DOES CONTROLLING FOR INVOICE AMOUNT IMPROVE PAYMENT TIME PREDICTIONS? A DISCRETE SURVIVAL APPROACH

by

Oleksandra Bilyk

A thesis submitted in partial fulfillment of the
requirements for the degree of

MA in Business and Financial Economics

Kyiv School of Economics

2025

Thesis Supervisor: ______________________________ Professor Olesia Verchenko

Approved by ____
Head of the KSE Defense Committee, Professor [Type surname, name]

____

____

____

Date ______________________________________

ACKNOWLEDGMENTS

The author wishes to express sincere gratitude to the entire Kyiv School of Economics community for genuine engagement and the tremendous effort you invest in providing students with the very best.

Special thanks go to Professor Elena Besedina and Professor Olesia Verchenko for their unwavering support and guidance from the very first day. Thank you for seeing the best in us and helping us to see it in ourselves.

The author is deeply grateful to the BFE and EA cohorts of 2023 for an unforgettable year and a half of shared learning and fun. A heartfelt thank you to Vadym and Petro—simply for everything.

The author also wishes to thank the Ampersand Foundation for its support and the commitment you bring to strengthening the community.

Finally, the author is very grateful to her mother, father, and grandmother for their faith in her abilities, and to her husband for his steadfast care and for always being on her side.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AUC Area Under the Curve

B2B Business-to-business

CPaaS Communicaiton Platform as a Service

ERP Enterprise Resource Planning system

FTSE The Financial Times Stock Exchange 100 Index

GLMM Generalized Linear Mixed Model

MASE Mean Absolute Scaled Error

ROC Receiver operating characteristic

CHAPTER 1. INTRODUCTION

For most of the business, account receivables serve as a major source to maintain working capital on the level that allows business to operate effectively. Therefore, collection of invoiced amounts becomes an essential daily process, moreover – an importnant business' need. The desire to optimize and improve the efficiency of debt collection results into creation of specialized teams and standardization of collection processes, i.e. internal collection policy. The next natural step is a significant increase in the efficiency of in-house collection, as well as partial transfer of debt to the collection agencies.

It is therefore important to explore approaches that can form the basis for improving cash-flow forecasting and prioritizing debt collection efforts at the company level. The motivation for this work stems from the author's three years of working experience in the Order-to-Cash department, and attempting to formulate the recommendations for debt collection prioritization using an empirical data. The limited nature of the main resource – one's working time – forces to prioritize one case over another under the guidence of internal Collection Policy. The Policy accounts for only two factors: customer's segment and time of overdue. This naturally follows from both the nature of the business – some clients are priority partners and bring in more revenue, and the nature of contractual relations, which clearlys define the payments terms and the possibility of applying late payments fees. Prioritization by amount owed, although not explicitly stated in the Policy, comes from common sense – collection efforts are directed primarily at the larger amounts.

But could the amount of an overdue account convey more information? Can we, using only internal historical data say more about the probability of payment of the debt based on the invoice amount? Positive answers to these questions would theoretically improve the effectiveness of internal collections, providing grounds for making a decision on when to transfer the debt to an external collection agency.

This study aims to find answers to the above questions for a specific company case. The dataset represents all account receivable invoices of one of the European legal entities of the Infobip company for the 2024 fiscal year. The research method chosen is discrete survival analysis, which is frequently used for analyzing time-to-event data, which in our case is time to payment. Originating in medical research, it is now successfully used in forecasting debt and loan repayments. By addressing the above mentioned questions, this research contributes both methodologically and practically. From a methodological perspective, we demonstrate how discrete survival methods could be adapted to incorporate time-varying components for real-world applications. From a practical standpoint, we gather evidence-based insights for invoice-level forecasting, which can support further improvement on more accurate cash-flow planning and collection efforts prioritization. Our work aim to bridge the gap between advanced statistical modeling and operational decision-making in order-to-cash process.

CHAPTER 2. LITERATURE OVERVIEW

Collection of account receivables is a crucial part of the daily operations of any business and is therefore the subject of research and ongoing improvement efforts. Despite the relevance of improving debt collection efficiency, the number of studies focused on predicting the actual and due dates of invoice payments is limited. However, it is also worth noting studies in a similar area, namely, corporate credit risk and credit management, as well as the prediction of bankruptcies and defaulted loans. These related domains also explore factors that can predict payment timing, the likelihood of late payment, and approaches to solving similar classification problems (e.g., the fact of payment or its time horizon). Default prediction could be interpreted as prediction of payment absence.

Research in payment prediction area over the past 15 years can be divided into two large groups based on research questions, which could be stated as "*whether* payment will be done" and "*when* payment will be done". Thus, some authors focus on predicting fact of payment within a specific time horizon, and this formulates the research question as a classification problem. To proritize collection activities, Appel et al. (2020) focus on likelihood of payment delay within 5 days time frame, and Zeng et al. (2008) aim to classify invoice payment time group (paid on time, 1-30 days late, 31-60 days late etc.) on the stage of invoice creation to identify high-risk invoices (90+ days late). Aforementioned works both use machine learning approach, and stand out by research goal precisely tailored to real business need. The strength of this focus is its high practical value, as the result of the implemented model is usually a clear input for making daily operational decisions. Such insights are important, but naturally limited in application. Other studies, which are aimed at modeling the time to event — Smirnov (2016), Dirick et al. (2017), Marletta (2022), Fimba and Fernandes (2024), Saini et al. (2024) — mostly include datasets from several companies, are more general and follow established frameworks of classic survival approach. The advantage of survival analysis as a method is that it provides both the ability to classify and estimate the time to an event.

Regading methods utilized within payment prediction research area, two main approaches can be distinguished. Survival analysis methods are more widely used in works which focusing on predicting time to event, mostly in research fields of credit risk management. Survival analysis, which originated in the field of epidemiology, is widely used in econometrics to estimate the probability of an event, the time horizon, and the factors that influence it. Fimba and Fernandes (2024) use Cox Proportional Hazards model for hazard rates and survival probability estimation of credit repayment delays. In benchmark study by Dirick et al. (2017) compared the perfomance of Accelerated Failure Time, Cox Proportinal Hazards and mixture cure models in predicting time of credit default. After assessing models on ten real-world datasets, Cox Proportional Hazards model with penalized splines showed best prediction accuracy. The authors use splines to model non-linear effect of covariates on hazard rates for continuous-time survival model. In our researh, we as well use splines but for modeling non-linear effect of discrete time on hazard rates. We will follow the methodology used by Saini et al. (2024) for predicting payment times as part of incoming payments forecasting on discrete time intervals.

Smirnov (2016) compares prediction accuracy of traditional survival Cox Proportional Hazards method with machine learning Random Survival Forest, adapted for survival data with right-censored observations (i.e. for some invoices payment was not done on moment of dataset collection). Their findings shows that machine learning outperformed statistical survival method. Moore and van Vuuren (2024) move forward and introduce Survival Boost algorithm designed to reinforce machine learning method with features derived by survival analysis model.

Recently more authors have been focusing exclusively on machine learning methods if research task is stated in the form of classification problem. Among best-performing algorithms are mentioned C4.5 Decision Tree Induction and PART algorithms (Zeng et al., 2008), Gradient Boosted Decision Trees XGBoost (Appel et al., 2020; Martikainen, 2023), Attention-enhanced Deep BiLSTM (Liu et al., 2020). Disadvantages of machine learning methods may include low interpretability and the need for regular model retraining to avoid a decrease in its effectiveness. Appel et al. (2020) show that the model has the best perfomance when trained on historical features withing last four months, and with widening timeframe the model degrades. At the level of implementation in real business processes, some of the above-mentioned methods require significant computing power and a strong expertise in machine learning domain.

Most reviewed studies consider time as a discrete variable, which is natural when focusing on "paid on time / paid late" classification regarding payment time, or predicting debt aging group with machine learning methods. Several researchers use debt age groups that are standard for business practice: paid on time, paid late for 1–30, 31–60, 61–90, and 90+ days. (Zeng et al., 2008; Martikainen, 2023), so treating it as discrete intervals. Smirnov (2016) test both survival and machine learning approaches on continuous data. Authors which utilize solely survival methods (Fimba and Fernandes, 2024; Marletta, 2022; Dirick et al., 2017) treat time as continuous value for their predictions of time to payment of default. Study of Saini et al. (2024) is the only one which utilize survival method on discrete time. Worth noting that discrete time is much more suitable for business applications, since initial data recorded in discrete intervals (e.g. day of payment), and the response required by the business is shaped by the categories of days, weeks, months and quarters.

Usage of large datasets of real companies undoubtedly provides the most valuable insights. Common to all studies is the use of invoice-level data as amount, issue date, and payment date as covariates, same applies to loans. Customer-specific data poses challange for most researches, particularly because disclosure of unique customer characteristics such as industry or financials is limited by privacy concerns, so manually engineered metrics are frequently used instead. Zeng et al. (2008) transform customer-specific data by calculating sum and average of invoices (both paid on time and late) and average number of overdue days for each customer. Martikainen (2023) accounts for number of prior invoices issued to customer. In addition to these, Appel et al. (2020) introduce standard deviation of late and outstanding invoices. Fimba and Fernandes (2024) in their prediction of time of delay in mortrage payments implement debtor-specfic latent variable using „5C factors" –

method used in banking for accessing credibility and accountability of debtors. Marletta (2022) in assessing factors which posses higher risk of credit default among small and medium Italian enterprises determines that lower risk assosiated with older firms and larger loans. Smirnov (2016) conducts their research on late business-to-business payments combining the most extensive data about debtors, incorporating inforamation about tax debt, non-submitted tax declarations, field of debtor's activity, and financial data (sales and current assets). Submission of annual tax report and area of debtor's activity, along with historical payment data, turned to be most influential predictors. Overall, most researchers highlight that historical data about payment behavior as the most significant for predicting payments, and significance some of customer-specific covariates as customer company size and operational area. Bellotti and Crook (2009) study the probability and timing of default of credit card accounts with the help of Cox Proportional Hazards and introduce macro series as time-varying covariates. Interest rates, unemployment, earnings and FTSE index shown to be the most influential macro drivers of default hazard, and their inclusion significantly improved model fit.

The difficulty of collecting a large and consistent dataset is worth mentioning separately. Appel et al. (2020) mentioned "legacy systems with branches all over the world," while several other authors point to the unreliability of manually entered data. From personal experience, the author can confirm that accounting for all indirect actions and factors associated with debt is a real challenge in the context of an outdated, limited, or, conversely, constantly changing ERP system that relies on manual labor.

In our research we will follow metrics engineering approach that was widely used in the literature. In their latest report international collection agency Atradius (2025) states that the top two reasons of delays of B2B payments are customer's liquidity issues (39%) and delays in payment process (30%). The second one aligns with insights from researc of Wu et al. (2025) that systematic payment delay up to 60 days of overdue may be done by debtors intentionally in order to gain an advantage in their liquidity's management. We hope to capture this customer-specific characteristic by including as control varaible total amount of debt which was overdue on the date when invoice was issued.

Studies that apply survival analysis to predict time to invoice payment and at the same time treating *time* as discrete value is still very limited. Incorporating a time-varying baseline hazard or time-varying covariate effects is presented in only few studies (Bellotti and Crook, 2009; Saini et al., 2024), despite a variability over time being an immanent characteristic of real-world payment dynamics. In this thesis, we extend the methodology proposed by Saini et al. (2024) by introducing an additional "simplified" model alongside the original specification for broader robustness check. Furthermore, we compile a unique, up-to-date dataset from real-world invoicing operations and test the theoretical framework on actual business data rather than synthetic or historical samples. We hope that this combination of methodological refinement and empirical validation will contribute to forecasting of payment behavior in supply chain finance.

CHAPTER 3. METHODOLOGY

Our research hypotheses are stated in next form:

- H_1: Adding time-varying effect of the *invoice amount* does improve the model comparing to basic model
- H_2: Adding time-varying effect of the *overdue debt amount* does improve the model comparing to basic model

To test these hypotheses, we will use discrete survival approach introduced by Saini et al. (2024) by estimating two extended models and comparing them with basic model. The choice of survival analysis as staticstical method for this research was driven by the nature of the dataset and the presence of successful research cases of its applications for modeling the time to default or payment. Discrete survival approach is suitable for estimation of probability an event occuring within specified time interval (Kleinbaum and Klein, 2016). B-spline basis functions are used for modeling the baseline hazard, and parameters estimation is done with generalized linear mixed model (GLMM), which allows response variable to have a non-normal distribution. The explanations of the theoretical part of this section are based on Tutz and Schmid (2016).

The corner concepts or survival statistical method are *time* and *event*. In our case, *time* is a number of days passed from the moment an invoice was issued until an *event* – the moment when invoice is paid. As we work with discrete time, let it take values in $\{1, \ldots, k\}$, which consists of k intervals: $[a_0, a_1)$, $[a_1, a_2)$, … , $[a_{q-1}, a_q)$, $[a_q, a_\infty)$, where $q = k - 1$. Thus, discrete time $T \in \{1, \ldots, k\}$ means that $T = t$ was observed on the time interval $[a_{t-1}, a_t)$. The components of the statistical model are designated as follows:

- T_{ij} – time to payment of an invoice j for a customer i
- b_i – customer i specific random effect

- x_{ij} – vector of explanatory variables for an invoice *j* for a customer *i*

Another integral part of survival analysis is the *hazard function*, which in application to our topic specifies the conditional probability of an invoice being paid at some moment (interval), given that it was not paid before. For fixed time $t \in [a_{t-1}, a_t)$ hazard function $\lambda(t|x) = P(T = t \mid T \geq t, x)$ provides the probability of an invoice being paid at time *t*, or rather on the interval $T \in \{t+1, \ldots, k\}$ given that $T \geq t$. As in chosen methodology, we follow generalized linear modeling framework for discrete hazard function parametrization:

$$\lambda(t|x_{ij}, b_i) = h(\beta_i + \gamma_{0t} + x_{ij}{}^T\gamma) \tag{3.1}$$

In the right side of formula (3.1) the *linear predictor* with general form of $\beta_i + \gamma_{0t} + x_{ij}{}^T\gamma$ composed of random effect β_i for customer *i*, time-varying intercept γ_{0t}, and dot product of covariates vector and coefficients vector $x_{ij}{}^T\gamma$. Function $h(\cdot)$ called the *response function* and connects linear predictor with hazard function. The value of response function, retrieved from a linear predictor for some time *t*, will be the *hazard rate* at time *t*, i.e. the probability of event occuring in this moment, given this hasn't happened before.

Given that hazard function represents the probability of an event occurring either at time *t* (given it is reached) or in a subsequent indivisible time interval, distribution functions are convenient as a response function. Following methodology of Saini et al. (2024), we chose *grouped proportional hazards model* as response function, which is known as discretized analogue of Cox Proportional Hazards model that use for modeling time to event for continuous time. Thus, hazard function will have next notation:

$$\lambda(t|x_{ij}, b_i) = 1 - \exp\left(-\exp(\beta_i + \gamma_{0t} + x_{ij}{}^T\gamma)\right) \tag{3.2}$$

In linear predictor term γ_{0t} as an intercept represents time-varying baseline hazard, which is not dependent on covariates and may be interpreted as common for all invoices. Constant baseline hazard for invoice being paid seems unrealistic, because it may be higher closer to invoice date or date agreed to be a payment term. According to chosen methodology, we model it with *B-splines basis functions*. Resulted baseline hazard function is a function in time which constructed of polynomial segments, smoothly joined together. Segments are limited by boundary knots, which are $t = 1$ and $t = max_{ij}T_{ij}$. We will use cubic B-splines, i.e. baseline hazard on each segment will be a weighted sum of three basis third-dergee polynomial functions $B_m(t)$. Thus, time-varying intercept γ_{0t} will be expressed in next form:

$$\gamma_{0t} = \gamma_0 + \sum_{m=1}^{3} \gamma_m B_m(t) \tag{3.3}$$

In R code we model we the baseline hazard using bs( ) function from splines package. To model time-varying effect of an invoice amount, we include interaction of B-splines and logarithm of invoice amount.

After introducing B-splines, we are ready to specify the linear predictor in full form. Notation for linear predictor η_{ijt} of an invoice *j* for a customer *i* being paid at time *t* will have the next form in application to our model:

$$\eta_{ijt} = b_i + \gamma_0 + \sum_{m=1}^{3} \gamma_m B_m(t) + \beta^T x_{ij} + \sum_{m=1}^{3} \delta_m (logAmount_{ij} \times B_m(t)) \tag{3.4}$$

Designations are as follows:

- $b_i \sim N(0, \sigma^2)$ is a random effect for *i* customer which captures customers' heterogeneity in payment behavior
- γ_0 is an intercept
- $\gamma_m B_m(t)$ is a product of cubic B-spline basis functions and it's coefficient, which models time-varying baseline hazard

- $\beta^T x_{ij}$ are fixed effects of an invoice *j* for a customer *i* (customer's segment, invoicr amount, invoice quarter, payment term, log invoice amount, log amount of overdue debt on the date of invoice)
- δ_m are coefficients for interactions between amount of an invoice *j* for a customer *i* and B-splines which allows to account for time-varying effects of invoice amount

Customer-specific effect chosen to be random to follow methodology by Saini et al. (2024), but it is determined by limitations in computing power as well. If set to fixed effect, it greatly increases model's complexity and requires to estimate separate coefficient for each customer (972 more coefficients in our case) instead of one variance term for random effects.

Following Tutz and Schmid (2016), we estimate model's parameters using maximum likelihood estimation approach, which is standard for generalized linear models. Coefficients obtained in this way make the observed data the most probable given our model specification. For discrete survival data, likelihood of an event occurring at some time *t* will be defined as a product of hazard rate $\lambda(t|x)$ for interval *t* and survival probability for all preceding intervals. In this way, we reformulate problem as a binary response problem, where each T_{ij} is a vector of T_{ij} components. This is the sequence of Bernoulli observations, each one being 0 except the last one being 1. I.e., at every interval (day) event (payment) did not occur, except the last interval, which is the day when payment received, and invoice closed. Thus, likelihood is a binary response model for T_{ij} number of observations. The likelihood of an invoice *j* for a customer *i* being paid at time $t = T_{ij}$ presented by next formula notation:

$$L_{ij} = \lambda\big(T_{ij} \mid x_{ij}, b_i\big) \prod_{s=1}^{T_{ij}-1} (1 - \lambda(s \mid x_{ij}, b_i) \qquad (3.5)$$

Formula (3.5) represents likelihood contribution of a single invoice, and by maximazing it for each invoice from the dataset we obtain estimated for linear predictor, which after gives us estimated hazard rate. In R we use glmer( ) function from lme4 package for model fitting. After model inference, we can construct payment probability for each specific $t = T_{ij}$:

$$\hat{P}(T_{ij} = t) = \hat{\lambda}(t) \prod_{s=1}^{t-1} (1 - \hat{\lambda}(s)) \tag{3.6}$$

Probability distribution computed with (3.6) allows us to forecast payment time for each invoice *j* for a customer *i*. Using probability disctribution, we can obtain expected payment time that will be the mathematical expectation:

$$\widehat{T_{ij}} = E(T_{ij}) = \sum_{t=1}^{\tau} t * \hat{P}\,(T_{ij} = t) \tag{3.7}$$

We specifying one basic and two extended discrete survival models to test our research hyposes. Models' fitting will be performed using training dataset. Steps represented by formulas (3.6) and (3.7) will be conducted on testing dataset, which allows us to test predictive performance of models in most close to real task and a practical method. Fixed effects $\beta^T x_{ij}$ of an invoice *j* for a customer *i* which are included in linear predictor part of discrete survival models are listed in the table below. All three models include random effect b_i for *i* customer, and time-varying intercept $\gamma_0 + \gamma_m B_m(t)$. Time-varying effect of amounts presented by interaction of B-spline and log-transformed invoice or debt amount. By *debt* we mean total outstanding amount that was overdue on the date of invoice (e.g. for customer with 15-days payment term, if invoices for January and February were not paid at the day when March invoice was issued, sum of their total will be considered as *debt*.)

Table 3.1. Discrete time survival models specification

Covariate	Basic model	Extended model 1	Extended model 2
Customer segment	X	X	X
Invoice quarter	X	X	X
Payment term	X	X	X
Log invoice amount	X	X	X
Time-varying log invoice amount		X	
Time-varying overdue log debt amount			X

To compare the predictive efficiency of discrete survival models, we estimate three Cox Proportional Hazards models for continuous time using similar covariates (Table 3.2). To address both research hypotheses, for robustness check we will conduct likelihood-ratio test, comparing goodness of fit of extended models with basic model within both survival methods.

Table 3.2. Continuous time Cox Proportional Hazards survival models specification

Covariate	Basic model	Extended model 1	Extended model 2
Customer segment	X	X	X
Invoice quarter	X	X	X
Payment term	X	X	X
Log invoice amount		X	
Log debt amount			X

Our expectations about signs of estimated coefficients based on insights from reviewed literature (Wu et al., 2020; Atradius, 2025) and will be common for both discrete survival and Cox Proportional Hazards models. Since we model time to payment as an event, in application to our reseacrh delays in payment will be associated with lower risk of payment, i.e. lower hazard rate, or lower probability of receiving payment at some time *t*. Therefore, we expect negative signs for both time-varying amount covariates. In other words, we assume that both higher invoice amount and overdue debt amount will be assosiated with longer payment delays.

To test validity of estimated extended models and identify potential misspecifications, we calculated Pearson's χ^2 and deviance statistics on grouped data evaluate how well predicted probabilities match observed data.

Additionally, we inspected Q–Q plots of standardized empirical Bayes estimates of customer-level random intercepts. This diagnostic evaluates if the distribution of customer-specific random effects is approximately normal, as it is one of the assumption of generalized linear mixed model.

To statistically test our hypotheses, we will conduct the likelihood ratio and Wald tests for both extended models. The likelihood ratio test will allow us to compare likelihood of each extended model with basic and test if adding of new covariate is stiatistically significant. The Wald test evaluates significance of individual coefficient estimates by comparing their estimated value to its standard error under the null hypothesis that the coefficient equals zero. We will report general statistics for models.

For goodness-of-fit assessment, we will additionally validate the extended models on out-of-sample testing dataset, collected from 2025 fiscal year invoices to examine predictive performance beyond the estimation period. This step helps to make sure that improvements observed for in-sample tests are not driven by overfitting and that the models generalize well to future observations.

CHAPTER 4. DATA

The data used in this research was provided by Infobip – an international company that provides CPaaS (Communicaiton Platform as a Service) and omnichannel solutions for businesses, which mostly used for marketing communications with clients.

The dataset is constructed based on account receivables invoices, issued by one of Inbobip's regional european entities in 2024 fiscal year and collected on 1st of September 2025. Following the methodology used by Saini et al. (2024), the dataset was shaped following the next three principles. Firstly, the outstanding invoices were excluded, which limits research results application to the scenario of standard and expected business flow, i.e. the quantity and price of servises provided are not disputed, and the customers are still operating. Secondly, the invoices paid on the date issued, and paid with 356 and more days overdue were excluded as well, to exclude extreme outliers. Finally, within the dataset each client has at least three invoices included, which allows to capture at least a minimum level of variety in customer payment behavior, and obtain customer-specific random effects estimates. Each invoice has next attributes:

- Customer Id – unique customer identifier.
- Invoice date – date of an invoice, majority of invoices dated with the last day of month.
- Total – invoice amount in unified currency, tax excluded. Most invoices reflect monthly service consumption.
- Segment – customer's segment, internal classification assigned based on customer's number of employees, yearly revenue, and strategical importance in the region. "A" segment denoting the biggerst clients, and "D" the smallest ones.
- Payment terms – term in days, during which customer must pay an invoice, and specified in agreement between customer and Infobip. Longer payment terms

mostly associated with higher customer segment. Values present in the dataset are 7, 15, 20, 25, 30, 45, 60, and 90 days.

- Due date – the date by which an invoice must be paid or it becomes overdue. Calculated as invoice date plus payment terms.
- Paid date – date on which payment was received and invoice closed.
- Days to payment – number of days between invoice date and paid date.
- Due in – a number which represents the number of days from the paid date to the due date if positive, and the number of days from the due date to the payment date if negative.
- Total overdue – additionaly calculated value of total amount of outstanding debt on the invoice date. Simply put, the sum of all invoices for which payment is overdue at the moment a new invoice is issued.

Total overdue feature was not initially presented in collected data. Following data preparation approach proposed in reviewed literature, we calculate it to capture customer-specific payment behavior historical data. Calculating total amount of debt that was overdue on the date of new invoice creation, we address one of the most undesirable cases when the debtor does not have the enough funds to pay off obligations to service providers.

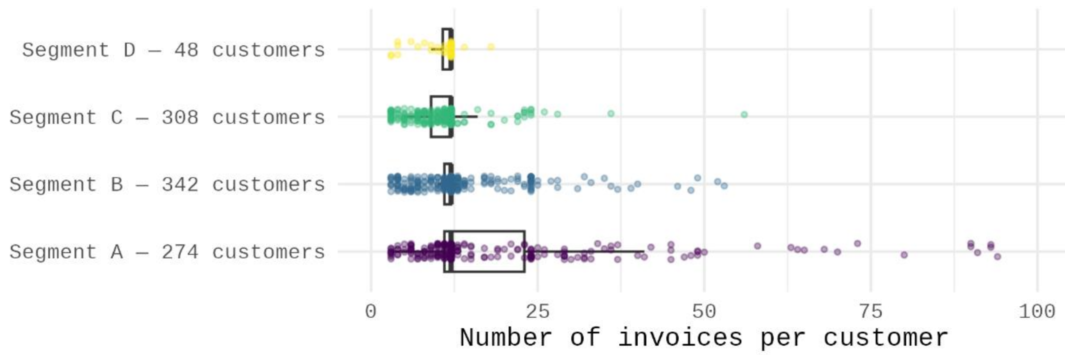
Final training dataset contains 13880 invoices issued for 972 customers. The descriptive statistics for invoice amount by customer segment is summarized in Table 4.1. Excess skewness and kurtosis statistics for all four segments indicate substentialy non-normal, right-skewed and peaked distribution of invoice amounts among all four segments. In this study we use logarithmic transformation of invoice amounts to handle outliers.

Table 4.1. Descriptive statistics for invoice amount by customer segment, training dataset

Segment	N	Mean	SD	Median	Min	Max	Skew	Kurtosis
Segment A	5515	131467.29	482044.18	9643.72	0.03	11588188.89	8.93	117.78
Segment B	4483	15610.89	52026.82	3308.12	1.15	861385.59	8.78	95.78
Segment C	3369	2347.32	4096.80	960.85	1.65	49603.47	4.96	38.58
Segment D	513	625.41	1099.63	211.31	0.94	6646.01	3.09	10.23

Figure 4.1. shows the number of clients by segment, as well as distribution of the number of invoices per client. The median number of invoices per client is 12, the same for all segments. Three clients in Segment A, with 128, 155, and 211 invoices, respectively, remain outside the boxplot.

Figure 4.1. Boxplot of number of invoices per customer by segment, training dataset



The distribution of time to payment across all dataset presented at Figure 4.2. Visually it is similar to the Poisson distribution with peak at the 30-day mark, which reflects the concentration of payments around standard and common payment term of 30 days.

Figure 4.2. Histogram of days to payment, training dataset

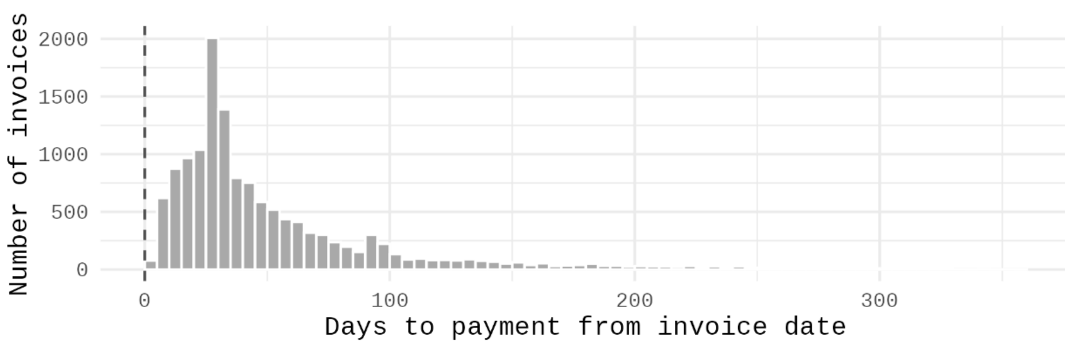


Figure 4.3. shows payment patterns for each segment in terms of actual payment date relative to the due date, which is stated in agreement, and represented by zero on the x-axis. A position on the negative side of the scale indicates late payment, which is equally common across all four segments and, at first glance, does not appear to have a strong correlation with the invoice amount.

Figure 4.3. Scatterplot of log-transformed invoice amount vs time of payment, by segment, training dataset

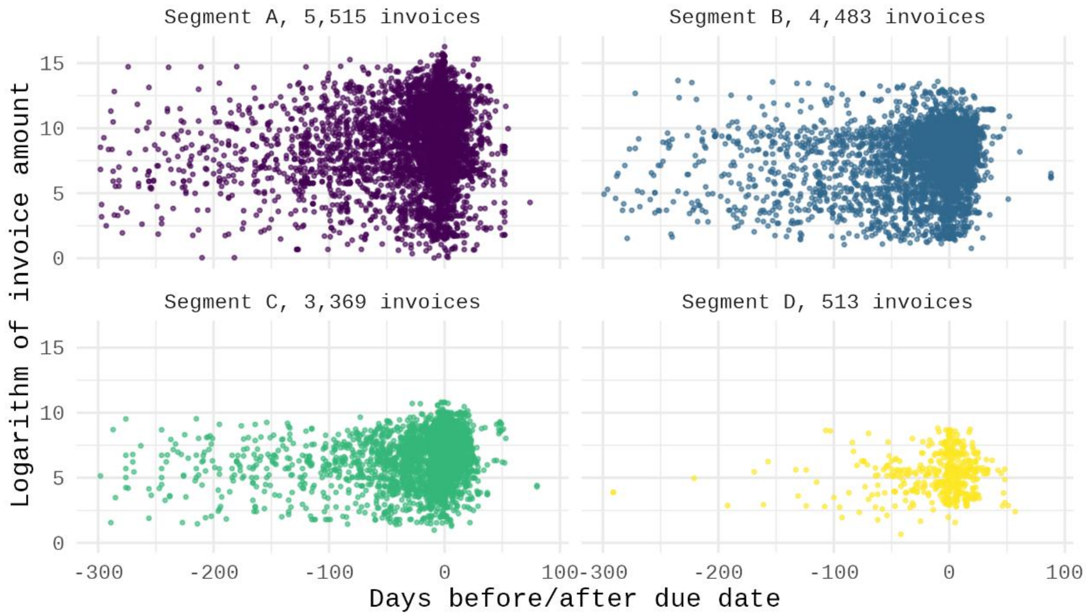
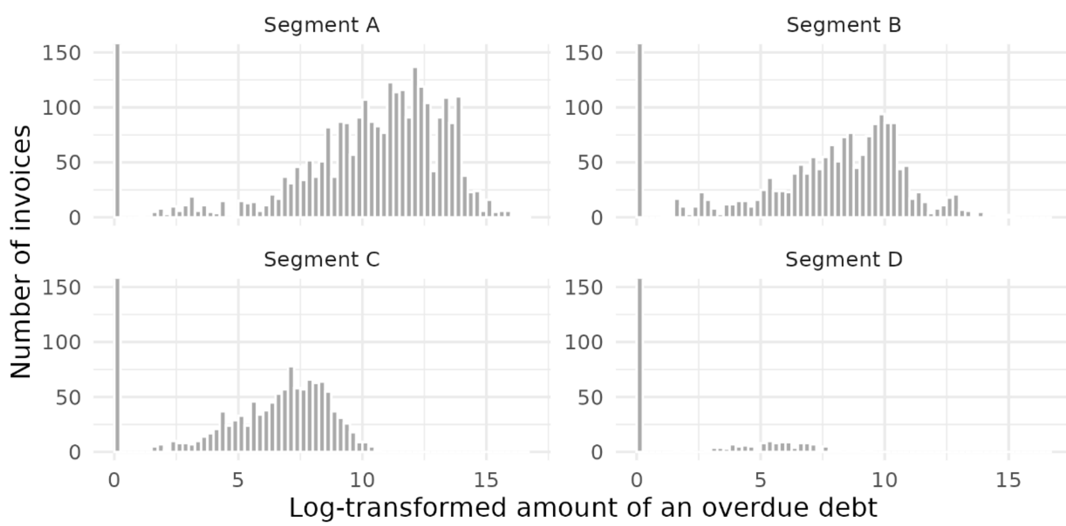


Figure 4.4. Histogram of log-transformed debt amount, by segment, training dataset



A more dramatic aspect of payment behavior is demonstrated by Figure 4.4. As can be seen, the higher the segment, the more pronounced the tendency to accumulate debt and pay it off late.

The testing dataset contains 6946 invoices issued for 920 customers. It was collected in the same way as training dataset, moreover, it includes only invoices for customers present in training dataset. It composed of paid invoices for first six months of 2025 fiscal year, it makes a perfect data to test out-of-sample prediction accuracy of the estimated models. The descriptive statistics for invoice amount and days to payment by customer segment is summarized in Table 4.2. It shows similar patterns of significantly right-skewed and peaked distribution across all dataset.

Table 4.2. Descriptive statistics for invoice amount and days to payments by customer segment, testing dataset

Segment	N	Mean	SD	Median	Min	Max	Skew	Kurtosis
Segment A	2996	122029.76	469756.43	7311.83	0.02	6467160.4	8.47	88.73
Segment B	2182	14193.68	42965.40	3121.32	1.98	684285.5	9.02	103.57
Segment C	1548	2708.87	5712.85	993.42	0.15	102989.2	8.13	105.44
Segment D	220	619.27	1062.73	222.17	0.38	5817.0	3.10	10.26

Segment	N	Mean days to payment	Median days to payment
Segment A	2996	41	51
Segment B	2182	31	46
Segment C	1548	30	43
Segment D	220	30	41

CHAPTER 5. RESULTS

5.1. Discrete survival models estimation result

In Table 5.1 we present the summary for estimated discrete survival models: coefficient estimates and standard errors . The extended table with confidence intervals included could be found in Appendix 1. (CIs) and p-values for the covariates of the model.

Table 5.1. Estimated discrete survival models summary

	Basic model		Extended 1 Log(Amount+1) * $B_m(t)$		Extended 2 Log(Amount overdue+1) * $B_m(t)$	
Covariate	CE	SE	CE	SE	CE	SE
Intercept	-6142***	0.088	-6124***	0.089	-6165***	0.086
Log(Amount+1)	0.151***	0.015	-0.201***	0.029	0.15***	0.015
Log(Amount+1) * $B_1(t)$			1847***	0.123		
Log(Amount+1) * $B_2(t)$			-2283***	0.2		
Log(Amount+1) * $B_3(t)$			2046***	0.207		
Log(Amount overdue+1) * $B_1(t)$					-0.734***	0.065
Log(Amount overdue+1) * $B_2(t)$					2278***	0.165
Log(Amount overdue +1) * $B_3(t)$					-1111***	0.136
Payment term	-0.471***	0.042	-0.466***	0.043	-0.475***	0.041
Segment B	0.249*	0.109	0.267*	0.109	0.235*	0.106
Segment C	0.376***	0.113	0.378***	0.113	0.364***	0.11
Segment D	0.75***	0.211	0.71***	0.212	0.734***	0.206
Quarter 2	0.075**	0.026	0.07**	0.026	0.076**	0.026
Quarter 3	0.047.	0.026	0.045.	0.026	0.059*	0.027
Quarter 4	0.166***	0.026	0.161***	0.026	0.184***	0.028
Spline $B_1(t)$	9916***	0.129	9945***	0.129	10005***	0.131
Spline $B_2(t)$	-5816***	0.19	-6133***	0.194	-6414***	0.199
Spline $B_3(t)$	10109***	0.169	10442***	0.175	10502***	0.181

Note: *** $p < .001$, ** $p < .01$, * $p < .05$. CE – coefficient estimate, SE – standard error

The positive sign of coefficients means higher payment hazard, i.e. earlier payment, and accordingly the negative sign implies payment delay. To interpret the numerical value of the estimated coefficients in terms of survival analysis, we calculate the hazard rate with formula $hazard\ rate = e^{\beta}$. Among all three models, the most consistent estimates are for payment terms, segment and invoice quartal. Thus, payment terms covariate is negative and highly significant, and each additional 15 days of contractual term (1SD) are associated with 38% decrease in payment hazard compared to the baseline. It aligns with findings of Wu et al. (2025) and common situation in business, when biggest clients leverage their importance to the vendor and abuse constant short-term delays in payment. Usually, it is these large customers who negotiate the longest post-payment terms. This is also evident from the values of the coefficients for C and D segments, which are associated with 45% and 105% increase in likelihood of payment in moment, conditional on non-payment so far. In terms of seasonality, invoices issued in the fourth quarter in all three models paid faster. Splines shape the baseline hazard over time so their coefficients cannot be interpreted in standalone business terms.

The sign for invoice amount covariate (Log(Amount+1)) in basic model does not match our expectations and implies positive effect of higher invoice amounts on payment hazard. In both extended models we should interpret the invoice amount effect, considering together coefficients of invoice amount and its' interaction with spline. For the first extended model, we can see that higher invoice amount accelerates payment in early and late horizon but slows it down in the middle. To some extent, the effect of the overdue debt amount in the second extended model is mirrored. High debt amount reduces immediate payment hazard on earlier stage, increasing it on mid-horizon to drop it again later. Estimates from both extended models indicate strong time-varying effects of invoice and overdue debt amount.

To test validity of estimated extended models and identify potential misspecifications, we performed a diagnostic check (Jiang and Nguyen, 2021) by calulating Pearson's χ^2 and deviance statistics on grouped data, to evaluate how well predicted probabilities match observed data, and plotting Q-Q plots for checking normality of distribution of customer-specific random effects.

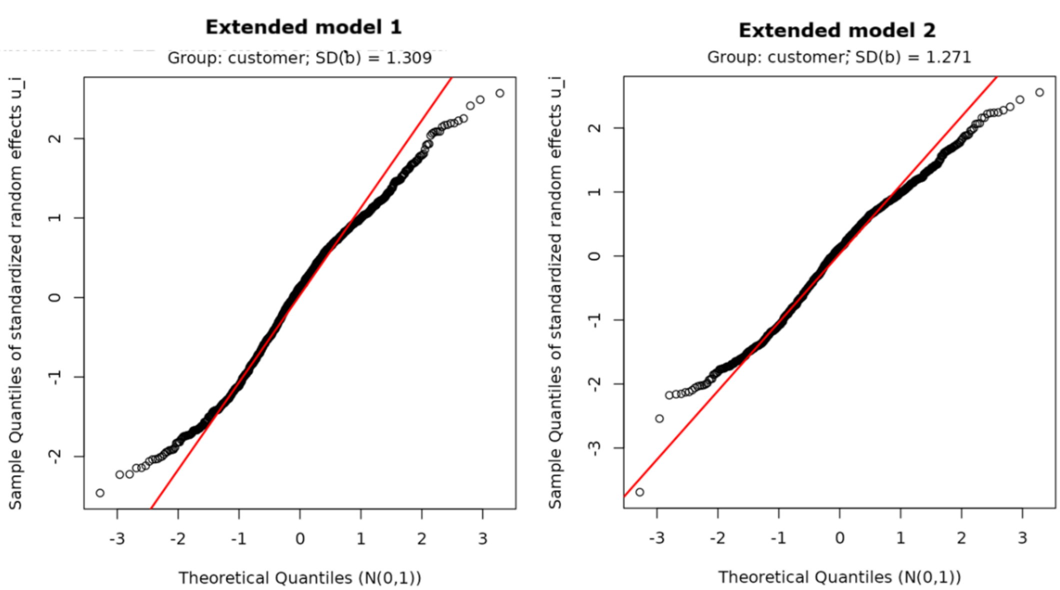
Pearson's χ^2 and deviance statistics were calculated for data grouped by customer segment and invoice quarter. Segment captures payment group's payment behavior patterns, and quartal represents seasonality. Thus, we have 16 groups (4 segments * 4 quarters, df = 16). Results presented in Table 5.2. shows that for both models' statistics is close to their respective degrees of freedom, so both models fit the grouped data well, but extended model 1 has better fit.

Table 5.2 Pearson's χ^2 and deviance statistics for extended discrete survival models (observations grouped by segment and invoice quarter)

Model	**df**	**Total Pearson χ^2**	**Total Deviance**
Extended 1 (added Log(Amount+1) * $B_m(t)$)	16	15.07	14.99
Extended 2 (added Log(Amount overdue+1) * $B_m(t)$)	16	12.96	12.91

The Q–Q plot shows how well the standardized random intercepts align with the theoretical normal distribution. Most points lie close to the diagonal, so the normality assumption holds, and slightly curved tails indicates a few extreme customers. Overall, the pattern supports the validity of the random-effects structure with minor deviations.

Figure 5.1. Q–Q plot of standardized customer-level random intercepts for extended discrete survival models



To test our hypotheses, we conduct the likelihood ratio and Wald tests for both extended models. Likelihood ratio of extended model against basic, and joint Wald test for time-varying added covariate are presented in Table 5.2. Statistics and its' significance level imply that both zero hypotheses are rejected, and including both invoice amount and overdue debt amount as time-varying covariates is statistically significant. Higher χ^2 in both Wald and likelihood ratio test shows that the improvement of model's fit is greater for extended model 1 with time-varying log invoice amount, comparing with extended model 2.

Table 5.3. Results of likelihood and joint Wald tests for extended discrete survival models

		Wald		**LRT**	
Term	**df**	**Statistics**	**p-value**	**Statistics**	**p-value**
Extended model 1					
Log(Amount+1) × $(B_1(t) + B_2(t) + B_3(t))$	3	228.3013	< .001***	217.51	< .001***
Extended model 2					
Log(Amount overdue +1) × $(B_1(t) + B_2(t) + B_3(t))$	3	193.1160	< .001***	185.94	< .001***

5.2. Models' inference and predictive performance

Charts below represent weekly payments predictions for 2025 fiscal year invoices (testing dataset). It is visually evident that discrete models give a significantly smaller spread of predicted values compared to continuous-time Cox Proportional Hazard models.

Figure 5.2. Weekly payment predictions, survival models

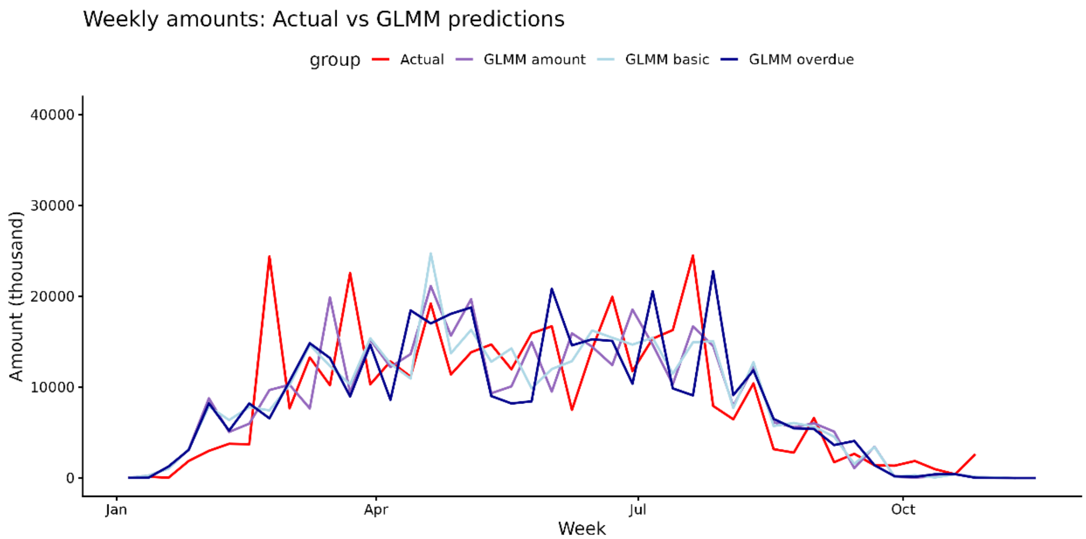
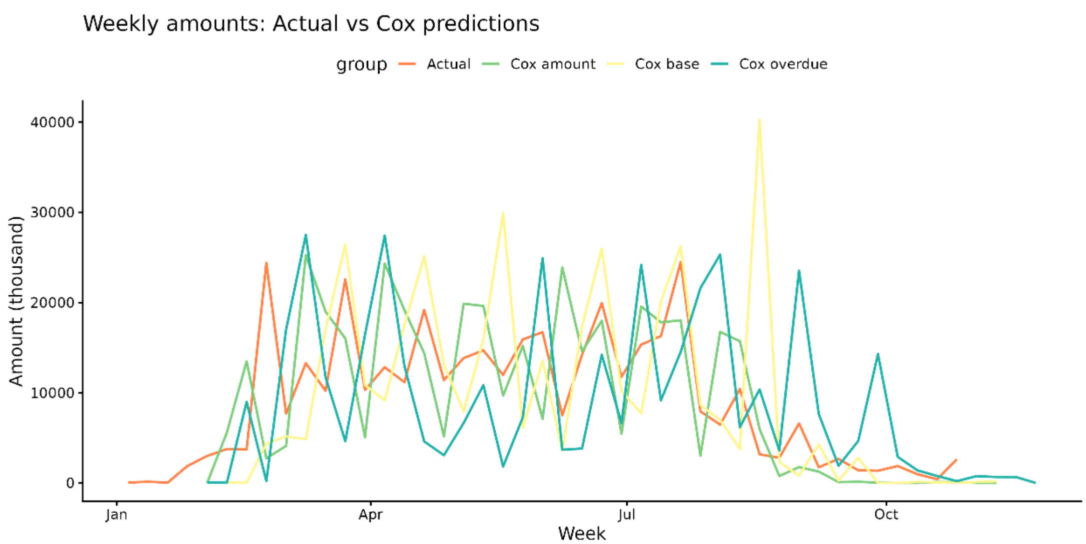


Figure 5.3. Weekly payment predictions, Cox Proportional Hazards models



To summarize predictivity accuracy of models we use next metrics:

- Mean Absolute Scaled Error (MASE) – was computed for both daily and weekly aggregated payment amounts. The closer value to 0, the better is forecast comparing with the naïve benchmark.

- Weekly win-rate – shows share of weeks when model's prediction error was smaller than the error from contractual invoice payment due date (comparing with the most primitive incoming cash-flow planning algorithm)

Table 5.4. Accuracy of predictions metrics summary

Model	**Daily MASE**	**Weekly MASE**	**Weekly win-rate**
Basic discrete survival	0,7010	0,6750	70%
Extended 1 (adding Log(Amount+1) * $B_m(t)$)	0,7660	0,7810	61%
Extended 2 (adding Log(Amount overdue+1) * $B_m(t)$)	0,8260	0,9090	61%
Basic Proportional Hazards	1,0190	0,8020	57%
Extended 1 (adding Log(Amount+1))	0,8570	1,0310	53%
Extended 2 (adding Log(Amount overdue+1))	0,8160	1,4510	50%

Basic discrete survival model shows the best scores across all three metrics. Both extended versions with time-varying interactions show worse results. All six models are more reliable on weekly predictions level. The simplest discrete survival model turns out to be the most precise when testing real data, even though likelihood ratio test showcased improvement of goodness of fit. Comparing two groups of models we can conclude that inclusion of time-varying baseline hazards and modeling for discrete time outperforms continuous Cox Proportional Hazards model within scope of our research.

5.3. Prediction accuracy for binary classification problem

Taking advantage of discrete survival analysis method, in addition to estimating time to event, we can reformulate the question as a binary classification problem and evaluate the accuracy of the models compared to logistic regression, which is the standard among statistical methods for this type of task. Using discrete survival models, fitted on 2024 year dataset, we are predicting if invoices from training 2025 year dataset will be paid before the due date by calculating cumulative probability of payment before the due date.

We also estimated logistic regression using training dataset with the same set of covariates as basic discrete survival model.

Figure 5.4. Receiver operating characteristic (ROC) curves – comparison with logistic regression

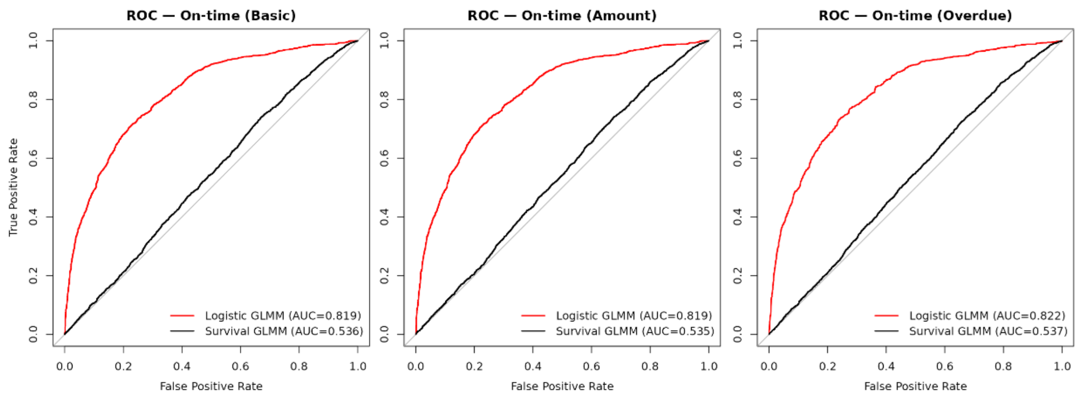


Figure 5.3 shows Receiver operating characteristic (ROC) curves for all three discrete survival models. ROC represents trade-offs of positive and false positive rates given by models. Higher value of AUC (Area Under the Curve) corresponds with better model performance, and it's visible that logistic regression is much more reliable compared to discrete survival. Despite that basic discrete survival function outperform the extended in prediction accuracy, all three functions showed precision close to coin flipping in binary classification task.

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

This research was aimed out to examine significance of controlling for invoice amount for payment time predictions within a discrete survival framework. We tested if adding a) time-varying effect of invoice amount, and b) a time-varying effect of overdue debt amount does improve the model compared to the basic specification. Both null hypotheses technically were rejected based on likelihood ratio tests, which shown that extended models provide a statistically better fit. However, predictive performance on the test dataset revealed a completely different picture: the basic discrete survival model consistently outperformed extended versions across all accuracy metrics (MASE and win rate). Using real-world data we prove one more time complexity and additional controls may improve model fit, but it doesn't necessarily to better forecasting.

We compered discrete survival models with continuous-time Cox Proportinal Hazards models. Incorporating a time-varying baseline hazard and modeling time as discrete intervals proved its suitability for capturing real-world payment dynamics. Within our research we identified that payment term length and customer segment remaining the strongest predictor of payment timing. Contrary to expectations, invoice amount showed a positive effect on payment hazard in the basic model, while extended models revealed complex time-varying patterns. Overdue debt amount exhibited a mirrored time-varying effect, reducing early payment likelihood but increasing mid-horizon hazard.

Discrete survival approach that we followed could be applied by finance and credit management teams for cash-flow forecasting, collection prioritization, and risk assessment. Businesses operating in B2B environments, especially those with large volumes of invoices and diverse customer segments, may benefit most from this approach. The model's interpretability make it suitable for integration into existing decision-support systems such as collection optimization and automation, and collection policy design.

Estimated on a real business dataset, our models could serve as a starting point for further research and improvements. Future versions of the model could incorporate additional customer-level behavioral features, exhaustive information about collection actions taken and disputes details (reasons, stage, resolution), and macroeconomic indicators as time-varying covariates. Another insufficiently explored and promising area is the combination of statistical and machine learning methods, which will allow to combine the advantages of interpretability with the power of massive and dynamic computing.

Summing up, this thesis utilize unique real-world dataset to showcase that discrete survival analysis is a powerful and versitile tool for predicting invoice payment times. It's crucial to find a proper balance between extended specifications, and simplicity, which remains key for robust forecasting performance. Hope our findings will help to bridge the gap between statistical modeling and actionable insights for real business needs.

REFERENCES

Appel Ana, Malfatti Gabriel, Cunha Renato, Cardoso Bruno, and de Paula Rogério. 2020. Predicting Account Receivables with Machine Learning. KDD MLF 2020 Virtual Conference. https://doi.org/10.48550/arXiv.2008.07363

Atradius (2025). Atradius Payment Practices Barometer Central and Eastern Europe (CEE) – 2025. https://group.atradius.com/dam/jcr:5fd3cc5c-ee13-426e-8512-807e018a60b3/payment-practices-barometer-central-and-eastern-europe-2025-en.pdf

Bellotti, T., amd Crook, J. (2009). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60(12), 1699–1707. https://doi.org/10.1057/jors.2008.130

Dirick, L., Claeskens, G., and Baesens, B. Time to default in credit scoring using survival analysis: a benchmark study. *J Oper Res Soc* 68, 652–665. 2017. https://doi.org/10.1057/s41274-016-0128-9

Fimba, A. B., and Fernandes, A. A. R. 2024. Survival Analysis Modeling On Credit Risk With Latent Variable Indicator Models. *Advanced Mathematical Models & Applications.*, *9*(1), 14–26. https://doi.org/10.62476/amma9114

Jiang, J., and Nguyen, T. (2021). Linear and generalized linear mixed models and their applications (2nd ed.). *Springer*. https://doi.org/10.1007/978-1-0716-1282-8

Kleinbaum, D. G., and Klein, M. (2016). Survival analysis: A Self-Learning Text, Third Edition. *Springer*. https://doi.org/10.1007/978-1-4419-6646-9

Liu, B., Zhang, Z., Yan, J., Zhang, N., Zha, H., Li, G., Li, Y., and Yu, Q. 2020. A Deep Learning Approach with Feature Derivation and Selection for Overdue Repayment Forecasting. *Applied Sciences*, 10 (23), 8491. https://doi.org/10.3390/app10238491

Marletta, A. (2022). The Evaluation Of Credit Risk Using Survival Models: An Application On Italian SMEs. *Statistica & Applicazioni - Vol. XIX, n. 2, 2021,* 161-178. https://boa.unimib.it/retrieve/e0f33dff-e2a2-463b-982a-c65c5ebe1695/03txt_Marletta%20%282%29.pdf

Martikainen, M. 2023. Predicting Late Payment Of Sales Invoices With Statistical Learning Methods [Master's Thesis, Lappeenranta-Lahti University of Technology LUT]. https://lutpub.lut.fi/bitstream/handle/10024/166112/Matti_Martikainen__Predicting_late_payment_of_sales_invoices.pdf

Microsoft Copilot. (2025). AI assistant used for code writing and debugging support during thesis development. Retrieved from https://copilot.microsoft.com

Moore, W. R., and Van Vuuren, J. H. (2024). A framework for modelling customer invoice payment predictions. *Machine Learning With Applications*, *17*, 100578. https://doi.org/10.1016/j.mlwa.2024.100578

Saini, S., Manai, G., van den Boom, W. et al. 2024. Invoice level forecasting with discrete survival methods for effective forecasting of account receivables in supply chain. *Discov Anal* 2, 5 (2024). https://doi.org/10.1007/s44257-024-00013-2

Smirnov, J. 2016. Modelling Late Invoice Payment Times Using Survival Analysis and Random Forests [Master's Thesis, University of Tartu]. https://dspace.ut.ee/server/api/core/bitstreams/64f3094d-4c75-42cb-9f34-84639575e7c6/content

Tutz, G., Schmid, M. 2016. Modeling discrete Time-to-Event data. *Springer*.

Wu, Jing and Lee, Hsiao-Hui and Birge, John R. (2020) The Hidden World of Trade Credit: The Flexibility Role of Late Payments. https://ssrn.com/abstract=3671400 or http://dx.doi.org/10.2139/ssrn.3671400

Zeng, S., Melville, P., Lang, C. A., Boier-Martin, I., & Murphy, C. 2008. Using predictive analysis to improve invoice-to-cash collection (pp. 1043–1050). https://doi.org/10.1145/1401890.1402014

APPENDIX

Appendix 1

	Basic model			Extended 1 Basic + Log(Amount+1) * $B_m(t)$			Extended 2 Basic + Log(Amount overdue+1) * $B_m(t)$		
Covariate	CE	SE	CI	CE	SE	CI	CE	SE	CI
Intercept	-6142***	0.088	(-6.315, -5.969)	-6124***	0.089	(-6.298, -5.951)	-6165***	0.086	(-6.334, -5.996)
Log(Amount+1)	0.151***	0.015	(0.122, 0.181)	-0.201***	0.029	(-0.258, -0.144)	0.15***	0.015	(0.121, 0.18)
Log(Amount+1) * $B_1(t)$				1847***	0.123	(1.606, 2.087)			
Log(Amount+1) * $B_2(t)$				-2283***	0.2	(-2.675, -1.891)			
Log(Amount+1) * $B_3(t)$				2046***	0.207	(1.64, 2.452)			
Log(Amount overdue+1) * $B_1(t)$							-0.734***	0.065	(-0.861, -0.607)
Log(Amount overdue+1) * $B_2(t)$							2278***	0.165	(1.955, 2.601)
Log(Amount overdue +1) * $B_3(t)$							-1111***	0.136	(-1.376, -0.845)
Payment term	-0.471***	0.042	(-0.554, -0.388)	-0.466***	0.043	(-0.55, -0.383)	-0.475***	0.041	(-0.556, -0.394)
Segment B	0.249*	0.109	(0.035, 0.463)	0.267*	0.109	(0.052, 0.481)	0.235*	0.106	(0.028, 0.443)
Segment C	0.376***	0.113	(0.155, 0.598)	0.378***	0.113	(0.155, 0.6)	0.364***	0.11	(0.149, 0.579)
Segment D	0.75***	0.211	(0.336, 1.165)	0.71***	0.212	(0.294, 1.125)	0.734***	0.206	(0.329, 1.138)
Quarter 2	0.075**	0.026	(0.024, 0.125)	0.07**	0.026	(0.02, 0.12)	0.076**	0.026	(0.024, 0.127)
Quarter 3	0.047.	0.026	(-0.004, 0.098)	0.045.	0.026	(-0.006, 0.096)	0.059*	0.027	(0.006, 0.113)
Quarter 4	0.166***	0.026	(0.114, 0.218)	0.161***	0.026	(0.11, 0.213)	0.184***	0.028	(0.13, 0.239)
Spline $B_1(t)$	9916***	0.129	(9.663, 10.169)	9945***	0.129	(9.692, 10.199)	10005***	0.131	(9.749, 10.261)
Spline $B_2(t)$	-5816***	0.19	(-6.188, -5.444)	-6133***	0.194	(-6.513, -5.752)	-6414***	0.199	(-6.803, -6.024)
Spline $B_3(t)$	10109***	0.169	(9.778, 10.441)	10442***	0.175	(10.098, 10.785)	10502***	0.181	(10.148, 10.857)

Note: *** $p < .001$, ** $p < .01$, * $p < .05$. CE – coefficient estimate, SE – standard error, CI – confidence interval