Model accuracy: Model Risk

Course roadmap

1. Project valuation: valuation metrics, planning and rules
2. Model quality and decision making. Benefit curve
3. Estimating model risk discounts
    - What is Model Risk?
    - Consideration of model risk in income assessment
    - Methods of assessing model quality degradation over time
    - Calculation of model risk
4. A/B testing and financial result verification
5. Unobservable model errors, metalearning

What is Model Risk?

Model Risk

Model risk

is the risk of loss resulting from using models in decision-making processes

Direct financial losses

Lost profit

Model Risk

1. **Model decay**
   Model performance decreases during model operation compared to performance at development stage

$$Model\ Risk = Benefit(Model) - Benefit\left(Model_{decay}\right)$$

Model Risk

2. **Ignoring the better challenger model**
   A feasible opportunity to build a more accurate model by enriching it with new data, factors or applying different algorithm

$$Model\ Risk = Benefit(Model_{better}) - Benefit(Model)$$

Model Risk

3. **Incorrect / sub-optimal model implementation**
   Decision rules within business process can incorrectly use model predictions, e.g. cutoffs for credit scoring can be tuned in a wrong way

$$Model\ Risk = Benefit\left(Model_{optimal}\right) - \ Benefit(Model)$$

Wrap-up

1. Model risk is the risk of loss resulting from using models in decision-making processes

Wrap-up

1. Model risk is the risk of loss resulting from using models in decision-making processes
2. Main sources of model risk:

Wrap-up

1. Model risk is the risk of loss resulting from using models in decision-making processes

2. Main sources of model risk:

    - Model decay

Wrap-up

1. Model risk is the risk of loss resulting from using models in decision-making processes
2. Main sources of model risk:
   - Model decay
   - Ignoring the better challenger model

Wrap-up

1. Model risk is the risk of loss resulting from using models in decision-making processes

2. Main sources of model risk:

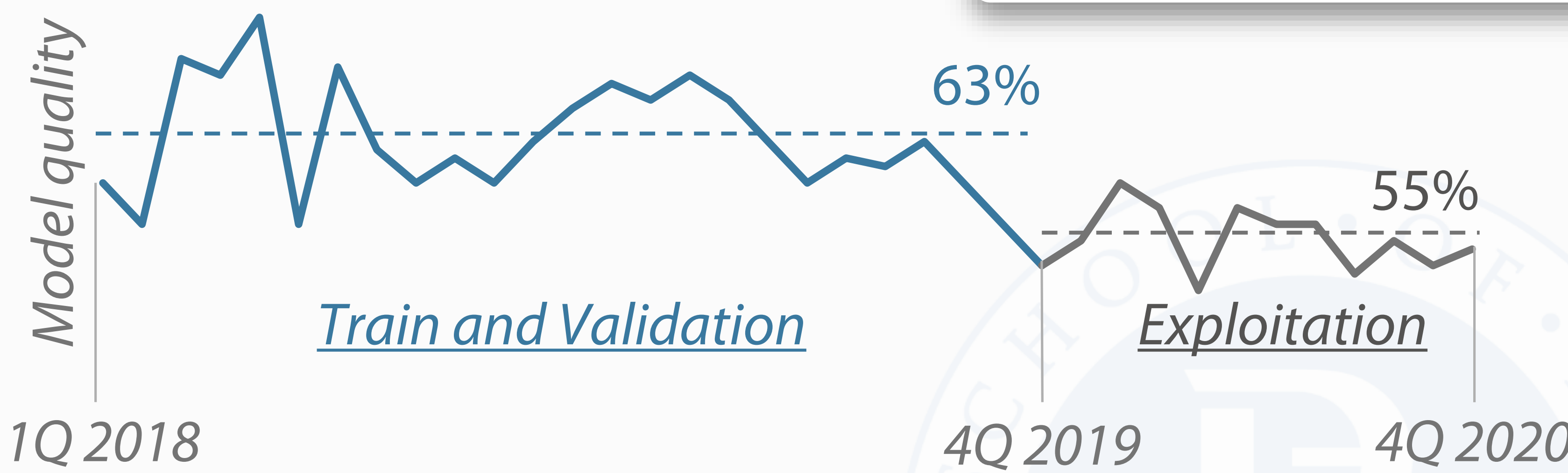
   - Model decay
   - Ignoring the better challenger model
   - Incorrect / sub-optimal model implementation

Consideration of model risk in income assessment

Model Risk

Model risk realization

Model quality dropped by 8% Gini



The risk of negative Gini changes during the exploitation of the model determines the value of the model risk

Consideration of model risk in income assessment

Model implementation

Model quality: 70%
Benefit: 1 M $

Models can be used in process for a long time and make a profit

Account the risk of falling model quality for every year

Consideration of model risk in income assessment

Model implementation

Model quality: 70%
Benefit: 1 M $

1 year of exploitation

Model quality: – 5%
Benefit: –10%

Account the risk of falling model quality for every year

Consideration of model risk in income assessment

Model implementation

Model quality: 70%
Benefit: 1 M $

Account the risk of falling model quality for every year

1 year of exploitation

Model quality: – 5%
Benefit: –10%

2 years of exploitation

. . .

Model quality: – 10%
Benefit : – 20%

Model quality degradation over time

Question? How much does the profit from the model decrease over time?

Answer depends on: How much the quality of the model deteriorates

Building a statistical forecast of the limiting drop in model quality

Wrap-up

1. Models could be used in processes for a long time and make profit

Wrap-up

1. Models could be used in processes for a long time and make profit

2. Model quality degrades over time

Wrap-up

1. Models could be used in processes for a long time and make profit
2. Model quality degrades over time
3. The amount of profit that the model will bring will decrease

Methods of assessing model quality degradation over time

Assessing model quality degradation over time

The best estimate of model stability over time is estimate of stability on out-of-time sample

Basic method of assessing model quality degradation over time

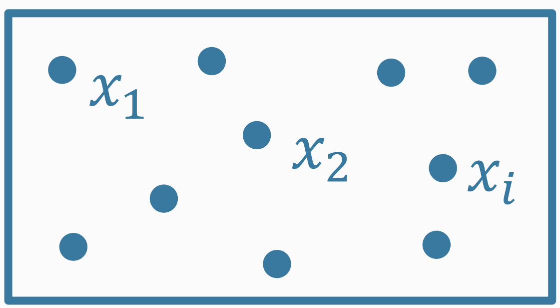
It is not always possible to leave a large sample for out-of-time

Bootstrapping

Inference about model quality on population
can be modelled by resampling the sample data

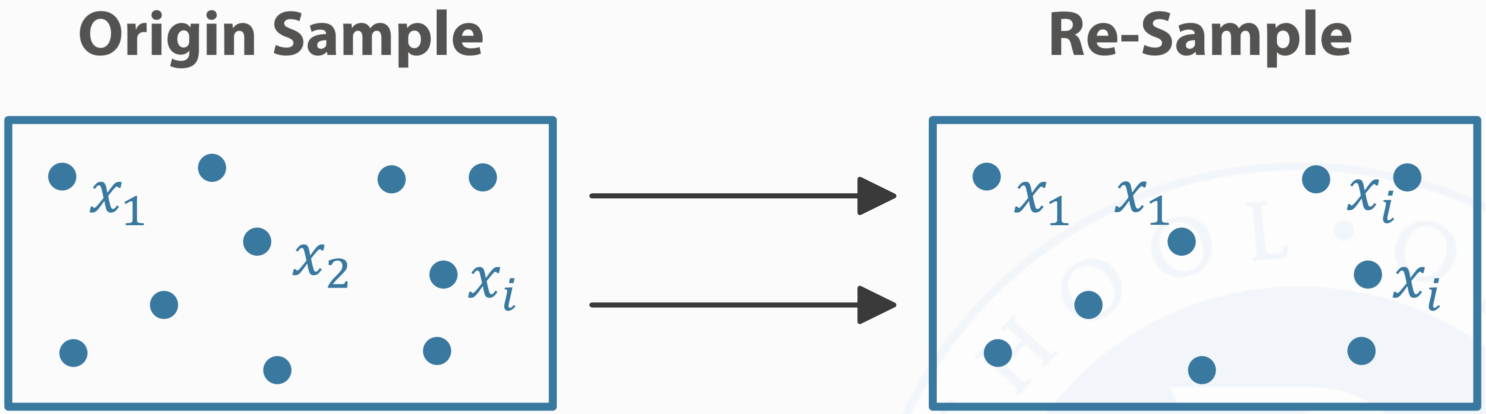
Basic method of assessing model quality degradation over time

Origin Sample



$$Size = N$$
$$Model\ Quality = Q$$

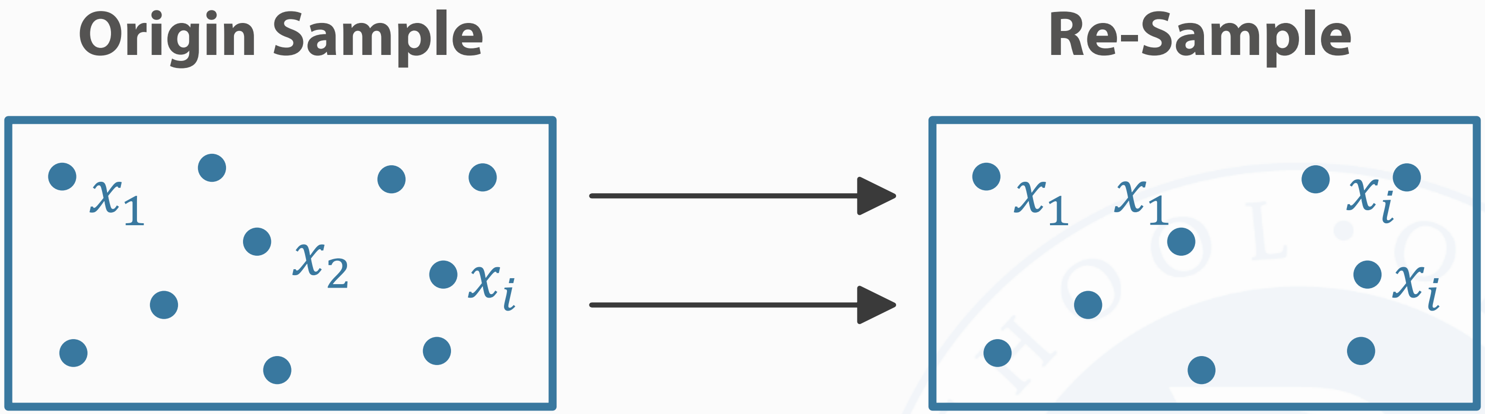
Basic method of assessing model quality degradation over time



Size = N

With the possibility of one or more values to be repeated

Basic method of assessing model quality degradation over time



$Size = N$

$Model\ Quality = Q^1$

Basic method of assessing model quality degradation over time

Repeat Re-Sample

Now we have distribution of model quality

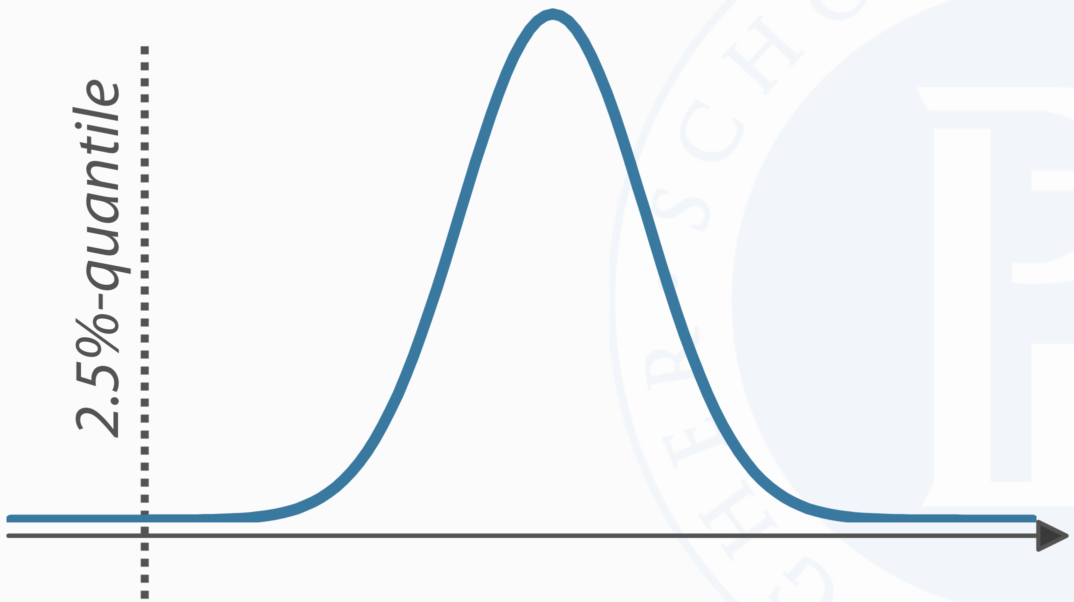
$$Model\ Quality = [Q^1, Q^2, \dots, Q^n]$$

Basic method of assessing model quality degradation over time

$$Model\ Quality = [Q^1, Q^2, \dots, Q^n]$$

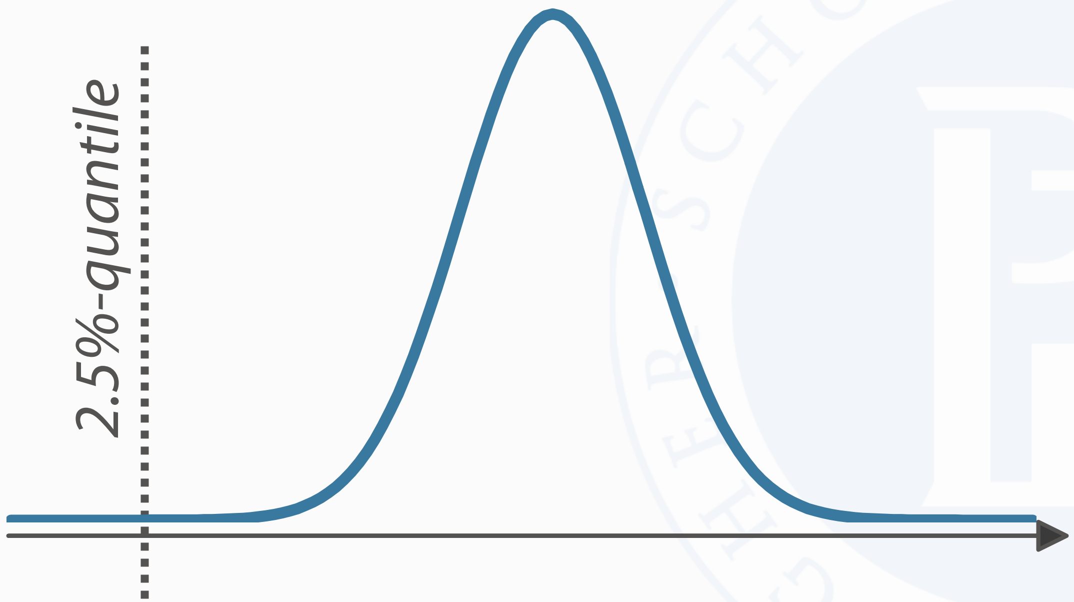
This represents an empirical distribution of model quality

From this distribution, one can derive confidence interval



Basic method of assessing model quality degradation over time

Information about the worst outcome can be obtained using a bootstrap to assess model quality on validation sample



Advanced method of assessing model quality degradation over time

Mann–Whitney U test

is a nonparametric test of the null hypothesis that:

Probability of X being greater than Y = Probability of Y being greater than X

Advanced method of assessing model quality degradation over time

Mann–Whitney U test

is usually used to determine if two independent samples were selected from populations having the same mean rank

Also it can be used to accurately assess the distribution of models quality metrics

Samples are the model scores for the non-target group and the target group

Advanced method of assessing model quality degradation over time

Mann–Whitney U test: calculation

1. Assign numeric ranks to all the observations, beginning with 1 for the smallest value. Where there are groups of tied values, assign a rank equal to the midpoint of unadjusted rankings.

Advanced method of assessing model quality degradation over time

Mann–Whitney U test: calculation

2. Add up the ranks for the observations which came from sample 1. The sum of ranks in sample 2 is now determinate.

   The sum of all the ranks equals $N\,(N+1)/2$ where N is the total number of observations

Advanced method of assessing model quality degradation over time

Mann–Whitney U test: calculation

3. Determine U

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

$$U = \min\{U_1;\ U_2\}$$

- n_i is the sample size for sample i
- R_i is the sum of the ranks in sample i

Advanced method of assessing model quality degradation over time

U is approximately normally distributed

$$\mu = \frac{n_1 n_2}{2} \qquad \sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Advanced method of assessing model quality degradation over time

The AUC and Mann–Whitney U test

$$AUC = \frac{U_1}{n_1 n_2}$$

In order to understand the distribution of the AUC, we can make use of our knowledge about the Mann-Whitney U

Wrap-up

1. Model quality inevitably decreases over time

Wrap-up

1. Model quality inevitably decreases over time
2. Large dataset allows to simulate the training
   of the model and its behavior and quality in time

Wrap-up

1. Model quality inevitably decreases over time
2. Large dataset allows to simulate the training of the model and its behavior and quality in time
3. If there is not enough data, then you can estimate the worst outcome with a bootstrap

Calculation of model risk

Calculation of model risk

1. **Model decay**

   Degraded model is unknown

More research required

2. **Ignoring the better challenger model**

   Alternative model available

3. **Incorrect / sub-optimal model implementation**

   Correct / optimal model available

Can be calculated

Calculation of model risk: model decay

1. Model decay

Degraded model is unknown **More research required**

Concept of
"Sensitivity of financial effect to model quality"

$$\Delta Benefit = \frac{\partial Benefit}{\partial Quality} \cdot \Delta Quality$$

Calculation of model risk: model decay

Concept of
"Sensitivity of financial effect to model quality"

$$\Delta Benefit = \frac{\partial Benefit}{\partial Quality} \cdot \Delta Quality$$

Change of benefit

Required value. Model risk

Calculation of model risk: model decay

Concept of

"**Sensitivity of financial effect to model quality**"

$$\Delta Benefit = \frac{\partial Benefit}{\partial Quality} \cdot \Delta Quality$$

Change of model quality

Known value

Research of model quality degradation over time

Calculation of model risk: model decay

Concept of
"Sensitivity of financial effect to model quality"

$$\Delta Benefit = \frac{\partial Benefit}{\partial Quality} \cdot \Delta Quality$$

Sensitivity of financial effect to model quality

More research required

Sensitivity of financial effect to model quality

Consider a binary classification model $X \longrightarrow Prob$

$X_1 \dots X_k$		$\rightarrow$ $Prob$	Y
34		**0.94**	**1**
20		0.25	0
21		0.16	1
50	…	**0.86**	**1**
41		0.51	0
25		**0.33**	**1**
19		0.27	0
82		**0.12**	**0**

$X_1 \dots, X_k$ – model factors

$Prob$ – model prediction

Y – model target

Sensitivity of financial effect to model quality

Swap Probabilities of several observations

$X_1 \ldots X_k \rightarrow Prob \qquad Y$

X_1	…	$Prob$	Y
34		**0.12**	**1**
20		0.25	0
21	. . .	0.16	1
50		**0.33**	**1**
41		0.51	0
25		**0.86**	**1**
19		0.27	0
82		**0.94**	**0**

Permutation $N\%$ observation leads to model quality degradation by $\widehat{N}\%$

Calculate $\boldsymbol{Benefit}(\widehat{N}\%)$ with adjusted threshold

Sensitivity of financial effect to model quality

Repeat permutation for different $N \in [0; 1]$ as many times as possible

$\widehat{N}\%$	$Benefit$	$Quality$
0%	**5 M$**	**70%**
2%	4.8 M$	68%
7%	4.5 M$	65%
10%	4.1 M$	62%
11%	3.9 M$	61%
20%	3.1 M$	40%
50%	1.5 M$	33%
90%	0.5 M$	15%

without permutation

Calculate $\boldsymbol{Benefit}(\widehat{N}\%)$ with adjusted threshold for every iteration

50% permutation

Sensitivity of financial effect to model quality

Calculation difference between benefit and quality for every observations with permutation and without

$\widehat{N}\%$	$Benefit$	$Quality$
0%	**5 M$**	**70%**
2%	4.8 M$	**68%**
7%	**4.5 M$**	**65%**
10%	4.1 M$	**62%**
11%	3.9 M$	**61%**
20%	3.1 M$	**40%**
50%	1.5 M$	**33%**
90%	0.5 M$	**15%**

→

$\Delta Benefit$	$\Delta Quality$
0 M$	**0%**
0.2 M$	**2%**
0.5 M$	**5%**
0.9 M$	**8%**
1.1 M$	**9%**
1.9 M$	**30%**
3.5 M$	**37%**
4.5 M$	**65%**

Sensitivity of financial effect to model quality

Find sensitivity of financial effect to model quality

$\Delta Benefit$	$\Delta Quality$
0 M$	0%
0.2 M$	2%
0.5 M$	5%
0.9 M$	8%
1.1 M$	9%
1.9 M$	30%
3.5 M$	37%
4.5 M$	65%

$$\Delta Benefit \sim F(\Delta Quality)$$

$$\Delta Benefit = \frac{\partial \boldsymbol{Benefit}}{\partial \boldsymbol{Quality}} \cdot \Delta Quality$$

Sensitivity of financial effect to model quality

Wrap-up

1. Presence of two (best alternative and optimal) models allows calculating the model risk:

Wrap-up

1. Presence of two (best alternative and optimal) models allows calculating the model risk:
    - Ignoring the better challenger model

Wrap-up

1. Presence of two (best alternative and optimal) models allows calculating the model risk:
    - Ignoring the better challenger model
    - Incorrect / sub-optimal model implementation

Wrap-up

1. Presence of two (best alternative and optimal) models allows calculating the model risk:
   - Ignoring the better challenger model
   - Incorrect / sub-optimal model implementation
2. Sensitivity of financial effect to model quality is main information for risk calculation of model decay

Wrap-up

1. Presence of two (best alternative and optimal) models allows calculating the model risk:
    - Ignoring the better challenger model
    - Incorrect / sub-optimal model implementation
2. Sensitivity of financial effect to model quality is main information for risk calculation of model decay
3. Permutation can help to simulate different model quality and benefit