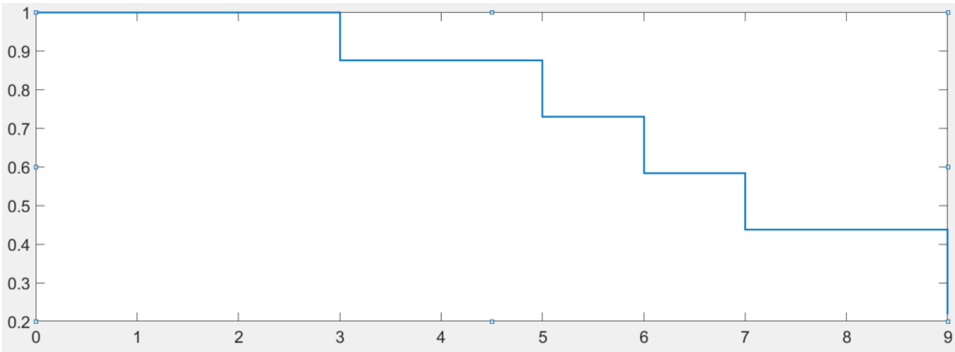
Table 3 Indicators of the survival curve

Time (t_i)	*Number of surviving enterprises* (n_i)	*Number of events* (d_i)	*Probability of surviving* (p_i)	*Cumulative survival probability* $S(t)$
2	10	2	80%	80%
3	8	1	87.5%	70%
4	7	0	100%	70%
5	7	1	85.7%	60%
6	6	1	83.3%	50%
7	5	1	80%	40%
8	4	0	100%	40%
9	4	1	75%	30%
10	3	0	100%	30%

According to the data in Table 3, the survival of enterprises in each year of the life cycle is as follows. In the second year, two companies go bankrupt, and the survival probability drops to 80%. In the third year, another company exits from the market, and the survival probability is 87.5%. No company goes bankrupt during the fourth year, and the survival probability remains 100%. In the fifth year, one company goes bankrupt, and the survival probability is 85.7%. In the sixth year, another company exits from the market, and the survival probability drops to 83.3%. In the seventh year, one company goes bankrupt, and the survival probability is 80%. No company exits from the market during the eighth year, and the survival probability is 100%. In the ninth year, one company goes bankrupt, and the survival probability drops to 75%. In the tenth year, no company goes bankrupt, and the survival probability is 100%. Comprehensive analysis shows that the cumulative survival probability during the 10-year life cycle is only 30%. In the end, a total of three companies survive.

According to Figure 2, there are apparent turning points in the 3rd, 5th, 6th, and 7th years, which may be related to industry competition and market changes. The median survival time is the 6th year, indicating that the enterprise may be a capital-intensive enterprise, or the government policy support is strong enough, so the enterprise has strong market potential and risk adaptability.

Figure 2 Kaplan-Meier survival curve (see online version for colours)



4.2 *Capture the impact of features on life cycle and compare GBRT prediction accuracy with linear regression*

i In the data collection part, the AFT model is used to study whether the characteristics of enterprises in the industry extend or shorten the life cycle. The collected data are shown in Table 4. The market intensity is 1 for high intensity and 0 for low intensity, and the enterprise scale is quantified into a specific value.

Table 4 Index values of 5 enterprises

Company number	*Life cycle (Year)*	*Enterprise scale* (X_1)	*Market competition intensity* (X_2)	*Exit status* (δ)
1	5.0	10	1	1
2	8.0	15	0	1
3	3.5	8	1	1
4	12.0	20	0	0
5	4.5	12	1	1

According to the data in Table 4, Company 1 survives for 5 years under high-intensity market competition, with a scale of 10, and finally exits from the market. Company 2 survives for 8 years under low-intensity market competition, with a scale of 15, and finally exits from the market. Company 3 survives for 3.5 years under high-intensity competition, with a scale of 8, and finally exits from the market. Company 4 shows stronger resilience in low-intensity competition, with a scale of 20, surviving for 12 years and not exiting from the market. Company 5 survives for 4.5 years under high-intensity market competition, with a scale of 12, and finally exits from the market. These data show that the impact of market competition intensity on enterprises' survival time and scale has significant differences.

The GBRT method is used to predict survival time. Table 5 lists the collected data. The enterprise competition intensity is calculated based on a score of 1–10.

Table 5 Feature indicators and life cycle of 10 enterprises

Company number	*Enterprise life cycle*	*Number of employees* (X_1)	*Market competition intensity* (X_2)
1	1	50	8
2	2	100	7
3	3	150	6
4	3	200	5
5	4	250	4
6	5	300	3
7	6	400	2
8	7	500	1
9	8	600	1
10	9	700	1

According to Table 5, Company 1 has 50 employees, a market competition intensity of 8 points, and a survival time of 1 year. Company 2 has 100 employees, a market competition intensity of 7 points, and a survival time of 2 years. Company 3 has 150 employees, a market competition intensity of 6 points, and a survival time of 3 years. Company 4 has 200 employees and survives for 3 years in an environment with a competition intensity of 5 points. Company 5 has 250 employees, a market competition intensity of 4 points, and a survival time of 4 years. Company 6 has 300 employees, a competition intensity of 3 points, and a survival time of 5 years. Company 7 has 400 employees, a market competition intensity of 2 points, and a survival time of 6 years. Company 8 has 500 employees, a competition intensity of only 1 point, and a survival time of 7 years. Company 9 has 600 employees and survives for 8 years under the condition of a competition intensity of 1 point. Company 10 has 700 employees, a competition intensity of 1 point, and a survival time of 9 years.

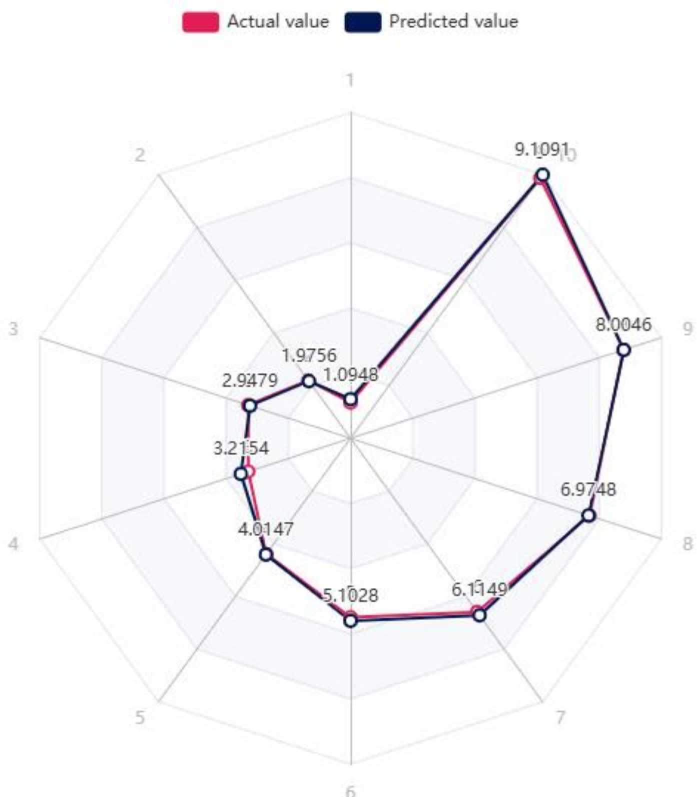
ii In the model construction part, the natural logarithm of the life cycle data in Table 4 is taken to calculate ln(T). The ln(T) values of companies 1–5 are 1.609, 2.079, 1.253, 2.485, and 1.504, respectively.

The maximum likelihood estimation is performed to obtain the values of the feature parameters and scale parameters. β_0 is equal to 0.5, and β_1 is 0.1370. This means that for every increase of one unit of enterprise scale, the logarithm ln(T) of the life cycle increases by 0.1370, and the corresponding life cycle T increases by about 1.1468. The larger the enterprise scale, the longer the life cycle. β_2 is equal to 0.0859, and the corresponding life cycle T increases by about 1.0897. The milder the competition intensity, the longer the life cycle. σ equals 0.1391. The final formula of the AFT model is: $\ln(T) = 0.5 + 0.1370\,X_1 + 0.0859\,X_2 + 0.1391\in$. Features of enterprise scale and market competition intensity both have a prolonging effect on the life cycle of an enterprise.

Calculated based on the data in Table 5, the MSE value is 0.015. The closer the value is to 0, the higher the prediction precision.

According to Figure 3, the predicted values of Companies 1-10 predicted by the GBRT model are 1.0948 years, 1.9756 years, 2.9479 years, 3.2154 years, 4.0147 years, 5.1028 years, 6.1149 years, 6.9748 years, 8.0046 years, and 9.1091 years, respectively. The deviation between the predicted values and the actual values is not significant, and they are highly overlapped, so the prediction precision is relatively high.

Figure 3 Predicted values of survival time by the GBRT model and actual values (see online version for colours)



The linear regression is used to predict survival time. The data in Table 4 is converted into a matrix and divided into X and Y. X is the dependent variable vector, and Y is the design matrix. The first column is the constant term 1 (used to estimate matrix β_0), and based on β in the least squares method, the values of β_0, β_1, and β_2 can be calculated. Table 6 lists the specific indicators.

Table 6 Indicator analysis of linear regression model

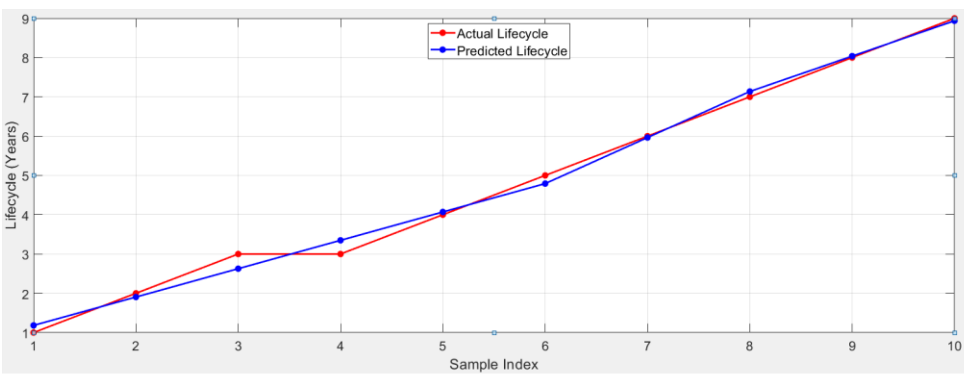
	Estimate	*SE*	*t*	*p*
Intercept	2.9088	0.6652	4.3730	0.0033
X_1	0.0090	0.0010	8.6218	0.0001
X_2	–0.2720	0.0873	–3.1143	0.0170

According to Table 6, the value of β_0 is 2.9088; the value of β_1 is 0.0090; the value of β_2 is –0.2720. The intercept's standard error (SE) value is 0.6652, with a t-value of 4.3730 and a p-value of 0.0033. The SE value of feature X_1 is 0.0010, with a t-value of 8.6218 and a p-value of 0.0001. The SE value of feature X_2 is 0.0873, with a t-value of –3.1143 and a p-value of 0.0170. The p-values of the intercept and the two features are all less than 0.05, rejecting the null hypothesis, which is significant. The obtained MSE value is 0.232, and the value of R^2 is 0.994. The equation of linear regression is $Y = 2.9088 + 0.0090X_1 - 0.2720X_2$.

By substituting the feature values of Company 1 to Company 10 into the linear regression equation, the predicted survival time of each company is calculated as follows: the predicted survival time of Company 1 is 1.1828 years; that of Company 2 is 1.9048 years; that of Company 3 is 2.6268 years; that of Company 4 is 3.3488 years; that of Company 5 is 4.0708 years; that of Company 6 is 4.7928 years; that of Company 7 is 5.9648 years; that of Company 8 is 7.1368 years; that of Company 9 is 8.0368 years; that of Company 10 is 8.9368 years.

According to Figure 4, the predicted values and actual values for these 10 companies are quite consistent, indicating that the predictions are relatively precise.

Figure 4 Predicted and actual value indicators of 10 companies (see online version for colours)



iii In the numerical results section, the two acceleration factors of enterprise scale and market competition intensity are 1.1468 and 1.0897 respectively, the MSE value of GBRT is 0.015, and the MSE value of the linear regression model is 0.232.

5 Discussion

According to the above conclusions, the acceleration factors of enterprise scale and market competition intensity are 1.1468 and 1.0897 respectively, and both are greater than 1, indicating that enterprises with larger scale have longer life cycles, and that enterprises with higher market competition intensity are more likely to survive for a longer period of time. Enterprise managers can optimise strategic decisions based on these results. For example, for large enterprises, they can further expand their market share and increase resource investment to consolidate their life cycles and ensure their continued competitiveness in the market. For enterprises facing high competition intensity, management can respond to competition by increasing innovation, improving product differentiation and optimising supply chain management, so as to stand out in the fierce market competition and extend their life cycles. In terms of model evaluation, the study pointed out that the MSE of the GBRT model is 0.015, which is much smaller than the MSE value of the linear regression model of 0.232, indicating that the prediction accuracy of GBRT is better than that of the traditional linear regression model. In actual operation, enterprise managers can use the GBRT model for life cycle prediction, especially when facing multiple variables and complex market environments. The high prediction accuracy of GBRT can help managers make more accurate decisions, formulate long-term strategies, and enhance the survival ability of enterprises at different stages of the life cycle. The survival analysis model that combines AFT and GBRT in this paper has more advantages than the traditional regression model. It can handle deleted data, consider nonlinear relationships, and provide more practical predictions. In practice, enterprises can apply survival analysis models to life cycle management, strategic planning, and risk assessment, especially when facing different market environments and competitive situations. Through accurate predictions of the enterprise life cycle, enterprises can formulate appropriate strategies at different stages, avoid overexpansion or waste of resources, and enhance their ability to develop sustainably.

6 Conclusion

According to the results of the Kaplan-Meier survival curve analysis, the life cycle of an enterprise is significantly affected by industry competition and market changes. The median survival time in the curve is the sixth year, which indicates that the enterprise is capital-intensive with strong market potential and risk resistance. The study shows that the acceleration factors of the two features of enterprise scale and market competition intensity are 1.1468 and 1.0897, respectively, both greater than 1, indicating that these features positively affect the enterprise's life cycle. In addition, the MSE of the GBRT model is 0.015, much smaller than the MSE value of the linear regression model of 0.232, fully demonstrating that the GBRT model has apparent advantages in prediction precision and further reflecting the superiority of the survival analysis model. However, there are still some limitations in the research of this paper, mainly reflected in the lack of dynamics. The current enterprise life cycle analysis usually relies on historical data, but in the digital economy environment, factors such as market environment, technological progress, and policy changes change frequently, and it is still a significant challenge to achieve real-time updates and predictions in a dynamic environment. Future enterprise life cycle prediction models need to have online learning capabilities to dynamically adjust according to real-time data and feedback. By continuously receiving new data, the model can adapt to changes in the market and environment promptly, thereby providing more accurate prediction results.

Conflicts of interest

The authors declare no conflict of interest.

References

Abiahu, M.C., Egbunike, P.A., Udeh, F.N., Egbunike, F.C. and Amahalu, N.N. (2019) 'Corporate life cycle and classification shifting in financial statements: evidence from quoted manufacturing firms in Nigeria', *Amity Business Review*, Vol. 20, No. 2, pp.75–91.

Aguirre Paz, L.M., Viteri Moya, J., Vásquez, R.A.D., Guerra, D.M.R. and Burkhon, D. (2024) 'Survival analysis based on fusion of decisions from multiple tree structure: a cutting-edge approach', *Fusion: Practice and Applications*, Vol. 14, No. 1, pp.263–272.

Bansal, M. (2022) 'Impact of corporate life cycle on misclassification practices: evidence from IFRS adoption in India', *Journal of Applied Accounting Research*, Vol. 23, No. 3, pp.628–649.

Eulaiwi, B., Al-Hadi, A., Hussain, S.M. and Al-Yahyaee, K.H. (2020) 'Investment committee, corporate cash holdings and corporate life cycle', *International Review of Finance*, Vol. 20, No. 3, pp.757–769.

Gyorffy, B. (2021) 'Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer', *Computational and Structural Biotechnology Journal*, Vol. 19, pp.4101–4109.

Habib, A. and Hasan, M.M. (2019) 'Corporate life cycle research in accounting, finance and corporate governance: a survey, and directions for future research', *International Review of Financial Analysis*, Vol. 61, pp.188–201.

Hejazi, R. and Salehi, M. (2019) 'Corporate life cycle stages and the relation between human resources investment and internal control system', *Empirical Research in Accounting*, Vol. 9, No. 2, pp.143–160.

Hong, N.H. (2020) 'The effect of business strategy on cost asymmetry according to corporate life cycle', *Journal of Digital Convergence*, Vol. 18, No. 5, pp.145–155.

Huang, L., Xing, Y., Lin, Q., Ruan, S. and Feng, M. (2024) *EsurvFusion: An Evidential Multimodal Survival Fusion Model Based on Gaussian Random Fuzzy Numbers*, arXiv preprint arXiv: 2412.01215.

Hussain, A., Akbar, M., Khan, M.K., Akbar, A., Panait, M. and Voica, M.C. (2020) 'When does earnings management matter? Evidence across the corporate life cycle for non-financial Chinese listed companies', *Journal of Risk and Financial Management*, Vol. 13, No. 12, p.313.

In, J. and Lee, D.K. (2019) 'Survival analysis: part II-applied clinical data analysis', *Korean J. Anesthesiol.*, Vol. 72, No. 5, pp.441–457.

Jing, B., Zhang, T., Wang, Z., Jin, Y., Liu, K., Qiu, W., Ke, L., Sun, Y., He, C., Hou, D., Tang, L., Lv, X. and Li, C. (2019) 'A deep survival analysis method based on ranking', *Artificial Intelligence in Medicine*, Vol. 98, pp.1–9.

Lanczky, A. and Gyorffy, B. (2021) 'Web-based survival analysis tool tailored for medical research (KMplot) development and implementation', *Journal of Medical Internet Research*, Vol. 23, No. 7, p.e27633.

Lee, S. and Lim, H. (2019) 'Review of statistical methods for survival analysis using genomic data', *Genomics and Informatics*, Vol. 17, No. 4, p.41.

Li, L., Liang, Y., Shao, M., Lu, S., Liao, S. and Ouyang, D. (2023) 'Self-supervised learning-based multi-scale feature fusion network for survival analysis from whole slide images', *Computers in Biology and Medicine*, Vol. 153, p.106482.

Ma, Q., Li, X., Chaudhry, P.E. and Chaudhry, S.S. (2020) 'Public relations and legitimacy: a study of new ventures on the corporate life cycle', *Systems Research and Behavioral Science*, Vol. 37, No. 4, pp.699–710.

Michalkova, L. (2021) 'Corporate life cycle and financial performance in terms of emerging markets', *European Proceedings of Social and Behavioural Sciences*, Vol. 121, pp.154–165.

Michalkova, L., Cepei, M., Valaskova, K., Vincúrová, Z. (2022) 'Earnings quality and corporate life cycle before the crisis. A study of transport companies across Europe', *Amfiteatru Economic*, Vol. 24, No. 61, pp.782–796.

Moshashaie, M. and Mirzajani, S. (2019) 'The relationship between corporate life cycle and cost of equity in listed firms on Tehran stock exchange', *Journal of Accounting and Social Interests*, Vol. 9, No. 1, pp.101–120.

Nagy, A., Munkacsy, G. and Gyorffy, B. (2021) 'Pancancer survival analysis of cancer hallmark genes', *Scientific Reports*, Vol. 11, No. 1, p.6047.

Ostad, M., Dastgir, M. and Khajavi, S. (2022) 'The accrual anomaly analysis at different stages of the corporate life cycle of companies listed on the Tehran stock exchange', *International Journal of Nonlinear Analysis and Applications*, Vol. 13, No. 1, pp.1269–1278.

Petris, P. (2023) 'Predicting the REIT corporate life cycle phase on a financial accounting basis', *Journal of Risk and Financial Management*, Vol. 16, No. 6, p.290.

Salinas-Escudero, G., Carrillo-Vega, M.F., Granados-García, V., Martínez-Valverde, S., Toledano-Toledano, F. and Garduño-Espinosa, J. (2020) 'A survival analysis of COVID-19 in the Mexican population', *BMC Public Health*, Vol. 20, pp.1–8.

Shang, K. and Asif, M. (2023) 'The design of a compound neural network-based economic management model for advancing the digital economy', *Journal of Organizational and End User Computing*, Vol. 35, No. 1, pp.1–15.

Singh, R., Gupta, C.P. and Chaudhary, P. (2023) 'Dividend policy and corporate life cycle: a study of Indian companies', *Managerial Finance*, Vol. 49, No. 11, pp.1722–1749.

Thu, L.B. and Khuong, N.V. (2023) 'How does the corporate life cycle influence Vietnamese firms corporate social responsibility?', *Cogent Economics and Finance*, Vol. 11, No. 1, p.2186043.

Tinguely, R.A., Montes, K.J., Rea, C., Sweeney, R. and Granetz, R.S. (2019) 'An application of survival analysis to disruption prediction via random forests', *Plasma Physics and Controlled Fusion*, Vol. 61, No. 9, p.095009.

Trihermanto, F. and Nainggolan, Y.A. (2020) 'Corporate life cycle, CSR, and dividend policy: empirical evidence of Indonesian listed firms', *Social Responsibility Journal*, Vol. 16, No. 2, pp.159–178.

Uttamagana, I.A.W. and Wirakusuma, M.G. (2023) 'Cost allocation effect of corporate social responsibility on market performance in the context of corporate life cycle', *International Journal of Business Management and Economic Research*, Vol. 14, No. 1, pp.2090–2095.

Wang, P., Li, Y. and Reddy, C.K. (2019) 'Machine learning for survival analysis: a survey', *ACM Computing Surveys* (*CSUR*), Vol. 51, No. 6, pp.1–36.

Wang, Z., Akbar, M. and Akbar, A. (2020) 'The interplay between working capital management and a firm's financial performance across the corporate life cycle', *Sustainability*, Vol. 12, No. 4, p.1661.

Wiegrebe, S., Kopper, P., Sonabend, R., Bischl, B. and Bender, A. (2024) 'Deep learning for survival analysis: a review', *Artificial Intelligence Review*, Vol. 57, No. 3, p.65.

Yarahmadi, J. (2019) 'Relationship between corporate life cycle and corporate social responsibility, and moderating role of financial resources', *Journal of Accounting Knowledge*, Vol. 10, No. 3, pp.69–87.

Zhang, H., Zhang, F., Gong, B., Zhang, X. and Zhu, Y. (2023) 'The optimization of supply chain financing for bank green credit using Stackelberg game theory in digital economy under internet of things', *Journal of Organizational and End User Computing*, Vol. 35, No. 3, pp.1–16.

Zhao, L. and Feng, D. (2020) 'Deep neural networks for survival analysis using pseudo values', *IEEE Journal of Biomedical and Health Informatics*, Vol. 24, No. 11, pp.3308–3314.

Zhao, T., Xiao, X. and Jiang, Y. (2019) 'Corporate social responsibility and corporate investment-a. study based on corporate life cycle', *Journal of Beijing Institute of Technology (Social Sciences Edition)*, Vol. 2019, No. 6, pp.107–115.

Zhong, Q., Mueller, J.W. and Wang, J.L. (2021) 'Deep extended hazard models for survival analysis', *Advances in Neural Information Processing Systems*, Vol. 34, pp.15111–15124.