Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Probability estimation model for the cancellation of container slot booking in long-haul transports of intercontinental liner shipping services

Hui Zhao\textsuperscript{a}, Qiang Meng\textsuperscript{a,}\textsuperscript{*}, Yadong Wang\textsuperscript{b}

\textsuperscript{a} Department of Civil and Environmental Engineering, National University of Singapore, Singapore 117576, Singapore
\textsuperscript{b} School of Economics & Management, Nanjing University of Science & Technology, Nanjing 210094, China

ARTICLE INFO

Keywords:
Container liner shipping
Container slot booking
Cancellation probability estimation
Time-to-event analysis

ABSTRACT

The intercontinental liner shipping services transport containers between two continents and they are crucial for the profitability of a global liner shipping company. In the daily operations of an intercontinental liner shipping service, however, container slot bookings from customers can be freely cancelled during a booking period, which causes loss of revenue and low utilization of ship capacity. Though a pain-point of the liner shipping industry, the container slot cancellation problem has not yet been well investigated in the literature. To fill this research gap, this study aims to estimate the probability for the cancellation of container slot booking in the long haul transports of the intercontinental liner shipping service by considering the primary influential factors of cancellation behavior. To achieve the objective, a container slot booking data-driven model is developed by means of a time-to-event modeling technique. To incorporate the effect of booking region on the cancellation probability, we introduce the frailty term in the model to capture the regionality of the container shipping market. Our case study with real slot booking data shows that the developed model performs well in forecasting the loaded containers of the slot booking requests. In addition, we shed light on how the internal factors of slot booking and external factors of shipping market influence the probability of cancellation.

1. Introduction

The liner container shipping companies transport containers on the fixed shipping routes with regular schedules (Wang et al., 2014; Wang et al., 2019a). Nowadays, there are about 400 international liner container shipping services in operation carrying about 80% of the cargoes in international trade by high-capacity container ships (UNCTAD, 2018). The intercontinental liner shipping services operated by the global liner container shipping companies transport containers from one continent to another. This kind of shipping service plays a vital role in the global supply chain and global trade (Yang et al., 2018; Wang et al., 2019b). Fig. 1 depicts an intercontinental shipping service between Asia and North America operated by OOCL.

A string of container ships is deployed in an intercontinental shipping service to periodically visit the corresponding ports for container loading and discharging operations. For example, six container ships are needed to provide the weekly service frequency for the service shown in Fig. 1. A loop of all the ports in the service is called a shipping “voyage” hereafter. A ship could complete about

\* Corresponding author.
E-mail address: ceemq@nus.edu.sg (Q. Meng).

https://doi.org/10.1016/j.trc.2020.102731
Received 10 November 2019; Received in revised form 27 May 2020; Accepted 18 July 2020
Available online 8 August 2020
0968-090X/© 2020 Elsevier Ltd. All rights reserved.
Transportation Research Part C 119 (2020) 102731

8–10 voyages a year, and each voyage of these ships is numbered by a voyage code in practice. A directed link between two adjacent ports of a voyage is called a shipping “leg”, and a shipping leg connecting two ports on different continents is referred to as a long-haul leg. Asian shipping companies usually name a long-haul leg from Asia to other continents as a head haul (HH), and the opposite direction as a back haul (BH). The long-haul legs are the longest legs in terms of distance in the shipping service and generate more profit than other legs. For the sake of presentation, we call this kind of transport the “long-haul transport”. Specifically, container transport via a HH leg is called a HH transport, and that via a BH leg is called a BH transport. It should be noted that although a HH leg connects two countries, e.g., China and the USA shown in Fig. 1, it can transport containers originating from the other countries/regions owning to the container transshipment operations (Schinas and Psaraftis, 1998).

The container slot booking process is the first phase to make an intercontinental liner shipping service provided by a shipping company, as illustrated below (see Fig. 2):

Given a planned voyage of a container ship serving an intercontinental liner shipping service, container slots allocated to a port are released for customers booking about 10–56 days before the container ship departs from the port in practice. About 3–7 days before ship departure time, the shipping company closes the slot booking, loads containers, and issues necessary shipping documents to customers. A container slot booking requested by a customer (shippers/freight forwarders) includes the number of slots, origin port/destination port as well as the other necessary cargo information. The shipping company should reserve the container slots of ships for the customer according to their booking information. The customers then prepare cargoes and transport their containers to ports. However, during the slot booking period, some of the bookings can be cancelled by a customer before ships depart from an original port, resulting in the low ship capacity utilization rate and low profit. Especially for the long-haul legs, the cancellation of slot bookings can incur more loss of profit due to their higher freight rates. This study focuses on the slot booking period for a long-haul leg in an intercontinental container liner shipping service since the long-haul transport is the most profitable in the intercontinental shipping service.

It is imperative for a liner container shipping company to have a solid model that can estimate the cancellation probability of container slot booking for the long-haul legs based on historical container booking data. This is because such a model enables the shipping company to predict the number of cancelled bookings and the number of containers that could be loaded on a ship. If the number of expected loaded containers is less than the allocated slots of a ship prior to ship departure, the shipping company will continue to accept slot bookings or prompt the shipping services to fill the ship capacity. Nevertheless, several factors affect the cancellation probability of container slot bookings, such as freight rate, shipping market condition, booking time, and customer information (Zhao et al., 2019). Therefore, this study aims to develop a container slot booking data-driven model for estimating the probability that a slot booking is cancelled before the loading time by considering the influence of the attributes of the slot booking (e.g., booking time, booked volume), the container shipping market (e.g., shipping index, market volatility) and customer information. The developed model can assist the container shipping company to improve the loading rate of the ship and to understand how the internal and external influential factors affecting the slot cancellation. The developed model also lays a foundation for the emerging studies on liner shipping revenue management (Meng et al., 2019). For example, it is essential to consider the cancellation of slot bookings in pricing policy which is an important method to raise revenue in container shipping revenue management (von Westarp and Schinas, 2016). Container shipping companies implement differential pricing to customers, thus rebates can be given to customers who have lower cancellation probabilities to encourage the fulfillment of slot booking. This model can also benefit the studies of overbooking in container shipping which is a good policy to reduce the waste of ship space incurred by cancellation.

1.1. Relevant studies

There are scant studies on the cancellation of container slot booking. One reason is that container shipping companies are generally

Fig. 1. The intercontinental liner container shipping service of OOCL: Asia to North America.
conservative and less likely to share their slot booking data for academia. There are only two articles directly related to container slot booking cancellation analyses. Wang and Meng (2018) used the statistical methods to forecast the net container slot booking, defined as the difference between the booked and cancelled slots, depending on historical container slot booking information. However, their methods could not consider any influential factors to estimate the individual-level cancellation probability. Zhao et al. (2019) extracted the possible attributes/factors that influence the customer behavior of slot booking cancellation from the real container slot booking data. Their empirical study shows that the relationship between influential factors and cancellation rate is highly nonlinear and calls on a further study to analytically model the relationship. To the best of our knowledge, there are no studies dealing with the estimation of the slot booking cancellation probability although it has been highlighted and requested by several scholars (e.g., Lee et al., 2015).

There are some similarities between container slot cancellation modeling and air ticket/cargo cancellation analyses. The air ticket or air cargo cancellation behaviors have been extensively examined in literature by using different methodologies (e.g., Garrow and Koppelman, 2004; Hueglin and Vannotti, 2001; Iliescu et al., 2008; Graham et al., 2010; Chiew et al., 2017; Morales and Wang, 2010). Although the airline ticket cancellation studies could provide some enlightenment for container slot cancellation modeling, estimating the probability of cancellation in container slot booking is a new and worthy research topic because of the following exclusive features.

First, the container transport demand is generated from international trades that exhibit regional characteristics. The intercontinental liner shipping services connect multiple trade zones covering various countries with different shipping policies, shipping market conditions, import/export volume, and cargo type. In other words, the container transportation demand from different countries exhibits diverse characteristics in container slot booking and cancellation. In this regard, the container slot cancellation behavior of customers in the same market region can be correlated and affected by the regional property of trade. Thus, it is more practical to assume that some container slot bookings are correlated. However, the studies on air ticket/cargo cancellation postulate that all the bookings are independent.

Second, the container slot bookings can be freely cancelled without any penalty in practice. For the seat booking in airline, some bookings are not allowed to be cancelled or could be cancelled with a high cancellation fee. However, the container slot cancellation fee is not widely adopted in the container liner shipping industry. The competition among container shipping companies is fierce, thus charging a penalty for cancellation could lead to loss of customers. For example, few container shipping companies (e.g., Maersk Line) had attempted to charging cancellation fee for some services, but they finally withdrew this fee because other shipping companies would not follow. In addition to cancellation fees, a few third-party platforms also provide shipping companies with some methods to avoid losses caused by the cancellations. One example is New York Shipping Exchange (NYSHEX), a digital freight marketplace, acting as a neutral contract arbiter for shippers and carriers in an attempt to prevent cancelled bookings and cargo rolling. Another third-party platform, 300cubits, also launched a business that used blockchain technology to develop shipping bitcoins to resolve cancellation and rolling issues. Unfortunately, this business was eventually discontinued due to insufficient participation by shipping companies. In other words, the third-party platform solutions are still at the infant development stage. At the current practices, as cancellations by customers are quite common, the shipping companies take the overbooking strategy to hedging the cancellation risk. Once the number of show-up containers exceeds the capacity of the ship, some containers can only be postponed to the next voyage. As a consequence, it is a common practice for a customer to book with several container shipping companies to guarantee the slots. This phenomenon results in a high cancellation rate in the container liner shipping industry.

Third, there are several key factors affecting the cancellation rate of container slot booking, including the attributes of slot bookings (e.g., volume), shipping market conditions (e.g., level of shipping index), and customer information. When estimating the probability of cancellation, both the internal and external factors should be considered simultaneously. However, the studies on the cancellation in airline rarely consider the external factors of the market.
1.2. Objectives and contributions

The objective of this study is to develop a container booking data-driven model for estimating the probability that a slot booking is cancelled prior to the loading time in a long-haul leg of an intercontinental liner container shipping service. The model focuses on the slot bookings of one direction (HH or BH) in an intercontinental container shipping service. The models of two shipping directions should be separately built because container transportation demand in the two directions has different characteristics (Zhao et al., 2019). The input data are the historical data of the slot bookings (e.g., booking time, loading time, origin/destination), the data of internal influential factors (e.g., freight rates), and the external data of shipping market (e.g., freight index and user information). These data enable us to build a tangible model that can estimate the cancellation probability of container slot bookings with consideration of the aforementioned characteristics.

To achieve our objective, we develop a time-to-event model with a hazard function. Time-to-event models are a family of models to analyze the time before an interested event happens (Hosmer et al., 2011). The event in this study is the cancellation of the slot booking prior to ship departure. We define the time to cancellation and transform the estimation of cancellation probability to the estimation of the distribution of time to cancellation. To consider both internal and external factors affecting the cancellation, we use the hazard regression approach to formulate their influences on the cancellation. Additionally, to reflect the specific feature of market segmentation in container liner shipping, we introduce an interesting frailty factor to reflect the correlation of customer’s slot booking behavior from the same market. In other words, the customers within a market region (i.e. country) are assumed to have correlated behaviors.

The contributions of this study are presented. First, to the best of our knowledge, this study is the first attempt to estimate the cancellation probability of container slot booking. As pointed out by a few researchers (e.g., Zurheide and Fischer, 2012), the lack of research into the cancellation probability estimation hiders the studies on container slot allocation and revenue management in the shipping industry. To some extent, this study fills up this gap. Second, this study develops a time-to-cancellation model with origin country frailty to consider the property of the market regions. We introduce the frailty term in survival analysis to capture the correlated behavior of cancellation caused by the regionality of international shipping market, which is a novel application of the frailty theory in the container liner shipping industry. This model can assist container shipping service operators to forecast the loading container volume using the data in the slot booking period. Third, real data from a container liner company are used to calibrate and evaluate the model. Thus, the results of this study can lead us to have more practical insights. Fourth, some behavioral insights into the container shippers’ cancellation behavior are analyzed, which is helpful in deciding the customer segmentation and identifying the peak time of cancellation during the booking period.

The remainder of this study is organized as follows. Section 2 elaborates container slot booking data and the research problem. Section 3 develops the container slot booking data-driven model. Section 4 presents a case study. Section 5 provides valuable insights on the cancellation behavior of customers. Section 6 concludes this study and points out some future research directions.

2. Data description and problem description

2.1. Container slot booking and cancellation data

Let us consider an intercontinental liner shipping service with HH and back BH directions, which is operated by a global liner shipping company. A number of container ships are deployed in the shipping service to maintain a regular service frequency (e.g., weekly service). Before a deployed ship departs from a particular port, customers will book the container slots of the ship to transport their containers from a loading port to a discharging port by providing the necessary container slot booking information. Customers

| Type of data          | Description of data                                      | Example                          |
|----------------------|---------------------------------------------------------|----------------------------------|
| Booking information  | Booking code and time                                   | Code of booking request          | 010863759                       |
|                      | Booking time: time of accepting the request             | 2017-03-21                       |
|                      | Shipping service name                                   | PCS                              |
|                      | Name of the booked ship                                | SCHUBERT                         |
|                      | Code of the booked voyage                              | 020                              |
|                      | Transport direction                                     | Head Haul (Asia to North America)| 2017-05-02                      |
|                      | The departure date of the ship (from the containers’ loading port) | 2017-05-02                      |
|                      | Booked TEU: number of booked slots (TEU)               | 4 TEUs                           |
|                      | Freight rate                                            | 750 USD/TEU                      |
|                      | Loading port                                            | Hong Kong                        |
|                      | Origin country                                          | Vietnam                          |
|                      | Destination port                                        | Oakland                          |
|                      | Destination country                                     | USA                              |
|                      | Cancellation time (only for cancelled bookings or no-shows) | 2017-04-01                      |

H. Zhao et al.
can request to cancel these bookings before the ship leaves the loading port. After receiving a container slot booking or cancellation request, the information technology (IT) department of the shipping company updates its slot booking database. The container slot booking and cancellation data format, illustrated by the first three columns in Table 1, can be extracted from the database. The data example shown in the last column of Table 1 is created with respect to the shipping service in Fig. 1.

Cancellations could be incurred at any time after the shipping company accepting the slot booking and before the containers are loaded to the ship. The cancellation time, shown by the last row of Table 1, is the time when the slot booking is cancelled by the customer. These time data are recorded only if a slot booking is cancelled. There are two situations of slot booking cancellations. In the first situation, customers inform the shipping company to cancel their bookings, and the cancellation time is the time when the shipping company receives the slot booking cancellation request. In the second situation, customers just do not transport their cargo to the port before the port cut-off time, and this kind of slot bookings is “no-show” slot bookings. For the no-shows, the cancellation times are assumed to be one day before the ship departure time, which is usually the port cut-off time in practice. For the convenience of presentation, we refer to both cancelled and no-show bookings as “cancelled bookings” in this study.

2.2. Influential factors

Let us examine the influential factors of the cancellation behavior of customers. Based on the empirical study by Zhao et al. (2019), the following factors have the impact on the slot booking cancellation: the time attributes of slot booking, the origin/destination of slot booking, the volume of the booked slots (in TEUs), the attributes of the booked voyage, the customer information, and the freight rate of booking. These factors are the necessary inputs for the model development. Additionally, we include the market region of slot booking to reflect the seasonality of container liner shipping, and add the average rates and the fluctuations of shipping market index as external factors in order to reflect market conditions. The influential factors considered by this study are presented in Table 2.

Next, we further analyze the influence factors presented in Table 2 as follows:

(i) Attributes of slot booking, customers and voyage.

We choose the booking time, booked slots in TEU as well as the freight rate for the slot booking as the first type of influence factors because they have an impact on the cancellation behavior of customers (Zhao et al., 2019). The booking date in Table 1 is separated into two attributes presented in Table 2, with BK_M as the month of booking and BKTDP as the time of booking. BKTDP is measured by days prior to departure. In these two time attributes, the BK_M reflects the season of booking and the BKTDP describes how early the booking request is proposed. The second group of covariates is the booked TEU which is the number of booked TEU proposed by customers. In addition, to investigate the influence of the attribute of voyage, we use the month of the ship departure time (DP_M in Table 2) to imply the seasonality of voyage time. The customer information is also an important influential factor of the cancellation behavior. Such information includes the types of the customer (e.g., freight forwarders, manufacturers), customer size (e.g., large retailers), customer location, and relationship with the shipping company. Unfortunately, these data are confidential for shipping companies and not available for this study. To incorporate the customer influence, we use the size of the customer’s booking to describe the customer information, which is measured by the customer’s booking volume in a year (in TEU). This booking volume of customer can measure the closeness of cooperation between customers and companies, and can also be used to judge large and small customers.

(ii) Regional characteristics of container shipping.

As aforementioned in Section 1.2, the international trades of different origins have various characteristics. The regional characteristics of cargo, namely cargo origins, are the factors for the cancellation probability estimation model development. Fig. 3 gives the examples of three bookings in the HH transport of the shipping service shown in Fig. 1 to clarify the definition of regional characteristics. The origin countries of Bookings 1, 2, and 3 are China, Philippines, and Vietnam respectively. The containers of these three bookings are all transported to the USA through the long-haul voyage. Bookings 2 and 3 are first transported to the ports in China by either land transportation or feeder shipping service and then transported to the USA through the HH transport. Although they are all

| Table 2 |
| Influential factors |
| Description |
| Notation |
| Attributes of slot booking |
| The month of the booking |
| BK_M |
| Booking time (days before ship departure) |
| BKTDP |
| Booked TEU |
| BK_TEU |
| The freight rate of slot booking |
| FRT |
| Attribute of voyage |
| Voyage month |
| DP_M |
| Customer information |
| Booking volume in a year |
| CST |
| Regional characteristics |
| Cargo origin |
| ORGN_CNTRY |
| Cargo destination |
| DEST_CNTRY |
| Shipping market |
| China (Export) Containerized Freight Index (CCFI) |
| CCFI |
| The variance of CCFI |
| CCFI_VAR |
transported via the ports in China, these three bookings belong to the different international trades. To reflect this regionality, the original country of a container, rather than the loading country, is also incorporated in the model development.

(iii) Shipping market.

In addition to the voyage information and booking information, we propose the influential factors of shipping market in order to reflect the influence of a specific market environment. Two aspects of the shipping market, namely, the average and the fluctuation of the freight rates in the shipping market are included as the inputs of our model development. As it is difficult to obtain the real freight rates of all the shipping companies, in this study, we use the China (Export) Containerized Freight Index (CCFI) which is a conventional index of the freight rates of the shipping market. The latest CCFI data of slot booking is used to reflect the average freight rate of shipping market. Moreover, we calculate the variance of the last three weeks’ CCFI data (CCFI_VAR in Table 2) to imply the fluctuation of the shipping market. A high CCFI_VAR implies a fluctuating market while a low CCFI_VAR indicates a stable market. With both CCFI and CCFI_VAR, the influence from the shipping market on the cancellation probability can be analyzed.

2.3. Estimating the cancellation probability

Let us detailedly introduce the research problem. For a given intercontinental container shipping service (e.g., the shipping service in Fig. 1) and a direction of long-haul transport (e.g., HH transport in Fig. 3), the historical slot booking and cancellation data are available. These data cover both the non-cancelled and cancelled bookings of the long-haul transport of the service. The factors influencing the behavior of cancellation are listed in Table 2 and they can be extracted from the historical data and CCFI index. The aim of this study is to develop a slot booking data-driven model to estimate the probability of slot booking cancellation by considering the key influential factors.

We denote a particular container slot booking by $B$ with the given booking information. The influential factors, shown in the third column of Table 2, are grouped into vector $Z_B$, namely:

$$Z_B = [B_K, M, B_K T D P, B_K, T E U, F R T, D P, M, O R G_N, C N T R Y, D E S T, C N T R Y, C C F I, C C F I, C F I, V A R, C S T]$$

This study is to develop a model $f(\cdot \mid \cdot)$ which can estimate the cancellation probability of booking $B$ considering the influential factors $Z_B$ as follows:

$$Prob_{cl}c(B) = f(Z_B)$$

$$Prob_{nc}l(B) = 1 - Prob_{cl}c(B)$$

where $Prob_{cl}c(B)$ is the probability that booking $B$ is cancelled during the booking period while $Prob_{nc}l(B)$ is the probability that the booking $B$ is not cancelled. This model depends on the slot booking data. With the model, the shipping company can estimate the cancellation probability of the accepted slot bookings.

3. Research methodology

The empirical study conducted by Zhao et al. (2019) has concluded that the logistic regression model cannot be used to estimate the slot booking cancellation probabilities. Therefore, we adopt the time-to-event modeling approach here because it has been shown to be
effective in forecasting cancellations in literature (e.g., Iliescu et al., 2008). To do so, a frailty term should be introduced to reflect the correlations of the cancellation behavior within an origin country.

### 3.1. Time-to-cancellation model with frailty

The time-to-event models are a class of analytical methods on modeling the feature of the time before an interested event happens (Hosmer et al., 2011). This model has been widely applied in the demography, biology, and medicine industries, but rarely applied in shipping industry. In this study, the event is the cancellation of a container slot booking. Accordingly, the time to cancellation (TTC, or lifetime) of a container slot booking is defined as the time period from its booking time to the cancellation time. Based on TTC, we can transform the problem of estimating the cancellation probability to the problem of modeling the distribution of the TTC (Yu et al., 2017). Therefore, in this study, the time-to-cancellation model (TTCM) aims at formulating the distribution of the TTC. We have two assumptions for the TTCM:

(i) Every booking has the possibility of being cancelled or not showing.

(ii) The cancellation probability of a booking is the probability that the TTC of the booking is less than the days from booking time to ship departure time. In other words, if the slot booking can “survive” to the ship departure time, it will not be cancelled.

We formulate the TTC as a random variable $T$ with the Probability Density Function (PDF) $f(t)$. According to the assumption (ii), the cancellation and non-cancellation probabilities of a booking $B$ with booking time $t_b$ can be calculated by

\[
\text{Prob}_{\text{clc}}(B) = P[T < t_b]
\]

\[
\text{Prob}_{\text{noclc}}(B) = P[T \geq t_b]
\]

There are two key functions in time-to-event analysis: survival function denoted by $S(t)$ and hazard function denoted by $\lambda(t)$. These two functions are elaborated in Appendix A. There is a natural relationship between PDF, hazard function, and survival function (Hosmer et al., 2011):

\[
\lambda(t) = \frac{f(t)}{S(t)}
\]  

(1)

Therefore, modeling the distribution of $T$ can be converted to the modeling of $\lambda(t)$ which implies the instantaneous rate of occurrence of the cancellation. The deducing of this relationship can be found in Appendix A.

We formulate the hazard function $\lambda(t)$ as the Cox proportional hazard function according to the Cox hazard regression method (Kaplan, 1977). The Cox regression method provides an effective and tractable form of hazard function to incorporate the covariates. This regression method assumes that there is a baseline hazard function and the effect of the covariates is proportional to this baseline hazard. Assume that the characteristics of an individual booking can be described as a vector of covariates $Z$. According to the Cox regression method, the hazard function is assumed to have the form:

\[
\lambda(t|Z) = \lambda_0(t)e^{\beta Z}
\]  

(2)

where $\beta$ is the vector of coefficients; $\lambda_0(t)$ is the baseline hazard function which is the hazard function when $Z = 0$. As expressed by Eq. (2), the hazard function $\lambda(t)$ is proportional to the baseline hazard function $\lambda_0(t)$, and the proportion is determined by the covariates $Z$ and coefficients $\beta$. $\lambda_0(t)$ can be specified in various analytical forms.

In order to consider the regionality of shipping market, we introduce the term “frailty” based on this classical hazard regression model. The frailty approach is a statistical modeling concept which aims to account for heterogeneity caused by unmeasured covariates from the clustered data (Li and Fan, 2002). In statistical terms, a frailty model is a random effect model for time-to-event data, where the random effect (i.e. the frailty) has a multiplicative effect on the baseline hazard function. Individuals in a cluster (origin country in this study) are assumed to share the same frailty, which reflects the behavioral correlations of the individuals.

In this study, the frailty term from the origin country $i$ is assumed to be $w_i$, and it is further assumed that the individuals in the same origin country share an identical frailty term. Thus, the hazard function of the individual slot booking $j$ from an origin country $i$ can be formulated by

\[
\lambda_0(t) = w_i \lambda_0(t)e^{\beta Z_i}
\]  

(3)

where $Z_i$ is the covariate vector containing the values of the influential factors. The corresponding survival function of a slot booking $j$ from origin country $i$ can be written as follows:

\[
S_j(t) = e^{-\Lambda_j(t)}
\]  

(4)

where $\Lambda_j(t) = \int_0^t \lambda_j(x)dx$ is the cumulative hazard function. In Eq. (3), the baseline hazard function $\lambda_0(t)$ reflects the shared feature of the bookings among all origin countries, while frailty term $w_i$ is a weight to reflect the influence from different origin countries (i.e. regionality). The $w_i$ is usually considered as the unobserved effect from the origin countries and assumed to be independent identically distributed random variables that follow a certain distribution. Conventionally, a gamma distribution with $E(w_i) = 1$ and $\text{Var}(w_i) = \theta$ is
assumed for computational convenience (Hougaard, 1995). The variation of frailty term \( \theta \) indicates the variability between different origin countries. It is also a useful statistic to rank the cancellation risk of various container shipping markets and assist shipping companies to make variant revenue management strategies for different market sections. The baseline hazard function \( \lambda_0(t) \), the parameter \( \theta \), and the coefficients \( \beta \) are all to be estimated in the TTC analysis model in this study.

The shape of \( \lambda_0(t) \) depends on the characteristics of the container slot booking and cancellation data. The form of baseline hazard function \( \lambda_0(t) \) is flexible, and can be specified as a non-parametric form or parametric form. This study uses both the two forms of baseline hazard functions to compare their performances. In the parametric form of baseline hazard function, it is conventional to assume that the TTC follows a Weibull distribution with two parameters, and the shape of the baseline hazard function is determined by the Weibull distribution. However, the baseline hazard function \( \lambda_0(t) \) that fits the slot booking data may have a complicated shape that cannot be captured by a given form of function. To formulate a more flexible shape of \( \lambda_0(t) \), we use the spline hazard functions as the non-parametric forms of \( \lambda_0(t) \). The spline function is capable of approximating complex shapes with a piecewise polynomial function. The spline function is formulated as a linear combination of basis functions which have several forms such as B-spline and M-spline basis functions. In this form of baseline hazard function, the parameters to be estimated are the coefficients of the linear combination. The formulations of the three baseline hazard functions are shown in Appendix B.

3.2. Maximum log-likelihood method to estimate the parameters

The maximum log-likelihood method is used to estimate parameter of the baseline hazard function \( \lambda_0(t) \), denoted by \( \mu \), the coefficients \( \beta \), and parameter \( \theta \) in the TTCM. Under the above two assumptions, the cancelled bookings and non-cancelled bookings contribute differently to the likelihood function since they provide different information about \( T \).

Fig. 4 depicts four container slot bookings (i.e. B1, B2, B3, and B4) that belonged to the same voyage and to be loaded on the same port. The horizontal axis is the time axis with time 0 means the time when the ship departures from the loading port. All the booking/cancellation time is measured as the number of days before the ship departure time 0. The lines (B1 to B4) in Fig. 4 show the observable “lifetime” (i.e. TTC) of the slot bookings. The start point of each line is the booking time, and the end point is either the cancellation time for cancelled booking or the loading time for the loaded booking. In the four slot bookings, B1 and B3 were cancelled before ship departure, and B2 was a no-show booking. B4 was a non-cancelled booking. The booking time and/or cancellation time of each booking are shown in the table of Fig. 4. As illustrated by Fig. 4, the observed lifetimes of B1, B2, and B3 are 7 days, 8 days, and 6 days respectively, which is exactly the TTC of these bookings. However, not all the bookings can provide the full information of the lifetime. For the bookings that are finally loaded onboard (i.e. B4), what we know is that its lifetime is no < 7 days, the time to loading (TTL), and it survives until the ship departure time 0.

Therefore, the likelihood contributed by cancelled and non-cancelled bookings should be formulated separately. We will first discuss the likelihood function of a general Cox hazard regression model formulated in the form of Eq. (2), and then extend it to the likelihood function for the model with frailty term \( w_i \). We denote the observed TTC/TTL of a slot booking \( j \) by \( t_j \). \( t_j \) is the time to cancellation for cancelled bookings (e.g., B1, B2, and B3 in Fig. 4), and time to loading for non-cancelled bookings (e.g., B4 in Fig. 4). For the cancelled bookings, the contribution of likelihood is the probability that the cancellation happens at the \( t_j \) days after booking, which can be formulated as follows:

![Fig. 4. Four examples for container slot booking and/or cancellation times.](image-url)
The non-cancelled bookings provide the information that no cancellation happens before the ship departure time. Thus, the likelihood contributed by a non-cancelled booking is one minus the cancellation probability denoted by \( P[T < t_i] \), namely:

\[
L_i = 1 - P[T < t_i] = S(t_i)
\]

Thus, the likelihood function of the whole sample set (i.e. both cancelled and non-cancelled) can be formulated by

\[
L = \prod_{i \in J} L_i(t_i)^{a_i} S(t_i)
\]

where \( a_i \) is the cancellation indicator which equals to 1 if the booking is cancelled and 0 otherwise, and \( J \) is the full sample set of the historical slot booking data.

In the case of the model formulated in Section 3.2, it is assumed that there are \( n \) origin countries denoted by \( i = 1, 2, \ldots, n \), and an individual booking in origin country \( i \) is denoted by \( j \) with \( j = 1, 2, \ldots, J_i \) where \( J_i \) denotes the number of bookings in origin country \( i \). As there are \( n \) origin countries, we start from the partial likelihood function of an origin country \( i \). According to Eq. (7), the partial likelihood of the slot bookings from origin country \( i \) can be expressed by

\[
L_i = \prod_{j=1}^{J_i} \lambda_{ij}(t_i)^{a_{ij}} S_{ij}(t_i)
\]

where \( \lambda_{ij} \) is the baseline hazard function. The detailed formulation of Eq. (9) is shown in Appendix C.

Based on Eq. (8), the full likelihood of the sample data, including slot bookings from all the origin countries, is the product of all the partial likelihoods. The logarithm of the full likelihood function is given by

\[
\ln (\beta, \theta, \mu) = \ln \left( \prod_{i=1}^{n} L_i \right)
\]

where \( \mu \) is the vector of parameters in the baseline hazard function \( \lambda_0(t) \). The detailed formulation of Eq. (9) is shown in Appendix C. The coefficients \( \beta \), the parameter for frailty term \( \theta \), and parameters \( \mu \) are all to be estimated. Thus, the maximum likelihood estimation problem in this study can be defined as follows:

\[
\{ \beta^*, \theta^*, \mu^* \} = \arg \max_{\beta, \theta, \mu} \ln (\beta, \theta, \mu)
\]

where \( \beta^* \), \( \theta^* \), and \( \mu^* \) are the optimal solutions. Eq. (10) is an unconstrained non-linear optimization problem. In the literature on time-to-event models, a conventional algorithm to solve this maximum log-likelihood problem is the Levenberg–Marquardt algorithm (Hosmer et al., 2011). However, this algorithm finds only a local minimum, which is not necessarily the global minimum. Therefore, it is interesting to use a global optimization technique to find a better solution for the maximum log-likelihood problem. Under this consideration, we choose the basin-hopping algorithm to solve the maximum log-likelihood problem. The basin-hopping algorithm is a stochastic global optimization method that tries to find the global minimum of an objective function. The idea of this algorithm is to search the global minimum around a local minimum. In each iteration, this algorithm first chooses a starting point, completes a local search the global minimum around a local minimum. In each iteration, this algorithm first chooses a starting point, completes a local search the global minimum around a local minimum. In each iteration, this algorithm first chooses a starting point, completes a local search the global minimum around a local minimum.

With the parameters estimated by the maximum log-likelihood method, the hazard function of TTCM can be expressed by

\[
\lambda_{ij}(t|\beta^*, \theta^*, \mu^*) = \hat{\lambda}_i(t|\mu^*) e^{\beta^* T_{ij}^{\theta^*}}
\]

where \( \beta^* \), \( \theta^* \), and \( \mu^* \) are the parameters obtained by solving the problem, and \( \hat{\lambda}_i \) is the estimated frailty term. Using the estimated \( \theta^* \), the frailty term \( \hat{\lambda}_i \) can be estimated using a Bayes method (Rueten-Budde et al., 2019).

With this estimated hazard function, the probability that a booking is cancelled before the ship departure is equivalent to the probability that the cancellation event incurred before the departure date, namely:

\[
P[T_{ij} < t_i^0] = 1 - S_{ij}(t_i^0|\beta^*, \theta^*, \mu^*)
\]

where \( t_i^0 \) is the booking time of a set slot booking \( j \) from origin country \( i \), measured as the days before ship departure, and \( S_{ij}(\cdot) \) is the survival function. \( T_{ij} \) is the random variable to describe the TTC of the booking.

It should be noted that although this study only considers the covariates listed in Table 2, the methodology is a general one and can be used for various data sets. More factors can be incorporated in the model development by extending the covariate vector \( Z_{ij}(\cdot) \) and re-estimating coefficients \( \beta \), e.g., such customer information as customer types, relationships, and sizes can be included in \( Z_{ij}(\cdot) \) of TTCM.
4. Case study

We use the real slot booking data of an international container shipping company to evaluate the performance of the developed model. The data are from the head-haul (westbound) transport of the intercontinental container shipping service from Asia to Europe as shown in Fig. 5.

Table 3 gives an overview of the data collected for the case study. It can be seen that the data cover the information of slot bookings from May 10, 2016 to April 13, 2017, and contain 15,684 slot booking requests with about 26% of them cancelled or failed to show up.

4.1. Covariates and frailty terms

The notation of the covariates and data used in the model of this case study are summarized in Table 4: 

\( t_{ij} \) in Table 4 is the lifetime of the slot booking in the historical data which is calculated as the difference between the slot booking time and cancellation time. \( \alpha_{ij} \) is a binary parameter to indicate whether the booking is cancelled (\( \alpha_{ij} = 1 \)) or not (\( \alpha_{ij} = 0 \)). For this case study, there are 10 origin countries listed in Table 3. Accordingly, there are 10 frailty terms \( w_i \), which are assumed to be i.i.d. random variables following the gamma distribution as discussed in Section 3. In addition, to assess the effect of frailty terms, we also consider a model without frailty term but treat the origin country as indicator covariates in \( Z_{ij} \).

The CCFI data used in this case study from May 10, 2016 to April 13, 2017 are shown in Fig. 6.

4.2. Model comparison and parameter estimation

To choose the most significant covariates and the best form of the baseline hazard function, we test the TTCM with all the covariates shown in Table 4, and various combinations of baseline hazard functions and frailty term. We randomly select 75% of the 15,684 slot booking data to calibrate models and others for evaluating their performances.

Table 5 lists the covariates and their significance testing results, baseline hazard functions, log-likelihood, and LCV/AIC results of the six models under different combinations of baseline hazard function forms and frailty term.

All the possible models are calibrated using the same set of training data and compared by the criteria of LCV/AIC. The smaller LCV/AIC indicates a better model performance (Commenges et al., 2007). As shown by Table 5, the LCV/AIC for the models with frailty are all smaller than the corresponding models without frailty. Models 1, 3, and 5, which have origin countries as frailty terms, perform better than Models 2, 4, and 6, considering origins as covariates when evaluated by either log-likelihood or LCV/AIC criteria. Therefore, the TTCM with frailty factors developed in this study performs better than the conventional proportional hazard regression models used by the existing studies on the forecasting of cancellations. This implies that the frailty model could effectively reflect the correlations of the bookings within an origin country, which is the special market feature in container shipping. For the form of the baseline hazard function, Model 1 with M-spline baseline hazard function outperforms Model 5 with Weibull form and Model 3 with B-spline form. Hence, we select Model 1, TTCM with frailty factor and M-spline baseline hazard function, to estimate the probability of cancellation in this case study.

For the selected model, the significance test results are shown in the second column of Table 5. For the TTCM with M-spline function and frailty factor, as shown by the first row of Table 5, the voyage month (i.e. DP_M) and booking month (i.e. BK_M) are insignificant at the significant level of 0.05. Considering the correlations between DP_M and BK_M, we eliminate the insignificant covariate, DP_M, from the model. For covariate BK_M, which is a set of 12 indicator variables from January to December, we test the
significance of the 12 indicators and find that January is the only significant indicator variable in all the months. Thus, the indicator variable "booked in January" is selected as a covariate in the model with notation BK_M_Jan. Finally, we select the most significant variables (i.e. BK_M_Jan, BKTDP, BK_TEU, CCFI_VAR, CCFI, FRT, and CST) of the Model 1 in Table 5, and recalibrate the Model 1 with these significant covariates.

Before the recalibration, we validate that the proportional hazard regression model can be used to model the data in this study. We first test the assumption of proportional hazard (PH). Readers interested in the statistical test of PH assumption are referred to the study by Grambsch and Therneau (1994). For all the covariates, only the freight rate cannot pass the test of PH assumption. Hence, the coefficient of freight rate is interpreted as "average effects" of the covariates (Hinkle et al., 2003). This average effect is sufficient to estimate the probability of cancellation, as we are only concerned with whether the TTC exceeds the ship’s departure time and not with its change over time. The freight rate is kept as a covariate in the model since its coefficient is significant. The second test is the correlation of the covariates because high correlation can cause multicollinearity in survival model. Table 6 shows the correlation coefficients of the covariates. All the correlation is negligible except for that between BK_M_Jan and CCFI with a moderate positive correlation at 0.503. This is not a high correlation according to the rule of thumb for interpreting the size of a correlation coefficient (Hinkle et al., 2003). As the correlation is moderate and both of the two factors are statistically significant, we decide to keep these two

![Image](image-url)
Table 5
Comparison of different baseline hazard functions and frailty models.

| Model number | Covariates and significance | Baseline hazard | Frailty term | Log-likelihood | LCV/AIC |
|--------------|-----------------------------|-----------------|--------------|----------------|---------|
| 1            | DP_M + BK_M + DEST_CTRY + BKTDP* + BK_TEU* + CCFI_VAR* + CCFI* + FRT* + CST* | M-splines | YES | −13264 | 1.614 |
| 2            | DP_M + BK_M + DEST + BKTDP* + BK_TEU + ORGN_CTRY + CCFI_VAR* + CCFI* + FRT* + CST* | M-splines | NO | −21526 | 1.689 |
| 3            | DP_M + BK_M + DEST + BKTDP + BK_TEU + CCFI_VAR + CCFI* + FRT + CST | B-splines | YES | −13669 | 1.624 |
| 4            | DP_M + BK_M + DEST + BKTDP + BK_TEU + ORIGIN_COUNTRY_CD + CCFI_VAR* + CCFI* + FRT + CST | B-splines | NO | −22642 | 1.698 |
| 5            | DP_M + BK_M + DEST + BKTDP* + BK_TEU* + CCFI_VAR* + CCFI* + FRT* + CST* | Weibull | YES | −13562 | 1.661 |
| 6            | DP_M + BK_M + DEST + BKTDP + BK_TEU + ORIGIN_COUNTRY_CD + CCFI_VAR* + CCFI* + FRT* + CST | Weibull | NO | −22862 | 1.705 |

Frailty: YES, if origin countries as frailty term; NO, if origin countries as indicator variables in covariates. *significant at level 0.05.

Table 6
The correlation coefficients of covariates.

| Correlation coefficient | CCFI | CCFI_VAR | BK_TEU | FRT | BK_M_Jan | BKTDP | CST |
|-------------------------|------|----------|--------|-----|----------|-------|-----|
| CCFI                    | 1.000| 0.264    | 0.018  | 0.184| 0.503    | 0.052 | −0.005|
| CCFI_VAR                | 1.000| 0.013    | 0.057  | 0.173| 0.052    | 0.010 | 0.096|
| BK_TEU                  | 1.000| −0.145   | 0.061  | 0.093| 0.093    | 0.096 | 0.096|
| FRT                     | 1.000| 0.126    | 0.089  | 0.121| 0.093    | 0.096 | 0.096|
| BK_M_Jan                | 1.000| 0.020    | 0.004  | 0.019| 0.019    | 0.019 | 0.019|
| BKTDP                   | 1.000| 0.037    | 0.093  | 0.093| 0.093    | 0.096 | 0.096|
| CST                     | 1.000|          |        |      |          | 0.037 |       |

Table 7
Estimated coefficients of the selected model (TTCM with frailty and M-spline baseline hazard function).

| Covariates     | coef | exp(coef) | P-value |
|----------------|------|-----------|---------|
| CCFI*          | −0.105| 0.900     | <0.001  |
| CCFI_VAR*      | 0.057 | 1.059     | <0.001  |
| BK_TEU*        | −0.021| 0.979     | 0.019   |
| FRT*           | 0.038 | 1.039     | <0.001  |
| BK_M_Jan*      | 0.208 | 1.232     | 0.0046  |
| BKTDP*         | −0.025| 0.975     | <0.001  |
| CST*           | −0.013| 0.987     | <0.001  |
| θ              | 0.702 |          | 0.0121  |

log-likelihood −13387.89
Spline form M-spline, 20 knots, 3 degree

*Significant at the level of 0.05.

Fig. 7. Expected loaded TEU, real loaded TEU, and total booked TEU from different origin countries.
factors. It should be noted that the estimated cancellation probability could be sensitive to the input data. The tests of significance and correlation should be conducted whenever the input data are updated, and the most effective covariates are selected.

We use the basin-hopping algorithm to solve the maximum log-likelihood problem, and the estimated parameters/coefficients are presented in Table 7. The implications of these parameters will be discussed in Section 5.

Herein, we will first use the calibrated parameters shown in Table 7 to validate the developed model in estimating the loaded containers for the container shipping service. We use the predicted expected number of loaded TEUs, which is the product of cancellation probability and booked TEUs, to evaluate the performance of TTCM (Cox 1972). The probability that the booking is cancelled before departure date is estimated by the model with parameters in Table 7. According to Eq. (12), the expected loading TEU of slot booking $i$ of origin country $j$ can be calculated by

$$ EL_{ij} = \left(1 - P[T_{ij} < t_{b_{ij}}]\right) \cdot B_{ij} = S_{ij}(T_{ij}, \hat{\beta}, \hat{\theta}, \hat{\mu}) \cdot B_{ij} $$ (13)

where $B_{ij}$ is the booked TEU of slot booking $j$ of origin country $i$; $1 - P[T_{ij} < t_{b_{ij}}]$ is the probability that the slot booking is not cancelled, and $EL_{ij}$ is the expected loaded TEU of this booking; $Z_{ij}$ is the vector of covariates of the booking; $\hat{\beta}$, $\hat{\theta}$, and $\hat{\mu}$ are the estimated parameters in the TTCM.

We use the testing data, which takes 25% of the slot booking data, to forecast the expected loaded TEU. To illustrate the estimated loaded containers, the slot bookings are aggregated by country as shown in Fig. 7 and by voyage as shown in Fig. 8.

We compare the estimated loaded TEU with the real loaded and booked TEU in these two figures, and the error rate is shown by the lines in Figs. 7 and 8. This error rate is calculated by

$$ \text{Error Rate}_k = 1 - \frac{\sum_{(i,j) \in G_k} EL_{ij}}{\sum_{(i,j) \in G_k} RL_{ij}} $$ (14)

where $k$ are the indexes of groups (i.e. origin countries or voyages), and the $G_k$ is the set of bookings that belong to a certain group $k$; $EL_{ij}$ is the expected loaded TEU estimated by the TTCM; $RL_{ij}$ is the real number of loaded TEU of the slot booking.

As shown in Fig. 7, in the predicted loaded TEU of the 10 countries, 9 countries have error rates <20%, and only one country (i.e. AU) has a high error rate. The average error rate is about 10.6%. Similarly, the average error rate of the estimated loaded TEU of the 46 voyages in Fig. 8 is about 12.6%, which shows the satisfactory prediction power of the TTCM. In other words, the test data show that the developed model is effective in estimating the loaded TEU for shipping companies in practice.

Fig. 8. Expected loaded TEU, real loaded TEU, and total booked TEU from different voyages.
5. Analysis of influential factors

In this subsection, we will discuss the implications of the estimated parameters shown in Table 7, and analyze the effects of the influential factors.

5.1. Cancellation risk of origin countries

The frailty terms are the effective indicators of the influence of origin countries on the cancellation probability (Rutten-Budde et al., 2019). An origin country with a higher $w_i$ indicates a high risk of cancellation of the slot bookings. Table 8 shows the Bayesian estimations of frailty terms of the 10 origin countries/regions.

With the estimated frailty values, the rank of the origins in terms of the slot booking cancellation risk can be obtained: the estimated frailty term of AU is the smallest in Table 8, indicating that the slot bookings from AU are less likely to be cancelled than other origins; HK shows the highest estimated frailty at 4.811, and the slot bookings from HK are at the highest risk of being cancelled than that from other origins. The estimated frailty of AU, JP, CN, TH, ID are all below 1 implying that if a slot booking is from these origins, the risk of being cancelled will be reduced when other influential factors are the same. In comparison, the frailty values of origins PH, SG, BD, VN, and HK are all larger than 1, indicating an increase in the risk of slot booking cancellation. The rank of cancellation risk can assist the container shipping companies to adjust their strategies on customer management and to pay more attention to the origins with a high risk of slot booking cancellation.

5.2. External influential factors

The effects of the covariates (i.e. the influential factors) can be evaluated by the coefficient and hazard ratio (HR) (Cox 1972). The hazard ratio $hr = e^\beta$ is the increasing ratio of hazard when raising the covariate which has the coefficient $\beta$. This ratio can reflect the proportional change of the hazard when increase the corresponding covariate by one unit. The hazard ratio of the covariates is shown in the second column of Table 7. For the external factors, the coefficient of CCFI is negative, which means that the bookings are less likely to be cancelled when the market shows a high freight rate. According to the hazard ratio 0.9, an increase of 100 on the index will decrease the hazard rate by about 10% when other factors stay the same. On the other hand, high variance of CCFI (i.e. CCFI_VAR in Table 7) indicates a higher risk of cancellation with HR at 1.059.

5.3. Internal influential factors

For the internal covariates in Table 7, the freight rate shows a coefficient at 0.038 and HR at 1.039. This means that the shipping company should consider the increase in cancellation if they decide to raise freight rates. Customer’s booking volume shows a negative effect indicating large customers are less possibly to cancel their bookings, which is reasonable in practice. The booked TEU also shows a negative coefficient at $-0.021$, indicating a decrease of cancellation hazard for bookings with larger booked TEU. However, for the slot booking data in this case study, this impact is insignificant in practice. Fig. 9 shows the histogram of the BK_TEU data, and more than 90% of the bookings booked less than 5 TEUs. Therefore, the range of the booked TEU is limited, and the change of cancellation risk caused by different number of booked TEU is not as significant as the other attributes.

For the internal covariates on time and season, the indicator variable of January (i.e. BK_M_Jan in Table 7) has a positive coefficient and HR at 1.232, which means the hazard of cancellation in January is relatively increased by 23% compared with other months. The other internal covariate on time is the booking time (i.e. BKTDP in Table 7), but its effect cannot be directly observed from HR since the conception of booking time depends on the increasing/decreasing intervals of $P_b(t_b)$. The probability of a slot booking $b$ will be cancelled before ship departure can be calculated by

$$P_b(t_b) = P_y[T_b < t_b] = 1 - \left[ \exp \left( -\hat{w}_i \bar{\Lambda}_0(t_b) \right) \right] \exp(\hat{\beta}_{\text{BK}} \cdot t_b + \Delta)$$

(15)

$$\Delta_i = \beta^{\top} \cdot Z_b - \beta_{\text{BK}} \cdot t_b$$

(16)

where $\bar{\Lambda}_0(t) = \int_0^t \lambda_0(x) \, dx$ is the cumulative baseline hazard function, $\hat{\beta}_{\text{BK}}$ is the estimated coefficient of booking time, $Z_b$ is the covariate vector of booking $b$, and $\Delta$ is the product of the covariates and their coefficients other than $\beta_{\text{BK}} \cdot t_b$. Accordingly, the probability of non-cancellation is $1 - P_b(t_b)$. The influence of booking time depends on the increasing/decreasing intervals of $P_b(t_b)$. That is, the earlier the booking is proposed in the decreasing interval the less possible it will be cancelled. However, in the increasing interval, earlier bookings are more likely to be cancelled than late bookings. Thus, we can analyze the relationship between booking time and cancellation probability by taking the derivative of $P_b(t_b)$, namely:
\[
d\frac{P_b(t_b)}{dt_b} = \hat{v}_i \cdot g(t_b) \cdot \exp\left[\hat{\beta}_{BK} \cdot t_b + \Delta_i - \hat{v}_i \Lambda_0(t_b) \cdot \exp\left(\hat{\beta}_{BK} \cdot t_b + \Delta_i\right)\right]
\]
(17)

\[
g(t_b) = \lambda_0(t_b) + \hat{\beta}_{BK} \Lambda_0(t_b)
\]
(18)

where the first and third parts of the derivative function are positive, but whether \(g(t_b)\) is positive or not depends on \(\hat{\beta}_{BK}\). If \(\hat{\beta}_{BK} > 0\), the probability of cancellation before ship departure is strictly increasing with booking time \(t_b\); otherwise the change of probability depends on the baseline hazard function \(\lambda_0(t_b)\) and cumulative baseline hazard function \(\Lambda_0(t_b)\). In this case study, the results in Table 7 shows \(\hat{\beta}_{BK} = -0.025\), thus the increasing/decreasing intervals of \(P_b(t_b)\) should be obtained by \(\lambda_0(t_b)\) and \(\Lambda_0(t_b)\). Using the baseline hazard function calibrated in this case study, we find \(g(t_b)\) is negative when \(58 < t_b < 61\). Therefore, the decreasing interval of function \(P_b(t_b)\) estimated by this study is \(t_b \in (58, 61)\), and \(P_b(t_b)\) is increasing in other intervals.

To illustrate the relationship between booking time and cancellation probability, we give the values of the other covariates and

| Country | AU | JP | CN | TH | ID | PH | SG | BD | VN | HK |
|---------|----|----|----|----|----|----|----|----|----|----|
| Estimated frailty term | 0.265 | 0.437 | 0.517 | 0.580 | 0.855 | 1.047 | 1.363 | 1.423 | 1.911 | 4.811 |

Table 8
Estimated frailty terms for origin countries/regions.

Fig. 9. Histogram for the booked TEU data of the slot bookings.

Fig. 10. Relationship between booking time (days from booking time to departure time) and cancellation probability: an example.
show the slot booking cancellation probability at different booking time in Fig. 10. The other covariates are set as the average value in the slot booking data in this case study. The origin country is Vietnam and the values of other covariates are shown in Fig. 10.

We use the coefficients in Table 7 to calculate the cancellation probability. The decreasing interval of \( P_b(t_b) \) is small enough (i.e. 3 days) to ignore in the practice of slot booking management, and generally the bookings proposed early have a higher probability of cancellation which can be calculated by \( P_b(t_b) \). As shown in Fig. 10, the cancellation probability will sharply drop if the slots are booked in about 10 days before ship departure. In the time interval from 45 to 65 days before ship departure, the change of cancellation probability is not obvious in comparison with other intervals. In the industrial practice, the time intervals are useful for the shipping companies to find the peak cancellation period in the slot booking stage, and pay more attention to manage the slot bookings which are booked at the time intervals with higher cancellation probability.

It should be noted that the influential factors discussed in this section are based on the database of a container shipping company and the index data of the shipping market. To ensure the comprehensiveness of our analysis, we empirically discuss the factors that are not included in our model due to the unavailability of data. According to our interview with some managers in a container liner shipping company, the cause of trade itself, the competent authority, and some emergencies are three notable influential factors. The influence of trade itself is mainly reflected in the behavior of the shipper or the consignee. If the cargoes are not completed on schedule, the packaging of the cargoes are defective, the parties to the trade temporarily reduce the cargo order, or the trade is suspended due to the unqualified goods, the container slot bookings can be cancelled. Such cancellations due to trade and cargoes are more common in container shipping practice. The second influential factor is the behavior of some competent authorities, such as customs and commodity inspection authorities. Low customs clearance efficiency or high inspection and quarantine standards may cause the cargo to fail to clear customs in time and thus fail to be on board the ship. The third influential factor is emergencies in some areas. For example, due to the outbreak of COVID-19, some goods intended to be exported and the slots booked for them had to be cancelled. We only discuss them empirically here because these factors are difficult to quantify, and the data are difficult to obtain. Quantitative research on these influencing factors is a valuable future research direction.

6. Conclusions

To deal with the pain-point of slot booking cancellation for intercontinental container shipping, this study developed a slot booking data-driven model to estimate the cancellation probability of container slot bookings in the slot booking time period. A TTCM based on TTC analysis method was developed to quantitatively consider the internal factors of a slot booking and external factors that influence the probability of cancellation. To improve the performance of the model and reflect the effects of market regionality, we used the frailty term to formulate the correlation of the cancellation behaviors of the bookings that belong to the same origin country. The case study shows that the developed model is useful in assisting the shipping companies to estimate the loaded containers by their slot booking data in the slot booking time period.

We recommend further studies to enhance the prediction power of the model by considering more customer information discussed in Section 2. It is another research direction to use the cancellation probability to optimize the pricing policy for container liner shipping company.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is supported by the research projects “Liner Shipping Container Slot Booking Patterns and Their Applications to the Shipping Revenue Management” (R-302-000-177-720) and “Container Haulage Problems: Model Development, Effective Algorithm Design and Applications” (R-302-000-226-720) funded by the NOL Fellowship Programme of Singapore.

Appendix A. Survival function and hazard function in time-to-cancellation model

**Survival Function.** \( S(t) = P[T \geq t] \) is defined as the survival function indicating the probability that the TTC of a booking is no less than \( t \). The relationship between survival function and the PDF of \( T \) can be formulated by

\[
S(t) = P[T \geq t] = \int_t^\infty f(x)dx \tag{A1}
\]

Based on the assumptions in Section 3 and the survival function, the probability that a slot booking is cancelled during the booking time period can be expressed by \( \text{Prob}_{\text{clc}}(B) = P[T < t_b] = 1 - S(t_b) \) where \( t_b \) is the booking time of slot booking \( B \) measured by days before departure. Accordingly, the probability that the slot booking \( B \) will not be cancelled can be formulated as \( \text{Prob}_{\text{nocl}}(B) = 1 - \text{Prob}_{\text{clc}}(B) = S(t_b) \).

**Hazard Function.** Hazard function indicates the rate of an interested event happening at a particular time (Hosmer et al., 2011). Hazard function can be presented as follows.
\[
\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(\tau \in T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{-d\ln[S(t)]}{dt}
\]  

(A2)

According to Eq. (A2), there is a natural relationship between \(\lambda(t)\) and \(S(t)\) as follows:

\[
S(t) = e^{-\Lambda(t)}
\]  

(A3)

where \(\Lambda(t) = \int_0^t \lambda(x)dx\) is called cumulative hazard function. Thus, the survival function can be obtained if we know the formulation of the hazard function \(\lambda(t)\). In this study, according to the above definition, the cancellation hazard function implies the instantaneous rate of occurrence of the cancellation conditioning that the booking has not been cancelled before time \(t\). In time-to-event analysis, hazard function \(\lambda(t)\) can be formulated to several forms, and the characteristics of the slot bookings, as well as the internal and external influential factors, can be considered when modeling the hazard function.

**Appendix B. Two forms of baseline hazard function**

The two forms of baseline hazard functions are elaborated as follows:

(i) Weibull baseline hazard function.

For the parametric form, it is conventional to assume that the TTC \(T\) follows a Weibull distribution of two parameters with the PDF

\[
f(t \mid \mu_1, \mu_2) = \mu_1 \mu_2 t^{\mu_1-1} \exp(-t^{\mu_2})
\]

(B1)

where \(\mu = \{\mu_1, \mu_2\}\) is the parameters of the Weibull distribution. According to the relationship between hazard function and PDF presented by Eq. (A2), the Weibull baseline hazard function is

\[
\lambda_0(t \mid \mu) = \frac{f(t)}{S(t)} = \mu_1 \mu_2 t^{\mu_1-1}
\]

(B2)

where \(\mu = \{\mu_1, \mu_2\}\) is to be estimated using the slot booking data. The Weibull distribution assumption is widely assumed in the studies on time-to-event analysis and shows effective performance for some data (e.g., [Haque and Washington, 2015; Li et al., 2015]).

(ii) Spline baseline hazard function.

We use both B-spline and M-spline functions in this study to estimate the baseline hazard functions. Given the degree and knots, a set of basis functions can be determined. The detailed formulation of the two splines and the concepts of degree and knots can be found in the study by [Ramsay (1988)]. Assume there are \(Q\) basis functions denoted by \(SP_q(t)\), \(q = 1, 2, ..., Q\), the spline baseline hazard function can be formulated as

\[
\lambda_0(t \mid \mu) = \sum_{q=1}^{Q} \mu_q SP_q(t)
\]

(B3)

where \(\mu = \{\mu_1, \mu_2, ..., \mu_Q\}\) are the parameters for the linear combination of the basis functions, and \(\mu\) is to be estimated using the slot booking data.

**Appendix C. Partial likelihood from origin country \(i\)**

Suppose that the data set that characterizes an individual booking is \(\{Z_{ij}, t_{ij}, \alpha_{ij}\} : i = 1, 2, ..., n; j = 1, 2, ..., J_i\}\), where \(Z_{ij}\) is the vector of attributes, \(t_{ij}\) is the TTC (or TTL), and \(\alpha_{ij}\) is the cancellation indicator of individual \(j\) in the origin country \(i\).

The cumulative hazard function of slot booking \(j\) from an origin country \(i\) can be formulated as

\[
\Lambda_{ij}(t) = w_i \Lambda_0(t) e^{Z_{ij}}
\]

(C1)

where \(\Lambda_0(t) = \int_0^t \lambda_0(x)dx\) is the baseline cumulative hazards. Accordingly, the survival function, defined by Eq. (A3), with frailty term can be obtained as follows

\[
S_{ij}(t \mid w_i) = e^{-\Lambda_{ij}(t)}
\]

(C2)

Substitute the hazard function \(\lambda_{ij}(t \mid w_i)\), to Eq. (7), we get the partial likelihood function

\[
L_{ij}(w_i) = \prod_{j=1}^{J_i} \lambda_{ij}(t_{ij} \mid w_i)^{w_i} S_{ij}(t_{ij} \mid w_i)
\]

(C3)

Note that the \(w_i\) is a random variable with gamma distribution. Therefore, we obtain the expectation over \(w_i\). According to Eq. (7),
for both cancelled and non-cancelled bookings in origin country $i$, the likelihood function can be presented as follows

$$L_i = E[L_i(w_i)]$$

$$= \int_0^{\infty} \left[ \prod_{j=1}^{n_i} \frac{\lambda_0(t_j) e^{\beta T_j} \zeta_j}{\Gamma(\frac{1}{\theta} + \sum_{j=1}^{n_i} t_j)} \right] g(w_i) dw_i,$$

where $g(w) = \frac{w^{\alpha_i-1} e^{-w}}{\Gamma(\alpha_i)}$ is the PDF of gamma distribution.

Since gamma function has a good property that $\Gamma(x + n) = \Gamma(x) \prod_{k=1}^{n} (x + n - k)$, $n \in \mathbb{Z}^+$, the partial likelihood can avoid the gamma function and has the following form

$$L_i = \prod_{j=1}^{n_i} \left[ \frac{\lambda_0(t_j) e^{\beta T_j} \zeta_j}{\Gamma(\frac{1}{\theta} + \sum_{j=1}^{n_i} t_j)} \right]$$

where $\Lambda_i = \sum_{j=1}^{n_i} t_j$ is the summation of the cancellation indicators of slot bookings in origin country $i$, and $\mu$ is the vector of parameters in the baseline hazard function $\lambda_0(t)$. The loglikelihood function can be written as follows:

$$l(\beta, \theta, \mu)$$

$$= \ln \left( \prod_{i=1}^{n} L_i \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ \alpha_i \ln(\lambda_0(t_j|\mu)) + \alpha_i \beta^T T_j \zeta_j - \frac{1}{\theta} \ln \theta - \left( \frac{1}{\theta} + \Lambda_i \right) \ln \left( \sum_{j=1}^{n_i} \lambda_0(t_j|\mu) e^{\beta T_j} \zeta_j + \frac{1}{\theta} \right) \right]$$

$$\quad + \sum_{i=1}^{n} \ln \left( \frac{1}{\theta} + \Lambda_i \right)$$

$$= \ln \left( \prod_{i=1}^{n} L_i \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ \alpha_i \ln(\lambda_0(t_j|\mu)) + \alpha_i \beta^T T_j \zeta_j - \frac{1}{\theta} \ln \theta - \left( \frac{1}{\theta} + \Lambda_i \right) \ln \left( \sum_{j=1}^{n_i} \lambda_0(t_j|\mu) e^{\beta T_j} \zeta_j + \frac{1}{\theta} \right) \right]$$

$$\quad + \sum_{i=1}^{n} \ln \left( \frac{1}{\theta} + \Lambda_i \right)$$

$$= \ln \left( \prod_{i=1}^{n} L_i \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ \alpha_i \ln(\lambda_0(t_j|\mu)) + \alpha_i \beta^T T_j \zeta_j - \frac{1}{\theta} \ln \theta - \left( \frac{1}{\theta} + \Lambda_i \right) \ln \left( \sum_{j=1}^{n_i} \lambda_0(t_j|\mu) e^{\beta T_j} \zeta_j + \frac{1}{\theta} \right) \right]$$

$$\quad + \sum_{i=1}^{n} \ln \left( \frac{1}{\theta} + \Lambda_i \right)$$

$$= \ln \left( \prod_{i=1}^{n} L_i \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ \alpha_i \ln(\lambda_0(t_j|\mu)) + \alpha_i \beta^T T_j \zeta_j - \frac{1}{\theta} \ln \theta - \left( \frac{1}{\theta} + \Lambda_i \right) \ln \left( \sum_{j=1}^{n_i} \lambda_0(t_j|\mu) e^{\beta T_j} \zeta_j + \frac{1}{\theta} \right) \right]$$

$$\quad + \sum_{i=1}^{n} \ln \left( \frac{1}{\theta} + \Lambda_i \right)$$

$$= \ln \left( \prod_{i=1}^{n} L_i \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ \alpha_i \ln(\lambda_0(t_j|\mu)) + \alpha_i \beta^T T_j \zeta_j - \frac{1}{\theta} \ln \theta - \left( \frac{1}{\theta} + \Lambda_i \right) \ln \left( \sum_{j=1}^{n_i} \lambda_0(t_j|\mu) e^{\beta T_j} \zeta_j + \frac{1}{\theta} \right) \right]$$

$$\quad + \sum_{i=1}^{n} \ln \left( \frac{1}{\theta} + \Lambda_i \right)$$

$$= \ln \left( \prod_{i=1}^{n} L_i \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[ \alpha_i \ln(\lambda_0(t_j|\mu)) + \alpha_i \beta^T T_j \zeta_j - \frac{1}{\theta} \ln \theta - \left( \frac{1}{\theta} + \Lambda_i \right) \ln \left( \sum_{j=1}^{n_i} \lambda_0(t_j|\mu) e^{\beta T_j} \zeta_j + \frac{1}{\theta} \right) \right]$$

$$\quad + \sum_{i=1}^{n} \ln \left( \frac{1}{\theta} + \Lambda_i \right)$$

References

Chiew, E., Dzainio, R.A., Garrow, L.A., 2017. Bayesian estimation of hazard models of airline passengers’ cancellation behavior. Transp. Res. Part A Policy Pract. 96, 154–167. https://doi.org/10.1016/j.trapa.2016.12.006.

Commenges, D., Joly, P., Gégout-Petit, A., Liqué, B., 2007. Choice between semi-parametric estimators of Markov and non-Markov multi-state models from coarsened observations. Scand. J. Stat. https://doi.org/10.1111/j.1467-9469.2006.00536.x.

Garrow, L.A., Koppelman, F.S., 2004. Predicting air travelers’ no-show and standby behavior using passenger and directional itinerary information. J. Air Transp. Manag. https://doi.org/10.1016/j.jairtraman.2004.06.007.

Graham, J.R., Garrow, L.A., Leonard, J.D., 2010. Business travelers’ ticketing, refund, and exchange behavior. J. Air Transp. Manag. https://doi.org/10.1016/j.jairtraman.2009.12.002.

Haque, M.M., Washington, S., 2015. The impact of mobile phone distraction on the braking behaviour of young drivers: A hazard-based duration model. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2014.07.011.

Hinkle, D.E., Wiersma, W., Jurs, S.G., 2003. Applied Statistics for the Behavioral Sciences. Houghton Mifflin College Division.

Hosmer, D.W., Lemeshow, S., May, S., 2011. Applied Survival Analysis: Regression Modeling of Time to Event Data: Second Edition, Applied Survival Analysis: Regression Modeling of Time to Event Data: Second Edition. https://doi.org/10.1002/97804708258019.

Hougaard, P., 1995. Frailty models for survival data. Lifetime Data Anal. https://doi.org/10.1007/BF99585760.

Huegl, C., Vannotti, F., 2001. Data mining techniques to improve forecast accuracy in airline business. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 438-442. https://doi.org/10.1145/502512.502578.

Iliescu, D.C., Garrow, L.A., Parker, R.A., 2008. A hazard model of US airline passengers’ refund and exchange behavior. Transp. Res. Part B Methodol. 42, 229–242. https://doi.org/10.1016/j.trb.2007.10.005.

Kay, R., 1977. Proportional hazard regression models and the analysis of censored survival data. Appl. Stat. 26, 227. https://doi.org/10.2307/2346962.

Lee, C.Y., Tang, C.S., Yin, R., An, J., 2015. Fractional price matching policies arising from the ocean freight service industry. Prod. Oper. Manag. https://doi.org/10.1111/poms.2013.23.

Li, R., Fan, J., 2002. Variable Selection for Cox’s proportional hazards model and frailty model. Ann. Stat. 30, 74–99. https://doi.org/10.1214/aoms/1015362185.
Li, R., Pereira, F.C., Ben-Akiva, M.E., 2015. Competing risk mixture model and text analysis for sequential incident duration prediction. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2015.03.009.

Meng, Q., Zhao, H., Wang, Y., 2019. Revenue management for container liner shipping services: Critical review and future research directions. Transp. Res. Part E Logist. Transp. Rev. https://doi.org/10.1016/j.trce.2019.06.010.

Morales, D.R., Wang, J., 2010. Forecasting cancellation rates for services booking revenue management using data mining. Eur. J. Oper. Res. 202, 554–562. https://doi.org/10.1016/j.ejor.2009.06.006.

Ramsay, J.O., 1988. Monotone Regression splines in action. Stat. Sci. 3, 425–441. https://doi.org/10.1214/ss/1177012761.

Rueten-Budde, A.J., Putter, H., Fiocco, M., 2019. Investigating hospital heterogeneity with a competing risks frailty model. Stat. Med. 38, 269–288. https://doi.org/10.1002/sim.8002.

Schinas, O.D., Psaraftis, H.N., 1998. New frontiers through short sea shipping. In: Transactions - Society of Naval Architects and Marine Engineers.

UNCTAD, 2018. Review of maritime transport 2018, New York and Geneva, United Nations.

von Westarp, A.G., Schinas, O., 2016. A fuzzy approach for container positioning considering sustainable profit optimization. Transp. Res. Part E Logist. Transp. Rev. 92, 56–66. https://doi.org/10.1016/j.tre.2016.04.014.

Wales, D.J., Doye, J.P.K., 1997. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. https://doi.org/10.1021/JP970984N.

Wang, S., Alharbi, A., Davy, P., 2014. Liner ship route schedule design with port time windows. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2014.01.012.

Wang, Y., Meng, Q., 2018. Integrated method for forecasting container slot booking in intercontinental liner shipping service. Flex. Serv. Manuf. J. https://doi.org/10.1007/s10696-018-9324-z.

Wang, Y., Meng, Q., Jia, P., 2019a. Optimal port call adjustment for liner container shipping routes. Transport. Res. Part B-Methodol. 128, 107–128. https://doi.org/10.1016/j.trb.2019.07.015.

Wang, Y., Meng, Q., Kuang, H., 2019b. Intercontinental liner shipping service design. Transport. Sci. 53 (2), 344–364. https://doi.org/10.1287/trsc.2018.0831.

Yang, D., Pan, K., Wang, S., 2018. On service network improvement for shipping lines under the one belt one road initiative of China. Transp. Res. Part E Logist. Transp. Rev. 117, 82–95. https://doi.org/10.1016/j.tre.2017.07.003.

Yu, Z., Wood, J.S., Gayah, V.V., 2017. Using survival models to estimate bus travel times and associated uncertainties. Transp. Res. Part C Emerg. Technol. https://doi.org/10.1016/j.trc.2016.11.013.

Zhao, H., Meng, Q., Wang, Y., 2019. Exploratory data analysis for the cancellation of slot booking in intercontinental container liner shipping: A case study of Asia to US West Coast Service. Transp. Res. Part C Emerg. Technol. 106, 243–263. https://doi.org/10.1016/j.trc.2019.07.009.

Zurheide, S., Fischer, K., 2012. A revenue management slot allocation model for liner shipping networks. Marit. Econ. Logist. 14, 334–361. https://doi.org/10.1057/mel.2012.11.