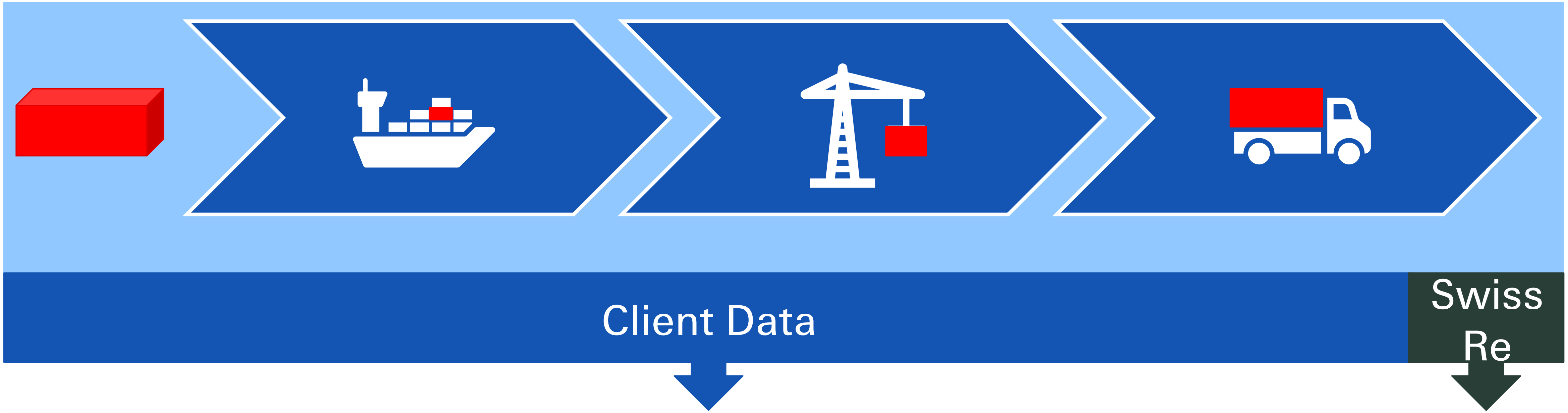
Swiss Re

Modeling Container Shipping Delay with Random Effects

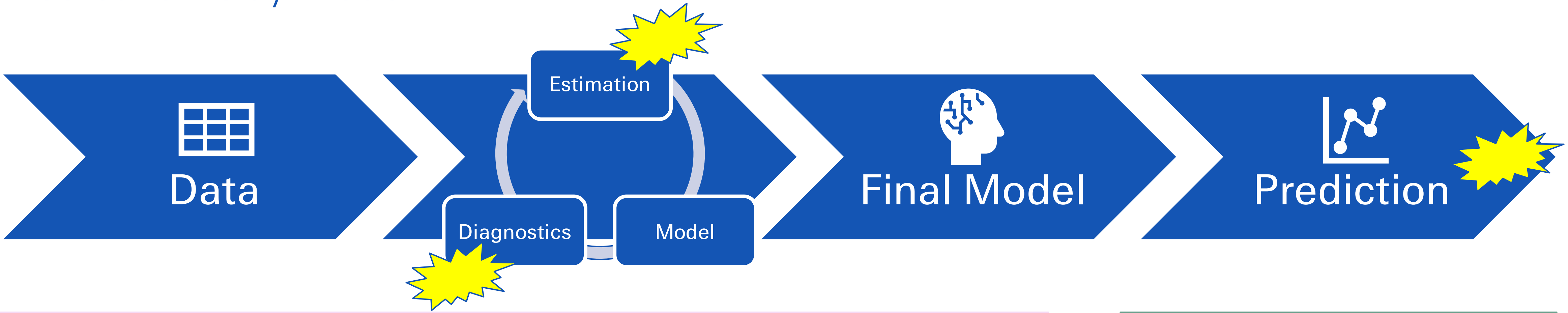
A Comparison of Methods

Container Shipping Data



Delayed	Port	Country	Carrier	Shipping Comp.	Container Size	Container Type	Year	Quarter	Cyclone
...									
TRUE	Genoa	Italy	ABC	XYZ	40 ft	RG (Reefer)	2019	Q1	FALSE
...									

Predictive Delay Model



Potential Issues:

Problem	Caused By
Model Matrix Singularity	Over-parametrization due to...
Roughly 200 Coefficients to Check	... High Cardinality of Factor
Interpretation of Reference Level	No Sensible Reference Level
Prediction for Unobserved Levels	Training Data is only a Sample

Proposed Solution:
Instead of independent coefficients per level, assume effects are randomly distributed, so-called random effects, e.g.,

$$\eta_{port} \sim N\left(0, \sigma_{port}^2\right).$$

Model Comparison

Further information in the Appendix, at a coffee break, or contact me via:

@JakobDambon
Jakob_Dambon@swissre.com
/in/jakob-dambon

Slides: github.com/jakobdambon/presentations

Model	Description and Details	Median Run Time[a] [min]	non-NA Predictions [%]	ROC (AUC) [%]
Logistic Regression	Standard Model	0.83	98.80[b]	63.7
Logistic Regression with **aggregated levels**	Leave 4 most frequent levels as is, aggregate rest as "others"	**0.01**	**100.00**[c]	66.0
Logistic Regression using **random effects**	Use Random Effects for five covariates	2.85	**100.00**	65.1
XGBoost[d]	Prior CV for hyper parameter tuning	109.22	98.80[b]	**72.0**

[a]Sum of training and prediction time; [b]Due to previously unobserved levels in the validation data; [c]Unobserved levels were defined as "others"; [d]**This method is only used for quantitative reference purposes. XGBoost uses high-order interactions and is much harder to interpret than any of the models above.**

Swiss Re

Thank you!
Any Questions?

Contact us

Jakob Dambon

Data Scientist
Jakob_Dambon@swissre.com

Follow us

Legal notice

Swiss Re

Appendix

Abstract

Over the last years, the global supply chain experienced several disruptions, which lead to delays of container shipments. From an insurance's and logistics' perspective, it is vital to recognize key factors of delay as they are associated with demurrage fees and further disruptions in the supply chain. Here, the interpretability of such delay models is essential to gain insights and to prevent future delays.

In an applied use case, we analyze the shipment data of an international operating retailer. We compare a wide range of generalized linear fixed and mixed effect models as well as more difficult to interpret boosting-based approaches with respect to two properties. First, we are interested in the key drivers of delay and the over all model interpretability. Second, we compare the predictive accuracy of all models in cross validation.

We showcase the advantages of generalized linear mixed effect models (GLMM) over similar fixed effect models in terms of interpretability and predictive accuracy. While the GLMM does not reach the same predictive performance as boosting approaches, it allows us to draw more precise conclusions about the main driving factors of container delays.

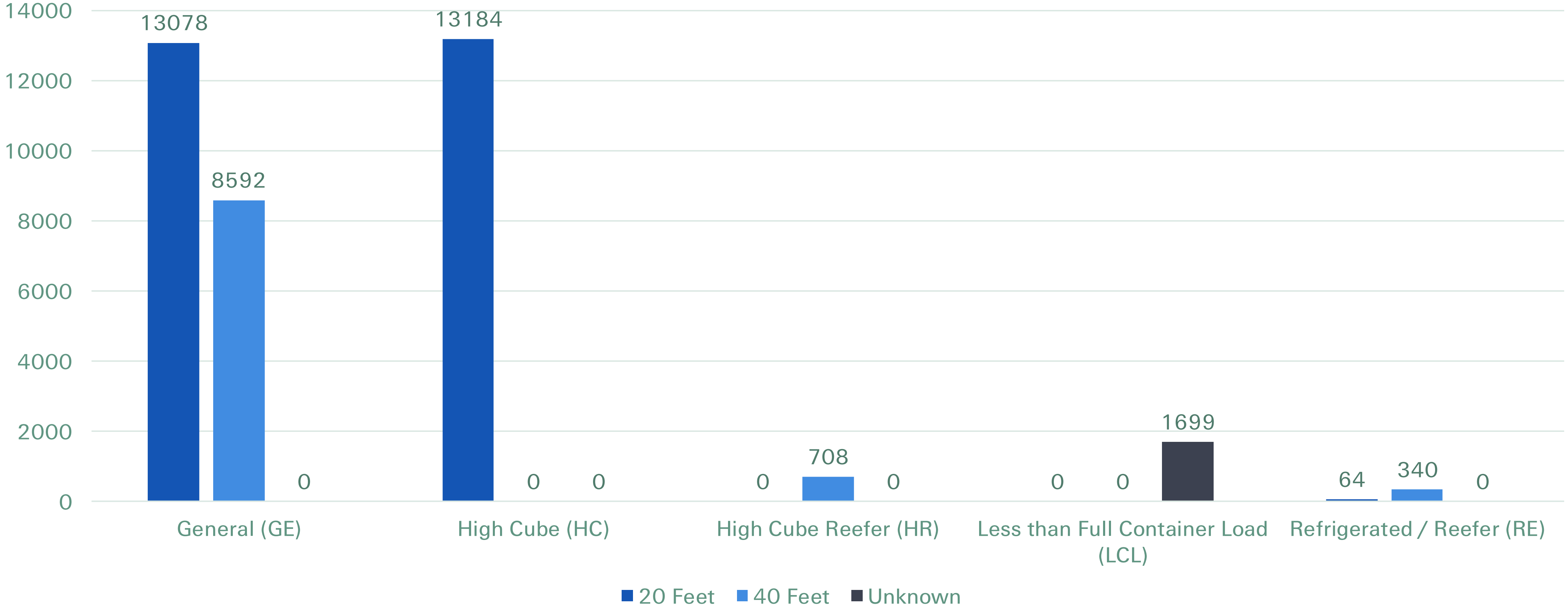
Data

We give an overview of the container shipping data. It roughly contains 38’000 observations. Each observation is one container of a certain size and type, arriving at a port by some given shipping company, then being unloaded, and waiting at the port until a carrier takes it away from the port. If the container takes more than 5 days between unloading and leaving the port, we label it as “delayed” as demurrage fees may apply. Otherwise, we label it “on-time”.

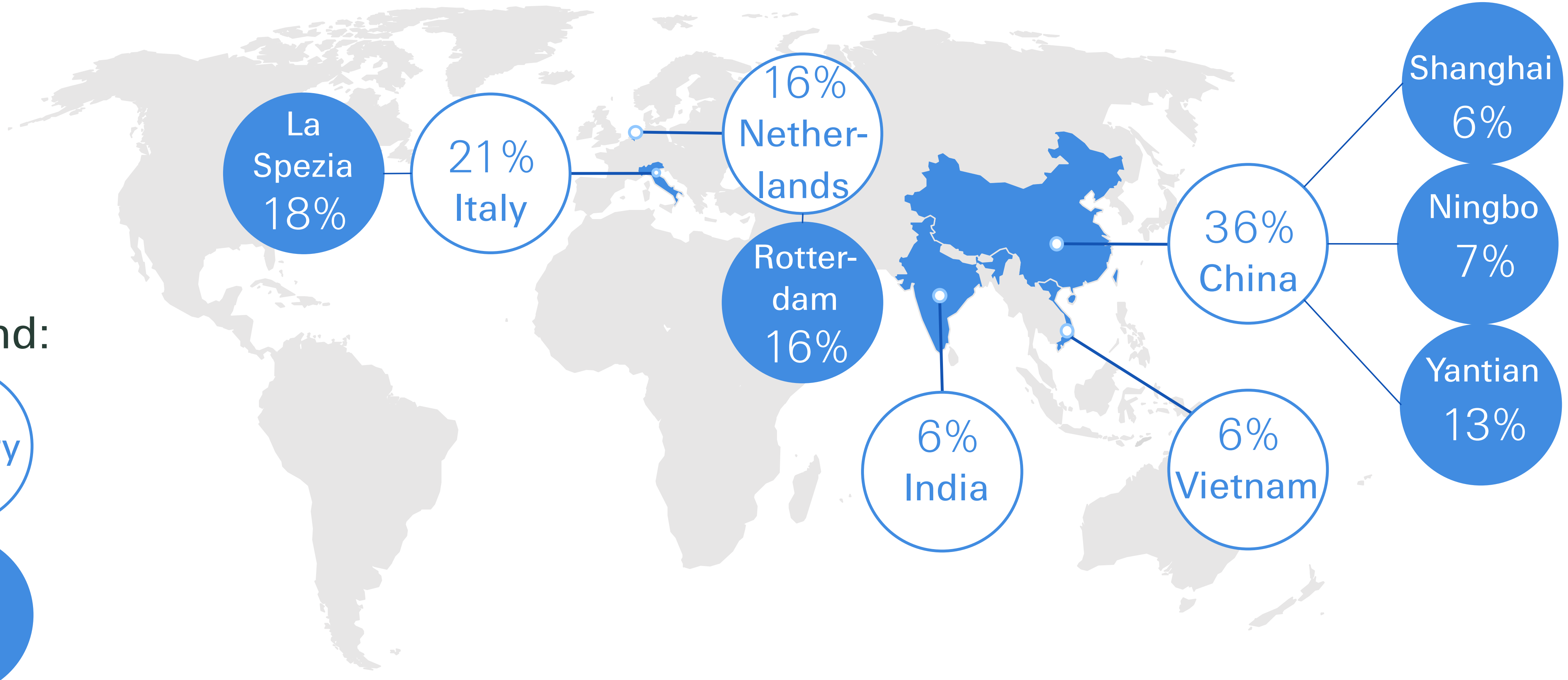
Additionally, we know the year and quarter where the unloading has taken place and we engineered a binary variable that indicates if a cyclone was hitting the port in a +/- 5 days time frame while the container was in port. For the latter, we were using Swiss Re’s CatNet® service.

A rough outline of the data has been given in the lightning talk (slide 2). In the following, we go through the data in more detail.

Container Types and Sizes (cts): Counts



Container per Country and Port:
Percentages of all shipments above 5%, other countries or ports not shown

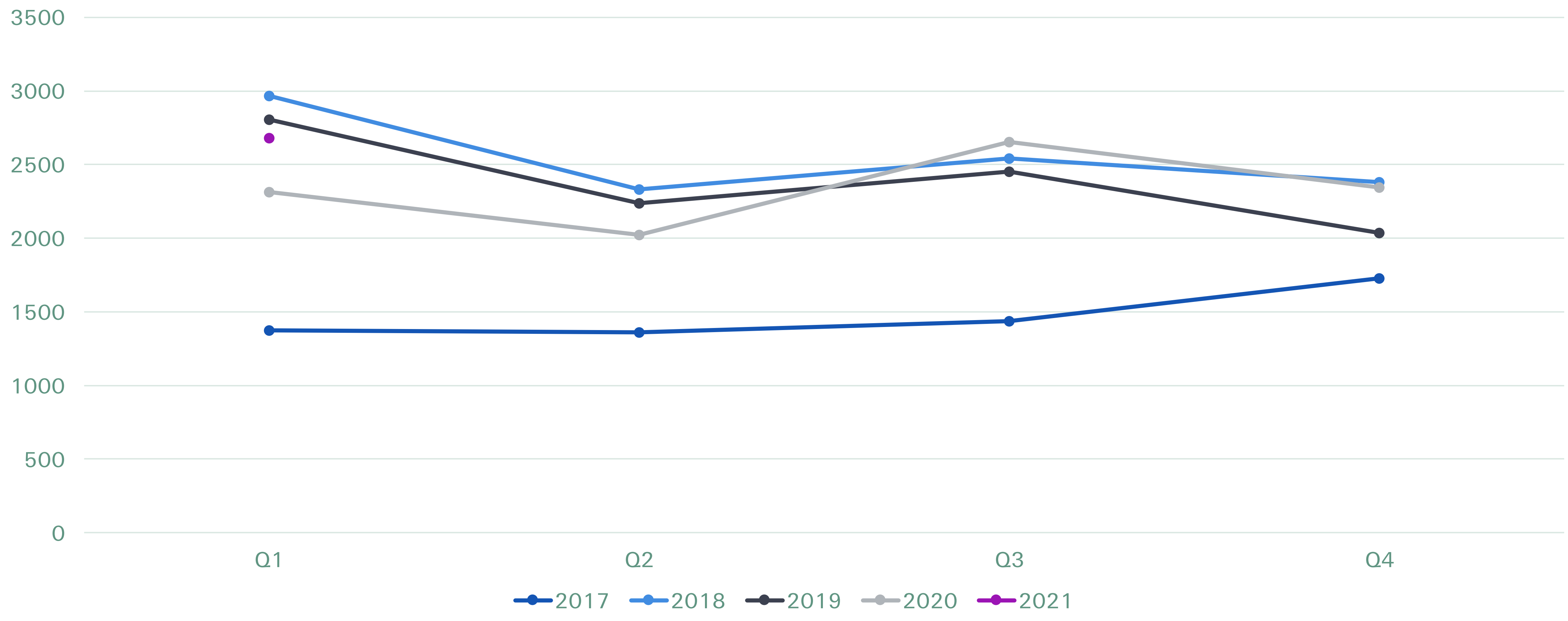


Legend:

xx% Country

Port yy%

Number of Containers per Quarter (yq):



High Cardinality of Levels

Ports	Carriers	Shipping Company
p_k	c_k	sc_k
$\#\{p_k\} = 114$	$\#\{c_k\} = 34$	$\#\{sc_k\} = 20$

Model

Our client was interested in an interpretable approach to investigate what the main drivers for container delay are. Naturally, we started with a logistic regression. From there, we tried different other approaches until we arrived at so-called mixed effect models. In this section, we present the individual models with the reasoning behind them.

Basic Logistic Regression

The logistic regression is the working horse in statistics when it comes to classification. With the presented covariates, the model is defined as

$$E(y|X) = g^{-1}(\beta_0 + \beta_1 \cdot cts_i + \beta_2 \cdot p_i + \beta_3 \cdot ctry_i + \beta_4 \cdot c_i + \beta_5 \cdot sc_i + \beta_6 \cdot yq_i + \beta_7 \cdot cyclone_i)$$

where:

- g is the logit link function
- cts_i is the container type and size and $cyclone_i$ is a dummy variable indicating if a cyclone was present at the port while the container was unloaded. Both covariates have a limited number of levels.
- p_i is the port, $ctry_i$ is the country, c_i is the carrier, sc_i is the shipping company, and yq_i is the year and the quarter. All these covariates have a high cardinality of levels, and we can expect unobserved levels.

Note that for prediction's sake, the country covariate does not add any benefit as the ports already cover all variability. However, for interpretation’s sake, we added it to the model.

Several models are over-parametrized and produce singularities. This results in uninterpretable coefficients, especially for the ports and countries. Hence, one should apply regularization (Ridge- or Lasso-Regression, Elastic Net) or some data aggregation. We apply latter on the next slide.

Grouping Data

Due to the many levels within some of the covariates and the consecutive over-parametrization of the model, the results of the logistic regression are not interpretable. One way to deal with this issue is to group specific levels within a factor. We only keep the 4 most frequent levels for covariates p (port), $ctry$ (country), c (carrier), sc (shipping company). All other levels are grouped into one new level "other" in each covariate, respectively.

Otherwise, the model stays the same. The additional benefit in doing so is that every new and unobserved level in that covariate can be grouped into "other", too. The main issue with this approach is the handling of the data, i.e., managing the data preparation before training, and the issue of a reference level. Below, we show the first couple of effects for a model trained on the data from 2017 and 2018.

```
                             Estimate Std. Error z value Pr(>|z|)    
(Intercept)                   0.81980    0.39986   2.050 0.040344 *  
container_type_sizeGE_40      0.01431    0.05052   0.283 0.777020    
container_type_sizeHC_40      0.04014    0.04720   0.850 0.395115    
container_type_sizeHR_40     -1.18763    0.17452  -6.805 1.01e-11 ***
container_type_sizeLCL        0.46360    0.10947   4.235 2.29e-05 ***
container_type_sizeRE_20     -1.16567    0.66825  -1.744 0.081099 .  
container_type_sizeRE_40     -1.93993    0.81726  -2.374 0.017611 *  
has_cyclone_windowTRUE       -0.02423    0.12331  -0.196 0.844240    
countryIndia                 -2.09451    0.20761 -10.089  < 2e-16 ***
countryItaly                 -0.80380    0.37412  -2.149 0.031673 *  
countryNetherlands            0.29454    0.38002   0.775 0.438311    
countryother                 -0.03177    0.06569  -0.484 0.628639    
```

Extract of Model Coefficients

Reference level is:
- General 20ft container
- no cyclone
- Country: China

Random Effects

Instead of modeling a coefficient using a constant for each level of a covariate (a so-called fixed effect), we assume that all effects of a covariate emerge from a zero-mean normal distribution. These effects are called random effects. For the 114 ports, instead of estimating 113 differences between the ports and some reference port, we model the effects of 114 port as drawn from a normal distribution, i.e.: $\beta_{port} \in R^{113} \Rightarrow \beta_{port} \sim N(0, \sigma^2_{port})$.

This eliminates the previous issues, namely:

- (*) There is no reference level required. The effect of each level is the difference to the population mean, and not to another level. On the next slide, there are examples given for the effects of some countries and quarters.
- If we observe a new level (say a new port), we can simply give a prediction for that port. Here we assume the population mean and the unaccounted variance for that covariate can be included by our estimate of σ_p^2.

We apply this approach to each covariate with high cardinality of levels. The covariate for container size and type as well as the cyclone remain the same. This yields a model with fixed and random effects which is called mixed effect model.

Interpretation of the Random Effect's variance: Apart of the estimated effects per level (*), the estimated variance $\hat{\sigma}_j{}^2$ of a covariate j is a proxy of its variable importance. A large $\hat{\sigma}_j{}^2$ indicates that a large variability in the effects has been discovered and modelled.

Mixed Effect Model Output

The mixed effect model consists of fixed and random effects. As training data, we use the same data set as previously shown with the grouped data model. We give an overview of the results below.

Fixed Effects:

Below, we show the estimated, remaining fixed effects in the model.

```
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -3.15068    0.61383  -5.133 2.85e-07 ***
container_type_sizeGE_40 -0.00728    0.05270  -0.138   0.8901
container_type_sizeHC_40  0.01107    0.04897   0.226   0.8212
container_type_sizeHR_40 -1.60602    0.26630  -6.031 1.63e-09 ***
container_type_sizeLCL    0.98488    0.57384   1.716   0.0861 .
container_type_sizeRE_20 -1.74634    0.78868  -2.214   0.0268 *
container_type_sizeRE_40 -2.02777    0.91325  -2.220   0.0264 *
has_cyclone_windowTRUE   -0.17388    0.12560  -1.384   0.1663
```

Reefer container (i.e., refrigerated containers) of all sizes and type (20ft, 40ft, and high-cube reefer) have *statistically significant negative log odds*, i.e., a delay is less likely compared to general 20ft containers.

Random Effects:

We consider two outputs of the model. First, the individual variances per covariate is given.

```
Random effects:
 Groups       Name        Variance Std.Dev.
 port         (Intercept) 0.9579   0.9787
 country      (Intercept) 1.4896   1.2205
 carrier      (Intercept) 1.9625   1.4009
 ship. comp.  (Intercept) 0.6562   0.8101
 year_quarter (Intercept) 0.1732   0.4162
```

Carriers and countries account for most of the variability in the data.

Second, the estimated effects for each level is given, e.g.:

```
$country
                       (Intercept)
Bangladesh              0.69269224
Begium                  1.55570196
Brazil                 -0.09061301
Cambogia               -1.28995503
Canada                  1.01121301
Chile                  -0.39719860
China                  -0.06655477
```

```
$year_q
        (Intercept)
2017Q1 -0.19769420
2017Q2 -0.07714598
2017Q3 -0.48014901
2017Q4  0.12692260
2018Q1  0.84331006
2018Q2 -0.33568054
2018Q3  0.27349707
2018Q4 -0.00122373
```

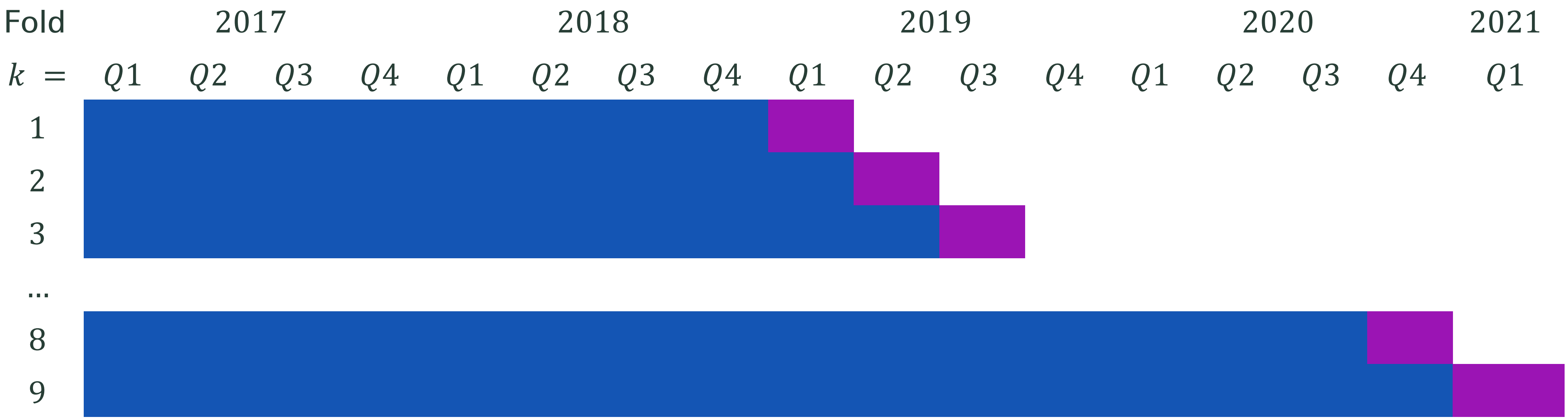
Predictive Performance: A Model Comparison

Besides the advantages of interpreting the model results, we want to compare the predictive performance of fixed and mixed effect models. In total, we compare the 3 linear models described before, i.e., a basic logistic regression, a logistic regression trained on grouped data, and a mixed effect model. As a last model for comparison, we use a XGBoost tree. Latter is a little bit out of scope for a comparison with linear models, as – on the one hand – XGBoost models allow for a high-order interactions. On the other hand, they are very difficult to interpret and take more computation time as usually there is a prior hyper-parameter tuning involved.

Predictive Performance Set-Up

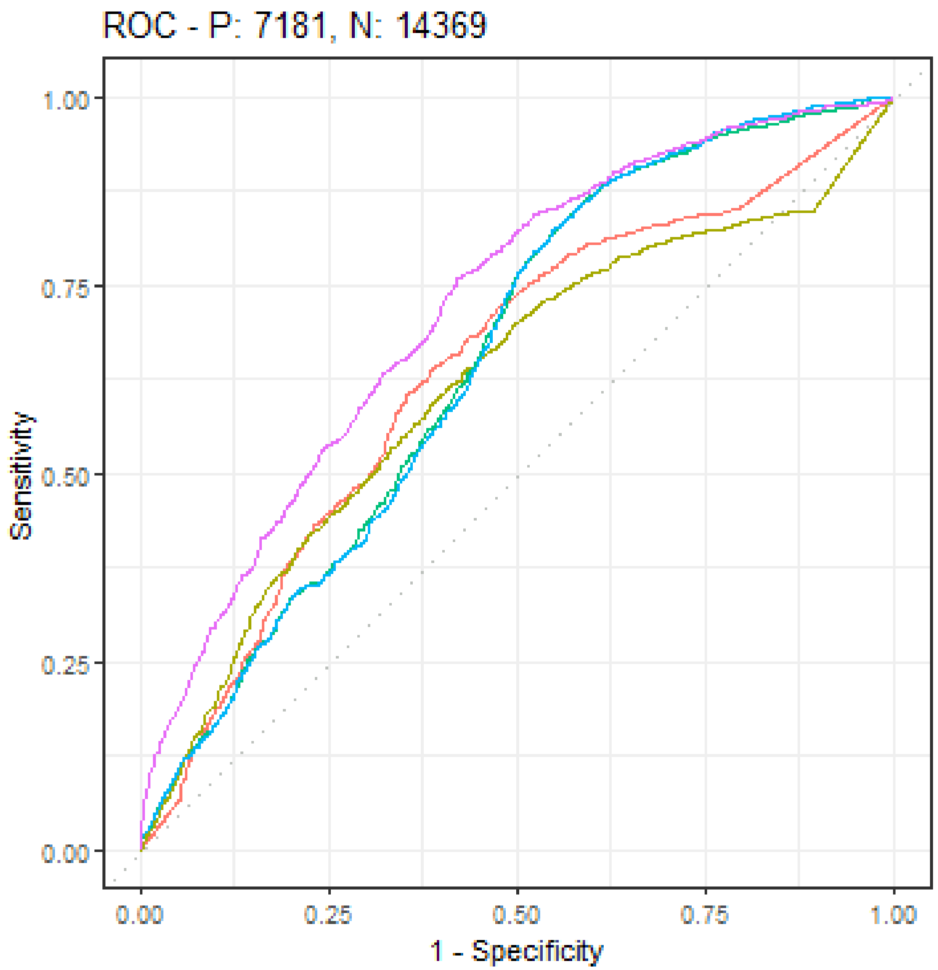
We use the temporal structure of the data to validate and compare the predictions. We use an extending window where:

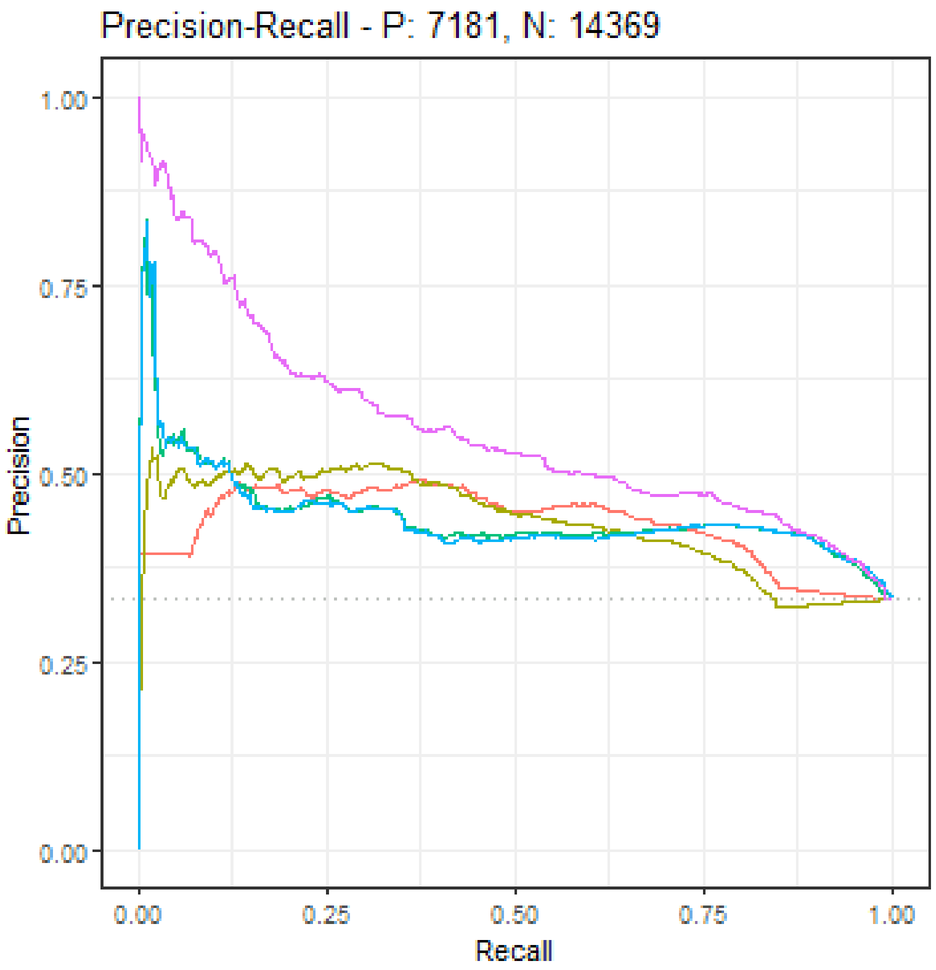
- For each fold $k = 1, \dots, 9$, we have a year and quarter $yq_k = 2019Q1, \dots, 2021Q1$. We use all the data from $2017Q1$ up to but not including yq_k as training data.
- We train the 4 mentioned model on that data set.
- With the models, predict the probabilities of delays for the out-of-sample (validation) data of yq_k.



ROC- and PR-Curve

Model	Description	Used R package	Color
Logistic Regression (Log. Reg.)	Only using fixed effects (slide 16)	`stats::glm`	
Log. Reg. with grouped data	As described on slide 17	`stats::glm`	
Log. Reg. with Random Effects	As described on slides 18 and 19	`lme4`	
		`gpboost`	
Xtreme Gradient Boosting	Only as benchmark; high degree of interaction, non-interpretable	`caret (xgboost)`	



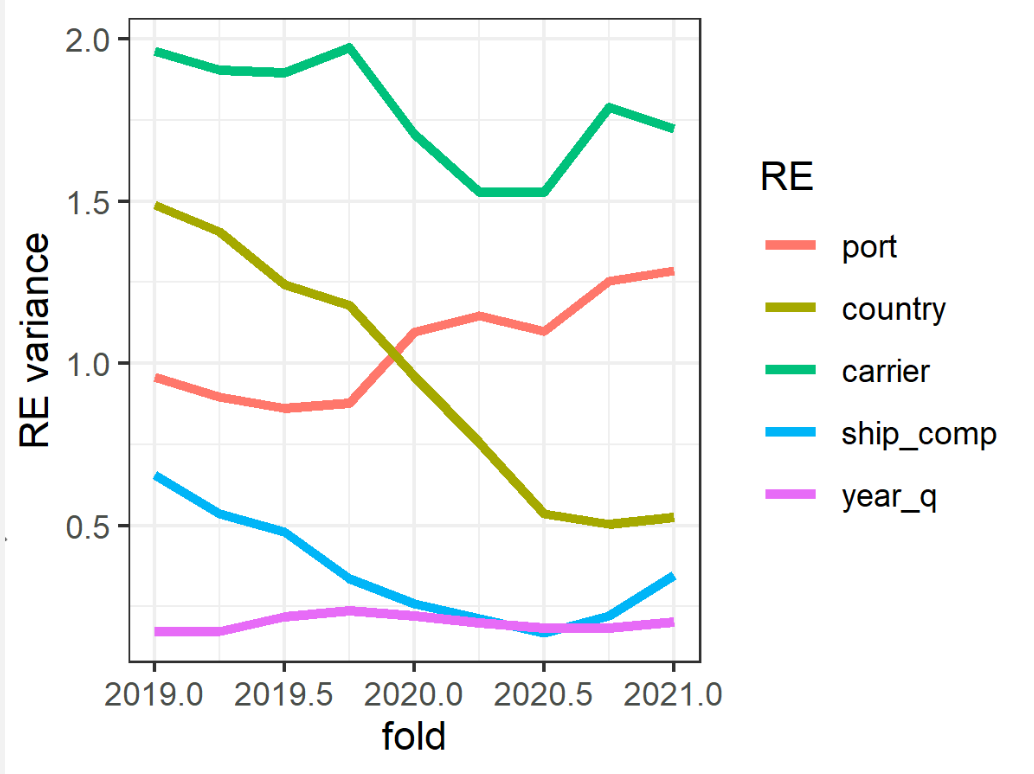


Further Analysis of Random Effects

Another way to gain insights from the mixed effect models is to examine the estimated, covariate-specific variances of the random effects. Due to our set-up for the predictive performance comparison, we can compare the variances per quarter and see their temporal development; see Figure on the right.

First, we see that carriers are the leading cause for a potential delay. Second, we observe that from 2019Q1 to 2020Q2, the variance of the countries random effect dropped, while the one of ports increased. Therefore, delays are now caused at individual ports rather than a whole country.

Estimated Random Effect (RE) Variances $\hat{\sigma}_j^2$ over Folds k



Summary

Random Effects offer many advantages when modeling factors of high cardinality. For instance, they share the same, easy way to interpret the coefficients and take it a step further as they do not model the difference to some reference level, but they model the effect per level directly. Further, they aggregate the variability per factor and provide a corresponding variance. They have a similar predictive performance as other linear models, while only requiring a moderate computation time.
Finally, we benefit in an easy way of implementation. Predictions of previously unknown levels can be computed using random effects without requiring grouping of levels or other data transformation. Therefore, we would like to encourage the further usage of mixed effect models, and in particular random effects, when new levels of covariates for predication can be expected and the interpretability or model estimation is challenged by the high cardinality of some factors.

References

- Chen, T., Guestin, C. (2016). XGBoost: A Scalable Tree Boosting System, KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 781-794 [DOI]( )
- Jiang, J. (2007). Linear and Generalized Linear Mixed Models and Their Applications, Springer.
- Sigrist, F. (2020) R Package gpboost, https://github.com/fabsig/GPBoost
- Viellechner, A., Spinler, S. (2020). Novel Data Analytics Meets Conventional Container Shipping: Predicting Delays by Comparing Various Machine Learning Algorithms, Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020, 1278-1287.

About the Speaker

2016:	MSc Mathematics,	ETH Zurich
2017-2021:	Research Associate,	Institute of Financial Services Zug
2021:	PhD Biostatistics,	University of Zurich
2021-now:	Data Scientist (80%),	Swiss Re
	& Lecturer (20%),	Lucerne University of Applied Sciences and Arts

@JakobDambon
Jakob_Dambon@swissre.com
/in/jakob-dambon