University of Glasgow

Investors Are (Not) Always Right

A Comparison of Transaction-based and Profitability-based Metrics for Financial Asset Recommendation

Dr. Javier Sanz-Cruzado

WORLD
CHANGING
GLASGOW

A WORLD
TOP 100
UNIVERSITY

The team

Dr. Javier Sanz-Cruzado
University of Glasgow

Dr. Richard McCreadie
University of Glasgow

Nikolaos Droukas
National Bank of Greece

Prof. Craig Macdonald
University of Glasgow

Prof. Iadh Ounis
University of Glasgow

Financial asset recommendation (FAR)

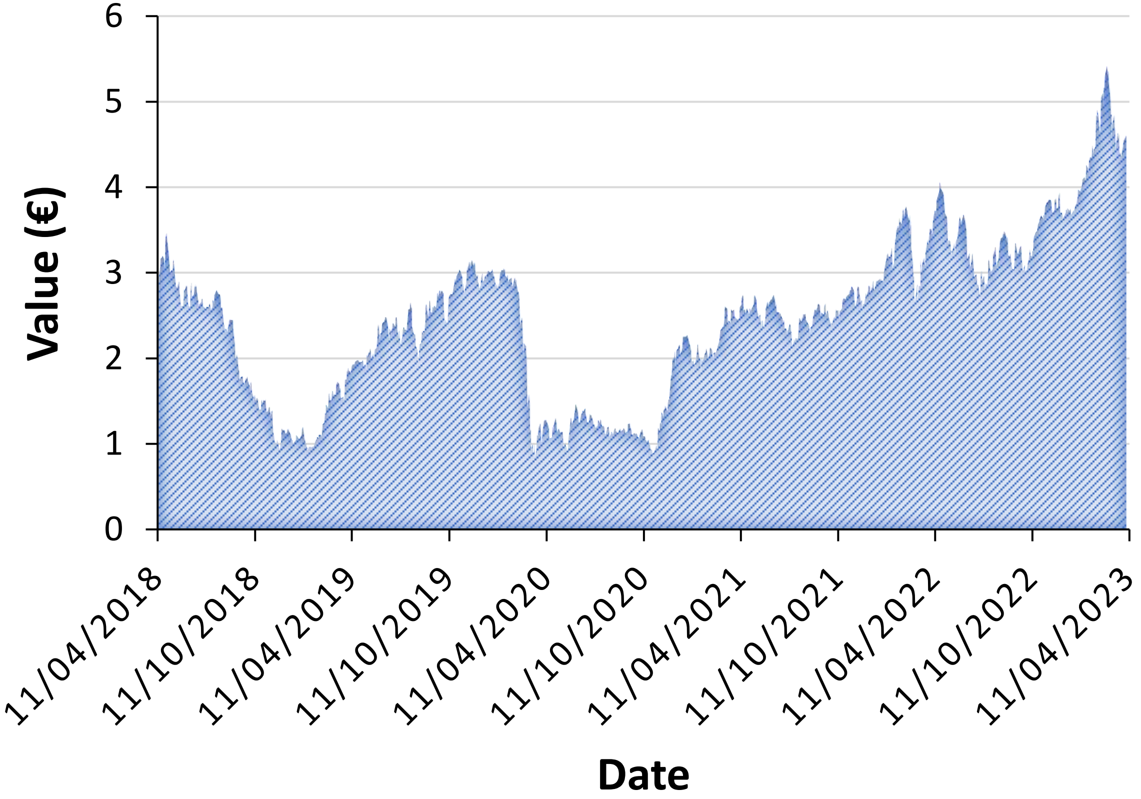
Motivation

- **Customer's goal:** Earn Money

- **Achieve this by investing in Financial Assets**
  - Stocks
  - Bonds
  - Mutual funds

- Identifying good assets is difficult and time consuming

NATIONAL BANK
OF GREECE



Solution

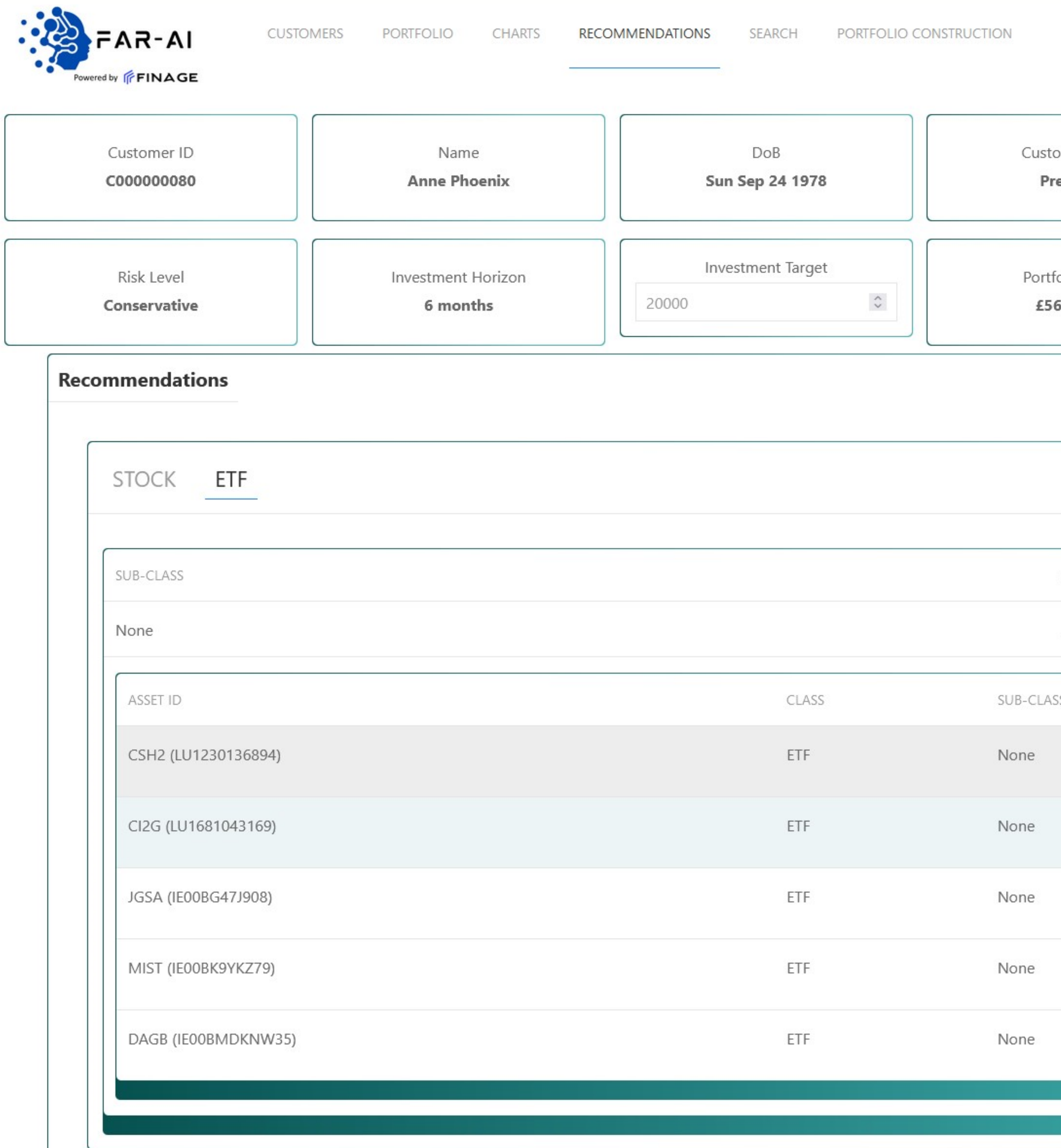
Financial asset recommendation

Given a customer

- Automatically rank financial assets
- Ranking the best assets first

Use cases:

- Assistance of financial advisors
- Robo-advisors
- Automatic trading



Customer ID
C000000080

Name
Anne Phoenix

DoB
Sun Sep 24 1978

Custo
Pre

Risk Level
Conservative

Investment Horizon
6 months

Investment Target
20000

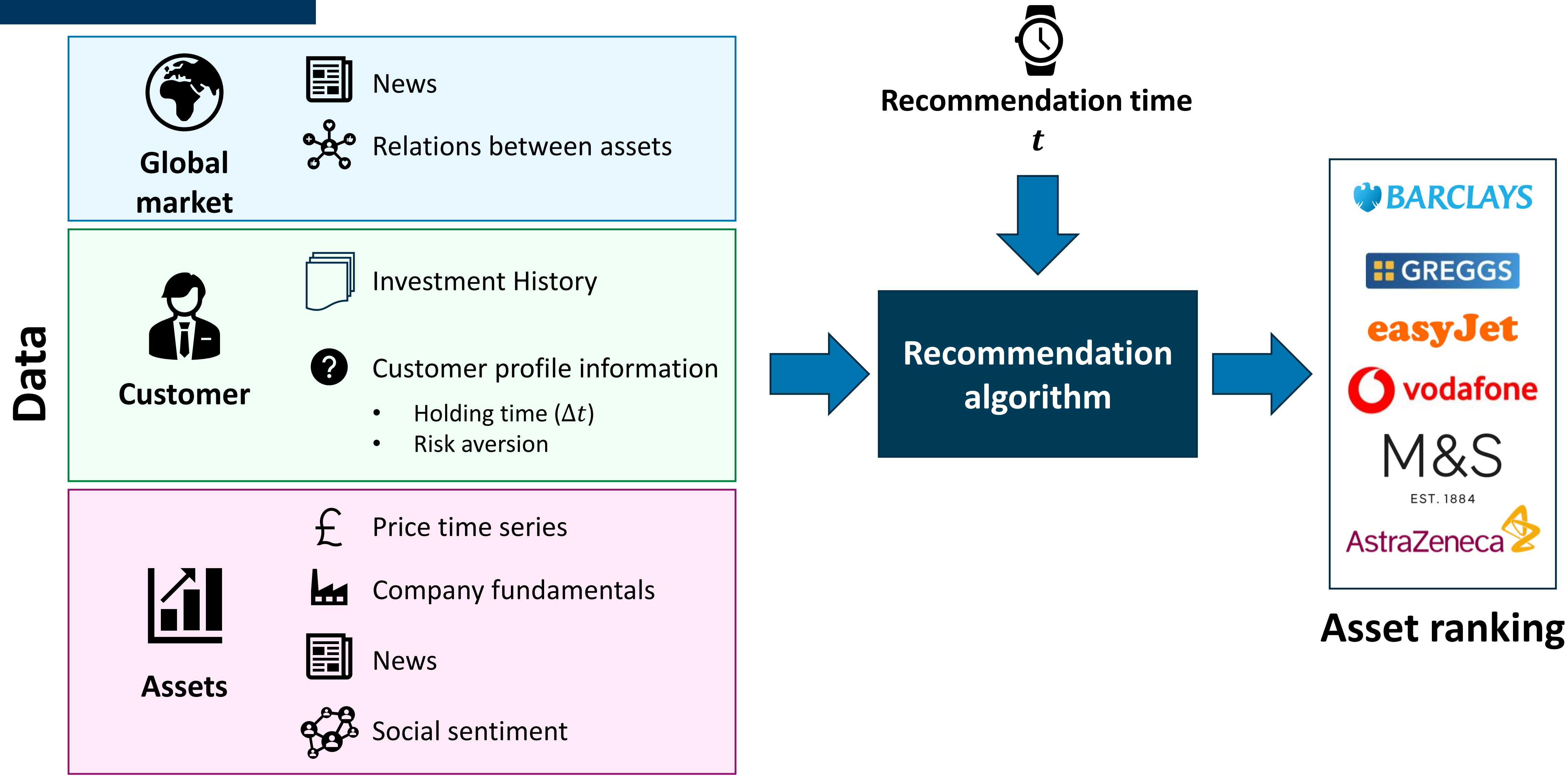
Portfo
£56

Recommendations

SUB-CLASS
None

ASSET ID	CLASS	SUB-CLAS
CSH2 (LU1230136894)	ETF	None
CI2G (LU1681043169)	ETF	None
JGSA (IE00BG47J908)	ETF	None
MIST (IE00BK9YKZ79)	ETF	None
DAGB (IE00BMDKNW35)	ETF	None

Task



Data

Global market
- News
- Relations between assets

Customer
- Investment History
- Customer profile information
  - Holding time (Δt)
  - Risk aversion

Assets
- Price time series
- Company fundamentals
- News
- Social sentiment

Recommendation time t

Recommendation algorithm

BARCLAYS
GREGGS
easyJet
vodafone
M&S EST. 1884
AstraZeneca

Asset ranking

What makes FAR interesting?

01

Multi-objective

- Customers want to increase their money
- But we also need to adapt to their personal situation, preferences and needs
  - Risk aversion
  - Holding time
  - Capacity

02

Multi-modal

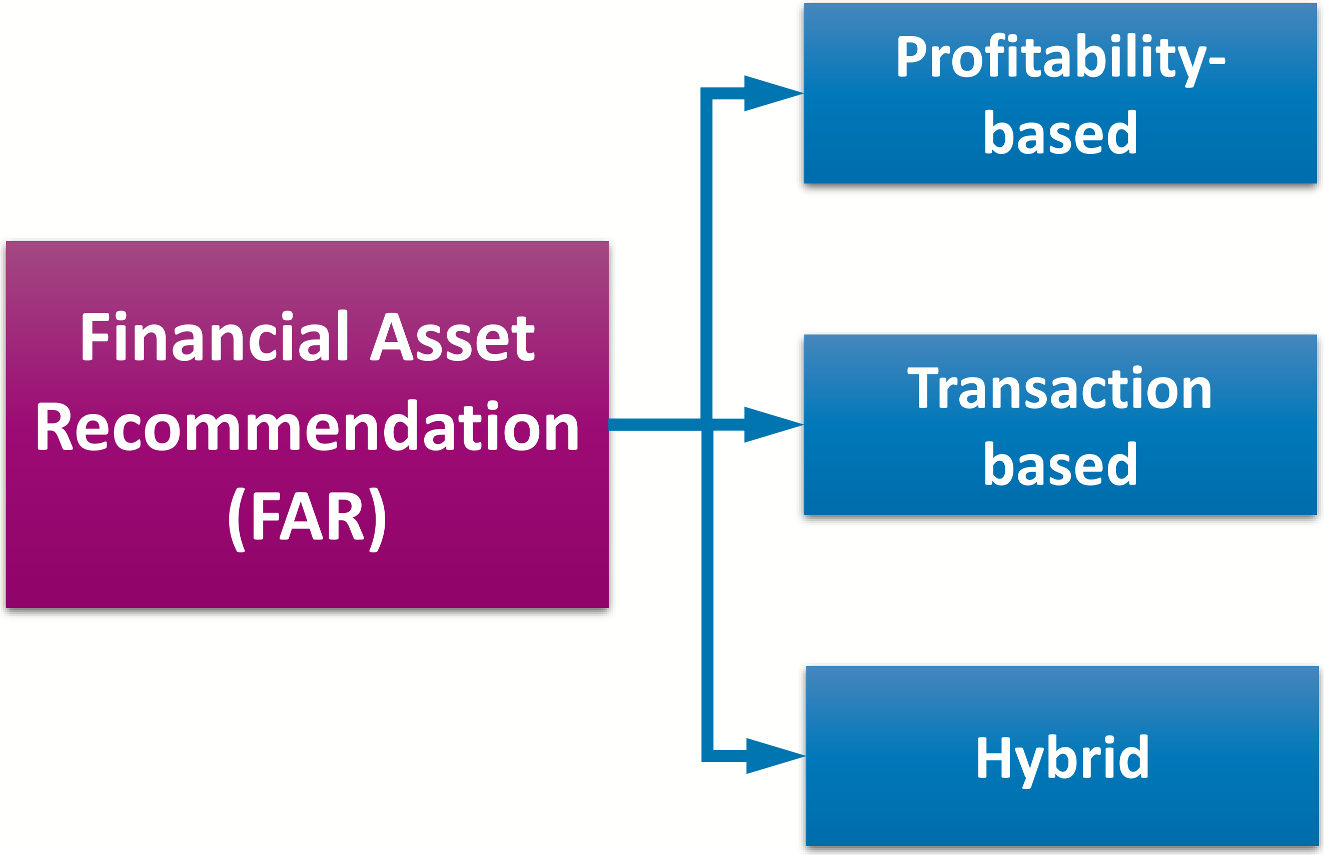
- Asset information
  - Pricing time series
  - Company fundamentals
  - News
- Customer information
  - Investor profile
  - Past investments

03

Time dependent

- Asset valuations are dynamic.
- Multiple factors affect price changes.
- Even external events
  - Pandemics, wars
  - Governmental regulations

Algorithms



Financial Asset Recommendation (FAR)

Profitability-based

- Asset dynamic asset information (prices, news)
- Non-personalized

Transaction based

- Uses what others have bought / sold
- Customer / asset similarity

Hybrid

- Combination of previous methods

Evaluation

How do we evaluate?

Expert based

Profitability based

Financial Asset Recommendation

Transaction based

Do our customers earn money?

- Aligned with customer interests
- Ignores past/future customer actual investments

Can we predict future investments?

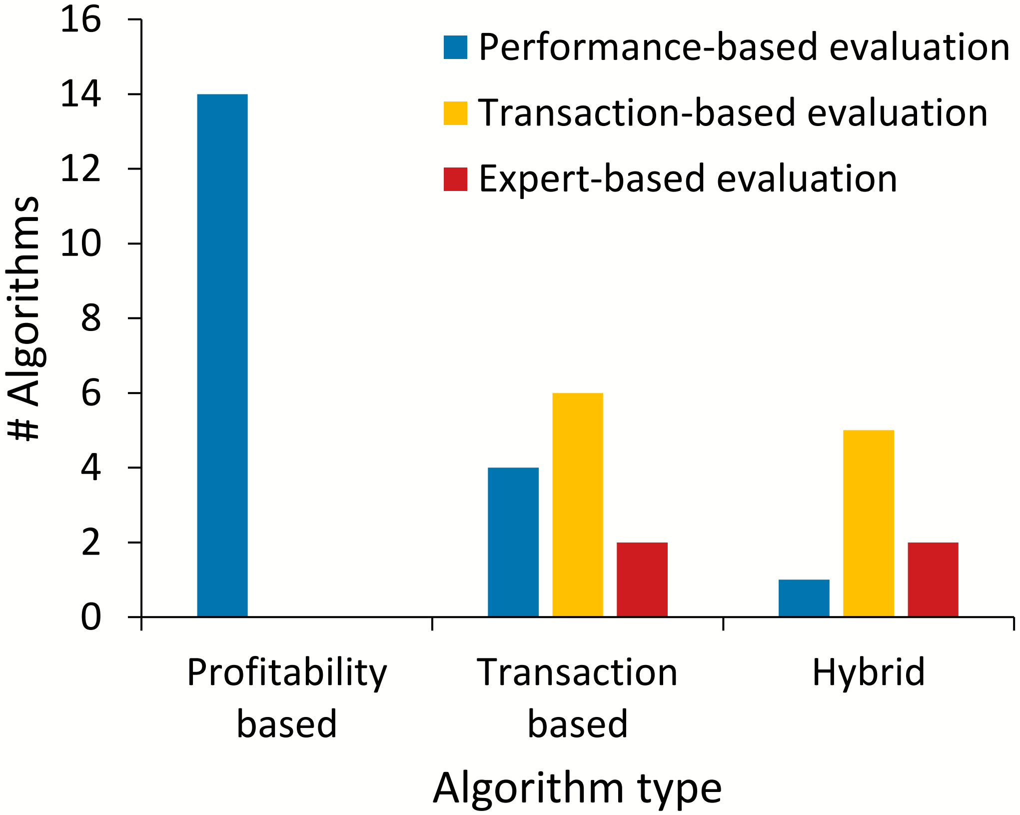
- Investment transactions indicate strong preference
- Relevant transactions: acquisitions
- Ignores temporal pricing information

In both cases, metrics look at a fixed time interval

- **Metrics:** Key performance indicators at a fixed time interval
  - Return on investment (ROI)
  - Net profit

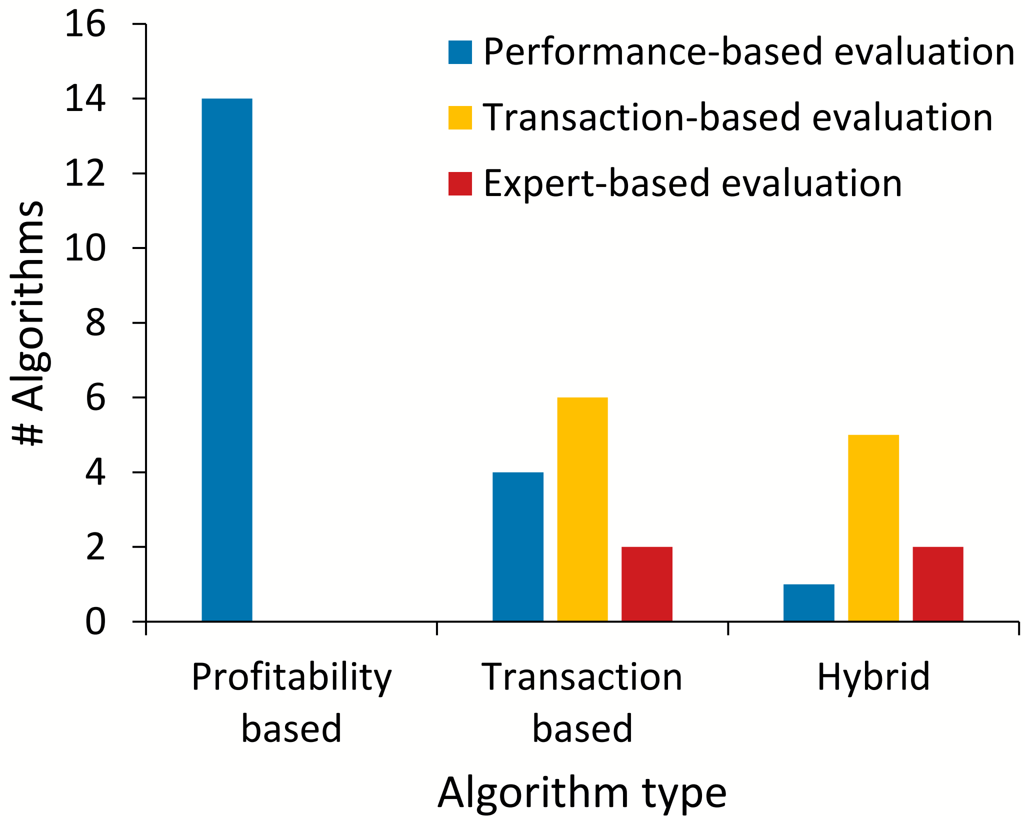
- **Metrics:** Recommender systems metrics
  - Precision
  - nDCG

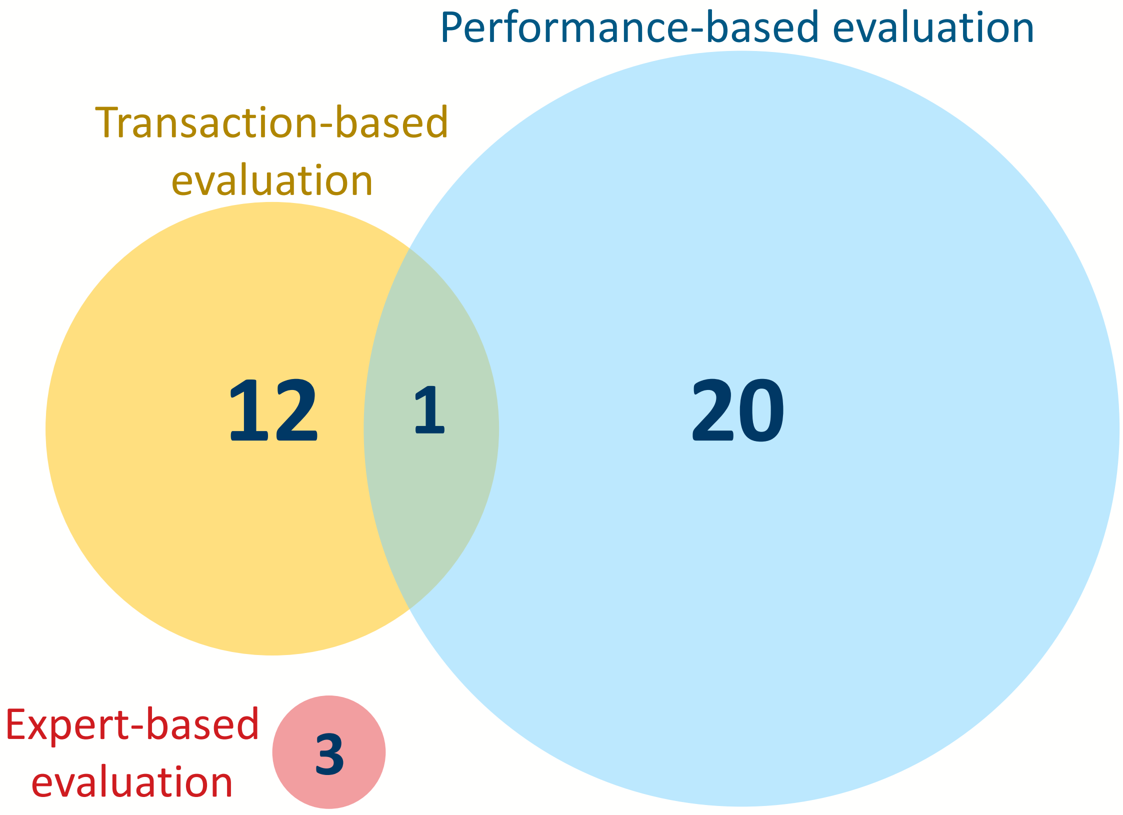
How have these metrics have been used historically?



- Evaluation is fragmented
- A majority of methods evaluate using profitability-based measures.
  - Aligned with customer interest
  - Transactions are difficult to get (proprietary datasets)
- Methods with transactions tend to evaluate using IR ranking measures.
- Expert-based evaluation is rarely used

How have these metrics have been used historically?





Transaction-based evaluation and profitability-based evaluation have barely been compared!

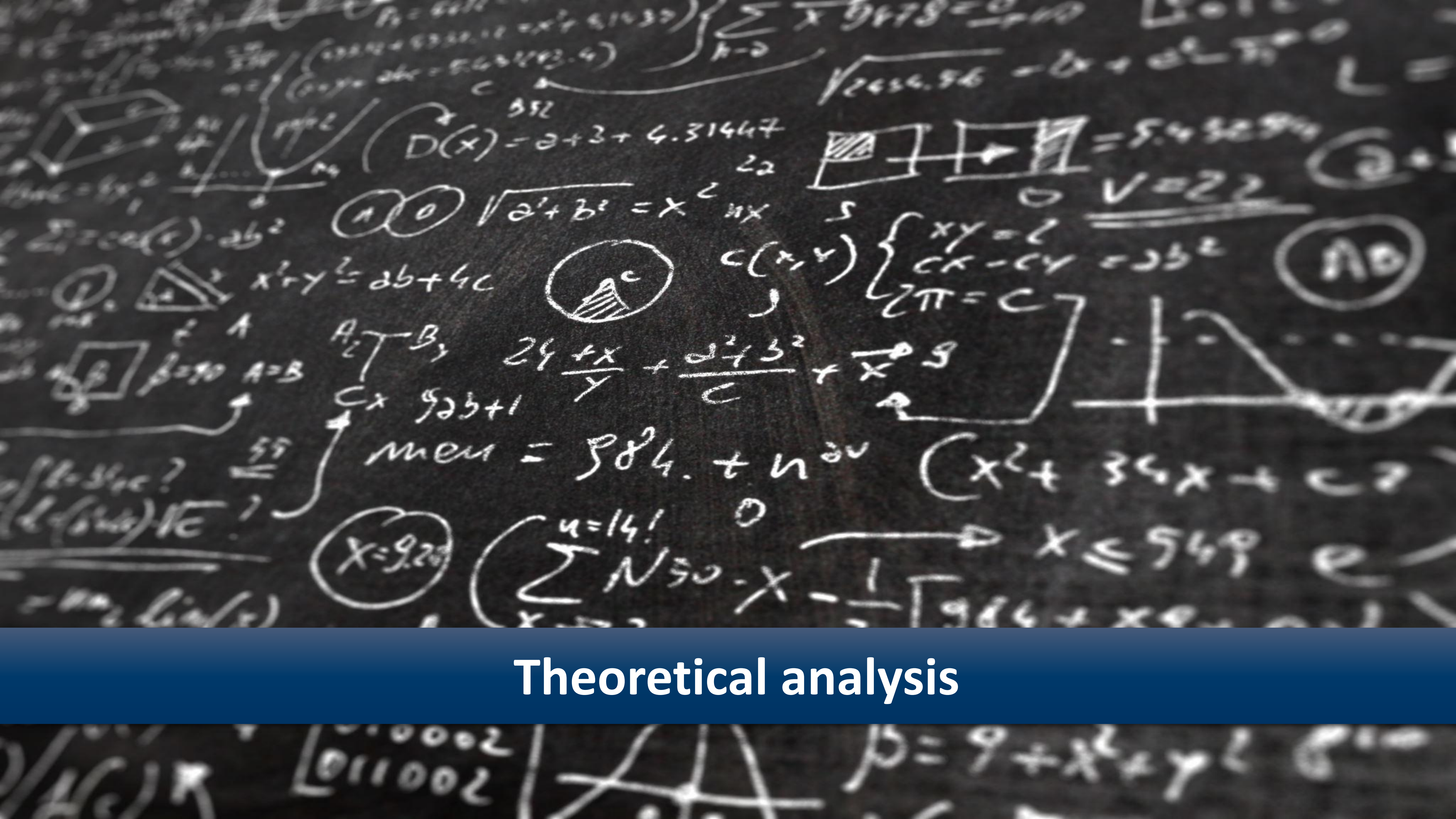
Is it correct to study only one perspective?

Let's assume that customers invest intelligently….

- Then, predicting their future investments would lead to high profitability
- And therefore, correlation between transaction and profitability-based metrics should be high

If correlation is high, we would only care about transaction-based metrics

RQ1. Can we indistinctively use transaction-based and profitability-based metrics for evaluating financial asset recommendations?



Theoretical analysis

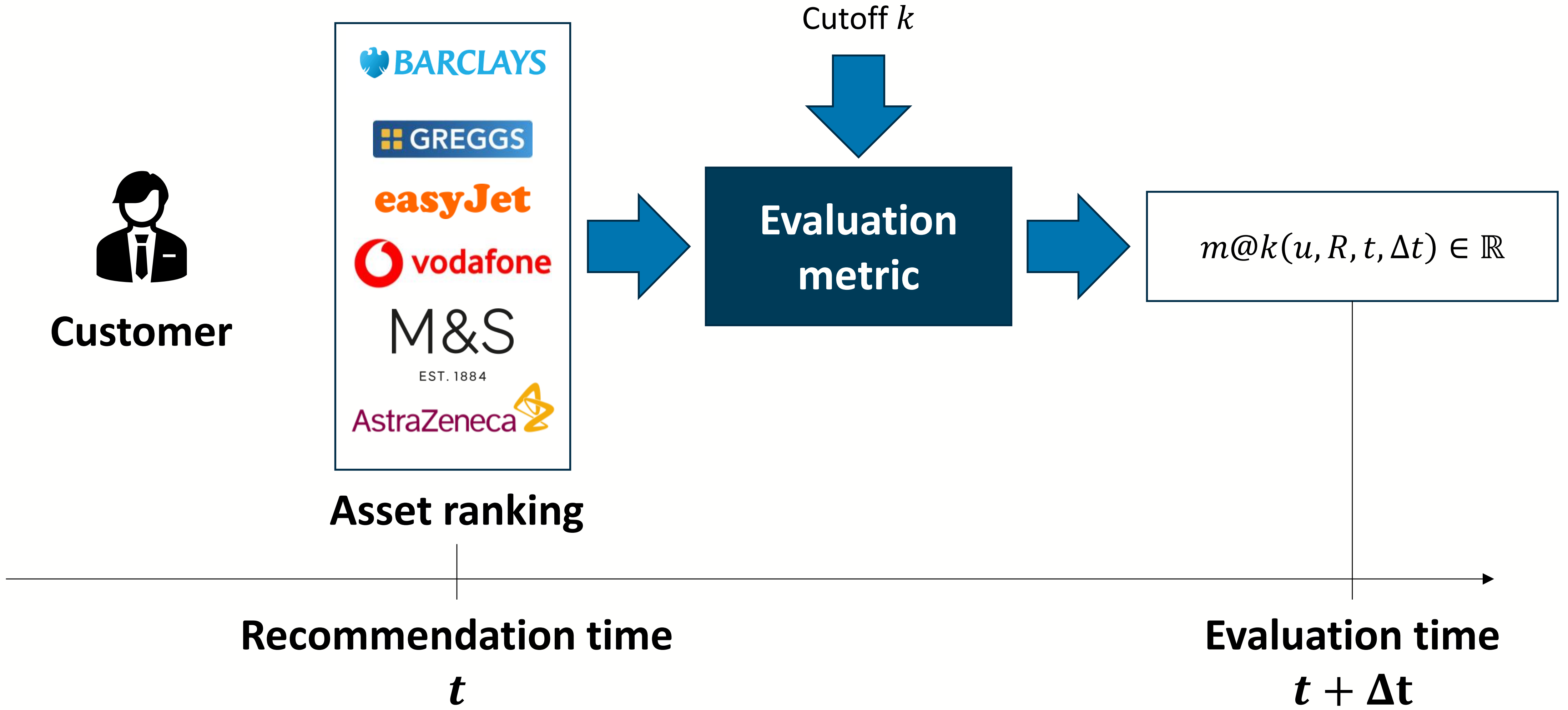
Theoretical comparison

What is the theoretical correlation between profitability-based and transaction-based metrics?

- We compute the correlation between **any pair** of metrics coming from these two families.

- Procedure:
  1. Define what we mean by evaluation metric.
  2. Define the properties of transaction-based metrics.
  3. Define the properties of profitability-based metrics.
  4. Compute their correlation.

What is an evaluation metric?



Customer

Asset ranking

Cutoff k

Evaluation metric

$m@k(u, R, t, \Delta t) \in \mathbb{R}$

Recommendation time
t

Evaluation time
$t + \Delta t$

Transaction-based metrics

Can we predict the preferences of retail investors?

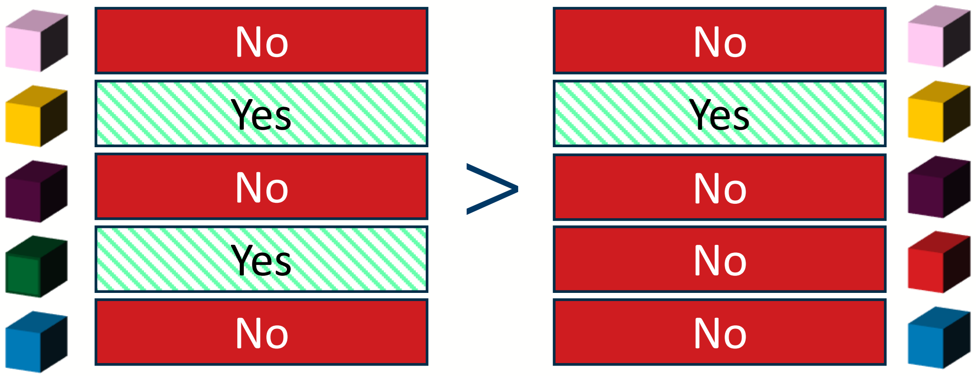
- Examples: P@k, nDCG@k

- Based on the concept of relevance
- Only based on customer actions

- We consider that an asset i is relevant for a customer u in the $[t, t + \Delta t]$ period if and only if:
  1. User u has not invested in i before time t.
  2. User u invests in i after time t, and before $t + \Delta t$.

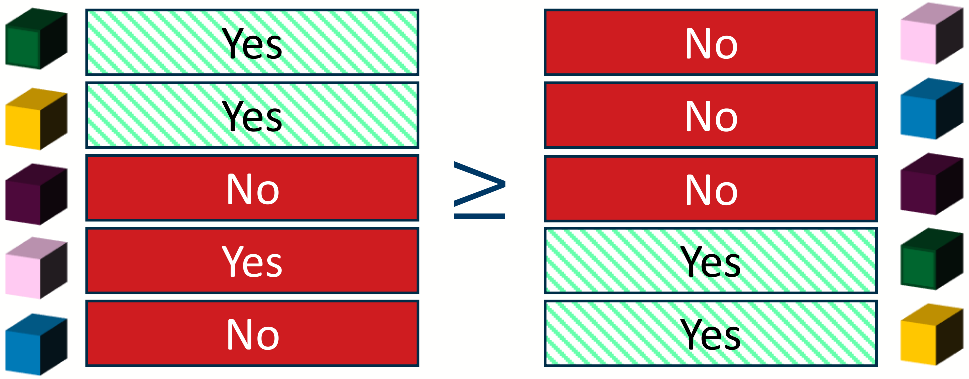
Properties of transaction-based metrics

Convergence



It’s better to have more relevant assets in the ranking

Top-weightedness

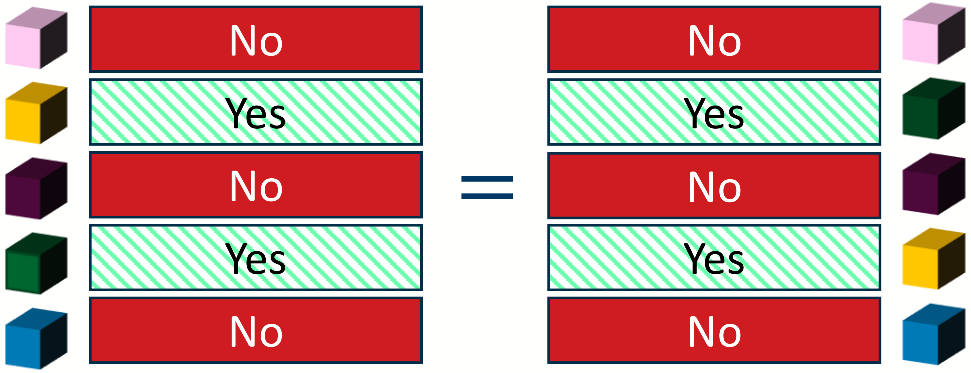


It’s better to have relevant assets in the same positions

🟩 Relevant

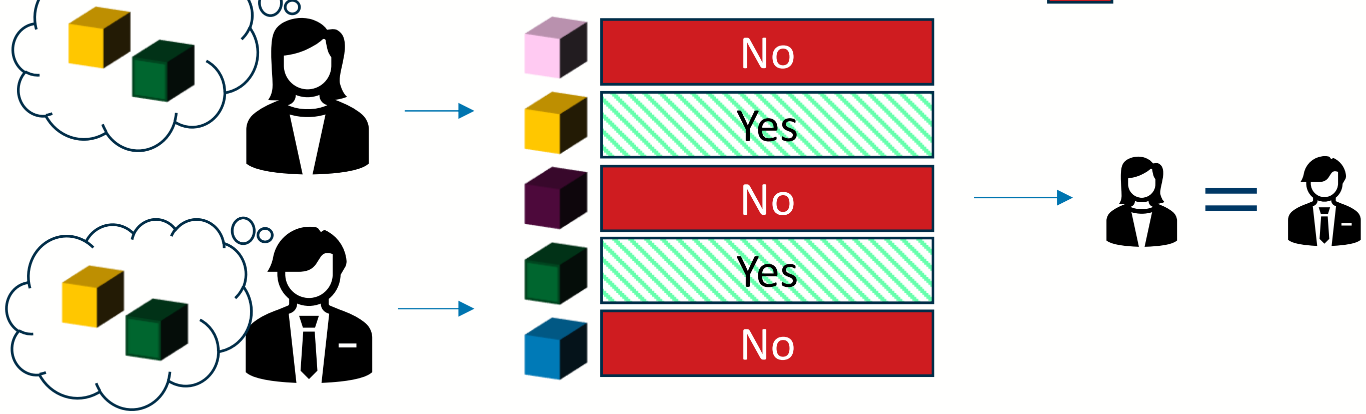
🟥 Not Relevant

Asset identity independence



Only the relevance matters, not the identity of the asset

Customer equivalence



If the same ranking is shown to two users who have same preferences, metric does not change

Profitability-based metrics

Do our customers earn money?

- Examples: Return on investment@k, Net profit@k

- Aligned with customer interests (earn money)
- Ignores the actual investments of customers

- We consider that an asset i is profitable for a customer u in the $[t, t + \Delta t]$ period if and only if its price increases between in the $[t, t + \Delta t]$, i.e.:

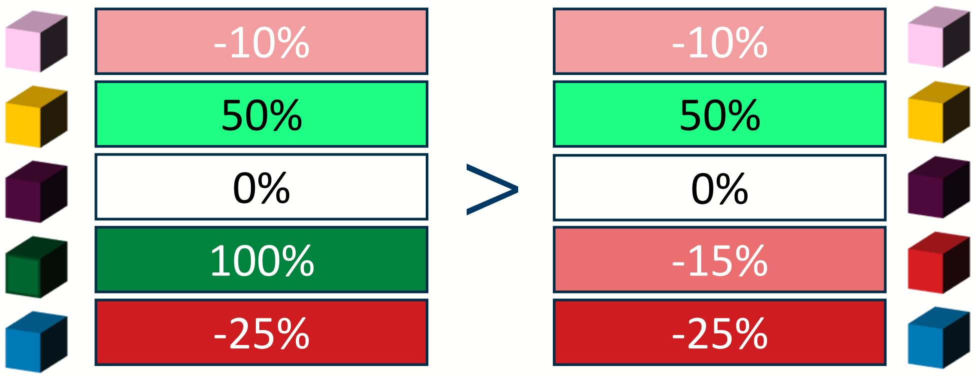
$$\mathrm{price}(i, t) < \mathrm{price}(i, t + \Delta t)$$

- **Profitability is graded:** the bigger the difference, the more the profitability.

University of Glasgow

Properties of performance-based metrics

Convergence



-10%		-10%
50%		50%
0%	>	0%
100%		-15%
-25%		-25%

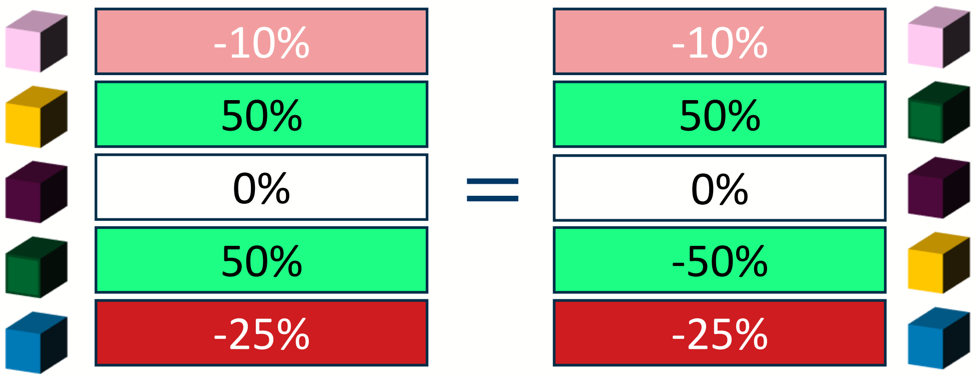
It's better to have more profitable assets in the ranking

Top-weightedness

100%		-25%
50%		-10%
0%	≥	0%
-10%		50%
-25%		100%

It's better to have more profitable assets in the top positions

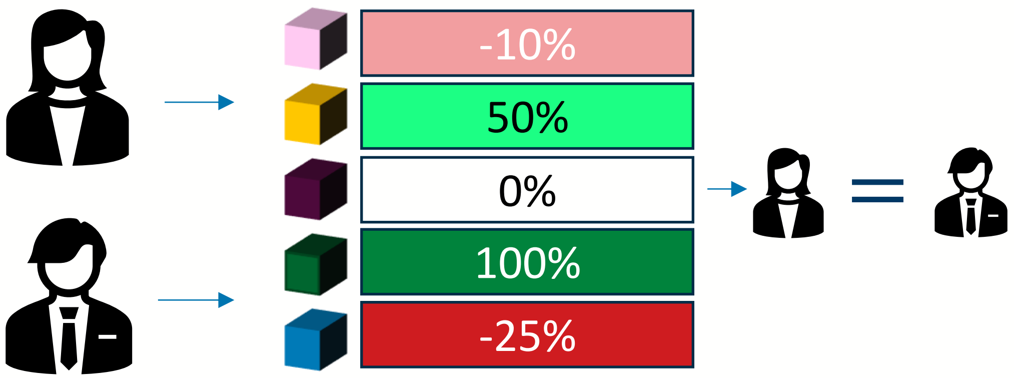
Asset identity independence



-10%		-10%
50%		50%
0%	=	0%
50%		-50%
-25%		-25%

Only the profitability matters, not the identity of the asset

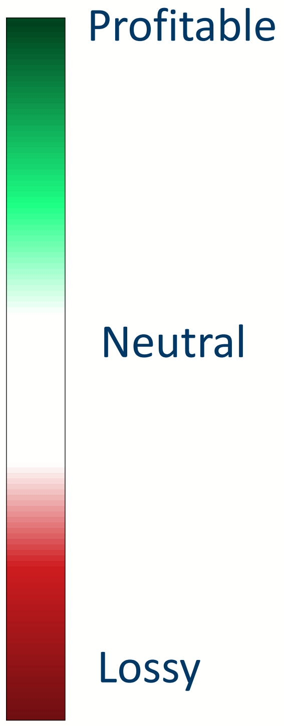
Customer independence



-10%
50%
0%
100%
-25%

=

If the same ranking is shown to two users, metric does not change



Theoretical correlation between metrics

Theorem

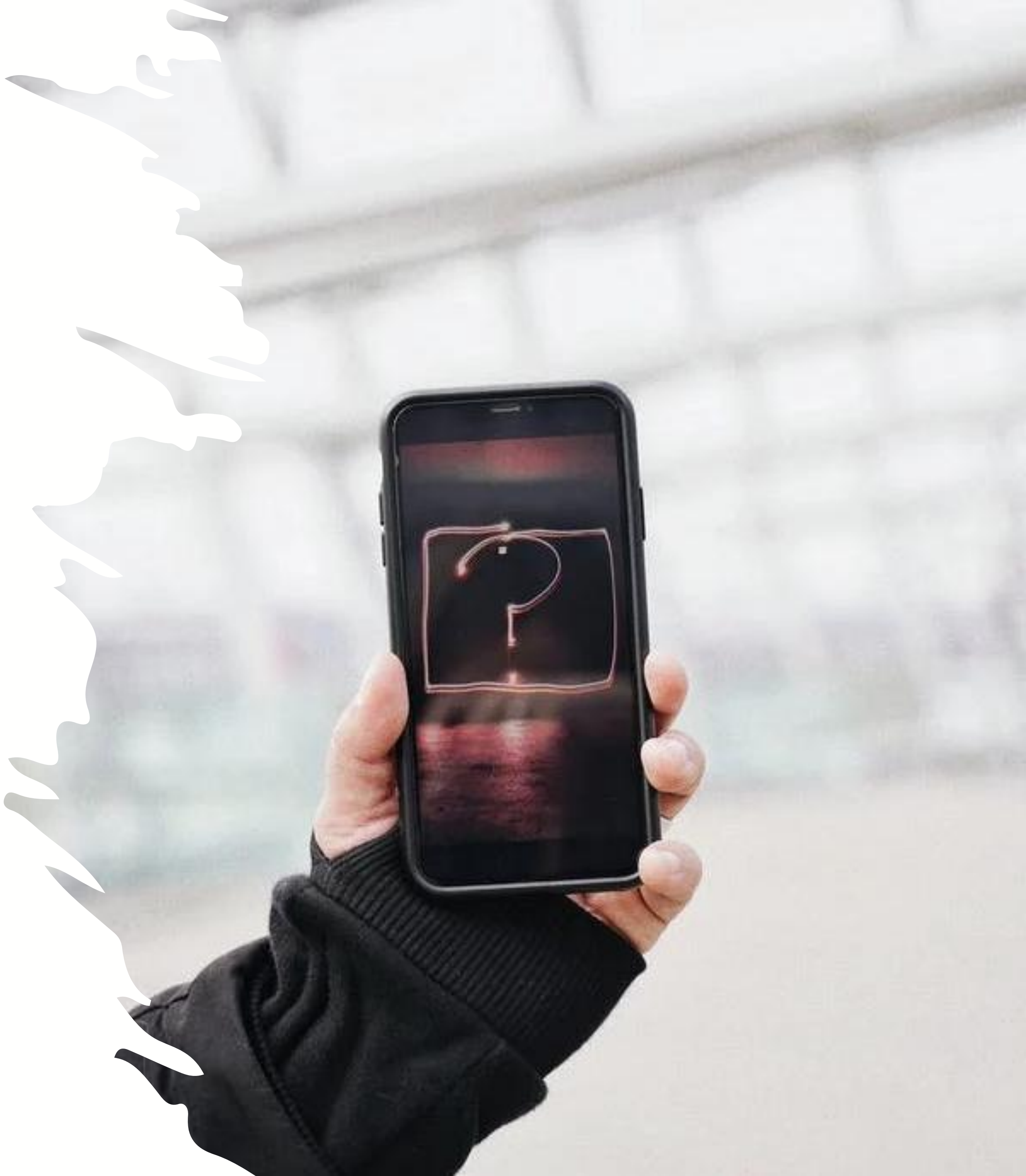
Transaction-based metrics and performance-based metrics are **independent**

- Given a date, an investment horizon
- Correlation over all the possible customers and models is 0

- Transaction-based metrics do not necessarily lead to profit…
- …but they do not lead to losses either.

We cannot theoretically exchange both families of metrics.

Questions?



Dr. Javier Sanz-Cruzado

Financial Recommendation Systems

✉ javier.sanz-cruzadopuig@glasgow.ac.uk

Empirical analysis

Empirical analysis

Hey, we have already seen that metrics are not correlated, why do we need to perform an empirical analysis?

- Theorem studies all possible customers / algorithms.
- Real-world datasets only explore a few customers.
- Investors can be subject to biases.
  - Popularity of the assets.
  - Knowledge of financial advisors.
  - Interests of financial institutions.
- Recommender systems limit their explorations following data.

We need to confirm our observations empirically

Research questions

RQ1 Can we indistinctively use transaction-based and profitability-based metrics for evaluating financial asset recommendations?

RQ2 Which algorithms optimize transaction-based metrics?

RQ3 Which algorithms optimize profitability-based metrics?

Dataset: FAR-Trans

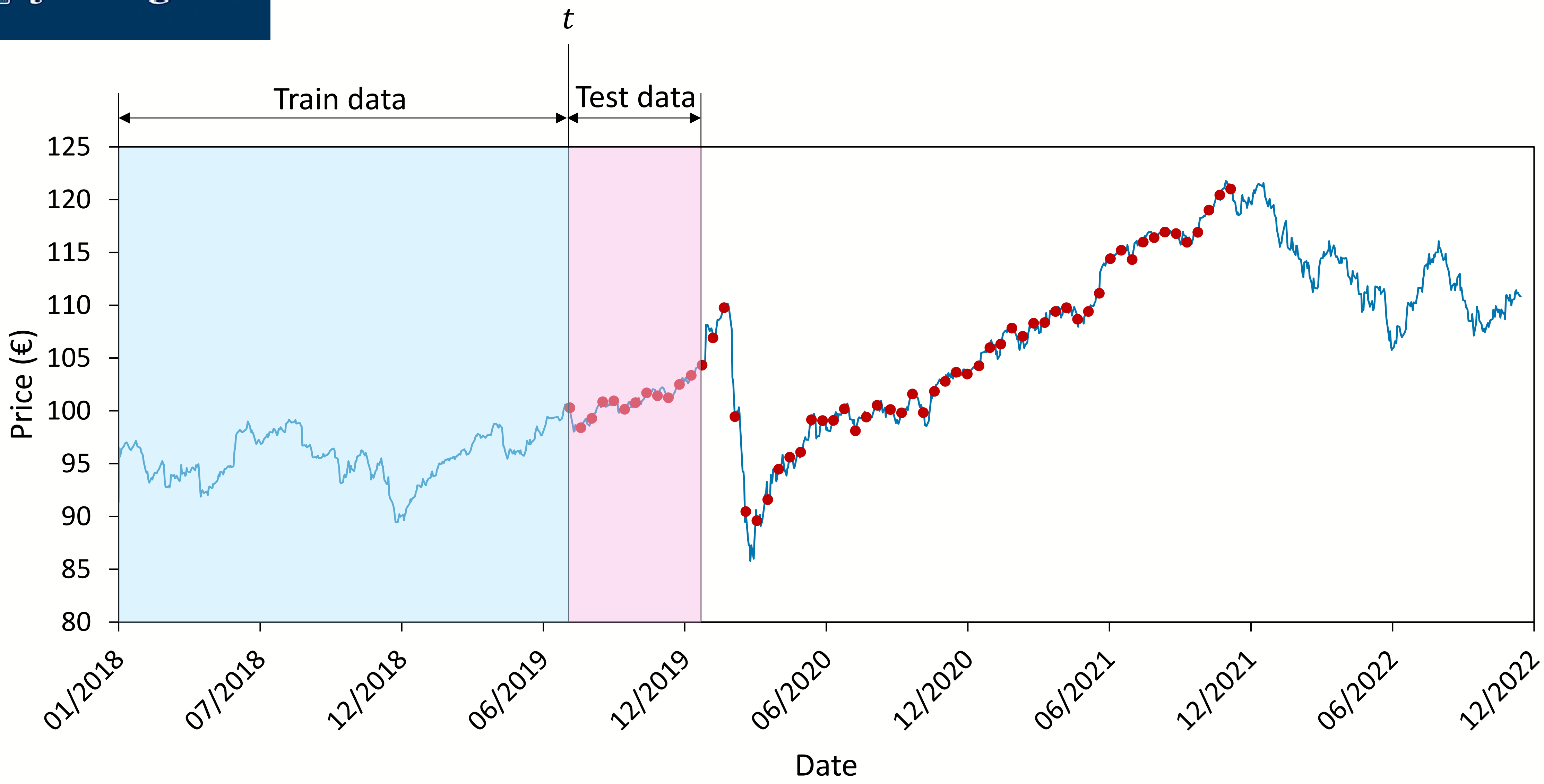
- **Greek market:** stock, bonds, mutual funds
- **Period:** 1st January 2018 – 30th November 2022

- **Combines:**
  - Time series data (pricing information)
  - Customer investments

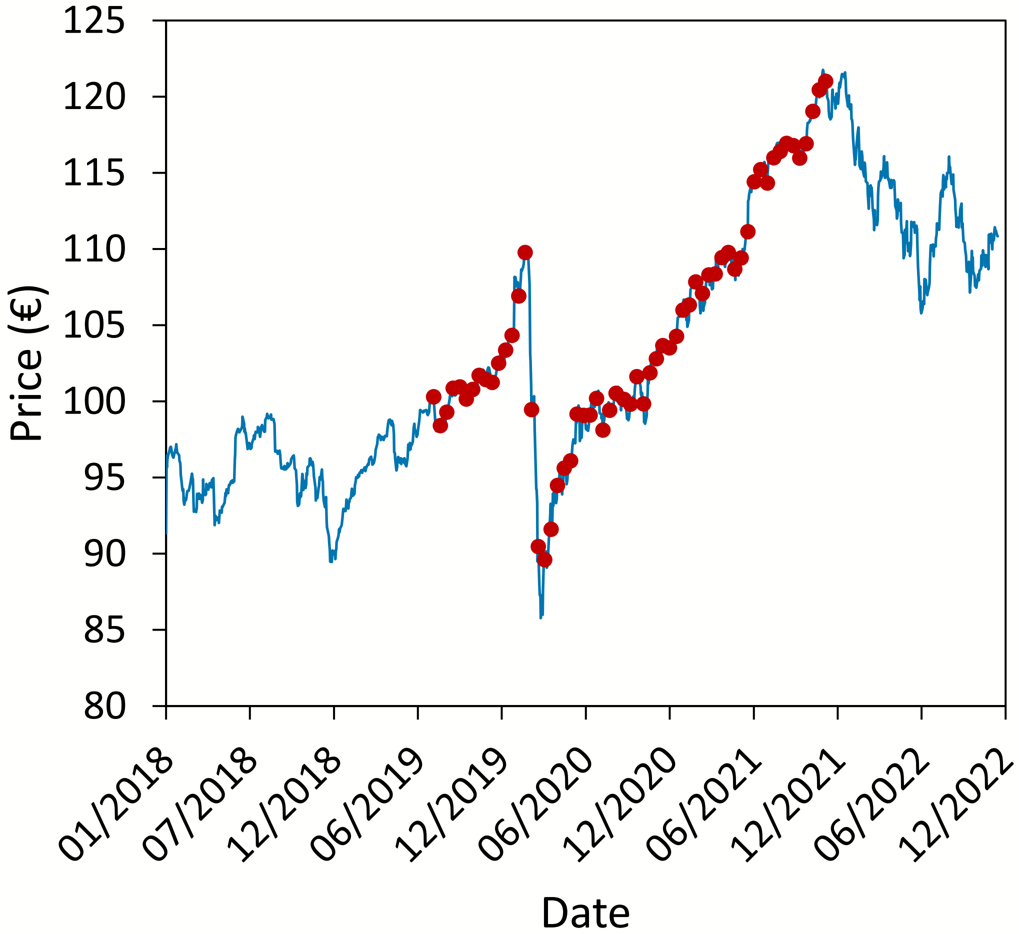
- **Statistics:**
  - 806 unique assets (321 with investments)
  - 29,090 unique customers
  - 703,303 price time points
  - 388,049 transactions (154,103 unique)

J. Sanz-Cruzado, N. Droukas, R. McCreadie. **FAR-Trans: An Investment Dataset for Financial Asset Recommendation**. IJCAI-2024 Workshop on Recommender Systems in Finance (Fin-RecSys), Jeju, South Korea, August 2024.

Dataset split



Dataset split



- **Total:** 61 time points

- **Length of test period:** 6 months

- **Starting date:** 1st August 2019
- **Ending date:** 23rd November 2021

- Varying market conditions
- Including Covid-19 period
- And 2022 market downturn

Metrics

- **Profitability-based:** monthly return on investment (Monthly ROI@10)
  - Relative increase of price w.r.t. the initial investment after some time Δt
  - Initial price: price at recommendation time
  - Final price: price at recommendation time + Δt
  - $\Delta t = 6$ months
- **Transaction-based:** nDCG@10
  - Higher nDCG indicates our model predicts future customer investments
  - Ranking-based IR/RecSys evaluation metric
  - Relevant transactions
    - New asset acquisitions (buys)
    - Up to 6 months after recommendation

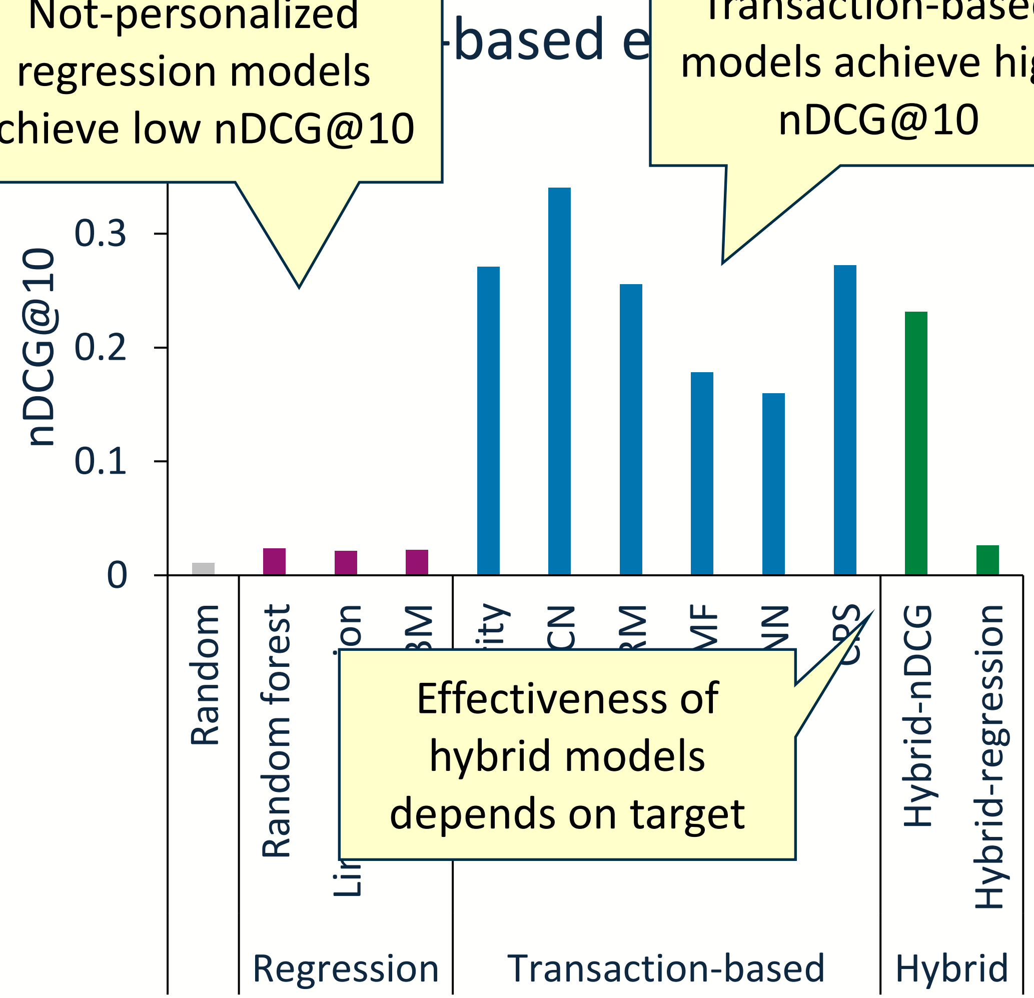
Algorithms

- **Profitability-based regression models**
  - Linear regression
  - Random forest
  - LightGBM

- **Transaction-based models**
  - **Not personalized:** popularity-based
  - **Collaborative filtering:** LightGCN, MF, UB kNN, association rule mining
  - **Demographic methods:** UB kNN with customer information

- **Hybrid:** using as features all the previous models,
  - LightGBM regression
  - LightGBM learning to rank (LambdaMART)

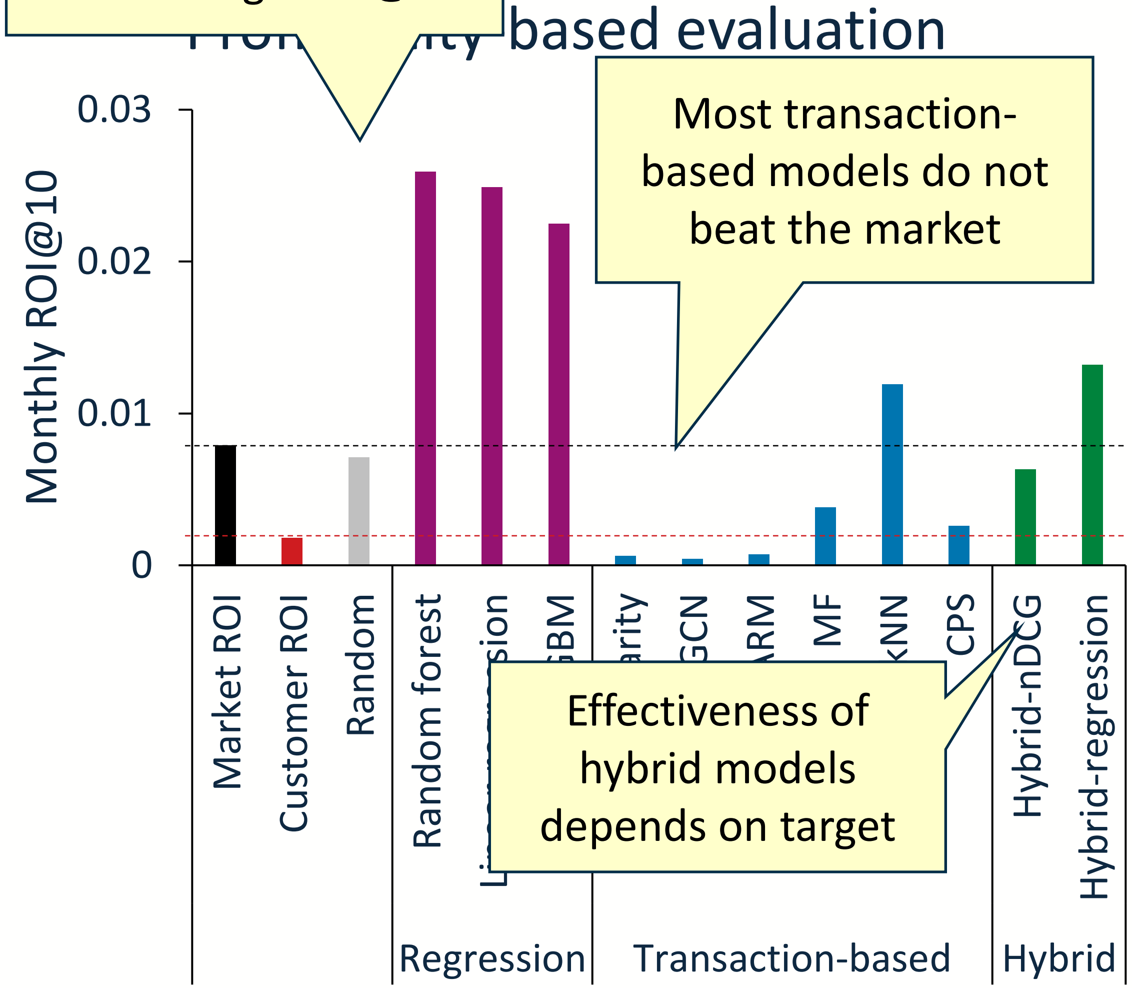
Results



Not-personalized regression models achieve low nDCG@10

Transaction-based models achieve high nDCG@10

Effectiveness of hybrid models depends on target



Not-personalized regression models achieve high ROI@10

Most transaction-based models do not beat the market

Effectiveness of hybrid models depends on target

Results

RQ2 Which algorithms optimize transaction-based metrics?

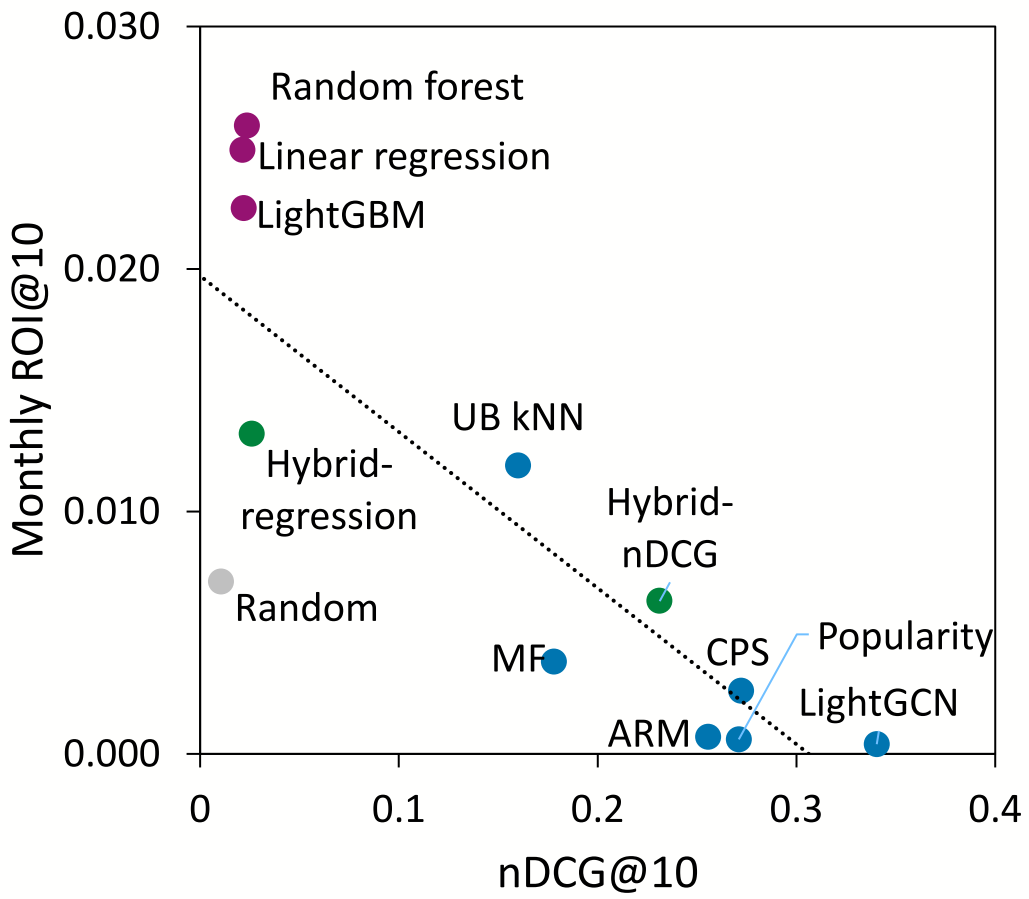
- Personalized transaction-based models optimize nDCG@10
- **Best model:** LightGCN

RQ3 Which algorithms optimize profitability-based metrics?

- Not-personalized profitability prediction methods optimize monthly returns @ 10
- **Best model:** Random forest regression

Comparison between metrics

What is the empirical correlation between metrics?



- Pearson correlation over all customer, date, algorithm triplets: **-0.13**
- Correlation between metrics is **negative**!
- If we improve future investment prediction, that could lead to losses!

We cannot exchange both families of metrics.

Factors affecting correlations

RQ4. What factors affect correlation between metrics?

1. Ability of customers to profit from market
2. Changes in market conditions
3. Customer investment holding time

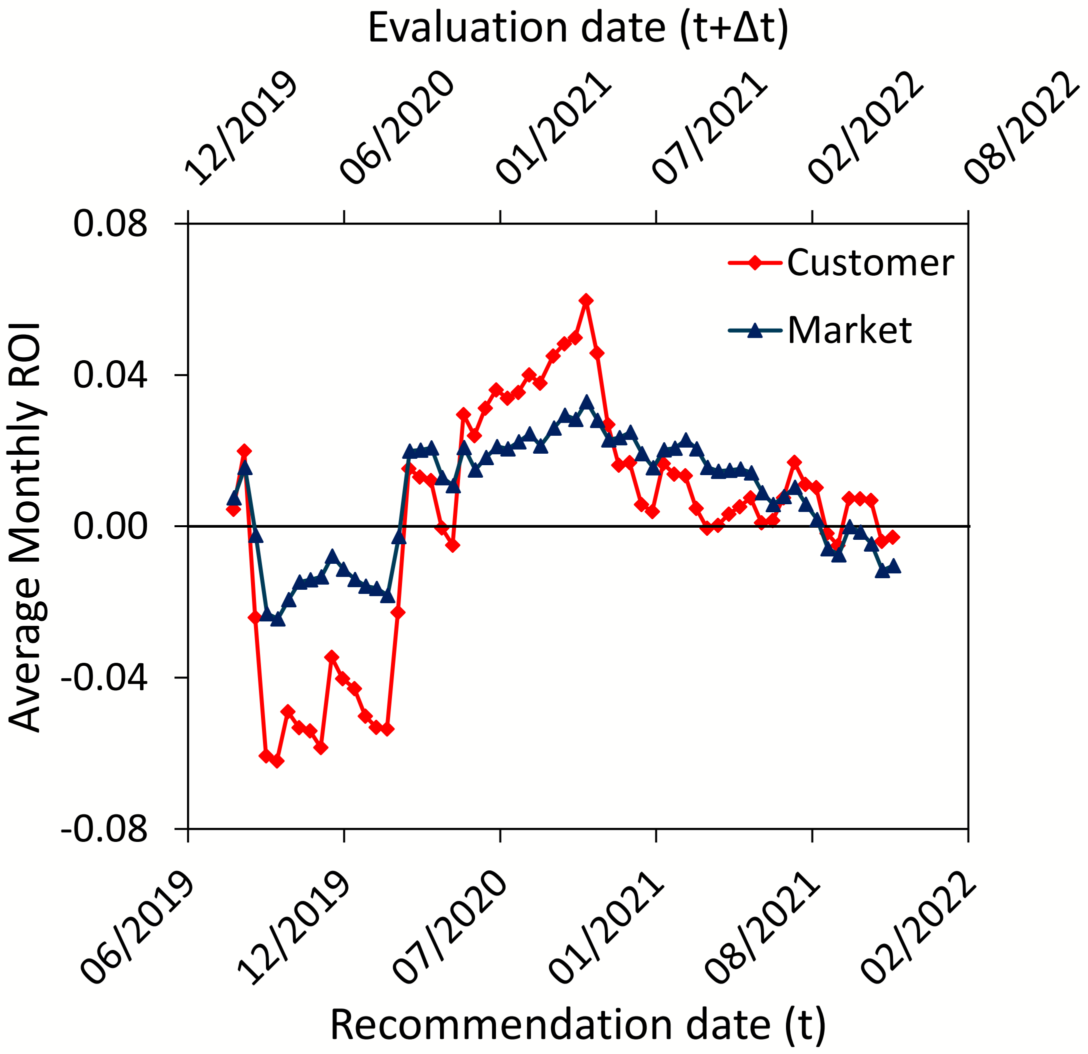
Ability of customers to profit from market

- **Previous hypothesis**
  - If customers invest intelligently….
  - Then, predicting their future investments would lead to high profitability
  - And therefore, correlation between both evaluation metrics should be high

- **But correlation is negative…**

Are our customers effective investors?

Do customers earn money?



- **Time horizon:** $\Delta t = 6$ months
- **Overall: No**
  - Market 0.79% Monthly ROI
  - Customers 0.18% Monthly ROI
- **Over time:** Depends on the chosen date

Then, is our initial hypothesis true?

Hypothesis testing

- **Simulation**
  - Create effective synthetic customers
  - Substitute the real customers by them
  - Re-run the experiments over them

- **Synthetic customer procedure creation**
  1. Estimate number of customers
  2. For each customer
     a) Choose the number of assets on which to invest
     b) Choose the time points of the investment
     c) Choose the assets on which to invest

- **Repeat the process 10 times**

Synthetic customer creation

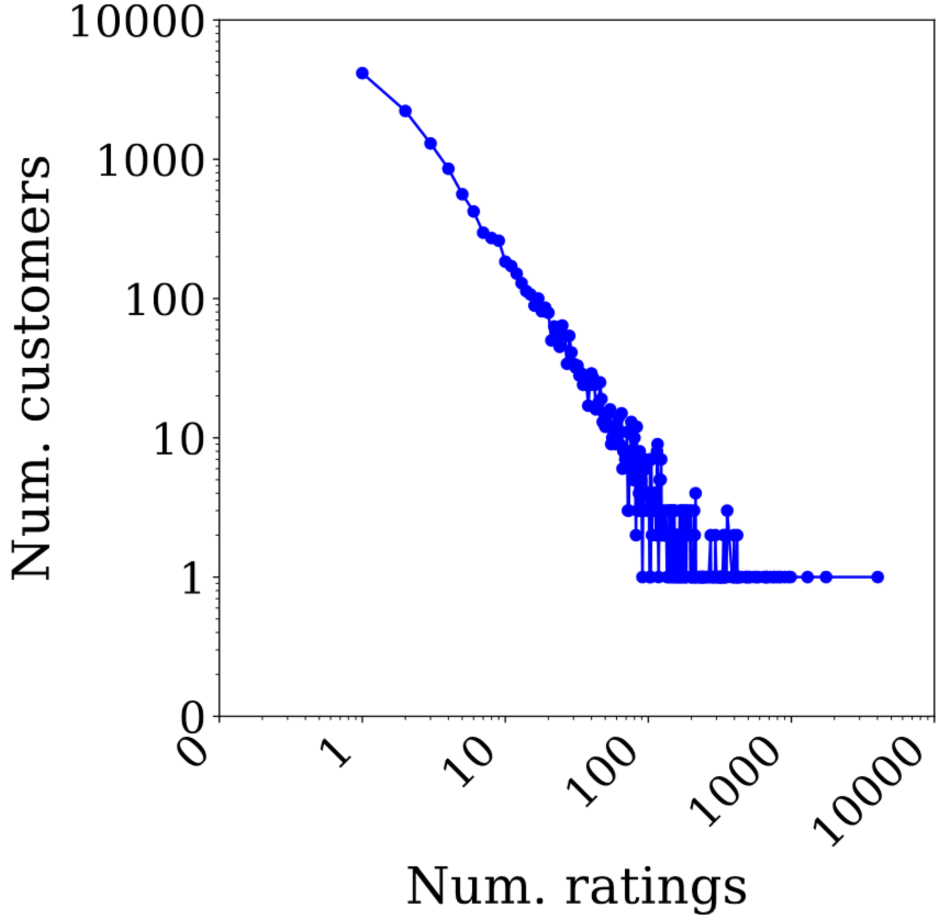
1. **Choose number of customers**
   - Same as in the real dataset: 29,090

2. **Choose number of assets on which a customer invest**
   a) Mimick the distribution of the original data
   b) We use a Gamma distribution $\Gamma(k, \theta)$
   c) Choose randomly the number of investments $n \sim \Gamma(k, \theta)$

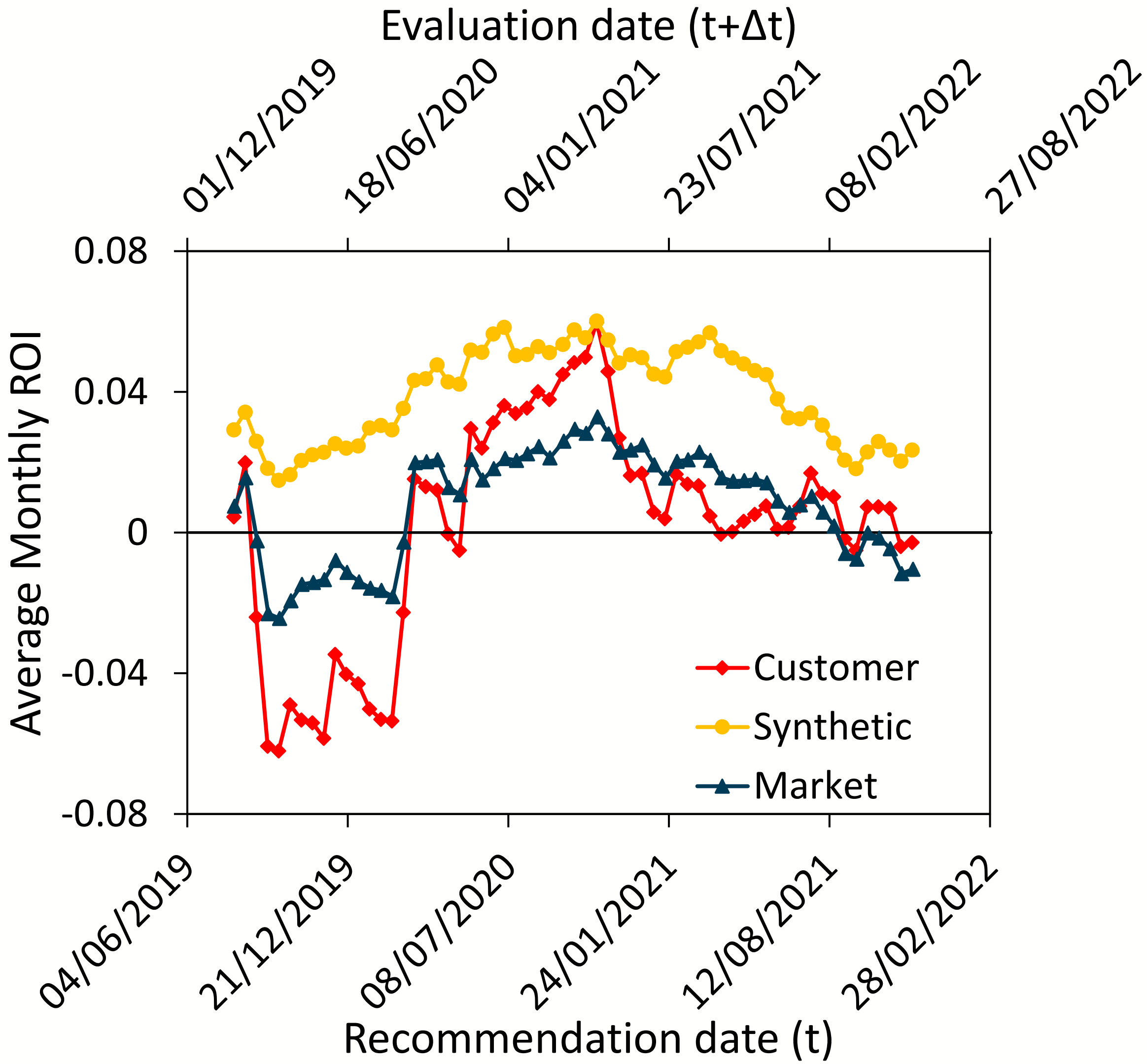
3. **Choose the time points of the investment**
   - Uniformly between January 1st 2018 and November 30th 2022

4. **Choose the assets**
   a) Choose among the top-50 most profitable assets between t and $t + \Delta t$ (Δt = 6 months)
   b) Choose proportionally to ROI

FAR-Trans dataset ratings distribution



Synthetic dataset statistics

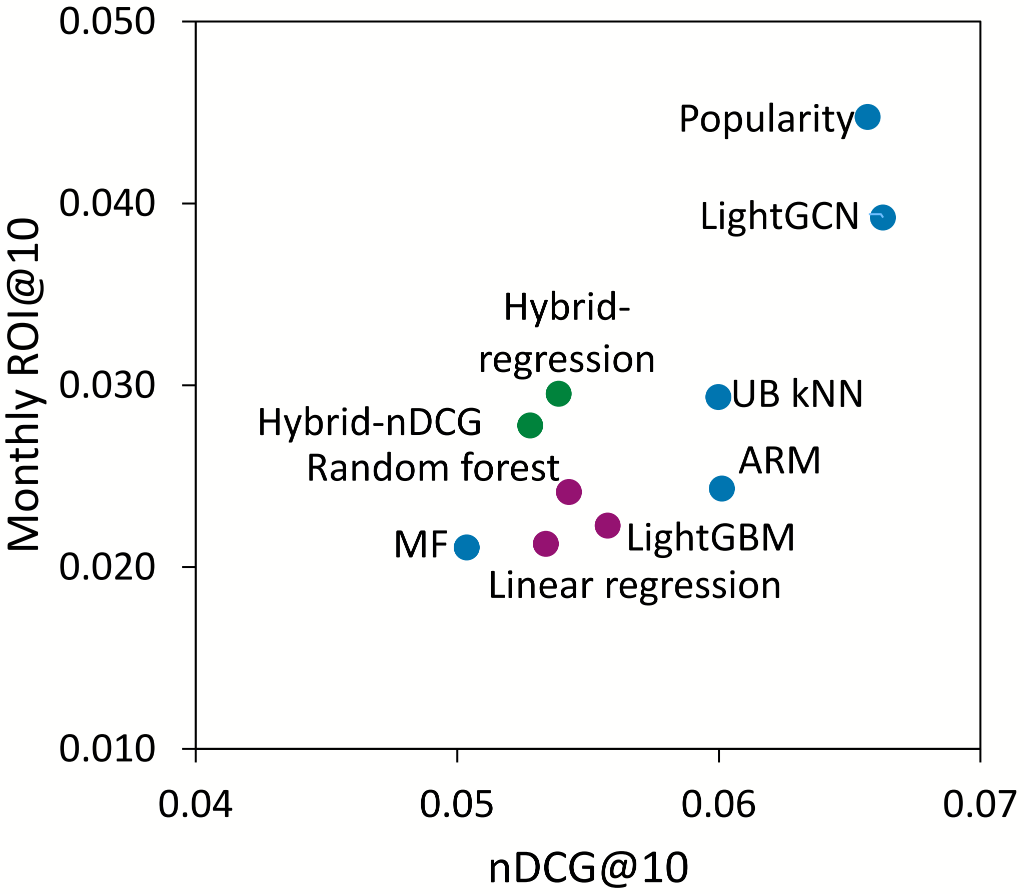


- The synthetic customers
  - Always beat the market
  - Always beat the real customers

Does this lead to positive correlation?

Experimental results (Synthetic customers)

What is the correlation between metrics in the synthetic dataset?

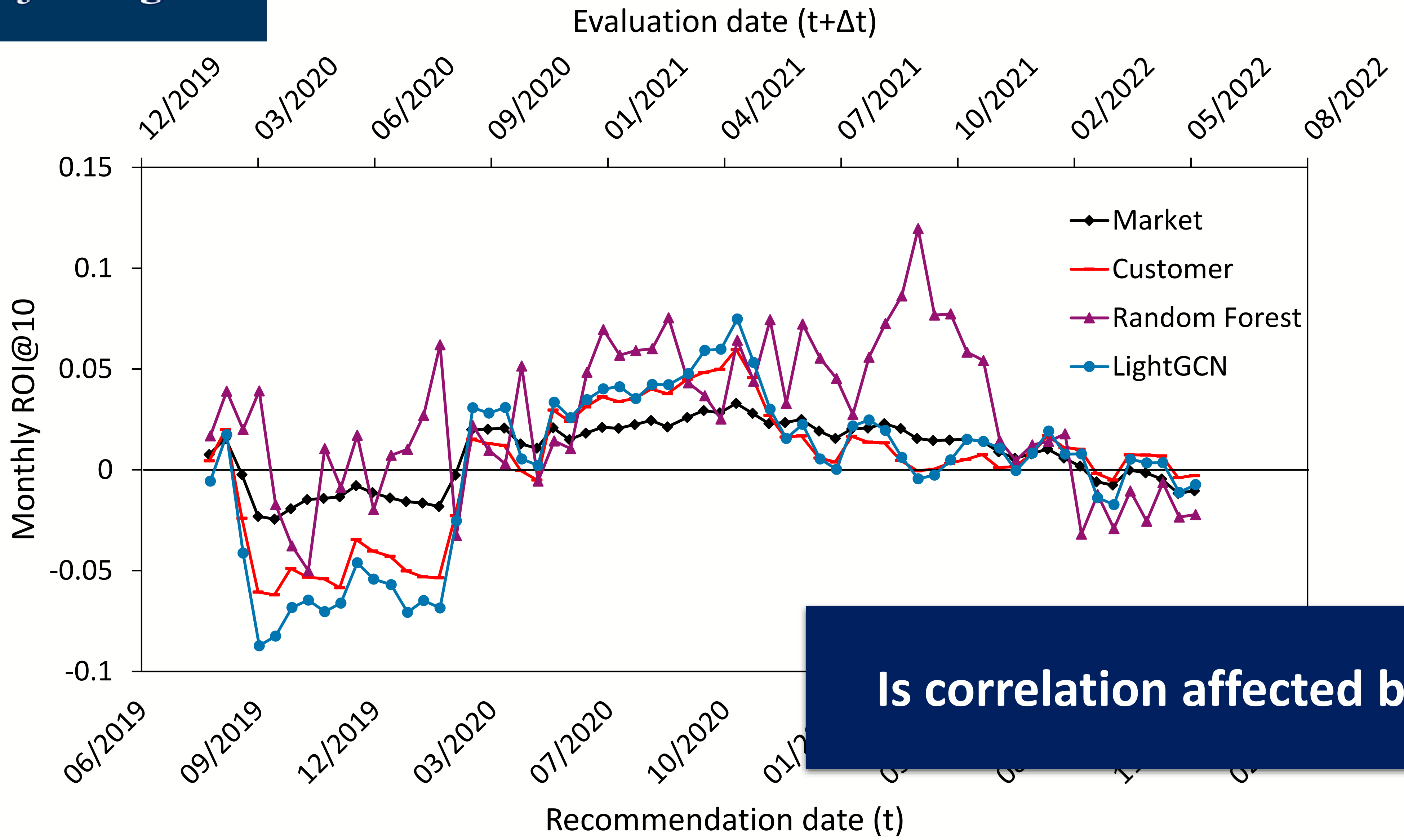


- Pearson correlation over all customer, date, algorithm pairs: **+0.13**
- Correlation between metrics is **positive**!

If customers are good investors, correlation is positive

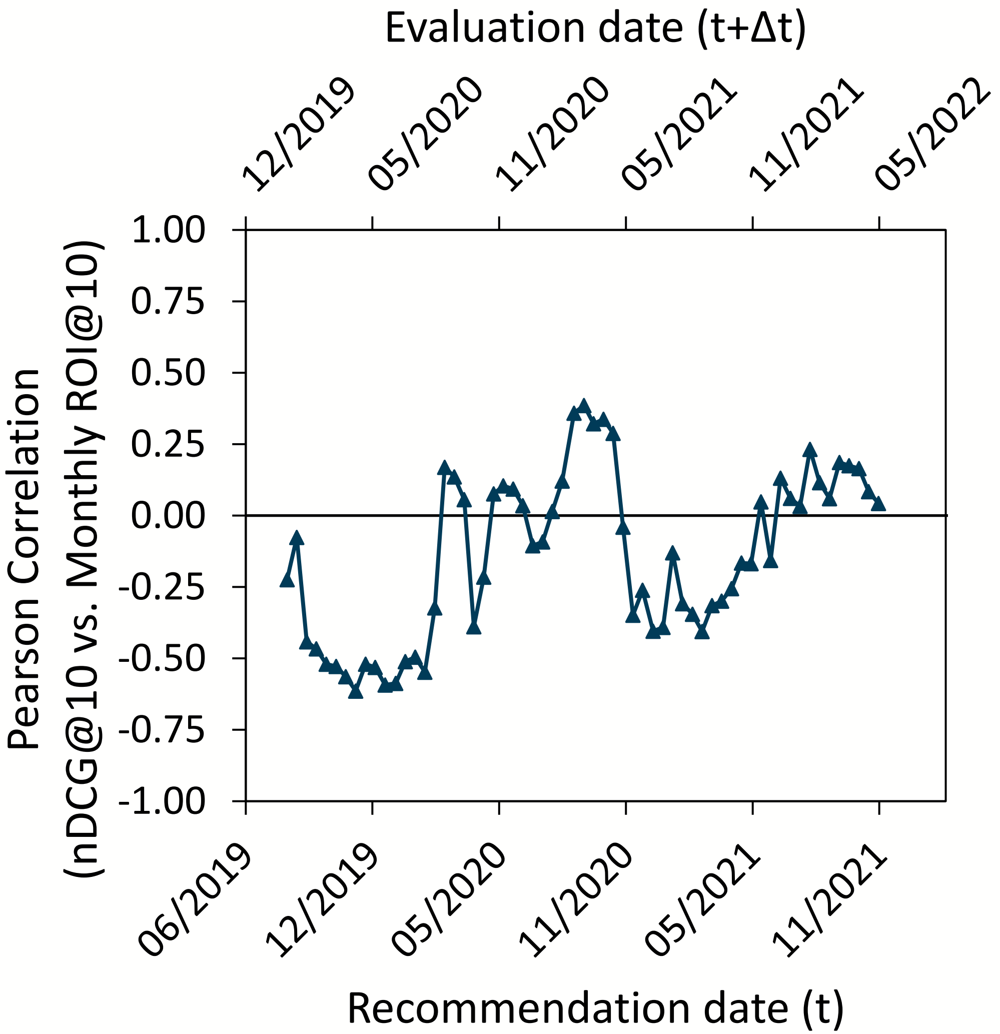
However, our customers are not always good investors

Changes in market conditions



Is correlation affected by time?

Correlation over time



- Correlation changes notably over time
- Between -0.5 and 0.5!

- Computing correlation over multiple dates hides these variations!

- Therefore, recommendation time affects the correlation

What is causing that?

Market ROI	**Customer ROI**	**Customer vs. Market ROI difference**

What is the cause of the variation in correlation?

- We plot the confusion matrices between correlation and multiple conditions

Market ROI

		Value	
		+	-
Correlation	+	47.5%	33.3%
	-	52.5%	66.7%

Market changes do not correspond to correlation changes

Customer ROI

		Value	
		+	-
Correlation	+	55.0%	19.0%
	-	45.0%	81.0%

When customers are not effective, correlation is negative (does not work when customers are effective)

Customer vs. Market

		Value	
		+	-
Correlation	+	84.0%	13.8%
	-	16.0%	86.2%

Great correspondence between ROI differences and correlation sign!

Changes in market conditions

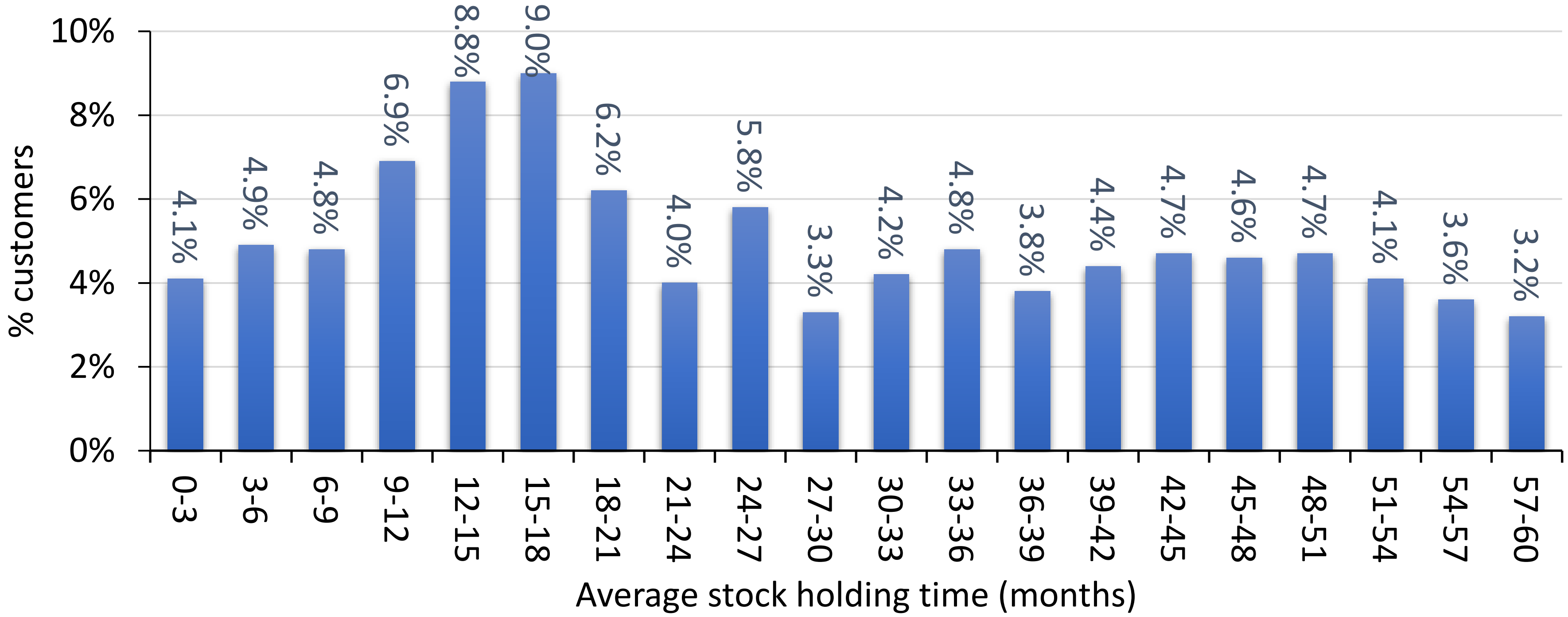
- **Time affects correlation between metrics**

- At different dates, we observe big variations.
- However, pure market conditions do not explain sign changes in correlation.

- Customer ability to beat the market does
  - When customers beat the market, correlation is likely to be positive
  - When customers do not, correlation is likely to be negative

Customer investment horizon

Is six months a reasonable future time target?



- Only 9% customers hold their investments for 6 months or less.
- Investments captured by nDCG might not necessarily align with a 6 month investment horizon.

How do results change for different horizons?

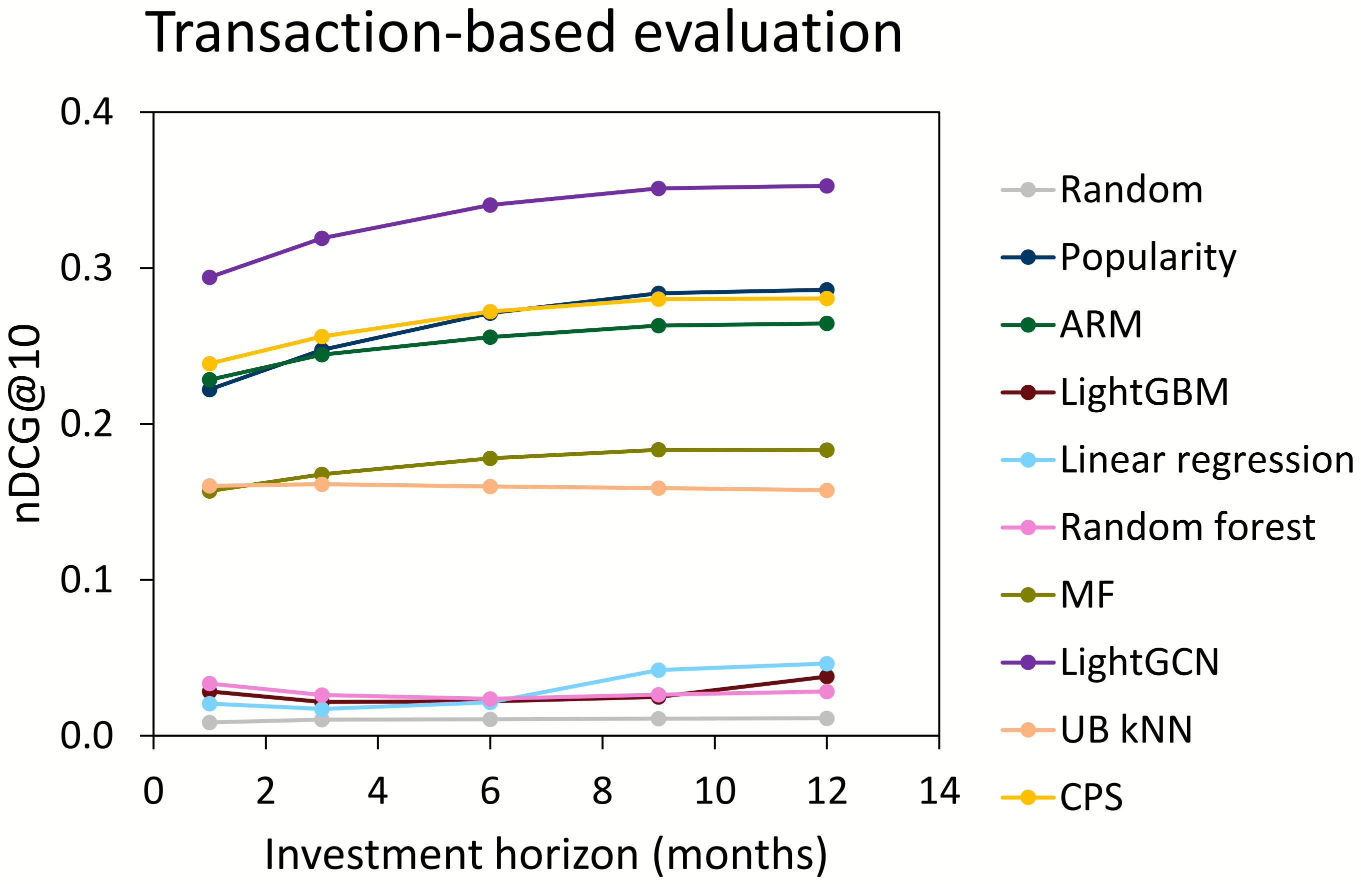
- We repeat our experiments for shorter and longer investment horizons.

- **New horizons:** 1, 3, 6, 9, 12 months

- As we increase the investment horizon:
  - Asset profitability changes.
  - Transactions in the test set increase.

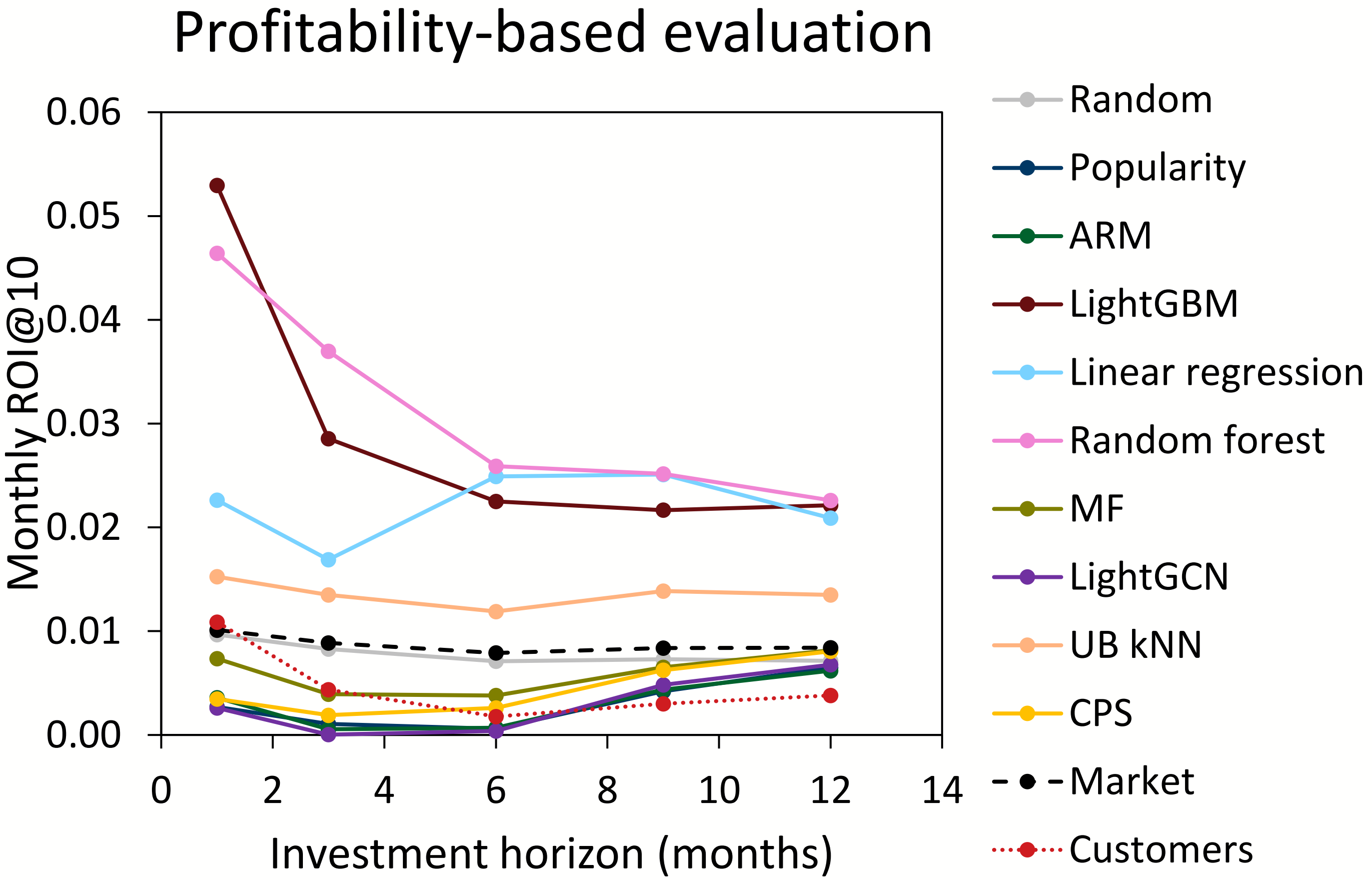
- How does this affect algorithms?
  - **Transaction-based algorithms:** training data is the same for different horizons.
  - **Profitability-prediction algorithms:** training examples are the same, but target changes.

Effectiveness (nDCG@10)



- Transaction-based models
  - As we increase the horizon, we just increase the test set.
  - Therefore, algorithms are just capable of capturing further transactions.
- Profitability-based models
  - Low nDCG@10 values
  - Still not-personalized
  - Rankings change when we modify horizon (target change)

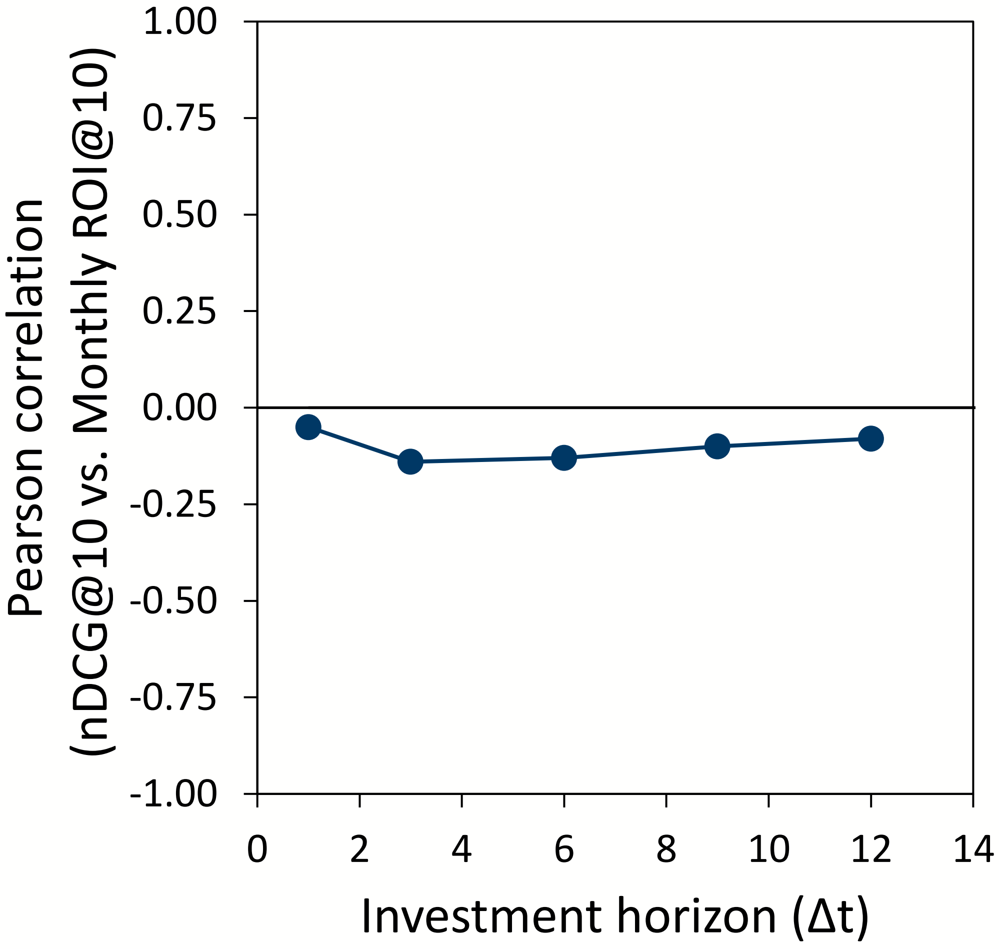
Effectiveness (Monthly ROI@10)



- Transaction-based models
  - Still under-perform the market
  - Exception: UB-kNN
  - Values slightly change

- Profitability-based models
  - Best ROI.
  - Large variations over time.
  - Random forest best overall (4 out of 5 horizons).
  - LightGBM best for 1 month.

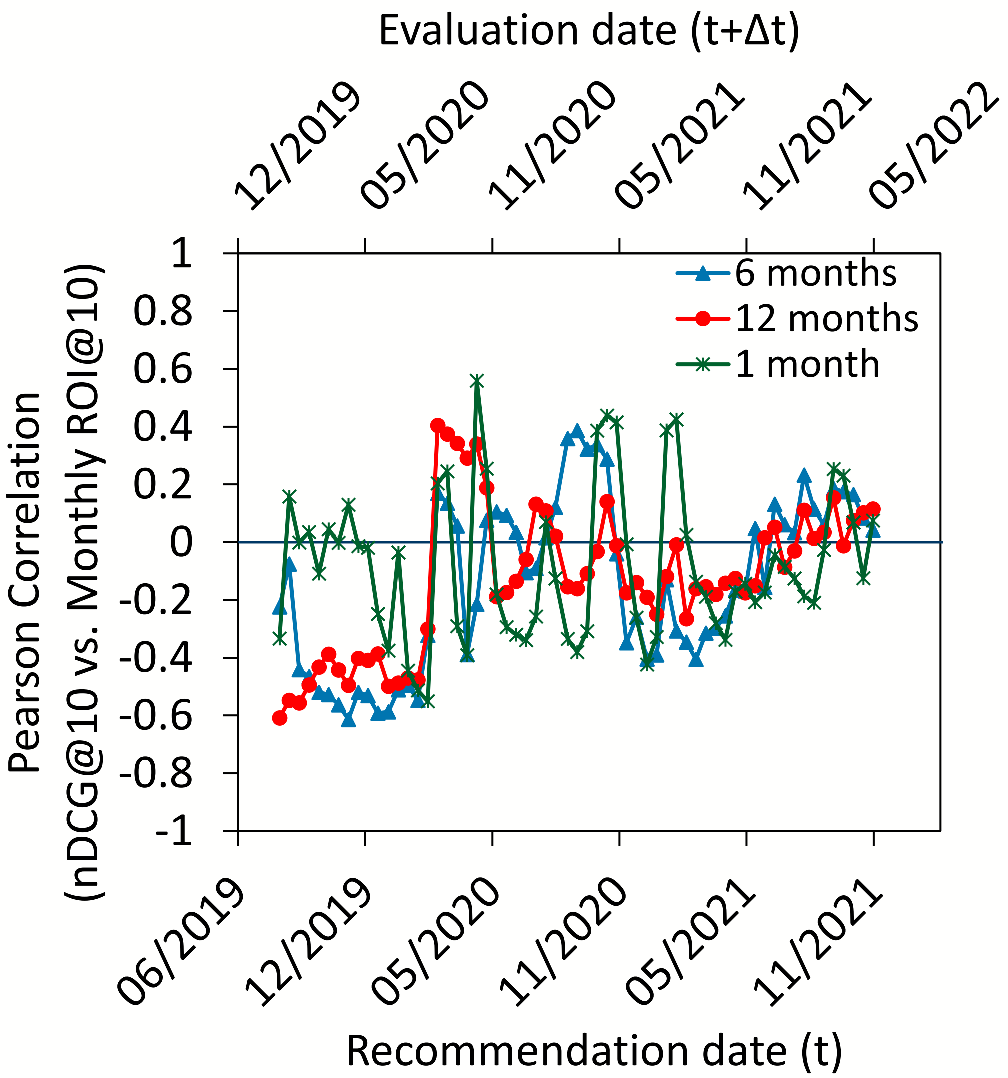
Does the correlation change?



Correlation does barely seem affected by investment horizon

Is that similarity consistent over time?

Analysis over time



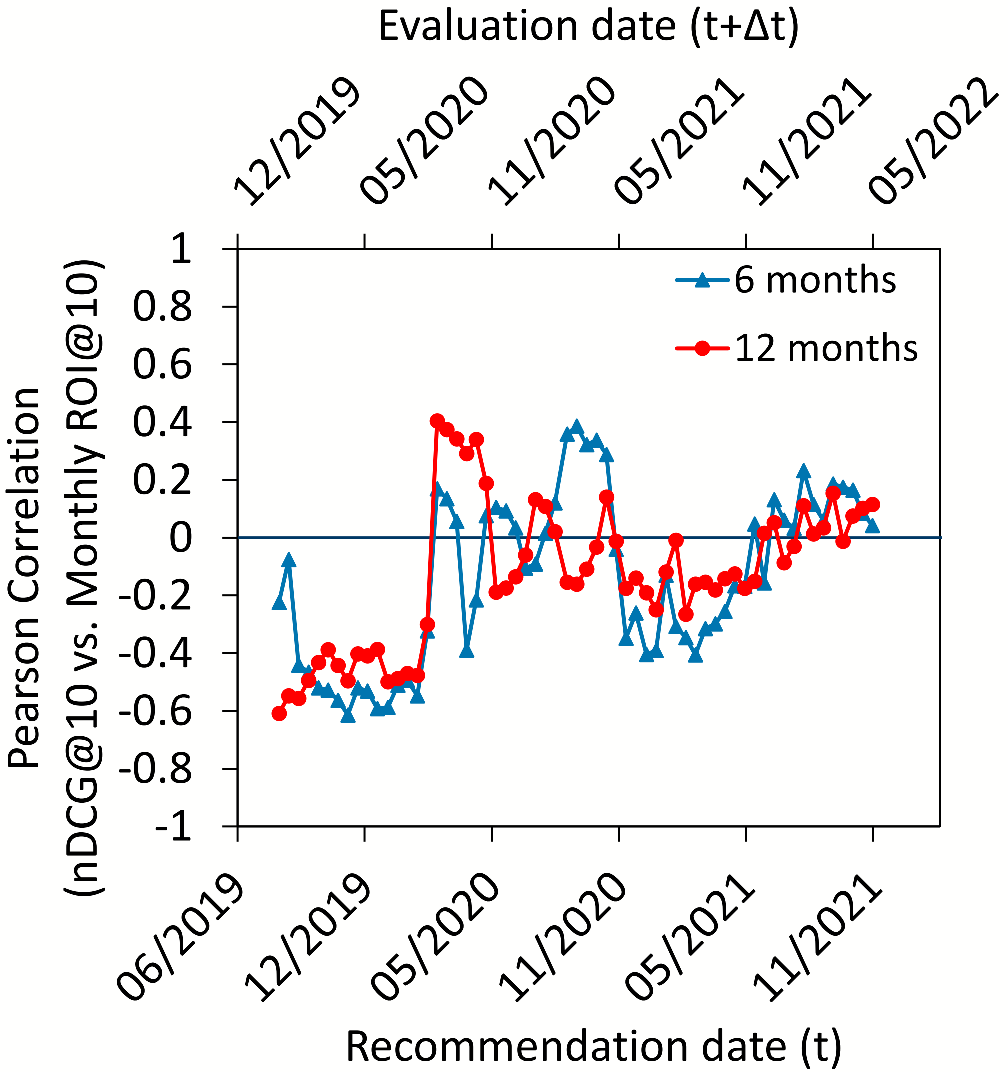
- At the same date, correlation changes notably when we change investment horizon

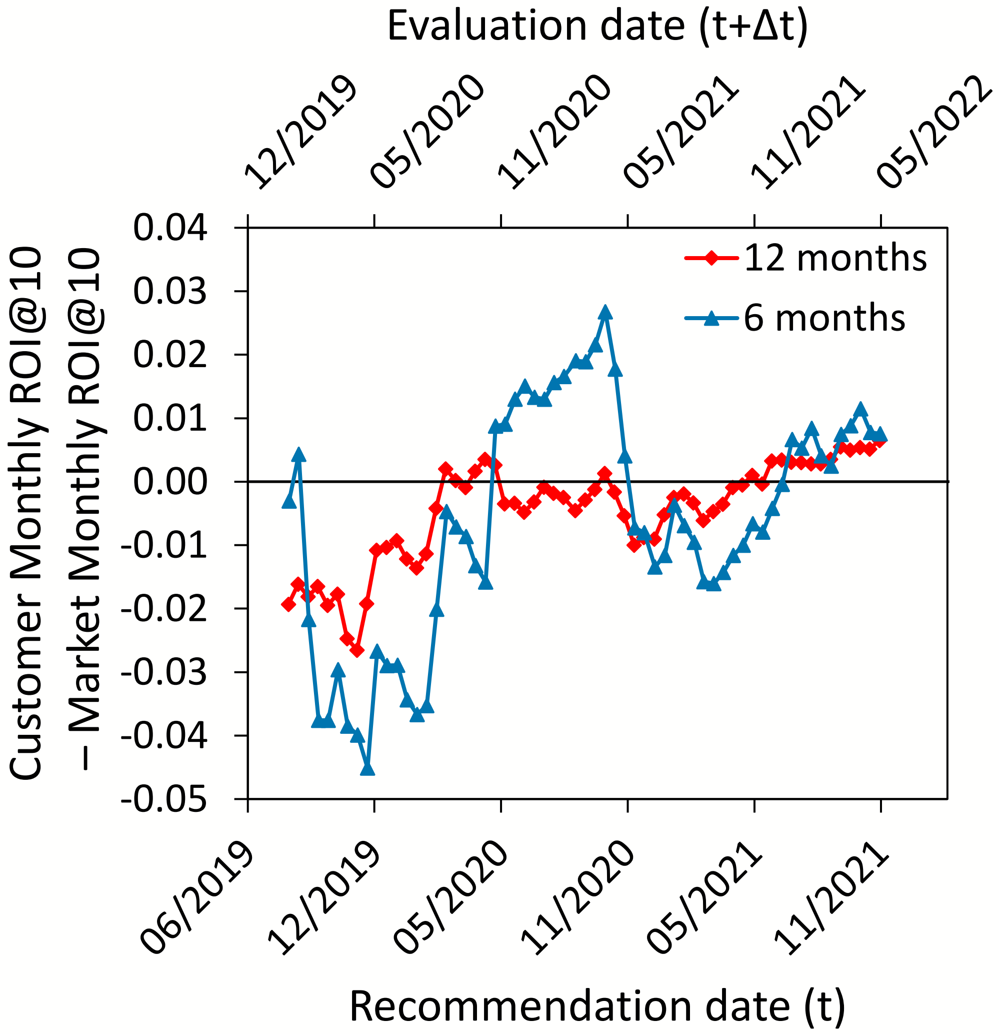
- At one date, we might find positive correlation when in other is negative

- Why?

Analysis over time

Differences between customers and market are different when we look at different horizons!





Investment holding time

- **Globally, correlation is not particularly affected by the investment horizon**

- However, computing correlation under all metrics hides variations

- For a single date, investment horizon has a large effect on correlation!
- We need to consider this individual effects!

- We again, checked that, for every investment horizon, differences between customer and markets explain most of correlation changes

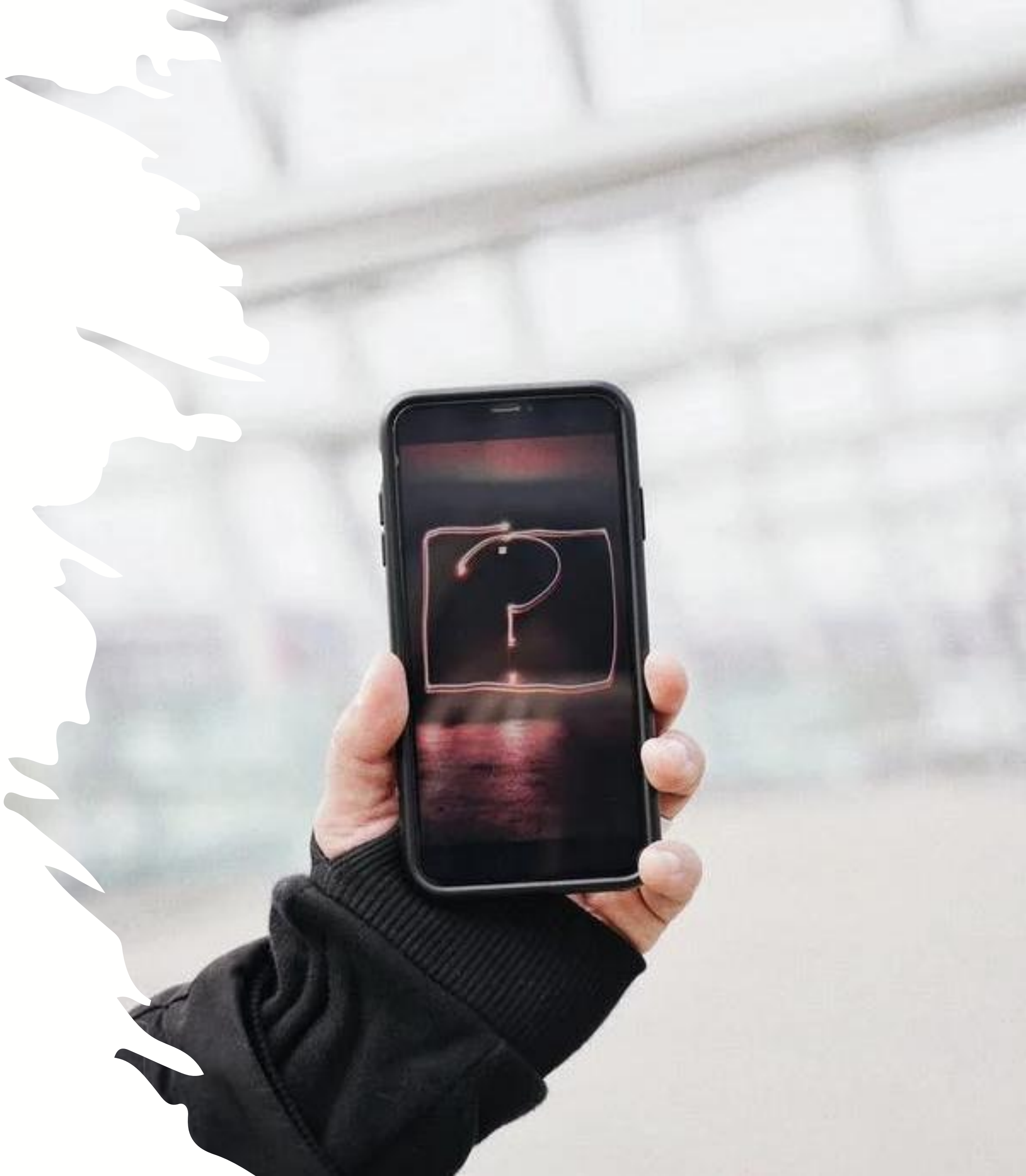
Conclusions

Conclusions

- **We cannot use transaction-based metrics in exchange of profitability-based metrics.**
  - Theoretically, they are independent.
  - Empirically, correlation is negative.
- **Reasons:**
  - Customers underperform the market average.
  - Customer effectiveness changes over time.
  - And is affected by different investment horizons.
- **Recommendations**
  - Don’t limit your evaluations to transaction-based metrics!
  - Consider changing market conditions when testing financial recommenders.
  - Customer strategies might confound our evaluation.

Questions?

Dr. Javier Sanz-Cruzado
Financial Recommendation Systems
✉ javier.sanz-cruzadopuig@glasgow.ac.uk



What is the cause of the variation in correlation?

- We plot the confusion matrices between correlation and multiple conditions

Market ROI

		Value	
		+	-
Correlation	+	37.6%	30.5%
	-	62.4%	69.5%

Market changes do not correspond to correlation changes

Customer ROI

		Value	
		+	-
Correlation	+	44.2%	20.9%
	-	55.8%	79.1%

When customers are not effective, correlation is negative (does not work when customers are effective)

Customer vs. Market

		Value	
		+	-
Correlation	+	70.7%	9.1%
	-	29.3%	90.9%

Great correspondence between ROI differences and correlation sign!