

Improved strategies for the maritime industry to target vessels for inspection and to select inspection priority areas

Sabine Knapp¹ and Christiaan Heij²

EI2019-21

Abstract

Inspection authorities such as Port State Control Memoranda of Understanding use different policies and targeting methods to select vessels for inspections and rely primarily on past inspection outcomes. One of the main goals of inspections is to improve the safety quality of vessels and to reduce the probability of future incidents. This study shows there is room for improvement in targeting vessels for inspections and in determining vessel-specific inspection priority areas (e.g., bridge management versus machinery related items). The proposed approach treats detention and incident types as separate risk dimensions and evaluates seven targeting methods against random selection of vessels using empirical data for 2018. The analysis is based on three comprehensive data streams that cover the world fleet and shows potential gains (reduction of false negative events) of 14-27 percent compared to random selection. This can be further improved by adding eight inspection priority risk areas that help inspectors to focus inspections by providing insight in the individual risk profile of vessels. Policy makers can further customize the approach by classifying the risk of vessels into categories and by selecting inspection targets and benchmark samples. A small application example is provided to demonstrate feasibility of the proposed approach for policy makers and inspection authorities.

Keywords

Vessel inspection policy; vessel-specific risk; detention risk; incident risk; priority areas; false negatives

¹ Corresponding author: Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands, phone +61-466827029, email knapp@ese.eur.nl

² Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands, phone +31-10-4081264, email heij@ese.eur.nl

1. Introduction

This manuscript evaluates the status quo practice of the maritime industry to primarily use past inspection outcomes to target risky vessels for inspections (Knapp, 2006). It builds on previous work of Knapp and Franses (2007a), Heij et al. (2011), and Heij and Knapp (2019) to select vessels for inspections with the highest benefit (reduction in risk) and to treat detention and incident risk as separate dimensions in order to reduce *false negative events*. A false negative event is an event where the targeting method classifies a vessel as low risk but in reality this vessel has high risk. If a vessel is classified as low risk based on detention risk only, the vessel does not get targeted but can experience an incident with (very) serious consequences. If this happens, the inspection regime failed the opportunity to inspect the vessel and to improve its safety quality because inspections reduce incident risk, as is well established in the literature (Knapp and Franses 2007b, Bijwaard and Knapp 2009). Perepelkin et al. (2010) and Ji et al. (2015) demonstrate that using incident data besides detention data improves the evaluation of registries and subsequent selection of ships for inspections. The included incident type risks are those classified as total loss and very serious and serious incidents (VSS) according to IMO definitions (IMO, 2000). Less serious incidents and near misses are excluded here due to their high degree of underreporting (Hassel et al., 2011).

The ultimate goal of inspections is raising the safety quality of vessels to prevent future incidents rather than preventing future detentions. By combining the two risk dimensions, false negative events can be reduced since the riskiest vessels benefit most from an inspection. This study is based on three data streams containing global incident, inspection, and ship particular data of the world fleet (73,905 vessels). The analysis confirms the importance of the two risk dimensions and the need to improve targeting ships for inspections, since during the year 2018, 60 percent of vessels that experienced VSS incidents were not selected for inspection up to three months prior to the incident. The analysis also highlights the need to improve selection of vessel-specific inspection priority risk areas, since 40 percent of the vessels there were inspected still had incidents and only 4 percent were detained. In addition, one can observe a very low correlation (-0.04) between the probabilities of detention and incidents (VSS) for the year 2018.

With the above philosophy in mind to treat detention and incident (VSS) as separate risk dimensions and with an established need for improvement, the presented approach extends Heij and Knapp (2019) who consider three combined methods and use three months of empirical data (the first quarter of 2018) to evaluate them. The approach presented here uses five combined methods and evaluates them along with the eight priority areas using empirical data for the whole year of 2018. Furthermore, the presented approach allows for the combination of an automated data-driven part with qualitative aspects. Data-driven approaches have been proposed previously by Knapp (2006), Knapp and Franses (2008), and Li et al. (2014) but exclude other components included here, such as accounting for two different risk dimensions and addition of inspection priority areas. The data-driven part here provides the risk profiles of individual vessels for eight vessel inspection priority risk areas and combines detention and incident (VSS) risk into one metric to enhance targeting. This data-driven part can be combined with other intelligence and expert knowledge of inspectors to finalize inspection selection and execution. An application example is provided to demonstrate feasibility.

2. Data streams and methodology

The analysis is based on a unique and comprehensive combination of data streams and methods and follows a systematic step by step approach as follows:

- **Step 1:** Use logit models to estimate risk formulas based on data from 2010 to 2014 covering the global fleet. This step builds on previous work by Knapp (2015).
- **Step 2:** Use the derived risk formulas of step 1 to estimate probabilities at ship level for each of the four quarters of the year 2018. Ten models are selected that cover targeting and vessel inspection priority areas resulting in a total of 2.9 million probabilities.
- **Step 3:** Use the probabilities of step 2 to calculate percentile ranks of vessels with the global fleet as benchmark sample and integrate these percentiles by means of five methods that combine detention and (VSS) incident percentiles. The percentile ranks of the inspection priority areas are also calculated but not combined with detention. In addition, percentile ranks are also used to classify vessels using five risk categories (1=very high to 5=very low risk). The output of this step is 4.4 million percentile ranks and form the basis for evaluation and validation.
- **Step 4:** Validate the targeting methods of step 3 against empirical data from 2018 using three validation variables (detention only, incidents only, and detention and incidents combined).

Table 1 lists the data sources, time frames, and number of observations for the data streams that form the basis for the four steps mentioned above. The global inspection data comprises data from over seventy countries from eight Port State Control Memoranda of Understanding (MoU's). Incident data had to be combined and manually reclassified using IMO definitions (IMO, 2000) since different data providers use different definitions of seriousness. In addition, for each incident the first event of the chain of events was identified since that is needed for the incident type models for the eight-vessel inspection priority risk areas. In addition, the various types of consequences are recorded for each incident.

Table 1: Data streams and data sources used in the analysis

Data streams used for:	Time frame and number of observations	Data types and sources
Estimating risk formulas (<i>incident, VSS and 8 incident types</i>)	Jan 2010 to Dec 2014 376,508 total observations 8,874 VSS incidents	Ship particular data from IHSM Global PSC inspection data Global incident data from IMO, IHSM and LLI
Estimating risk formulas (<i>detention</i>)	Jan 2010 to Dec 2014 158,187 inspections 6,458 detentions	
Estimating probabilities (<i>detention and incident types</i>)	Dec 2017 to Sept 2018 73,905 vessels per period 295,620 for 4 periods	Quarterly data feeds of incident, inspection and ship particular data from IMO, IHSM and LLI
Validating methods	Jan 2018 to Dec 2018 1,868 inspections 756 VSS incidents	Global quarterly incidents and detention feeds (756 incidents when counting incident by quarter by IMO without duplicates, and 817 incidents when counting duplicates)

Note: All data is for the world fleet with 73,905 vessels, excluding fishing vessels and tugs, and with the following status codes: in service, commission, laid-up, launched, in casualty or repair, converting, US Reserve Fleet. Further, VSS denotes very serious and serious incidents including total loss.

To ensure that results are not biased due to underreporting of less serious incidents and near misses (Hassel et al., 2011), this study concentrates on very serious (including total loss) and serious incidents

(VSS). Ship particular data contains standard information such as ship type, age, size, flag, company (e.g. beneficial owner, class society, safety management company), construction (engine information, ship yard country), previous incidents and inspection outcomes. Tugs and fishing vessels are excluded, and ship types are grouped into six main groups: general cargo, dry bulk, container, tanker, passenger, and other types. To estimate probabilities at ship level, the data stream contains quarterly data feeds for 2018 from the same sources shown in Table 1 for 73,905 vessels, which are out-of-sample data compared to the data for 2010-2014 used to estimate the risk formulas.

The risk formulas used in this analysis are based on logit models following the methodology of the selection of variables from Knapp (2015), Knapp (2006) and Knapp and Franses (2007a, b, 2008). Table 2 lists the resulting risk models that form the basis for step 2 of the methodology. The incident type models serve as proxy to inspection related focus areas, where separate models are used for collisions, powered groundings, main engine failures, and drift groundings.

Table 2: Models chosen for this analysis and their related risk priority inspection areas

Acronym	Model type	Use of model
DET	Detention	For targeting – to combine with VSS
VSS	Incident (very serious and serious)	For targeting – to combine with detention
COLL POWGRD	Collision (VSS) and Powered grounding (VSS)	Collision and powered groundings are both proxies to passage planning, bridge management, crew qualification
ENG DRFTGRD	Main engine failures (VSS) and Drift grounding (VSS)	Engine failures and drift groundings are both proxies to main engine failures, black outs, emergency procedures
FIRE	Fire and explosion (VSS)	proxy to fire related aspects, emergency procedures
HULL	Hull failure (VSS)	proxy to maintenance related issues including tanks and water integrity
LIFE	Loss of life (VSS)	proxy to occupational safety, safety management, life boats
POL	Pollution (VSS)	proxy to pollution prevention and emergency response

The logit model estimates the probability (p) of an event of interest such as detention or incidents (VSS) by means of $p = \exp(xb)/(1+\exp(xb))$, where ‘exp’ denotes the exponential function and ‘ x ’ is the set of vessel-specific variables (e.g. ship type, size, age, flag, classification society, beneficial owner, engine designer, ship yard country). Over 500 variables (including counting dummies for categorical ones) are considered initially. The database to estimate the incident type models has one observation per vessel per year whereas the database to estimate the detention model has multiple observations per vessel per year since vessels can be inspected various times per year. The models are specified by backward elimination by removing insignificant factors (at the 5% significance level). The largest of the resulting models for incidents (VSS) contains 172 variables while the smallest contains 16 (for fire and explosion). All models are estimated by quasi-maximum likelihood (Greene 2008) to allow for possible misspecification of the assumed underlying distribution function for logit models. The employed logit models are described in more detail in Knapp (2015) and Heij and Knapp (2019).

Since the effect of risk factors changes over time as they proxy how industry responds to market conditions and legislative changes, the risk formulas need to be updated every five years. Based on the data from 2010 to 2014, the effects of vessel age and size for VSS incident risk are opposite

to those for detention risk. Since the detention model reflects actual PSC MoU decisions in practice, this indicates that no incident information is part of the targeting routine. This finding relates to one of the main messages of this paper that past incident information is relevant for targeting vessels for inspection to reduce future incidents (VSS). It also demonstrates that the inspection data is biased since it reflects the various targeting policies of coastal states.

Step 2 involves the estimation of ship-specific probabilities using quarterly data feeds for 2018 and using the risk formulas obtained under step 1. This was achieved by means of a customized software program resulting in 2.9 million probabilities that form the basis for the development and testing of the targeting methods shown in Table 3. The probabilities are converted into percentile ranks using as benchmark the global fleet in each of the four quarters.

Table 3: Targeting methods evaluated

Targeting methods	Description
Detention (only)	Vessels are ranked by percentile ranks from detention probabilities only.
Incidents (only)	Vessels are ranked by percentile ranks from TLVSS incident type probability (TLVSS= total loss, very serious and serious).
Combined methods:	
Method A (max)	Vessels are ranked by the highest of the two base percentile ranks
Method B (min)	Vessels are ranked by the lowest of the two base percentile ranks
Method C (weight)	Vessels are ranked by a weight of 50/50 incident to detention
Method D (weight)	Vessels are ranked by a weight of 75/25 incident to detention
Method E (weight)	Vessels are ranked by a weight of 25/75 incident to detention

Percentile ranks provide a useful way for policy makers to understand where a particular vessel stands with respect to all other vessels in the benchmark sample, which could be adjusted to regional preferences (e.g. all vessels that arrived in a particular region over the last 3 years) rather than using the global fleet as benchmark. Table 3 shows the seven targeting methods that were evaluated based on 4.4. million percentile ranks.

In order to classify vessels based on their percentile ranks, five risk categories are chosen as shown in Table 4. The suggested target inspection coverage is flexible and can be set by policy makers considering the country’s (or group of countries in the case of a PSC MoU) risk appetite, regional priorities and resources.

Table 4: Risk categories and target inspection coverage

Risk Category	Percentile rank	Suggested target inspection coverage
RC1: Very high risk	90 to 100	100%
RC2: High risk	80 to 89	100%
RC3: Medium risk	70 to 79	90%
RC4: Low risk	60 to 69	10% random
RC5: Very low risk	0 to 59	5% random

To test feasibility, the suggested target inspection coverage was applied to ship arrival data for 2018 of one country containing 34 thousand arrivals in port with average daily arrival rate of 90 vessels and average daily inspection rate of 8 vessels. Applying the above target inspection coverage using method B, the same average figures were obtained, that is daily inspection rate of 8 vessels, making the suggested coverage feasible. At the global level and based on unique IMO numbers, the quarterly inspection rate is 21.2% while the yearly inspection rate is 42.6%.

3. Evaluation of targeting methods

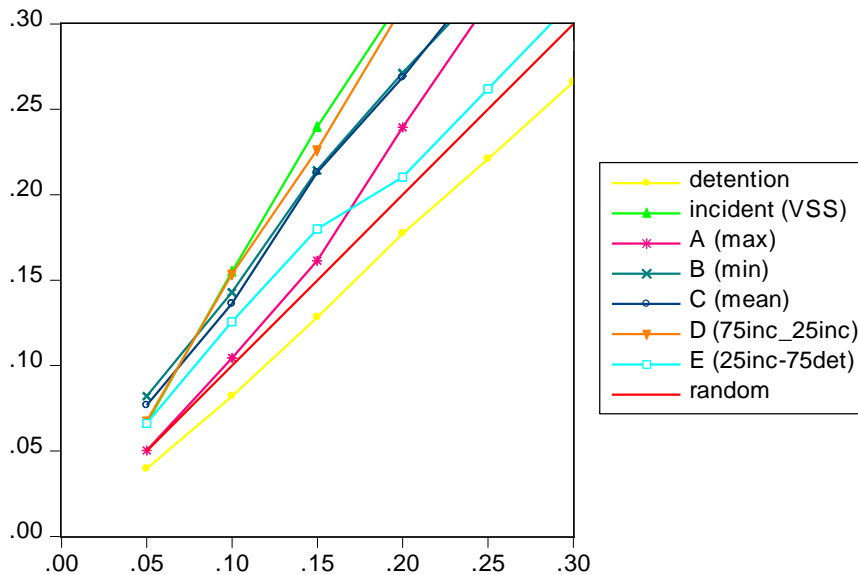
To evaluate the various proposed targeting methods, three evaluation variables were considered: incidents (VSS); detentions; incidents and detentions combined (the vessel was either detained or had an incident within the relevant time period). The evaluation time periods are specified as follows, since estimated probabilities are only valid up to a maximum of three months:

- P1: Probabilities estimated as of late Dec 2017 – empirical data from Jan to March 2018
- P2: Probabilities estimated as of late March 2018 – empirical data from April to June 2018
- P3: Probabilities estimated as of late June 2018 – empirical data from July to Sept 2018
- P4: Probabilities estimated as of late Sept 2018 – empirical data from Oct to Dec 2018

One way to visualize how well the targeting methods perform compared to random selection of vessels is via ROC (receiver operating characteristic) curves that plot the *true positive rate (TPR)* on the *Y-axis* against the *false positive rate (FPR)* on the *X-axis*. Figures 1 to 3 provide ROC curves for the three evaluation variables, zooming into the Top30% of all vessels, which represents the top three risk categories (RC1 to RC3). Any curve above the diagonal line (random selection) constitutes an improvement. Appendix A provides the complete ROC curves.

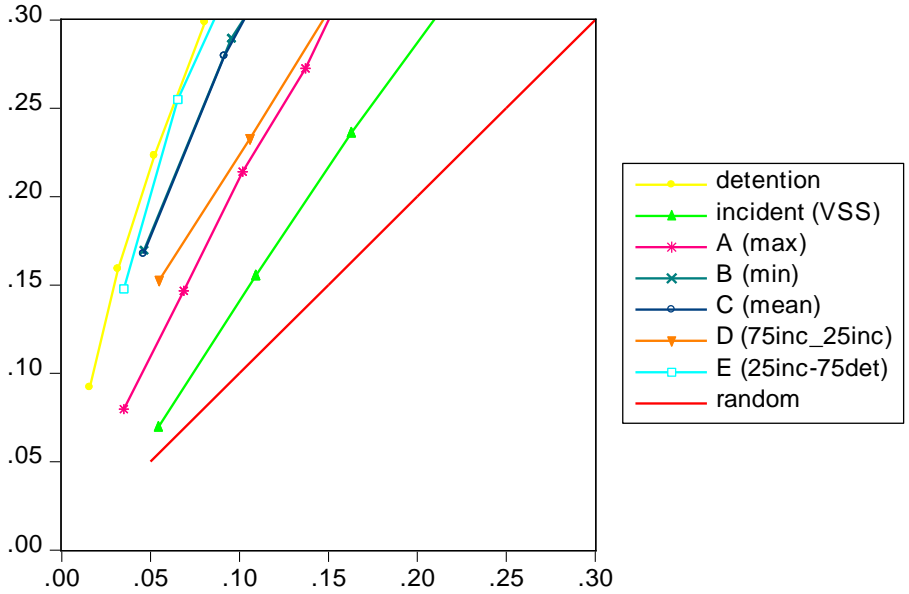
One can observe that all methods perform better than random selection with the exception of the detention method (using detention only) for the evaluation variable incidents (VSS). This is understandable given the small correlation (-0.046) between the two which also confirms that vessels with a high probability of detention do not necessarily have a high probability of incident (VSS) and that the two need to be treated as different risk dimensions.

Figure 1: ROC curve top 30% – evaluation variable incidents (VSS): total incidents 756



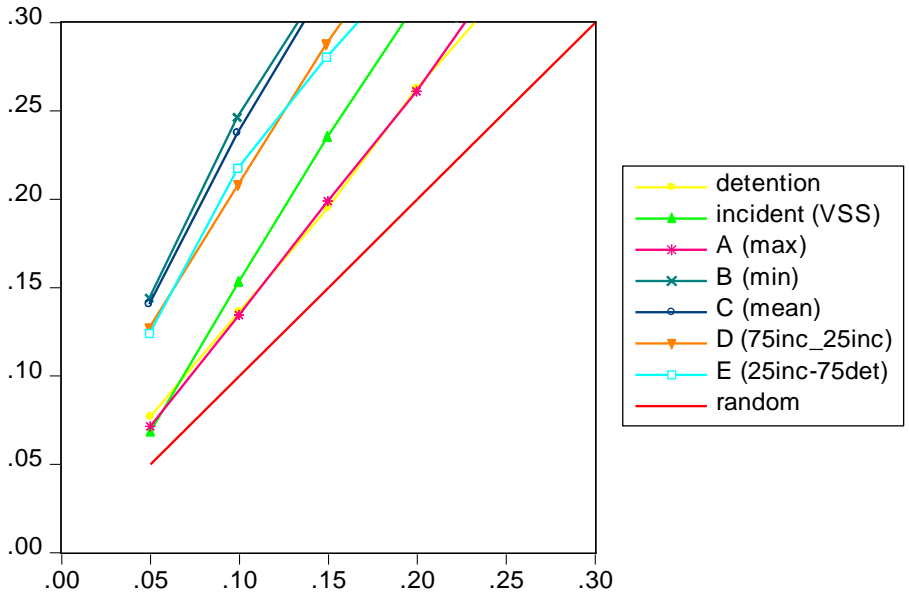
Note: Y-axis: true positive rate, X-axis: false positive rate

Figure 2: ROC curve top 30% – evaluation variable detention: total detentions 1,868



Note: Y-axis: true positive rate, X-axis: false positive rate

Figure 3: ROC curve top 30% – evaluation variable VSS and detention combined: total count 2,589



Note: Y-axis: true positive rate, X-axis: false positive rate

At the global level, 60.2% of all incidents (VSS) were not selected for inspection up to three months prior to the incident. Of the 39.8% vessels that were inspected, 4.4% were detained indicating that vessel inspection priority risk areas could be improved in order to focus inspection efforts and to reduce incidents from happening. Restricting this to very serious incidents (VS) only, 75.4% were not selected for inspection and 3.3% were detained. After excluding cases with heavy or severe weather conditions, only 42.1% (VSS) and 34.3% (VS) of vessels with incidents were selected for inspection up to three months prior to the incident. Method D on the other hand would have selected 44.8% of all vessels with VSS incidents and 39.3% of all vessels with VS incidents in the top three risk categories (RC1 to RC3). Taking different inspection rates into

account and if random selection is set to factor 1 for comparison reasons, the 44.8% classification rate of method D translates to 1.49 compared to random selection (44.8% divided by 30%).

Besides visualization by ROC curves, the improvement over random selection of vessels in terms of reduction of the false negative rate is quantified in Table 5. Note that the false negative rate is the opposite of the true positive rate (in the sense that both rates add up to 100 percent). To test the significance of differences in success rates across methods, the Satterthwaite Welch t-test is performed (Appendix B shows detailed results for some methods).

Table 5: Reduction in false negatives compared to random selection (Top 5% to Top 30% of vessels)

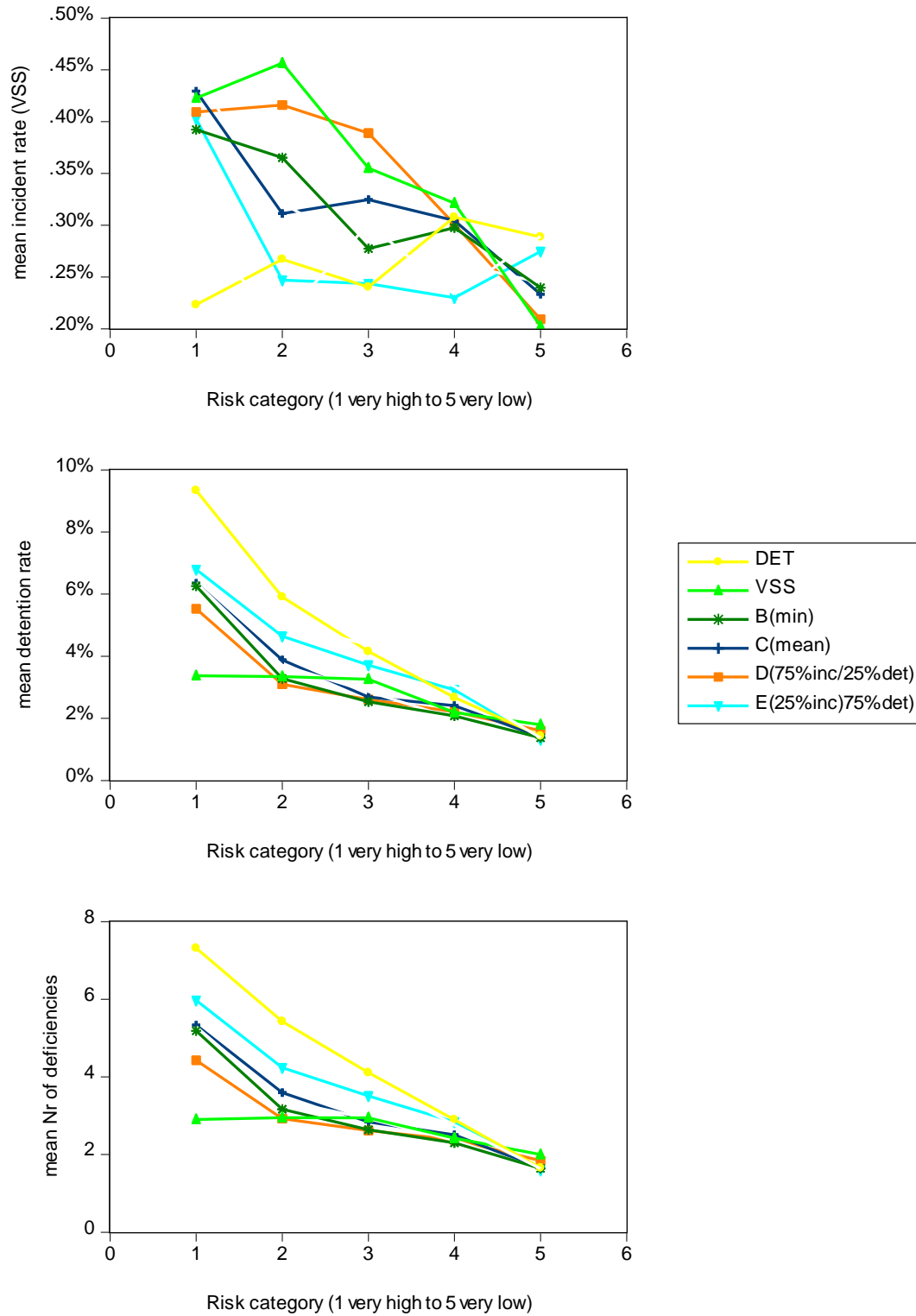
Evaluation variable	% Top	Emp. Count	DET	VSS	A	B	C	D	E
VSS (295,620)	5	756	1.0%	-1.6%	0.0%	-3.2%	-2.7%	-1.8%	-1.6%
	10	756	1.8%	-5.5%	-0.5%	-4.3%	-3.6%	-5.3%	-2.6%
	15	756	2.2%	-8.9%	-1.1%	-6.4%	-6.3%	-7.6%	-3.0%
	20	756	2.3%	-11.5%	-3.9%	-7.1%	-6.9%	-11.0%	-1.0%
	25	756	2.9%	-14.0%	-6.2%	-7.4%	-8.1%	-12.7%	-1.2%
	30	756	3.4%	-14.1%	-6.0%	-7.2%	-8.8%	-14.4%	-1.4%
Detained (62,537)	5	1868	-4.2%	-2.0%	-3.0%	-12.0%	-11.8%	-10.3%	-9.8%
	10	1868	-5.9%	-5.5%	-4.7%	-19.0%	-17.9%	-13.3%	-15.5%
	15	1868	-7.3%	-8.6%	-6.4%	-22.1%	-21.7%	-16.4%	-17.3%
	20	1868	-9.9%	-11.1%	-7.3%	-24.6%	-22.3%	-18.0%	-19.0%
	25	1868	-11.1%	-13.1%	-9.3%	-26.5%	-22.8%	-19.1%	-19.2%
	30	1868	-14.0%	-6.2%	-10.5%	-27.7%	-23.4%	-20.2%	-19.3%
VSS and detained combined (295,620)	5	2,589	-2.7%	-1.8%	-2.2%	-9.4%	-9.1%	-7.7%	-7.4%
	10	2,589	-3.6%	-5.3%	-3.4%	-14.6%	-13.8%	-10.8%	-11.8%
	15	2,589	-4.5%	-8.5%	-4.9%	-17.5%	-17.1%	-13.8%	-13.0%
	20	2,589	-6.3%	-11.0%	-6.1%	-19.6%	-17.8%	-15.9%	-13.6%
	25	2,589	-7.0%	-13.2%	-8.2%	-21.0%	-18.5%	-17.1%	-13.9%
	30	2,589	-8.8%	-15.4%	-9.0%	-21.9%	-19.1%	-18.5%	-14.0%

Note: DET=detention, VSS=incident, A (max), B(min), C(mean), D(75%incident/25%detention), E(25% incident/75% detention)

The results confirm that all methods are significantly better than random selection with the exception of the method DET (using detention only) for the evaluation variable VSS. For the evaluation variable incidents (VSS), method D performs best at the Top30% level which combines the first three risk categories (RC 1 to 3) while method VSS performs best at the Top10% level (RC1) followed closely by method D. The Satterthwaite Welch t-test however confirms no significant difference between method VSS and method D at the Top10 or Top30% level but confirms that method B and using detention only vary significantly compared to method VSS or method D. Method D gives more weight to incidents but also accounts for detention risk to capture vessels that have low percentile ranks for VSS but high for detention. For the evaluation variable detention and combining detention with incidents, method B (min) performs best at the Top10% and 30% level. This is also confirmed by the Satterthwaite Welch t-test where method B varies significantly compared to method DET, VSS and method D at the Top10 and Top 30% level.

Figure 4 provides the mean deficiency rate and detention rate of inspected ships and the mean incident rate of all vessels for each of the suggested risk categories which should be higher for higher risk categories.

Figure 4: Incident rate (%), detention rate (%), and mean number of deficiencies per risk category



Note: mean incident rate = Sum of incidents/total nr of unique vessels by RC, mean detention rate = sum of detentions/sum of inspections by RC, mean nr of deficiencies = sum of deficiencies/total nr of unique inspected vessels

The final part of the analysis compares observed incident types with the eight-vessel inspection priority risk areas. The 817 incidents (including vessels that had more than one incident per quarter, which is therefore more than the 756 incidents mentioned in Table 5) are manually checked to identify the first event of what is normally a chain of events. Since an incident can have multiple events and consequences, this leads to 886 outcomes (84 for VS) linked to the inspection priority risk areas. Vessels with high risk (RC1 to RC3) are identified and Table 6 shows the percentage of these vessels to the total of relevant incidents for each category. Some incident types such as grounding, stranding and loss of life, have few observations and it is not possible to distinguish between powered and drift grounding with the available data, hence the comparison is made with grounding/stranding for this type of category using the same count.

Table 6: % of vessels identified in risk category 1 to 3 by inspection priority risk areas (2018)

Observed incident type/first event	Corresponding Inspection priority risk areas	VSS		VS	
		Total	% RC1-RC3	Total	% RC1-RC3
Pollution	Pollution	52	30.8%	3	0.0%
Loss of life	Loss of life	31	12.9%	31	12.9%
Collision/contact	Collision and Contact	202	31.2%	5	0.0%
Fire and explosion	Fire and Explosion	72	34.7%	12	25.0%
Engine/mechanical failures	Engine related failures	317	59.9%	4	50.0%
Propulsion/Steering gear failure	Drift grounding	32	34.4%	-	n/a
Hull related/stranding	Hull related failures	160	33.1%	19	47.4%
Grounding/standing	Powered grounding	19	63.2%	10	70.0%
	Drift grounding	19	57.9%	10	60.0%

Note: For incident type related to grounding and stranding, inspection priority risk areas for powered groundings and drift grounding are matched since they cannot be easily distinguished from the observed data.

In the future, vessel inspection priority risk areas can be extended by adding for instance models related to the Maritime Labour Convention (MLC) and by producing MLC type deficiency probabilities. Another possible improvement is related to occupational safety type incidents and human error. The empirical data showed 45 such cases (21 for VS) that cannot be easily matched against any risk inspection priority areas at this stage but could be in the future if there is a separate risk model for occupational safety related incidents.

The inspection priority risk areas can also be used to further improve targeting vessels for inspection in addition to using combined methods A-E. This idea was tested using the inspection priority risk areas of the 817 vessels that had incidents with results shown in Table 7.

Table 7: Improvement in targeting by adding priority risk areas (relative true positive hit rate)

methods	random	using methods alone	adding inspection priorities showing high risk ranking (RC1 to RC3) with at least:							
			1	2	3	4	5	6	7	8
detention incident (VSS)	1	0.88	2.71	2.36	2.04	1.73	1.36	1.20	1.04	0.98
method A	1	1.49	2.59	2.19	1.86	1.66	1.54	1.49	1.49	1.49
method B	1	1.20	2.65	2.26	1.90	1.60	1.33	1.23	1.22	1.20
method C	1	1.24	2.63	2.31	2.04	1.82	1.55	1.45	1.35	1.31
method D	1	1.29	2.60	2.27	1.97	1.74	1.49	1.42	1.34	1.32
method E	1	1.49	2.58	2.21	1.89	1.69	1.55	1.50	1.50	1.49
method E	1	1.06	2.68	2.35	2.03	1.76	1.44	1.32	1.20	1.15
Additional vessels selected			343	252	177	117	58	33	16	9

Note: The relative hit rate corrects for different inspection rates and is calculated as follows: %correctly classified/% of vessels inspected which is 30% for RC1 to RC3.

Table 7 shows the improvement compared to random selection by using the various methods alone and by combining them with inspection priorities that have high risk ratings (RC1 to RC3). For instance, for method B, improvement over random selection is 0.24 (1.24-1) compared to using method B alone for targeting. Using at least one of the inspection priorities that have high risk ranking in addition to method B, overall improvement is 1.63 (2.63-1) over random selection or 1.39 (2.63-1.24) compared to using method B alone. Based on Table 7, using at least 4 or more inspection priorities with higher risk rankings in addition to a base targeting method alone (eg. such as method B) seems to provide a good balance between improvement and the number of added vessels that would be selected for inspection as shown in the last row of Table 7. The selection of how many of the risk priorities need to show a higher risk rating (e.g 1, to 8) to be considered for inspection is up to policy makers and available resources. The inspection priority areas are more refined models (all restricted to VSS incidents) than the base VSS incident model and are worth been considered is more than 3 or 4 show high risk rating as they are also correlated (please refer to Appendix C) indicating if a vessel has higher risk areas in one or two areas, it most likely will also show higher risk ranking in other areas.

4. Application example for inspectors

The approach presented in the previous section provides a data-driven or quantitative approach to assist selecting vessels for inspection and focusing inspection efforts with the aim to reduce false negative events. The data-driven part can be combined with other intelligence and expert knowledge of inspectors to finalize inspection selection and execution. The procedure can be split up into the following three main steps, where the first two steps can be fully automated and the final step allows addition of qualitative knowledge and other relevant intelligence.

- **Step 1:** Use risk formulas that are updated every 5 years to estimate ship-specific probabilities based on up-to-date data feeds that are received daily or weekly.
- **Step 2:** Calculate percentile ranks relative to the relevant benchmark sample (e.g. global fleet or vessels that visited the relevant region during the last 3 or 5 years) and classify vessels into risk categories. Consider inspection priority risk areas to focus inspection activities and to possibly further improve selection of vessels for inspection.
- **Step 3:** Combine the outcomes of steps 1 and 2 with expected arrival data in a particular port or wider area of interest and plan the inspection visits based on priorities and capacities, taking the data-driven outcome as guidance. To finalize the inspection planning, use other available intelligence and expert knowledge, for instance, knowledge about specific companies or vessels known to inspectors or the region, market economic conditions, or new legislative requirements.

Risk dimensions can be shown graphically as visual assistance to inspectors. Figure 5 provides an example of 11 vessels that all had incidents, of which only one (container vessel 5) has been selected for inspection (with a resulting detention). Such graphs can be generated automatically showing all vessels in port liable for inspection for a particular day or time period. The graph shows instantly where each vessel stands with respect to the others. In this particular example, the targeting method using a Top-30% rule for detention only would have missed vessels 3, 4, 5, 6, 7 and 10. If only incidents (VSS) are used for targeting, the Top-30% rule would have missed

vessels 2, 5, 6 and 9. When using combined methods with the addition of inspection priority areas, all vessels could be considered for inspection.

This type of visualization could be complemented by a Table (refer to Table 8 for an example) that lists the percentile ranks of selected methods and the percentile ranks for the vessel inspection priority risk areas. Note that in this example, all methods are shown but in reality one or two would be chosen such as method B, D, VSS and detention to cover all priorities and focuses.

Figure 5: Example of visualization of risk dimensions of vessels in port

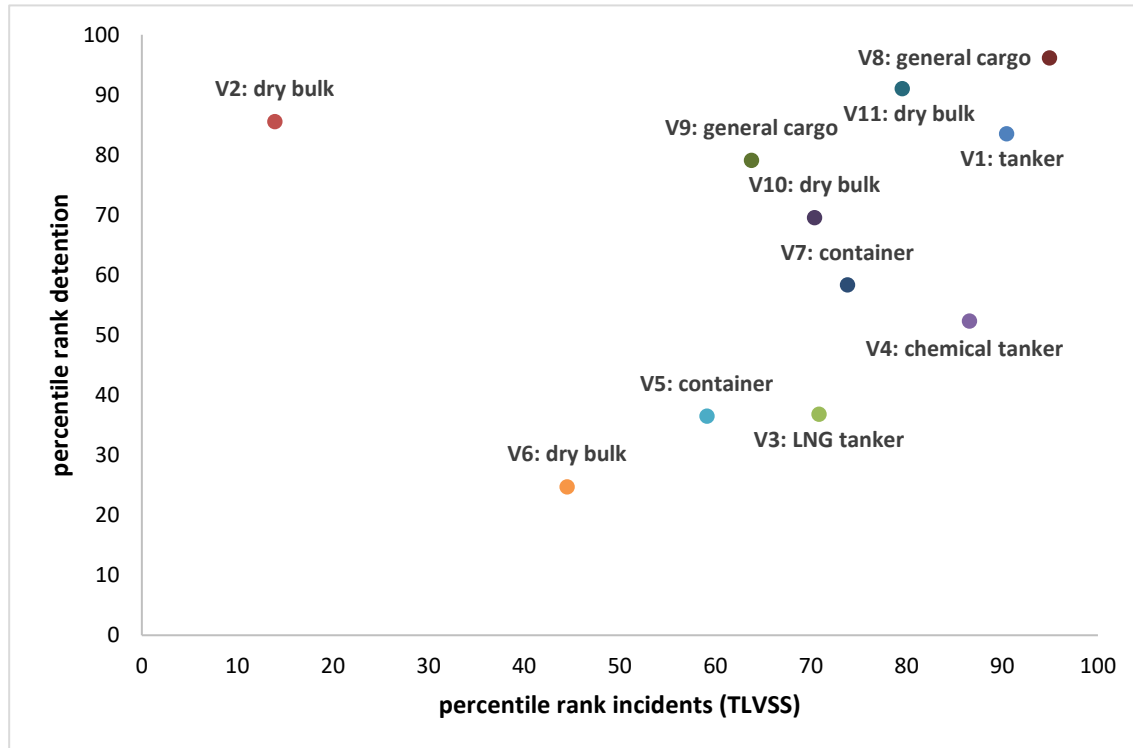


Table 8: Summary of percentile ranks of vessels in port

Vessel ship type	percentiles for targeting methods							percentiles for vessel inspection priorities risk areas							
	DET	VSS	A(max)	B(min)	C(mean)	D(75/25)	E(25/75)	COLL	DRFTGRD	POWGRD	ENGINE	FIRE	HULL	POL	LIFE
1 oil tanker	83.51	90.45	83.47	93.87	93.05	93.70	90.61	84.58	73.53	77.77	87.81	96.40	72.62	90.16	95.06
2 dry bulk	85.58	13.93	76.21	24.32	50.93	27.68	71.71	13.28	51.20	42.49	64.34	29.82	25.78	3.06	12.84
3 LNG tanker	36.84	70.85	55.63	55.42	57.00	65.44	45.47	82.35	77.62	73.85	80.87	53.42	28.78	95.19	91.15
4 chemical tanker	52.34	86.56	77.66	71.01	76.96	83.85	63.49	72.25	76.37	75.77	85.20	96.91	78.66	21.65	78.17
5 container	36.51	59.11	40.84	55.09	48.20	54.61	41.81	79.22	55.55	42.01	42.91	74.94	50.38	57.49	54.65
6 dry bulk	24.75	44.50	25.28	40.46	27.94	37.50	25.05	61.21	38.59	71.39	45.17	69.66	32.83	38.96	12.78
7 container	58.35	73.82	59.53	76.72	72.79	74.32	64.97	69.99	40.58	67.44	49.69	14.58	80.09	21.18	75.91
8 general cargo	96.15	94.94	92.89	99.02	98.63	98.33	98.53	83.53	89.63	93.62	93.31	8.43	86.70	87.61	94.71
9 general cargo	79.12	63.79	66.71	81.28	79.14	72.11	80.40	18.36	20.46	44.36	24.03	29.14	10.03	69.05	15.67
10 dry bulk	69.54	70.38	54.77	85.47	77.60	74.91	74.16	59.66	97.41	46.48	59.02	45.05	95.32	35.90	39.41
11 dry bulk	91.06	79.55	84.27	91.72	91.98	88.10	93.13	73.46	78.15	79.22	58.47	68.22	82.25	66.59	66.49

Note: DET=detention, VSS=incident, COLL= collisions, DRFTGRD=drift groundings, POWGRD=powered grounding, ENGINE=engine related failures, FIRE=fire and explosion, HULL = hull related failures, POL = pollution, LIFE = loss of life

With respect to the inspection priority areas, the following is of interest given that all vessels experienced incidents but only one was selected for inspection and detained.

- Vessel 1 experienced a steering gear failure and stranded in February 2018. The percentile ranks for engine related failure is high (87.81) and drift grounding is medium (73.53). It arrived in port in January 2018 but was not inspected.
- Vessel 3 experienced main engine failures and was towed to port in late June 2018. The percentile rank for engine failure is high (80.87). It arrived three times in port (April, May and June) but was not selected for inspection.
- Vessel 4 experienced main engine failure and piston crown damage in late April 201. Its percentile rank for engine failure is high (85.20). It arrived in port four times prior to within 90 days of the incident and was not inspected.
- Vessel 8 experienced loss of life by mid Sept 2018. Percentile ranks of all methods are very high (above 90, and 94.71 for loss of life). It arrived in port twice in August, the last time just two weeks before the incident, but was not selected for inspection.
- Vessel 10 ran aground late in November 2018, was re-floated and departed. It has very high percentile ranks for drift grounding (97.41) and hull related failures (95.32). It arrived in port twice before the incident and was not inspected.
- Vessel 11 experienced engine problems late in October 2018. While its percentile rank of engine failure is low, the one for drift grounding is medium (78.15).

5. Discussion and conclusions

This study considers and evaluates the status quo assumption of maritime inspections such as PSC inspections and industry inspections to primarily use past inspection outcomes – in particular past detention and deficiencies – to target vessels for inspections. One of the main goals of inspections is to improve the safety quality of vessels and to reduce the probability of future incidents. In terms of targeting efficiency, the reduction of *false negative events* is the main focus. The main contributions of this study are summarized as follows:

- While it has been established that inspections decrease future incident risk, this study confirms room for improvement since of all vessels that experienced incidents (VSS) in the year 2018, only 40% had been inspected and 4% were detained. An alternative approach to target vessels for inspections treats detention and incidents as separate risk dimensions. The very low correlation (-0.04) between the probabilities of detention and incident (VSS) at ship level confirms that these two dimensions measure different risk aspects.
- Five combined targeting methods are developed and tested against random selection of vessels using empirical data for 2018. The results show a potential gain (reduction of false negative events) of 14 to 27% compared to random selection which can be further improved if adding vessel inspection priority risk areas to the targeting routine. The study demonstrates that combined methods have the potential to reduce false negative events as the chance to catch risky vessels is improved. In the future, different weights for incident type risk and detention could be tested, especially if a longer time period for testing becomes available.
- The use of percentile ranks makes it possible to combine two risk dimensions and allows customized benchmarking of vessels, for instance, by considering vessels for a particular region of interest rather than the global fleet used here. Since the percentile ranks are based

in the benchmark sample and the probability estimates, they are dynamic in nature and correct for improvements of the fleet since vessels are always compared against each other for given overall safety quality at any given time.

- The study presents risk categories and associated inspection target coverage, which are tested against arrival data of one particular country. However, these categories and target coverages are flexible and can be set by policy makers depending on their regional preferences and inspection capacities, along with the selection of the benchmark sample.
- Incident data still has many quality issues and it remains difficult to use these data to validate targeting methods or to determine incident types related to inspection priority risk areas. This was partly overcome by restricting the data to VSS incidents and by manually classifying incident data from at least three different sources. In addition, it is impossible to measure the number of vessels that did not have incidents due to inspections or that would not have had an incident if they had been inspected. For the year 2018, 60% of all vessels with a VSS incident were not inspected in the 3 months prior to the incident, and of the 40% that were inspected only 4% were detained. This indicates that inspection efforts and priorities can be improved by focusing them in a better way.
- The eight considered vessel inspection priority risk areas provide the means to help inspectors in focusing their efforts, as these areas provide insight into the individual vessel risk profile. In the future, these areas could be extended further, for example, by adding Maritime Labour Convention deficiency type probabilities.
- The data-driven part to assist with targeting and to decide how to focus inspection priorities can be automated. It is further simple to visualize risk dimensions and risk priorities at the level of vessels that are expected in a particular port or area. This can act as one component in the more complex process to select vessels for inspections and can be combined with qualitative aspects including other intelligence available and based on the inspector's expert knowledge.

It should be acknowledged that the evaluation of new methods as presented here has two important restrictions. First, the inspection and detention data are the product of current inspection strategies across the globe and are therefore biased towards current targeting regimes. Second, the incident data has several limitations. A true test of alternative inspection targeting approaches can only be obtained by actually implementing such approaches for some time to guide the inspection decisions and by recording the outcomes.

Acknowledgement

We thank the data providers for this manuscript, including IMO, LLI, and IHS Markit.

References

- Bijwaard, G., and S. Knapp (2009), "Analysis of Ship Life Cycles – The Impact of Economic Cycles and Ship Inspections." *Marine Policy* 33 (2): 350-369.
- Greene, H.W. (2008), *Econometric Analysis*, 6-th Edition. New Jersey: Pearson Prentice Hall.
- Hassel, M., B.E. Asbjørnslett, and L.P. Hole (2011), "Underreporting of Maritime Accidents to Vessel Accident Databases." *Accident Analysis and Prevention* 43 (6): 2053-2063.

- Heij, C., and S. Knapp (2019), "Shipping Inspections, Detentions and Accidents: An Empirical Analysis of Risk Dimensions." *Maritime Policy and Management*, in print
- Heij, C., G. Bijwaard, and S. Knapp (2011), "Ship Inspection Strategies: Effects on Maritime Safety and Environmental Protection." *Transportation Research Part D* 16 (1): 42-48.
- IMO (2000), "Reports on Marine Casualties and Incidents - Revised Harmonized Reporting Procedures." IMO Document MSC/Circ. 953, MEPC/Circ. 372. Dated December 14, 2000. London: IMO.
- Ji, X., J. Brinkhuis, and S. Knapp (2015), "A Method to Measure Enforcement Effort in Shipping with Incomplete Information." *Marine Policy* 60: 162-170.
- Knapp, S. (2006), *The Econometrics of Maritime Safety – Recommendations to Improve Safety at Sea*. Rotterdam: ERIM Ph.D. Series Research in Management.
- Knapp, S., and P.H. Franses (2007a), "A Global View on Port State Control - Econometric Analysis of the Differences across Port State Control Regimes." *Maritime Policy and Management* 34 (5): 453-482.
- Knapp, S., and P.H. Franses (2007b), "Econometric Analysis on the Effect of Port State Control Inspections on the Probability of Casualty." *Marine Policy* 31 (4): 550-563.
- Knapp, S., and P.H. Franses (2008), "Econometric Analysis to Differentiate Effects of Various Ship Safety Inspections." *Marine Policy* 32 (4): 653-662.
- Knapp, S. (2015), "Methodology and Implementation Aspects for Risk Formulas at Ship Level." Technical Report for AMSA, Contract 15AMSA116, November 24, 2015.
- Li, K.X., J. Yin, and L. Fan (2014), "Ship Safety Index." *Transportation Research Part A* 66: 75-87.
- Perepelkin, M., S. Knapp, G. Perepelkin, and M. De Pooter (2010), "An Improved Methodology to measure Flag Performance for the Shipping Industry." *Marine Policy* 34 (3): 395-405.

Appendix A: ROC curves without zoom

Figure A.1: ROC curve – evaluation variable incidents (VSS): total incidents 756

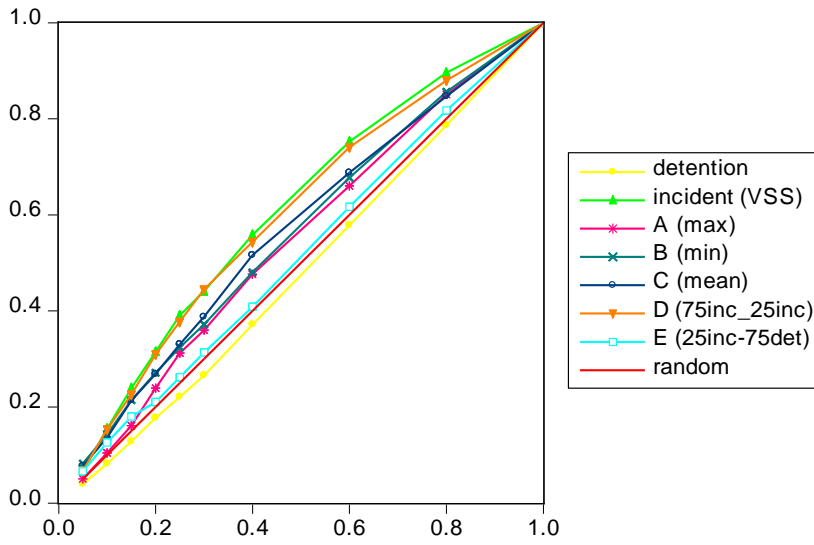


Figure A.2: ROC curve – evaluation variable detention: total detentions 1,868

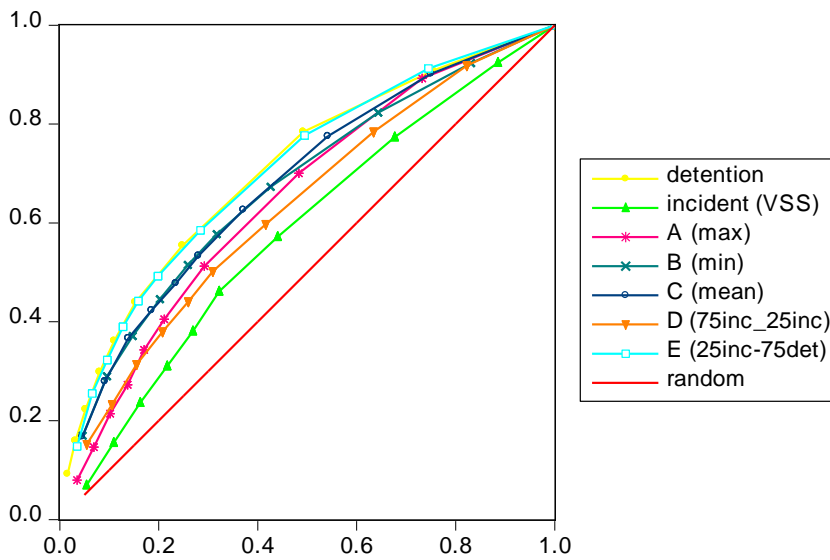
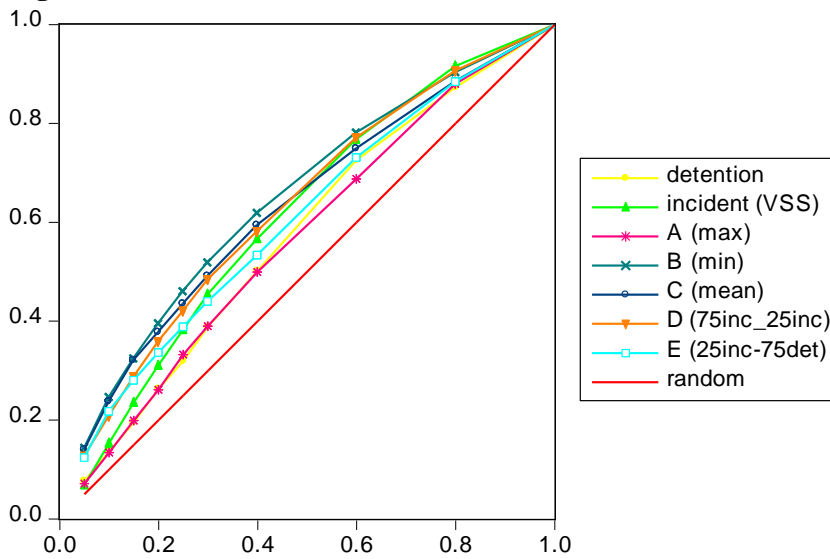


Figure A.3: ROC curve – evaluation variable VSS and detention: total count 2,589



Appendix B: P-values of the Satterthwaite Welch t-test

Method	Random	DET	VSS	B(min)
Very serious and serious incidents (756)				
Top10				
DET	0.2115	-	-	-
VSS	0.0016	0.0000	-	-
B(min)	0.0118	0.0002	0.5158	-
D(75inc/25det)	0.0020	0.0000	0.9433	0.5628
Top 30				
DET	0.0878	-	-	-
VSS	0.0000	0.0000	-	-
B(min)	0.0066	0.0000	0.0064	-
D(75inc/25det)	0.0000	0.0000	0.8767	0.0040
Detention (1,868)				
Top10				
DET	0.0000	-	-	-
VSS	0.0000	0.7531	-	-
B(min)	0.0000	0.0000	0.0000	-
D(75inc/25det)	0.0000	0.0000	0.0000	0.0001
Top 30				
DET	0.0000	-	-	-
VSS	0.0000	0.1777	-	-
B(min)	0.0000	0.0000	0.0000	-
D(75inc/25det)	0.0000	0.0001	0.0128	0.0000
Detention and incidents combined (2,589)				
Top10				
DET	0.0001	-	-	-
VSS	0.0000	0.0894	-	-
B(min)	0.0000	0.0000	0.0000	-
D(75inc/25det)	0.0000	0.0000	0.0000	0.001
Top 30				
DET	0.0000	-	-	-
VSS	0.0000	0.0000	-	-
B(min)	0.0000	0.0000	0.0000	-
D(75inc/25det)	0.0000	0.0000	0.0278	0.0144

Appendix C: Correlation of percentile ranks of vessel inspection priority risk areas

Area	COLL	DRFTGRD	POWGRD	FIRE	HULL	LIFE	MENGINE	POL
COLL	1.00							
DRFTGRD	0.72	1.00						
POWGRD	0.60	0.54	1.00					
FIRE	0.46	0.24	0.49	1.00				
HULL	0.36	0.27	0.56	0.77	1.00			
LIFE	0.41	0.23	0.29	0.62	0.53	1.00		
MENGINE	0.67	0.85	0.43	0.17	0.13	0.23	1.00	
POL	0.38	0.21	0.32	0.49	0.54	0.64	0.20	1.00

Note: COLL= collisions, DRFTGRD=drift groundings, POWGRD=powered grounding, ENGINE=engine related failures, FIRE=fire and explosion, HULL = hull related failures, POL = pollution, LIFE = loss of life