Risk evaluation methods at individual ship and company level

Christiaan Heij[1] and Sabine Knapp[2]

Econometric Institute, Erasmus University Rotterdam

Econometric Institute Report 2011-23

Abstract

Safety management and risk profiling to identify substandard ships are of importance to the shipping industry. Whereas current methods rely heavily on detention risk and flag state performance, we extend the risk assessment by considering various risk dimensions and by evaluating a wide range of risk factors. Apart from detention risk, we consider also the risk of various types of accidents (total loss, very serious, and serious) and damage (hull and machinery, cargo, pollution, loss of life, and third party liabilities). Risk factors include ship particulars like ship type and classes of companies and owners, as well as historical information on past accidents, inspections, and changes of particulars. We present methods to summarize and visualize various risk dimensions and we pay particular attention to the identification of potentially risky companies. The empirical results are obtained by combining rich data sets with information on ship arrivals, inspections, and accidents for the period from 2006 to 2010. The presented methods and results are of interest to various stakeholders in the industry, such as charterers, insurance companies, maritime administrations, and the International Maritime Organization.

Keywords

Risk evaluation, shipping industry, accident risk, detention risk, company risk

[1] Corresponding author. Address: Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands. Phone: +31-10-4081264; Fax: +31-10-4089162; E-mail address: heij@ese.eur.nl

[2] Disclaimer for Knapp: The views expressed in this article represent those of the author and do not necessarily represent those of the Australian Maritime Safety Authority (AMSA). E-mail address: knapp@ese.eur.nl

1 Introduction

Safety management and risk profiling methods to identify and eliminate substandard ships are of importance for the shipping industry. Although this industry is highly regulated by more than fifty international conventions of the International Maritime Organization (IMO) and the International Labor Organization (ILO), enforcement of the standards can be weak due to the international nature of world shipping. A series of large-scale accidents in the 1970's caused IMO's attention to shift from prescriptive legislation to preventive action (Knapp and Franses, 2009). The International Safety Management (ISM) code, which came into force in 1998 and which was amended several times (IMO, 2010a), demands that the safety management of each vessel is performed by a dedicated company, the so-called 'Document of Compliance' (DoC) company. The IMO recently adopted a resolution (IMO, 2009) calling for the development of performance indicators, including statistical methods to measure and evaluate safety risk.

At present, risk profiling of substandard ships belongs to the tasks of port state control (PSC) regimes and vetting industries. Ten PSC regimes are active in conducting safety inspections, by applying generally agreed 'Memoranda of Understanding' to target individual ships for inspections. The effectiveness of these inspection strategies has been evaluated by Knapp and Franses (2007), Perepelkin et al. (2010), and Heij et al. (2011). Current port state control inspection strategies focus mostly on flag state performance related to detention risk, thereby ignoring accident risk that is sometimes accounted for by vetting inspection regimes. Quality differences between DoC companies may be relevant in risk evaluation, not only for PSC regimes but also, for example, for ship owners to limit liability in case of an accident. Although the relevance of companies is widely recognized, the statistical evaluation of company effects on risk is complicated by the large number of companies. Since the year 2000, about 72,000 companies have been founded that are associated with ship operations, including safety management, ownership levels, and commercial operations.[3] The latest development at PSC level, the European Union directive (EU, 2009; IMO, 2010b) which came into force in 2011, does not yet include statistical methods to quantify risk at individual ship and company level. As far as we know, there exist only a single contribution to evaluate

[3] In this paper, we will use the term company to denote a DoC company. For companies associated with ownership, we consider beneficial ownership as defined by IHS-Fairplay.

detention risk at the company level (Mueller, 2007), and no contributions to analyze accident risk at the company level.

The two main objectives of this paper are the following. First, we propose a methodology to help in answering the challenge posed by the IMO to develop statistical methods that are appropriate for the measurement and evaluation of safety risks in the shipping industry. Second, we analyze detention and accident risk both at the individual ship level and, in particular, at the company level. Our analysis involves over 5,000 companies, and we present various methods to evaluate the safety performance of these companies. The methodology and the results are of interest to various stakeholders in the shipping industry, such as charterers, maritime administrations, and the IMO. Charterers obtain a better picture of the safety level of ships for charter, and maritime administrators are helped to select ships for PSC inspections, flag state inspections or ISM company audits. We present various results on the individual company level, but we can not disclose any company names for reasons of confidentiality. The proposed methods can also be applied to analyze other risk factors, for example, owner-specific risk that is of interest to insurance companies such as P&I Clubs or marine underwriters to manage their risk portfolios.

The paper is structured as follows. Section 2 discusses the employed data sets on ship arrivals, inspections, detentions, and accidents. Section 3 presents the applied statistical methods. On the level of individual ships, detention and accident risks are modeled by means of logit models that account for measurable risk factors, including age, size, ship type, flag, company class, owner class, and others. To assess the risk on the company level, we compare the model-based risk (which accounts for measurable risk factors and which is averaged over all the vessels of the same company) with the empirically observed company risk. We propose three methods to evaluate company-specific risk: binomial p-values, difference of means, and log-odds residuals. If the actually observed risk of a company exceeds the expected (model-based) risk to a large extent, then this company deserves special attention as it is potentially relatively more risky than average. Section 3 presents also a simple graphical tool for two-dimensional visualization of several risk indicators. One dimension is for detention risk, and the other for the monetary value at risk that is the monetary value aggregated over five types of potential damage. Section 4 contains the empirical results for company-specific detention and accident risk and for the visualization of risks for selected sets of vessels, companies, and owners. Section 5 concludes. Four appendices present further background on the binomial p-value method, the interpretation of high risk scores, the risk performance of small and large companies, and risk scores of selected vessel groups.

2 Data

2.1 Data sources

For our analysis of detention and accident risk, we combine four data sets that all relate to the period from January 1 of 2006 until December 31 of 2010. Two data sets were provided by the Australian Maritime Safety Authority (AMSA). One of these sets contains data of well over 100,000 individual ship arrivals with ship particular information. The other AMSA data set consists of nearly 15,000 port state control inspections that are used to estimate and evaluate detention risk. Data on various types of accidents with nearly 20,000 observations were provided by AMSA, IHS Fairplay (IHSF), Lloyds Maritime Intelligence Unit, and the IMO. These data are used for the estimation and evaluation of the risk of two types of accidents: serious accidents, and total loss and very serious accidents. The accident data are merged with IHSF world fleet data of about 130,000 vessels, and the resulting dataset of about 280,000 observations includes information on various types of accident damage (hull and machinery, cargo, third party liability, pollution, and loss of life).

Several steps are performed to transform the original, raw data sets into a workable format. The accident data originate from various sources and need to be harmonized. Accidents are manually reclassified according to IMO definitions for seriousness (IMO, 2000), which are very serious (including total loss), serious, and less serious. The five damage types are manually reclassified according to Knapp et al. (2011). Company information of vessels is currently still rather scarce, notwithstanding various efforts at IMO level to improve the fleet coverage of company information and to implement company numbering schemes (IMO, 2011). Special care was given to the collection of relevant company information from the various data sources.

2.2 Risk indicators and risk factors

We distinguish two kinds of risk in our analysis, that is, preventive risk related to inspections and operative risk related to accidents. If inspection strategies are effective, inspected vessels of substandard quality have a relatively high chance to be detained. Therefore, the vessel-specific detention probability upon inspection indicates potential future risk. We obtain inspection and detention data from the AMSA arrival data set for 2006-2010. The total number of inspections and detentions is respectively 14,947 and 989, and these numbers are 9,088 and 646 for dry bulk carriers, and 5,859 and 343 for all other ship types. In our analysis of company effects, we remove inspection data of vessels with unknown company, leaving

1,020 companies with a total of 14,346 inspections and 920 detentions.[4] After aggregation of the five-year data to company level, the mean number of inspections per company is 14 and the mean number of detentions per company is .90, with a mean detention probability of about 8.3% per inspection.[5] Table 1 provides an overview of risk indicators on the company level. This table contains also model-based information that will be discussed in Section 3.1.

The operative shipping risk is measured in terms of accidents that are classified according to the seriousness of accidents (IMO, 2000), of which we consider the two most serious types, one for very serious accidents including total loss (denoted by TLVS) and the other for serious accidents (denoted by S). By joining various accident data sets, we obtain 278,194 observations in total, with 952 TLVS accidents and 5,895 serious accidents. If a vessel has multiple accidents of the same type and in the same calendar year, then only one of these accidents is taken into consideration. If the data are restricted to vessels for which the company is known, this leaves 129,733 vessels of 5,287 companies, with 153 TLVS accidents and 1,230 serious accidents.[6] The large reduction in the available number of data for vessels and accidents is due to missing company information for many vessels. After aggregation to company level, the mean empirical accident rates are about 0.2% for TLVS and about 1.4% for serious accidents. Table 1 provides some further information. Accidents can also be classified according to the type of damage resulting from the accident. We distinguish five types of damage (Knapp et al, 2011): hull and machinery, cargo, third party liability, pollution, and loss of life. Our damage type data set contains 278,194 observations, and the number of accidents with a specific damage type is 4,004 for hull and machinery, 72 for cargo, 2,498 for third party liability, 840 for pollution, and 617 for loss of life.

The data sets contain information on various factors that may affect the risk of detention and accidents. Our selection of factors is motivated by previous studies, and we refer to Knapp and Franses (2007), Cariou et al (2007), Bijwaard and Knapp (2009), and Heij et al (2011).The datasets contain information on various factors that may affect the risk of detention and accidents. Our selection of factors is motivated by previous studies, and we

[4] We omitted 400 companies with a single inspection, as it is not well possible to evaluate company-specific detention risk from a single observation. Of these 400 inspections, 43 resulted in detention and 357 not, so that the detention rate in this class is 43/400 = 10.8%.

[5] This number is obtained as average over the companies; as companies differ in the number of inspections (from 2 to 209), the outcome differs from the detention rate at inspection level that equals 920/14,346 = 6.4%.

[6] We omitted 662 companies with a single vessel, as it is not well possible to evaluate company-specific accident risk from a single observation.

Table 1: Descriptive statistics of empirical and model-based risk indicators on company level

	Companies	Min	Max	Sum	Mean
Detentions					
Number of vessels	1,020	1	659	63,167	61.93
Number of inspections	1,020	2	209	14,346	14.06
Number of detentions (empirical)	1,020	0	20	920	.90
Detention probability (empirical %)	1,020	.00	66.67	8,476	8.31
Number of detentions (model-based)	1,020	.01	16.16	941	.92
Detention probability (model-based %)	1,020	.58	38.61	7,521	7.37
Total loss and very serious accidents					
Number of vessels	5,287	2	659	129,733	24.54
Number of accidents (empirical)	5,287	0	2	153	.03
Accident probability (empirical %)	5,287	.00	33.33	972	.18
Number of accidents (model-based)	5,287	.00	3.18	226	.04
Accident probability (model-based %)	5,287	.00	6.03	1,197	.23
Serious accidents					
Number of vessels	5,287	2	659	129,733	24.54
Number of accidents (empirical)	5,287	0	11	1,230	.23
Accident probability (empirical %)	5,287	.00	50.00	7,508	1.42
Number of accidents (model-based)	5,287	.00	10.79	1,707	.32
Accident probability (model-based %)	5,287	.09	42.12	8,399	1.59

The (empirical and model-based) probabilities are expressed as percentage and they are obtained as averages per company, where the averages are taken over all inspections (for detentions) or all vessels (for accidents) of the same company. The model-based number of detentions (or accidents) is obtained by multiplying the relevant company-specific probability with the number of inspections (or vessels) of the company.

refer to Knapp and Franses (2007) and Cariou et al (2007) for detention risk factors, and to Bijwaard and Knapp (2009) and Heij et al. (2011) for accident risk factors. These factors include a wide range of ship particulars, including ship type (general cargo, dry bulk, container, tanker, passenger, and other), age, size, and group indicators for flag (traditional maritime nations, emerging maritime nations, new open registries, old open registries, international open registries), company (developed, transition, emerging, unknown), owner (developed, transition, emerging, unknown), class (IACS, that is, the International Association of Classification Societies, non-IACS, and unknown), and ship yard countries (risk assessment by professional inspectors as very high, high, medium, or low). The flag groups are taken from Alderton and Winchester (2002) and the company and owner groups from UNCTAD (2010). Past changes of flag, company, owner or class are also relevant. Another set of risk factors is derived from the ship history, including past inspections, detentions, and deficiencies, as well as past accidents including less serious accidents.

For the AMSA inspection data set with 1,020 companies, the mean vessel age is 12 years and the mean size is about 34,000 GRT. The distribution over ship types is 14% general cargo, 54% dry bulk, 10% container, 14% tanker, 2% passenger, and 5% other. For the accident data set with 5,287 companies, the mean vessel age is 19 years and the mean size is about 14,000 GRT. The distribution over ship types is 39% general cargo, 17% dry bulk, 6% container, 22% tanker, 5% passenger, and 10% other.

3 Methods

3.1 Model-based risk at individual ship level

In the foregoing section, we distinguished eight types of risk, that is, detention risk and seven accident risks: for TLVS and S accidents, and for five types of damage. The inspection data are for the Australian region where dry bulk vessels account for more than half of the inspections, so that we analyze detention risk for dry bulk apart from the other five ship types. For each of the nine resulting types of risk, we use the various data sources to estimate a logit model for the occurrence of an event, that is, a detention (in case of the inspection data) or an accident of specified type. Each logit model contains a set of risk factors that is obtained by first estimating the model that includes all possibly relevant factors, followed by down-testing until all remaining factors are significant (at the 5% level).[7]

Table 2 presents a partial list of results, and we mention some outcomes of interest.[8] Detention and accident risk increase with the age of the vessel. Size increases accident risk,

---

[7] For given values (x_1, …, x_k) of the k explanatory variables (including dummy coding for grouping variables), the logit model expresses the probability of an event by the logistic expression $P(\text{event}) = \exp(x\beta)/(1+ \exp(x\beta))$, where $x\beta = x_1\beta_1 + \ldots + x_k\beta_k$. The unknown parameters (β_1, …, β_k) are estimated from the data by quasi-maximum likelihood (Greene, 2008) to correct standard errors for possible misspecification of the assumed logistic distribution.

[8] We omit further details that are less relevant for this paper, as we focus on using the model-based probabilities in risk assessment. More detailed estimation results are available upon request from the authors.

Table 2: Partial list of factors affecting various risk aspects on individual ship level

		Detention		Accident		Damage type				
		DB	Other	TLVS	S	HM	CAR	TPL	POL	LL
Ship particulars										
Age		+	+	+	+	+		+	+	
Size		--	--	+	+	+	+	+	+	+
Ship type:	General cargo	n/a	+	BM	BM	BM	BM	BM	BM	BM
	Dry bulk	n/a	n/a	--	--	--	--	--		--
	Container	n/a	+		--	--			+	--
	Tanker	n/a	+	--	--	--	--	--	+	--
	Passenger	n/a	BM		+			+		+
	Other	n/a	n/a	+		--		+	+	+
Flag:	Traditional	BM	BM	BM	BM	BM	BM	BM	BM	BM
	Open registry		--	+	--			--		
	Emerging				--	--		--		--
	Unknown				--	--		--		
Company:	Developed	BM	BM	BM	BM	BM	BM	BM	BM	BM
	Transition			--	--	--		--	--	--
	Developing	--	+	--	--	--		--	--	--
	Unknown			+	+	+	+	+	+	+
Owner:	Developed	BM	BM	BM	BM	BM	BM	BM	BM	BM
	Transition			+	+			+	+	+
	Developing			+			--	+		+
	Unknown	--		+	+			+	+	+
Class:	IACS	BM	BM	BM	+	+		+		+
	Non IACS				+	+				
	Unknown				BM	BM	BM	BM	BM	BM
Changes:	Company	+		+	+	+			+	+
	Owner				+			+	+	
	Class			+	+	+	+		+	+
Ship history										
Past detentions		--		+	+	+				+
Past accidents:	TLVS					--				
	Serious			+	+	+		+	+	+
	Less serious			+	+	+	+	+	+	+
Number of observations		9,088	5,859	278,194		278,194				
Number of events		646	343	952	5,895	4,004	72	2,498	840	617

A '+' ('--') denotes a positive (negative) effect that is significant at 5% level; 'BM' denotes a benchmark group, and 'n/a' means that a variable does not apply. 'DB' denotes Dry Bulk, 'TLVS' total loss and very serious, 'S' serious, 'HM' hull and machinery, 'CAR' cargo (all ship types except passenger vessels), 'TPL' third party liability, 'POL' pollution (for tankers), and 'LL' loss of life (for passenger vessels). 'IACS' means International Association of Classification Societies.

but it decreases detention risk. Dry bulk vessels, container ships and tankers tend to have relatively smaller accident risk as compared to general cargo vessels. Of the four considered company groups, companies with unknown residency carry higher accident risk, and companies in the transition and developing residency groups carry lower risk. As concerns owner effects, accident risk is the lowest for the developed group. Changes in ship particulars and past risk events result in increased risk.

The model-based risk probabilities can be compared with the empirical risk frequencies. Logit models have the property that the average of the estimated individual probabilities is equal to the empirical risk rate, defined as the number of events divided by the total number of observations. After aggregation to company level, this does no longer hold true because the companies differ in size (number of inspections or number of vessels). Table 1 provides information on the model-based risk at company level, which is obtained by averaging the individual model probabilities per company. For the inspection and accident data, the average is taken respectively over all inspections and over all vessels of the company. The model-based mean detention risk (7.4%) is slightly smaller than the empirical detention risk (8.3%), whereas the model-based accident risk is slightly higher than the empirical accident risk (0.23% as compared to 0.18% for TLVS, and 1.59% as compared to 1.42% for S). These differences arise because large companies have larger weight than small companies before aggregation, whereas all companies have equal weight after aggregation. Appendix C provides a more detailed risk comparison of small and large companies.

3.2 Visualization of risk dimensions

The risk of a vessel or company can be evaluated in several ways, as was seen in the previous subsection. A simple tool to visualize the various risk dimensions into a two-dimensional graph is the so-called heat plot. One (horizontal) dimension of this graph is for preventive risk in terms of the detention probability (PDET), and the other (vertical) dimension integrates various accident risks in terms of the monetary value at risk (MVR). Heat plots can be of help to target ships for inspection and also for ISM audits to evaluate company performance.

The MVR provides an estimate of the expected total monetary value of five damage types (hull and machinery, cargo, third party liability, pollution, and loss of life), taking into account both the unconditional probability of an accident (of type TLVS or S) and the conditional probability of each damage type if an accident occurs. The MVR is computed as

$\text{MVR} = p_{acc} \times \sum_{j=1}^{5} p_j V_j$, where p_{acc} is the probability of an accident of type TLVS or S, p_j is the conditional probability of damage type j in case of an accident, and V_j is the monetary value of this damage type. The values V_j are constructed as in Knapp et al. (2011) and are based on the value of assets and cargo, third party liability limits, maximum insurance coverage for oil pollution, and insured value of life. The numerical values of p_{acc}, p_j, and V_j all depend on the vessel under consideration, and the involved probabilities are obtained from the logit models for accidents described in Section 3.1.

Heat plots based on the AMSA arrival data set will be presented in Section 4.3. These plots are either for individual arrivals or for company and owner averages, where the average is taken over all arrivals of the same company or owner. Each vessel (or company or owner) takes a risk position along the detention dimension and along the MVR dimension, corresponding to a point in the heat plot with coordinates (PDET, MVR). The color of this point shows the relative risk position, cold (blue) for low risk and hot (red) for high risk. These colors are defined in terms of the empirical two-dimensional cumulative distribution (CDF) of (PDET, MVR) for the set of all vessels. High risk corresponds to large CDF values, and low risk to small CDF values. The risk graduation is also expressed in terms of four numerical values, denoted by SW, NE, W, and S. Here SW (‘south-west’) is the CDF, that is, the percentage of vessels having smaller risk along both dimensions, that is, with detention risk smaller than PDET and with monetary value at risk smaller than MVR. In a similar way, percentages of vessels are given in the regions NE (‘north-east’, with larger risk along both dimensions), W (‘west’, with smaller detention risk), and S (‘south’, with smaller MVR). For example, if the two risk dimensions are independent, then a vessel with median values for PDET and MVR will have W = 50, S = 50, SW = 25, and NE = 25.

3.3 Methods for the evaluation of company-specific risk

3.3.1 Binomial p-value

As is shown in Table 1, the number of observations per company varies between 2 and 209 for detention risk and between 2 and 659 for accident risk. For a given company, let n denote the number of observations, y the number of events (detentions or accidents), and p_i the event risk for the i-th observation of this company. The risk probability p_i varies per observation, as they depend on the risk factors of the vessel. The risk of the company can be expressed in terms of the model-based probability to observe at least y events, under the assumption that the observations are independent. If this right-tail probability is small, this indicates that the company is potentially risky, as the actual number of events is large as compared to what is expected from the model-based probabilities that account for measurable risk factors like age, size, group, and past performance. Stated otherwise, a small right-tail probability indicates that the actual risk is higher than the model predicts. As we prefer a higher risk to correspond with higher risk scores, the company risk is measured in terms of the left tail probability, that is, the probability to observe at most y-1 events. The risk measure therefore corresponds to the p-value for a left-tail test on the proportion of events.

The above method for identifying potentially risky companies was proposed for the analysis of detention risk by Mueller (2007). Exact calculation of the p-value becomes intricate for two or more events, as the formulas depend on company size and our data set contains over 5,000 companies of widely ranging size. Approximate p-values are obtained from the binomial distribution with n experiments and with average event probability $\overline{p} = \sum_{i=1}^{n} p_i / n$. The binomial p-values would be exactly correct if the n risk probabilities are all the same, and Appendix A illustrates that the binomial approximation is accurate enough also for the within-company risk probability variations in our data. As the risk probabilities are small, Appendix A indicates that calculations could perhaps be simplified even further by using the Poisson distribution, but we will use the binomial distribution.

As was stated above, a large p-value indicates that the company is potentially risky. It should be realized that, even if all companies are actually equally risky, large p-values could occur by chance if we consider large sets of companies. One should therefore exercise care in interpreting high risk scores, as is further discussed in Appendix B.

3.3.2 Difference in means

Company-specific risk can also be evaluated by comparing the empirical risk rate with the model-based risk rate that accounts for vessel-specific risk factors like age, size, group, and past performance. Using the notation of the previous subsection, the empirical risk rate of a company is defined as the number of events divided by the number of observations, that is, $\overline{y}$ = y/n, and the model-based risk rate is the average event probability $\overline{p} = \sum_{i=1}^{n} p_i / n$. The binomial p-value method evaluates company risk in terms of the probability to observe $n\overline{y}$ or more events if the expected number of events is $n\overline{p}$. An alternative is to evaluate risk in terms of the difference in means $\overline{y} - \overline{p}$. The amount of variance of this difference depends on company size, so we need to correct for heteroskedasticity. For simplicity, and similar to the binomial p-value method, we approximate the distribution of y by the binomial distribution with n experiments and fixed event probability $\overline{p}$. The variance of $\overline{y} - \overline{p}$ then is $\overline{p}(1-\overline{p})/n$, and homoskedastic difference in mean scores[9] are obtained by

$$dm = \sqrt{n/(\overline{p}(1-\overline{p}))}(\overline{y} - \overline{p}).$$

Relatively risky companies correspond with relatively large positive values of dm. As many companies are quite small (n < 10 for more than half of the companies) and event probabilities are small (the mean of $\overline{p}$ is .07 for detentions and less than .02 for accidents), $n\overline{p}$ is often quite small and well below 1. This implies that the difference scores dm will not be normally distributed. In our applications in Section 4, we will not impose cut-off values to define 'high risk', and instead we will simply rank the difference scores and consider the companies with the highest risk scores.

3.3.3 Log-odds residuals

The model-based risk probabilities are obtained from logit models, as was described in Section 3.1. The logit model relates event risk p to risk factors x by $p = \exp(x\beta)/(1+\exp(x\beta))$. This expression is equivalent to $\ln(p/(1-p)) = x\beta$, where $\ln(p/(1-p))$ is called the log-odds, that is, the logarithm of the odds p/(1-p) of the probabilities of 'success' and 'failure'. The empirical log-odds is defined as $\ln(\overline{y}/(1-\overline{y}))$, where $\overline{y} = y/n$ is the empirical risk rate. The empirical log-odds is defined only for $0 < \overline{y} < 1$, and this restriction involves no loss for the risk classification of companies because companies with $\overline{y} = 0$ can not be declared to be risky and companies with $\overline{y} = 1$ are not present in our data. A company can be considered to be more risky the more its empirical log-odds exceeds its model-based log-odds. This suggests evaluating company risk in terms of the log-odds difference

$$dlo = \ln(\overline{y}/(1-\overline{y})) - \ln(\overline{p}/(1-\overline{p})).$$

In order to get a proper interpretation of the log-odds difference, we have to correct for heteroskedasticity and also for the sample selection bias caused by excluding companies

[9] One could also calculate studentized ('leave-one-out') residuals from the mean to detect high-risk companies, but as the number of observations is large, we will simply use the scores dm.

without events (that is, with $\bar{y}$ = 0).[10] Define z = ($\bar{y}$/(1-$\bar{y}$) / ($\bar{p}$/(1-$\bar{p}$)), so that dlo = ln(z) ≈ z-1 (using the first-order Taylor expansion at z = 1). As $\bar{y}$ and $\bar{p}$ are small (with mean values of about .08 for detentions and less than .02 for accidents), the variance of dlo is approximately equal to that of $\bar{y}/\bar{p}$. The variance of $\bar{y}$ is approximated by $\bar{p}(1-\bar{p})/n \approx \bar{p}/n$, so that $1/(n\bar{p})$ approximates the variance of dlo. We therefore correct for heteroskedasticity by multiplying dlo with the factor $\sqrt{n\bar{p}}$. We approximate the bias correction term as follows, by neglecting higher order terms of $\bar{y}$ and $\bar{p}$ and by using the binomial approximation with probability $\bar{p}$ for y, the number of events. The probability that $\bar{y}$ = 0 is $p_0 = (1-\bar{p})^n$, and if we condition on the information that $\bar{y}$ > 0 then the conditional probabilities P($\bar{y}$ = k/n | $\bar{y}$>0) are equal to P($\bar{y}$ = k/n)/(1–p_0), so that E($\bar{y}$ | $\bar{y}$>0) = $\bar{p}$/(1–p_0) and E($\bar{y}^2$ | $\bar{y}$>0) = ($\bar{p}(1-\bar{p})/n + \bar{p}^2$)/(1–$p_0$) ≈ ($\bar{p}$/n)/(1–$p_0$). We use this result for $z \approx (1/\bar{p})\bar{y}/(1-\bar{y}) \approx (1/\bar{p})(\bar{y}+\bar{y}^2)$, so that E(z | $\bar{y}$>0) ≈ (1+1/n)/(1–p_0) and hence E(dlo | $\bar{y}$>0) ≈ -1 + (1+1/n)/(1–p_0). If we combine this bias correction term with the heteroskedasticity correction factor $\sqrt{n\bar{p}}$, we obtain the following regression equation.

$$\sqrt{n\bar{p}}\left(\ln\left(\frac{\bar{y}}{1-\bar{y}}\right)-\ln\left(\frac{\bar{p}}{1-\bar{p}}\right)\right) = \mu + \alpha\sqrt{n\bar{p}} + \beta\frac{(1+\frac{1}{n})\sqrt{n\bar{p}}}{1-(1-\bar{p})^n} + \varepsilon$$

A company is relatively risky if its residual from the above regression equation is relatively large, as in this case the empirical log-odds are larger than would be expected from the risk factors that apply for this company.

We compare the above expression with the difference in means statistic of Section 3.3.2. As $\bar{y}$ and $\bar{p}$ are small, we get dm = $\sqrt{n/(\bar{p}(1-\bar{p}))}(\bar{y}-\bar{p}) \approx (\bar{y}-\bar{p})\sqrt{n/\bar{p}}$, and the left-hand side of the regression equation can be approximated by $\sqrt{n\bar{p}}\ln(\bar{y}/\bar{p}) \approx \sqrt{n\bar{p}}(\bar{y}/\bar{p}-1) = (\bar{y}-\bar{p})\sqrt{n/\bar{p}}$. This shows that the dependent variable in the above regression equation is approximately equal to the dm statistic. The two terms on the right-hand side of the equation are needed to correct for sample bias caused by the condition that $\bar{y}$ > 0, whereas this condition is not needed for the dm statistic.

[10] As the dataset contains no companies with $\bar{y}$ = 1, this restriction is not evaluated further.

4 Results

4.1 Company specific detention risk

The risk of companies consists of two parts. First, the risk differs between companies because of differences in fleet composition. Such differences can easily be identified by taking the average per company of the model-based risk obtained from the logit models in Section 3.1. We will now focus on the risk part that remains after correcting for the observed risk factors, by applying the three methods to estimate company-specific risk discussed in Section 3.3. In this subsection, we target companies with relatively high detention risk, and accident risk is considered in Section 4.2.

The results for detention risk are shown in the top part of Table 3 (columns for 'all companies'). Of the 1,020 companies, 553 are never detained, leaving 467 companies for the log-odds analysis. This also implies that 553 of the p-values are equal to zero. The non-zero p-values have a rather uniform distribution without clear right-tail cut-off points. The mean difference scores are skewed to the right but approximately normal for the set of companies with at least one detention, and the log-odds residuals are skewed to the left. The three indicators of company-specific detention risk are highly correlated. The correlation between p-value and mean difference is .92, between p-value and log-odds .87, and between mean difference and log-odds .81. The same finding holds true for relatively risky companies, that is, with relatively high scores on the three indicators, as can be seen from the three scatter diagrams on the top row of Figure 1.

We determine a set of companies with relatively high company-specific detention risk in the following way. For each risk indicator, we select the companies with risk scores belonging to the top 5%. As the total number of companies is 1020, with 467 companies for the log-odds indicator, the number of selected companies is chosen as 50 for the p-value and mean difference indicators and as 25 for the log-odds indicator. The three sets of high-risk companies show considerable overlap. A total of 40 companies have high risk according to both the p-value and the mean difference indicator. Of the 25 companies with high log-odds risk, 24 have high p-value and 20 have large mean difference score. The latter set of 20 companies has top 5% scores for all three risk indicators, and some results for this set of potentially high-risk companies are summarized in Table 3 (columns for 'top-risk'). The model-based detention risk is somewhat larger than average, but the difference in empirical risk is much more substantial. This indicates that the relatively high risk of these companies is not due to adverse measurable risk factors of the fleet composition but due to unobserved

Table 3: Descriptive statistics of three indicators of company risk

	All companies				Top-risk			
	Nobs	Min	Max	Mean	Nobs	Min	Max	Mean
Detentions								
P-value (left tail)	1,020	.00	1.00	.27	20	.94	1.00	.97
Mean difference	1,020	-2.63	4.69	.04	20	2.16	3.96	2.78
Log-odds residual	467	-6.49	1.71	.00	20	1.01	1.71	1.24
Detention probability (empirical %)	1,020	.00	66.67	8.31	20	12.25	66.67	39.16
Detention probability (model-based %)	1,020	.58	38.61	7.37	20	4.61	14.13	8.96
Total loss and very serious accidents								
P-value (left tail)	5,287	.00	1.00	.02	28	.71	1.00	.95
Mean difference	5,287	-1.58	31.43	-.04	28	1.68	24.43	7.32
Log-odds residual	144	-1.72	.92	.00	28	-.22	.92	.12
Accident probability (empirical %)	5,287	.00	33.33	.18	28	1.00	33.33	12.21
Accident probability (model-based %)	5,287	.00	6.03	.23	28	.03	1.22	.26
Serious accidents								
P-value (left tail)	5,287	.00	1.00	.10	14	.96	1.00	.99
Mean difference	5,287	-2.99	52.10	-.08	14	2.96	7.25	4.73
Log-odds residual	840	-7.36	5.65	.00	14	.97	1.71	1.27
Accident probability (empirical %)	5,287	.00	50.00	1.42	14	5.97	20.00	11.88
Accident probability (model-based %)	5,287	.09	42.12	1.59	14	.40	5.63	1.78

'Nobs' denotes the number of observations. The definition of 'top-risk' companies is provided in the text.

factors that may be related to the company. This type of company information is relevant for inspection authorities and for the evaluation of company performance in establishing safety standards. We mention that 9 of the 20 (45%) relatively high-risk companies are inspected less than 10 times, whereas the population share is 72% so that small companies (in terms of number of inspections) are not over-represented in the high-risk class. Further, 4 of the 20 (20%) have less than 10 vessels, as compared to a population share of 22%, so that small companies are also not over-represented if measured by the number of vessels.

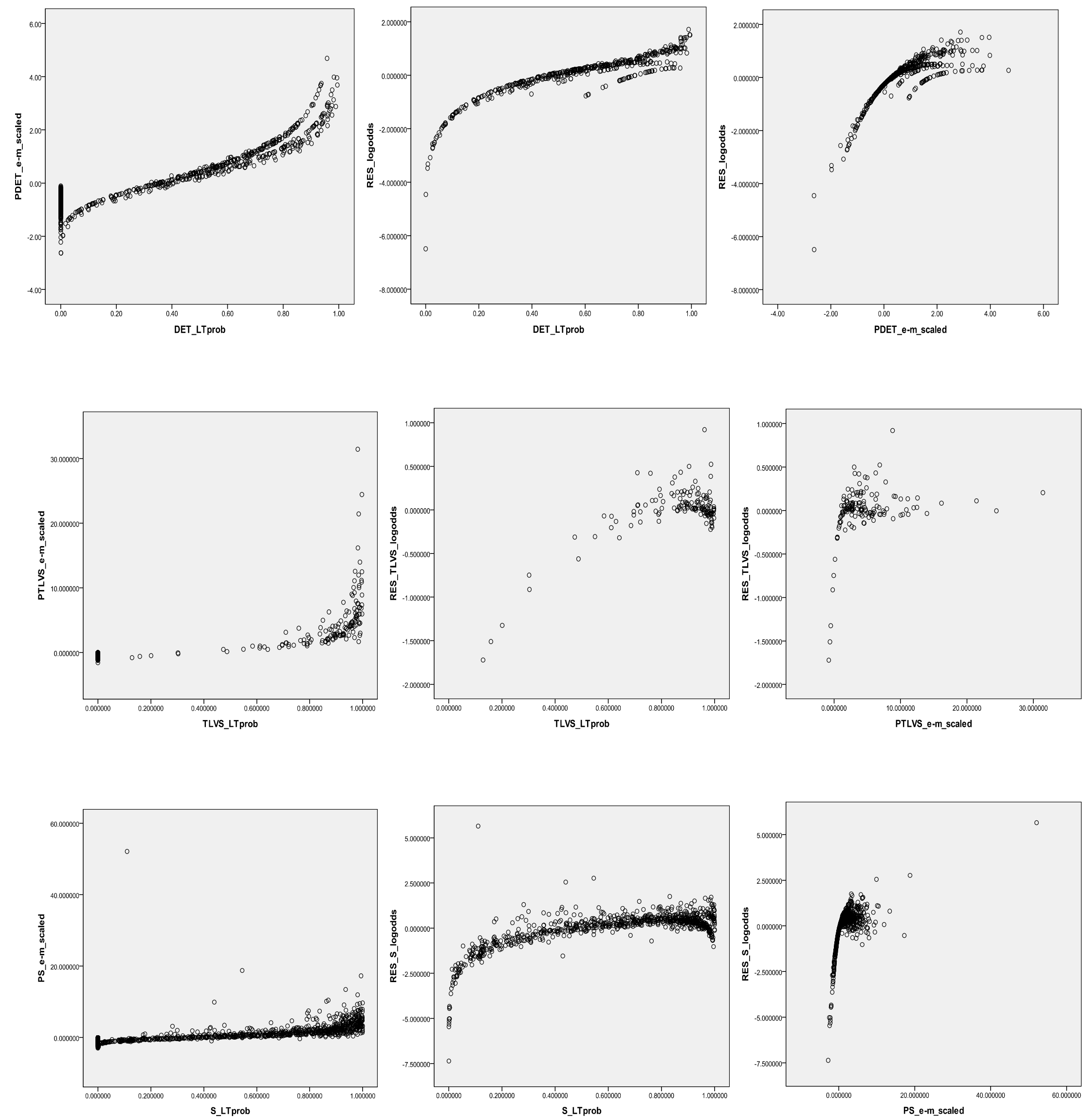

Figure 1: Scatter diagrams of three indicators of company risk

The top row is for detention risk, the middle row for risk of total loss or very serious accidents, and the bottom row for risk of serious accidents. The diagrams show p-value (horizontal) against mean difference (vertical) on the left, p-value (horizontal) against log-odds residuals (vertical) in the middle, and mean difference (horizontal) against log-odds residuals (vertical) on the right.

4.2 Company specific accident risk

We apply the same methodology as in the previous section now for the identification of companies with a potentially high risk for accidents of the types TLVS (total loss and very serious) and S (serious). The results for both types of accident risk are shown in the bottom part of Table 3 (columns for ‘all companies’). Of the 5,287 companies, 144 incur an accident of type TLVS and 840 of type S. This means that the number of log-odds residuals and of non-zero p-values is relatively small. The non-zero p-values show a clear peak at the right tail (near 1), especially for TLVS. The mean difference scores are skewed to the right and the log-odds residuals are somewhat skewed to the left. Most of the indicators of company-specific detention risk are highly correlated. For TLVS accidents, the correlation between p-value and mean difference is .77, between p-value and log-odds .69, and between mean difference and log-odds .28. For accidents of type S, these correlations are respectively .68, .69, and .56. Figure 1 shows that the three indicators do in general also agree for high risk scores, although the association is somewhat weaker than for detention risk. The company-specific risk-indicators of TLVS are only weakly related to those of S accidents, with an average over the 9 correlations for pairs of TLVS and S risk indicators of .10.

A set of companies with relatively high company-specific accident risk is determined as follows. For the mean difference indicator we select the top-250 companies with the highest risk scores (about 5% of the 5,287 companies). For the log-odds indicator, we select the top-50 (top-10) companies with the highest log-odds scores for S-type (TLVS-type) accidents (about 5% of respectively 840 and 144 companies). As the p-value indicator is non-zero for respectively 840 and 144 companies, we select roughly 10% of the highest p-values by selecting a top-100 (top-20) for S-type (TLVS-type) accidents. The association between the top-250 for the mean difference indicator and the two other indicators is very strong: for TLVS, all top-20 p-values and all top-10 log-odds also belong to this group, and for S this holds for 90 out of 100 p-values and for 47 out of 50 log-odds. The association between high p-values and high log-odds is weak, with an overlap of 2 of 10 for TLVS and of 14 of 50 for S. For TLVS accidents, we define a group of 28 top-risk companies by requiring that at least two of the three indicator scores belong to their respective top-groups. For S accidents, we select a group of 14 top-risk companies by requiring that all three indicators score in their respective top-groups. Results for these two sets of potentially high-risk companies are summarized in Table 3 (columns for ‘top-risk’). The model-based accident risk is somewhat larger than average, but the difference in empirical risk is much more substantial. The conclusion is the same as for detention risk, that is, the relatively high accident risk of these companies is not due to adverse measurable risk factors of the fleet composition but due to unobserved factors that are possibly related to the company.

4.3 Visualization of relative risk positions

For maritime practitioners, it is of interest to get an idea of the relative risk position of a vessel (or company or owner) within a specified population of vessels (or companies or owners). We use our arrival dataset to provide some illustrations of the use of heat plots for this purpose, which were discussed in Section 3.2. The employed data set covers both inspected and non-inspected arrivals, with a total of over 110,000 arrivals and nearly 6,000 companies. As detention risk (PDET) and the monetary value at risk (MVR) contain some severe right-tail outliers, we truncate these values to improve the visibility of the heat plots. Another option would be to use quantile scores, but we prefer to use truncated values as practitioners are interested in the magnitude (and not only in the ranking) of detention risk and of the monetary value at risk. After truncation, both axes of the heat plot are scaled by defining scaled values of MVR by $100 \times MVR/(MVR_{max} - MVR_{min})$, where $MVR_{max} - MVR_{min}$ denotes the range of the truncated MVR values, and PDET is scaled in an analogous way. The correlation between truncated detention risk and MVR is quite low, only 3.5% at the company level, so that the two risk measures do indeed measure different dimensions of risk.

The first illustration concerns the set of ten dry bulk arrivals in the port of Newcastle in Australia between July 1 and 3 of 2010, within the population of all 5,970 dry bulk arrivals in this port in the period 2006-2010. The values of PDET contain no outliers, but those of MVR do contain some outliers (the maximum is 10 standard deviations above the mean). We truncate 131 MVR values (2.2%) at 5 million dollar (about 3 standard deviations above the mean). One of the ten selected arrivals has an MVR of 6.9 million and hence belongs to the truncated class. Table 4 contains some summary statistics, and relative risk statistics are provided in Table A.4. The heat plot is shown in the top diagram of Figure 2. The vessel with an MVR of 6.9 million has also the highest detention risk (2.0%) in the set of ten arrivals. Within the population of all arrivals in Newcastle, this arrival falls deep in the hot zone, as 88.7% of all arrivals have a lower score on both PDET and MVR and there are no arrivals for which both PDET and MVR are higher than this one. The second-highest SW score is 61.3% for a vessel with an MVR of 1.6 million and detention risk of 1.5%. Seven of the ten arrivals carry relatively low risk.

Table 4: Descriptive statistics of monetary value at risk and detention probability

	Nobs	Share	Monetary value at risk			Detention probability		
			Min	Max	Mean	Min	Max	Mean
Arrivals								
All arrivals	110,245		.0	14.0	1.0	.1	11.9	1.3
Inspected arrivals	14,947		.0	14.0	.9	.1	11.9	1.3
Detained arrivals	989		.0	11.8	1.1	.2	9.6	1.9
Dry bulk arrivals in Newcastle								
All (2006-2010)	5,970		.0	14.0	.9	.1	5.6	1.3
July 1-3 of 2010	10		.1	6.9	1.1	.5	2.0	1.2
Companies								
All	1,888	.8	.0	8.3	.7	.1	7.1	1.5
High share cape size dry bulk	9	71.4	.8	2.8	1.4	.6	1.4	1.0
Owners								
All	1,405	.7	.0	10.4	.8	.1	11.9	1.5
High share cape size dry bulk	12	48.4	.7	4.7	2.2	.8	1.9	1.2

'Nobs' denotes the number of observations, and 'Share' denotes the percentage share of cape size dry bulk vessels. The monetary value at risk is expressed in millions of US dollars, and the detention probability is given as percentage.

The other two illustrations are on aggregated level, one for companies and the other for owners. Within a population of 1,888 companies, we consider the risk position of the nine companies with a share of at least 50% of cape size dry bulk vessels.[11] Further, within a population of 1,405 owners, we consider the risk position of the twelve owners with a share of at least 25% of cape size dry bulk vessels.[12] At the aggregate level, PDET and especially

[11] Stated more precisely, a company belongs to this class if at least 50% of its arrivals consist of cape size dry bulk vessels aged between 5 and 15 years and belonging to the following groups: developed UNCTAD vessel, developed owner, medium ship yard risk, open registry flag, and IACS class. The average share of cape size dry bulk arrivals of the nine companies is 71%, with a minimum of 51% and a maximum of 100%.

[12] Stated more precisely, an owner belongs to this class if at least 25% of its arrivals consist of cape size dry bulk vessels aged between 5 and 15 years and belonging to the following groups: developed UNCTAD vessel, developed company, medium ship yard risk, open registry flag, and IACS class. The average share of cape size dry bulk arrivals of the twelve owners is 48%, with a minimum of 25% and a maximum of 100%.

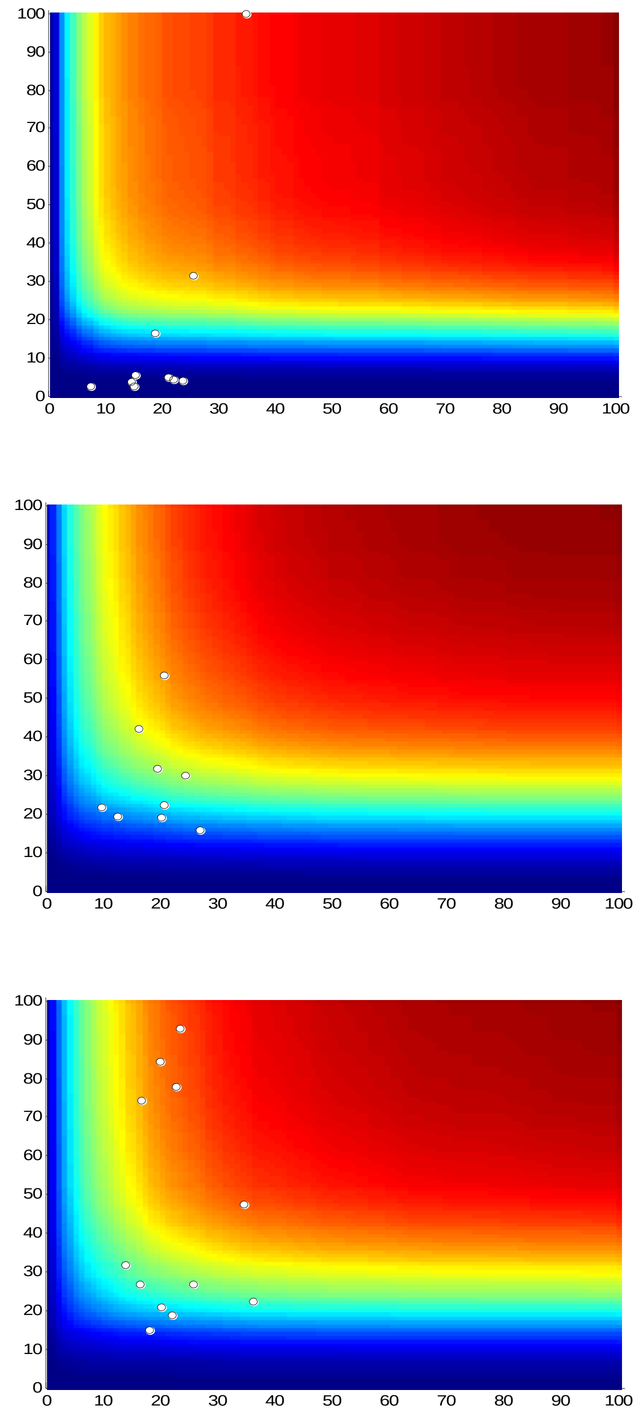

Figure 2: Heat plots for monetary value at risk (vertical) and detention risk (horizontal)

The top diagram is for 10 dry bulk vessels that arrived in Newcastle between July 1 and 3 of 2010; the middle diagram is for 9 companies with a fleet that consists for at least 50% of cape size dry bulk vessels; the bottom diagram is for 12 owners with arrivals that consist for at least 25% of cape size dry bulk vessels.

MVR have some right-tail outliers, and we truncate PDET at .05 (the maximum value is .07 for companies and .12 for owners) and MVR at 5 million. For the company data, 15 values of PDET and 10 values of MVR are truncated (both less than 1%), and for the owner data the number of truncations is 12 for PDET (less than 1%) and 19 for MVR (1.4%).

The resulting heat plots are shown in Figure 2, and Tables 4 and A.4 provide related risk statistics. The heat plot for companies identifies one company as relatively risky and three or four as mildly risky, whereas four companies have relatively low risk. The monetary value at risk of the nine selected companies is relatively high (the minimum value of .8 million is larger than the overall average of .7 million), and the detention risk is relatively low (the maximum value of 1.4% is smaller than the overall overage of 1.5%). Of the twelve owners, five are relatively risky and seven are relatively safe. Again, the MVR values are relatively large and the PDET values relatively low.

5 Conclusion

This paper provides suggestions to answer challenges posed by the IMO (IMO, 2009) for the development of performance indicators, including statistical methods to measure and evaluate safety risk. Measurable risk factors are incorporated in logit models for the risk of detention and for several types of accident risk, and the effect of these factors is estimated by combining various data sources on ship arrivals, inspections, detentions, and accidents. Special attention is given to the identification of potentially risky Document of Compliance companies that are responsible for safety management. Further, we propose a simple tool to visualize several risk aspects in terms of a heat plot that combines detention risk with five types of accident in terms of the monetary value at risk. The proposed methods of risk analysis can also be applied to flags (relevant for the IMO) and beneficial ownership (relevant for insurance companies and P&I clubs).

The practical usefulness of the discussed methodologies can be enhanced further by improving the quality of the data. At present, much relevant information is still lacking, for example, on companies and ownership. IMO numbering schemes are currently under review (IMO, 2011) and this might lead to improvement in the future. Further, the inspection data set is limited to Australia (AMSA region), due to lack of information from other regions. The IMO call for the development of a global integrated ship information system (IMO, 2010c) can lead to important improvements of the datasets also in this respect.

Acknowledgements

We thank IHS Fairplay (former Lloyds Register Fairplay), Lloyd's Maritime Intelligence Unit, and the International Maritime Organization for the use of accident and world fleet data, and the Australian Maritime Safety Authority for the use of their inspection and arrival data and for access to the report of Mueller (2007). We also thank Dave Nagle for his advice on data management.

References

Alderton, T., Winchester, N., 2002. Flag states and safety: 1997-1999. *Maritime Policy and Management* 29(2), 151-162.

Bijwaard, G., Knapp, S., 2009. Analysis of ship life cycles – The impact of economic cycles and ship inspections. *Marine Policy* 33(2), 350-369.

Cariou, P., Mejia, M.Q., Wolff, F.C., 2007. On the effectiveness of port state control inspections. *Transportation Research Part E* 44(3), 491-503.

Greene, H.W., 2008. *Econometric Analysis* (6-th ed.). Prentice Hall, New Jersey.

Heij, C., Bijwaard, G., Knapp, S., 2011. Ship inspection strategies: Effects on maritime safety and environmental protection. *Transportation Research Part D* 16(1), 42-48.

Knapp, S., Bijwaard, G., Heij, C., 2011. Estimated incident cost savings in shipping due to inspections. *Accident Analysis and Prevention* 43(4), 1532–1539.

Knapp, S., Franses, P.H., 2007. A global view on port state control - econometric analysis of the differences across port state control regimes. *Maritime Policy and Management* 34(5), 453-483.

Knapp, S., Franses, P.H., 2009. Does ratification matter and do major conventions improve safety and decrease pollution in shipping? *Marine Policy* 33(5), 826-846.

Mueller, W., 2007. Statistical modelling of ship detention based on ship inspection data 1996-2005. CSIRO Internal Report CMIS 2007/128, Commonwealth Scientific and Industrial Research Organisation, Canberra. (Report provided in confidence.)

Perepelkin, M., Knapp, S., Perepelkin, G., De Pooter, M., 2010. A method to measure flag performance for the shipping industry. *Marine Policy* 34(3), 395-405.

Other references

European Union, 2009. Port state control. Directive 2009/16/EC of the European Parliament and the Council of 23 April 2009, L 131/57, 28.05.2009.

IMO, 2000. Reports on marine casualties and incidents, revised harmonized reporting procedures. MSC/Circ. 953, MEPC/Circ. 372, adopted December 14, 2000.

IMO, 2009. Strategic plan for the organization (2010-2015). Assembly Resolution A.1011(26), adopted November 26, 2009.

IMO, 2010a. International safety management code and guidelines on implementation of the ISM Code, 2010 Edition.

IMO, 2010b. Harmonization of port state control activities. Paris MoU New Inspection Regime, FSI 18/Inf.2, March 26, 2010.

IMO, 2010c. Global integrated ship information system. Assembly Resolution A.1029(26), adopted November 29, 2009.

IMO, 2011. IMO ship identification number scheme and IMO unique company and registered owner identification number scheme. C 106/4/3/Rev.1, May 31, 2011.

UNCTAD, 2010. Review of Maritime Transport 2010. UNCTAD, Geneva.

Appendix A – On the binomial p-value method

Consider a company for which n observations are available, with event probability p_i for the i-th observation. As all probabilities are in principle different, computation of the probability of k events requires a summation of n!/(k!(n-k)!) terms, and computation of all event probabilities requires a total of 2^n summations. As n can be as large as 659 in our dataset, exact computations are infeasible. We illustrate that exact computations are also not needed, as binomial probabilities with the average event probability provide a good approximation. The binomial probabilities are exactly correct if all p_i take the same value, and we will therefore consider a (small) company with widely differing probabilities. In the set of 643 companies with at most nine inspections, the range between the maximum and minimum detention risk per company has mean .054 and standard deviation .046. We consider a company where this range is .147, about two standard deviations above the mean (only 25 of the 643 companies have a larger range). This company has six inspections for five vessels, as one vessel is inspected twice, and the detention probabilities are .020, .046, .046, .068, .107, and .167.

Table A.1 shows the results of various methods to compute the probability distribution of the number of detentions. The first row shows the exact distribution, and the second row contains the binomial distribution with average event probability .076. The last two rows show approximations of the binomial distribution with n = 6 and p = .076 by means of the Poisson distribution (with mean λ = np) and the normal distribution (with mean μ = p and variance σ^2 = np(1-p)). The table shows that the binomial approximation is quite accurate and that even the Poisson approximation performs reasonably well, whereas the normal approximation is not accurate. As the probabilities p_i are estimated and not exact, the differences between the binomial and exact probabilities become even less significant.

Table A.1: Outcomes of four methods to compute detention probabilities for a company

Detentions	0	1	2	3	4	5	6
Exact	.6176	.3159	.0606	.0056	.0003	.0000	.0000
Binomial	.6229	.3068	.0630	.0069	.0004	.0000	.0000
Poisson	.6344	.2887	.0657	.0100	.0001	.0000	.0000
Normal	.7435	.2425	.0140	.0001	.0000	.0000	.0000

Appendix B – On the interpretation of high risk scores

Section 3.3 describes three risk scores for the identification of potentially risky companies. Even if all companies are equally risky, large risk scores will occur by chance if we consider large sets of companies. As an illustration, we consider the p-value method for the risk of serious accidents of 5,287 companies. We compare the actual p-values obtained from the accident data set in Section 4.2 with the p-values that would be obtained if all companies were identical, in the sense that each company has the same number of vessels and the same accident probability. From Table 1 we obtain a mean number of vessels of 24.5 and a mean model-based accident probability of .0159, so we will consider the situation 5,287 companies that each have $n = 25$ vessels with equal accident probability $p = .0159$.

The top part of Table A.2 shows the probability distribution and the corresponding left-tail p-values for the number of accidents for a single company, that is, for the binomial distribution with parameters $n = 25$ and $p = .0159$. The p-value increases with a few jumps, and we define p-value intervals by means of the midpoints between jump points. Table A.2 shows the expected counts in each of seven intervals under the assumption of identical companies. For example, we expect to find 38 companies with a p-value larger than .966 and 3 with a p-value larger than .996, even though all companies are equally risky. Table A.2 shows also the counts for the actual p-values of Section 4.2, with 55 p-values larger than .966 and 15 p-values larger than .996. The cumulative distributions are shown in more detail, especially for the right tail, in Figure A.1. The message is that a high risk score of a company does not automatically imply high risk, as the high score might be a matter of bad luck. Instead, the number of high risk scores is relevant, and a set of companies with high risk scores may contain companies that perform badly so that such companies deserve attention.

Table A.2: Approximate expected company frequencies for high risk of serious accidents

Detentions	0	1	2	3	4	5	>= 6
Identical companies							
Probability (%)	66.985	27.057	5.246	.650	.058	.004	.000
P-value (left tail)	0	.670	.940	.993	.999	1.000	1.000
P-value interval lower bound	0	.000	.805	.966	.996	.999	1.000
P-value interval upper bound	0	.805	.966	.996	.999	1.000	1
Expected count	3,541	1,431	277	34	3	0	0
Cumulative count (right tail)	5,287	1,746	315	38	3	0	0
Actual companies							
P-value (left tail) in interval	4,447	494	253	75	8	9	1
Cumulative count (right tail)	5,287	840	346	93	18	10	1
Difference cumulative counts	0	-906	31	55	15	10	1

The figures for 'identical companies' are based on the assumption that all 5,287 companies have the same number of vessels (25) and the same accident probability (0.0159). The figures for 'actual companies' show the number of empirical p-values (obtained for the actual companies) in the indicated intervals, as well as the corresponding right tail cumulative counts and the difference of these counts with the cumulative counts obtained for identical companies. For the p-value intervals, the lower bound .000 (1.000) allows for any positive value not strictly equal to 0 (any value larger than 0.9999).

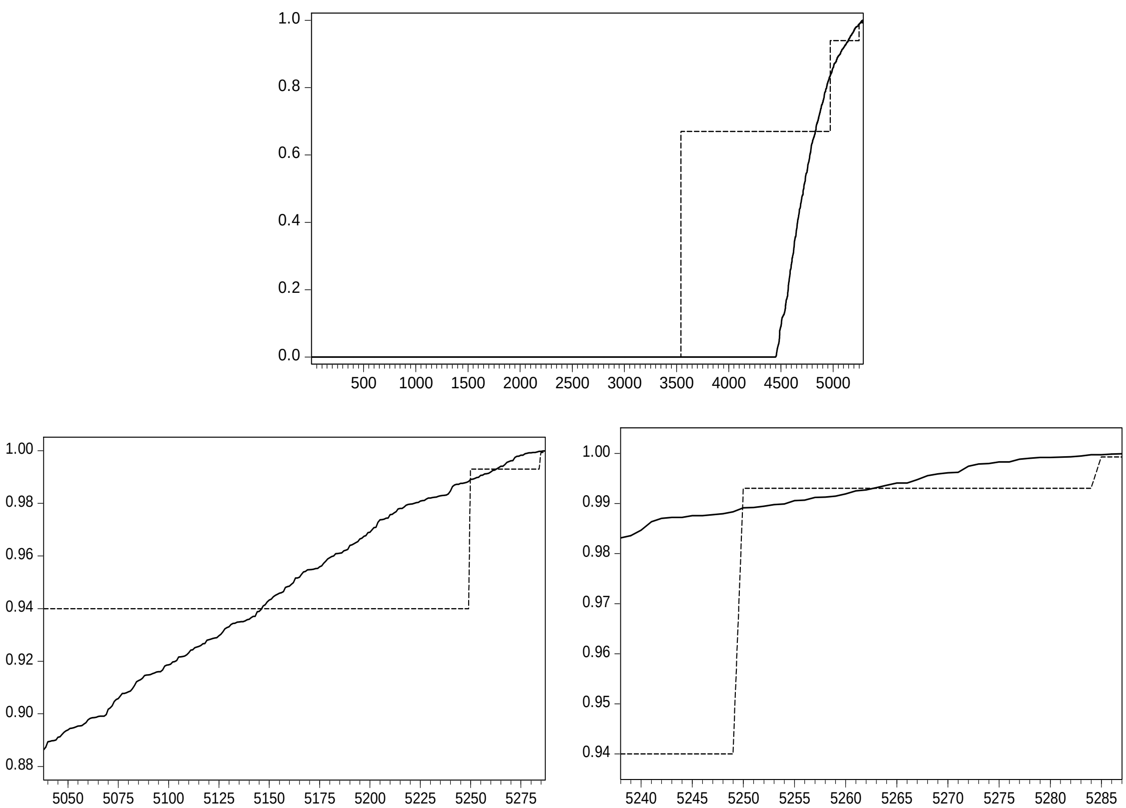

Figure A.1: Cumulative distributions of empirical and simulated left-tail p-values

The simulated p-values (dotted lines) are based on the assumption that all companies have the same number of vessels and the same accident probability. The empirical p-values (solid lines) apply for the actual companies. The two bottom diagrams show parts of the upper diagram, on the left (right) for the largest 250 (50) p-values.

Appendix C – Risk results for large and small companies

The random variation in empirical company risk depends on size of the company, that is, on the number of inspections for detention risk and on the number of vessels for accident risk. The standard deviation of the empirical risk rate is inversely proportional to the square root of the company size size. This means that the risk evaluation of large companies is less uncertain than that of small companies. We present risk scores for two groups, small companies with less than ten observations and large companies with at least ten observations. This results in 637 small companies and 383 large ones for detention risk, and 2,721 small companies and 2,566 large ones for accident risk.

Table A.3 provides a summary of results. Small companies have a somewhat larger detention risk than large companies. This holds true for both the empirical risk (4.5% difference) and for the model-based risk (1.6% difference), and the differences are shown in more detail in the box plots in the top row of Figure A.2.[13] If company size is defined in terms of the number of vessels, the differences between small and large companies diminish to 2.2% for the empirical risk and to 0.4% for the model-based risk. The three indicators for company-specific risk do not differ much between the two groups. Further, the groups of small and large companies have comparable TLVS accident risk scores, and for serious accidents the only notable difference is that small companies have a somewhat larger empirical risk (1.6%) as compared to large companies (1.2%). The box plots for both types of accidents in Figure A.2 do also indicate roughly similar distributions for both groups of companies.

[13] We compared the empirical detention counts for small companies with the expected counts obtained for the overall detention probability of 8.31% for all companies (see Table 1). The actual counts for 0, 1, 2, and 3 or more detentions are respectively 442, 156, 33, and 6, and the expected counts are respectively 454, 153, 26, and 4. High detention counts occur slightly more often than expected, but the differences are not large.

Table A.3: Risk comparison of small and large companies

	Small companies			Large companies		
	Min	Max	Mean	Min	Max	Mean
Detention						
Number of vessels	1	438	32	1	659	112
Number of inspections	2	9	4	10	209	31
Detention prob. (empirical %)	0.00	66.67	10.98	0.00	28.58	6.52
Detention prob. (model-based %)	0.58	38.61	8.08	0.80	20.23	6.48
P-value (left tail)	0.00	0.99	0.22	0.00	1.00	0.35
Mean difference	-1.14	4.69	0.09	-2.63	3.98	-0.05
Log-odds residual	-2.24	1.52	0.29	-6.49	1.71	-0.21
Accident						
Number of vessels	2	9	5	10	659	46
TLVS: Acc. prob. (empirical %)	0.00	33.33	0.19	0.00	33.33	0.18
Acc. prob. (model-based %)	0.00	6.03	0.24	0.02	3.98	0.21
P-value (left tail)	0.00	1.00	0.01	0.00	0.99	0.04
Mean difference	-0.44	24.43	-0.02	-1.58	31.43	-0.06
Log-odds residual	-0.22	0.26	-0.02	-1.72	0.92	0.00
S: Acc. prob. (empirical %)	0.00	50.00	1.64	0.00	50.00	1.19
Acc. prob. (model-based %)	0.09	40.73	1.57	0.13	42.12	1.61
P-value (left tail)	0.00	1.00	0.07	0.00	1.00	0.14
Mean difference	-1.18	17.26	0.04	-2.99	52.10	-0.21
Log-odds residual	-1.02	1.10	0.23	-7.36	5.65	-0.07

A company is classified as small (large) if it has less than 10 (at least 10) inspections (in case of detention risk) or vessels (in case of accident risk). The number of small (large) companies is 637 (383) for detention risk and 2,721 (2,566) for accident risk. 'TLVS' denotes total loss and very serious accidents, and 'S' serious accidents.

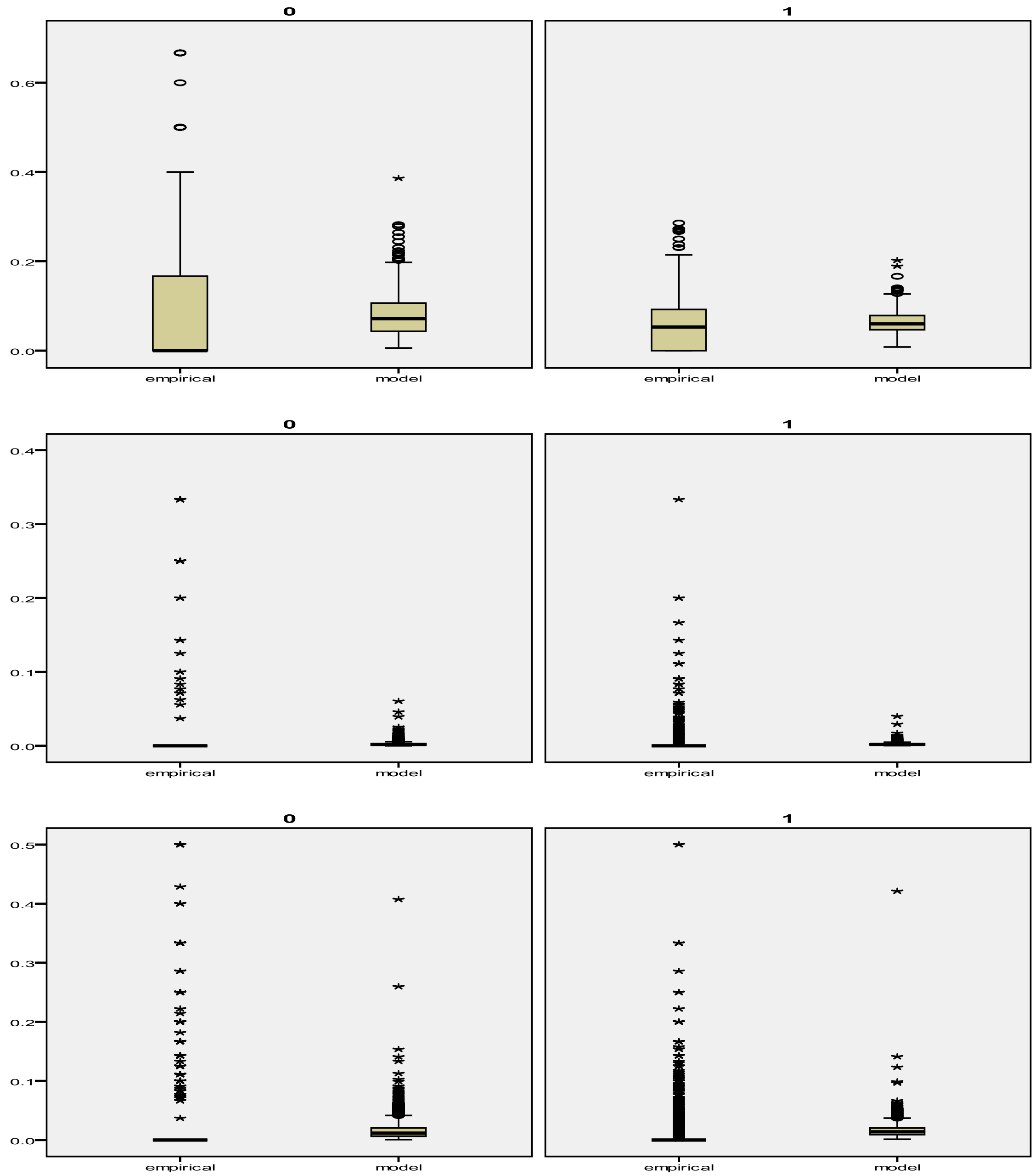

Figure A.2: Box plots of risk of small companies (left) and large companies (right)

In each of the six panels, the left box is for empirical risk and the right box for model-based risk, and in each of the three rows, the left panel is for small companies (with less than 10 inspections for the top row and with less than 10 vessels in the middle and bottom row) and the right panel is for large companies (with at least 10 inspections for the top row and with at least 10 vessels in the middle and bottom row). The top row is for detention risk (637 small and 383 large companies), the middle row is for total loss and very serious accidents and the bottom row for serious accidents (both with 2,721 small and 2,566 large companies).

Appendix D – Risk scores of selected groups of vessels, companies, and owners

Table A.4: Risk scores of selected groups of vessels, companies, and owners

Group	Share	Narr	MVR	PDET	SW	NE	W	S
Vessels								
(Dry bulk arrivals			6.9	2.0	88.7	.0	88.7	100.0
in port Newcastle,			1.6	1.5	61.3	5.8	71.7	83.9
July 1-3 of 2010,			.8	1.2	38.9	14.1	49.1	75.7
10 vessels)			.2	1.4	25.3	20.0	68.5	36.8
			.3	1.3	24.1	22.5	57.5	44.1
			.2	1.3	22.4	23.3	62.3	36.8
			.3	1.0	12.2	36.3	31.7	44.1
			.2	.9	10.4	41.9	31.7	36.8
			.1	1.0	7.5	49.0	31.7	26.8
			.1	.5	.6	88.0	3.9	8.7
Companies								
(At least 50%	71	52	1.5	1.3	42.1	7.3	47.2	87.6
cape size dry bulk	100	5	.8	1.4	41.7	13.8	55.5	72.5
arrivals 2006-10,	52	147	2.8	1.1	37.4	2.7	38.4	96.3
9 companies)	64	217	1.6	1.1	32.0	7.5	35.8	88.8
	53	238	1.1	1.1	31.2	11.9	38.4	80.9
	100	3	1.0	1.1	27.7	15.0	35.8	76.9
	51	344	2.1	.9	22.2	4.9	23.7	93.7
	100	1	1.0	.7	10.2	20.2	13.1	76.9
	52	25	1.1	.6	7.2	17.4	8.9	80.9
Owners								
(At least 50%	33	3	2.4	1.8	67.8	2.8	71.5	93.5
cape size dry bulk	33	36	1.1	1.9	59.1	6.3	73.0	79.8
arrivals 2006-10,	26	147	1.3	1.4	47.5	9.3	54.4	83.8
12 owners)	26	35	4.7	1.3	45.9	1.3	46.2	98.4
	100	2	3.9	1.2	45.7	2.1	46.2	97.4
	42	12	4.2	1.1	36.4	1.9	36.7	97.9
	100	3	.9	1.2	34.7	15.4	43.1	76.2
	42	72	1.0	1.1	30.7	15.3	36.7	78.7
	75	12	3.7	.9	26.8	2.5	27.0	97.3
	46	41	.7	1.0	23.2	23.8	29.8	69.5
	33	109	1.3	.9	20.4	13.4	23.2	83.8
	25	95	1.6	.8	16.4	10.5	17.7	88.3

‘Share’ denotes the percentage share of dry bulk cape size in arrivals, and ‘Narr’ is the number of such arrivals. ‘MVR’ is the monetary value at risk, expressed in millions of US dollars, and the detention probability ‘PDET’ is given as percentage. ‘SW’, ‘NO, ‘W’, and ‘S’ are expressed as percentage; the meaning of these risk scores is explained in the text, and within each group the data are ordered with decreasing SW scores.