Generalized *F* Distribution Model with Random Parameters for Estimating Property Damage Cost in Maritime Accidents

Jinxian Weng[a], Dong Yang[b, *], Gang Du[c]

[a]*College of Transport and Communications, Shanghai Maritime University, Shanghai, China 201306*
[b]*Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong, China*
[c]*Department of Business Management, School of Business Administration, Faculty of Economics and Management, East China Normal University, Shanghai 200062, China*

Abstract
This study develops a generalized *F* distribution model with random parameters to estimate the ship property damage cost in shipping traffic accidents using ten years' shipping accident data in the Fujian waters. Model results show that sinking and capsizing can incur the largest property damage cost, followed by collisions, contact, grounding and fire/explosion. There is a smaller ship property damage cost when the ship is moored or docked. The poor visibility has the least impact on the increment of ship property damage cost. The results also reveal that the bigger property damage cost is associated with shipping accidents occurring in the Straits/sea areas and under the strong wind/wave condition and nighttime period. One important finding is that the lookout failure exhibits a bigger effect than the operation error. The results of this study are beneficial for policy-makers in proposing efficient strategies to reduce property damage cost in shipping accidents. The developed model is useful for insurance companies in determining the appropriate ship insurance rates.

Keywords: Generalized *F* distribution; Ship property damage; Shipping Traffic Accident; Random parameter; Cost

1. Introduction
A shipping accident may lead to a large amount of economic loss as well as public criticism. Fig. 1 shows the total losses of shipping accidents by top 10 regions from 2006 to 2015. It can be seen that the shipping accident loss is the largest in the South China Sea. In general, the consequence (i.e., human life loss, property damage cost) of a shipping accident is related to various influencing factors including the accident type, ship type, shipping traffic, weather condition, accident time, navigational status, accident location, human error and so on. It is incumbent upon policy-makers to implement efficient navigational safety strategies with the objective of decreasing the human life loss as well as property damage cost if an accident does occur. Considering the limited available resources and budgets, policy-makers have to prioritize safety strategies. This can be achieved with the help of a comprehensive understanding of the contributory factors that affect the shipping accident consequence.

*Dr. Dong Yang. Email: dong.yang@polyu.edu.hk

Numerous studies have been carried out to analyze the historical accident data. One major focus has been placed on the analysis of influencing factors on the shipping accident occurrence likelihood and consequence (e.g., Jin et al., 2001; Roberts et al., 2010; Jin, 2014; Weng and Yang, 2015). Among these studies, many statistical models like negative binomial regression models (e.g., Jin et al., 2001; Talley et al., 2006; Weng and Yang, 2015), Poisson regression models (e.g., Talley et al., 2006) and zero-inflated negative binomial regression models (e.g., Weng et al., 2016) have been developed to evaluate the shipping accident consequence in terms of human life loss. It should be pointed out that the occurrence of shipping accident may not only lead to crew injuries/deaths but also result in property damages to the involving ships. Nevertheless, few efforts have been made to the analysis of the ship property damages resulting from shipping accidents. Because of the variations in exogenous factors, the cost of ship property damages could not be a fixed value at different times. In general, the ship property damage cost should follow a specific type of distribution such as Weibull distribution, normal distribution, gamma distribution, log-logistic distribution and so on.

In order to accurately model a complex shipping accident evaluation system where multiple factors exert various effects to the accident consequence, we aim to develop a generalized *F* distribution model to evaluate the property damage cost resulting from shipping accidents. Note that the traditional distribution types (e.g., Weibull distribution, log-logistic distribution, lognormal distribution, gamma distribution and etc) are all special forms of the generalized *F* distribution. Hence, the generalized *F* distribution model is more appropriate to explore the property damage cost resulting from shipping accidents. In reality, the effect of a specific influencing factor may not be the same for each observation. In order to incorporate the unobserved heterogeneity, we introduce the random parameters for the model.

Therefore, the objective of this study is to develop a generalized F distribution model with random parameters to estimate the ship property damage cost resulting from shipping accidents. The effects of influencing factors on the ship property damage cost are examined in this study. The contributions of this study are three-fold. First, it makes an initial attempt to develop a mathematical model that can be used to accurately describe the relationship between ship property damage cost and influencing factors in shipping accidents. To the best of our knowledge, the literature regarding the models for estimating ship property damage cost in shipping accidents is rather limited. Second, with regard to the methodology, the developed F distribution model with random parameters is a generalized model, which covers the traditional distribution models like Weibull, log-logistic, lognormal and gamma models thus has a broader applicability. Third, from the practical implication viewpoint, the results of this study are beneficial for policy-makers in proposing efficient strategies to reduce the property damage cost caused by shipping accidents. The developed model can also assist insurance companies in determining the appropriate ship insurance rates.

2. Literature review

In recent decades, shipping safety has attracted a lot of attentions from academia. Many academic papers addressing the occurrence probability and consequence of shipping accidents have been published in different reputable journals, including Accident Analysis & Prevention, Journal of Navigation, Safety Science, Maritime Policy & Management and so on. Table 1 tabulates the summary of the existing works published in major safety/shipping journals. It can be seen from Table 1 that the fishing vessel accident is one study object which many papers addressed (e.g., Jin et al., 2001; Jin and Thunberg, 2005; Perez-Labajos et al., 2006; Roberts et al., 2010). Besides the fishing vessel accident, Talley et al. (2006), Talley et al. (2008) and Yip et al. (2015) analyzed the passenger ship shipping accidents. Eliopoulou and Papanikolaou (2007), Ugurlu et al. (2015) explored the effects of contributory factors in oil tanker accidents. Eliopoulou et al. (2013) examined the casualty resulting from cellular type container ships by employing a simple data analysis. These shipping accident related studies only took into account specific ship type because of data limitations. Chung et al. (2017) identified the role of burnout in seafarer health and well-being and its effect on safety.

Many other studies have been conducted to investigate the relationship between the contributory factors and shipping accident risks considering all ship types. For example, many researchers estimated the occurrence probability and consequence of shipping accidents at different water areas such as US coast area (Jin et al., 2001; Jin and Thunberg, 2005; Talley et al., 2006, 2008), the Istanbul strait (Akten, 2004; Ozsoysal and Ozsoysal, 2006; Uluscu et al., 2009; Arslan and Turan, 2009; Aydogdu et al., 2012, Ugurlu et al., 2016), the Singapore strait (Debnath and Chin, 2010; Weng et al. 2012), Hong Kong water area (Yip, 2008); UK Coast area (Roberts et al., 2010), South China Sea (Wang et al., 2014; Weng et al. 2016) and Gulf of Bothnia (Goerland et al. 2017). The estimation of human life loss and accident severity has been the major study object in these studies. As another major consequence concerning maritime authorities and insurance companies, the property damage to ships in shipping accidents has been rarely addressed in the existing literature, as shown in Table 1.

So far, many statistical approaches have been applied to evaluate the shipping accident consequences. For example, Jin et al. (2001) estimated total losses and crew injuries in commercial fishing vessel accidents using Probit and negative binomial regression techniques. Talley et al. (2006) applied the Tobit, negative binomial and Poisson regression techniques to determine the total loss, injuries and deaths/missing people in passenger vessel accidents, respectively. Talley et al. (2008) modeled the severity of cruise vessel accidents using the Tobit regression method. Yip (2008) employed the negative binomial regression technique to describe the injuries and fatalities caused by ship accidents in Hong Kong waters. Weng et al. (2016) employed the zero-inflated negative binomial regression technique to evaluate the human life loss resulting from shipping accidents in the South China Sea. Luo et al. (2017) examined the factors affecting both first and recurrent accidents, by focusing on the duration between two accidents with Cox proportional hazard models.

Unlike the human life loss (the number of deaths/injuries), the ship property damage cost is not an integer variable. Hence, the traditional distribution models such as Weibull distribution, normal distribution, gamma distribution, log-logistic distribution may not be applicable for determining the ship property damage cost in shipping accidents. The generalized *F* distribution model, which contains all the above-mentioned distribution types as its special cases (Hogg and Ciamp ,1985; Cox, 2008), could be more suitable for the estimation of ship property damage cost resulting from shipping accidents.

3. Methodology

In general, the generalized *F* distribution is a four-parameter family, generalizing the central *F* with noninteger degrees of freedom ($2m_1$, $2m_2$) by adding location (λ) and scale ($\sigma > 0$) parameters according to the standard accelerated failure time model. In reality, the cost of ship property damage caused by shipping accidents might be affected by the explanatory variables such as the accident type, accident location, weather conditions, ship type, accident time, etc. For simplicity, we assume that the location parameter λ can be considered as the linear function of these explanatory variables, namely

$$\lambda = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k = \boldsymbol{X\beta} \tag{1}$$

where $\boldsymbol{X}$ is the vector of the explanatory variables influencing the ship property damage cost and $\boldsymbol{\beta}$ is the vector of coefficients.

Therefore，the density function of a generalized *F* distribution is expressed by

$$\begin{aligned} f(y_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}, m_1, m_2, \sigma) &= \frac{e^{-\lambda m_1/\sigma} y^{(m_1/\sigma)-1} (m_1/m_2)^{m_1}}{\sigma B(m_1, m_2)[1+(m_1/m_2)(e^{-\lambda} y)^{1/\sigma}]^{(m_1+m_2)}} \\ &= \frac{1}{B(m_1, m_2)} \left(\frac{\frac{m_1}{m_2} e^{\omega}}{1+\frac{m_1}{m_2} e^{\omega}} \right)^{m_1} \left(\frac{1}{1+\frac{m_1}{m_2} e^{\omega}} \right)^{m_2} \frac{1}{\sigma y} \end{aligned} \tag{2}$$

where $B(m_1, m_2)$ is the beta function evaluated at m_1 , m_2 >0; $\omega = [\log y_i - \lambda]/\sigma = [\log y_i - \boldsymbol{X}_i \boldsymbol{\beta}]/\sigma$. As argued by Hogg and Ciamp (1985), the generalized *F* distribution includes many commonly used distributions as special cases (e.g., gamma distribution, lognormal distribution, exponential distribution and etc). For example, the log-logistic distribution can be considered as a special generalized *F* distribution with parameters $m_1 = m_2 = m$. A Weibull distribution is the generalized *F* distribution with parameters m_1=1 and $m_2 \to \infty$. The detailed relationship between generalized *F* distribution and other commonly used distributions has been illustrated by Hogg and Ciamp (1985).

The estimates of $\boldsymbol{\beta}$, m_1, m_2 and σ can be determined by maximizing the log-likelihood function $\ln L(\boldsymbol{\beta}, m_1, m_2, \sigma) = \sum_{i=1}^{N} \ln f(y_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}, m_1, m_2, \sigma)$ with respect to $\boldsymbol{\beta}$, m_1, m_2 and σ. In this study, we can use the Expectation-Maximisation (EM) algorithm (Dempster et al., 1977) that is similar to the quasi-Newton methods to determine the maximum likelihood estimates. The iteration steps are described below:

(1) Initialize values $\gamma^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}, \ldots, \beta_k^{(0)}, m_1^{(0)}, m_2^{(0)}, \sigma^{(0)}\right)^T$ for $\gamma = \left(\beta_0, \beta_1, \ldots, \beta_k, m_1, m_2, \sigma\right)^T$.

(2) Compute the probabilities $f(y_i \mid \boldsymbol{X}_i, \boldsymbol{\beta}, m_1, m_2, \sigma)$ in the data sample using the initial values.

(3) Solve for $\frac{\partial}{\partial \boldsymbol{\gamma}} \ln L\left(\boldsymbol{\beta}, m_1, m_2, \sigma\right) = 0$ to get the estimates of $\boldsymbol{\gamma}$, denoted as $\boldsymbol{\gamma}^{(1)} = \left(\beta_0^{(1)}, \beta_1^{(1)}, \ldots, \beta_k^{(1)}, m_1^{(1)}, m_2^{(1)}, \sigma^{(1)}\right)^T$.

(4) Replace the initial parameters $\gamma^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}, \ldots, \beta_k^{(0)}, m_1^{(0)}, m_2^{(0)}, \sigma^{(0)}\right)^T$ by $\boldsymbol{\gamma}^{(1)} = \left(\beta_0^{(1)}, \beta_1^{(1)}, \ldots, \beta_k^{(1)}, m_1^{(1)}, m_2^{(1)}, \sigma^{(1)}\right)^T$ and go back to step (2).

(5) Iterate until a criterion of convergence is reached. At the convergence, the values of the parameters to be determined are the maximum likelihood solutions.

The generalized *F* distribution model assumes that the effect of an individual variable is the same for each observed case. In other words, it assumes that the ship property damage cost is homogeneous across different observations, which might be inconsistent with the reality. Generally, one effective approach to incorporate the unobserved heterogeneity is to introduce random parameters for the generalized *F* distribution model. More specifically, the difference between the fixed and random parameters for the generalized *F* distribution model is that the latter is added by a randomly distributed term ξ_n. Namely, a random parameter can be expressed by

$$\beta_n^{'} = \beta_n + \xi_n \tag{3}$$

where $\beta_n^{'}$ is the random parameter term varying with different observations; β_n is the fixed parameter and the random error term ξ_n follows $N(0, \sigma_n^2)$.

In general, this method allows for the correlation across random parameters, the examination of which may yield insights into the data generating process for the underlying heterogeneity (Hojati et al., 2013). It should be noted that not all the parameters should be considered as random parameters. In this study, we use a method to determine which parameter should be regarded as a random parameter for the generalized *F* distribution model. More specifically, this method starts with the generalized *F* distribution model with all fixed parameters (no random parameters). For each explanatory variable, we calculate the log-likelihood of the model by means of the (EM) algorithm mentioned above if the corresponding parameter is regarded as a random parameter. The random parameter that could yield the smallest value for the Akaike Information Criterion (*AIC*) will be used by the model. When one more random parameter is considered, and then we calculate the log-likelihood again. Hence, the random parameters are added one by one to the model until no random parameters could improve the model performance.

4. Data collection

In this study, we collected historical ship traffic accident data between the years of 2004 and 2014 from the Fujian Maritime Safety Administration. There was a detailed record for each shipping accident occurred in the Fujian water area. Each recorded data contains three types of information: (i) Accident characteristics; (ii) Environmental characteristics and (iii) Human casual factors. Accident characteristics include the detailed information regarding accident type (collision, fire/explosion, grounding, contact, sinking, capsizing), accident location (coastal_habour_port area, straits/sea area), ship type (cargo/container ship, fishing ship, LNG/LPG/Oil/Chemical ship, other ship types), ship tonnage and the navigational status (underway, moored/docked). Environmental characteristics contain the information regarding visibility (good visibility, restricted visibility, poor visibility), strong_wind/wave (yes or no) and accident time (daytime period, nighttime period). Judgment_error, lookout_failure and opearation_error are the three human casual factors related to shipping accidents.

A total of 1248 ship accidents including 650 collision accidents were collected from 2004 to 2014 in Fujian water area. Based on the collected data, Table 2 provides detailed summaries on the statistics of 20 variables described above. It can be seen that 52.1% of shipping accidents are collisions while only a minority of shipping accidents suffer a contact (17.1%), grounding (10.4%), fire/explosion(4.5%), sinking (6.4%) and capsizing (0.6%). There is also a small proportion of accidents occurring under adverse weather conditions (25.5% for strong_wind/wave conditions) and under poor visibility conditions (17.4% for poor conditions and 6.73% for restricted conditions). According to the collected data, it can be found that the mean for the navigational status is 0.822, suggesting that the majority of shipping accidents occurred when the ship was underway. The low mean (0.030) for the LNG/LPG/Oil/Chemical ship suggests that there is only a small proportion of accidents involving in dangerous ships (i.e., LNG/LPG/Oil/Chemical ships) while cargo/container ship (83.2%) is the major ship type involving in shipping accidents, shown in Table 2. In addition, it can be seen from the table that there is a very small proportion of accidents associated with the judgment error (10.5%) while more than half of accidents are found to be associated with the operation error (55.1%). The mean property damage cost caused by shipping accidents equals to 69.854×10^4 RMB.

Table 3 further gives the results for the analysis of variance (ANOVA) test and Levene's test under different circumstances. It can be seen that the ship property damage cost for the collision accidents (77.8×10^4 RMB) is slightly higher than that for non-collision accidents (61.2×10^4 RMB). The ship property damage cost is found to be larger if a shipping accident occurs during the nighttime period or under the adverse weather condition and the poor visibility condition. Table 3 also shows that the average cost of ship property damage caused by shipping accidents occurring straits/sea areas is significantly larger than for the accidents occurring coastal/harbor/port areas (110.8

$\times 10^4$ RMB versus 41.1×10^4 RMB). According to Table 3, the variance of ship property damage cost in shipping accidents is statistically different for accident times, accident locations and accident types including collision, contact, grounding and sinking at the 0.05 significance level. However, the mean and variance of property damage cost are not affected by the accident type of capsizing at the significance level of 0.05.

Nevertheless, the above one-way ANOVA and Levene's tests in Table 3 are univariate statistical techniques that only allow the analysis of a single categorical factor at a time. This may give rise to biased or incorrect results due to the isolation of a single factor for analysis while other factors are held fixed. For example, it is inconsistent with our expectation that a shipping accident suffering a fire/explosion has lower property damage cost than the accident with no fire/explosion, as shown in Table 3. In reality, the mean and variance of shipping accident consequence may be affected by multiple factors at the same time. Therefore, it is more appropriate to take multiple factors into account simultaneously to describe such relationship.

5. Results & Discussions

5.1 Model results

The generalized F distribution procedure in the Limdep software (Version 9, Econometric Software Inc., NY, USA) was performed to calibrate the proposed generalized F distribution model using the collected 1278 shipping accident records. It should be pointed out that some variables will be removed from the model if a low statistical significance is given. However, these excluded variables are still of interest and expected to have some effects on the ship property damage cost in this study. Moreover, the removal of these variables may reduce prediction accuracy. Therefore, all the variables are kept in the model formulation in this study. In order to demonstrate that the generalized F-distribution model with random parameters provides the better fit to these data, other four distribution types are also tested for comparison, namely, Weibull distribution model, Weibull distribution model with gamma heterogeneity, gamma distribution model and Log-logistic distribution model. When several models are available, one can compare the model performance using the most regularly used measures such as the log-likelihood, *AIC* and the Bayesian Information Criterion (*BIC*). The *AIC* penalises a model with a larger number of parameters and is defined as $AIC = -2\ln L + 2q$, where $\ln L$ represents the fitted log-likelihood and q is the number of parameters for the model. The *BIC* penalizes a model with a larger number of parameters and a larger sample size, and is defined as $BIC = -2\ln L + q\ln N$, where N is the sample size. In general, lower values of the *AIC* and *BIC* statistics are preferable.

Table 4 tabulates the log-likelihood, the *AIC* and the *BIC* for each distribution model. It can be seen that different distribution models give different *AIC* and *BIC* values. It can be found that the generalized F distribution model with random parameters provides the best fit for the estimation of ship property damage cost with the lowest *AIC* statistic (4414.90) and the lowest *BIC* statistic (4548.26), compared with those from the other eight models. These results confirm the appropriateness of using the generalized F distribution model with random parameters to estimate the ship property damage cost resulting from shipping accidents.

Table 5 gives the results of the generalized F distribution model with random parameters for estimating ship property damage cost in shipping accidents. From the table, it can be seen that only four coefficients including the coefficient for grounding, the coefficient for accident location, the coefficient for visibility, and the coefficient for navigational status are considered as random parameters. Furthermore, the coefficient for collisions is positive and statistically significant at a 0.01 level, indicating that collisions are strongly associated with the property damage cost resulting from shipping accidents. The model results also show that the ship property damage cost in shipping accidents is significantly influenced (at a significance level of 0.01) by the following explanatory variables: fire/explosion, sinking, capsizing, accident location, cargo/container ship, fishing ships, navigational status, ship tonnage, accident time and lookout failure. The coefficient associated with the judgment error is statistically significant at a level of 0.10 while other three variables including grounding, visibility and operation error have minor effects on the property damage cost in shipping accidents because of their small coefficients.

5.2 Sensitivity analysis

Although the signs of the estimated coefficients for the generalized *F* distribution model with random parameters could provide useful information on whether changes in given explanatory variables increase or decrease the loss of human life resulting from shipping accidents, they cannot provide further information on the extent to how much the ship property damage cost changes.

To gain further insight into the quantitative effects of the explanatory variables on the property damage cost, we can use the exponents of the estimated coefficients to interpret the results. The exponents of the coefficients, with all coefficients typically evaluated at their mean values, translate to a percent change in ship property damage cost resulting from a change from 0 to 1 or from 1 to 2 for the categorical variables (e.g., collision, navigational status) and 10000 units increase for the continuous explanatory variable (i.e., ship tonnage). For example, the exponential coefficient of accident time is 1.311 ($\approx e^{0.271}$), suggesting that the ship property damage cost is about 31.1 % higher for shipping accidents occurring at night than for those occurring during the daytime period.

Fig. 2 depicts the effects of influencing factors on the cost of property damage to ships. According to the percentage changes in Fig. 2, most of the results are consistent with the univariate analysis results in Table 3. The figure shows that the property damage cost in shipping accidents suffering collisions is 79% larger on average than that for accidents suffering no collisions. From Fig. 2, we can see that largest ship property damage cost is strongly associated with the ship sinking and capsizing. More specifically, the ship property damage cost will dramatically increase by 20 times if a ship eventually sinks into the water or suffers a capsizing. To prevent the occurrence of sinking and capsizing, one efficient method is to enhance ship structural safety. Another method is to maintain ship stability during damage. The extreme case of sinking and capsizing does not occur as long as the ship has stability even if the ship is structurally damaged (Lee et al., 2005). Followed by the ship sinking and capsizing, the occurrence of fire/explosion could also significantly increase the ship property damage cost. More specifically, Fig. 2 shows that the ship property damage cost resulting from fire/explosion accidents are 127% higher than from accidents where no fire/explosions are involved.

Consistent with previous studies (e.g., Akten, 2004; Talley et al., 2006; Weng and Yang, 2015), the accident time is strongly correlated with the serious accident consequence. Fig.2 highlights that the ship property damage cost for night-time accidents is generally larger than for daytime accidents (by 31%), which is close to the finding from the study of Talley et al. (2006). The larger ship property damage cost at night might be attributed to the fact that the search and rescue efficiency will be greatly reduced at night compared to the daytime period. Similar to the study of Jin (2014), the distance to the shore is also a significant factor influencing the shipping accident damage cost in this study. It is found that shipping accidents occurring at the straits/sea area (i.e., far away from the shore) can cause a bigger property damage cost with the figure of 93%. This result is consistent with our expectation because there will be a reduced rescuing efficiency for the shipping accidents occurring at the straits/sea area.

It can be also seen from Fig. 2 that the navigational status can influence the property damage cost. More specifically, there will be an increase of 52% on the ship property damage cost when the ship is underway, as compared with the situation when the ship is docked or moored. Compared to the navigational status, the factor of visibility has a smaller effect on the ship property damage cost. It is found that the poor visibility will increase the ship property damage cost only by 8% ($= 2 \times 4\%$) compared with the good visibility in this study. The smaller effect of visibility on the ship property damage cost can be explained by the fact that visibility is associated with the accident time and navigational status. Our historical shipping accident data show that there is a Pearson correlation of 0.364 between the visibility and accident time, and a correlation rate of 0.317 between the visibility and navigational status. Therefore, the effect of visibility has been at least partially captured by the factors of accident time and navigational status.

Fig.2 shows that the ship property damage cost varies with the ship type significantly. As expected, the involvement of dangerous ships such as LNG/LPG/Chemical ships is associated with bigger ship property damage cost. This result is reasonable because the LNG/LPG/Chemical ships are almost large-sized ships (with a mean tonnage of 16522 ton) and they carry higher economic value of goods in this study. As evidenced by many researchers (e.g., Weng and Yang, 2015), the probability of being a serious accident is expected to be greater when the large-sized ship is involved in accidents, further resulting in a bigger ship property damage cost. One important finding from Fig. 2 is that an accident involving fishing ships may cause the bigger ship property damage cost by 59% than that not involving fishing ships. This result may be attributed to the fact that the majority of fishing ships (i.e., 93.3% of fishing ships) are involved in collision accidents in this study. However, fishing ships are not compulsorily equipped with the automatic identification system (AIS) while this system could make ships “visible” to each other. In addition, the majority of fishing ships may not comply with collision avoidance rules (e.g., International Regulations for Preventing Collisions at sea). Obviously, the aggressive navigation behavior of fishing ships could increase the risk of serious collisions, further leading to a bigger property damage cost.

Many researchers (e.g., Karahalios, 2014) argued that the most likely cause in a shipping accident would be human errors on the ships. In reality, the human error could not only affect the occurrence likelihood of shipping accident but also influence the accident consequence. Our results show that the lookout failure is the type of human errors that has the largest impact in increasing the ship property damage cost, followed by the judgment error and operation error. More specifically, the presence of lookout failure could increase the ship property damage cost by 38% while the existence of operation error can only contribute to an increase of 14% on the ship property damage cost. Therefore, more attentions should be paid to avoid the occurrence of lookout failures in order to reduce shipping accident consequence.

5.3 Under-reporting analysis

Underreporting of maritime accidents is a problem not only for the authorities endeavouring to improve maritime safety through legislation, but also for the entities that use maritime casualty statistics in risk and accident analysis (Hassel et al., 2011). It is known that a certain number of unreported maritime accidents are confirmed by first-hand reports from ship operators. There could be several possible reasons for this, including local reporting procedures not being understood by the crew or ship-owner. In general, the reliability of accident analysis results may be affected by the proportion of missing data. In this study, no other database was collected to confirm how many accident reports were missing. Nevertheless, we can evaluate the effects of the missing data by randomly adding a part of data for the model development. More specifically, the collected shipping accident data from 2004 to 2014 is considered as the first data sample while the second data sample includes the missing data.

In general, we can use the likelihood ratio test to compare the fit of two models, one of which (the null model) is a special case of the other (the alternative model). In this study, the likelihood ratio test statistic (often denoted by *D*) can be defined by

$$D = -2[LL(\beta_T) - LL(\beta_a) - LL(\beta_b)] \tag{4}$$

where $LL(\beta_T)$ is the log-likelihood at convergence of the model *T* using the true data (the collected data plus the missing data), $LL(\beta_a)$ is the log-likelihood at convergence of the model *A* using the collected data, $LL(\beta_b)$ refers to the log-likelihood of the model *B* using the missing data. The statistic *D* is chi-square distributed with the freedom that equals to the summation of the number of estimated parameters in the models *A* and *B* minus the number of estimated parameters in the model *T*.

Fig. 3 depicts the effects of missing data on the reliability/applicability of the model results based on collected data. It can be seen that the likelihood ratio test statistic *D* is a small value and the corresponding *p*-value is a big value if there is only a small proportion of missing data (e.g., 5%~15% of data). This implies that the ship property damage cost estimation model based on the missing data is equivalent to the developed model using the collected data. In other words, the developed ship property damage cost estimation model is still reliable and applicable when only a small proportion of data was missing. However, it can be clearly seen that the test statistic *D* will significantly increase if there is a large proportion of missing data, as shown in Fig. 3. For example, the *D* equals to 21.0 and the corresponding *p*-value is 0.01 when 30% of the collected data are missing in reality, suggesting that the developed model could not be applied to accurately estimate the property damage cost in shipping accidents under this situation.

6. Conclusions

This study has developed a generalized *F* distribution model with random parameters to estimate the ship property damage cost resulting from shipping accidents using archived shipping accident data in the Fujian waters in China. Both fixed and random parameters for the Weibull distribution, log-logistic distribution, gamma distribution were also examined for the model comparison. Based on ten-years shipping accident datasets from 2004 to 2014, the model parameters were calibrated using the Expectation-Maximisation (EM) algorithm.

The model comparison results show that the generalized *F* distribution model with random parameters performs the best to estimate the ship property damage cost resulting from shipping accidents. The generalized *F* distribution model results show that the ship property damage cost is greater by: (i) 79% for collisions; (ii) 52% and 7% for contact and grounding accidents, respectively; (iii) 127% for fire/explosion accidents; (iv) 93% for accidents occurring at straits/sea areas; (v) 64% for accidents involving LNG/LPG/Oil/Chemical ships; (vi) 52% if the ship is underway; (vii) 26% and 31% for accidents occurring in the strong wind/wave condition and nighttime period, respectively. From the model results, it can be concluded that sinking and capsizing can incur the biggest ship property damage cost. The poor visibility has the least impact on the increment of ship property damage cost because the effect has been partially captured by the accident time and navigational status. One important finding from this study is that the lookout failure exhibits bigger effects on the ship property damage cost than the operation error. This study implies that it is essential to prevent ships from capsizing or sinking once an accident occurs so that the resulting loss can be reduced significantly.

Government agencies and shipping companies could employ the developed model to assess various safety enhancing measures and strategies. The developed model in this study can also assist insurance companies to determine the appropriate ship insurance rates. One limitation of this study is that some other influencing factors such as vessel age and the time spent in port/navigation were not included in the collected shipping accident database. Future study will take into account these factors for modeling by collecting more data from various data sources.

7. Disclaimer

The views expressed in this paper only reflect the opinions of the authors and must not be considered as official opinions from any national or international maritime authorities.

8. Acknowledgements

The authors sincerely thank the anonymous referees for their helpful comments and valuable suggestions, which considerably improved the exposition of this work. This study is sponsored by ‘Shuguang Program’ supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission (grant no. 16SG41).

References

Akten, N., 2004. Analysis of shipping casualties in the Bosphorus. Journal of Navigation, 57, 345–356.

Arslan, O. and Turan, O., 2009. Analytical investigation of marine casualties at the Strait of Istanbul with SWOT–AHP method. Maritime Policy & Management, *36*(2), 131-145.

Aydogdu, Y. V., Yurtoren, C., Park, J. and Park, Y., 2012. A study on local traffic management to improve marine traffic safety in the Istanbul Strait. Journal of Navigation, 65(1), 99–112.

Chung, Y., Lee, P.T., Lee, J., 2017. Burnout in seafarers: its antecedents and effects on incidents at sea. Maritime Policy & Management, 44(7), 916-931.

Cox, C., 2008. The generalized *F* distribution: An umbrella for parametric survival analysis. Statistics in Medicine, 27, 4301–4312.

Debnath, A. K., Chin, H. C., 2010. Navigational traffic conflict technique: a proactive approach to quantitative measurement of collision risks in port waters. Journal of Navigation, 63(1), 137–152.

Dempster, A. P., Larid., N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B (Methodological), 39(1), 1–38.

Eliopoulou E., Papanikolaou, A., 2007. Casualty analysis of large tankers. Journal of Marine Science and Technology, 12: 240–250.

Eliopoulou E., Hamann R., Papanikolaou A., Golyshev P., 2013. Casualty analysis of cellular container ships. Proceedings of the IDFS 2013, Shanghai, 25–27.

Goerlandt, F., Habtamnesh Goite, Osiris A. Valdez Banda, Anders Höglund, Paula Ahonen-Rainio, and Mikko Lensu. An analysis of wintertime navigational accidents in the Northern Baltic Sea. Safety science 92, 66-84.

Hassel, M., Asbjornslett, B.E., Hole, L.P., 2011. Underreporting of maritime accidents to vessel accident databases. Accident Analysis and Prevention, 43 (6), 2053–2063.

Hogg, S.A., Ciampi, A., 1985. GFREG: a computer program for maximum likelihood regression using the generalized F distribution. Computer Methods and Programs in Biomedicine, 20(2), 201-215.

Hojati, A.T., Ferreira, L., Washington, S., Charles, P., 2013. Hazard based models for freeway traffic incident duration. Accident Analysis and Prevention, 52, 171 – 181.

Karahalios, H. 2014. The contribution of risk management in ship management: the case of ship collision. Safety Science, 63, 104-114.

Jin, D., 2014. The determinants of fishing vessel accident severity. Accident Analysis and Prevention, 66, 1–7.

Jin D, Kite-Powell H, and Talley, W., 2001. The safety of commercial fishing: determinants of vessel total losses and injuries. Journal of Safety Research, 32(2), 209-228.

Jin D, Thunberg, E., 2005. An analysis of fishing vessel accidents in fishing areas off the northeastern United States. Safety Science, 43(8), 523-540.

Lee,D., Lee, S., Park, B., Kim, S., 2005. A study on the framework for survivability assessment system of damaged ships. Ocean Engineering, 32, 1122–1132.

Luo, M., Shin, S.H. and Chang, Y.T., 2017. Duration analysis for recurrent ship accidents. Maritime Policy & Management, 44(5), 603-622.
Talley, W. K., Jin, D., Kite-Powell, H.L., 2006. Determinants of the severity of passenger vessel accidents. Maritime Policy and Management, 33, 173–186.
Talley, W. K., Jin, D., Kite-Powell, H.L., 2008. Determinants of the severity of cruise vessel accidents. Transportation Research Part D, 13, 86–94.
Ozsoysal, R. and Ozsoysal, O. A., 2006. Maritime casualties through the Bosphorus. Naval Engineers Journal, 118(1), 77–82.
Perez-Labajos, C., Azofra M., Blanco B., Achutegui J., and Gonzalez J., 2006. Analysis of accident inequality of the Spanish fishing fleet. Accident Analysis and Prevention, 38(6), 1168-1175.
Roberts, S. E., Jaremin, B. and Marlow, P. B., 2010. Human and fishing vessel losses in sea accidents in the UK fishing industry from 1948 to 2008. International Maritime Health, 61(3), 143–153.
Uğurlu, Ö., Erol, S. and Başar, E., 2016. The analysis of life safety and economic loss in marine accidents occurring in the Turkish Straits. Maritime Policy & Management, *43*(3), 356-370.
Uğurlu, Ö., Köse, E., Yıldırım, U. and Yüksekyıldız, E., 2015. Marine accident analysis for collision and grounding in oil tanker using FTA method. Maritime Policy & Management, *42*(2), 163-185.
Uluscu, O. S., Ozbas, B., Altiok, T. and Or, L., 2009. Risk analysis of the vessel traffic in the Strait of Istanbul. Risk Analysis, 29(10), 1454–1472.
Yip, T.L., 2008. Port traffic risks – A study of accidents in Hong Kong waters. Transportation Research Part E, 44, 921–931
Yip, T.L., Jin, D., Talley, K.W., 2015. Determinants of injuries in passenger vessel accidents. Accident Analysis and Prevention, 82, 112-117.
Wang, J., Li, M., Liu, Y., Zhang, H., Zou, W., Cheng, L., 2014. Safety assessment of shipping routes in the South China Sea based on the fuzzy analytic hierarchy process. Safety Science, 62, 46–57.
Weng, J., Ge, Y., Han, H., 2016. Evaluation of shipping accident casualties using zero-inflated negative binomial regression technique. Journal of Navigation, 69(2), 433-448.
Weng, J., Meng, Q. and Qu, X., 2012. Vessel collision frequency estimation in the Singapore Strait. Journal of Navigation, 65, 207–221.
Weng, J., Yang, D., 2015. Investigation of shipping accident injury severity and mortality. Accident Analysis and Prevention, 76, 92-101.

Table 1 Recent literature focusing on the analysis of shipping accident

Authors	**Ship type**	**Study object**	**Water area**	**Methodology**
Jin et al. (2001)	Fishing vessel	Total loss and number of fatal and non-fatal crew injuries	U.S. Coast area	Probit and negative binomial regression
Jin and Thunberg (2005)	Fishing vessel	Accident probability	Northeast fishing areas of the United States	Logit regression
Roberts et al. (2010)	Fishing vessel	Fatal work-related accidents	UK Coast area	Descriptive method
Eliopoulou & Papanikolaou (2007)	Tankers	Degree of accident severity	worldwide	Descriptive method
Talley et al. (2006)	Passenger ships	Total loss, injuries and deaths/missing people	US Coast area	Tobit, negative binomial & Poisson regression
Talley et al. (2008)	Cruise vessel	Property damage and injury severities	US Coast area	Tobit regression method
Eliopoulou et al. (2013)	Containerships	Total loss	No specific region	Descriptive method
Akten (2004)	No special ship type	Marine casualties	Istanbul Strait	Descriptive method
Ozsoysal and Ozsoysal (2006)	No special ship type	Casualty and dominant factors of casualty	Istanbul Strait	Descriptive method
Uluscu et al. (2009)	No special ship type	Human Casualty, Property-infrastructure damage, Environmental Damage and Traffic Effectiveness	Istanbul Strait	Expert judgment, linear regression
Aydogdu et al. (2012)	No special ship type	Traffic management for improving maritime safety	Istanbul Strait	Marine Traffic Fast Time Simulation (MTFTS)
Debnath and Chin (2010)	No special ship type	Influencing factors of shipping collision risks	Singapore Port waters	Ordered probit regression models
Yip (2008)	Multiple type ships	Injuries and fatalities	Hong Kong Waters	negative binomial regression
Yip et al. (2015)	Passenger vessel	Crew and passenger injuries	U.S. Coast	Poisson regression
Jin (2014)	Fishing vessel	Vessel damage and crew injury severity		ordered Probit model
Weng and Yang (2015)	All ship types	Shipping accident severity and human life loss in fatal shipping accidents	Worldwide water areas	binary logistic regression, zero-truncated binomial regression
Weng et al. (2016)	All ship types	Human life loss	South China Sea	zero-inflated negative binomial
Goerland et al. (2017)	All ship types	Prevailing sea ice, meteorological conditions, winter operation type, angles and speeds	Gulf of Finland and Gulf of Bothnia	Visual data mining

Table 2 Variable descriptions

Explanatory variables	Descriptions	Mean	Stdev
Dependent variable			
Ship property damage cost	The cost of property damages to the ships involving in shipping accident ($\times 10^4$RMB)	69.854	171.707
Accident characteristics			
Collision	1 if a collision is occurred, 0 otherwise	0.521	0.500
Contact	1 if a contact is occurred, 0 otherwise	0.171	0.376
Grounding	1 if a grounding is occurred, 0 otherwise	0.104	0.311
Fire/explosion	1 if a fire/explosion is occurred, 0 otherwise	0.045	0.207
Sinking	1 if a sinking is occurred, 0 otherwise	0.064	0.245
Capsizing	1 if a capsizing is occurred, 0 otherwise	0.006	0.075
Accident location	0 for coastal_habour_port area, 1 for straits/sea area	0.412	0.492
Ship type			
Cargo/container ship	2 if two cargo/container ships are involved in a collision accident; 1 if a cargo or container ship is involved in the shipping accident, 0 otherwise	0.832	0.371
Fishing ship	2 if two fishing ships are involved in a collision accident; 1 if a fishing ship is involved in the shipping accident, 0 otherwise	0.024	0.153
LNG/LPG/Oil/Chemical ship	2 if two LNG/LPG/Oil/Chemical ships are involved in a collision accident; 1 if a LNG/LPG/Oil/Chemical ship is involved in the shipping accident, 0 otherwise	0.030	0.172
Other ship types	2 if two ships of other types are involved in a collision accident; 1 if a ship of other types is involved in the shipping accident, 0 otherwise	0.114	0.319
Navigational status	1 if underway, 0 if moored or docked	0.822	0.383
Ship tonnage	The tonnage of ship involving in the shipping accident (ton)	5699	15195.110
Environmental characteristics			
Visibility	2 for poor visibility, 1 for restricted visibility, 0 for good visibility conditions	0.415	0.769
Strong_wind/wave	1 if yes, otherwise 0	0.255	0.436
Accident time	1for the nighttime period, 0 for the daytime period	0.506	0.500
Human casual factors			
Judgment_error	1 for the judgment errors, otherwise 0	0.105	0.307
Lookout_failure	1 for the lookout failures, otherwise 0	0.349	0.477
Operation_error	1 for the operation errors, otherwise 0	0.551	0.498

Table 3 Statistical comparisons of ship property damage cost in shipping accidents

Categorical variable	Attributes	Mean ($\times 10^4$RMB)	StDev ($\times 10^4$RMB)	Test statistic	
				ANOVA test	Levene's test
Collision	=0(No)	61.2	129.92	2.90	1.77
	=1(Yes)	77.8	185.68		
Contact	=0(No)	78.6	184.52	15.90*	14.33*
	=1(Yes)	27.4	71.93		
Grounding	=0(No)	73.7	176.71	5.50*	4.65*
	=1(Yes)	36.5	115.56		
Fire/explosion	=0(No)	70.6	242.31	0.49	2.85*
	=1(Yes)	54.2	118.59		
Sinking	=0(No)	61.0	161.41	49.88*	11.30*
	=1(Yes)	198.5	250.20		
Capsizing	=0(No)	69.4	171.84	1.81	1.86
	=1(Yes)	156.9	126.16		
Navigational status	=0(Moored)	34.1	68.07	11.81*	10.54*
	=1(Underway)	77.6	185.83		
Visibility	=0(Good)	63.2	147.65	3.10*	2.76*
	=1(Restricted)	82.0	115.80		
	=2(Poor)	94.2	262.17		
Strong_wind/wave	=0(No)	60.2	135.59	11.64*	12.12*
	=1(Yes)	98.1	247.04		
Accident time	=0(Daytime)	63.1	204.68	1.86	4.08*
	=1(Nighttime)	76.4	266.26		
Accident location	=0(Coast/habour/port)	41.1	112.26	51.85*	45.75*
	=1(Straits/sea area)	110.8	225.39		

*Significance level of 0.05.

Table 4 Comparisons of best-fitting ship property damage cost estimation models with different distribution types

Model performance	*Weibull*	*Weibull with gamma heterogeneity*	*Gamma*	*Log-logistic*	*Generalized* F *distribution*	*Weibull with random parameters*	*Gamma with random parameters*	*Log-logistic with random parameters*	*Generalized* F *distribution with random parameters*
Log-likelihood	-2298.82	-2210.58	-2248.8	-2218.77	-2199.26	-2296.00	-2234.34	-2212.02	-2181.45
AIC	4639.64	4465.16	4537.60	4477.54	4442.52	4636.00	4512.68	4474.04	4414.90
BIC	4747.36	4578.01	4645.32	4580.13	4555.37	4748.85	4625.53	4602.27	4548.26

Table 5 Estimated coefficients of the variables included in the generalized *F* distribution model with random parameters

Variables	Coefficients	Standard errors	*p*-value
Intercept	0.453	0.215	0.02
Collision	0.580	0.163	<0.01
Contact	0.421	0.135	<0.01
Grounding*	0.102	0.163	0.50
Std. Dev#	0.067	0.036	0.05
Fire/explosion	0.820	0.197	<0.01
Sinking	3.066	0.289	<0.01
Capsizing	3.028	0.583	<0.01
Accident location*	0.656	0.093	<0.01
Std. Dev#	0.095	0.036	<0.01
Cargo/container ship	0.374	0.097	<0.01
Fishing ship	0.462	0.144	<0.01
LNG/LPG/Oil/Chemical ship	0.495	0.162	<0.01
Other ship types	0.348	0.109	<0.01
Navigational status*	0.419	0.104	<0.01
Std. Dev#	0.104	0.053	0.05
Ship tonnage	2.39×10^{-5}	2.93×10^{-6}	<0.01
Visibility*	0.041	0.049	0.35
Std. Dev#	0.045	0.030	0.10
Strong_wind/wave	0.230	0.100	0.02
Accident time	0.271	0.079	<0.01
Judgment_error	0.203	0.132	0.10
Lookout_failure	0.319	0.093	<0.01
Operation_error	0.129	0.081	0.11
m_1	31.075	16.763	0.05
m_2	8.501	4.803	0.06
σ	3.042	1.548	0.05

*Random parameter; #Standard deviation of normally distributed parameter.

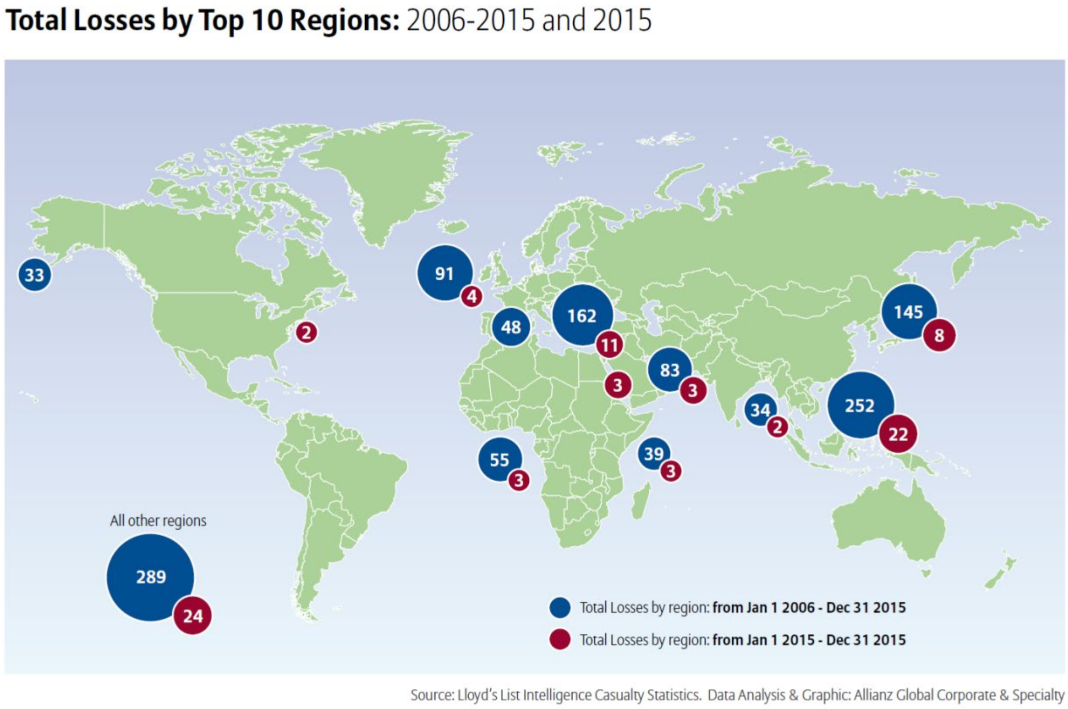

Fig. 1 Spatial distribution of shipping accident losses in top 10 regions

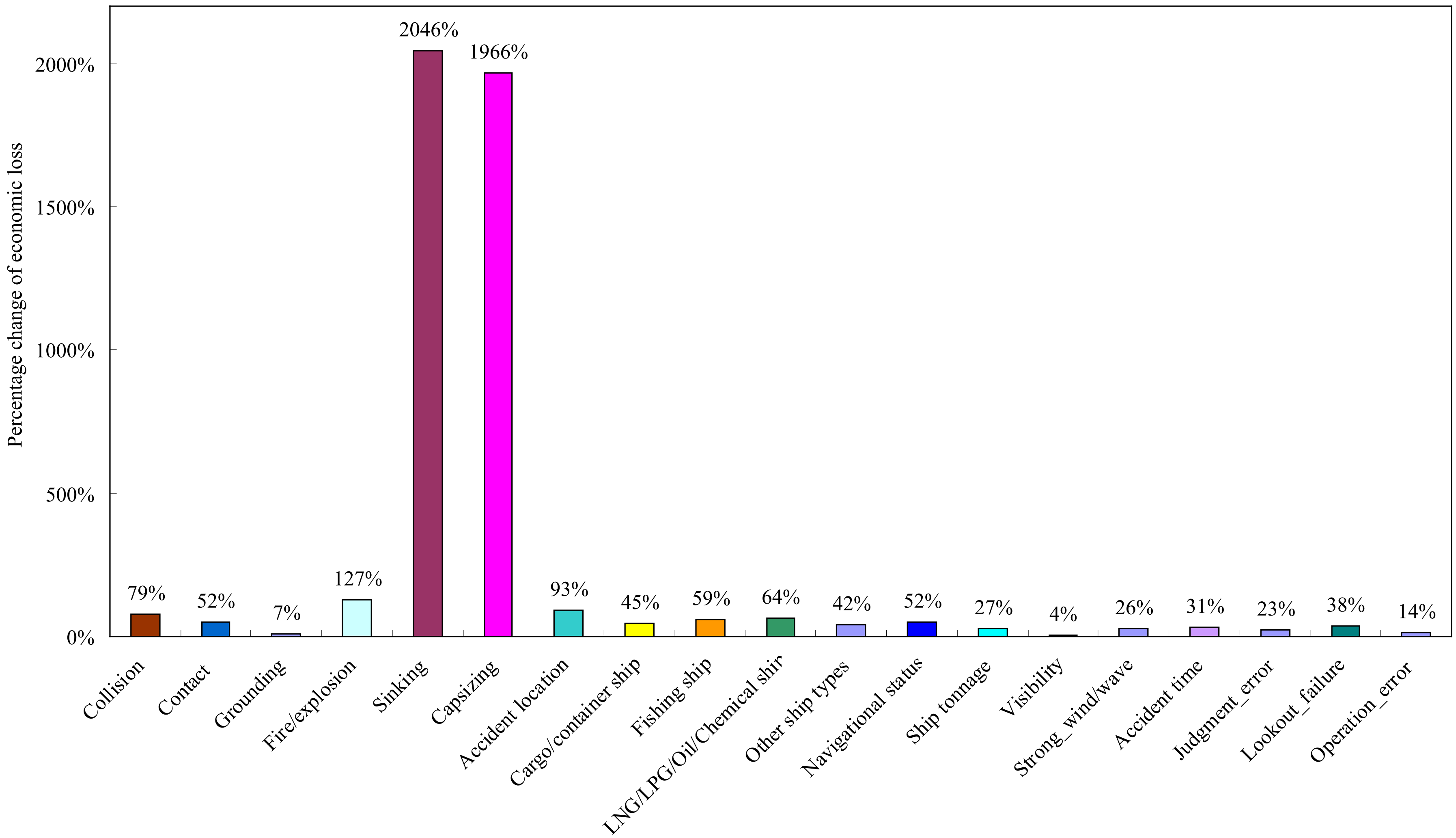

Fig. 2 Relative changes in ship property damage cost

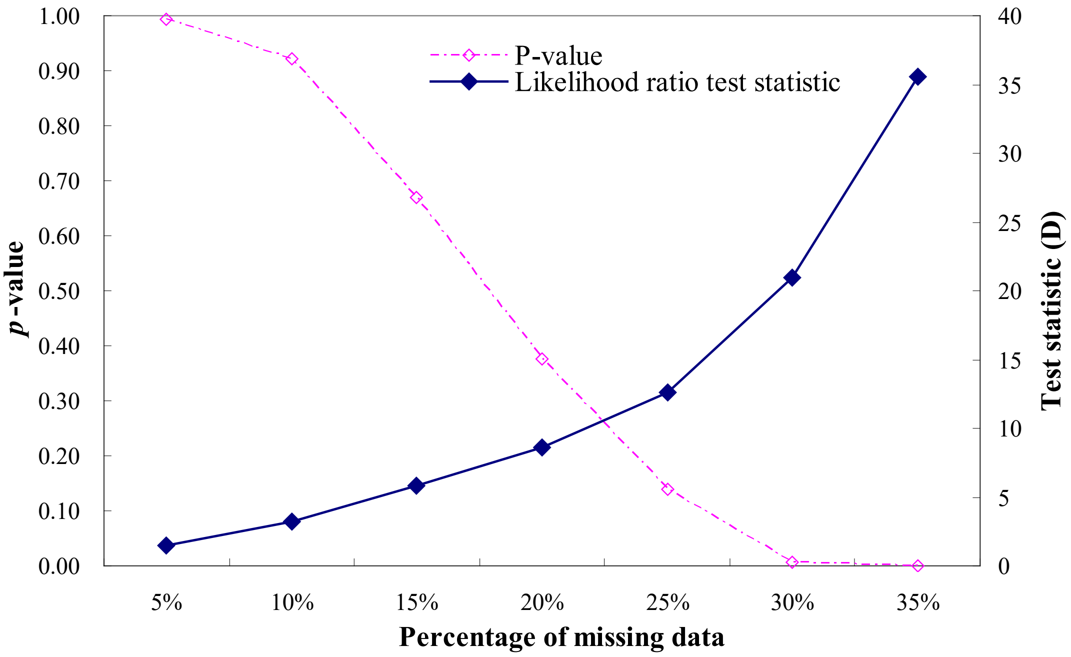

Fig. 3 Effects of missing data on the model reliability and applicability