INTESA SANPAOLO

Quant roles in the Market, Financial and CI&B Risks Area

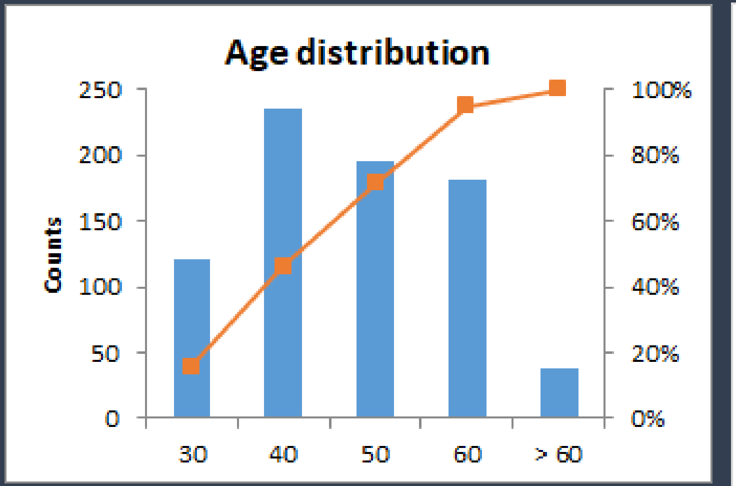
Mission

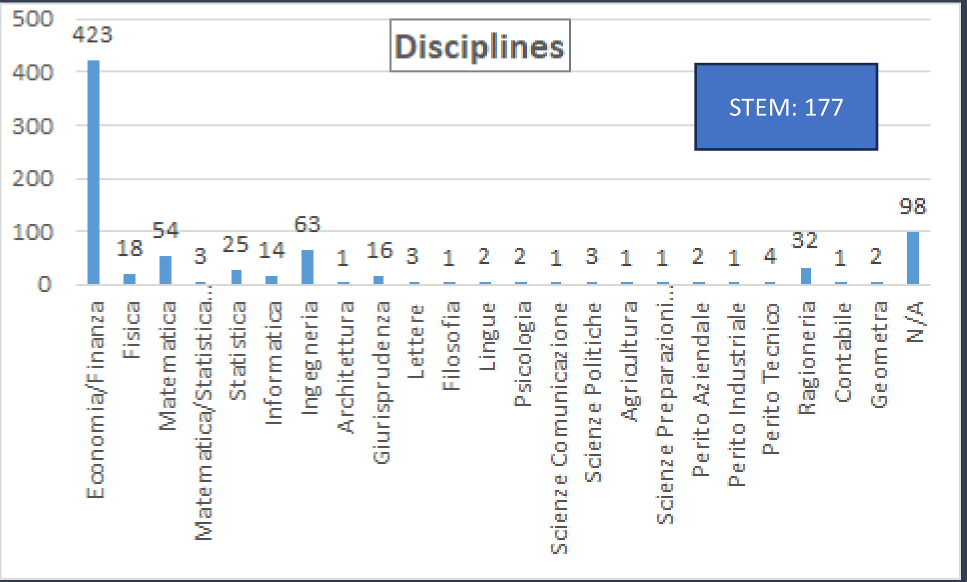
The Chief Risk Officer Area, working at ISP Group level, is managed by the Chief Risk Officer and reports to the Managing Director and CEO of the Group

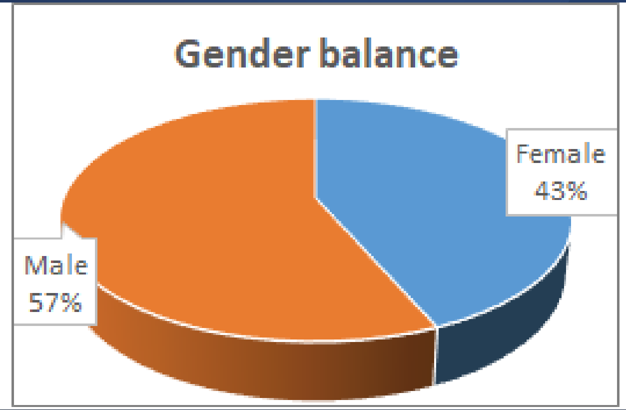
- ***Govern the macro-process of definition, approval, control and implementation of the Group's Risk Appetite Framework (RAF)*** *with the support of the other corporate functions involved*
- Consistent with corporate strategies and objectives, **assist the Bodies in defining and implementing guidelines and policies on risk management**
- Coordinate the **implementation of guidelines and policies on risk management** by the relevant Group business units, also in the various corporate contexts
- **Guarantee the measurement and control of Group exposure to the various types of risk**, also verifying the implementation of guidelines and policies as above
- **Perform II level monitoring and controls on credit quality**, composition and evolution of the various loan portfolios and on proper classification and measurement of single positions ("single name" controls)
- **Perform II level monitoring and controls for monitoring ICT and security risk**, as well as risks other than credit risk
- **Continuously and iteratively validate risk measurement and management systems** – used both for the determination of capital requirements and for non-regulatory purposes – in order to assess their compliance with regulatory provisions, operational company and reference market demands, and manage the internal validation process at Group level; in this context, ensure the definition and oversight of a framework for model risk governance

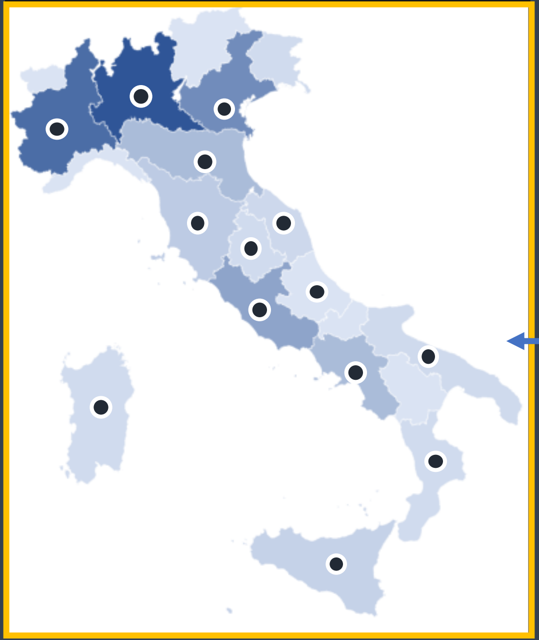
Team

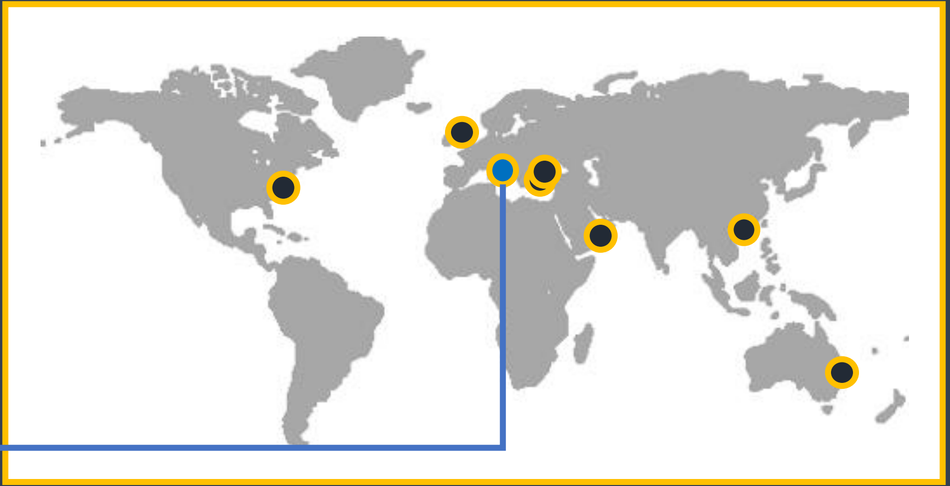
769 People











New York, London, Istanbul, Doha, Dubai, Hong Kong, Shanghai, Sidney

Market and Financial Risk Management

INTESA SANPAOLO

Market and Financial Risk Management
Risk Inventory and Taxonomy

- Credit Risk
- Market Risk
- Banking Book Financial Risk
- Liquidity Risk
- Insurance Risk
- Strategic Risk
- Pension Fund Risk
- Reputational Risk
- Model Risk
- Financial Statement Risk
- Non-Financial Risk - Operational Risk

Market Risk
- *Generic and specific position risk*
- *Issuer's specific risk*
- *Counterparty risk OTC/listed derivatives and SFT*
- *Trading book securitization risk*
- *Valuation risk*

Banking Book Financial Risk
- *Interest rate risk banking book*
- *Participation risk*
- *Foreign exchange risk*
- *Real estate risk*

Liquidity Risk
- *Short term liquidity risk*
- *Structural liquidity risk*

Market and Financial Risk Management

Main activities

Market and Counterparty risks

- Development and management of risk methodology and measurement, both for regulatory and managerial purposes
- Definition and monitoring of risk limits
- Capital absorption supervision, backtesting calculation total exposure
- Management reporting and escalation

Valutation risk

- Oversight policy for the valuation of financial instruments at fair value, prudent value, independent price verification
- Definition and monitoring of risk limits

Market and Financial Risk Management

IRRBB e Liquidity risks

- Development and management of IRRBB and Liquidity risk measurement systems
- Definition and monitoring of risk limits, evaluation of available reserves
- Reporting management and further compliance

Development of Risk Management Framework

- Development of the market risk management framework and design of its functional architecture
- Monitoring of front-to-risk supply chains with particular reference to the alignment of parameters and pricing models
- Observatory on digital evolution, innovation lab and market risk management resolution

Structure

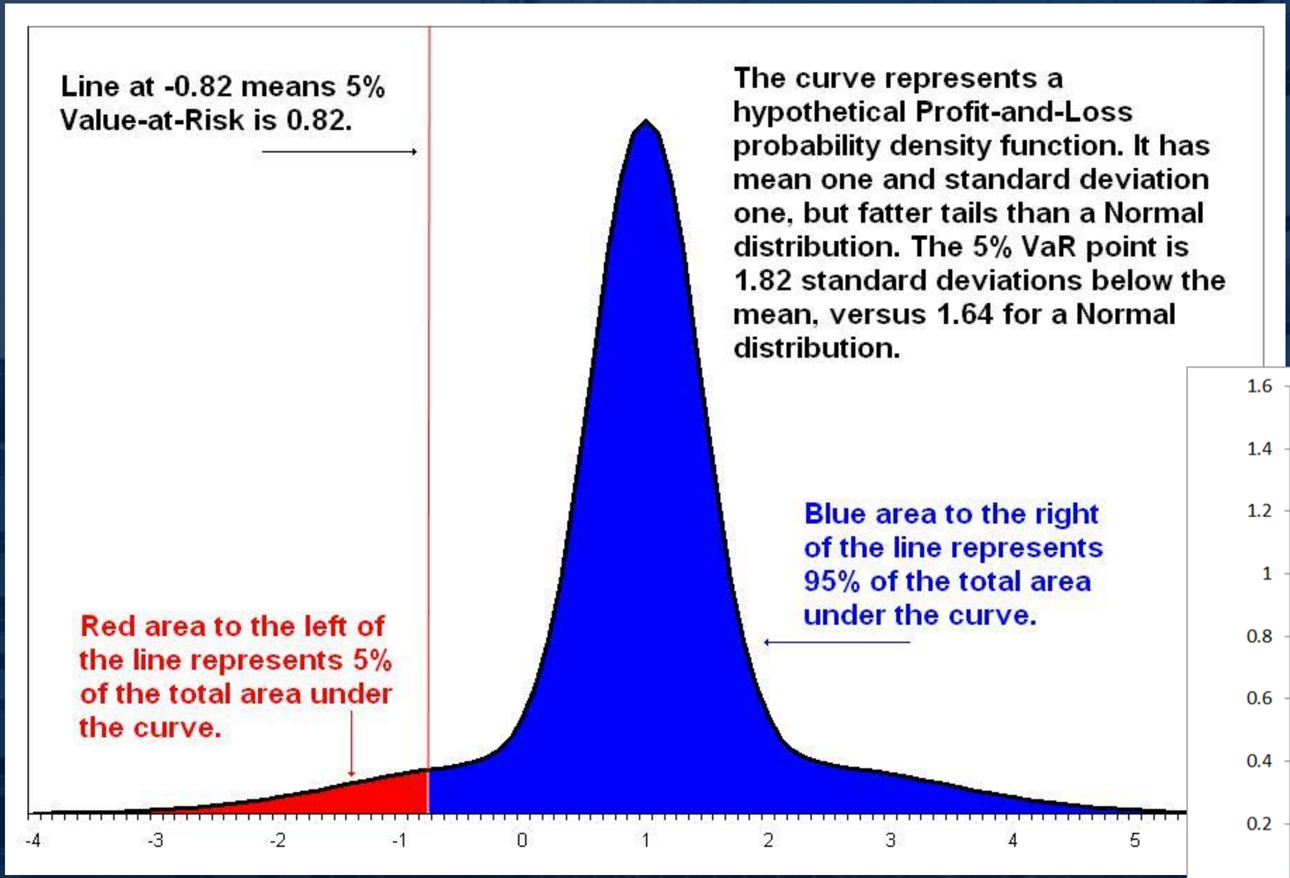
- 130 people
- 4 sub-departments
- 9 offices

Risk governance, new products and business models

- Management of regulatory body, definition of guidelines and rules documents
- Business models supervision
- Management of the market data management system
- Data quality and financial parameter correctness evaluation
- Control of the approval process for new products

Market and Financial Risk Management

Dealing with uncertainty

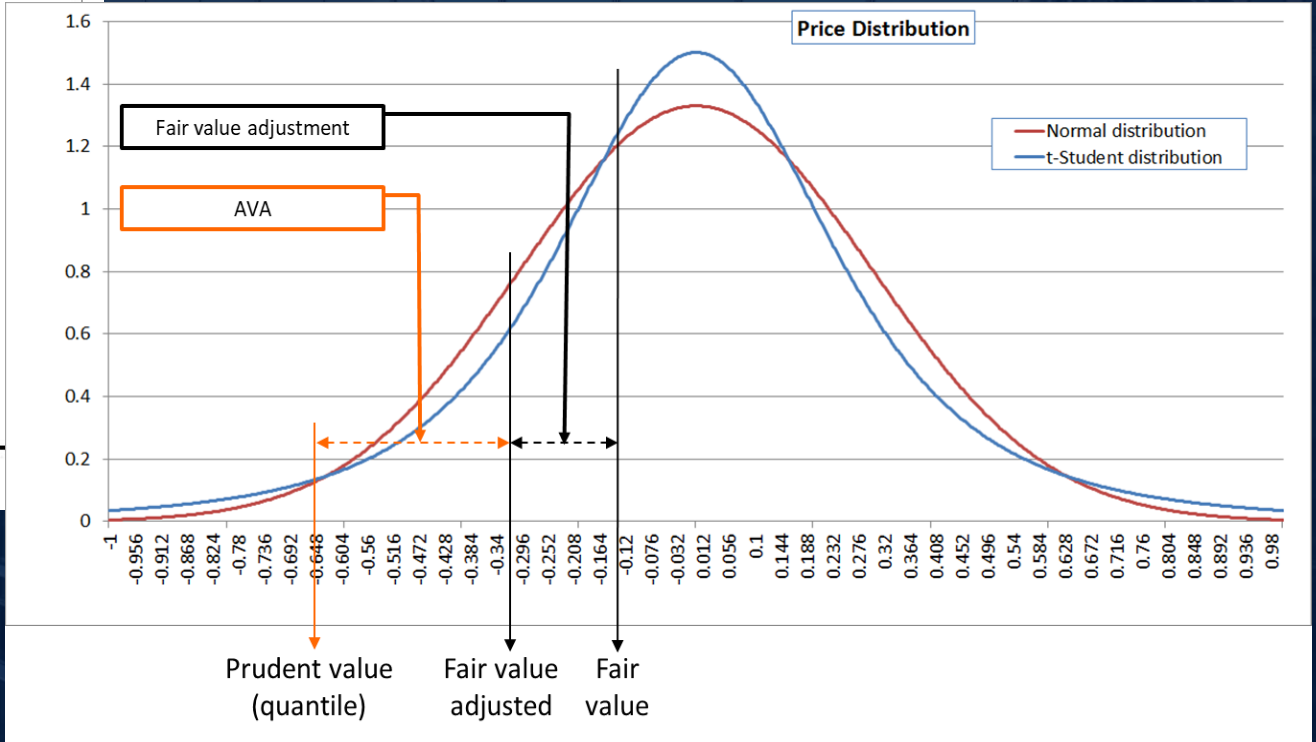


Distribution of profits & losses across a given time horizon

(source: Wikipedia)

Fair value, prudent value and risk measures are different things

Distribution of exit prices at the valuation date



INTESA SANPAOLO

Market and Financial Risk Management

The risk manager toolbox – Skills from university to financial industry

The toolbox

- Scientific approach
- Logic and rationality
- Identify the most important drivers in a complex problem
- Make analyses and experiments, find evidence
- Modelling
- Programming and using softwares
- Communication with different people
- Documentation from internal report to research articles

```
for m1 = 1,M do begin
 for m2 = 1,M do begin
  for u1 = u_min,u_max do begin
   for u2 = u_min,u_max do begin
    if u1 gt u2 then begin
     for v1 = v_min,v_max do begin
      if v1 lt u1 then begin
       for v2 = v_min,v_max do begin
        if v2 ge v1 then begin
         KE_B = double(m1*u1^2+m2*u2^2)
         KE_A = double(m1*v1^2+m2*v2^2)
         if (KE_B gt KE_A) and (KE_A ge 0.965*KE_B) then begin
          x_axis[index]=index
          LM_B = double(m1*u1+m2*u2)
          LM_A = double(m1*v1+m2*v2)
          y_LM_Diffs[index]=LM_B-LM_A
          Total_LM=Total_LM+LM_B-LM_A
          y_LM_Total[index]=double(Total_LM/(index+1))
          index=index+1
          if index gt 65535 then goto, end_of_loop
         endif
        endif
       endfor
      endif
     endfor
    endif
   endfor
  endfor
 endfor
endfor
```

The strategy

- Look for the right place with the right people and learn on the job
- Think differently, make questions
- Work hard, work as a team

INTESA SANPAOLO


INTESA SANPAOLO
Our selection process
INTESA SANPAOLO

Our selection process

- RESEARCH AND APPLICATION
- REMOTE VIDEO INTERVIEW
- INTERVIEW
- ON-BOARDING

The remote video interview

The digital video interview is the **first selection step** for most of our open positions.

Apply for a professional position; if your profile meets the requirements, you will receive an invitation to carry out your video interview.

Turn on your PC or smartphone, take your time **whenever and wherever** you feel most comfortable and answer the questions of our recruiters by recording your answers.

You can **practice your responses** and record your video many times before sending your final release.

Tips for the Video interview

Pay attention to the moment and to the location

Dress as if it were a face to face interview

Stay relaxed and smile

Prepare a list of key points and concepts

Read the job offer again

Practise with some tests

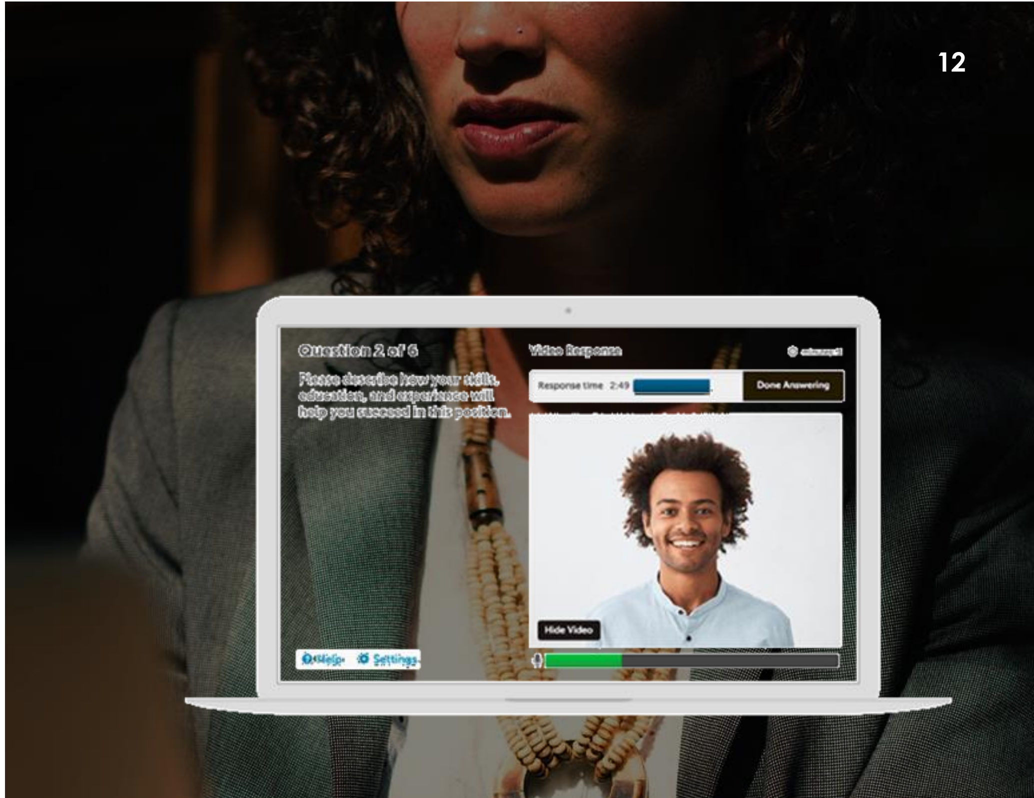
Check your answers before confirming

Focus on the recruiter's questions

Look at the camera

Describe in details your experiences

Be yourself



Innovative and Simple

Two chances for a first good impression!

Wherever it's convenient for you

Whenever it's convenient for you

From any device

The Interview

Real time behavioural Interview

During the interview we ask candidates to describe past situations and tasks that are relevant in terms of knowledge, skills, and abilities. The assumption is that past behavior is the best predictor of future performance in similar situations.

Technical Interview

During the interview the technical line managers will check candidate's competences and knowledge related to the vacant position.

INTESA SANPAOLO

Internships: open positions

CRO Financial Market & CIB Risk- Stage curriculare ed extracurriculare

- Link: https://jobs.intesasanpaolo.com/job/Milano-CRO-Financial-Market-&-CIB-Risk-Stage-curriculare-ed-extra-curriculare/1030111201/
- Always open, no deadline (hopefully)
- M.Sc. preferred
- Time to activate the stage: 1-2 months, start date flexible
- Length: typically 6 months full time
- Where: embedded in a single office, driven by a tutor
- M.Sc. Thesis: few projects available, to be discussed
- When to apply:
    - curricular stage: apply once max 3 exams + thesis are left,
    - Extra-curricular stage: apply a couple of months before graduation

Q&A

Learning Market Data Anomalies

Padova, 6th May 2024

Marco Bianchetti[1,2]

Joint work with, Giuseppe Crupi[3], Marco Scaringi[1], Manola Santilli[1]

[1]Financial and Market Risk Management Department, Intesa Sanpaolo
[2]University of Bologna, Department of Statistical Sciences «Paolo Fortunati»
[3]Iason Italia srl

Acknowledgements and disclaimer

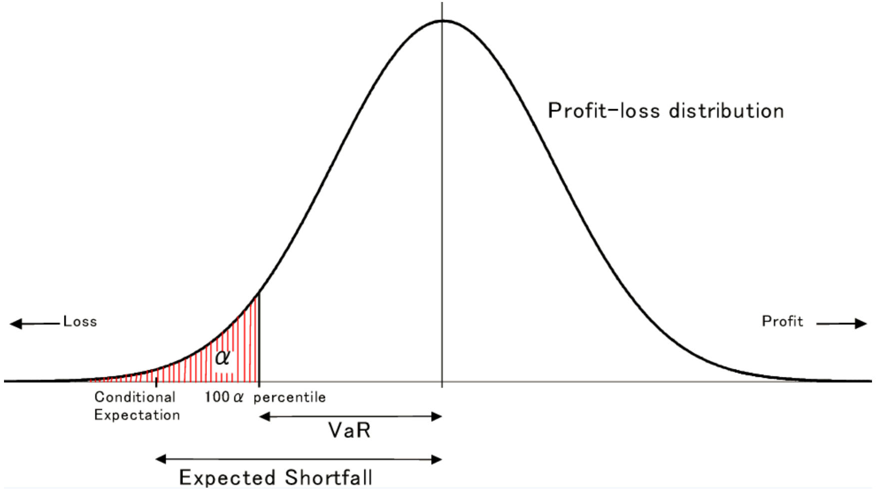
The authors gratefully acknowledge fruitful interactions with many colleagues at Intesa Sanpaolo, E. Biadene and M. Pavan at Deloitte Consulting, F. Codazzi and Prof. D. E. Galli at Physics Department, Università degli Studi di Milano.

Summary

1: Introduction

Market Risk framework and market data management

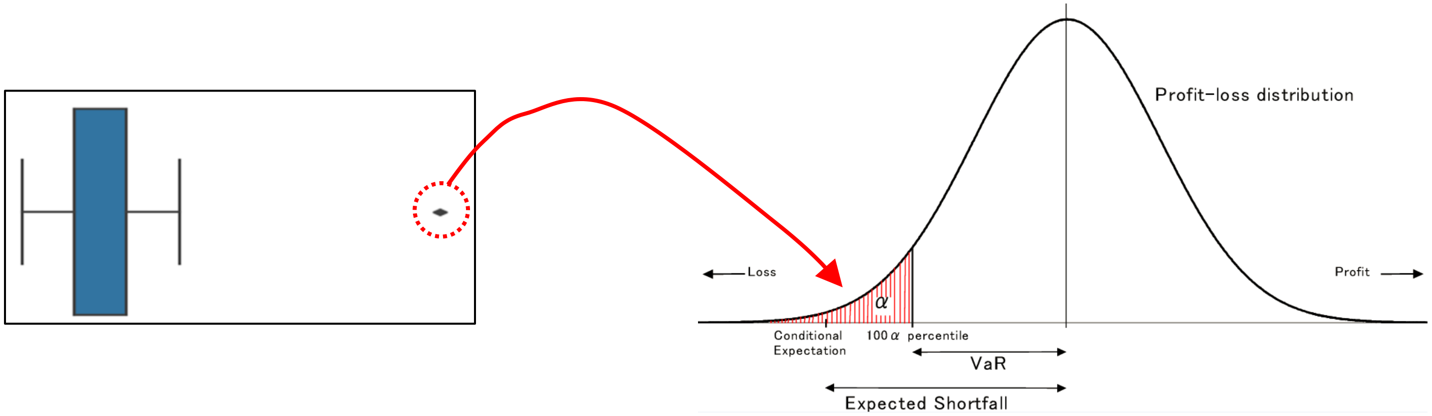
- **Market risk** is defined as "*the risk of losses in on and off-balance-sheet positions arising from movements in market prices*" (see BCBS, sec. MAR 10.1). It is measured through a number of **market risk measures**, e.g. value at risk (VaR), expected shortfall (ES), etc.
- The calculation of market risk measures for a portfolio of financial instruments (derivatives, bonds, funds, etc.) is based on the **distribution of profits and losses** (P&Ls) across a number of **market risk scenarios** (e.g. 250) for a given **time horizon** (e.g. 1 day) and a given **confidence level** (e.g. 99%). Such scenarios can be generated using different methodologies, but, in any case, one typically needs **historical series of market data for each market risk factor underlying the financial instruments included in the portfolio**.



- Real trading **portfolios of large financial institutions** may contain **10^5-10^6 financial instruments** (even more) across different currencies and asset classes (interest rates, inflation, credit, equity, commodity and forex), leading to **millions or billions of data points** to compute daily market risk measures.
- Even if these figures may significatively change across institutions, depending on their size, business model and risk methodologies, a **robust market data management process** is clearly a cornerstone of the whole market risk management framework.

1: Introduction
Anomalies in Market Risk framework

- Some data may present **anomalous values** because of a wide range of reasons, e.g. bugs in the related production processes, sudden and severe market movements, etc. These anomalies may have important **consequences on risk measures** which, by definition, measure tail risk.



- Hence, it is crucial to integrate the daily data quality process with **semi-automatic and statistically robust tools** able to smartly analyze all the available information and **identify possible relevant anomalies.**

1: Introduction

Market data anomaly detection

In this work we deal with different **unsupervised machine learning models** to detect possible anomalies in market data widely used in market risk measures. They are

- **Isolation Forest (IF)**
- **Autoencoder (AE)**
- **Long Short-Term Memory (LSTM)**

In particular, we apply these models to marked data sets widely used in market risk management, i.e. **interest rate curves** and **volatility surfaces**

Our approach is completely **general** since it may be applied to market data for different

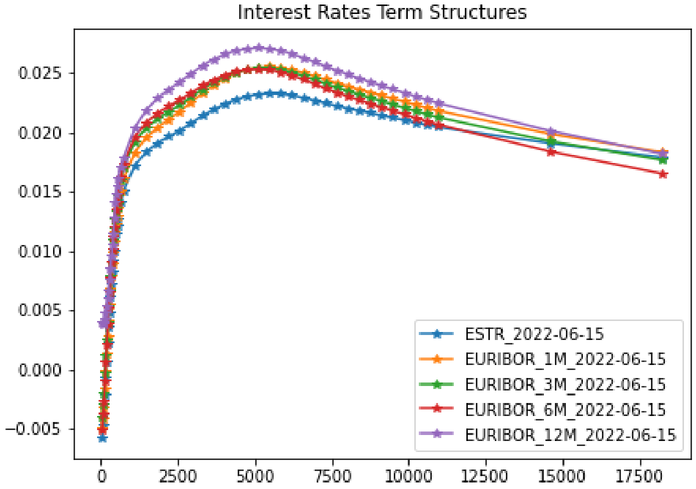
- **asset class**, e.g. interest rates, equity, credit, etc…;
- **typology**, i.e. market quotations (e.g. IRS swap rates, equity options prices matrix) or market-implied quantities (e.g. zero rate curves or implied volatilities);
- **dimension**, e.g. 1-dimensional interest rates curves or bond yields curves, 2-dimensional volatility surfaces, 3-dimensional swaption volatility cubes.

See references, in particular, A. Sokol, Autoencoder Market Models for Interest Rates, 2022 https://ssrn.com/abstract=4300756

2: Dataset #1

Interest rate curves

The dataset is a collection $\{C_1^x, \ldots, C_N^x\}$ of historical series of **EUR interest rate curves** with different **5 tenors** $x \in \{€STR, Euribor1M, Euribor3M, Euribor6M, Euribor12M\}$ and length $N = 2691$ businness days. Each curve C_i^x at a given business date t_i is a vector (called **term structure**) of zero coupon rates $\{r_{i,j}^x\}$ with fixed underlying tenor x and maturity (a.k.a. **term** or **pillar**) t_j. Each curve C_i^x is built from the corresponding market instruments (Deposits, Futures, FRAs, IRSs with the same tenor) through a mathematical/numerical procedure known as **bootstrapping**.



IR Term structures as of 15th June 2022

$$C_1^x = \{r_{1,1}^x, \ldots, r_{1,j}^x, \ldots, r_{1,50}^x\}$$

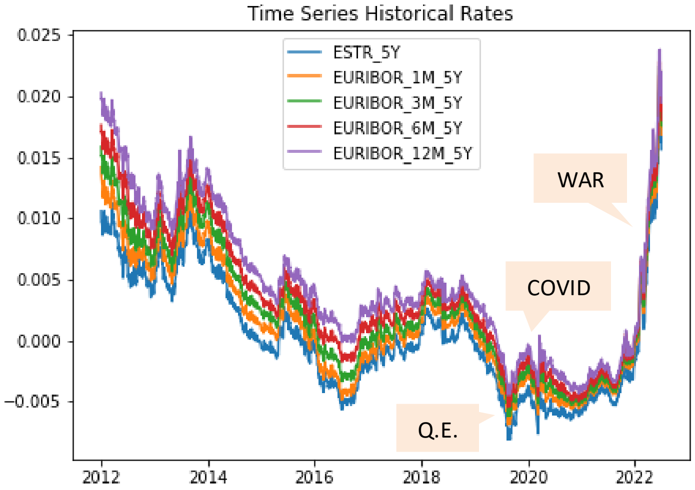
$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$C_N^x = \{r_{N,1}^x, \ldots, r_{N,j}^x, \ldots, r_{N,50}^x\}$$

$$r_{i,j}^x = r^x(t_i, t_j)$$

range 2012-2022

→ 5 x 50 x 2691 dataset



IR Historical series of 5Y pillars

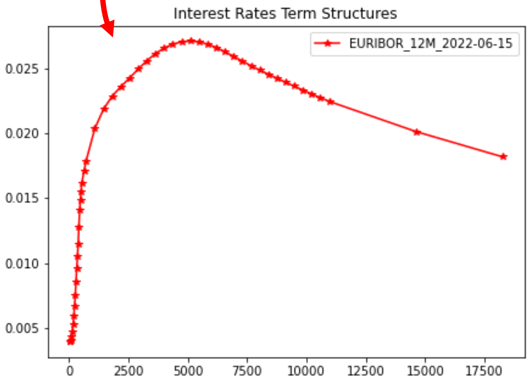
See e.g. F. Ametrano, and M. Bianchetti,, Everything You Always Wanted to Know About Multiple Interest Rate Curve Bootstrapping but Were Afraid to Ask (April 2, 2013). Available at SSRN: https://ssrn.com/abstract=2219548

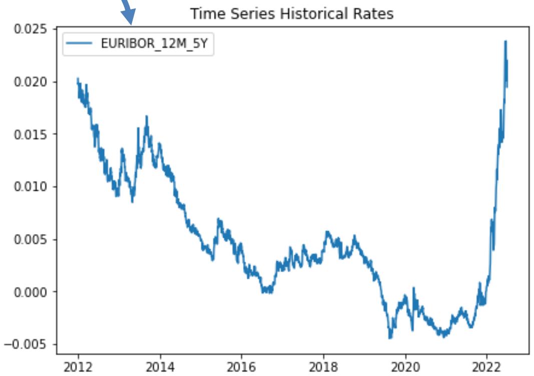
2: Dataset #1

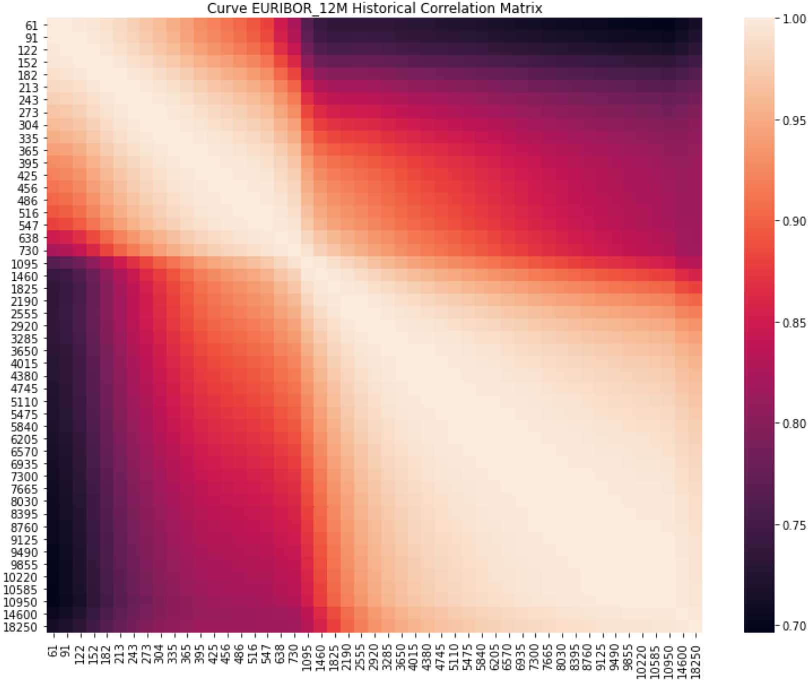
Interest rate curves: focus on Euribor12M

BUSINESS_DATE	30D	...		5Y	6Y		...		50Y
2022-06-24	-0.005812	-0.004610	-0.003415	-0.002139	-0.000775	0.000433	0.001664	0.002793	0.003822
2022-06-23	-0.005871	-0.004663	-0.003438	-0.002108	-0.000700	0.000539	0.001814	0.002951	0.003998
2022-06-22	-0.005874	-0.004581	-0.003405	-0.001958	-0.000448	0.000896	0.002352	0.003639	0.004831
2022-06-21	-0.005835	-0.004526	-0.003375	-0.001843	-0.0002	[illegible]	0.002586	0.003898	0.005111
2022-06-20	-0.005836	-0.004502	-0.003383	-0.001845	-0.0002	[illegible]	0.002569	0.003870	0.005071
2022-06-17	-0.005823	-0.004676	-0.003582	-0.002029	-0.000494	0.000855	0.002353	0.003659	0.004837
2022-06-16	-0.005829	-0.004710	-0.003635	-0.002053	-0.000489	0.000916	0.002504	0.003873	0.005134
2022-06-15	-0.005824	-0.004751	-0.003761	-0.002179	-0.000703	0.000681	0.002223	0.003528	0.004777
2022-06-14	-0.005882	-0.004822	-0.003908	-0.002239	-0.000726	0.000715	0.002328	0.003734	0.005109
...				...					...

$r^{12M}_{i,j}$







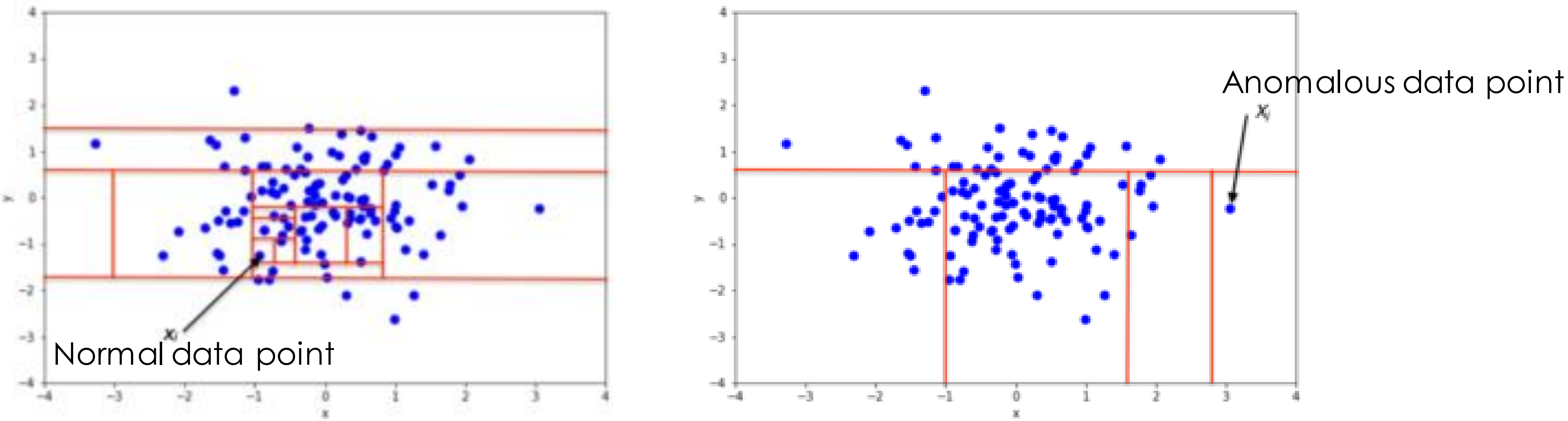
3: Methodologies

Overview

Algorithm	Model	Main Characteristics	Anomaly detection target	Anomaly dependence on the history	Anomaly parameter
Isolation Forest	**1 dim. Isolation Forest (1DIF)**	**Ensamble decision tree-based algorithm**; anomalies are identified using lines (planes) parallel to the axes to separate data points. The model assigns higher anomaly scores to points that need few splits to be isolated. The model is applied to **each single pillar of the curve**	Hisorical normalized **zero rate returns**. Only local information (no information about the shape of the curve)	Full: the score of each data point depends on the relative position of **all the other data points**	**Contamination threshold** (2% of the top scored points)
	3 dim. Isolation Forest (3DIF)		Includes also **first and second derivatives** to carry non-local information about first neighbors		
Neural Networks	**Autoencoder (AE)**	**Neural network used to reduce the dimension of inputs into** a **smaller representation** (compression and noise neglection). Then data are reconstructed in their original dimension **reproducing in a good way normal data, in a bad way anomalous data.**	**The entire interest rate curve** (all pillars). The predicted curve is compared with the actual one.	**No direct time dependency** (only via training)	**Anomaly threshold** (98° percentile of RMSE distribution)
	Long Short-Term Memory (LSTM)	**Recurrent neural network**; the behaviour of the data points in **the past affects the prediction** of subsequent one. Every node consists in four steps, in which **the previous information is stored** depending on its importance in the prediction.		**Via the previous n curves** (and via training)	

4: Isolation Forest
Methodology

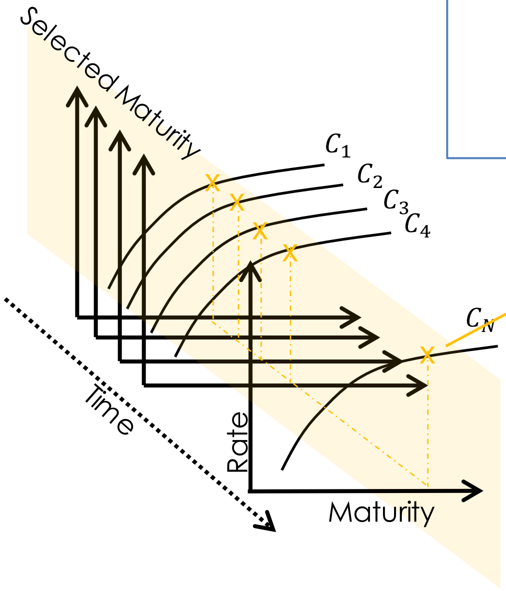
- The Isolation Forest algorithm is an **ensamble of isolation trees**, which splits the data space randomly selecting an attribute using lines orthogonal to the origin and assigns **higher anomaly scores to data points that need few splits to be isolated**.
- The output is a **hierarchical tree structure**. Anomalies require short paths within the tree to reach a terminating node starting from the root. The results are averaged among all the trees.
- The percentage of (top ranked) data labelled as anomaly is an **hyperparameter** (**contamination threshold**).
- Data preprocessing: **min max normalization of daily returns.**



We thank E. Biadene and M. Pavan for the results reported in this section

4: Isolation Forest

Anomaly detection using 1-dimensional Isolation Forest (1DIF)



$$C_1^x = \{r_{1,1}^x, \dots, r_{1,j}^x, \dots, r_{1,50}^x\}$$

$$\dots$$

$$C_N^x = \{r_{N,1}^x, \dots, r_{N,j}^x, \dots, r_{N,50}^x\}$$

$$R_j^x = \{r_{1,j}^x, \dots, r_{N,j}^x\} \in \mathbb{R}^N$$

i-th pillar historical series

1-dim **Isolation Forest** (one feature)

For each single pillar t_j the 1DIF model assigns a **score** to each data point and labels as **outlier** the top k% scored.

$$S_j = \{s_{1,j}, \dots, s_{N,j}\} \in \mathbb{R}^N \ (score)$$

$$L_j = \{l_{1,j}, \dots, l_{N,j}\} \in \mathbb{R}^N, \ l_{i,j} \in \{0,1\}$$

$$\frac{\sum_{i=1}^{N} l_{i,j}}{N} \approx k$$

Remarks

- We look at **each single pillar separately**, not at the whole curve.
- We take into account **one single feature**, i.e. the rate level.

4: Isolation Forest

Anomaly detection using 3-dimensional Isolation Forest (3DIF)

zero rate

$$C_1^x = \{r_{1,1}^x, \dots, r_{1,j}^x, \dots, r_{1,50}^x\}$$

...

$$C_N^x = \{r_{N,1}^x, \dots, r_{N,j}^x, \dots, r_{N,50}^x\}$$

1st derivative (centered finite difference)

$$\dot{C}_1^x = \{\dot{r}_{1,1}^x, \dots, \dot{r}_{1,j}^x, \dots, \dot{r}_{1,50}^x\}$$

...

$$\dot{C}_N^x = \{\dot{r}_{N,1}^x, \dots, \dot{r}_{N,j}^x, \dots, \dot{r}_{N,50}^x\}$$

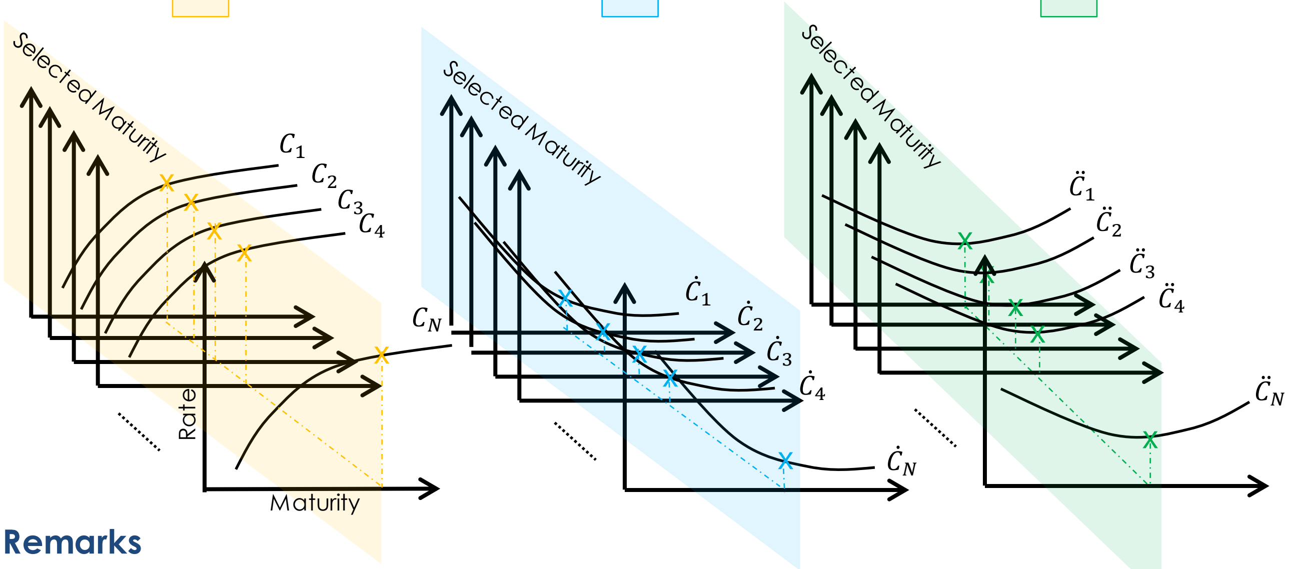
2nd derivative (centered finite difference)

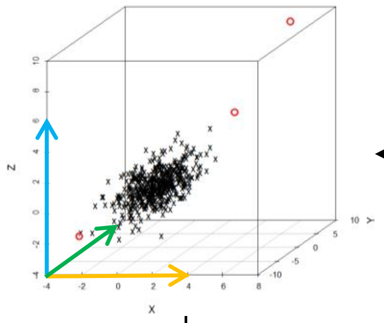
$$\ddot{C}_1^x = \{\ddot{r}_{1,1}^x, \dots, \ddot{r}_{1,j}^x, \dots, \ddot{r}_{1,50}^x\}$$

...

$$\ddot{C}_N^x = \{\ddot{r}_{N,1}^x, \dots, \ddot{r}_{N,j}^x, \dots, \ddot{r}_{N,50}^x\}$$

$$P_j^x = \begin{Bmatrix} r_{1,j}^x & & r_{N,j}^x \\ \dot{r}_{1,j}^x, & \dots, & \dot{r}_{N,j}^x \\ \ddot{r}_{1,j}^x & & \ddot{r}_{N,j}^x \end{Bmatrix}$$





3-dim
Isolation Forest
(3 features)

$$S_j = \{s_{1,j}, \dots, s_{N,j}\} \in \mathbb{R}^N \ (score)$$

$$L_j = \{l_{1,j}, \dots, l_{N,j}\} \in \mathbb{R}^N, \ l_{i,j} \in \{0,1\}$$

$$\frac{\sum_{i=1}^N l_{i,j}}{N} \approx k$$

Remarks

- We look at each **single pillar plus first neighbours**, not at the whole curve.
- We take into account **three features**, i.e. rate level, slope and curvature.

4: Isolation Forest

Time-window selection

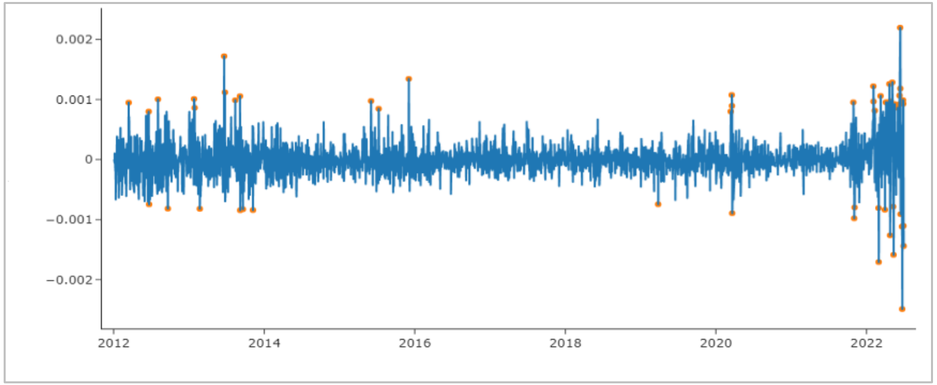
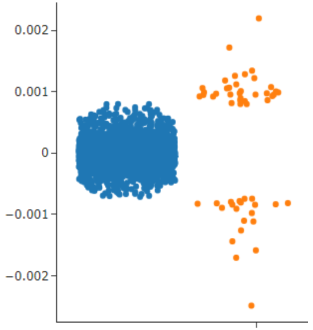
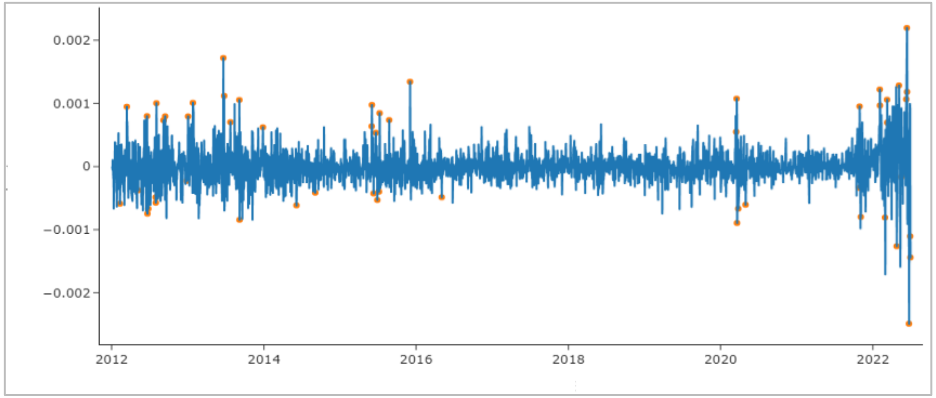
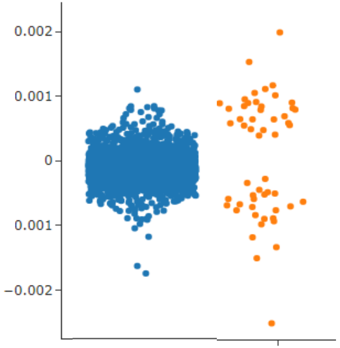
- Two different ways to deal with the **incoming data**

Increasing time window length	1-dim **Isolation Forest**	**fixed** time window length
$R_i^x = \{r_{1,i}^x, r_{2,i}^x, r_{3,i}^x, \dots, r_{N,i}^x\}$		$R_i^x = \{r_{1,i}^x, r_{2,i}^x, r_{3,i}^x, \dots, r_{N,i}^x\}$
$R_i^x = R_i^x \cup r_{N+1,i}^x$ →	l_{N+1}	← $R_i^x = \{\ \ , r_{2,i}^x, r_{3,i}^x, \dots, r_{N,i}^x, r_{N+1,i}^x\}$
$R_i^x = R_i^x \cup r_{N+2,i}^x$ →	l_{N+2}	← $R_i^x = \{\ \ ,\ \ , r_{3,i}^x, \dots, r_{N,i}^x, r_{N+1,i}^x, r_{N+2,i}^x\}$
$R_i^x = R_i^x \cup r_{N+3,i}^x$ →	l_{N+3}	← $R_i^x = \{\ \ ,\ \ ,\ \ , \dots, r_{N,i}^x, r_{N+1,i}^x, r_{N+2,i}^x, r_{N+3,i}^x\}$

4: Isolation Forest

Results: Euribor 12M, 5Y pillar

● Outlier
● Inlier

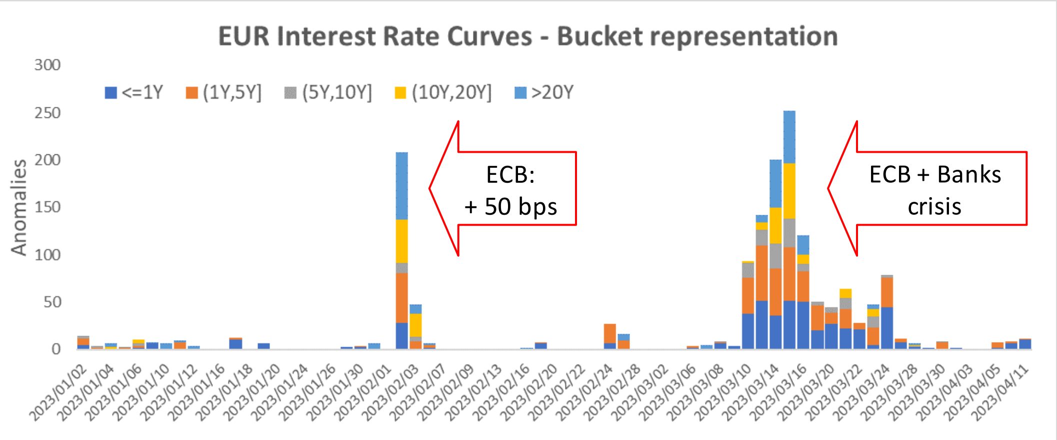
	Anomalies in zero rate daily returns	Distribution
1DIF		
3DIF		

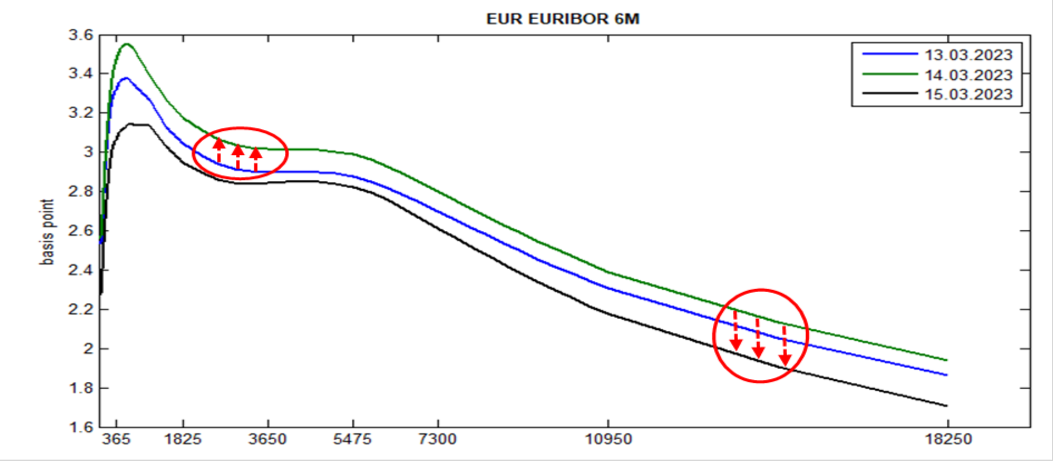
Remarks

- **Trivial result**: anomalies are nothing but the largest returns in the sample (i.e. the top 2% scored)
- **Non-trivial result**: the inclusion of additional features results in a **reshuffling of the anomalies** (the total number is still the top 2% scored). A point may be anomalous because of slope and curvature, not only because of level

4: Isolation Forest

Real-time application





1) Aggregated number of signals reported by **Isolation Forest** for **all EUR interest rate curves**, (€STR OIS, EURIBOR 1M, 3M, 6M, 12M) for 5 maturity buckets from **02.01.2023** to **11.04.2023** (69 business days).

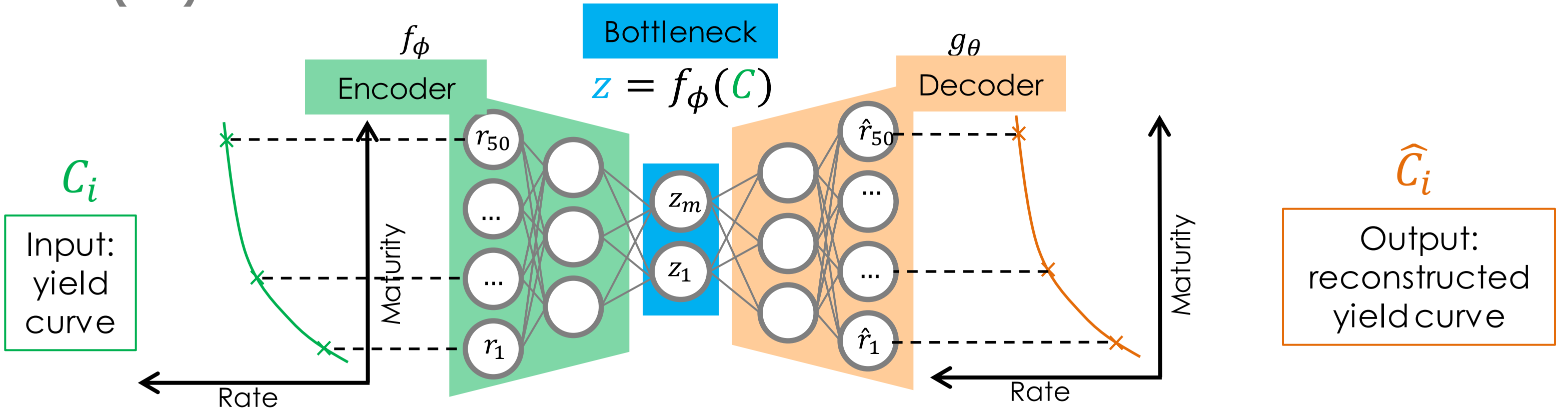
We notice that **signals are concentrated in shorter** (<=1Y and (1Y,5Y]) **and longer** ((10Y,20Y] and >20Y) **parts of the curves**, which are more sensitive to market events for liquidity reasons.

As expected, **signals are concentrated into periods of higher market volatility**. For instance, in March 2023 the ECB announced a series of rate hikes in order to bring down inflation.

2) **EUR EURIBOR 6M** curve in the three business days with the highest number of signals (13-14-15 Mar. 2023). The first scenario (14 vs 13) is a large upward bump of the curve for which IF detects signals in the short-medium terms, while the following scenario (15 vs 14) is a parallel down shift of the whole curve (rebound) for which IF detects signals almost on the whole curve.

5: Neural Networks

Autoencoder (AE)



- The **Autoencoder** (AE) is a neural network able to extract the salient features from the dataset $\{C_i\}_{i=1}^N$ through a data **compression and decompression** procedure. The magnitude of the reconstruction error is a measure of anormality.
- The reconstruction $\widehat{C}_i$ of an input sample $C_i \in \mathcal{X} \subseteq \mathbb{R}^{50}$ is performed in two steps:
  - **Encoder:** function $f_\phi: \mathcal{X} \to \mathcal{Z}$ mapping input data into the **latent space** $\mathcal{Z} \subseteq \mathbb{R}^m$ where $m \ll 50$.
  - **Decoder**: function $g_\theta: \mathcal{Z} \to \mathcal{X}$ mapping latent space data to the original data space such that

$$\hat{C} = g_\theta(z) = g_\theta\left(f_\phi(C)\right)$$

- AE learns the map $f_\phi \circ g_\theta$ looking for the **optimal AE parameters** $\{\hat{\phi}, \hat{\theta}\}$ which minimize the **reconstruction error** (RMSE objective function) such that $\widehat{\boldsymbol{C}} \approx \boldsymbol{C}$.

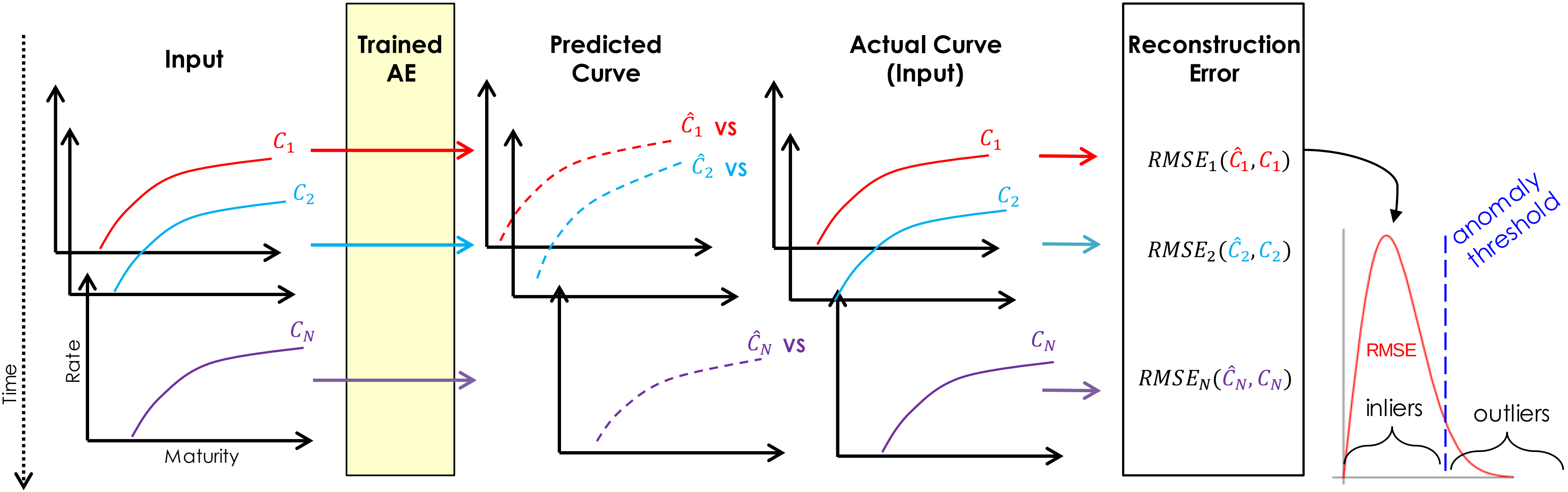
$$\{\hat{\phi}, \hat{\theta}\} = \underset{\phi,\theta}{\operatorname{argmin}} \sum_{i=1}^{N} RecErr(C_i; \phi, \theta), \qquad RecErr(C_i; \phi, \theta) = \left\| g_\theta\left(f_\phi(C_i)\right) - C_i \right\| = RMSE(\hat{C}_i - C_i) = \left[\frac{1}{n=50} \sum_{j=1}^{n=50} \left(\hat{r}_{i,j} - r_{i,j}\right)^2\right]^{\frac{1}{2}}$$

5: Neural Networks

Autoencoder: anomaly detection

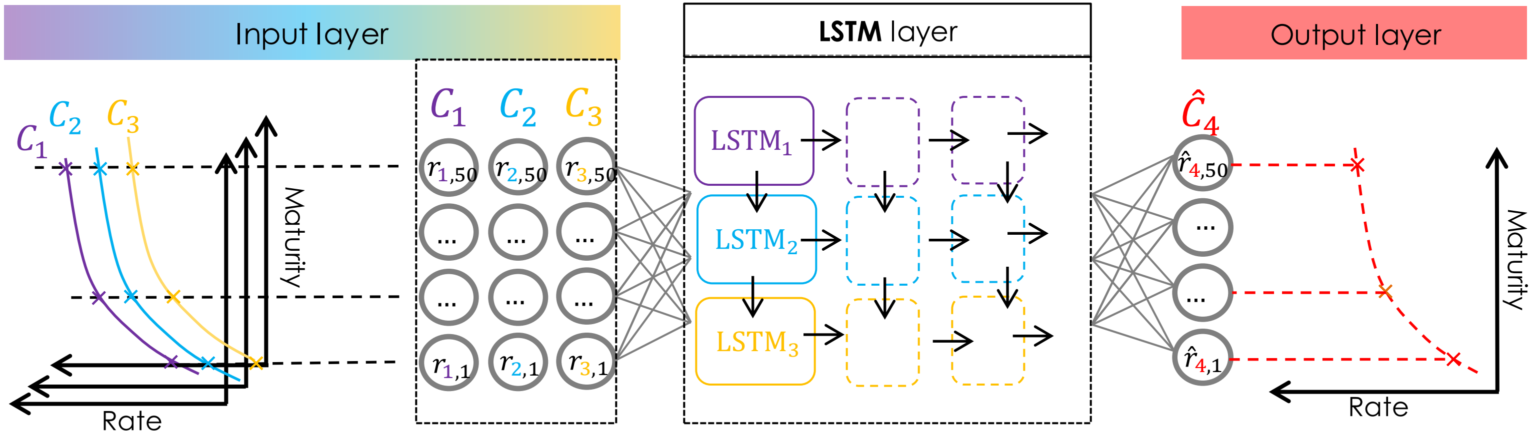
Anomaly detection using autoencoders proceeds through the following steps:

- feed a **trained AE** with the dataset $\{C_i\}_{i=1}^N$ of yield curves, **one by one**;
- produce the output **reconstructed yield curves** $\{\hat{C}_i\}_{i=1}^N$ and their **RMSEs**;
- label «**anomalies**» the curves which lie above a given percentile (**anomaly threshold**) of the RMSE distribution;
- manually **check** the detected anomalies and **fine tune** the anomaly threshold.



5: Neural Networks

Long-Short Term Memory (LSTM)



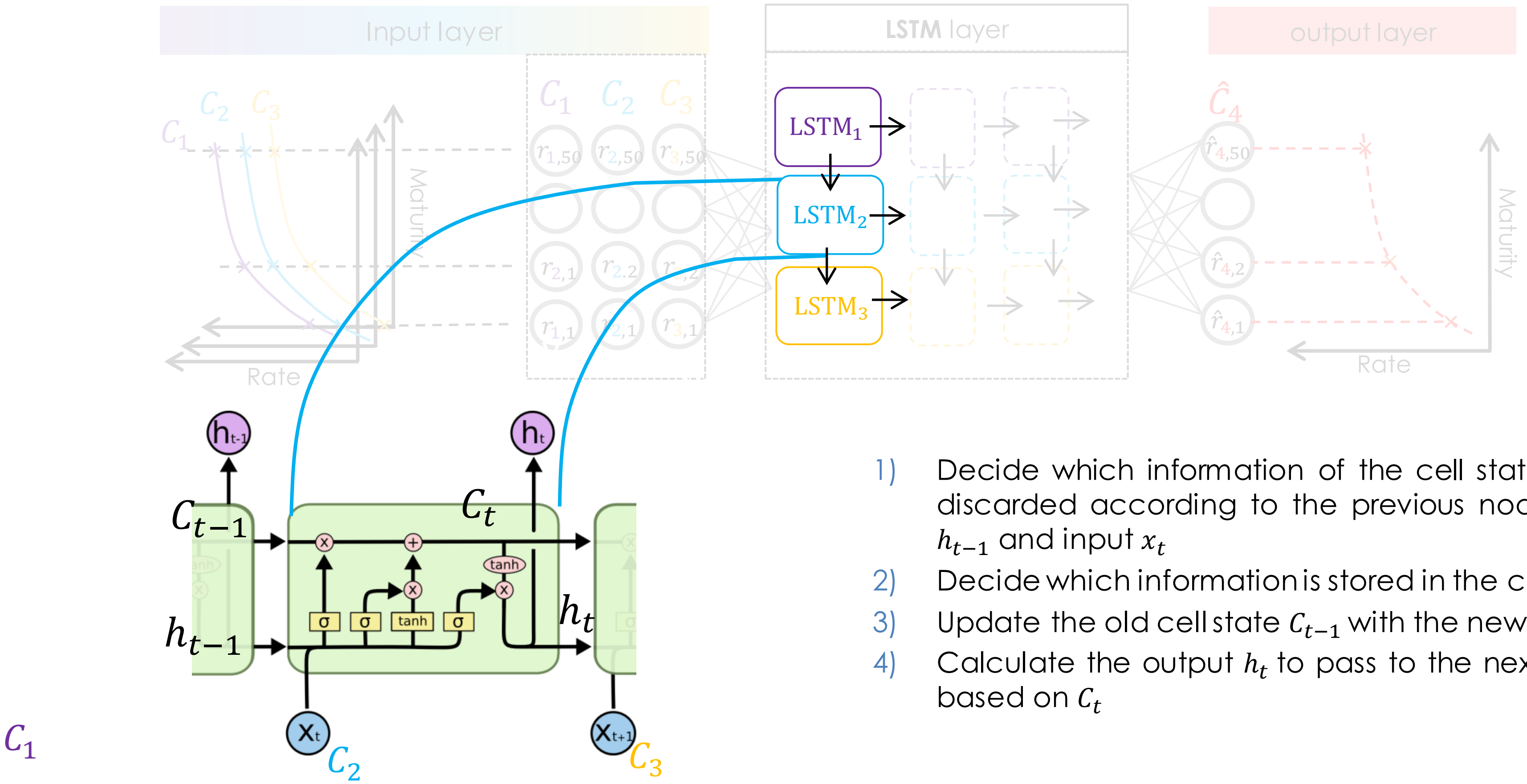
- The LSTM is a **recurrent neural network** (RNN) algorithm, which leverages on the information of the previous data to learn patterns and forecast future data. It combines **long-term** and **short-term** information through a complex **gate control** structure.
- The input is a subset of yield curves $C_{i-k}, C_{i-k+1}, \dots, C_{i-1}, k > 1$ ($k = 1$ would be similar to the AE). LSTM learns the map f_p looking for the **optimal LSTM parameters** p which minimize the **reconstruction error** (RMSE objective function) such that $\hat{C} \approx C$.

$$\{\hat{p}\} = \underset{p_{LSTM}}{\operatorname{argmin}} \sum_{i=1}^{N} RecErr(C_i; p),$$

Input curves for curve $\hat{C}_i$ reconstruction

$$RecErr(C_i; p) = \left\| f_p(C_{i-k}, C_{i-k+1}, \dots, C_{i-1}) - C_i \right\| = RMSE(\hat{C}_i - C_i) = \left[\frac{1}{50} \sum_{j=1}^{50} (\hat{r}_{i,j} - r_{i,j})^2 \right]^{\frac{1}{2}}$$

5: Neural Networks
Long-Short Term Memory (LSTM)



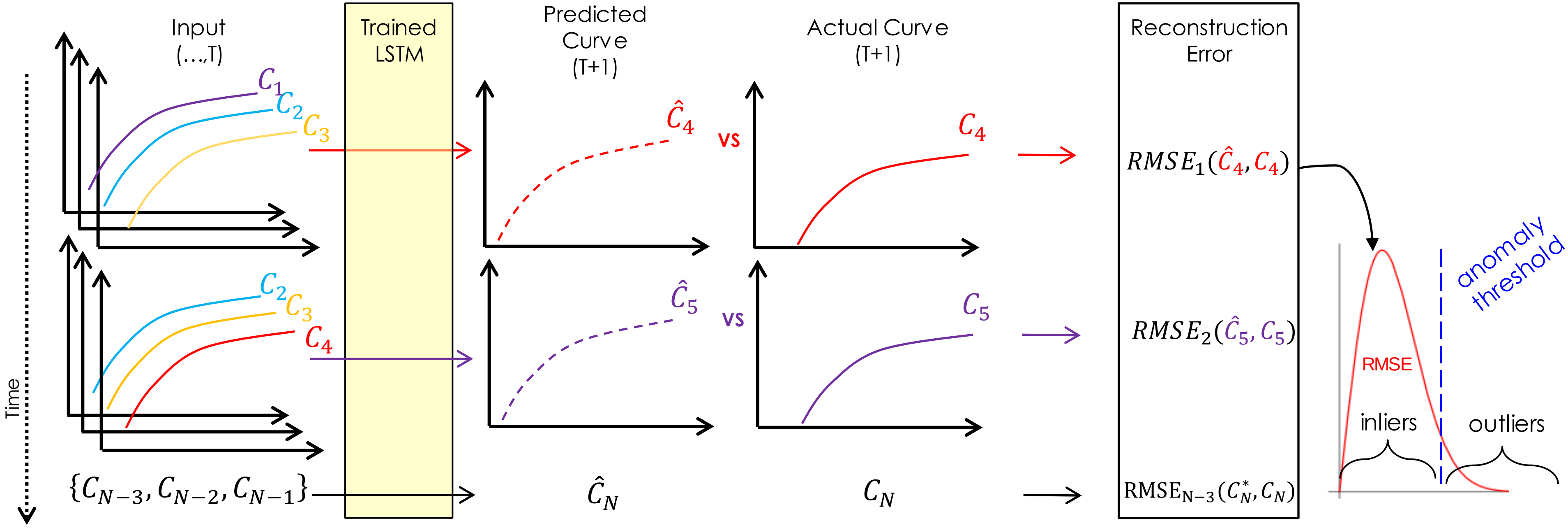
1) Decide which information of the cell state C_{t-1} is discarded according to the previous node state h_{t-1} and input x_t
2) Decide which information is stored in the cell state
3) Update the old cell state C_{t-1} with the new C_t
4) Calculate the output h_t to pass to the next node, based on C_t

5: Neural Networks

LSTM: anomaly detection

Anomaly detection using LSTM proceeds through the following steps:

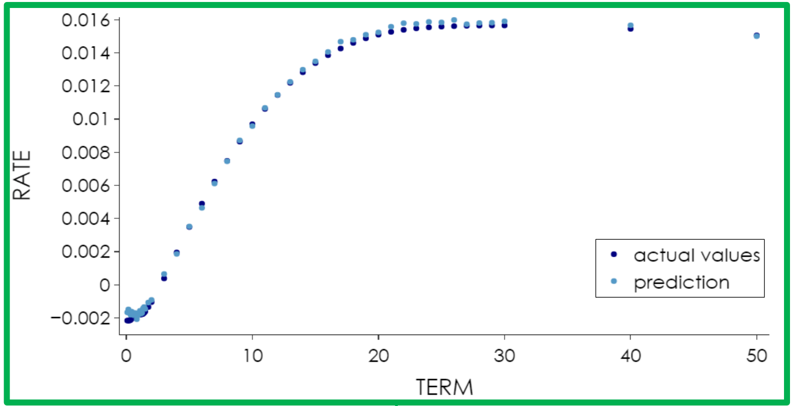
- feed a **trained LSTM** with the **dataset** $\{C_i\}_{i=1}^N$ of yield curves, packed in subsets of *k* elements;
- produce the output **reconstructed yield curves** $\{\hat{C}_i\}_{i=1}^N$ and their RMSEs;
- label «**anomalies**» the curves which lie above a given percentile (**anomaly threshold**) of the **RMSE distribution**;
- **manually check** the detected anomalies and **fine tune** the anomaly threshold and subset *k*.



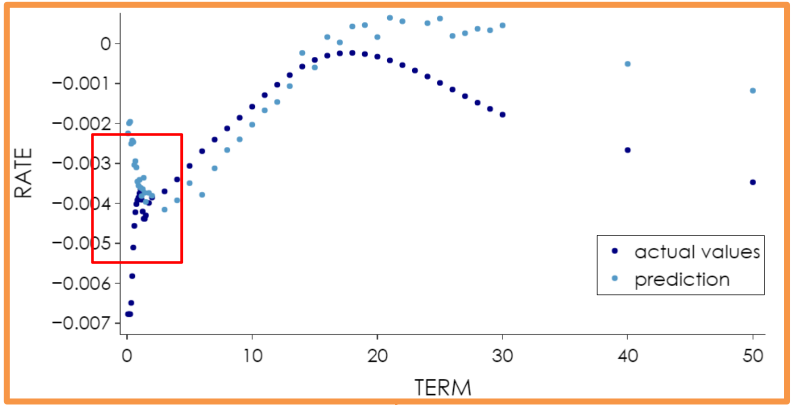
5: Neural Networks

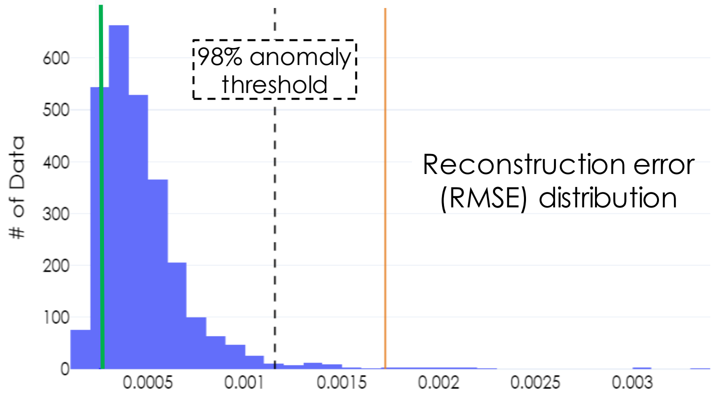
Results: curve reconstruction example

Normal curve (20 July 2018)



Anomalous curve (12th March 2020)





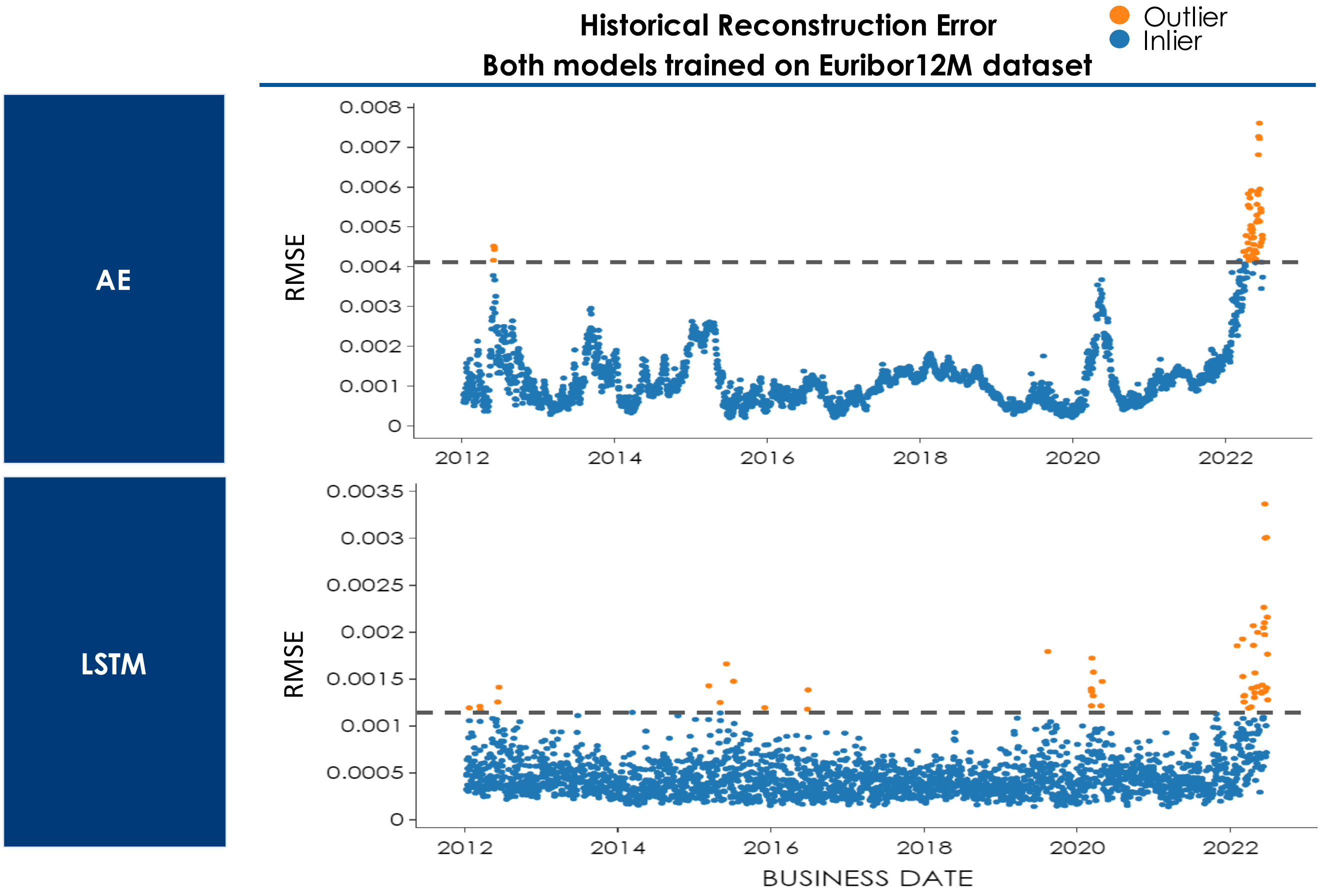
Remarks

- Optimized AE by multiple runs with different hyperparameters' combinations and nested cross-validation. Anomalous threshold = 98%, $k = 5$.
- Is normal really normal: **YES**, very common yield curve shape in the dataset.
- Is anomalous really wrong? **NO**, the «fold» in the yield curve shape around 1Y pillar is due to well-known bootstrapping details.
- **Lesson: human check post ML processing is crucial**

5: Neural Networks

Results with AE and LSTM: Euribor12M curve

Historical Reconstruction Error
Both models trained on Euribor12M dataset



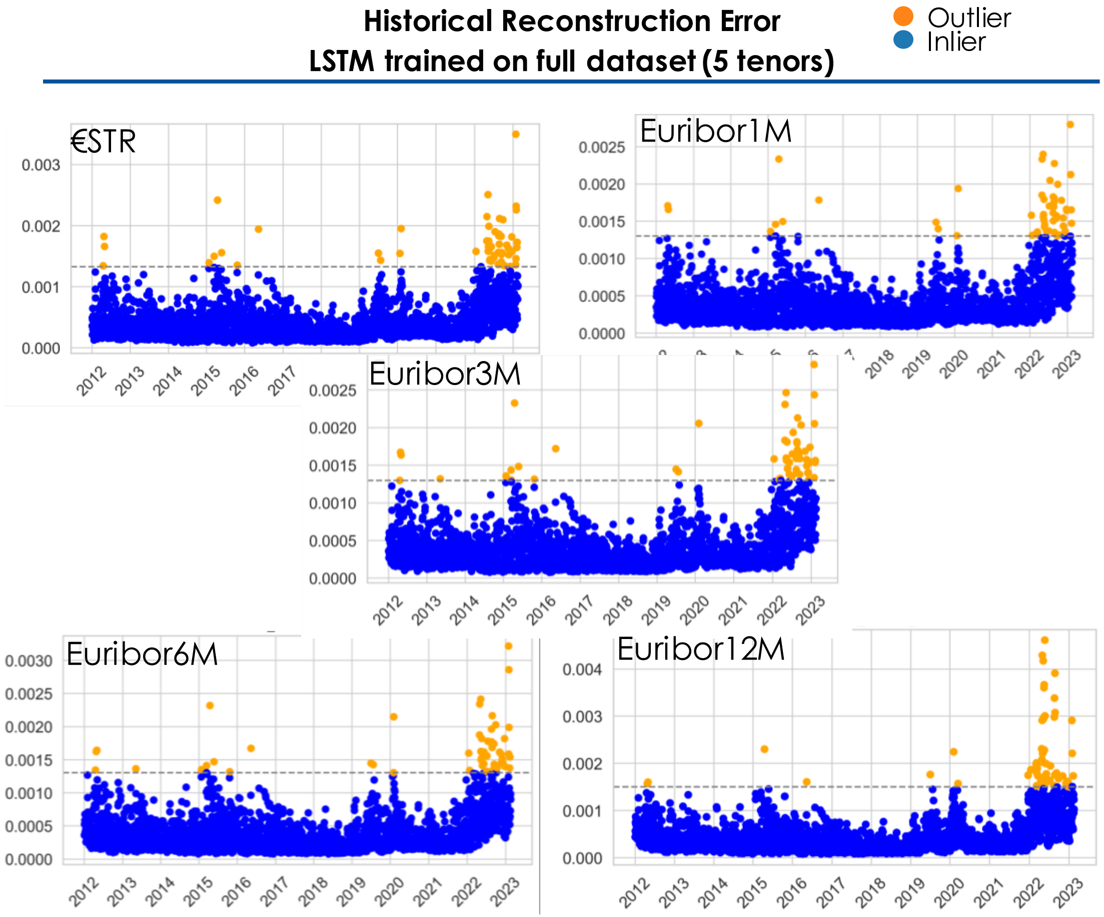
Remarks

- **AE** displays **larger reconstruction errors** than LSTM;
- AE is **highly sensitive to trends in data**, in fact essentially all anomalies are detected in the last period of sharp rate increase due to ECB monetary policy.
- **LSTM** displays **smaller reconstruction errors** than AE (see vertical scale),
- **Anomalus clusters are found in different periods**, thanks to the additional information provided by the $k = 5$ curves pack feeding.

5: Neural Networks

Results with LSTM, all EUR yield curves

Historical Reconstruction Error
LSTM trained on full dataset (5 tenors)



LSTM

Remarks

- LSTM trained to **full dataset**, including all EUR yield curves with 5 tenors.
- **Five anomaly thresholds** computed after the training separately for each tenor (at 98% of each RMSE distribution).
- As a consequence, **anomalies are detected looking at the whole behaviours of all 5 yield curves**, i.e. taking into account the corresponding basis spreads.
- Overall behaviours are **comparable**, since yield curve basis are less volatile than rates themselves, i.e. most of the times yield curves **move parallel to each other**.

5: Neural Networks

Dataset #2: Swaption volatility cubes

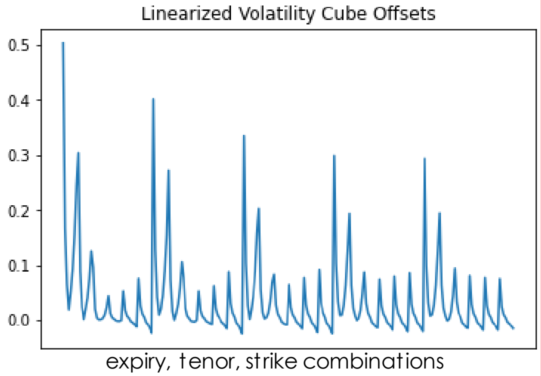
The data set is constituted by an historical series of $N = 614$ $\{\Sigma_1, \Sigma_2, \dots, \Sigma_N\}$ Black-implied **Interest Rate Swaption volatility cubes** in the range 2020-2022. Cube dimensions are Swaptions' **tenor** (length of the underlying Swap), **expiry** (exercise date) and **strike** (as offset w.r.t. to the at-the-money strike). Each cube contains **240 points**.

$$\Sigma_1 = \{\sigma^1_{i,j,k} - \sigma^1_{i,j,0}\}$$
$$\Sigma_2 = \{\sigma^2_{i,j,k} - \sigma^2_{i,j,0}\}$$
$$\dots$$
$$\Sigma_N = \{\sigma^N_{i,j,k} - \sigma^N_{i,j,0}\}$$

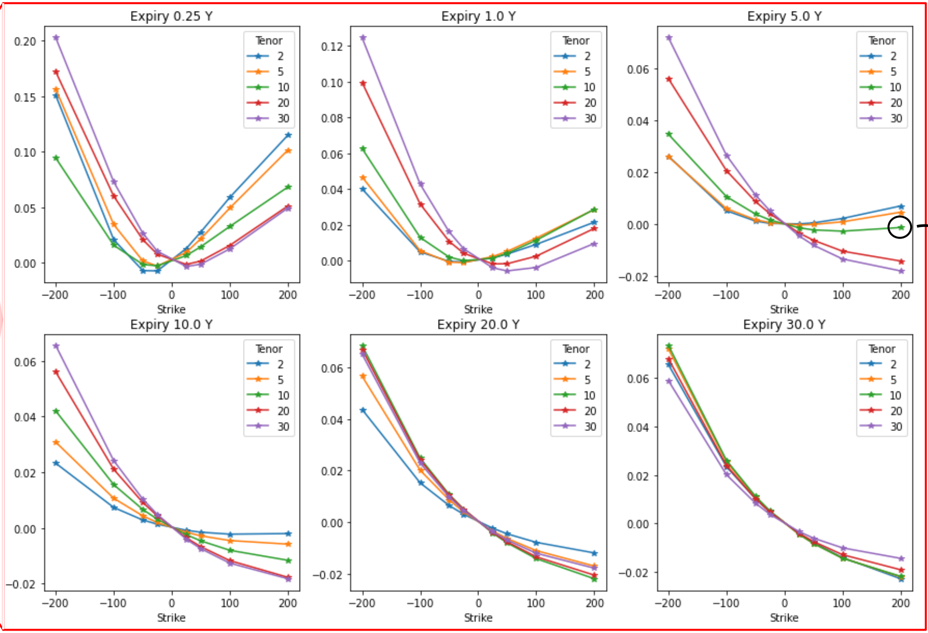
$\sigma^n_{i,j,k}$: volatility, k=0 ATM
$(\#\Sigma_i = \mathbf{240})$

$i \in \{3M, 1Y, 5Y, 10Y, 20Y, 30Y\} \rightarrow$ underlying swap **tenors**
$j \in \{2Y, 5Y, 10Y, 20Y, 30Y\} \rightarrow$ option **expiries**
$k \in \{-200, -100, -50, -25, 0, 25, 50, 100, 200\} \rightarrow$ **strike** shifts (basis points w.r.t ATM)

$\Sigma_i \rightarrow$



Volatility Cube as of 2nd April 2021

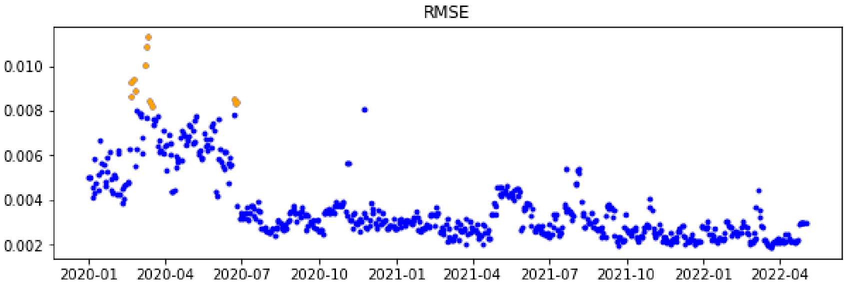


$\sigma^i_{10Y,5Y,200}$
$-\sigma^i_{10Y,5Y,0}$

5: Neural Networks

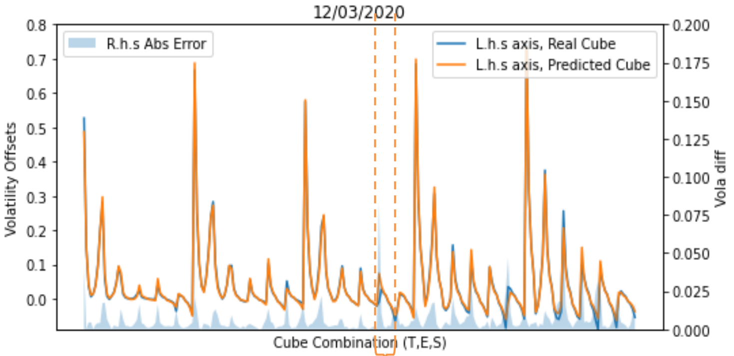
Dataset #2: AE results

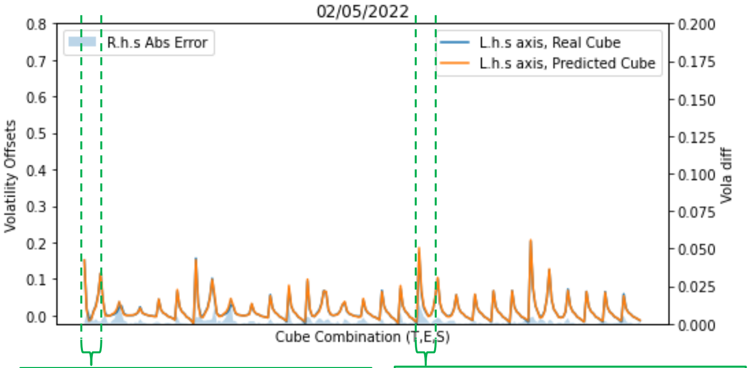
Outlier
Inlier



Remarks

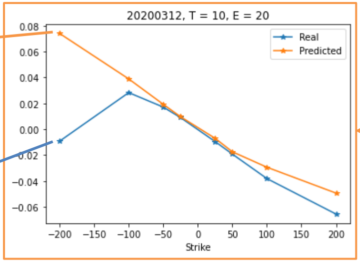
- Top left: RMSEs full cube. Each cube contains **240 points**.
- The approach is **scalable** to data sets with larger dimensions.
- The worst anomalous cube (12/3/2020) effectively contains **wrong data** $\sigma^{20200312}_{10Y,20Y,-200}$ (clearly non-arbitrage free) which was correctly spotted by the AE.





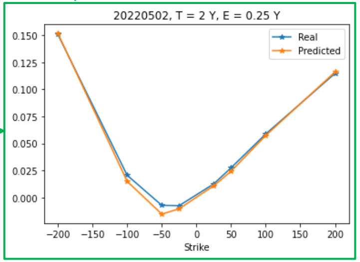
$\sigma^{20200312}_{10Y,20Y,-200} - \sigma^{20200312}_{10Y,20Y,0}$
predicted data

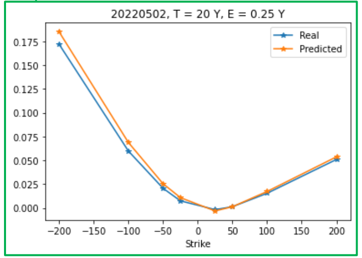
$\sigma^{20200312}_{10Y,20Y,-200} - \sigma^{20200312}_{10Y,20Y,0}$
real data



Anomalous cube (12 Mar. 2020)

Normal cube (2 May 2022)



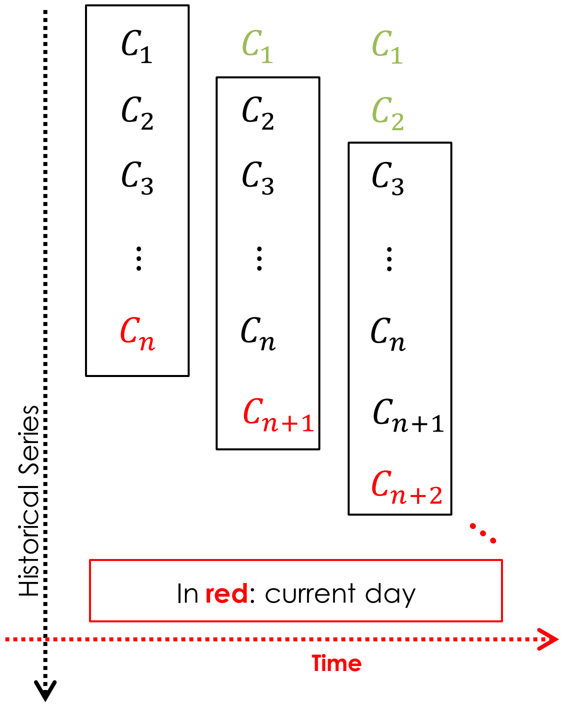


6: Retraining Strategies

Introduction

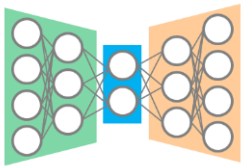
Algorithms trained on historical datasets subject to frequent updates my be trained and retrained in different ways. Since training neural networks is expensive, one typically resorts to training strategies. In this section we refer to Autoencoders, but the generalization to other models is straighforward.

Dataset representation across time

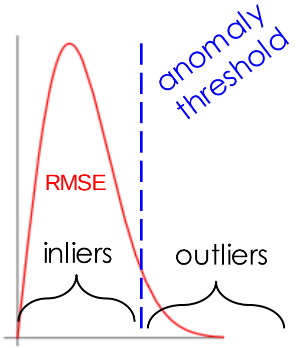


Anomaly detection steps

Model definition and training
Definition of neural network's architecture and hyperparameters' calibration



Threshold definition
A percentile of the RMSE distribution (reconstruction errors)



Anomaly detection test
A data is anomalous if its corresponding RMSE exceedes the threshold

6: Training Strategies

Introduction

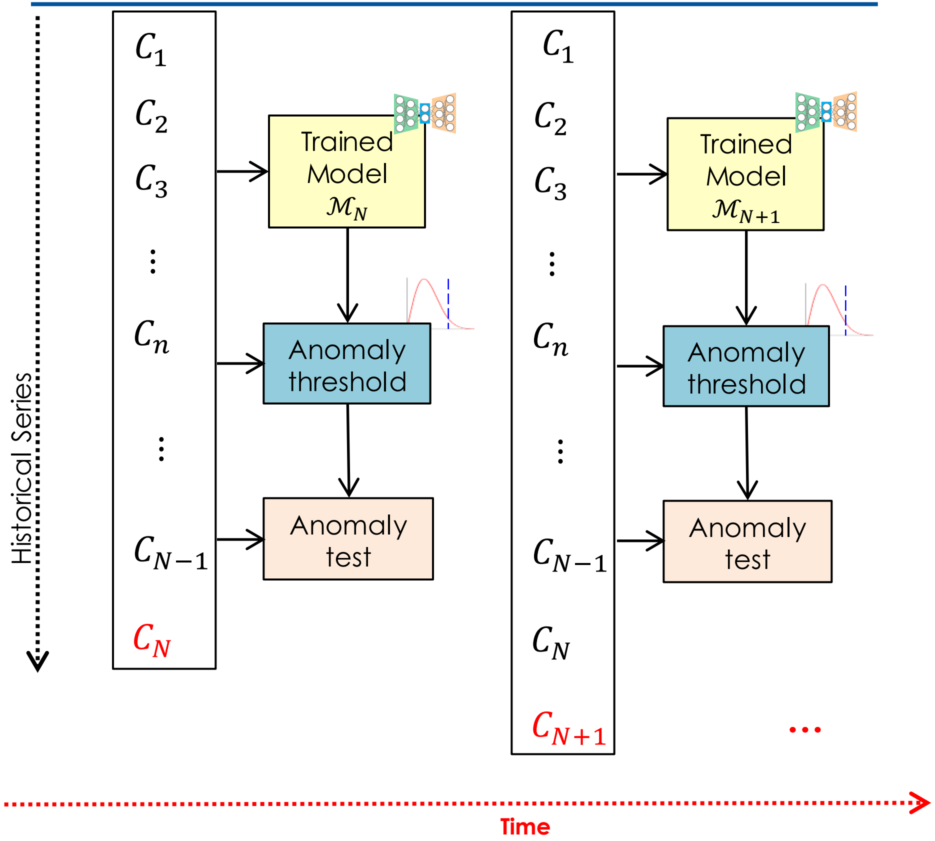
This table summarizes the different training schemes descripted in the following slides.

Training Scheme	Description
Static	Classic Static method where the NN is trained using all the data and all of them are used to evaluete the anomaly threshold.
Sliding Window (SW)	The sliding window consist of a sequence of n element that are updated over time. Using them is particularly advantageous as they can capture the local expected behaviour and limit the processing time of algorithms. This enables the analysis to be limited to an interval over the entire set of collected data, and new data points can be periodically added to the window while older ones become less relevant and are discarded. SWAD is particularly useful in situations where anomalies occur over a short period and require real-time detection, because it allows to use a lower quantity of data and that means a rapid alert in case of anomaly
Periodic Training (PR)	The Periodic Retrain method involves retraining the Neural Network at regular intervals. This approach primarily relies on the dataset's structure and characteristics rather than a mathematical concept. Notably, retraining with step size is a rapid method that requires minimal computation and provides an immediate response from the code.
Triggered Training (TR)	The main purpose of this mechanism is to identify instances when the output of the NN classifier is not suitable and trigger the retraining algorithm accordingly.

6: Training Strategies

Static Scheme

Scheme



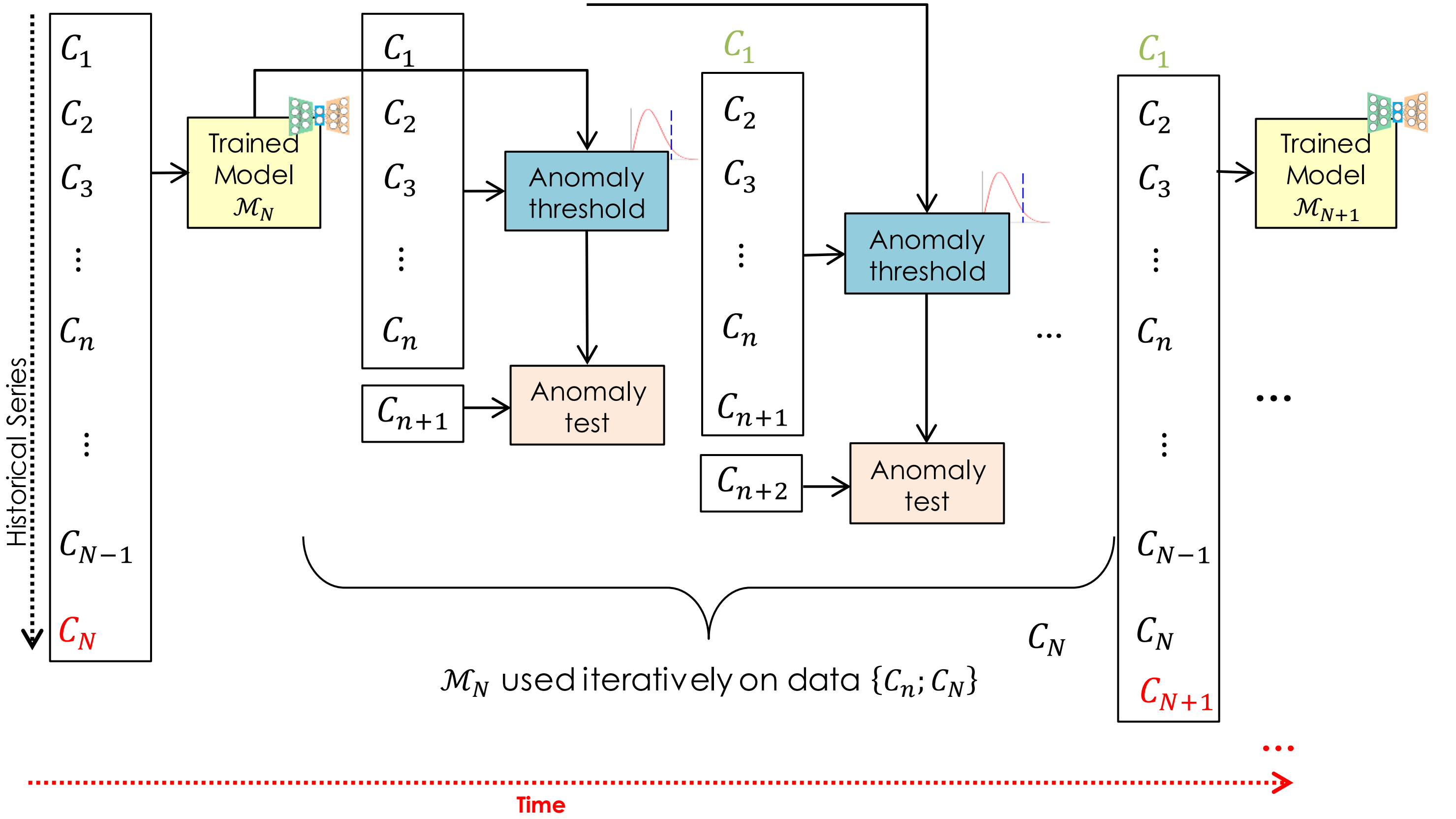
Remarks

- Each day the historical series steps forward, the newest data is added, the oldest data is kept.
- Each day the training is repeated using the whole data set.
- The threshold is updated everyday.
- Depending on the length N, different historical periods, i.e. curve shapes, may be included or not in the analysis, affecting the threshold and the anomalies detected.

6: Training Strategies

Sliding Window Scheme (SW)

Scheme



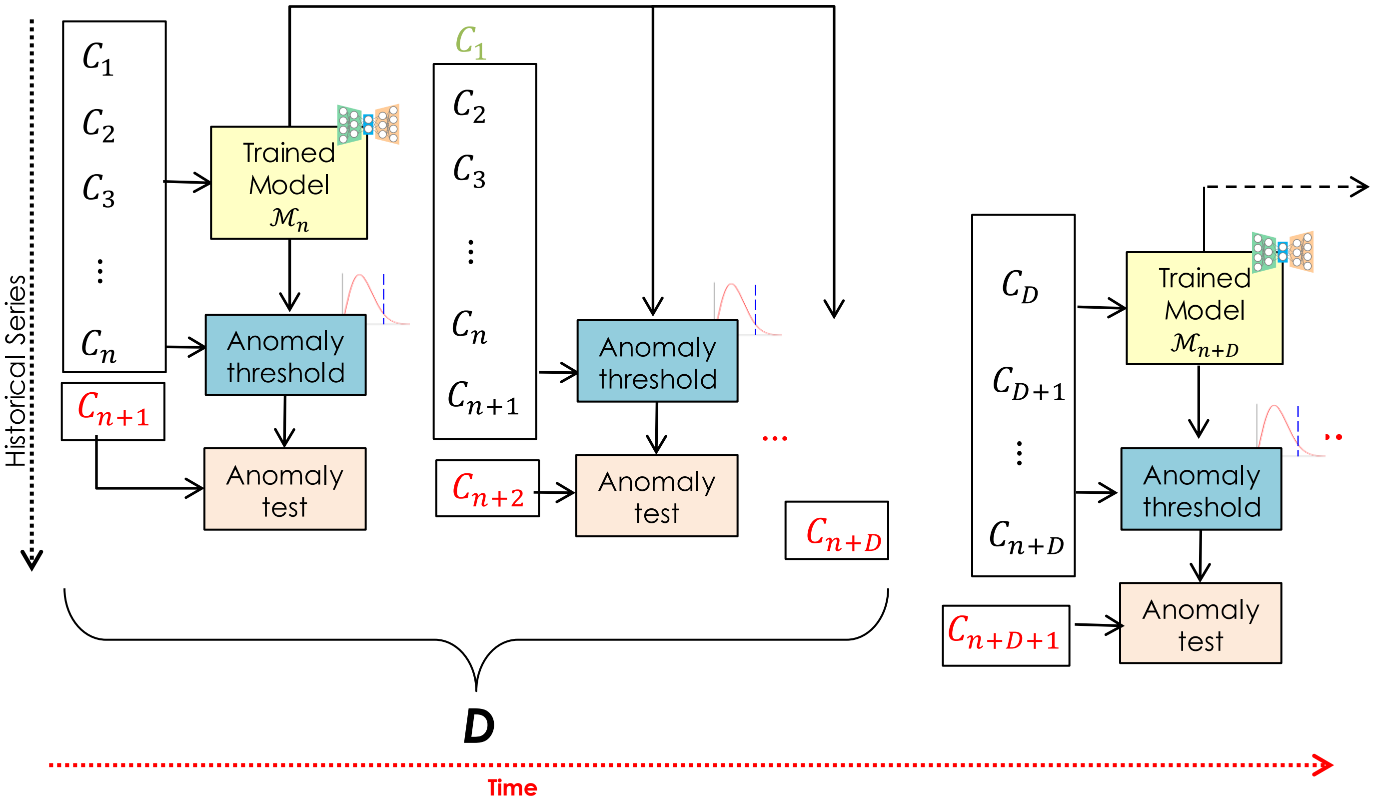
Remarks

- Model $\mathcal{M}_N$ is trained on day N using the whole data set.
- $\mathcal{M}_N$ is used to define the anomaly threshold using the RMSE of the previous $n < N$ data and check the following $(n+1)^{th}$ data. The process is iterated until $n = N - 1$.
- The day after at $N + 1$ the process is repeated, retraining a new model $\mathcal{M}_{N+1}$.

6: Training Strategies

Periodic Retraining scheme (PR)

Scheme



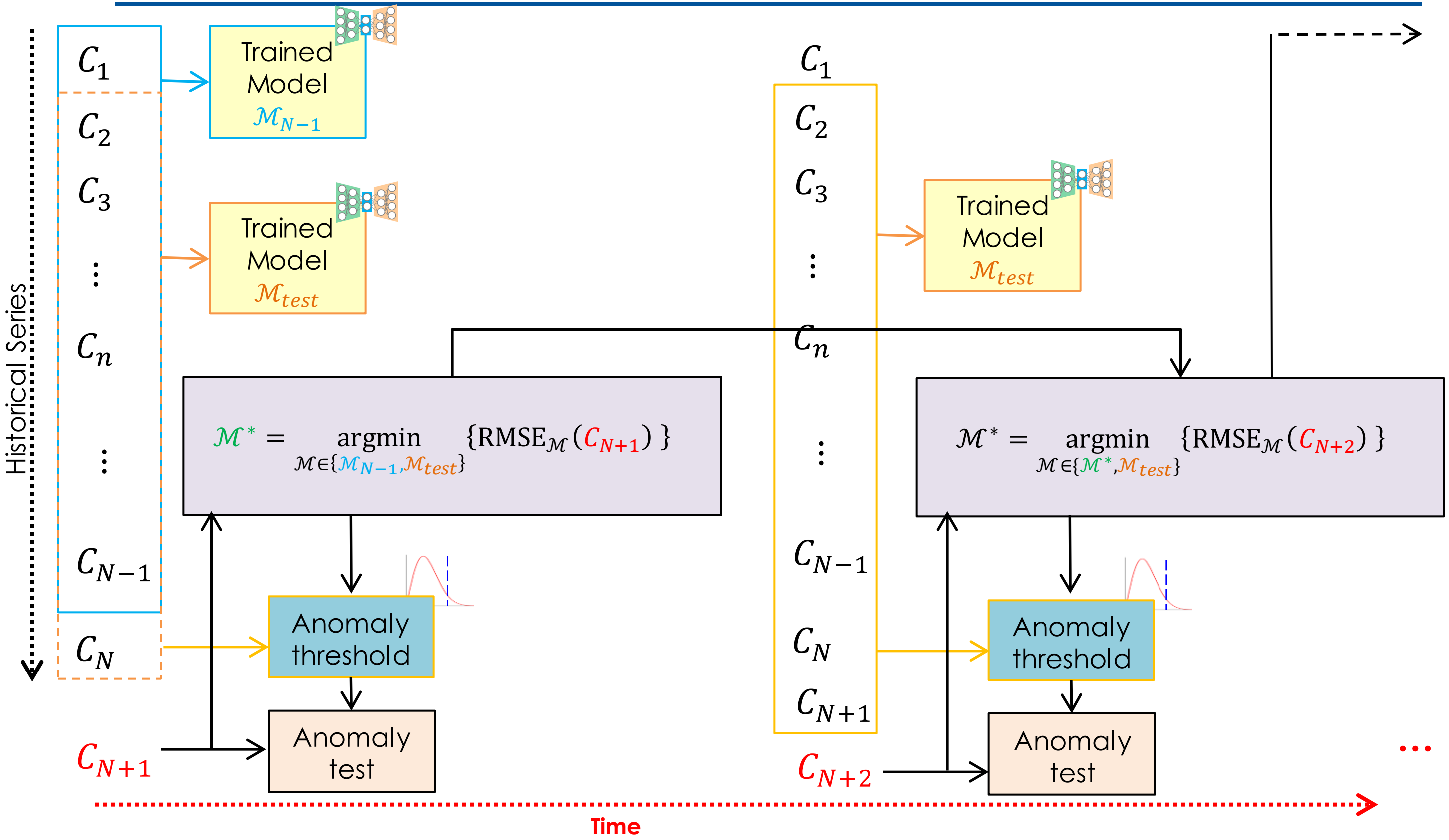
Remarks

- Model $\mathcal{M}_n$ is trained on day $n+1$ using the previous n data.
- $\mathcal{M}_n$ is used to define the anomaly threshold using the RMSE of the previous n data and check the following $(n+1)^{th}$ data. The process is iterated D times until $n+D$ data.
- The day after at $n+D$ the model is retrained. If $D=1$ we have a daily retraining.

6: Training Strategies

Triggered Training scheme (TR)

Scheme



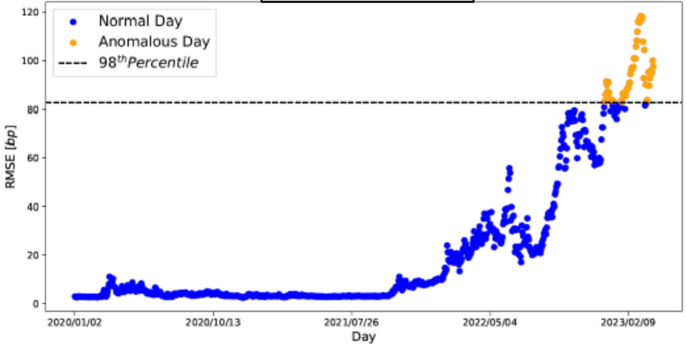
Remarks

- Model $\mathcal{M}_{\text{test}}$ is trained on day N using the previous N data
- Model $\mathcal{M}_{test}$ is compared with the previous model $\mathcal{M}_{N-1}$ and the best model $\mathcal{M}^*$ is selected
- Model $\mathcal{M}^*$ is used to define the anomaly threshold and check the following $(N+1)^{th}$ data.
- The day after the procedure is repeated

6: Training Strategies

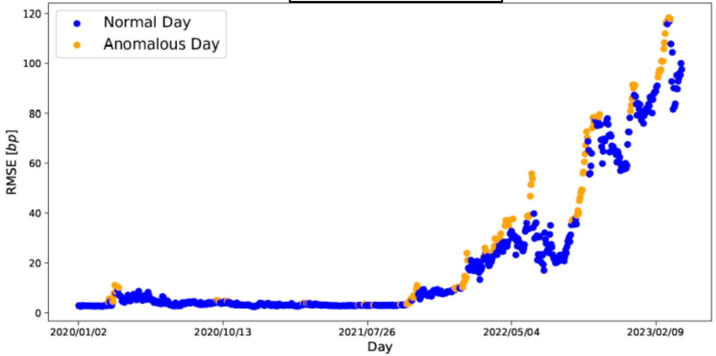
Results with Autoencoder

Static



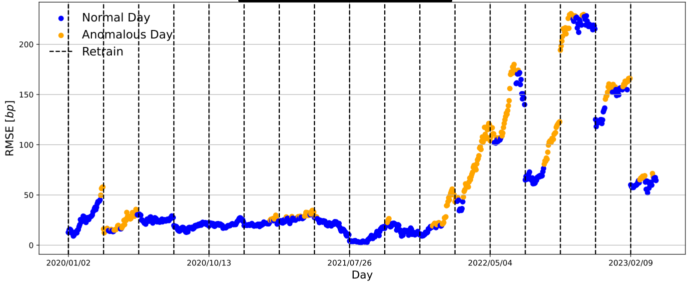
Basic model due to the single threshold: high impact of regime changes.

SW



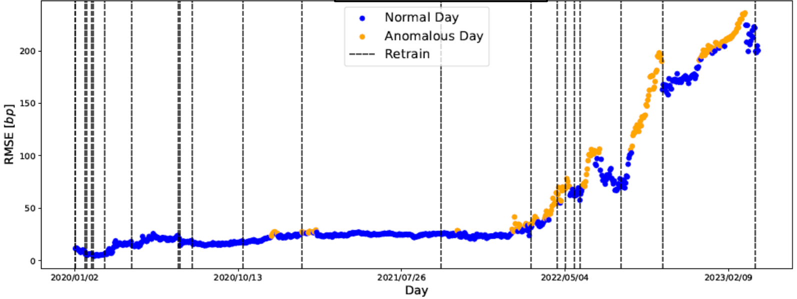
Rolling windows help to find anomalous curves based on the threshold defined using the recent 250 curves.

PR



Every D = 50 days the NN is retrained on the n = 250 previous curves and the RMSE experiences a significant reduction.

TR



Retraining occurs only when the new candidate model outperforms the one adopted so far. RMSE is smooth and retrainings follow regime changes

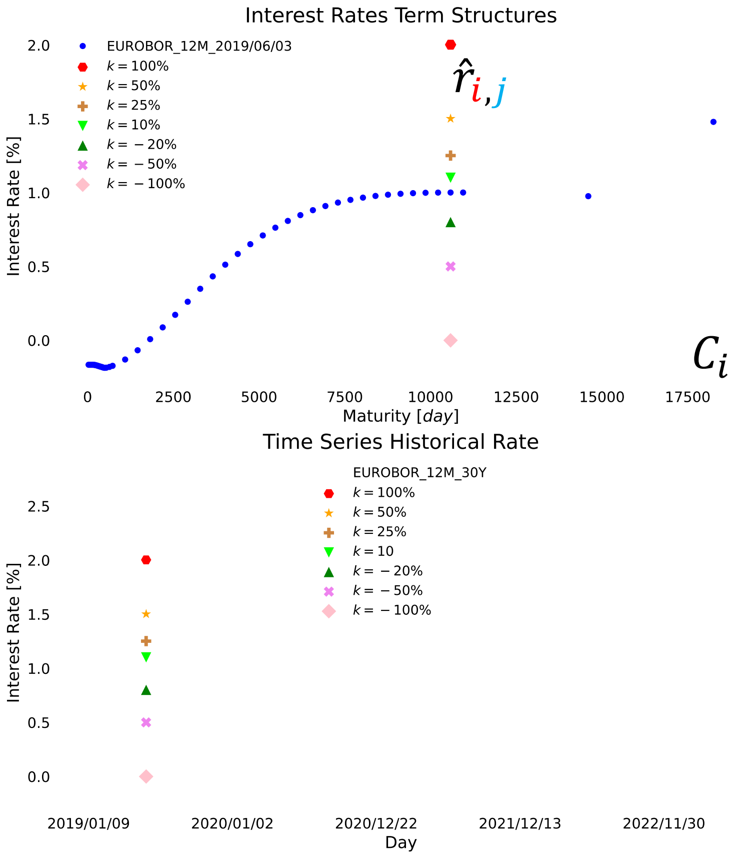
7: Artificial Anomalies

Rationale

- Since we deal with **unsupervised models**, we do not have labelled dataset with truly detected anomalies. Hence, we cannot rely on a priori knowledge of the expected outcomes of anomaly detection to test the models' performances.
- Therefore, we constructed a **labelled dataset** including **artificial anomalies**, to check the **detection performance of the different models.**
- Artificial anomalies may be constructed in several ways. We selected two cases.
    - **Single pillar anomalies**: we inject anomalies as percentage bumps on a single pillar of a randomly selected subset of yield curves.
    - **Group anomalies at curve level**: we inject anomalies sampling from a gaussian distribution centered to each pillar and with standard deviation proportional to the historical standard deviation of the pillar itself.

7: Artificial Anomalies

Single Maturity – fixed *k*-percentage anomalies



Introduce **single maturity anomalies** by replacing the original zero rate value with a corrupted one.

Given the **i-th** curve of the dataset,

$$C_i = \{r_{i,1}, r_{i,2}, \dots, r_{i,50}\},$$

we select a **maturity** j and we **apply a percentage shock** k to obtain

$$\widehat{C}_i = \{r_{i,1}, r_{i,2}, \dots, \hat{r}_{i,j} \dots, r_{i,50}\},$$

where

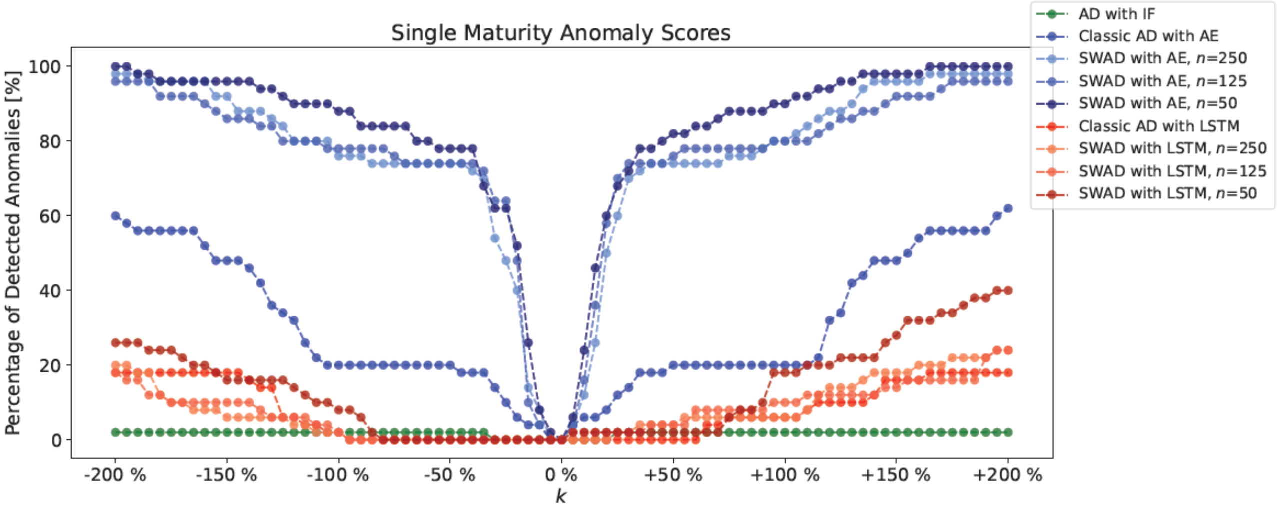
$$\hat{r}_{i,j} = (1 + k)\, r_{i,j}$$

and $k \in \{k_1, \cdots, k_N\}$ ($N = 7$ in the l.h.s. figures). The larger k the larger the anomaly.

7: Artificial Anomalies

Single Maturity – fixed k-percentage anomalies

Fraction of Detected Anomalies



Fraction of detected anomalies (true positives) for 9 different shock sizes k and 9 different combinations of training strategies and neural networks. We injected anomalies on 30Y pillar for 50 randomly selected «normal» curves in the dataset.

Remarks

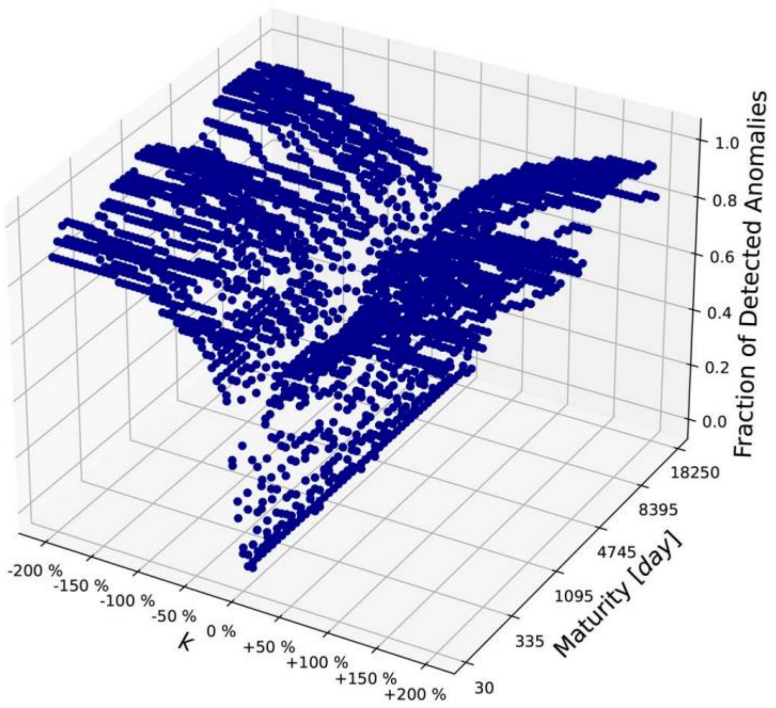
- This approach allows to focus exclusively on **detecting the labeled anomalous curves**, using the fraction of detected anomalies as a measure of models' performance.
- As expected, the fraction of detected anomalies (true positives) $\rightarrow 0$ when the bump size $k \rightarrow 0$.
- Viceversa, the fraction of true positives increses with the bump size.
- **Autoencoders** shows the best performance to identify the artificial anomalies.

7: Artificial Anomalies

Single Maturity – fixed k-percentage anomalies

Fraction of Detected Anomalies

For the best Autoencoder of slide 31: fraction of detected anomalies as a function of bump size k and the **pillar** affected by the injected anomaly.



Remarks

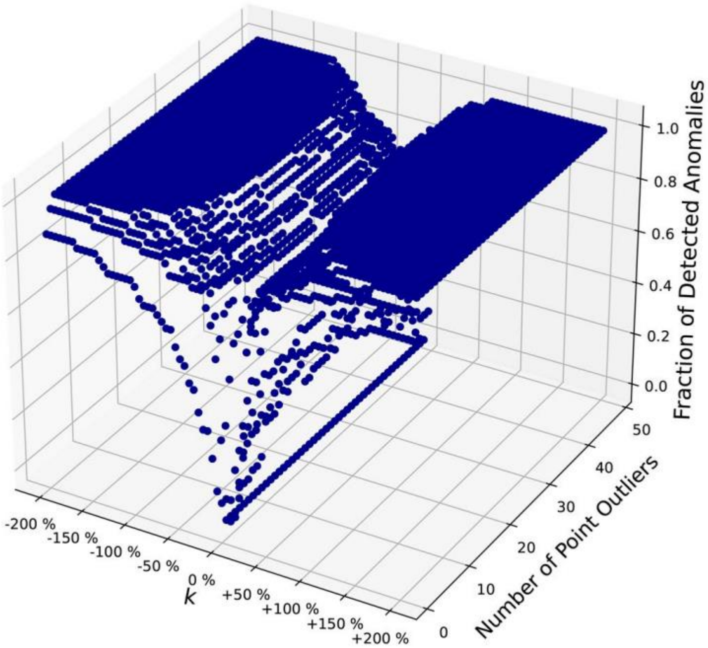
- The symmetrical shape confirms that the behaviour is **indipendent from the corrupted maturity.**

7: Artificial Anomalies

Multiple Maturity – fixed k-percentage anomalies

Fraction of Detected Anomalies

For the best Autoencoder of slide 31: fraction of detected anomalies in function of k and the **number of (randomly selected) pillars** affected by the injected anomaly (from 0 to 50).

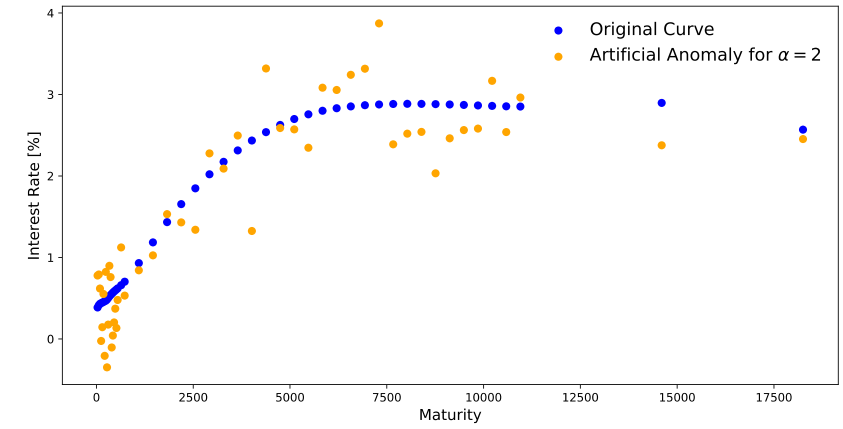


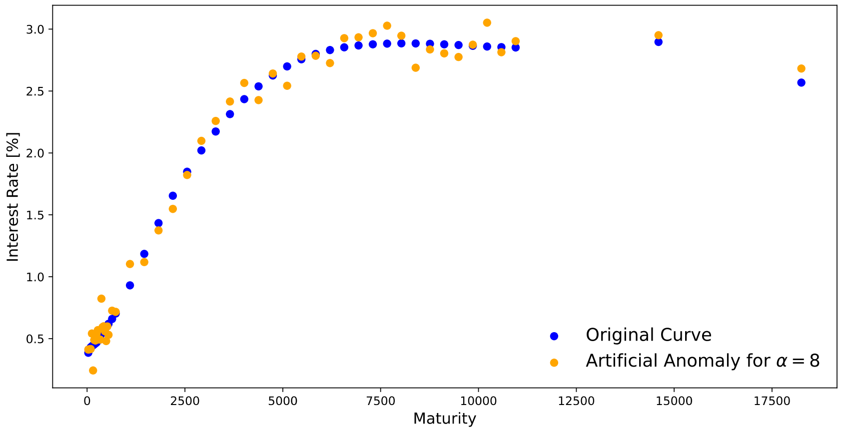
Remarks

- The higher the parameter k, the higher the number of detected anomalies, even with a low number of point outliers.
- When a larger number of point anomalies is introduced (>40) **the model interprets the anomalous curve as a normal one**, since the original curve is just shifted by an amount k.

7: Artificial Anomalies

Group anomalies at curve level





Introduce of a **Group Anomalies** by corrupting the original values **of the whole curve**.

Given the **i-th** curve of the dataset

$$C_i = \{r_{i,1}, r_{i,2}, \dots, r_{i,50}\}$$

we build

$$\widehat{C_i} = \{\hat{r}_{i,1}, \hat{r}_{i,2}, \dots, \hat{r}_{i,50}\}$$

$$\hat{r}_{i,j} = \mathcal{N}\left(r_{i,j}, \frac{\sigma_j}{\alpha}\right) \qquad j = 1, \dots, 50$$

where the **normal distribution** is centered in each zero rate value with historical standard deviation

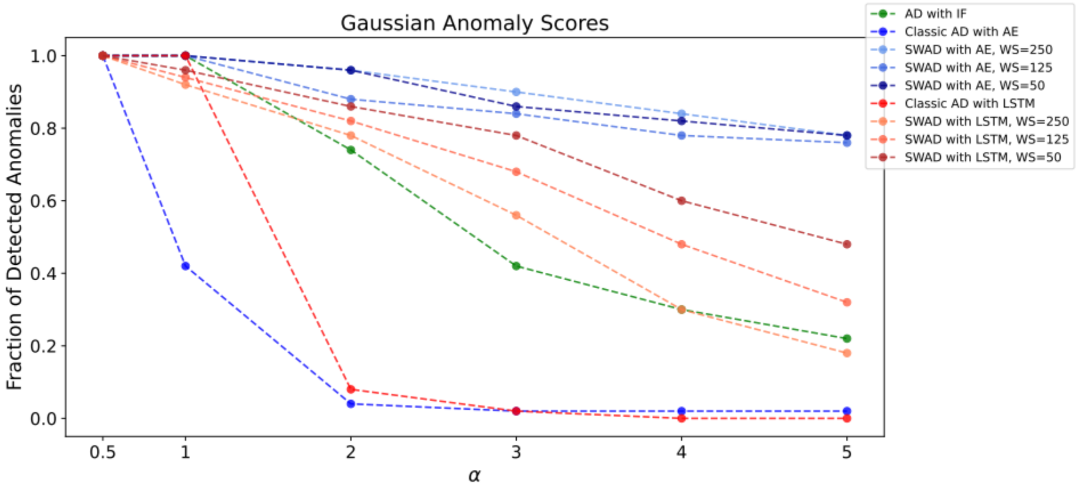
$$\sigma_j = \mathrm{Std}(r_{1,j}, r_{2,j}, \dots, r_{N,j})$$

and α is a control parameter. The larger α the smaller the anomaly.

7: Artificial Anomalies

Curve level anomalies

Fraction of Detected Anomalies



We injected anomalies in **50 randomly selected «normal» curves** and we tested for different models and different schemes the percentage of the 50 curves that the models detect as anomalous for different values of α.

Remarks

- As before, this approach enables to focus exclusively on detecting the labeled anomaly curves, using the fraction of detected anomalies as a metric to evaluate the performance of our models.
- Clearly, the smaller/higher the parameter α the lhigher/smaller the fraction of detected anomalies.
- **Autoencoder** is the best in class model to identify the artificial anomalies

8: Conclusions

- We applied different **unsupervised machine learning** techniques in the context of **market data anomaly detection.** Our framework is general and applicable to any kind of market data (in terms of asset class, dimension,…). We tested two different data sets: **interest rate curves** (1-dimensional) and Swaptions' **implied volatilities** (3-dimensional).
- **Isolation Forests,** which looks separately at **single pillars**, were found to work better when **multiple features**, i.e. level, slope and curvature, are used.
- **Neural networks** allow to take into account whole curve shapes. We found that, in case of long historical series, **LSTM** works better that **AE**, since it considers the information carried by the most recent data, and is able to distinguish among **different regimes**.
- We developed and tested **different approaches** to use Neural Networks in **real time**, comparing daily vs periodically **retraining strategies**, also where the model's update is triggered by decreasing reconstruction ability.
- Finally, we tested the models' robustness with **artificial anomalies**, finding that Autoencoders show better results w.r.t. LSTM when shorter historical series are used.

Thanks for the attention!

9: References

- BCBS, Basel Committee on Banking Supervision, https://www.bis.org/basel_framework/
- 2022, Sokol, Alexander, Autoencoder Market Models for Interest Rates, https://ssrn.com/abstract=4300756
- 2022 – AIFIRM - Big Data & Advanced Analytics per il Risk Management
- 2020 – Mueller, Roeder - Anomaly Detection in Market Data Structures via Machine Learning
- 2019 – Zhu, Chan, Bright, Applying Machine Learning for Troubleshooting Credit Exposure and xVA Profiles
- 2018 – Kondratyev, Learning Curve Dynamics with Articial Neural Networks https://ssrn.com/abstract=3404863
- 2018 – Ruff et al. - A Unifying Review of Deep Shallow Anomaly Detection
- 2017 – Brummelhuis, Luo, CDS Rate Construction Methods by Machine Learning Techniques https://ssrn.com/abstract=2967184