Bond Liquidity with AI

MODELLING BID-ASK SPREADS USING DEEP LEARNING

Presentation at Budapest University of Technology and Economics

László Arany, PhD (Risk Management and Liquidity Core Research, MSCI)

December 8, 2020

What is Liquidity Research about?

MSCI

Agenda

1

Bond Markets and Bid-Ask Spreads

Why do we care about the bid-ask spread?

2

Data

What kind of data did we use and how did we pre-process it?

3

Models

What kind of models did we build and why?

4

Evaluation

How do you know which of two models is better?

5

Results

What can you expect from a deep learning model on a regression task?

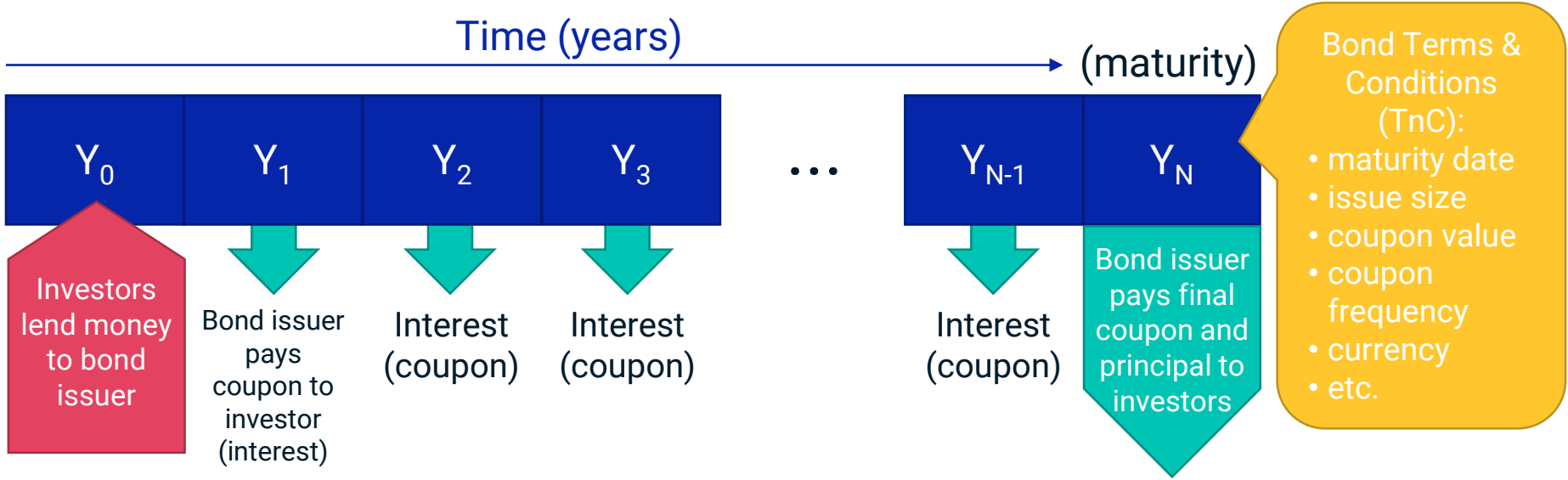
MSCI

Bond Markets and Bid-Ask Spreads

WHY DO WE CARE ABOUT THE BID-ASK SPREAD?

What is a (corporate) bond?

- Debt: the corporation (bond issuer) borrows money from investors (bond holders) and pays it back with interest.

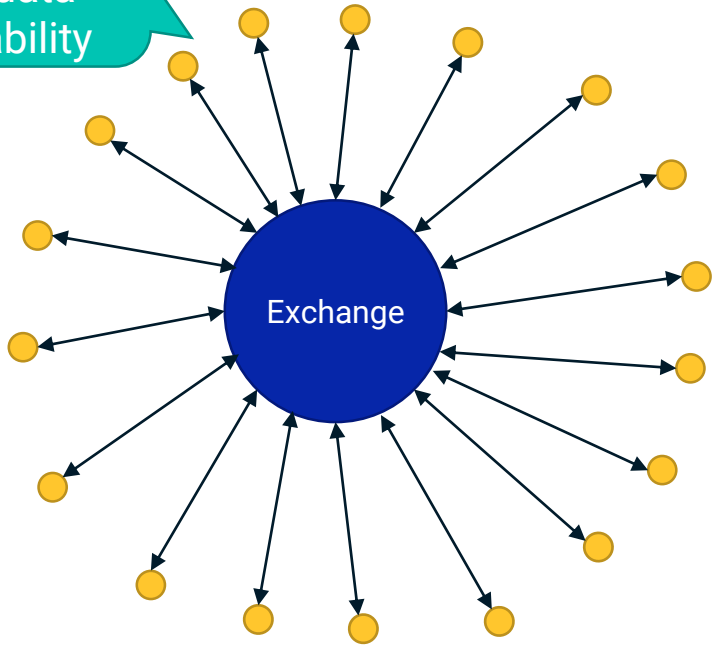


- **A bond can be sold during its term or held to maturity.**

Exchanges vs dealer markets

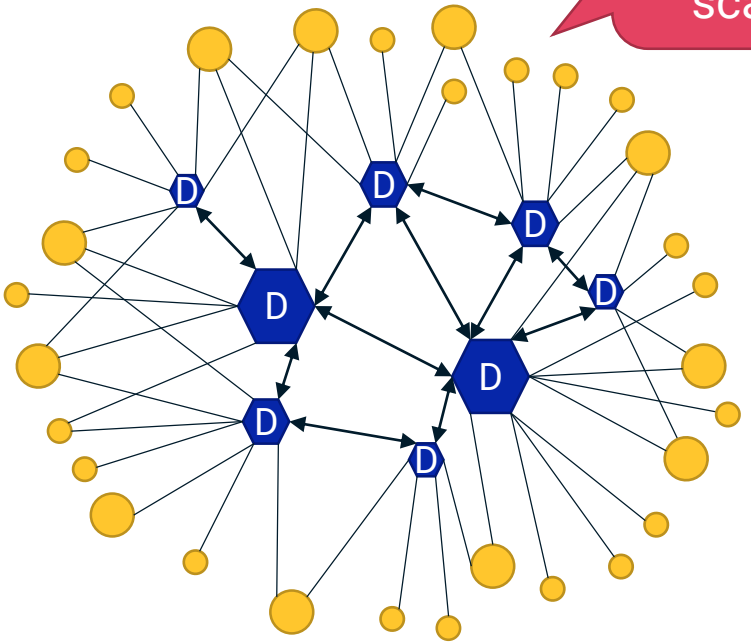
Higher transparency and data availability

Exchange markets



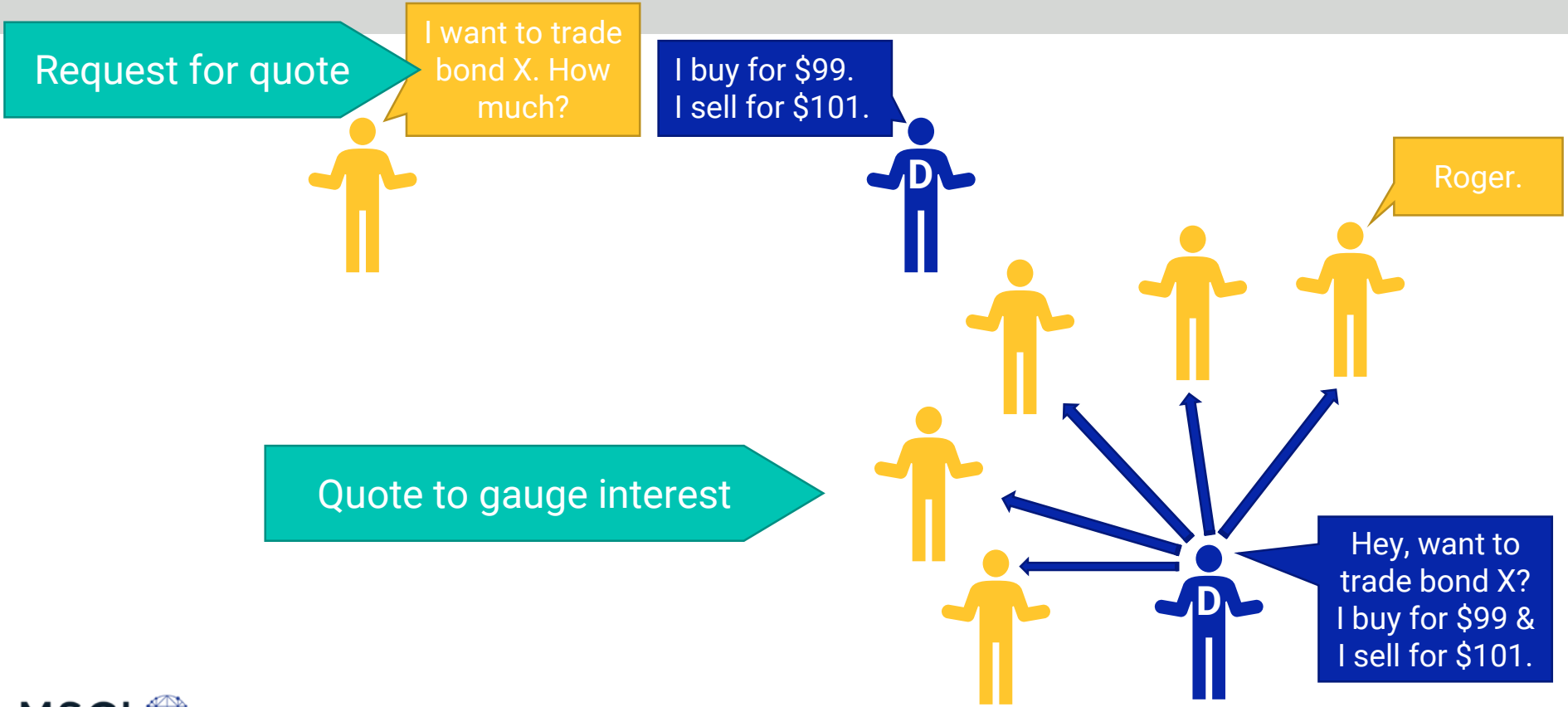
Dealer markets

Opaque and data is scarce!

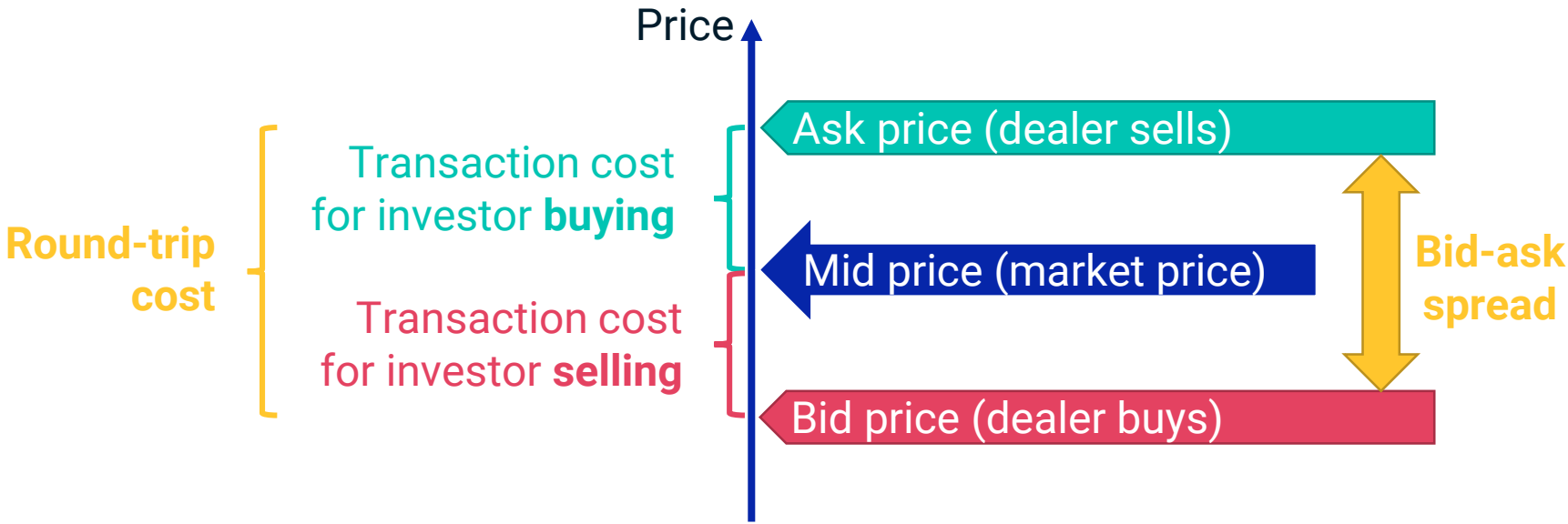


MSCI

Trading in dealer markets – what are quotes?



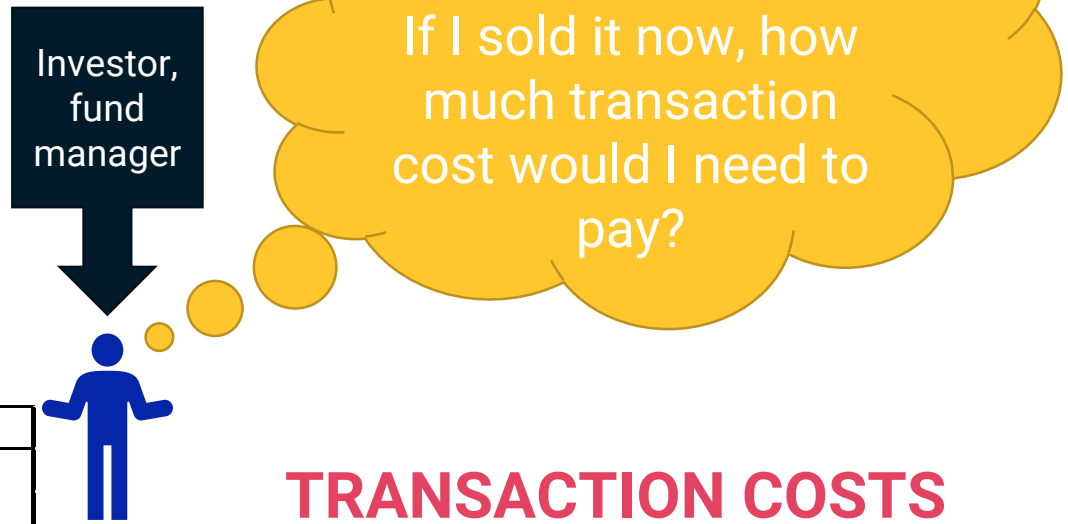
How dealers make money: the bid-ask spread



Why do investors care about bid-ask spreads?

- Investor buys bond with bid-ask spread of 1% (or 100bps) → 50 bps (0.5%) cost
- Price of the bond increases by 2% → Expect to sell at profit = 1%.
- But what if the bid-ask spread increases to 250 bps?
- Profit 0.2% instead of 1%.

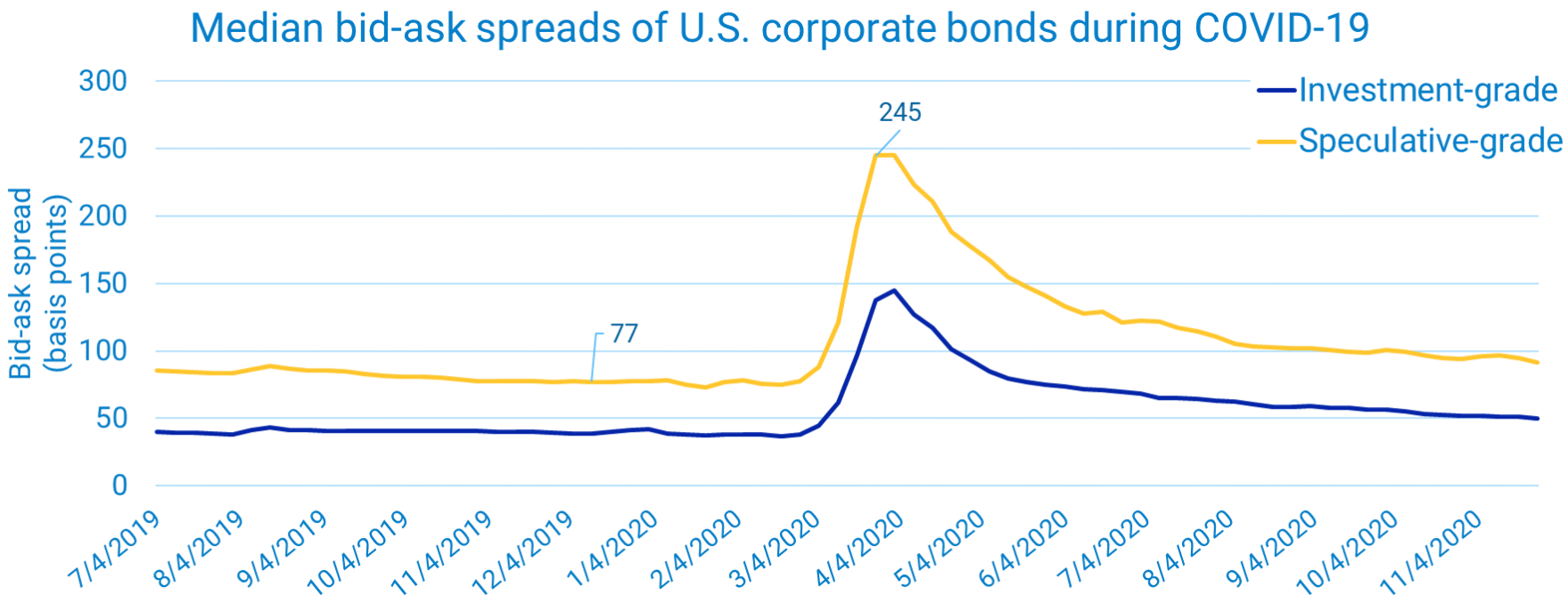
	T0	T1 expected	T1 actual
Bid-ask spread	100 bps	100 bps	250 bps
Ask price	100.5	102.51	103.275
Mid price	100	102	102
Bid price	99.5	101.49	100.725
Profit		1.0%	0.2%



TRANSACTION COSTS EAT INTO PROFITS!

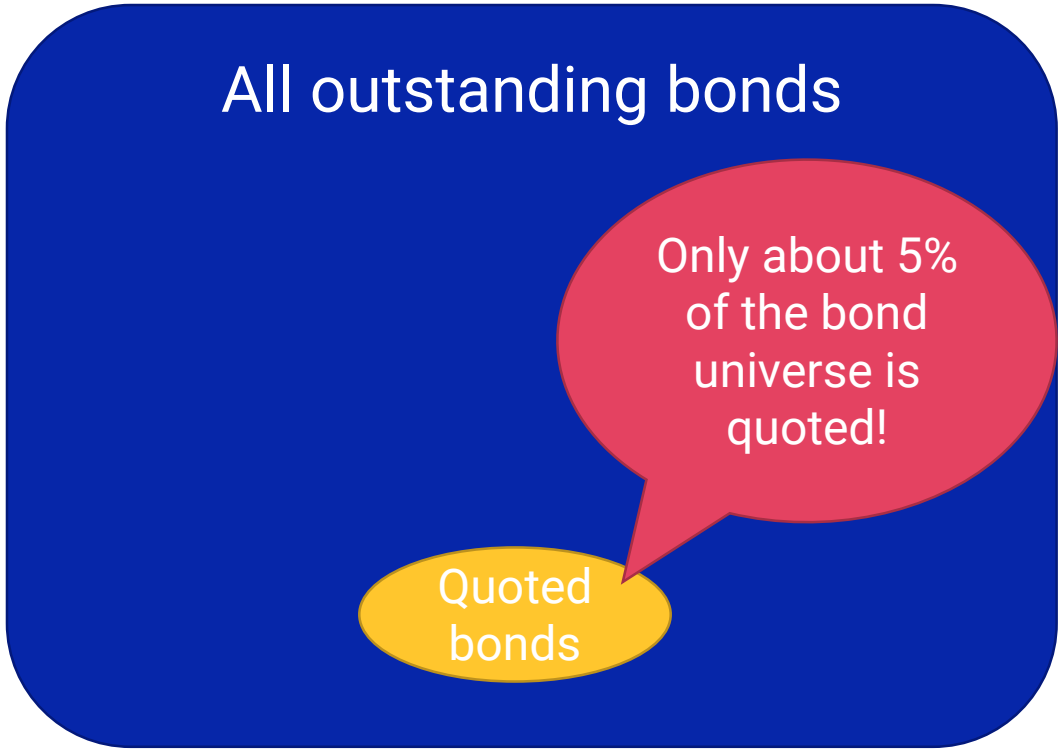
MSCI

Liquidity during the COVID-19 crisis

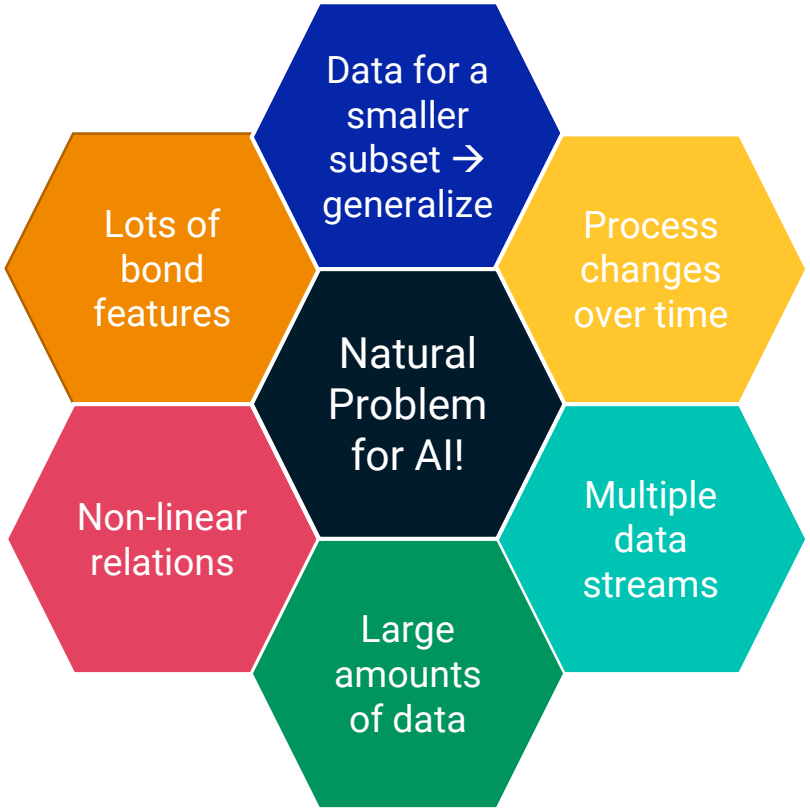


So how do we know the bid-ask spread?

- Trade reporting (delayed, difficult)
- Request quotes from dealers
- Dealers send out quotes to gauge interest → can be parsed
- Not all bonds are quoted/traded, only a fraction
- Many bonds are in buy-and-hold portfolios → held to maturity
- Similar bonds tend to have similar liquidity → we can use quoted bonds to predict bid-ask spreads of non-quoted bonds

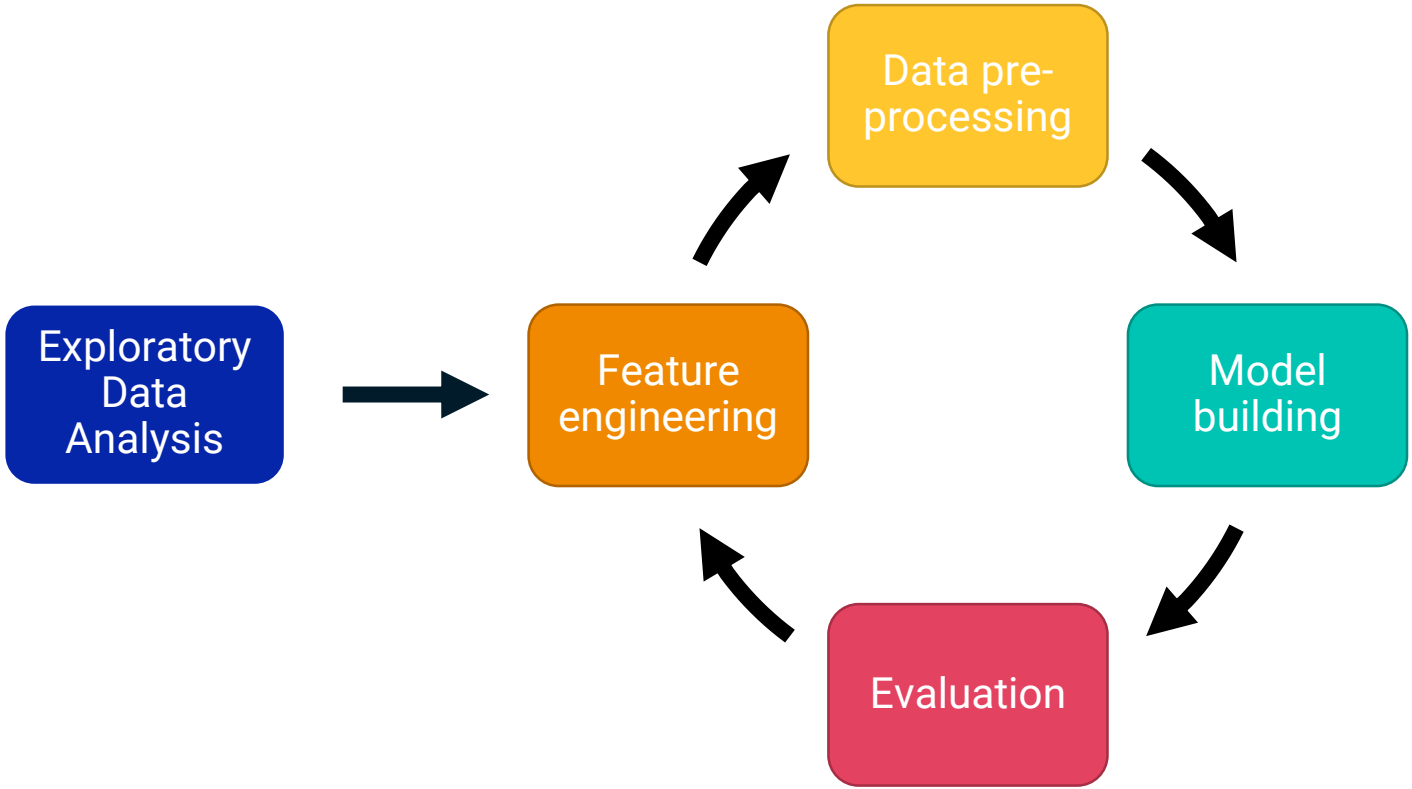


A natural problem for AI!



MSCI

The modelling process



MSCI

Data

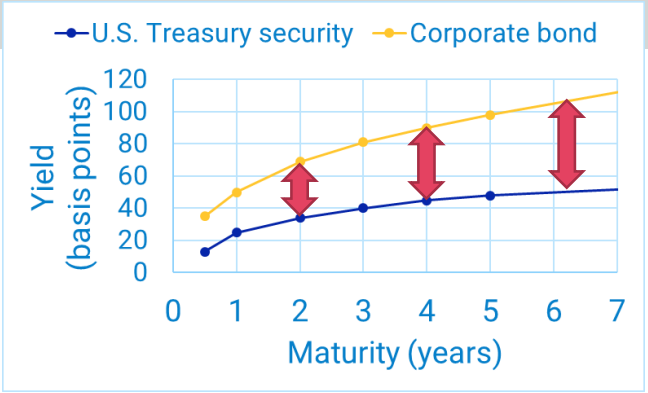
WHAT KIND OF DATA DID WE USE AND HOW DID WE PRE-PROCESS IT?

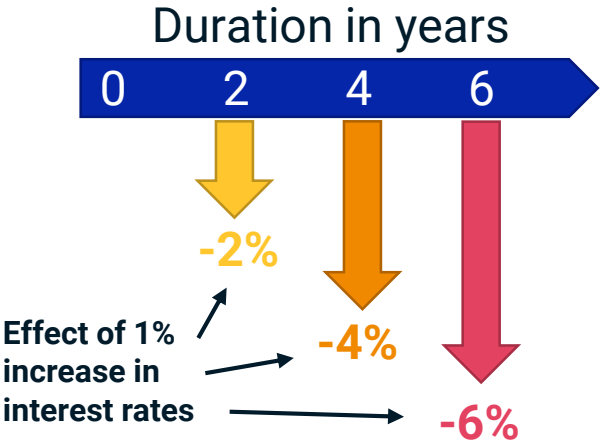
The “label”

- A single positive continuous number: a regression problem

The “features”

- Bond meta data **100+ features**
  - coupon and other interest terms, bond age, outstanding amount, region and country, sector and industry, bond rating, etc.
- Credit spread **single number + time history**
  - a measure of the credit quality of the company
  - ~ how much more interest (%) the corporate bond must pay than government bonds (calculated from market price)
  - depends on probability of default (failing to make payments)
  - depends on recovery (% of bond price investors will get if the company’s assets are liquidated due to default)
- Duration **single number + time history**
  - a measure of bond price sensitivity to interest rate changes
  - ~ % change in bond price due to an interest rate change of 1%
  - depends on time to maturity, interest terms, etc.

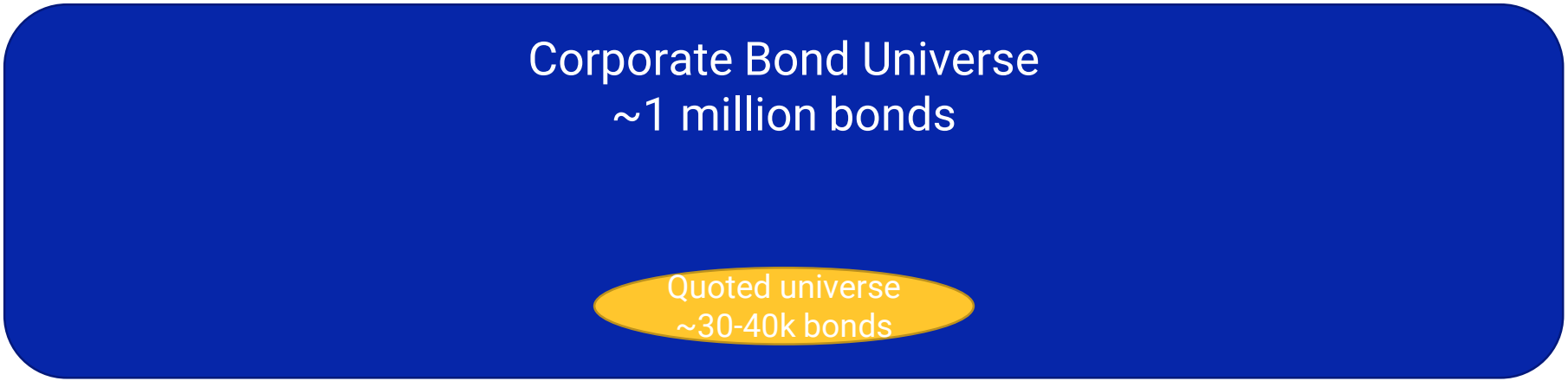




MSCI

The amount of data

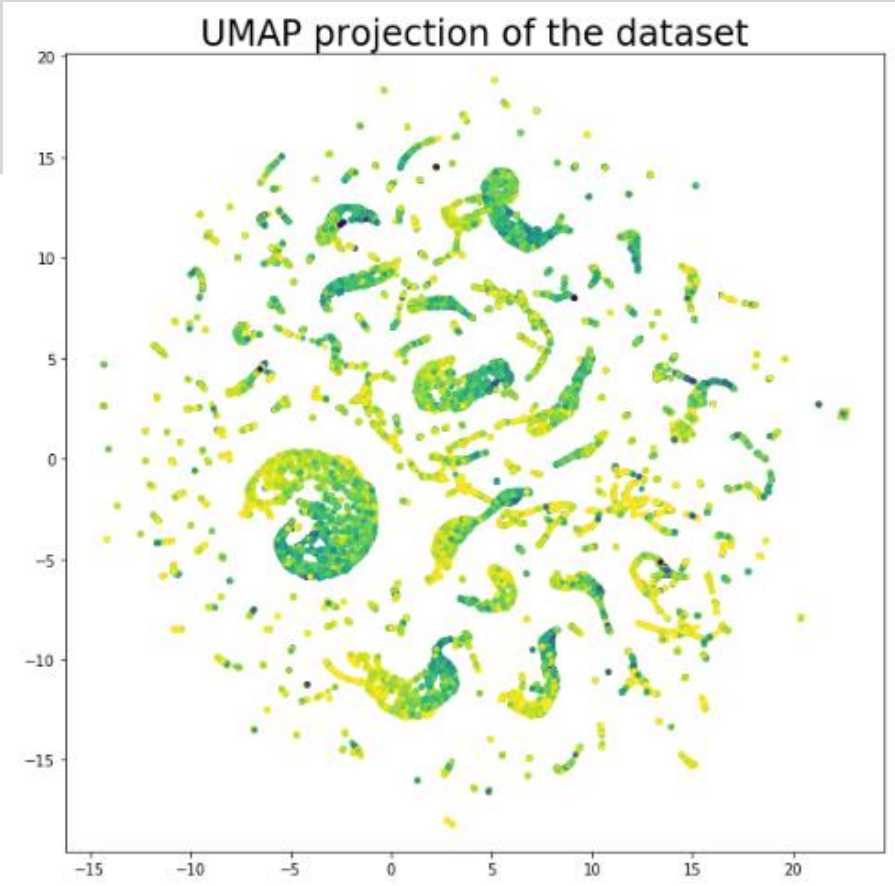
- ~30-40k quotes per day for corporate bonds – “labelled” data
- ~500k with all calculated features (credit spread, duration, etc) – “unlabelled” data
  - requires market price!
- ~100-150 columns (“features”)
- ~8-10 GB of data per year



Data pre-processing

- Add calculated features
- Add derivatives for time history features
- Outlier filtering
- Caps & floors (Winsorizing)
- Feature scaling
  - Standardising, min-max scaling, etc
- Handling missing values
- Train-valid-test split (60/20/20)

- Exploratory Data Analysis (EDA) and data cleaning / pre-processing is about **70%** of a machine learning project!



Uniform Manifold Approximation and Projection for Dimension Reduction

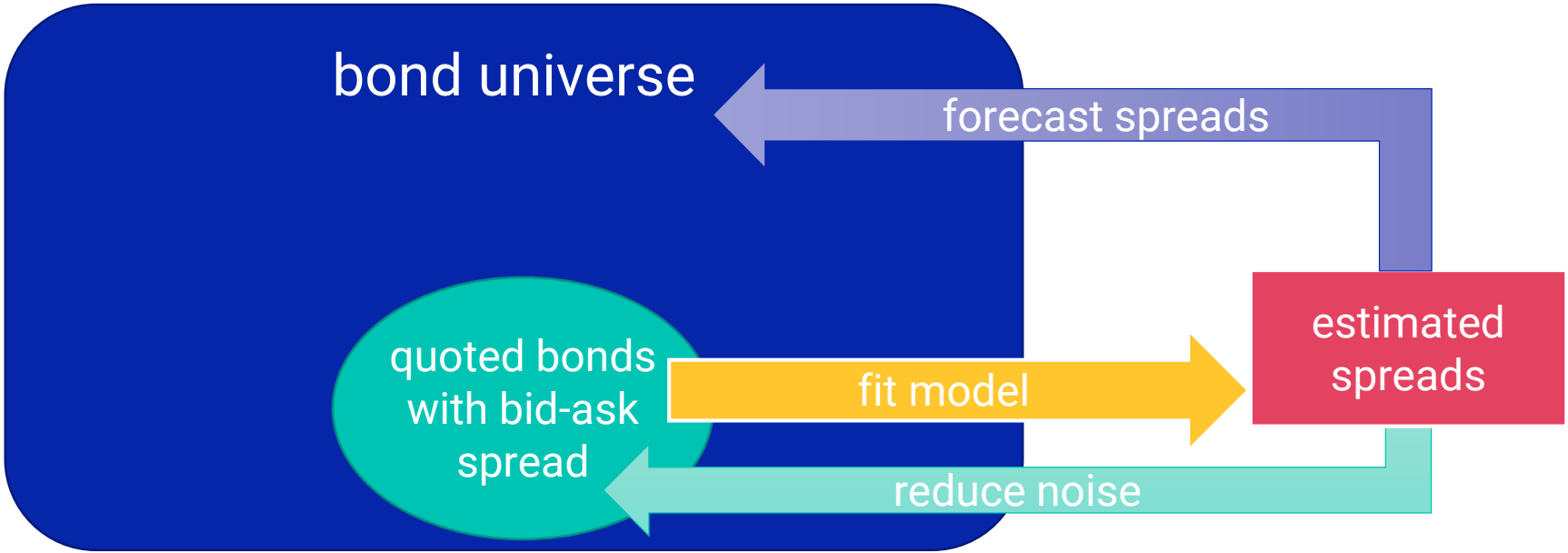
MSCI

MSCI

Models

WHAT KIND OF MODELS DID WE BUILD AND WHY?

What is the modelling task?



Baseline models

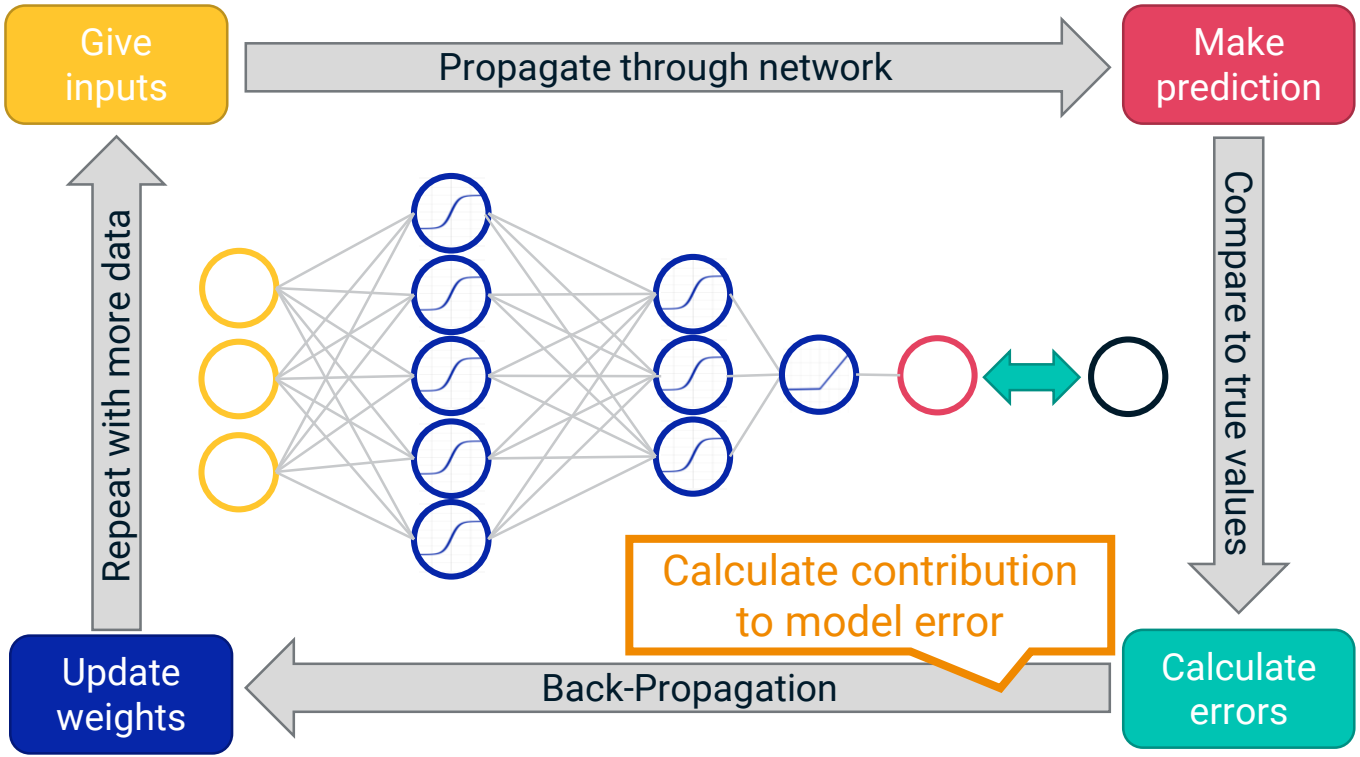
- Naïve estimator: the bid-ask spread tomorrow will be the same as today
- Mean prediction: the bid-ask spread of each bond will be predicted as the mean of the quoted universe
- Linear regression: simple linear regressions using subsets of variables
- Current MSCI model:
  - **Cross-sectional** model – based on data from a single day
  - **Two-stage regression** model

$$BAS = (\beta_1 \cdot C + \beta_2 \cdot D) \cdot e^{\gamma_0 + \gamma_1 \cdot O + \gamma_2 \cdot A + \gamma_3 \cdot B + \gamma_4 \cdot S}$$

  - First stage: linear regression
  - Second stage: regression on the first stage's residuals
  - Second stage regressors are multipliers

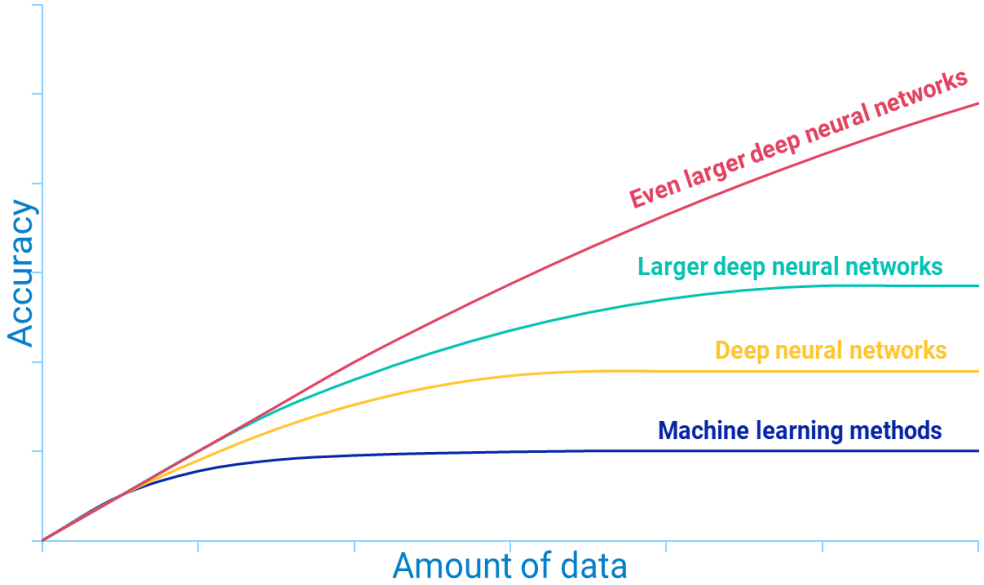
- First stage
  - C – credit spread
  - D – duration
- Second stage
  - O – outstanding amount (issue size)
  - A – relative age of the bond (proportion)
  - B – dummy variable IsBank (whether the issuer is a bank)
  - S – dummy variable – IsSubordinated? (whether it is a subordinated issue)

What is a Deep Neural Network and how does it work?



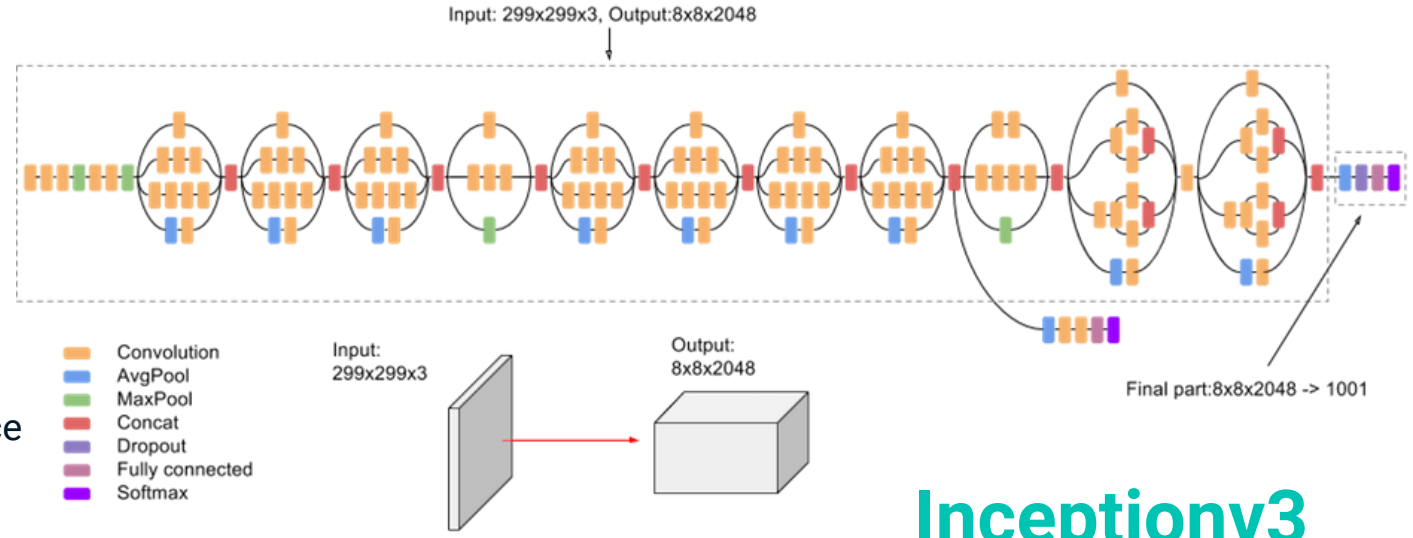
Why is deep learning powerful?

- Knowledge-based methods:
  - Rule-based, parametric
- Data-driven methods:
  - Machine learning (linear regression, SVM, k-means, random forest, etc):
    - More dependence on the preparation of data
  - Deep learning:
    - Representation learning instead of feature engineering – works together with modelling
- Hybrid models:
  - Used for most machine learning and deep learning applications
  - E.g. knowledge-based caps and floors on outputs, filters on inputs



This is a novel problem!

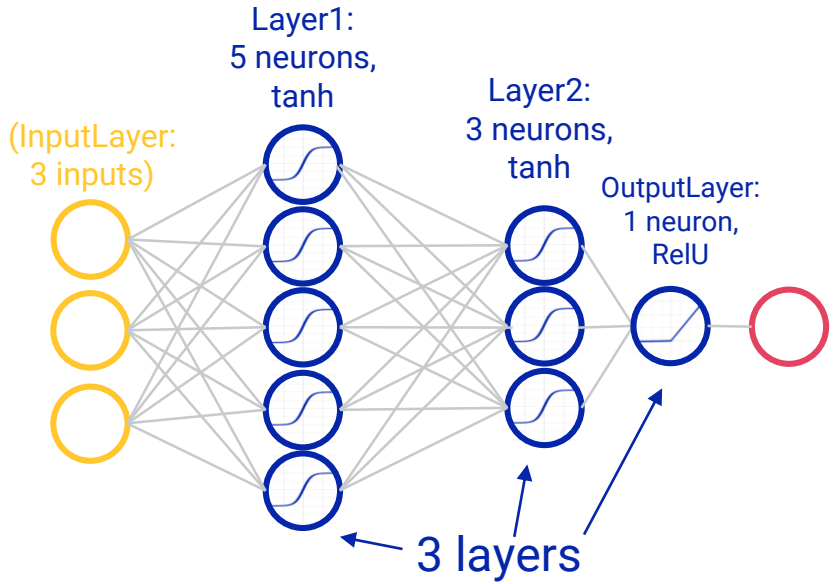
- Standard machine learning projects:
  - lots of people,
  - lots of data,
  - giant networks,
  - lots of hyperparameter optimization
  - on high-performance machines
- A novel problem:
  - there are no published network designs!
  - find architecture that works



Inceptionv3

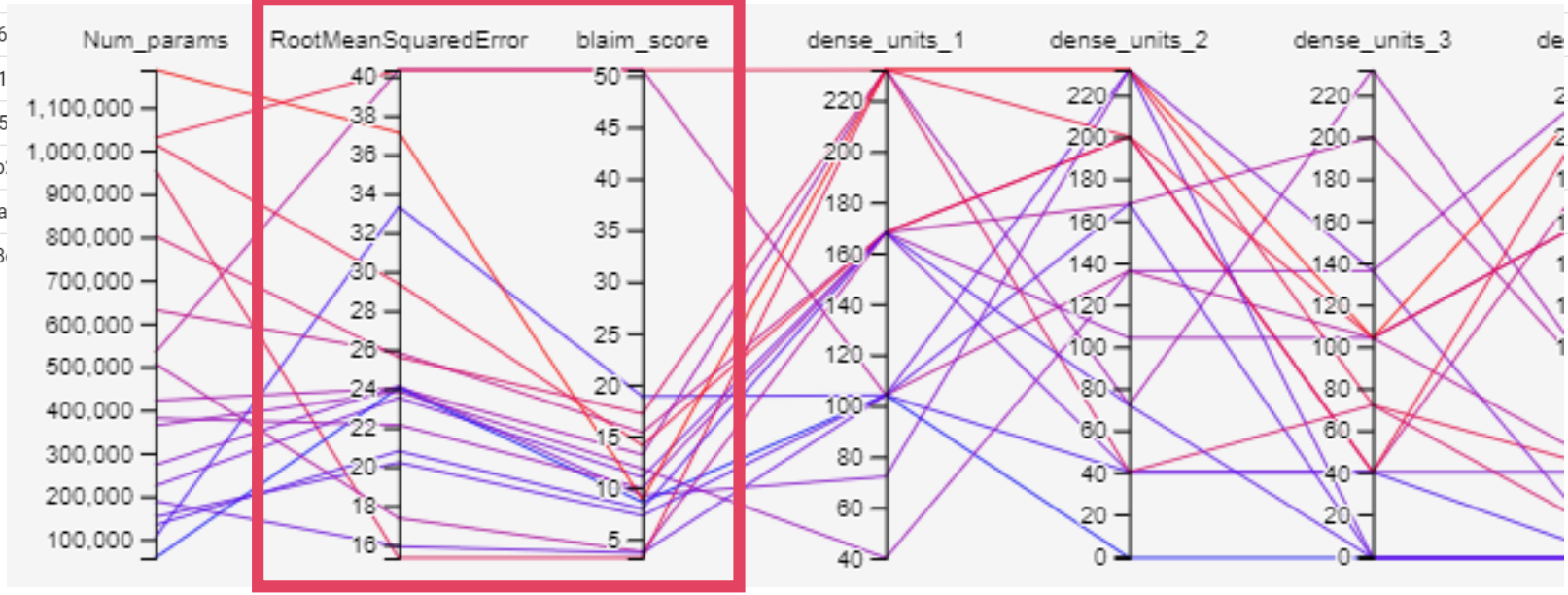
Dense networks of increasing complexity

- 2-layer basic network
- 5-layer manual “logical” network
  - dropout
  - ResNets (skip connections)
- Hyperparameter optimization
  - Choosing the parameters of the network itself
  - number of layers
  - number of neurons in each layer
  - activation function
  - learning rate
  - etc.
- Hyperopt results:
  - 4-6 dense layers
  - large dropout rates (50-75%), i.e. only a small subset of features are actually used
  - larger first layer (128/256), then small layers afterwards (64/32)



Hyperparameter optimization

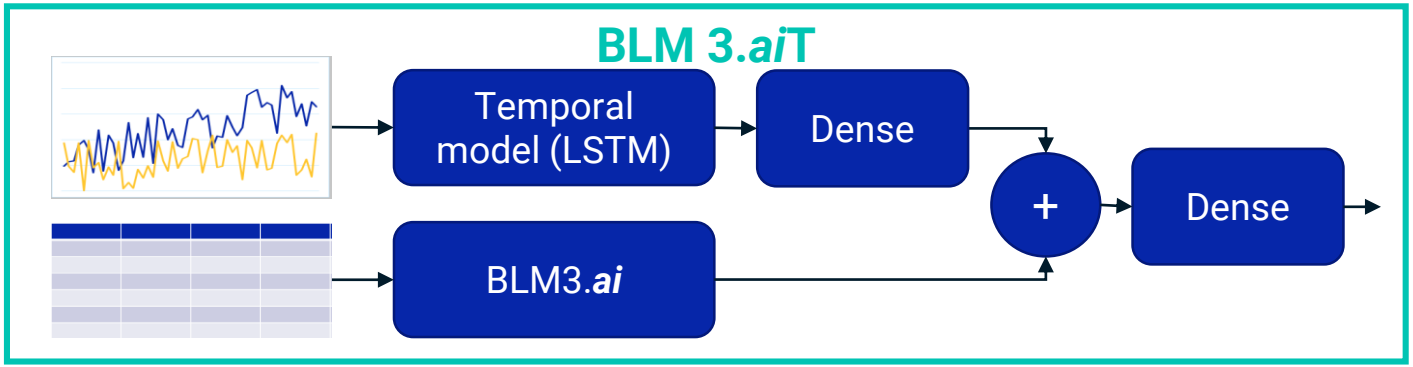
Show Metrics	Trial ID	RootMeanSquaredError	blaim_score	Num_params	lr	num_layers	dense_units_1	dense_units_2	dense_units_3	dense_units_4	dense_units_5	dense_units_6
☐	070139d1558ea3...	33.308	18.838	1.0439e+5	0.010000	3.0000	104.00	40.000	40.000	-1.0000	-1.0000	-1.0000
☐	22fec23ebb41d42...	40.335	50.554	5.3229e+5	0.010000	7.0000	104.00	136.00	104.00	40.000	136.00	168.00
☐	24dfe8d382ec343...	25.819	15.339	6.3091e+5	0.00010000	7.0000	168.00	200.00	40.000	200.00	72.000	72.000
☐	2c92cebd382b6											
☐	3b93ae656c741											
☐	3ba3e9b0b5d35											
☐	4bf31e773b61b											
☐	799ae1c10a2da											
☐	7e95f3d4694d8											



MSCI

Temporal models

- Using Long Short-Term Memory (LSTM) cells
- Allows for connecting time history variables
  - includes time evolution, looks back several days rather than just cross-sectional modelling
  - can capture dynamics / derivatives
- Simple version: concatenate LSTM and dense network outputs at the end



- (Advanced version: use hyperopted network output as initial states of LSTM)

MSCI

Evaluation

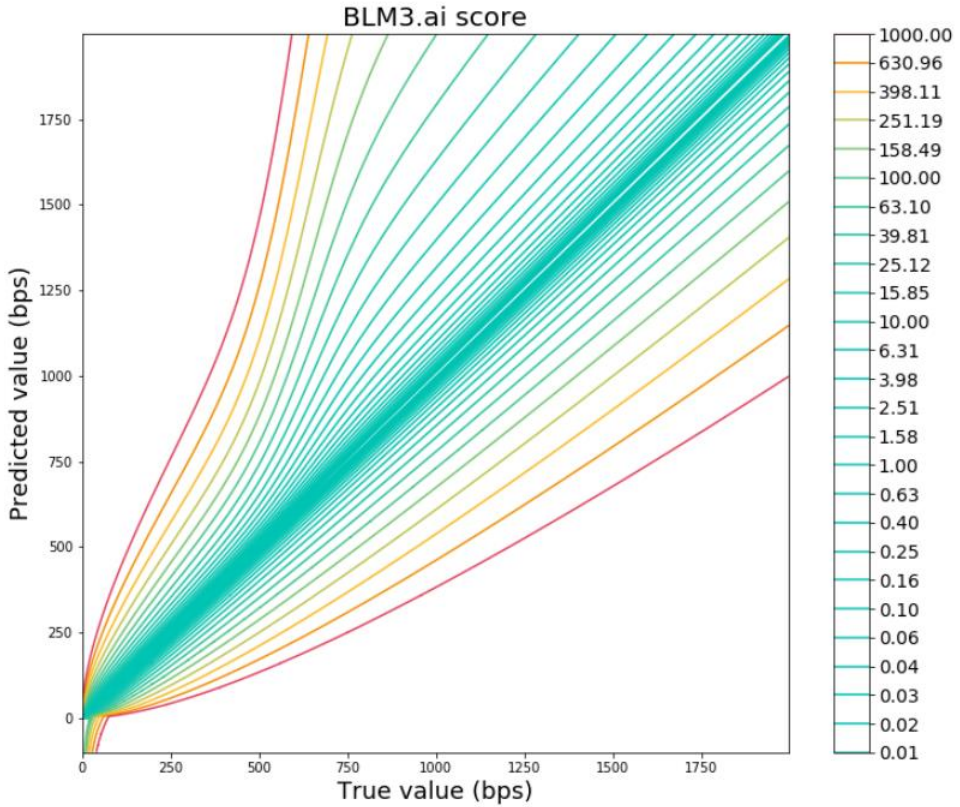
HOW DO YOU KNOW WHICH OF TWO MODELS IS BETTER?

Getting closer to the real world

1. Evaluation metric bias:
   - should be based on client preference rather than a mathematical formula
2. Cross sectional models vs future prediction
   - single day vs predicting for the days ahead
3. In-sample leakage
   - is our test set really independent?
4. Statistical evaluation
   - factor in randomness

Evaluation metrics bias

- How to evaluate models?
- RMSE? R^2? Average absolute error? Median abs. error? Percentiles?
- Custom metric based on client need (i.e. expected usage), not just a mathematical expression!



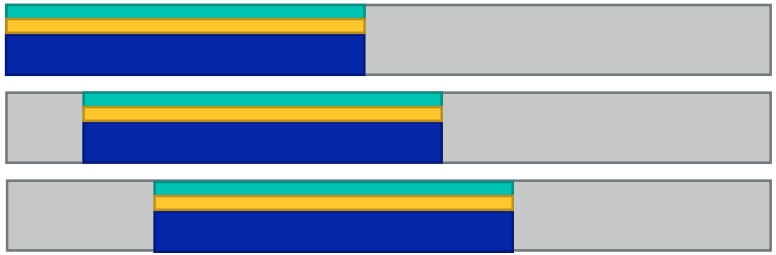
MSCI

Cross-sectional vs future prediction scenarios

- Train – validation – test set split

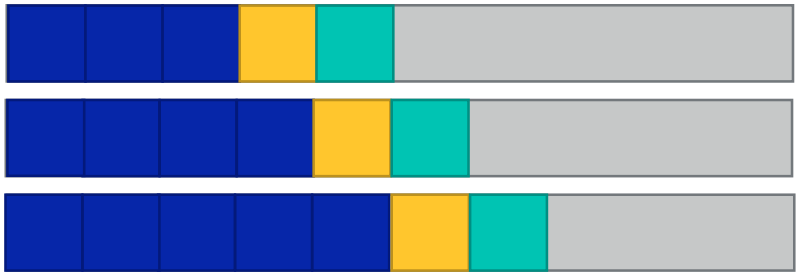
1. Cross section every day

2. Larger cross sectional model (e.g. 5 days)



3. Constant-size moving window

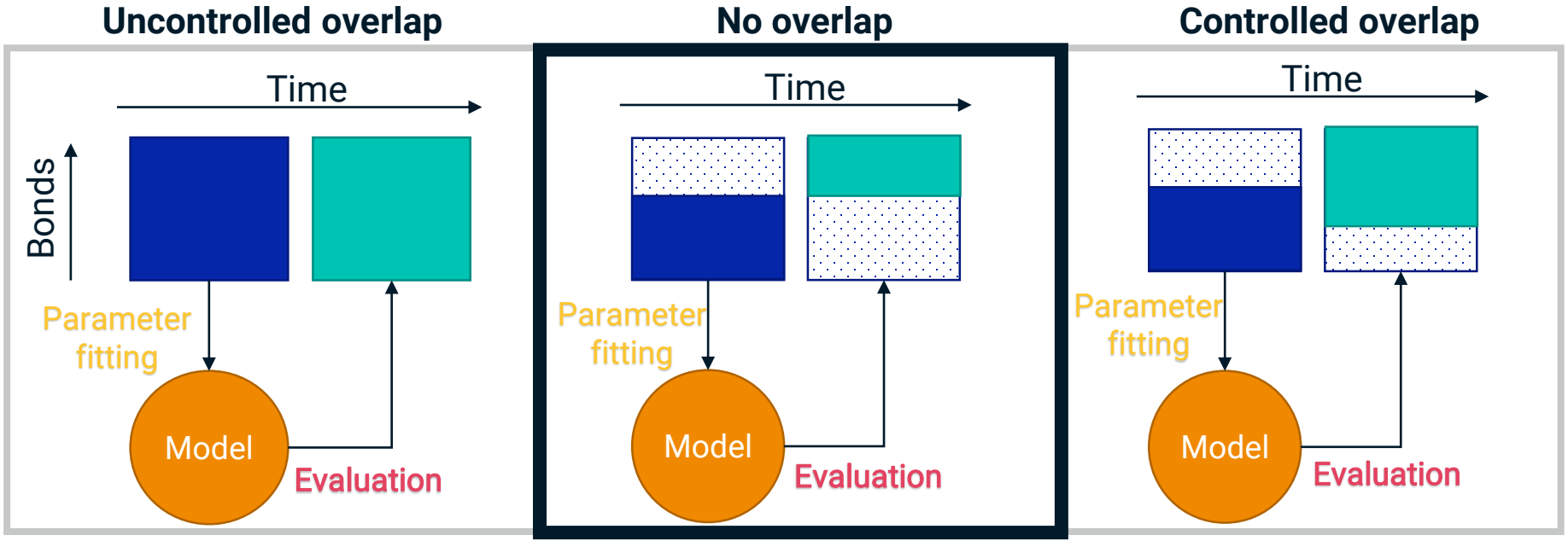
4. Increasing training set moving window



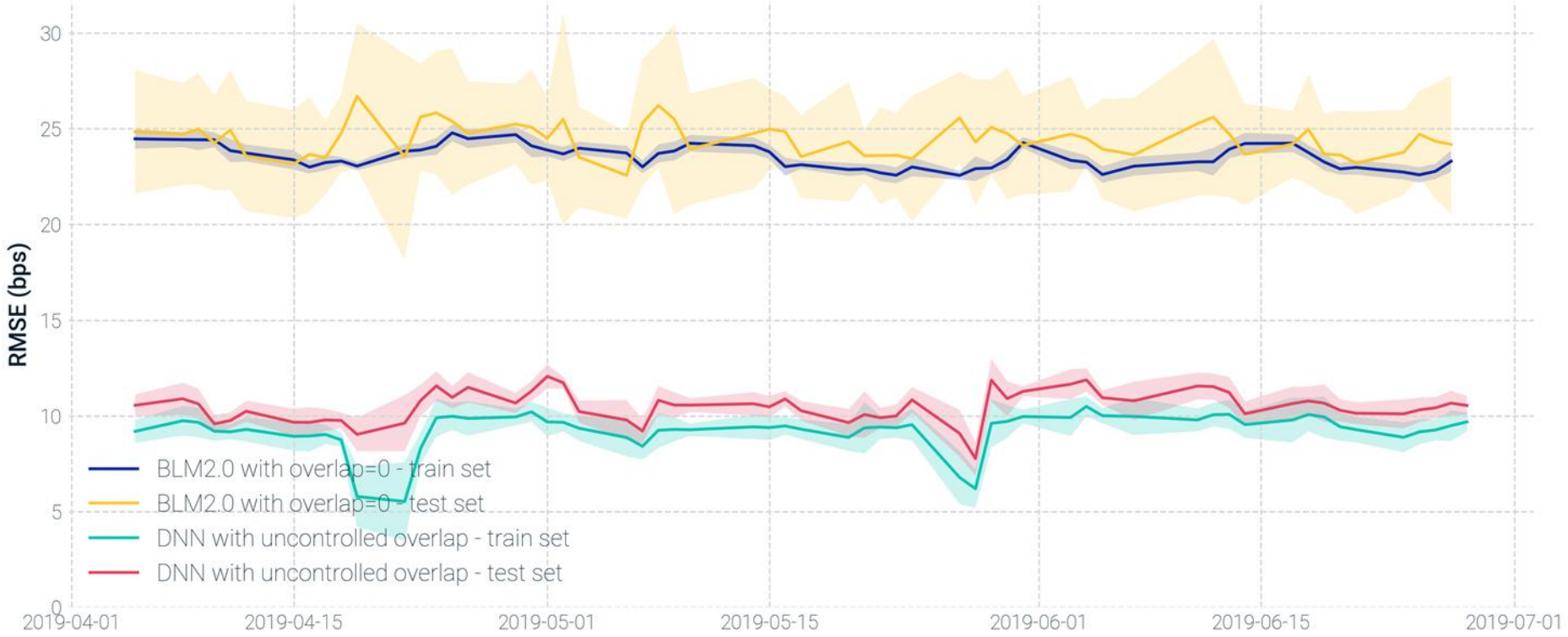
MSCI

In-sample leakage

- Issue: evaluation using quotes for the same bonds that were used for training

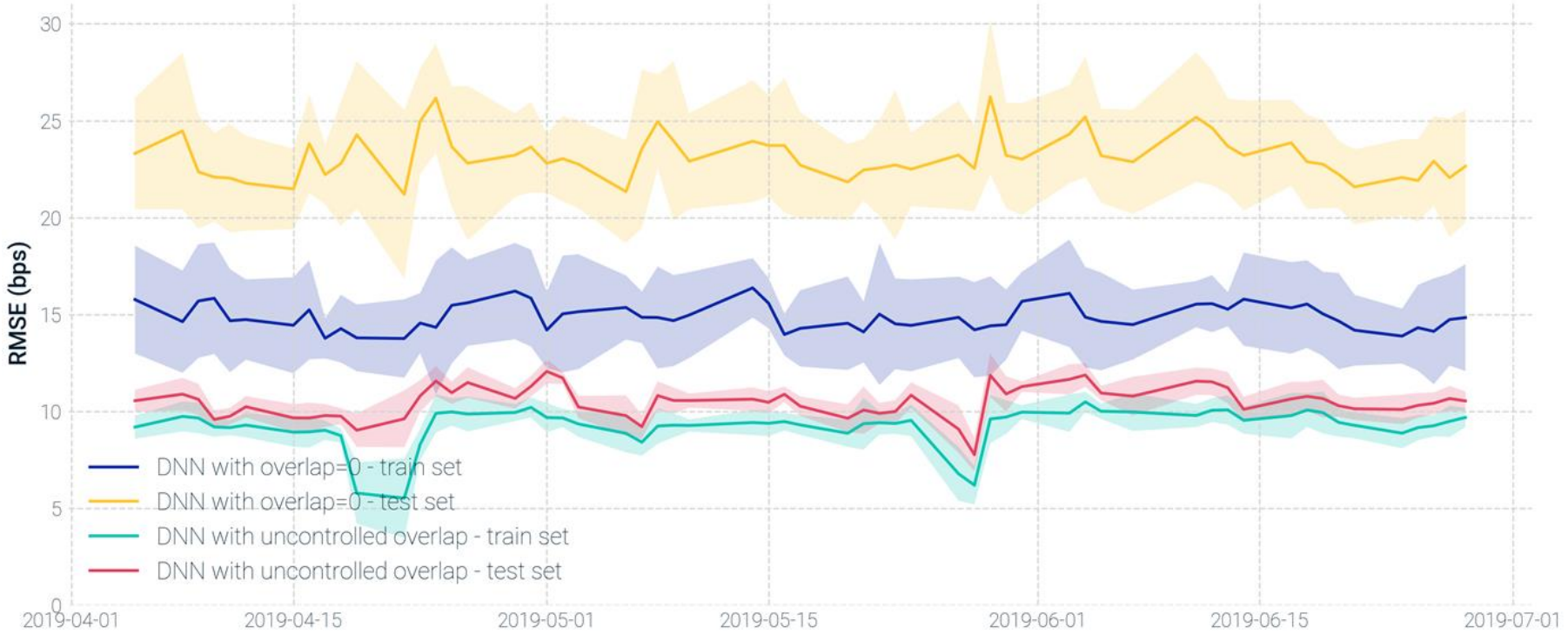


In-sample leakage and overfitting - 1



MSCI

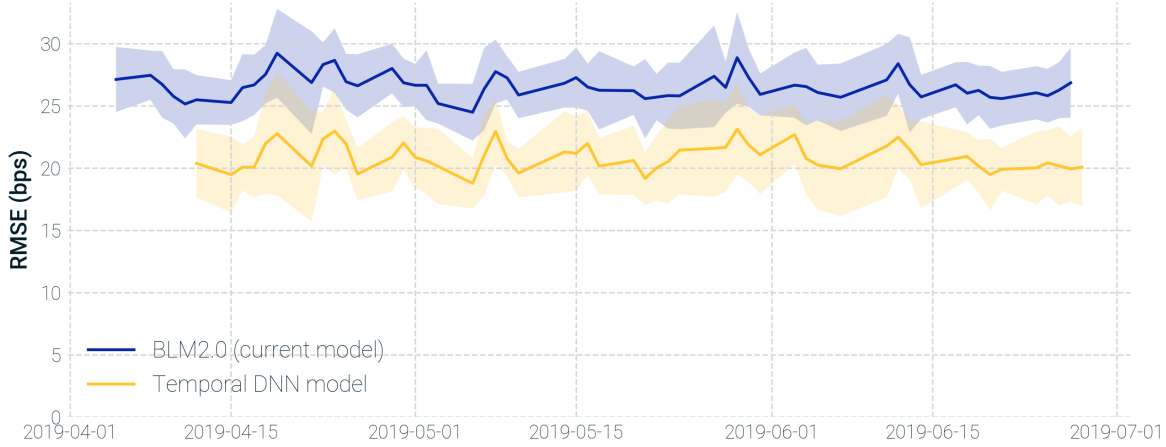
In-sample leakage and overfitting - 2



MSCI

Statistical evaluation of test set results

- Multiple factors cause uncertainty / randomness:
  - Biased train – validation – test set selection
  - Initial weights of the network (less important)
  - Hyperopt process
- Evaluation should be statistical
  - 10-20-50 runs each day
- Backtesting: improvements that are stable over time increase confidence (even if bands overlap)



MSCI

MSCI

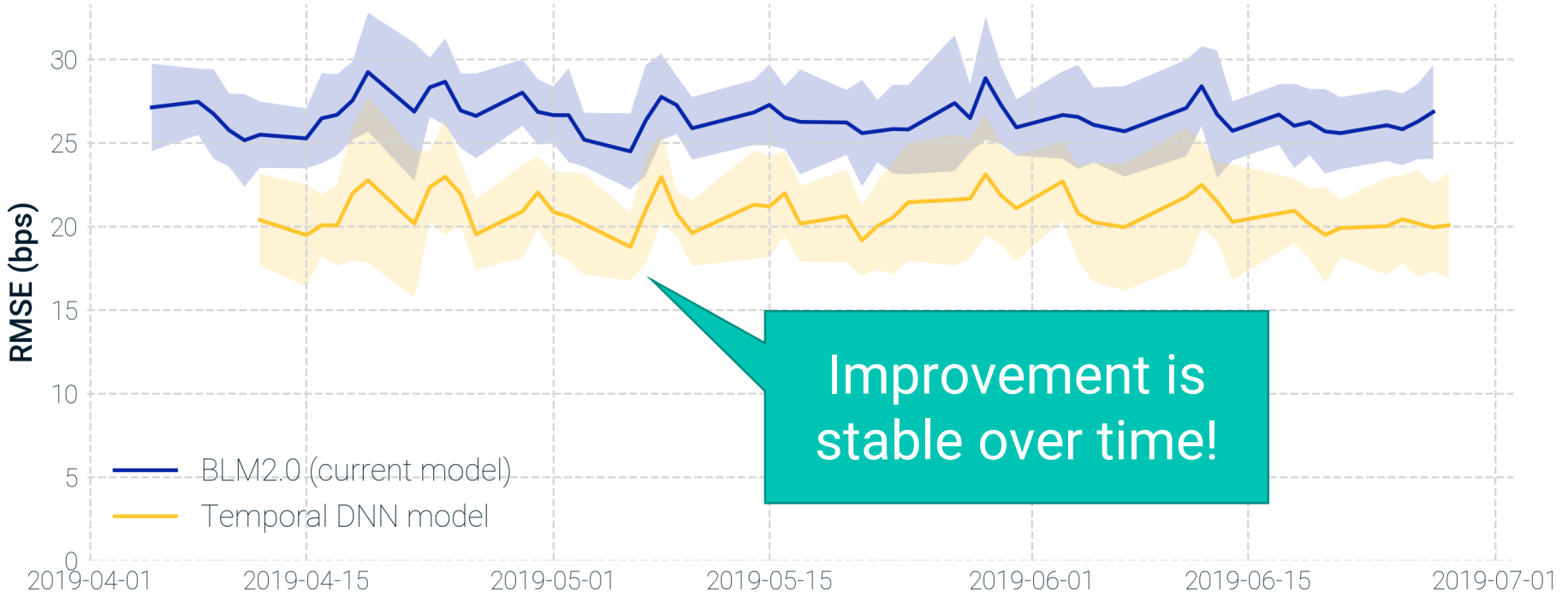
Results

WHAT CAN YOU EXPECT FROM A DEEP LEARNING MODEL ON A REGRESSION TASK?

What did we expect to achieve?

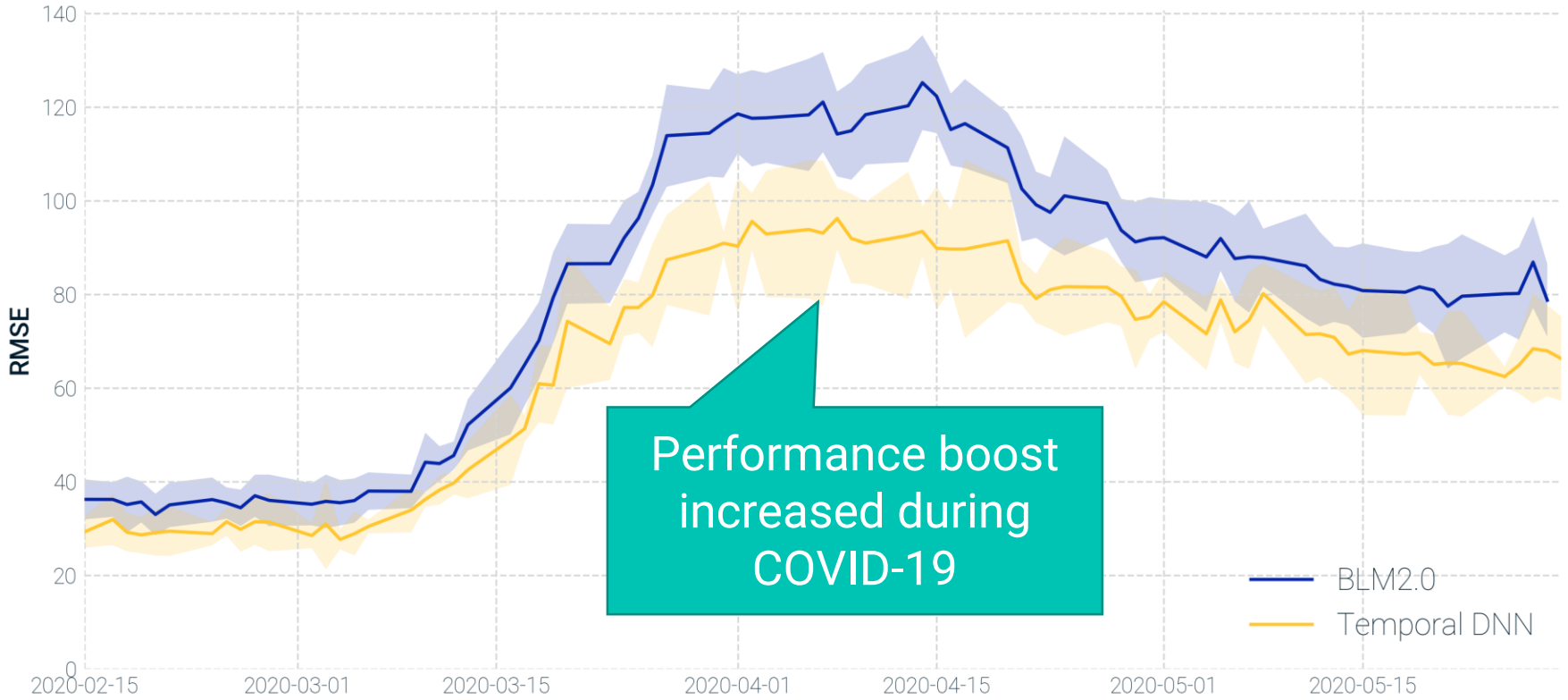
1. Linear regression tends to be OK for regression tasks
2. Our existing model is more sophisticated than linear regression:
   - two-stage regression
   - can model non-linear relationships
3. It was not clear in advance that any improvement can be achieved
4. At the outset:
   - 10% improvement → successful project
   - 15% improvement → quite good results
   - 20% improvement → expected best-case scenario when backtesting over the long term

Stable and significant improvement in calm periods



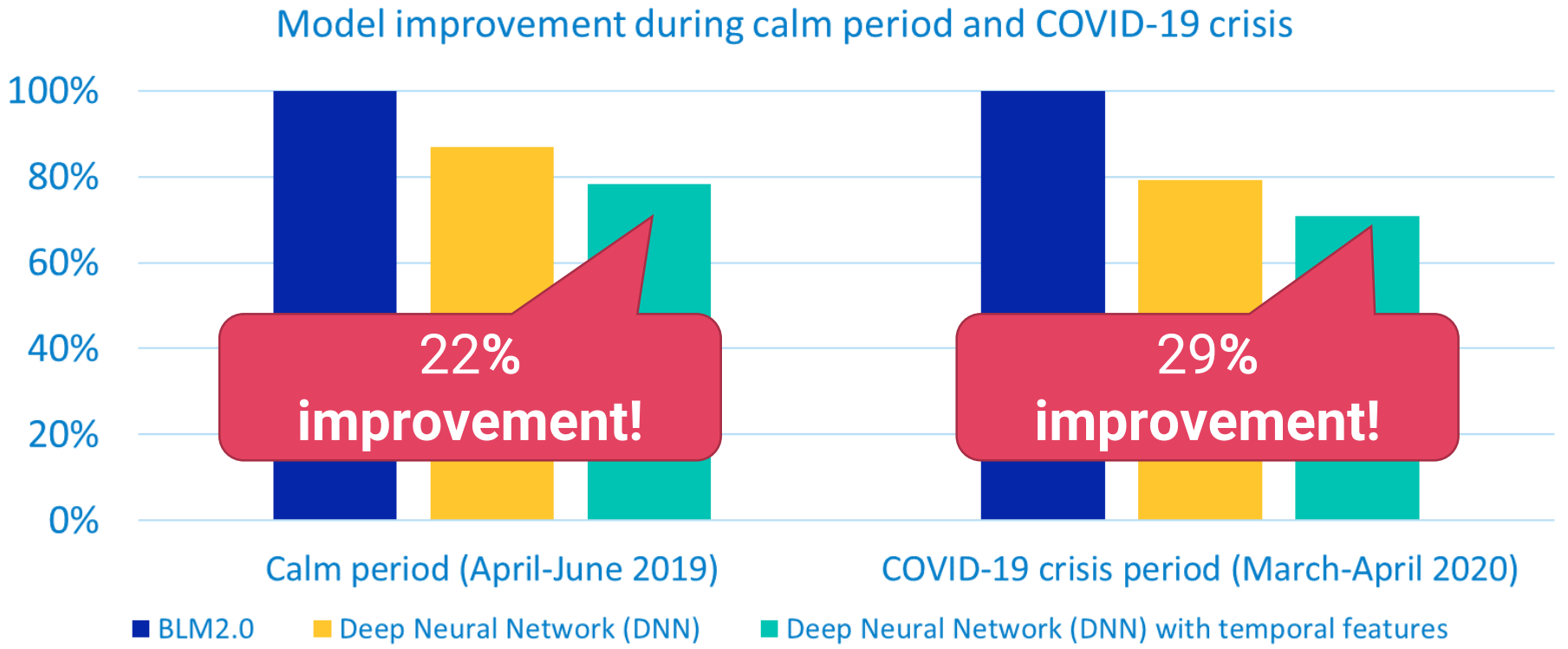
MSCI

Improvement more significant during a crisis period

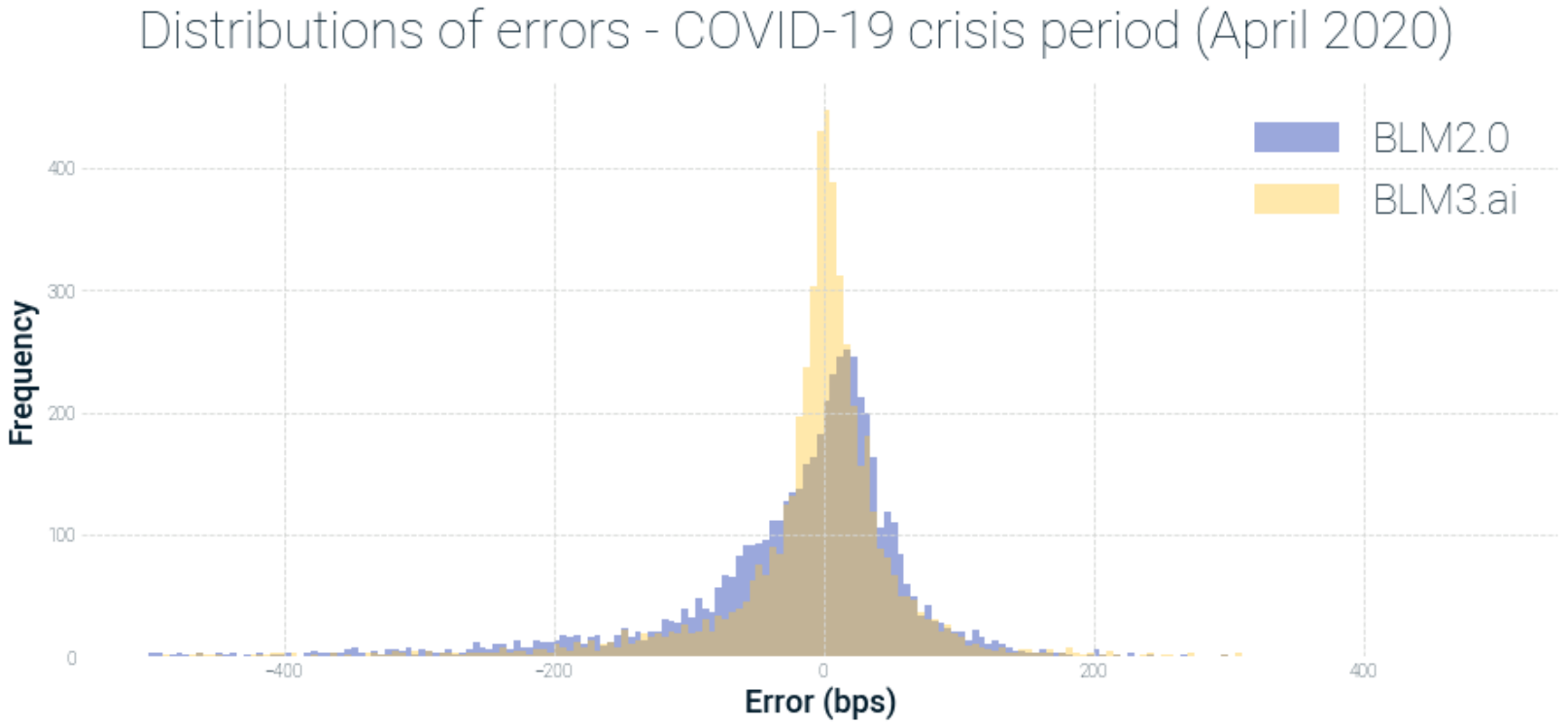


MSCI

Overall performance



RMSE is not the full picture...



MSCI

Thank you for your attention!

FOR MORE DETAILS ON THIS: LASZLO.ARANY@MSCI.COM

IF INTERESTED IN AN INTERNSHIP: JOIN_MSCI_BUDAPEST@MSCI.COM

About MSCI

MSCI is a leading provider of critical decision support tools and services for the global investment community. With over 45 years of expertise in research, data and technology, we power better investment decisions by enabling clients to understand and analyze key drivers of risk and return and confidently build more effective portfolios. We create industry-leading research-enhanced solutions that clients use to gain insight into and improve transparency across the investment process. To learn more, please visit www.msci.com.

Notice and disclaimer

This document and all of the information contained in it, including without limitation all text, data, graphs, charts (collectively, the "Information") is the property of MSCI Inc. or its subsidiaries (collectively, "MSCI"), or MSCI's licensors, direct or indirect suppliers or any third party involved in making or compiling any Information (collectively, with MSCI, the "Information Providers") and is provided for informational purposes only. The Information may not be modified, reverse-engineered, reproduced or redisseminated in whole or in part without prior written permission from MSCI. All rights in the Information are reserved by MSCI and/or its Information Providers.

The Information may not be used to create derivative works or to verify or correct other data or information. For example (but without limitation), the Information may not be used to create indexes, databases, risk models, analytics, software, or in connection with the issuing, offering, sponsoring, managing or marketing of any securities, portfolios, financial products or other investment vehicles utilizing or based on, linked to, tracking or otherwise derived from the Information or any other MSCI data, information, products or services.

The user of the Information assumes the entire risk of any use it may make or permit to be made of the Information. NONE OF THE INFORMATION PROVIDERS MAKES ANY EXPRESS OR IMPLIED WARRANTIES OR REPRESENTATIONS WITH RESPECT TO THE INFORMATION (OR THE RESULTS TO BE OBTAINED BY THE USE THEREOF), AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, EACH INFORMATION PROVIDER EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES (INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF ORIGINALITY, ACCURACY, TIMELINESS, NON-INFRINGEMENT, COMPLETENESS, MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE) WITH RESPECT TO ANY OF THE INFORMATION.

Without limiting any of the foregoing and to the maximum extent permitted by applicable law, in no event shall any Information Provider have any liability regarding any of the Information for any direct, indirect, special, punitive, consequential (including lost profits) or any other damages even if notified of the possibility of such damages. The foregoing shall not exclude or limit any liability that may not by applicable law be excluded or limited, including without limitation (as applicable), any liability for death or personal injury to the extent that such injury results from the negligence or willful default of itself, its servants, agents or sub-contractors.

Information containing any historical information, data or analysis should not be taken as an indication or guarantee of any future performance, analysis, forecast or prediction. Past performance does not guarantee future results.

The Information should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. All Information is impersonal and not tailored to the needs of any person, entity or group of persons.

None of the Information constitutes an offer to sell (or a solicitation of an offer to buy), any security, financial product or other investment vehicle or any trading strategy.

It is not possible to invest directly in an index. Exposure to an asset class or trading strategy or other category represented by an index is only available through third party investable instruments (if any) based on that index. MSCI does not issue, sponsor, endorse, market, offer, review or otherwise express any opinion regarding any fund, ETF, derivative or other security, investment, financial product or trading strategy that is based on, linked to or seeks to provide an investment return related to the performance of any MSCI index (collectively, "Index Linked Investments"). MSCI makes no assurance that any Index Linked Investments will accurately track index performance or provide positive investment returns. MSCI Inc. is not an investment adviser or fiduciary and MSCI makes no representation regarding the advisability of investing in any Index Linked Investments.

Index returns do not represent the results of actual trading of investible assets/securities. MSCI maintains and calculates indexes, but does not manage actual assets. Index returns do not reflect payment of any sales charges or fees an investor may pay to purchase the securities underlying the index or Index Linked Investments. The imposition of these fees and charges would cause the performance of an Index Linked Investment to be different than the MSCI index performance.

The Information may contain back tested data. Back-tested performance is not actual performance, but is hypothetical. There are frequently material differences between back tested performance results and actual results subsequently achieved by any investment strategy.

Constituents of MSCI equity indexes are listed companies, which are included in or excluded from the indexes according to the application of the relevant index methodologies. Accordingly, constituents in MSCI equity indexes may include MSCI Inc., clients of MSCI or suppliers to MSCI. Inclusion of a security within an MSCI index is not a recommendation by MSCI to buy, sell, or hold such security, nor is it considered to be investment advice.

Data and information produced by various affiliates of MSCI Inc., including MSCI ESG Research LLC and Barra LLC, may be used in calculating certain MSCI indexes. More information can be found in the relevant index methodologies on www.msci.com.

MSCI receives compensation in connection with licensing its indexes to third parties. MSCI Inc.'s revenue includes fees based on assets in Index Linked Investments. Information can be found in MSCI Inc.'s company filings on the Investor Relations section of www.msci.com.

MSCI ESG Research LLC is a Registered Investment Adviser under the Investment Advisers Act of 1940 and a subsidiary of MSCI Inc. Except with respect to any applicable products or services from MSCI ESG Research, neither MSCI nor any of its products or services recommends, endorses, approves or otherwise expresses any opinion regarding any issuer, securities, financial products or instruments or trading strategies and MSCI's products or services are not intended to constitute investment advice or a recommendation to make (or refrain from making) any kind of investment decision and may not be relied on as such. Issuers mentioned or included in any MSCI ESG Research materials may include MSCI Inc., clients of MSCI or suppliers to MSCI, and may also purchase research or other products or services from MSCI ESG Research. MSCI ESG Research materials, including materials utilized in any MSCI ESG Indexes or other products, have not been submitted to, nor received approval from, the United States Securities and Exchange Commission or any other regulatory body.

Any use of or access to products, services or information of MSCI requires a license from MSCI. MSCI, Barra, RiskMetrics, IPD and other MSCI brands and product names are the trademarks, service marks, or registered trademarks of MSCI or its subsidiaries in the United States and other jurisdictions. The Global Industry Classification Standard (GICS) was developed by and is the exclusive property of MSCI and Standard & Poor's. "Global Industry Classification Standard (GICS)" is a service mark of MSCI and Standard & Poor's.

Privacy notice: For information about how MSCI collects and uses personal data, please refer to our Privacy Notice at https://www.msci.com/privacy-pledge.

Contact us

AMERICAS		EUROPE, MIDDLE EAST & AFRICA		ASIA PACIFIC	
Americas	+1 888 588 4567 *	Cape Town	+ 27 21 673 0100	China North	10800 852 1032 *
Atlanta	+ 1 404 551 3212	Frankfurt	+ 49 69 133 859 00	China South	10800 152 1032 *
Boston	+ 1 617 532 0920	Geneva	+ 41 22 817 9777	Hong Kong	+ 852 2844 9333
Chicago	+ 1 312 675 0545	London	+ 44 20 7618 2222	Mumbai	+ 91 22 6784 9160
Monterrey	+ 52 81 1253 4020	Milan	+ 39 02 5849 0415	Seoul	00798 8521 3392 *
New York	+ 1 212 804 3901	Paris	0800 91 59 17 *	Singapore	800 852 3749 *
San Francisco	+ 1 415 836 8800			Sydney	+ 61 2 9033 9333
São Paulo	+ 55 11 3706 1360			Taipei	008 0112 7513 *
Toronto	+ 1 416 628 1007			Thailand	0018 0015 6207 7181 *
* = toll free msci.com clientservice@msci.com esgclientservice@msci.com				Tokyo	+81 3 5290 1555

MSCI

The process for submitting a formal index complaint can be found on the index regulation page of MSCI's website at: https://www.msci.com/index-regulation.

Advanced temporal model design