BIG DATA ANALYTICS
project presentation

Unique, optimal investment strategies for user-specified parameters

By "Big Data Big Dreams"

- *Ariq Bintang*
- *Luca Gewehr*
- *Marco Hafid*

Coming up...

1. **Background**
2. Introduction
3. Research question
4. Data
5. Analysis
6. Results
7. Looking forward

Background

Unclear investment expectations.
Ineffective, costly, inaccessible portfolio management.

Coming up...

1. Background
2. **Introduction**
3. Research question
4. Data
5. Analysis
6. Results
7. Scaling & Cloud Deployment

Introduction (1/2)

Objective	Determine the unique, optimal investment strategies for user-specified parameters.
In support	We consider a range of sub-questions.

Ideation	• **Which financial instruments** would be most relevant? ➢ Exclude individual assets, given survivorship bias and hindsight bias. ➢ Focus on complementary indices of equities, bonds and commodities. • **Which indices** to include in our investment universe? ➢ Brainstormed a longlist of 30+ indices.
Data	• **Price data of selected indices and currency pairs**: ➢ ~~*Refinitiv Eikon*~~, ~~*Wharton (WRDS)*~~, ~~*Yahoo Finance*~~, Bloomberg Terminal. • **Swiss inflation data**: ➢ World Bank. • **CHF money market rates**, and **spot interest rates on Swiss bonds**: ➢ Swiss National Bank.

Introduction (2/2)

Methodology	• **Data preprocessing**: ➢ Data cleaning, integration and transformation. ➢ Data preparation (feature engineering). • **Derive unique optimal investment strategies** for user-specified parameters: ➢ Apply (machine learning) algorithms to the relevant dataset. ➢ Balance model performance and computational efficiency.
Final result	A reliable tool for decision-making in investment management.

Coming up...

1. Background
2. Introduction
3. **Research question**
4. Data
5. Analysis
6. Results
7. Scaling & Cloud Deployment

Research question

What is the unique optimal investment strategy that corresponds exactly to a given user-specified set of investment parameters?

Sub-questions:

1. Identifying the most relevant investment parameters.
2. Choosing the appropriate securities to be considered.
3. Defining and balancing criteria for model accuracy and computational efficiency.
4. Determining restrictions to possible combinations of securities.
5. Developing a method to determine an optimal investment strategy.
6. Identifying the optimal estimation method and corresponding specification
7. Validating the robustness of the optimal investment strategy.
8. Considering the impact of inflation and foreign exchange movements.
9. Evaluating the theoretical underpinnings and assumptions of the optimization model.

Coming up...

1. Background
2. Introduction
3. Research question
4. **Data**
5. Analysis
6. Results
7. Scaling & Cloud Deployment

Data
Data sources and data collection

Leveraging data from multiple sources:
sources that are reliable and comprehensive, thus well-suited for our research objectives.

File name (raw data)	Data source	Description	Date range (frequency)	File size
Bloomberg_Terminal-spreadsheet_builder.xlsx	Bloomberg Terminal	price data of selected indices and currency pairs	1 Jan 1973 – 16 May 2023 (daily)	4.363 KB
API_FP.CPI.TOTL.ZG_DS2_en_excel_v2_5454868.xls	World Bank Open Data	Swiss inflation data (CPI in %),	1960 – 2022 (annual)	315 KB
snb-chart-data-rendeidglfzch-en-all-20230502_1430.xlsx	Swiss National Bank data portal	CHF money market rates	4 Jan 1988 – 28 Apr 2023 (daily)	359 KB
snb-chart-data-zimomach-en-all-20230502_1430.xlsx	Swiss National Bank data portal	CHF spot interest rates on Swiss Confederation bond issues	3 Jan 2000 – 28 Apr 2023 (daily)	177 KB

Collecting and integrating the data correctly is a significant task:
loading the raw data into R from different sources, each with different data formats.

Data
Data cleaning and data preparation (1/2)

R data frames with raw data, collected from multiple sources			
index_prices_local_currency	**CHF_FX**	**swiss_inflation**	**CHF_rf_rates**
daily price data of selected indices, denoted in their local currency	*daily price data of selected currency pairs*	*annual Swiss inflation data (CPI, %)*	*daily CHF money market rates and spot interest rates on Swiss confederation bond issues*

(1) aligning dates across different data frames

(2) sorting and filtering data

(3) revising column names

(4) recalculating inflation values into percentages

(5) removing indices that did not contain sufficiently long dated price data and were not essential

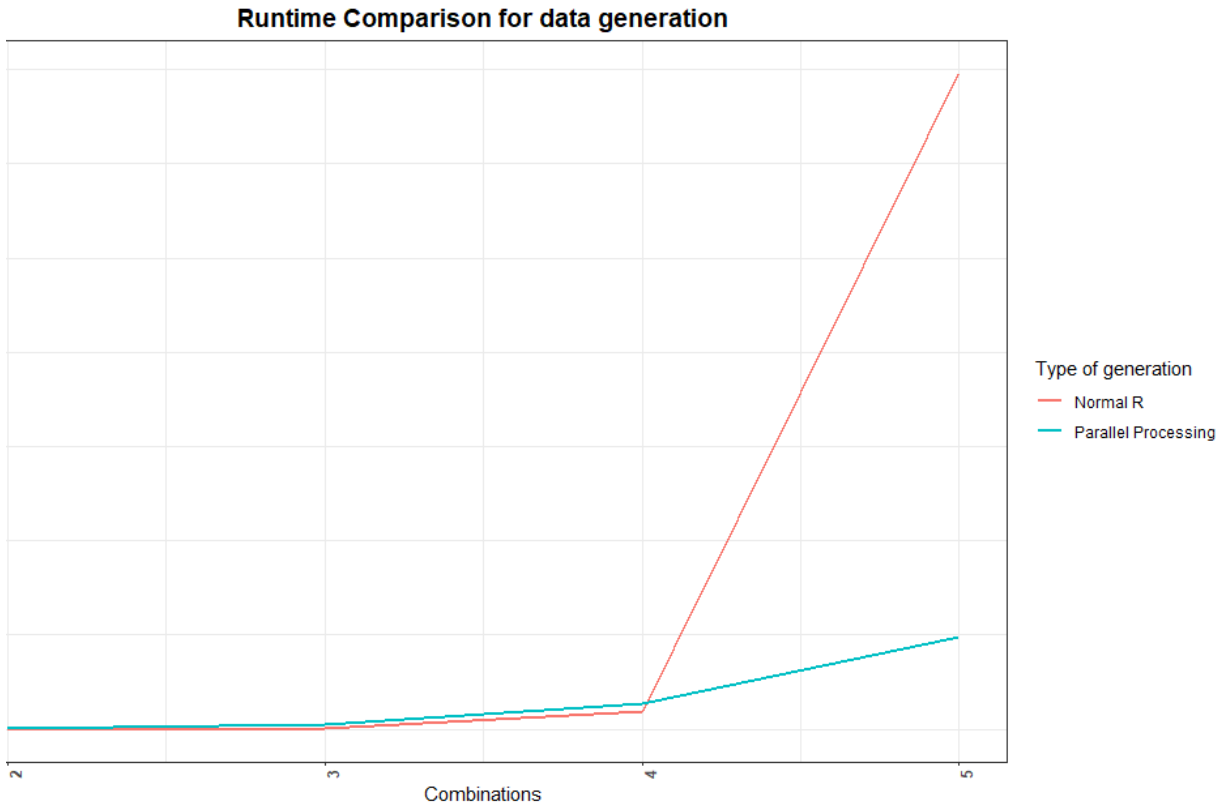
(6) removing dates that still contained N/A values

Data that is clean, consistent and prepared for further analysis or transformations:

- number of columns (indices) reduced from 49 to 26;
- rows (dates) include only observations for which all remaining indices display values.

Data

Data cleaning and data preparation (2/2)

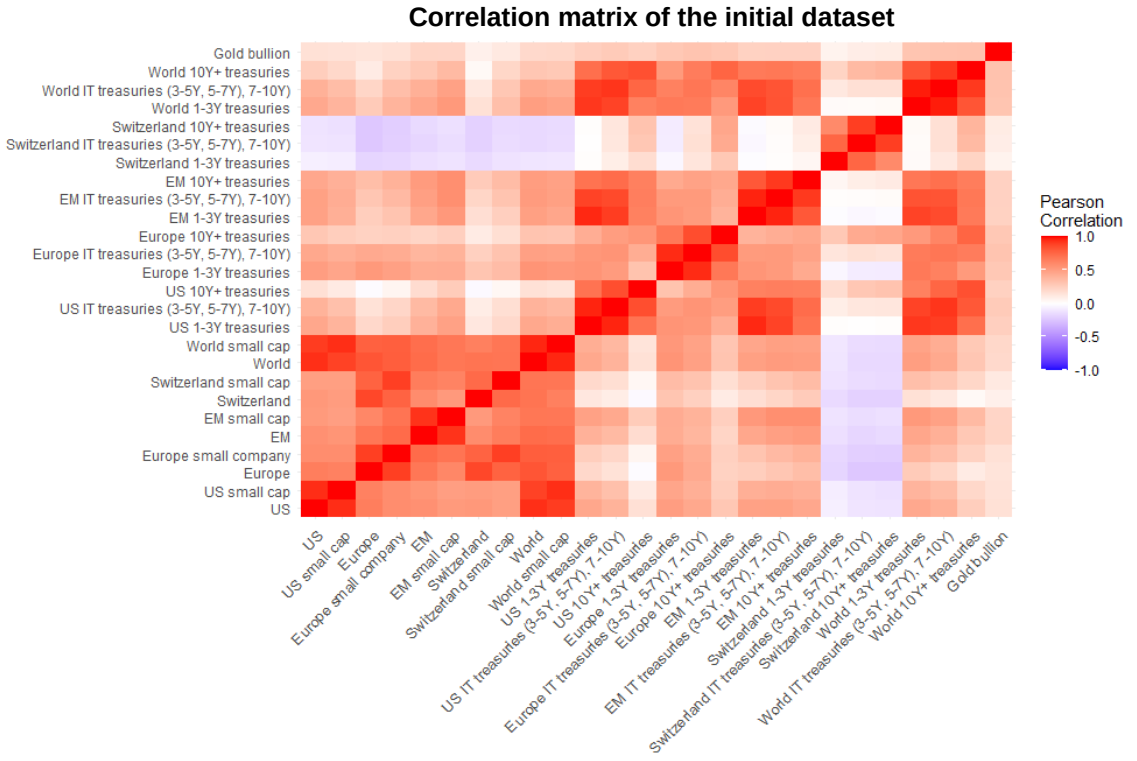


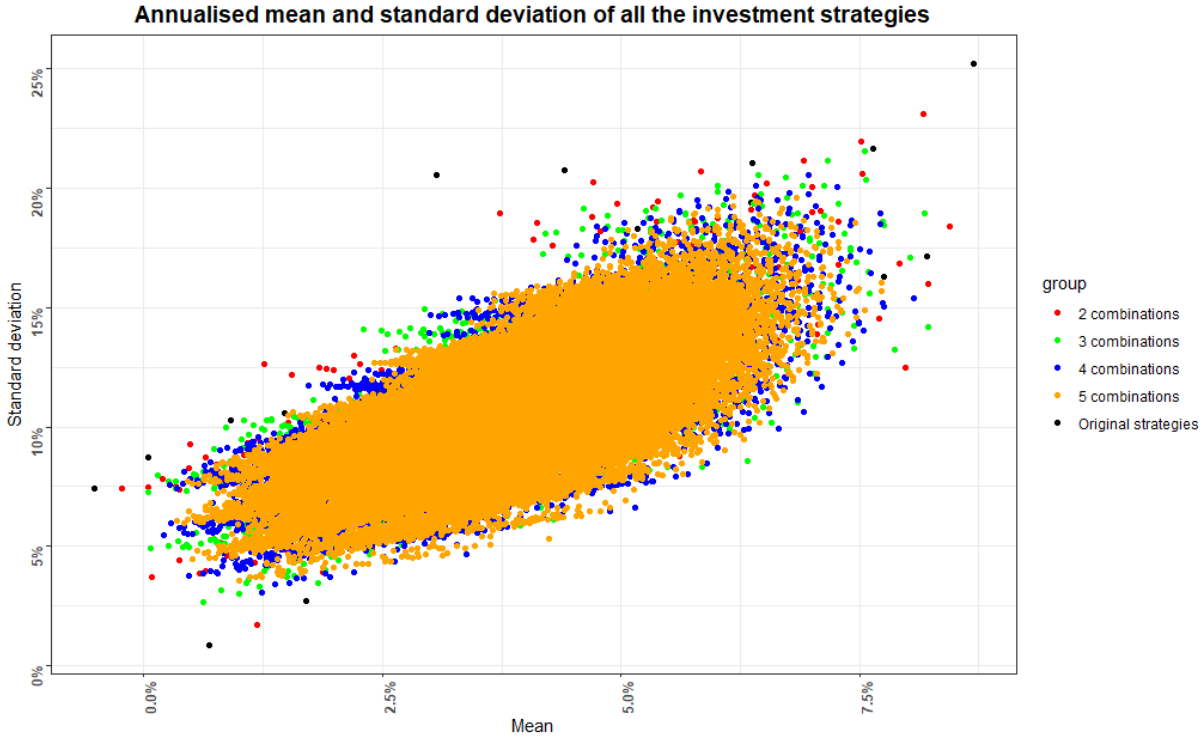
2	351	14
3	2951	115
4	17901	694
5	83681	3245
6	313911	12171
7	971711	37675
8	2533986	98248
9	5658536	219394

Coming up...

1. Background
2. Introduction
3. Research question
4. Data
5. **Analysis**
6. Results
7. Scaling & Cloud Deployment

Analysis


Correlation matrix of the initial dataset
Gold bullion
World 10Y+ treasuries
World IT treasuries (3-5Y, 5-7Y), 7-10Y)
World 1-3Y treasuries
Switzerland 10Y+ treasuries
Switzerland IT treasuries (3-5Y, 5-7Y), 7-10Y)
Switzerland 1-3Y treasuries
EM 10Y+ treasuries
EM IT treasuries (3-5Y, 5-7Y), 7-10Y)
EM 1-3Y treasuries
Europe 10Y+ treasuries
Europe IT treasuries (3-5Y, 5-7Y), 7-10Y)
Europe 1-3Y treasuries
US 10Y+ treasuries
US IT treasuries (3-5Y, 5-7Y), 7-10Y)
US 1-3Y treasuries
World small cap
World
Switzerland small cap
Switzerland
EM small cap
EM
Europe small company
Europe
US small cap
US
Pearson Correlation
1.0
0.5
0.0
-0.5
-1.0


Annualised mean and standard deviation of all the investment strategies
Standard deviation
0%
5%
10%
15%
20%
25%
Mean
0.0%
2.5%
5.0%
7.5%
group
2 combinations
3 combinations
4 combinations
5 combinations
Original strategies

Coming up...

1. Background
2. Introduction
3. Research question
4. Data
5. Analysis
6. **Results**
7. Scaling & Cloud Deployment

Results

Column	Worst_Period_End_Value
US	106.95353
US small cap	118.89280
Switzerland	85.80458
Switzerland small cap	111.33754
World	86.70190
World small cap	102.65978
US 1-3Y treasuries	90.38016
US IT treasuries (3-5Y, 5-7Y), 7-10Y)	107.61577
US 10Y+ treasuries	111.47750
Europe 10Y+ treasuries	85.21149
EM 1-3Y treasuries	108.08380
EM IT treasuries (3-5Y, 5-7Y), 7-10Y)	116.25697
EM 10Y+ treasuries	115.20969
Switzerland 1-3Y treasuries	95.40698
Switzerland IT treasuries (3-5Y, 5-7Y), 7-10Y)	97.21826
Switzerland 10Y+ treasuries	96.01914
World 1-3Y treasuries	85.48971
World IT treasuries (3-5Y, 5-7Y), 7-10Y)	97.02807
World 10Y+ treasuries	95.42560
Gold bullion	100.54892

Input parameters: 10 year time period, 85% threshold

Best of the worst: US small cap

Coming up...

1. Background
2. Introduction
3. Research question
4. Data
5. Analysis
6. Results
7. **Scaling & Cloud Deployment**

Scaling and Cloud Deployment
(Work in progress)

Leveraging the power of cloud platforms and data processing technologies:

1. Employ better (machine learning) algorithms for our investment optimization model.
2. Evaluate and improve the out-of-sample performance of our model.
3. Improve computational performance.
4. Store increasingly large data sets more efficiently.

Functionality	Amazon AWS	Google Cloud Platform	Microsoft Azure	Open-Source Software
Object Storage	Amazon S3	Google Cloud Storage	Azure Blob Storage	Hadoop Distributed File System (HDFS)
Data Warehousing / Analysis	Amazon Redshift	BigQuery	Azure Databricks	Apache Hadoop (MapReduce), Apache Spark (Spark SQL)
Compute Instances	AWS EC2	Google Compute Engine	Azure Virtual Machines	-
Machine Learning Platform	AWS SageMaker	Google Colab, Google Cloud AI	Azure Machine Learning	Apache Spark (Spark MLlib), H2O.ai, TensorFlow, PyTorch
Database Management System	Amazon RDS	Google Cloud SQL	Azure SQL Database	SQLite
Data Orchestration	AWS Step Functions	Google Cloud Composer	Azure Logic Apps	Apache Airflow

Thank you!

Questions?