Machine learning (and more) applied to market regimes, changepoints and anomaly detection in quantitative wealth and investment management QWIM

Cristian Homescu

April 2021

Contents

1 Motivation for the project

There is much evidence that crash and bubble periods display much different patterns than normal markets, suggesting that forecasting models (and investing approaches) ought to be based on multiple regimes.

It was shown that asset performance over long time periods can be separated into distinctive periods, called regimes, which display common characteristics. Regime-based asset allocation has been shown to add value over rebalancing to static weights and, in particular, reduce potential drawdowns by reacting to changes in market conditions. regime based asset allocation can effectively respond to changes in financial regimes at the portfolio level, in an effort to provide better long-term results than more static approaches can offer.

Baltas and Karyampas ("Forecasting the equity risk premium: The importance of regime-dependent evaluation," 2018):

"Is superior econometric predictability across the business cycle synonymous with predictability at all times?"

It appears that recently introduced forecasting models for equity risk premium ERP, which have been shown to generate econometrically superior ERP forecasts, have forecasting ability which is regime-dependent. They give rise to significant relative losses during market downturns, when it matters the most for asset allocators to retain assets and their client base intact. Conversely, any economic benefit occurring during market upswings is diminished for high risk-averse and leverage-constrained investors.

1.1 Market states in $\mathbb{QWIM}$

It was observed empirically that there are two separate market states:

- low uncertainty (relatively stable and resilient) market
- high uncertainty (relatively chaotic and fragile) market

Markets in **"low uncertainty"** state:

- statistically well behaved
- can be modeled using standard statistical tools
- volatility is stable and low
- correlations relatively stable
- tail events ($\geq$3 std deviations in either direction) quite rare.

Markets in **"high uncertainty"** state:

- not statistically well behaved
- vols and correlations change significantly on regular basis
- Tail events happen with much more regularity

To account for the two market states, practitioners use a relatively similar concept of **"risk on, risk off"**:

The **"high uncertainty"** state can incorporate multiple instances and multiple types of significant changes in time series:

- market regimes
- changepoints
- bubbles and crashes

1.2 Structural breaks: market regimes

Regime changes, some transitory, some recurring (recessions versus expansions) some permanent (structural breaks), are prevalent across a wide range of financial markets and in behavior of many macro variables. Examples of regimes considered in academia and/or practitioners:

- bull vs. bear market regimes
- inflationary vs. recessionary regimes

- high vs. low volatility regimes
- mean reverting vs. trending regimes

Regime shifts are challenging for investors because they cause portfolio performance, risk and behavior to depart significantly from ranges implied by long-term averages of means and covariances. Regime-based asset allocation was shown to deliver improved performance and risk profile

Goof performance of investment strategies greatly enhanced with introduction of regime switching models $\mathbb{RSM}$s . $\mathbb{RSM}$s characterize market states using estimates of parameters of some underlying model, and use a transition matrix to quantify probability of moving from one state to another.

$\mathbb{ML}$ may be effective at detecting change (even in chaotic system), for example through robust anomaly detection. It can be enhanced to compute probability of observation in previously observed "market regimes" (defined as clusters in $\mathbb{ML}$). Thus clustering algorithms can identify regimes in datasets. What they have in common with regular regime switching models is ability of producing probabilities of "switching" into another regime. $\mathbb{ML}$ can also feed on large amounts of data to detect preconditions of a break

1.3 Structural breaks: bubbles and crashes

Chaotic systems of the real world are comparable to stock market indices evolution.Llog-periodic power law singularity (LPPLS) model captures well bubbles and crashes LPPLS framework successfully captures, ex-ante, most prominent bubbles across different time scales (Black Monday, Dot-com, and Subprime Crisis).

1.4 Structural breaks: changepoints

Change point detection (CPD) is the problem of finding abrupt changes in data when a property of the time series changes . Segmentation, edge detection, event detection, and anomaly detection are similar concepts within $\mathbb{ML}$ space.

Traditional changepoint detection methods only look for statistically-detectable boundaries that are defined as abrupt variations in the generative parameters of a data sequence. . However, it is observed that breakpoints occur on more subtle boundaries non-trivial to detect with these statistical methods, but detectable using deep learning

2 Practical details for the project

The main purpose of the project described in this document is to provide exposure to students on important (and interesting) practical topics in quantitative wealth and investment management QWIM.

The level of complexity depends on the number of hours designated for the project. For example, 50-60 hours for a regular project, and 100-120 hours for a thesis/capstone project. Upon request, the scope (and the corresponding number if hours) of any given project can be extended.

The students would work on the project as part of a team (usually with 2-3 students).

All QWIM projects were selected such that the students' efforts have a good chance of producing results relevant to the industry, and at least as good as the results presented in the QWIM literature. Thus for each project we may consider (on an optional basis, based primarily on students' preference) to submit a corresponding article to journals widely followed by practitioners and academics in investment and wealth management, with participating students included as the leading coauthors of the submitted article.

The main challenge for each project is to identify the criteria for what would be considered **"good enough"**. Similar to projects in the industry, the meaning of "good enough" is based on a combination of comprehensive literature review, discussions within team and with me (and/or my colleagues) and analysis of results. Emphasis is placed on creating a narrative (with the aid of an interactive visualizer) for convincing the intended audience that what was done in the project delivers **"good enough"** outcome.

2.1 Interaction with students

For each project I would make myself available for meetings on a weekly basis (for discussions and guidance). Some of my colleagues have also expressed interest to participate in such meetings. Due to our work schedule and deliverables, most of the discussions will have to be scheduled outisde working hours (in weekends or evenings). The meetings will take place through video conferencing such as WebEx, Zoom, Google Meet, Microsoft Teams, etc., based on the team's preference. If the meetings are through WebEx, I would provide a link, while the student team will provide a link for any other video conferencing tool.

The students working on a given project can also send questions by email (my recommendation is to aggregate the questions from team members into an email sent once a day). We aim to provide answers within 1-2 days, either by email or through a phone discussion.

2.2 Data

Due to compliance reasons all projects would be based on publicly available, non-proprietary and non-confidential data (indices, ETFs, mutual funds, etc.). Since neither I nor my team are allowed to provide these datasets, I can only provide a list of suggested datasets. This list is included in a later section named Practical Info.

The datasets were selected to have the following features:

- be good proxies for most representative asset and subasset classes
- to be widely available
- to be as liquid as possible
- to have daily granularity
- to encompass periods with as many market regimes as possibles (most proposed daily datasets are from 1990 or 1991)
- time series have "nicer" statistical properties compared to time series of, say, individual stocks or bonds

2.3 Private GitHub repository for the QWIM project

The team will create a private GitHub repository, which will store relevant project materials, including codes. The team will use Git Desktop application as source control repository linked to the GitHub repository.

2.4 Deliverables

The project deliverables include literature survey, numerical results, analysis and visualization. For each project references will be provided for a comprehensive literature survey, and students are encouraged to identify additional relevant literature. Regarding the implementation, the project will primarily use existing codes:

- Python and R packages from official repositories (PyPi for Python and CRAN for R)
- machine learning platforms such as TensorFlow, PyTorch, CNTK, Chainer, mlr3, H20, PlaidML, mlpack, etc.
- implementations of articles through codes available in repositories such as GitHub, BitBucket, GitLab, etc.

Visualization of data and results visualization will be interactive and it will be based on Shiny R framework; to reduce programming effort, a template for such a Shiny visualizer will be provided in the team private GitHub repository.

The deliverables are:

- written report including literature survey and numerical results
- interactive visualizer (most likely Shiny-based visualizer using R and Python packages)
- (optional) presentation slides, and/or RMarkdown presentation, and/or Jupyter Notebook(s)

2.5 (Optional) Article submission to leading journals

On an optional basis (based primarily on students' preference), a version of the report can be prepared for submision to leading journals such as Journal of Financial Data Science, Journal of Portfolio Management, Journal of Asset Management, Journal of Investment Strategies, Quantitative Finance, Journal of Wealth Management, Journal of Investing, Journal of Machine Learning in Finance, etc.

3 Project tasks and timelines

For each project the main tasks are:

1) literature review

2) decide on the appropriate metrics and quantitative methods within context of "good enough" for the project

3) write-up summary of literature review: methods, metrics, testing procedures

4) identification of Python and/or R packages which are most appropriate for the selected methods and metrics

5) code design to decide on main code components

6) implementation of code components

7) interactive visualization of numerical results

8) project report containing description of methods, metrics, and tests, and analysis of results.

3.1 Suggested timelines for project tasks

The table below suggests a timeline for the project tasks and the corresponding percentages of project time:

Table 1: Suggested timeline for project tasks

Task ID	Task Name	Percentage of project time
1	Literature review	15%
2	Identification of "good enough" metrics and quantitative methods	5%
3	Write-up of summary of literature review	5%
4	Identification of appropriate packages in Python and/or R	10%
5	Code design for main components of project coding framework	5%
6	Implementation of coding framework and components	40%
7	Interactive visualizer using the provided Shiny template	10%
8	Project report and presentation	10%

3.2 Literature review

The first task is based on a comprehensive literature survey, included in the preliminary document of the project. Students are encouraged to identify additional relevant literature.

This task may be the most important of the project, since it provides an overview of what was done, what works well and less well, and what appear to be the most promising avenues to complete the project.

Emphasis is placed on information contained in the Main References, with analysis of the other References performed only as needed.

When reading the literature, there are 4 main directions to consider:

1) methods

2) metrics

3) testing procedures

4) numerical results

The primary focus would be on the the references included in "Main References" subsection of the document for your QWIM project. Then, to the extent there is time, to consider the other references included in the project document. In the same time, you are encouraged to identify other references that might be considered "Main references", and to share those references with me for discussion.

For the articles in Main References category, the suggested approach would be the following:

- For each article focus primarily on Abstract, Conclusion, and Numerical Results

- Do this for all articles considered to be Main References, such that you gain a high-level understanding of what is currently done in the literature
- Select the metrics that you may want to use in order to quantify the meaning of "good enough" for the project.
- Select the quantitative methods which appear to be most likely to be "good enough" for the project.
- Perform a "deeper dive" into the articles containing the approaches you consider the most promising,

For the articles which are not in "Main References" category, read Abstract, Conclusion, and Numerical Results, to see whether any of those articles might need to be considered for inclusion in your summary.

3.3 Write-up summary of literature review

The write-up summary summarizes the methods, metrics, testing procedures, and numerical results identified during the literature review. The write-up could also be incorporated within reports and/or presentations for the QWIM project.

3.4 Identification of appropriate Python and/or R packages

Based on the literature review and on diiscussions, we identify the most potentially useful methods, metrics and testing procedures. Then wee identify the most appropriate implementations of the selected methods and metrics.

The primary sources of implementatins are existing codes from:

- Python and R packages from official repositories (PyPi for Python and CRAN for R)
- machine learning platforms such as TensorFlow, PyTorch, CNTK, Chainer, mlr3, H20, PlaidML, mlpack, etc.
- implementations of articles through codes available in repositories such as GitHub, BitBucket, GitLab, etc.

3.5 Code design

An important task is to have a code design session to decide in advance on the main code components, which are meant to be modular and encapsulated, such that the entire team can work on the codes.

Examples of such main code components are presented below:

- get data
- forecast
- optimization
- calculate metrics
- perform tests
- obtain results
- construct interactive visualizer

The code design procedure consists of:

1) visual display of major components of the coding framework
2) UML diagrams for each of the components.

The Appendix contains an illustrative example within context of a QWIM project on forecasting of financial time series. The first figure shows the major components, while the second figure shows UML diagrams of those components (the names of data members and methods are currently generic, and one would need to change them to appropriate names)

While these figures were obtained through Microsoft Visio using a code design file (.vsd file), there are other software tools (either online or installed locally) which can be used to create such code design diagrams. NOTE: if you have access to Microsoft Visio and you want to use it for code design diagrams, you can ask me for the .vsd file which was exported into the PDF from which I have extracted the snapshots.

List of software tools for code design diagrams, which are either free (open source) or have a free type of account

- Modelio (either desktop version or online version)
- LucidChart (online)
- draw.io (either desktop version or online version, now called app.diagrams.net)
- Visual Paradigm (online)
- UMLet (either desktop or online version)
- Curated list of UML tools – 2019 edition
- Top online UML modeling tools in 2019

3.6 Implementation of coding framework and components

The implementation is done using identified packages or codes, in Python and/or R. The project will primarily use existing codes:

- Python and R packages from official repositories (PyPi for Python and CRAN for R)
- machine learning platforms such as TensorFlow, PyTorch, CNTK, Chainer, mlr3, H20, PlaidML, mlpack, etc.
- implementations of articles through codes available in repositories such as GitHub, BitBucket, GitLab, etc.

3.7 Interactive visualizer

While visualization of data and numerical results can be done through various tools (including Jupyter notebooks or Dash in Python), my recommendation is to consider an interactive visualizer based on Shiny framework in R. A template for the Shiny visualizer will be provided in the private GitHub repository set up by the team for the project.

Some information about Shiny:

- Shiny from RStudio: tutorials and gallery
- Why R Shiny Trumps UI and JavaScript Based Visualization Tools
- Shiny's Holy Grail: Interactivity with reproducibility

3.8 Project report and presentation

The report containing description of methods, metrics, and tests, and analysis of results.

While the report can be written using various tools (including Microsoft Word), my recommendation is to use LyX to write both the project report and the project presentation. Two LyX templates for creating reports and, respectively, presentations will be provided in the private GitHub repository set up by the team for the project.

Some information about Shiny:

- LyX features
- LyX tutorial with PDF here
- LyX Tutorial video Part One and Part Two
- LyX tutorial video Part One and Part Two and Part Three and Part Four
- Introduction to LyX
- Insert figures in LyX
- Essentials of LyX

4 Literature Review

Main references:
Costa and Kwon ("A regime-switching factor model for mean-variance optimization," 2020),
Dal Pra et al. ("Regime Shifts in Excess Stock Return Predictability: An Out-of-Sample Portfolio Analysis," 2018),
Demos and Sornette ("Birth or burst of financial bubbles: which one is easier to diagnose?" 2017),
Filimonov et al. ("Modified profile likelihood inference and interval forecast of the burst of financial bubbles," 2017),
Fons et al. ("A novel dynamic asset allocation system using Feature Saliency Hidden Markov models for smart beta investing," 2021),
Gerlach et al. ("Dissection of Bitcoin's Multiscale Bubble History from January 2012 to February 2018," 2018),
Lattanzi and Leonelli ("A changepoint approach for the identification of financial extreme regimes," 2019),
Mizuno et al. ("Detecting Stock Market Bubbles Based on the Cross-Sectional Dispersion of Stock Prices," 2020),
Nystrup et al. ("Dynamic Allocation or Diversification: A Regime-Based Approach to Multiple Assets," 2018),
Nystrup et al. ("Dynamic portfolio optimization across hidden market regimes," 2018),
Nystrup et al. ("Learning hidden Markov models with persistent states by penalizing jumps," 2020),
Nystrup et al. ("Detecting change points in VIX and S&P 500: A new approach to dynamic asset allocation," 2016),
Nystrup et al. ("Regime-Based Versus Static Asset Allocation: Letting the Data Speak," 2015),
Pharasi et al. ("Market states: A new understanding," 2020),
Simonian ("Mixed Ag: A Regime-Based Analysis of Multi-Asset Agriculture Portfolios," 2020),
Simonian and Wu ("Factors in Time: Fine-Tuning Hedge Fund Replication," 2019),
Sornette et al. ("Can We Use Volatility to Diagnose Financial Bubbles? Lessons from 40 Historical Bubbles," 2017),
Wheatley et al. ("Are bitcoin bubbles predictable? combining a generalized metcalfe's law and the LPPLS model," 2018),

4.1 Market states, regimes and structural breaks of time series for investment strategies

References:
Bae et al. ("Dynamic asset allocation for varied financial markets under regime switching framework," 2014),
Costa and Kwon ("Risk parity portfolio optimization under a Markov regime-switching framework," 2019),
Costa and Kwon ("A regime-switching factor model for mean-variance optimization," 2020),
Fischer and Murg ("A combined regime-switching and Black Litterman model for optimal asset allocation," 2015),
Kim et al. ("Global Asset Allocation Strategy Using a Hidden Markov Model," 2019),
Komatsu and Makimoto ("Dynamic Investment Strategy with Factor Models Under Regime Switches," 2015),
Mulvey et al. ("Machine learning, economic regimes and portfolio optimisation," 2018),
Nystrup et al. ("Regime-Based Versus Static Asset Allocation: Letting the Data Speak," 2015),
Nystrup et al. ("Dynamic portfolio optimization across hidden market regimes," 2018),
Nystrup ("Regime-Based Asset Allocation: Do Profitable Strategies Exist?" 2014),
Nystrup et al. ("Dynamic Allocation or Diversification: A Regime-Based Approach to Multiple Assets," 2018),
Papenbrock and Schwendner ("Handling risk-on/risk-off dynamics with correlation regimes and correlation networks," 2015),
Platanakis et al. ("Portfolios in a Regime Shifting Non-Normal World: Are Alternative Assets Beneficial?" 2017),
Seidl ("Markowitz versus Regime Switching: An Empirical Approach," 2012),
Sheikh and Sun ("Regime Change: Implications of Macroeconomic Shifts on Asset Class and Portfolio Performance," 2012),
Simonian and Wu ("Minsky vs. Machine: New Foundations for Quant-Macro Investing," 2019),

4.2 Regime-based asset allocation

References:
Ahmad et al. ("Regime dependent dynamics and European stock markets: Is asset allocation really possible?" 2015),
Bae et al. ("Dynamic asset allocation for varied financial markets under regime switching framework," 2014),
Berger and Gencay ("Short-run wavelet-based covariance regimes for applied portfolio management," 2020),
Blin et al. ("A Macro Risk-Based Approach to Alternative Risk Premia Allocation," 2017),
Flint and Mare ("Regime-Based Tactical Allocation for Equity Factors and Balanced Portfolios," 2019),
Fons et al. ("A novel dynamic asset allocation system using Feature Saliency Hidden Markov models for smart beta investing," 2021),
Kritzman et al. ("Regime Shifts: Implications for Dynamic Strategies," 2012),

Lezmi et al. (“Portfolio Allocation with Skewness Risk: A Practical Guide,” 2018),
Liszewski (“Asset allocation under multiple regimes,” 2016),
Nystrup et al. (“Dynamic portfolio optimization across hidden market regimes,” 2018),
Nystrup et al. (“Regime-Based Versus Static Asset Allocation: Letting the Data Speak,” 2015),
Nystrup et al. (“Dynamic Allocation or Diversification: A Regime-Based Approach to Multiple Assets,” 2018),
Oliveira and Valls Pereira (“Asset Allocation With Markovian Regime Switching: Efficient Frontier and Tangent Portfolio With Regime Switching,” 2018),
Sheikh and Sun (“Regime Change: Implications of Macroeconomic Shifts on Asset Class and Portfolio Performance,” 2012),
van Vliet and Blitz (“Dynamic strategic asset allocation: Risk and return across the business cycle,” 2011),
Vo and Maurer (“Dynamic Asset Allocation under Regime Switching, Predictability and Parameter Uncertainty,” 2013),

4.3 Detection and usage of bubbles, crashes and business cycles for investment strategies

References:
Astill et al. (“Real-Time Monitoring for Explosive Financial Bubbles,” 2018),
Bianchi (“The Great Depression and the Great Recession: A view from financial markets,” 2020),
Cram (“Late to Recessions: Stocks and the Business Cycle,” 2020),
Engle and Ruan (“Measuring the probability of a financial crisis,” 2019),
Gerlach et al. (“Crash-sensitive Kelly Strategy built on a modified Kreuser-Sornette bubble model tested over three decades of twenty equity indices,” 2020),
Gobel and Araujo (“Indicators of economic crises: a data-driven clustering approach,” 2020),
Kole and van Dijk (“How to Identify and Forecast Bull and Bear Markets?” 2016),
Kreuser and Sornette (“Super-Exponential RE bubble model with efficient crashes,” 2019),
Mehta (“The Mechanism behind the Bursting of Financial Bubbles and Market Crashes,” 2020),
Mizuno et al. (“Detecting Stock Market Bubbles Based on the Cross-Sectional Dispersion of Stock Prices,” 2020),
Smug et al. (“Predicting Financial Market Crashes Using Ghost Singularities,” 2017),
Sornette (“Dragon-kings and Predictions: Diagnostics and Forecasts for the World Financial Crisis,” 2014),
Sornette and Cauwels (“Financial bubbles: mechanisms and diagnostics,” 2014),
Sornette et al. (“Real-time prediction and post-mortem analysis of the Shanghai 2015 stock market bubble and crash,” 2015),
Sornette et al. (“Resolving Persistent Uncertainty by Self-Organized Consensus to Mitigate Market Bubbles,” 2016),
Sornette et al. (“Can We Use Volatility to Diagnose Financial Bubbles? Lessons from 40 Historical Bubbles,” 2017),
Viebig (“Exuberance in Financial Markets: Evidence from Machine Learning Algorithms,” 2020),
Wang and Zong (“Are Crises Predictable? A Review of the Early Warning Systems in Currency and Stock Markets,” 2020),
Yan and Huang (“Financial cycle and business cycle: An empirical analysis based on the data from the U.S,” 2020),
Yao and Li (“A study on the bursting point of Bitcoin based on the BSADF and LPPLS methods,” 2021),
Zhang et al. (“LPPLS bubble indicators over two centuries of the S&P 500 index,” 2016),

5 Practical Info

5.1 Software

The recommended version of software are as follows:

- R
  - version 3.6.3, or
  - version 4.0.5, or
  - Microsoft R Open latest version (currently 4.0.2)
- Python
  - latest version of 3.8 (currently 3.8.9), or
  - latest version of Python 3.7 (currently 3.7.10), or
  - Anaconda Python 3 latest version (currently 2020.11 for Python 3.8)
- R IDE
  - RStudio Desktop Open Source latest version (currently 1.4.1106)
- Python IDE:
  - PyCharm Community Edition latest version (currently 2021.1), or
  - Visual Studio Code latest version (currently 1.55.2), or
  - your favorite IDE
- LyX latest version (currently 2.3.6.1)
- JabRef latest version:
  - latest official version (currently 5.2), or
  - latest development version 5.3 (from JabRef website)
- TexLive latest version (currently 2021)
- appropriate R packages (to be selected during the project)
- appropriate Python packages (to be selected during the project)
- Git for desktop latest version (currently 2.7.2)
- Notepad++ latest version (currently 7.9.5)
- SourceTree latest version (currently 3.4.4)

5.2 Examples of Datasets

The datasets were selected to have the following features:

- be good proxies for most representative asset and subasset classes
- to be widely available
- to be as liquid as possible
- to have daily granularity
- to encompass periods with as many market regimes as possibles (most proposed daily datasets start from early 1990s)
- time series have "nicer" statistical properties compared to time series of, say, individual stocks or bonds

The following datasets are suggested

Table 2: Daily data sets

Name	Description	Name	Description
BCOMTR	Bloomberg Commodity Index Total Return	RU20VATR	iShares Russell 2000 Value ETF
HFRIFWI	HFRI Fund Weighted Composite Index	RUMCINTR	iShares Russell Mid-Cap ETF
LBUSTRUU	Bloomberg Barclays US Aggregate Bond Index	RUMRINTR	iShares Micro-Cap ETF
LG30TRUU	Bloomberg Barclays Global High Yield Total Return Index Value Unhedge	RUTPINTR	iShares Russell Top 200 ETF
LMBITR	Bloomberg Barclays Municipal Bond Index Total Return Index Value Unhedged USD	S5COND	S&P 500 Consumer Discretionary Index
NDDUE15X	Amundi MSCI Europe Ex UK Ucits ETF Dr	S5CONS	S&P 500 Consumer Staples Index
NDDUJN	MSCI Japan Index	S5ENRS	S&P 500 Energy Index
NDDUNA	iShares MSCI North America UCITS ETF	S5FINL	S&P 500 Financials Sector GICS Level 1 Index
NDDUPXJ	MSCI Pacific ex Japan UCITS ETF	S5HLTH	S&P 500 Health Care Index
NDDUUK	iShares MSCI UK ETF	S5INDU	S&P 500 Industrials Index
NDDUWXUS	MSCI World ex USA total net return	S5INFT	S&P 500 Information Technology Index
NDUEEGF	SPDR MSCI Emerging Markets UCITS ETF	S5MATR	S&P 500 Materials Index
RU10GRTR	iShares Russell 1000 Growth ETF	S5RLST	S&P 500 Real Estate Index
RU10VATR	iShares Russell 1000 Value ETF	S5TELS	S&P 500 Communication Services Index
RU20GRTR	iShares Russell 2000 Growth ETF	S5UTIL	S&P 500 Utilities Index
RU20INTR	Russell 2000 Total Return	SPXT	Proshares S&P 500 EX Technology ETF

Table 3: Monthly data sets

Name	Description	Name	Description
IBXXSHY1	iShares 0-5 Year High Yield Corporate Bond ETF	M2USEV	MSCI USA Enhanced Value Index
IDCT20RT	ICE U.S. Treasury 20+ Year Bond Total Return Index	M2USRWGT	MSCI USA Risk Weighted Index
LBUSTRUU	Bloomberg Barclays US Agg Total Return Value Unhedged USD	M2USSNQ	MSCI USA Sector Neutral Quality Index
LC07TRUU	Bloomberg Barclays U.S. Universal Total Return Index Value Unhedged	MID	S&P 400 Mid Cap Index index
LD01TRUU	Bloomberg Barclays 1-3 Yr Credit Total Return Index Value Unhedged US	MXEA	MSCI EAFE Index
LT01TRUU	Bloomberg Barclays US Treasury 1-3 Year Index	MXEF	MSCI Emerging Markets Index
LUICTRUU	Bloomberg Barclays U.S. Intermediate Credit Total Return Index	MXUSMVOL	MSCI USA Minimum Volatility Index
LULCTRUU	Bloomberg Barclays U.S. Long Credit Index	MXWD	MSCI All Countries World Index
M1CXBRU	iShares Core MSCI International Developed Markets ETF	MXWOUIM	MSCI All Countries World Index
M1USMVOL	MSCI USA Minimum Volatility (USD) Index	NDDUUS	MSCI Daily Total Return Net USA USD Index
M2US000$	iShares Edge MSCI USA Momentum Factor ETF	SPX	S&P 500 Index

References

Ahmad, W., Bhanumurthy, N. R., and Sehgal, S. (2015). "Regime dependent dynamics and European stock markets: Is asset allocation really possible?" In: *Empirica* 42(1), pp. 77–107.

Astill, S., Harvey, D. I., Leybourne, S. J., Sollis, R., and Taylor, A. M. R. (2018). "Real-Time Monitoring for Explosive Financial Bubbles." In: *Journal of Time Series Analysis* 39(6), pp. 863–891.

Bae, G. I., Kim, W. C., and Mulvey, J. M. (2014). "Dynamic asset allocation for varied financial markets under regime switching framework." In: *European Journal of Operational Research* 234(2), pp. 450–458.

Baltas, N. and Karyampas, D. (2018). "Forecasting the equity risk premium: The importance of regime-dependent evaluation." In: *Journal of Financial Markets* 38(March), pp. 83–102.

Berger, T. and Gencay, R. (2020). "Short-run wavelet-based covariance regimes for applied portfolio management." In: *Journal of Forecasting* 39(4), pp. 642–660.

Bianchi, F. (2020). "The Great Depression and the Great Recession: A view from financial markets." In: *Journal of Monetary Economics* 114, pp. 240–261.

Blin, O., Ielpo, F., Lee, J., and Teiletche, J. (2017). "A Macro Risk-Based Approach to Alternative Risk Premia Allocation." In: *Factor Investing*. Elsevier, pp. 285–316.

Costa, G. and Kwon, R. H. (2019). "Risk parity portfolio optimization under a Markov regime-switching framework." In: *Quantitative Finance* 19(33), pp. 453–471.

Costa, G. and Kwon, R. H. (2020). "A regime-switching factor model for mean-variance optimization." In: *Journal of Risk* 22(4), pp. 31–59.

Cram, R. G. (2020). "Late to Recessions: Stocks and the Business Cycle." In: *SSRN e-Print*.

Dal Pra, G., Guidolin, M., Pedio, M., and Vasile, F. (2018). "Regime Shifts in Excess Stock Return Predictability: An Out-of-Sample Portfolio Analysis." In: *The Journal of Portfolio Management* 44(3), pp. 10–24.

Demos, G. and Sornette, D. (2017). "Birth or burst of financial bubbles: which one is easier to diagnose?" In: *Quantitative Finance* 17(5), pp. 657–675.

Engle, R. F. and Ruan, T. (2019). "Measuring the probability of a financial crisis." In: *Proceedings of the National Academy of Sciences* 116(37), pp. 18341–18346.

Filimonov, V., Demos, G., and Sornette, D. (2017). "Modified profile likelihood inference and interval forecast of the burst of financial bubbles." In: *Quantitative Finance* 17(8), pp. 1167-11861–20.

Fischer, E. O. and Murg, M. (2015). "A combined regime-switching and Black Litterman model for optimal asset allocation." In: *Journal of Investment Strategies* 4(3), pp. 1–36.

Flint, E. J. and Mare, E. (2019). "Regime-Based Tactical Allocation for Equity Factors and Balanced Portfolios." In: *South African Actuarial Journal* 19(1), pp. 27–52.

Fons, E., Dawson, P., Yau, J., Zeng, X.-j., and Keane, J. (2021). "A novel dynamic asset allocation system using Feature Saliency Hidden Markov models for smart beta investing." In: *Expert Systems with Applications* 163, pp. 113720+.

Gerlach, J.-C., Demos, G., and Sornette, D. (2018). "Dissection of Bitcoin's Multiscale Bubble History from January 2012 to February 2018." In: *arXiv e-Print*.

Gerlach, J.-C., Kreuser, J. L., and Sornette, D. (2020). "Crash-sensitive Kelly Strategy built on a modified Kreuser-Sornette bubble model tested over three decades of twenty equity indices." In: *SSRN e-Print*.

Gobel, M. and Araujo, T. (2020). "Indicators of economic crises: a data-driven clustering approach." In: *Applied Network Science* 5(1) (44).

Kim, E.-c., Jeong, H.-w., and Lee, N.-y. (2019). "Global Asset Allocation Strategy Using a Hidden Markov Model." In: *Journal of Risk and Financial Management* 12(4), p. 168.

Kole, E. and van Dijk, D. (2016). "How to Identify and Forecast Bull and Bear Markets?" In: *Journal of Applied Econometrics* 32(1), pp. 120–139.

Komatsu, T. and Makimoto, N. (2015). "Dynamic Investment Strategy with Factor Models Under Regime Switches." In: *Asia-Pacific Financial Markets* 22(2), pp. 209–237.

Kreuser, J. and Sornette, D. (2019). "Super-Exponential RE bubble model with efficient crashes." In: *The European Journal of Finance* 25(4), pp. 338–368.

Kritzman, M., Page, S., and Turkington, D. (2012). "Regime Shifts: Implications for Dynamic Strategies." In: *Financial Analysts Journal* 68(3).

Lattanzi, C. and Leonelli, M. (2019). "A changepoint approach for the identification of financial extreme regimes." In: *arXiv e-Print*.

Lezmi, E., Malongo, H., Roncalli, T., and Sobotka, R. (2018). "Portfolio Allocation with Skewness Risk: A Practical Guide." In: *SSRN e-Print*.

Liszewski, O. (2016). "Asset allocation under multiple regimes." MA thesis. Erasmus University.

Mehta, P. (2020). "The Mechanism behind the Bursting of Financial Bubbles and Market Crashes." In: *SSRN e-Print*.

Mizuno, T., Ohnishi, T., and Watanabe, T. (2020). "Detecting Stock Market Bubbles Based on the Cross-Sectional Dispersion of Stock Prices." In: *Proceedings of the 23rd Asia Pacific symposium on intelligent and evolutionary systems*. Ed. by H. Sato, S. Iwanaga, and A. Ishii. Vol. 12. Springer International Publishing, pp. 194–202.

Mulvey, J. M., Hao, H., and Li, N. (2018). "Machine learning, economic regimes and portfolio optimisation." In: *International Journal of Financial Engineering and Risk Management* 2(4), p. 260.

Nystrup, P. (2014). "Regime-Based Asset Allocation: Do Profitable Strategies Exist?" MA thesis. Technical University of Denmark.

Nystrup, P., Hansen, B. W., Larsen, H. O., Madsen, H., and Lindstrom, E. (2018a). "Dynamic Allocation or Diversification: A Regime-Based Approach to Multiple Assets." In: *The Journal of Portfolio Management* 44(2), pp. 62–73.

Nystrup, P., Hansen, B. W., Madsen, H., and Lindstrom, E. (2015). "Regime-Based Versus Static Asset Allocation: Letting the Data Speak." In: *The Journal of Portfolio Management* 42(1), pp. 103–109.

Nystrup, P., Lindstrom, E., and Madsen, H. (2020). "Learning hidden Markov models with persistent states by penalizing jumps." In: *Expert Systems with Applications* 150, p. 113307.

Nystrup, P., Madsen, H., and Lindstrom, E. (2018b). "Dynamic portfolio optimization across hidden market regimes." In: *Quantitative Finance* 18(1), pp. 83–95.

Nystrup, P., William Hansen, B., Madsen, H., and Lindstrom, E. (2016). "Detecting change points in VIX and S&P 500: A new approach to dynamic asset allocation." In: *Journal of Asset Management* 17, pp. 361–374.

Oliveira, A. B. and Valls Pereira, P. L. (2018). "Asset Allocation With Markovian Regime Switching: Efficient Frontier and Tangent Portfolio With Regime Switching." In: *SSRN e-Print*.

Papenbrock, J. and Schwendner, P. (2015). "Handling risk-on/risk-off dynamics with correlation regimes and correlation networks." In: *Financial Markets and Portfolio Management* 29(2), pp. 125–147.

Pharasi, H. K., Seligman, E., and Seligman, T. H. (2020). "Market states: A new understanding." In: *arXiv e-Print*.

Platanakis, E., Sakkas, A., and Sutcliffe, C. (2017). "Portfolios in a Regime Shifting Non-Normal World: Are Alternative Assets Beneficial?" In: *European Financial Management Association Annual Meeting Athens*.

Seidl, I. (2012). "Markowitz versus Regime Switching: An Empirical Approach." In: *The Review of Finance and Banking* 04(1), pp. 033–043.

Sheikh, A. Z. and Sun, J. (2012). "Regime Change: Implications of Macroeconomic Shifts on Asset Class and Portfolio Performance." In: *The Journal of Investing* 21(3), pp. 36–54.

Simonian, J. (2020). "Mixed Ag: A Regime-Based Analysis of Multi-Asset Agriculture Portfolios." In: *The Journal of Portfolio Management* 46(6), pp. 135–146.

Simonian, J. and Wu, C. (2019a). "Factors in Time: Fine-Tuning Hedge Fund Replication." In: *The Journal of Portfolio Management* 45 (3), pp. 159–164.

Simonian, J. and Wu, C. (2019b). "Minsky vs. Machine: New Foundations for Quant-Macro Investing." In: *The Journal of Financial Data Science* 1(2), pp. 94–110.

Smug, D., Ashwin, P., and Sornette, D. (2017). "Predicting Financial Market Crashes Using Ghost Singularities." In: *SSRN e-Print*.

Sornette, D. (2014). "Dragon-kings and Predictions: Diagnostics and Forecasts for the World Financial Crisis." In: *SSRN e-Print*.

Sornette, D., Andraszewicz, S., Murphy, R. O., Rindler, P. B., and Sanadgol, D. (2016). "Resolving Persistent Uncertainty by Self-Organized Consensus to Mitigate Market Bubbles." In: *SSRN e-Print*.

Sornette, D. and Cauwels, P. (2014). "Financial bubbles: mechanisms and diagnostics." In: *arXiv e-Print*.

Sornette, D., Cauwels, P., and Smilyanov, G. (2017). "Can We Use Volatility to Diagnose Financial Bubbles? Lessons from 40 Historical Bubbles." In: *SSRN e-Print*.

Sornette, D., Demos, G., Zhang, Q., Cauwels, P., Filimonov, V., and Zhang, Q. (2015). "Real-time prediction and post-mortem analysis of the Shanghai 2015 stock market bubble and crash." In: *Journal of Investment Strategies* 4(4).

van Vliet, P. and Blitz, D. (2011). "Dynamic strategic asset allocation: Risk and return across the business cycle." In: *Journal of Asset Management* 12(5), pp. 360–375.

Viebig, J. (2020). "Exuberance in Financial Markets: Evidence from Machine Learning Algorithms." In: *Journal of Behavioral Finance* 21(2), pp. 128–135.

Vo, H. T. and Maurer, R. (2013). "Dynamic Asset Allocation under Regime Switching, Predictability and Parameter Uncertainty." In: *SSRN e-Print*.

Wang, P. and Zong, L. (2020). "Are Crises Predictable? A Review of the Early Warning Systems in Currency and Stock Markets." In: *arXiv e-Print*.

Wheatley, S., Sornette, D., Huber, T., Reppen, M., and Gantner, R. N. (2018). "Are bitcoin bubbles predictable? combining a generalized metcalfe's law and the LPPLS model." In: *SSRN e-Print*.

Yan, C. and Huang, K. X. D. (2020). "Financial cycle and business cycle: An empirical analysis based on the data from the U.S." In: *Economic Modelling* 93, pp. 693–701.

Yao, C.-Z. and Li, H.-Y. (2021). "A study on the bursting point of Bitcoin based on the BSADF and LPPLS methods." In: *The North American Journal of Economics and Finance* (101280), Early View.

Zhang, Q., Sornette, D., Balcilar, M., Gupta, R., Ozdemir, Z. A., and Yetkiner, H. (2016). "LPPLS bubble indicators over two centuries of the S&P 500 index." In: *Physica A: Statistical Mechanics and its Applications* 458, pp. 126–139.

Appendix A: Code design diagrams

The code design procedure consists of:

1) visual display of major components of the coding framework
2) UML diagrams for each of the components.

We present examples below.

Figure 1: Visual display of major components of the coding framework

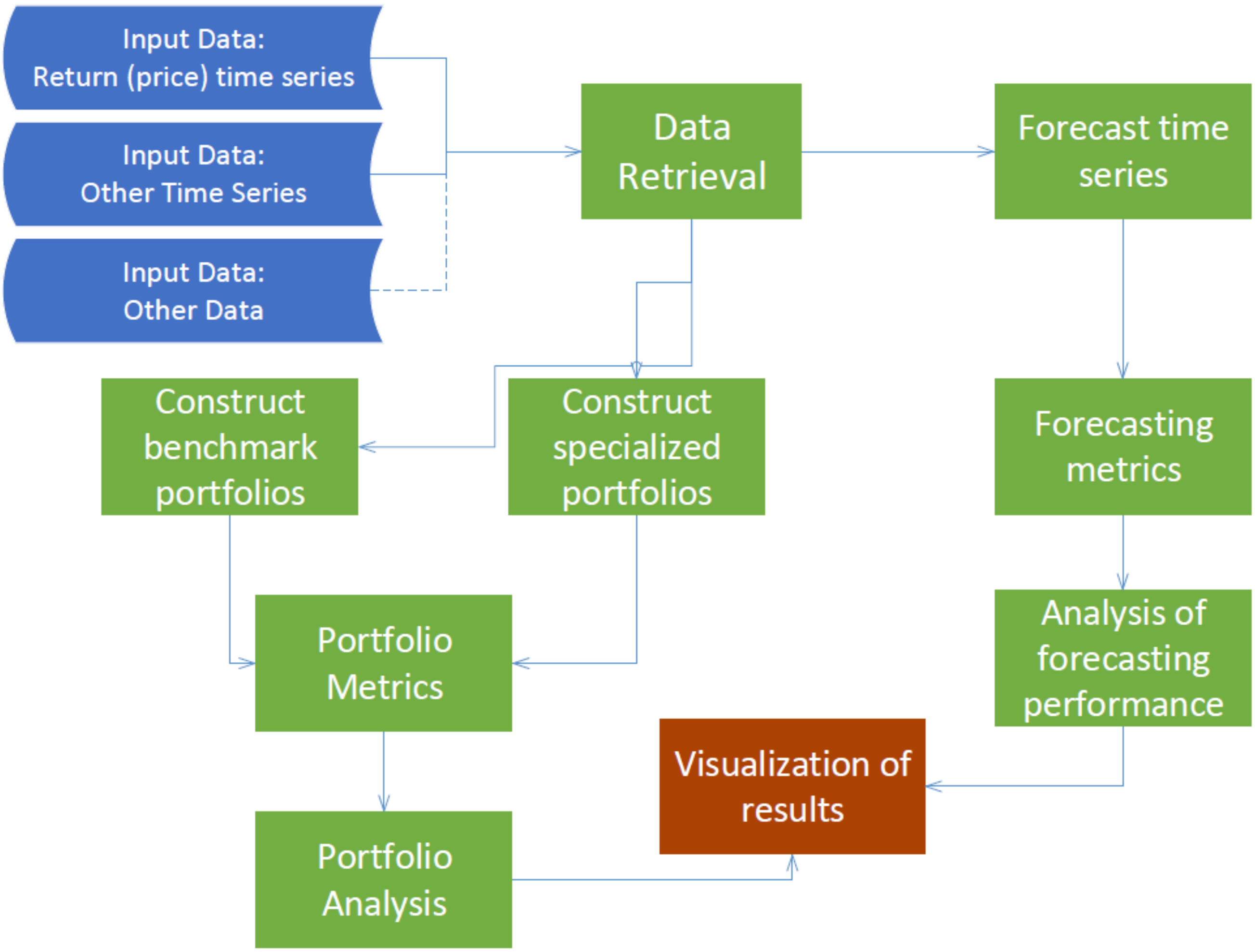
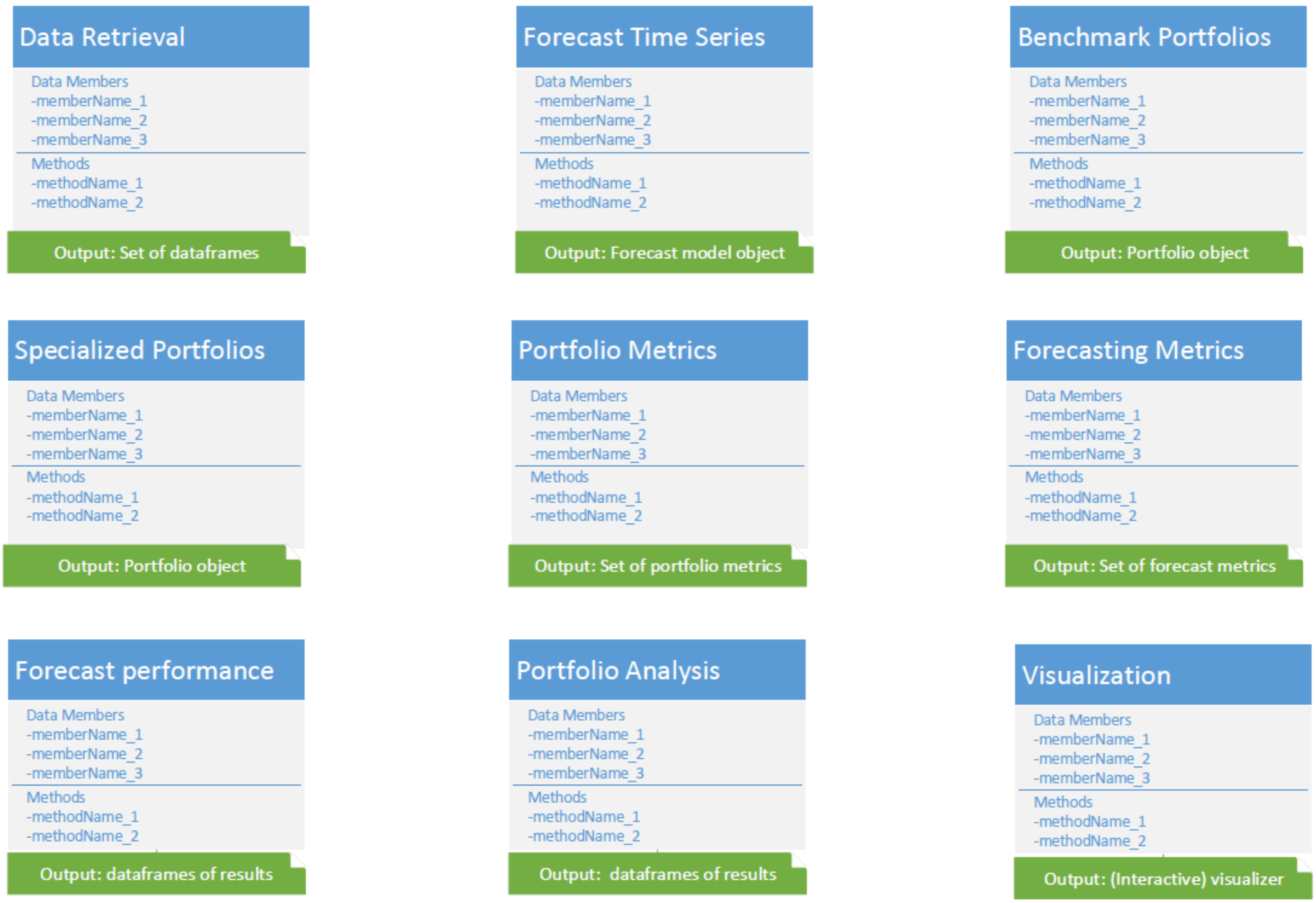


Figure 2: UML diagrams for each of the components



Appendix B: Incorporating comparison of portfolio metrics using benchmark portfolios

For your QWIM project it is likely that you would incorporate comparison of portfolio metrics using benchmark portfolios selected from most common "optimal portfolio" types used in the industry and in academia.

To exemplify, see below how such comparison of portfolio metrics can be used within context of a QWIM project based on forecasting. Such analysis based on comparison of portfolio metrics for 2 optimal portfolios can be incorporated within any other QWIM project.

Portfolio optimization methods

List of portfolio optimization methods may include:

- equal weighting
- mean variance optimization (Markowitz)
- minimum variance optimization
- maximum diversification
- risk budgeting/risk parity
- hierarchical risk parity
- Black-Litterman
- robust versions of some the above portfolio optimization methods

Relevant info includes:

- Portfolio Optimization: A General Framework for Portfolio Choice
- Performance of risk-based asset allocation strategies
- Revisiting the Portfolio Optimization Machine Portfolio
- Construction Techniques Applied to Traditional Multi Asset Portfolios

Python and R packages/codes for portfolio optimization

- PyPortfolioOpt:webpage (see also official PyPi webpage)
- Riskfolio-Lib: webpage (see also official PyPi webpage)
- PortfolioAnalytics: webpage (see also official CRAN webpage)
- MLFinLab: webpage
- Codes with location given in Machine Learning in Asset Management: Part 2: Portfolio Construction—Weight Optimization

Portfolio metrics:

List of portfolio metrics may include:

- Sharpe ratio
- Sortino ratio
- Information ratio
- Maximum Drawdown
- expected shortfall

- maximum loss, etc.

Relevant info includes:

- Portfolio metrics
- Picking the Right Risk-Adjusted Performance Metric
- Risk-Adjusted Performance Measurement – State of the Art
- An Investor's Guide to the Risk Versus Return Conundrum
- How sharp is the Sharpe ratio? Risk-adjusted Performance Measures

Python and R packages/codes for portfolio metrics

- tidyquant: webpage (see also official CRAN webpage)
- PerformanceAnalytics:webpage (see also official CRAN webpage)
- Pyfolio:webpage (see also official PyPi webpage)
- Riskfolio-Lib: webpage (see also official PyPi webpage)
- empyrical: webpage (see also official PyPi webpage)
- JFE: see official CRAN webpage
- QuantStats: webpage (see also official PyPi webpage)
- ffn:webpage (see also official PyPi webpage)
- bt: webpage (see also official PyPi webpage)
- MLFinLab: webpage

Example of comparison of portfolio metrics within context of QWIM project on forecasting

Select a few portfolio optimization methods (choose from the ones implemented in Python and/or R packages mentioned above, such as PyPortfolioOpt) which are based on expected returns and expected covariance matrix.

Then one would create 2 optimal portfolios (let's call them Benchmark and Enhanced) using same portfolio optimization method and information available at the date of portfolio construction:

- expected returns and expected covariance matrix calculated from historical data
- expected returns and expected covariance matrix calculated from forecasted values (obtained using the forecasting model).

Then one would compare side-by-side various portfolio metrics for the 2 optimal portfolios.

NOTE: If you have N forecasting methods used in your coding framrwork, then for each optimizaton method you would end up with (1+N) optimal portfolios

Let's say that you want to construct portfolios at date of June 20, 2019, and you have data as below

- Range of entire dataset: January 1st, 1990 - August 1, 2020
- Range of Training dataset: January 1st, 1990- February 20, 2017
- Range of Test dataset: February 20, 2017 - August 1, 2020

Then for Benchmark portfolio:

- vector of expected means is calculated based on historical data available at June 20, 2019 (namely from 1990 to June 19, 2019)
- expected covariance matrix is calculated based on historical data available at June 20, 2019 (namely from 1990 to June 19, 2019)

For Enhanced portfolio:

- vector of expected means is calculated based on forecasted values available at June 20, 2019 and obtained using the forecasted model trained on given training dataset (which is from 1990 to 2017)
- expected covariance matrix is calculated based on forecasted values available at June 20, 2019 and obtained using the forecasted model trained on given training dataset (which is from 1990 to 2017)

Then you would compare various portfolio metrics among the two portfolios. These metrics can be calculated on following time periods:

- from date of portfolio construction (June 20, 2019) to last date for which you have data (August 1, 2020)
- from starting date of dataset (January 1st, 1990) to last date for which you have data (August 1, 2020)
- from starting date of dataset (January 1st, 1990) to date of portfolio construction (June 20, 2019)

So you would have side-by-side comparisons of portfolio metrics for each of the above 3 time periods.

Portfolio metrics can be calculated using various Python and/or R packages mentioned above.