Master of Science in Mathematical Statistics

Dissertation: Final Draft

Systematic Asset Allocation Using Flexible Views for South African Markets

Ann Sebastian

Supervisor: A/Prof. T. Gebbie

Mathematical Statistics,
Department of Statistical Sciences,
University of Cape Town, Rondebosch

2020

Abstract

We implement a systematic asset allocation model using the Historical Simulation with Flexible Probabilities (HS-FP) framework developed by Meucci [142, 144, 145]. The HS-FP framework is a flexible non-parametric estimation approach that considers future asset class behavior to be conditional on time and market environments, and derives a forward looking distribution that is consistent with this view while remaining as close as possible to the prior distribution. The framework derives the forward looking distribution by applying unequal time and state conditioned probabilities to historical observations of asset class returns. This is achieved using relative entropy to find estimates with the least distortion to the prior distribution. Here, we use the HS-FP framework on South African financial market data for asset allocation purposes; by estimating expected returns, correlations and volatilities that are better represented through the measured market cycle. We demonstrate a range of state variables that can be useful towards understanding market environments. Concretely, we compare the out-of-sample performance for a specific configuration of the HS-FP model relative to classic Mean Variance Optimization(MVO) and Equally Weighted (EW) benchmark models. The framework displays low probability of backtest overfitting and the out-of-sample net returns and Sharpe ratio point estimates of the HS-FP model outperforms the benchmark models. However, the results are inconsistent when training windows are varied, the Sharpe ratio is seen to be inflated, and the method does not demonstrate statistically significant out-performance on a gross and net basis.

Keywords: online learning, technical analysis, portfolio selection, backtesting, overfitting, in-sample, out-of-sample, Johannesburg Stock Exchange

Acknowledgments

I would like to express my gratitude to my supervisor A/Prof. Tim Gebbie for his guidance, encouragement, great suggestions and thorough knowledge of a wide variety of topics that assisted me in completing this dissertation. I respect Tim for his dedication and support. The type of supervisor that assists any time of the day or day of the week. I am privileged to work with a great mind like Tim over the past two years and could not have asked for a better supervisor.

I would also like to thank academics and practitioners such as Andrew Paskaramoothy, Igor Rodionov, Dieter Hendricks, Raphael Nkomo and Lara Dalmeyer for related conversations and discussions related to the work. Particular thanks are extended to Graham Barr for thoughtful comments and critique of an early draft of the summary pre-prints. To my STANLIB colleagues (you know who you are) thanks for all the support!

A big thank you to my parents and in-laws who offered countless hours of support and encouragement. To my feisty baby daughter Daneira Ava Sebastian-Kocheapen who carried me through this dissertation and made this journey much more challenging, interesting and rewarding. Thanks Daneira for giving me an opportunity to be the right kind of role model. To my pillar of strength my husband who had to walk this journey with me on a daily basis through COVID and pregnancy through the highs and lows. I could not have completed this dissertation without you! Most importantly thanks and praise to the almighty God through whom all things are possible.

Related Publications and Pre-prints

Much of the content contained in this paper has been submitted in the form of a pre-print to the arXiv.org e-Print archive and ResearchGate is titled “Systematic Asset Allocation Using Flexible Views for South African Markets” [179, 180].

Declaration

I, Ponni Ann Sebastian, declare that this dissertation titled, “Systematic Asset Allocation Using Flexible Views for South African Markets” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Signed by candidate

Date:

Contents

1 Introduction **14**
1.1 Background . . . 14
1.2 Motivation . . . 15
1.3 Research Objective . . . 17
1.4 Dissertation Structure . . . 17

2 Choice and Utility Theory **18**
2.1 Risk Preference . . . 20
2.1.1 Risk Aversion Coefficient . . . 21
2.2 Indifference Curve . . . 23

3 Markowitz Portfolio Theory **24**
3.1 Diversification . . . 25
3.2 Efficient Frontier . . . 26
3.3 Tobin's Two-Fund Separation Theorem . . . 30
3.4 Problems with MPT . . . 32
3.4.1 Sensitivity and Error-Maximization . . . 32
3.4.2 Failure of Diversification . . . 34

4 Asset Pricing Models **36**
4.1 Stochastic Discount Factor . . . 36
4.2 Arbitrage Pricing Theory . . . 37
4.2.1 Derivation of APT . . . 39
4.3 Extending APT . . . 40
4.3.1 The Non-Linear APT . . . 40
4.3.2 The Conditional APT . . . 41
4.4 Specification of Factors . . . 42
4.4.1 Statistical Factor Models . . . 42
4.4.2 Macroeconomic Factor Models . . . 43
4.4.3 Fundamental Factor Models . . . 43

5 Law of Large Numbers **44**
5.1 Derivation of HS-FP Framework using LLN . . . 44
5.1.1 Classic Historical Estimation . . . 44
5.1.2 Generalization to Flexible Probabilities . . . 45
5.2 LLN and Ergodic Theorem . . . 46

6 HS-FP Model Setup **48**
6.1 Flexible Probabilities . . . 49
6.1.1 Time Conditioned Probabilities . . . 51
6.1.1.1 Rolling Window Probabilities . . . 51
6.1.1.2 Exponentially Decayed Probabilities . . . 52
6.1.2 State Conditioned Probabilities . . . 53
6.1.2.1 Crisp Conditioned Probabilities . . . 54
6.1.2.2 Kernel Conditioned Probabilities . . . 56
6.1.3 Time and State Conditioned Probabilities . . . 58
6.1.3.1 Entropy-Based Conditioning Procedure . . . 59
6.2 Combining Multiple State Variables . . . 62
6.2.1 Equal Weighting . . . 63

6.2.2 Degree of Conditioning and Correlation Weighting 63

7 Benchmark Models 66
7.1 Equal Weighted (EW) 66
7.2 Mean Variance Optimization (MVO) 67

8 Implementation 68
8.1 Data 68
8.1.1 Data on Asset Classes 68
8.1.2 Data on State Variables 72
8.1.2.1 Macroeconomic State Variables 73
8.1.2.2 Financial, Risk and other State Variables 75
8.2 Methodology 76
8.2.1 Data Preparation 76
8.2.2 Forecasting Return and Covariance 76
8.2.3 Portfolio Optimization and Backtesting 77

9 Results 78
9.1 Performance Measures 78
9.2 Comparing Results of Single State Variables 80
9.3 Comparing Results of Multiple State Variables 81
9.4 Comparing Out-of-Sample HS-FP Results to Benchmark Models . . . 82
9.4.1 Comparison of Allocations over Time 82
9.4.2 Comparison of Out-of-Sample Statistics 83
9.4.3 Comparison of Out-of-Sample Rolling Returns 84
9.5 Comparison with Multi-Asset Funds 86

10 Robustness of Out-of-Sample Results 87
10.1 Transaction Cost 87
10.2 Training Window 88
10.3 Probabilistic Sharpe Ratio 90
10.4 Probability of Backtest Overfitting 91
10.5 Deflated Sharpe Ratio 95

11 Conclusions 97
11.1 A Systematic Asset Allocation Framework 97
11.2 Implementation and Performance 98
11.3 Approach in Practice 99
11.4 Future Work 99

12 References 101

13 Appendix A 112
13.1 Derivation of MV Frontier 112
13.2 Derivation of MV Frontier with Risk-Free Asset 112
13.3 Glivenko-Cantelli Theorem 113
13.4 Deflated Sharpe Ratio Calculation 114

14 Appendix B 116
14.1 Code patterns 116
14.1.1 Workflow for HS-FP Framework 116
14.1.2 Conditional HS-FP Function 122
14.1.3 Crisp Probability Function 123

14.1.4 HS-FP CDF Function 124
14.1.5 Minimum Relative Entropy Function 125
14.1.6 Effective Number of Scenarios Function 128
14.1.7 HS-FP Mean Covariance 129

List of Figures

1 Shows the utility function shapes for different risk preference. The y axis represents the investor's utility and the x axis shows the wealth levels. The figure shows how investor's utility from wealth changes for changes in wealth in different types of investor's namely risk averse, risk neutral and risk loving. 21
2 The investor derives different levels of utility on each of the indifference curves (C1, C2, and C3). However, the investor derives the same level of utility on any point say S1 and S2 on the utility curve C1. Additionally, if an investor has to decide between points S4 and S6 an investor would choose point S6 because the investor obtains the same amount of return for a lower level of risk on point S6. Hence, the utility of a risk-averse investor increases as you move leftwards from indifference curve C1 to C3. 23
3 The dots on the efficient frontier represent optimal portfolios for different levels of risk tolerance. The dot labeled GMV represent the Global Minimum Variance portfolio. The dots labeled D, A, MVP, MRR and GMV are all optimal portfolios. The dot labeled MVP is the Mean Variance Portfolio is the optimal portfolio for an investor who has risk tolerance depicted by his highest indifference curve IC1. The dot labeled MRR is the Maximum Risk adjusted Return portfolio. The difference in risk levels between the GMV and any other portfolios represents unsystematic risk 27
4 When the risk-free asset is included the efficient frontier becomes a straight line called the Capital Market Line (CML) which is tangent to the risky asset efficient frontier at a single point MSR (Maximum Sharpe Ratio). The MSR dot represents the portfolio which is the optimal mix of risky assets in the mean variance space. All efficient portfolios are a mixture of the MSR tangency portfolio and the risk-free asset. 31
5 Shows an example of how to obtain time conditioned probabilities using: 1) the rolling window approach where $\lambda = 60$ months, and 2) an exponential decay approach where $\tau = 24$ months. The grey region represents the weight or probability applied relative to the corresponding calendar date. 53
6 Here we consider month-on-month changes in inflation, CPI (Table 3), as a state variable. We plot the raw CPI on y_1 axis which we then smooth, score and finally standardize into the market state variable which we plot on y_2 axis, and condition by target state $CPI = -1.04$, which is the latest CPI market state value as shown by the red line. . 57
7 Shows an example of how to obtain state conditioned probabilities using: 1) crisp conditioning probability, and 2) kernel conditioning probability with h = 0.5 and $\gamma = 2$. The state variable used to condition is CPI and the target state is CPI=-1.04. The grey region represents the weight or probability applied relative to the corresponding calendar date. 58

8 The first three panels show an example of obtaining time, state, time and state conditioned probabilities using a single state variable. The state variable used to condition is CPI and the target state is CPI=-1.04. The grey region represents the weight used relative to the corresponding calendar date.The first panel shows time conditioning by exponential decayed probability where τ= 24 months, the second panel shows state conditioning by crisp conditioning, the third panel shows time and state conditioning using entropy pooling approach. Finally the last panel shows how the Resource 20 Index returns (see Table 1) can be conditioned by time and state using entropy pooling approach. 62
9 Shows an example of how to obtain time and state conditioned probabilities using multiple state variables CPI and GDP. In the first panel we condition only by the latest market state of CPI state variable and overlay the time and state conditioned probability. In the second panel we condition only by the latest market state of GDP state variable (see Table 3) and overlay the time and state conditioned probability. In the third panel we show the ensemble probability when conditioning by both time and multiple state variables. Finally in the last panel we show the Resources 20 Index (see Table 1) conditioned by both time and the multiple state variables CPI and GDP. 65
10 Rolling 1 Year return for Asset Classes: A comparison of the 1 Year rolling return for the industrial, financial, resource, bonds and currency asset classes over the period February 1999 until Dec 2017. 69
11 Rolling 1 Year Volatility for Asset Classes: A comparison of the 1 Year rolling volatility for the industrial, financial, resource, bonds and currency asset classes over the period February 1999 until Dec 2017. . . . 70
12 Rolling 1 Year Sharpe Ratios for Asset Classes: A comparison of the 1 Year rolling Sharpe ratio for the industrial, financial, resource, bonds and currency asset classes over the period February 1999 until Dec 2017. 71
13 Correlation Matrix: A full sample correlation matrix for the industrial, financial, resource, bonds and currency asset classes over the period February 1999 until Dec 2017. 72
14 Standardized values for macroeconomic state variables over the period 31 Jan 1999 to 31 Dec 2017. 74
15 Standardized values for financial, risk and other state variables over the period 31 Jan 1999 to 31 Dec 2017. 75
16 Optimal Asset Allocation: The out of sample optimal allocation to each asset class over time is shown for the three portfolio construction methods considered; the Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimization (MVO) (see Section 7.2) and Equal Weighting (EW) (see Section 7.1) respectively. 83
17 Cumulative Return for Asset Allocation Models: A comparison of the out of sample cumulative returns for the Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimal (MVO) (see Section 7.2) and Equally Weighted (EW) (see Section 7.1) asset allocation models. 84
18 Rolling Relative Returns for Asset Allocation Models: A comparison of the 1 Year out of sample relative returns for the Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimal (MVO) (see Section 7.2) and Equally Weighted (EW) (see Section 7.1) asset allocation models. 85

19 Shows the distribution of Logits and Probability of Backtest Overfitting (PBO) for the HS-FP model, the Logits is the logarithm of odds, and in our analysis the odds represents the odds that the optimal strategy chosen IS will underperform OOS. The Logits are then used to calculate the PBO. When the distribution of Logits is centered in a significantly positive value with left tail marginally covering the region of negative logits this implies the backtest results are conducive to good OOS results. 93
20 Shows the pairs of (SR IS,SR OOS) for optimal model configuration or the performance degradation associated with HS-FP model backtest and the OOS probability of loss. 94
21 Shows the stochastic dominance of HS-FP with the cumulative distribution function of best OOS SR across all combinations (optimized) and OOS SR (non-optimized), as well as the second order stochastic dominance (SD2) for every OOS SR. 95

List of Tables

1 Asset Classes: The list of asset classes is given with their Bloomberg codes and the proxies used to represent these asset classes. 68
2 Indicative Asset Class Performance: The annualised returns, annualised volatilities, and finally, annualised Sharpe ratios are provided as summary statistics for the asset classes over the period February 1998 until December 2017. 69
3 State Variable Signals: The list of state variables used as investment signals, along with their reporting frequency and Bloomberg codes are provided; these include variables from both domestic and global econometric candidate signals and cover macroeconomic, risk and trend based variables. The list is indicative and not exhaustive. The sampling frequencies are Monthly (M) and quarterly (Q). 73
4 Single State Variable Summary Statistics I: The state variables listed in Table 3 are used individually as the only state variable to condition the HS-FP model; popular summary statistics such as gross annualised return, annualised volatility, annualised Sharpe ratio, maximum draw down, CVaR and turnover (see Sec 9.1 for details regarding the calculation of these ratios) are shown for the March 2003 - Dec 2017 out of sample period. 81
5 Multiple State Variable Summary Statistics I: The state variables listed in Table 3 are used in combination to condition the HS-FP model; two methods to combine information from each state variable are assessed. Equal weighting (EQ) (see Section 6.2.1) where the conditioned probability from each state variable is equally weighted to give final probability and DCC (see Section 6.2.2) where the conditioned probability from each state variable is weighted by Degree of Conditioning and Correlation (DCC) between the different state variables. Popular summary statistics such as gross annualised return, annualised volatility, annualised Sharpe ratio, maximum draw down, CVaR and turnover are shown for the March 2003 - Dec 2017 out of sample period. 81
6 Asset Allocation Models Summary Statistics I: Popular summary statistics for the three portfolio construction methodologies covered in this paper are considered, Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimal (MVO) (see Section 7.2) and Equally Weighted (EW) (see Section 7.1) respectively. The gross annualised return, annualised volatility, annualised Sharpe ratio, maximum draw down, CVaR and turnover summary statistics are provided for the out of sample period of March 2003 until December 2017. 83
7 Summary Statistics for Asset Allocation Models relative to Balanced Funds: Popular summary statistics such as annualised return, annualised volatility and annualised Sharpe ratio for the Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimal (MVO) (see Section 7.2) and Equally Weighted (EW) (see Section 7.1) asset allocation models are compared to South African and Global Multi-Asset Flexible Category funds over the period Jan 2013-Dec 2017. 86

8 Summary Statistics for Transaction Cost = 0, 20, 30 and 50 bps: Summary statistics such as annualised return, annualised volatility, annualised Sharpe ratio, maximum drawdown, conditional value at risk and average monthly turnover are shown for the three asset allocation models (see Section 6, Section 7.2 and Section 7.1); when the transaction cost is set to 0, 20, 30 and 50 bps. The out of sample performance is observed over the March 2003 - Dec 2017 period. Here we see that when transaction costs are increased it does reduce the HS-FP's Sharpe ratio relative to benchmark models, however it remains the highest Sharpe ratio, even at a transaction cost as high as 50 bps. 88

9 Summary Statistics for Training Window = 5, 8, 10 and 12 years: Summary statistics such as annualised return, annualised volatility, annualised Sharpe ratio, maximum drawdown, conditional value at risk and average monthly turnover are shown for the three asset allocation models (see Section 6, Section 7.2 and Section 7.1); when the initial training window is set to 5, 8, 10 and 12 years. The out of sample returns are observed over the March 2003 - Dec 2017 period. Here we see that when training windows are altered it does impact the Sharpe ratio of HS-FP relative to benchmark models and when a training window of 12 years is used the HS-FP does underperform the MVO benchmark model. . . 90

10 PSR Matrix: The table represents the PSR matrix, the rows of the table represent the candidate strategy and the columns represent the benchmark strategy. The table should be interpreted as the probability that the candidate strategy SR exceeds the benchmark strategy SR. . 91

1 Introduction

1.1 Background

One of the most important decisions in the asset management industry is deciding how much to allocate to the various asset classes. Formally, asset allocation is the process of deciding how much to allocate to asset classes by gathering and processing market related data in order to deliver on client objectives [199]. Asset allocation is considered the most important determinant of portfolio outcomes [30]. It is generally the most difficult decision for investors to make and is complicated by the restrictive nature of constraints in portfolio construction. The optimal asset allocation should be determined based on a client's: unique needs, return specification, risk tolerance, investment horizon and fee preference.

In practice there are two popular approaches to specify a client's optimal asset allocation: the long term strategic asset allocation approach, and the dynamic asset allocation approach. Within the dynamic asset allocation approach there are different investment management styles that can be used, such as, the discretionary or quantitative approaches. Here a discretionary approach is solely based on the view of a portfolio manager regarding asset prices. On the other hand, a quantitative approach is a systematic, rules based investment process, some of which use stochastic models to generate asset prices. Historically, majority of asset management houses utilize a discretionary approach where the ultimate decision of how much to allocate to an asset class is based on the view of the portfolio manager.

The discretionary view is often grounded on the portfolio manager's belief of mispricing opportunities based on a fundamental valuation exercise. In order to make the decision the portfolio manager(s) requires assistance from an expansive team of specialists such as sector analysts, asset region and macro specialists. These roles are necessary due to the extensive fundamental and macroeconomic data that needs to be manually analyzed for the vast number of companies, sectors and asset classes that exist in both a local and global context. Due to this substantial manpower requirement discretionary funds are generally offered to clients at higher costs than quantitative active funds.

When John Bogle launched the first passive fund his main objective was to provide clients with access to investment products at the lowest possible cost. The passive funds which utilized clearly defined rules to replicate the market, required minimal man power and hence could be offered at significantly lower costs to the comparable active funds. Once these funds became mainstream with a live track record a fascinating observation was noticed, majority of active funds underperformed their passive fund counterparts. Interestingly, the average underperformance of active funds after fees was seen to be more than the cost reduction of passive funds. This sparked the interest in systematic funds: Was the real benefit of passive funds the non-emotional rules based investment approach? That information was dilute by the very nature of human intermediation?

The advancements in quantitative portfolio theory, confluence of technologies and rise in computing power have had altering effects on the asset management industry. Since the creation of the first quantitative portfolio theory there has been further enhancements in the estimation and attribution behind the science of investments. The technological advancements have benefited the industry through access to machines that can efficiently structure, analyze and summarize large complex data sets in a matter of seconds. These advancements have benefited clients by providing insight on the true drivers of portfolio risk, covariation and returns. This has resulted in a broader range of investment products such as smart beta funds and other systematic solutions with a commonality of lower costs and systematic investment approach. Human assistance is still necessary to identify what data is useful and what problems require attention.

For investors to decide how much to allocate to various asset classes they need to gather views on each of the asset classes. Asset class views are notoriously correlated, dynamic and noisy. The key inputs required in any quantitative portfolio construction process grounded in modern portfolio theory are the underlying market risk driver(s), return views and views on the risk and dependence structure of the asset class [50]. However, these inputs are unknown and need to be estimated by market participants.

Market participants can estimate these asset characteristics from historical data or simulated data. Here we follow a data-informed approach and use historical data. Some of the benefits of using historical data is that it is simple to implement and represents the actual asset class history so assumptions on asset class distribution, dependence structure and other features need not be explicitly made. However, naively using measured historical data of financial markets presupposes that past experience will somehow match future experience. This is unlikely.

The reality is that we only have a single measured realization of reality and do not explicitly know the probability of the sequence of history we have measured, nor do we have a concrete measure of how quickly information dissipates, how agents and market participates adapt to the data they measure, nor how quickly our models can be expected to break-down. This opens all modeling using measured financial market history to potentially wild forms of generalisation error.

1.2 Motivation

Quantitative approaches can generally use complex statistical models and are often labeled as a black box approach. The popularity of discretionary asset management has been mainly driven by one aspect, a charismatic well articulated portfolio manager who can explain his/her views in a simple manner to an investor. The portfolio manager in explaining his/her view will often refer to the future macro environment tugging on specific market variables such as GDP, inflation, volatility and so on to substantiate the validity of the their view. This aspect is generally missing in quantitative asset management where the prevailing perception is that of a technically complicated process that a man on the street cannot understand. A quantitative approach is required that is simple to understand and can incorporate popular market variables into the asset allocation process.

When asset class returns are observed through a market cycle it can be seen that the returns vary through time and market regimes [85]. Different asset classes perform differently in different market conditions and no single asset class dominates in all market conditions. Additionally, state variables exist that are useful in distinguishing market conditions and hence forecasting asset class performance. State variables are observable indicators that measure the state of the market. There is a variety of state variable indicators ranging from macroeconomic, financial, risk and other categories. Flavin and Wickens [75] provide a detailed overview of all previous literature on utilizing macroeconomic and financial variables to forecast asset class returns. Although there has been an abundant academic literature supporting asset class return predictability, as summarised by Rapach and Zhou in [162], and that there is in general consensus that in sample asset class returns do display predictability [36] there is far less consensus regarding out of sample asset class return predictability [201].

A lot of interest has been gathering in the quantitative portfolio management community regarding the robustness of backtest results and the need to assess the potential degree of backtest overfitting in a suggested investment strategy. This interest has been primarily driven by the observation that many backtests of quantitative strategies have good in-sample performance but when these strategies are applied out-of-sample the results are significantly poorer. It has been argued that this observation arises from data snooping and chasing the best in-sample strategy which then forms the basis for the investment approach utilized on an out-of-sample basis. An investment strategy grounded on these mistakes are flawed from the onset. There needs to be a clear dissection into the performance of any backtest to assess whether the performance of the strategy was due to luck as a result of unintended biases or due to the skill of the investment strategy.

A systematic approach to asset allocation brings discipline by implementing a rules based approach to include relevant information about the market into the return prospects of asset classes [2]. For these reasons an asset allocation model is required that is adaptive to changing market conditions and can incorporate historical data with investor views in a mathematically robust manner. Investors typically trust recent history more than those from far in the past. With this in mind we considered the "Historical Simulation with Flexible Probabilities" (HS-FP) framework developed by Meucci [142, 144, 145]. The key features of this framework are that it is, first, a non-parametric approach, that it is second, able to incorporate historical data with investor views on state variables; hereby generating time and state conditioned probabilities. The type of investor view we make use of relates to the qualitative question investors often ask: *"When in recent history were state variables most similar to today's levels?"*

The aim here is to then estimate empirical distributions which reflect the current market regime. This predictive requirement can be implemented using the most recent mixed distribution estimate to form subsequent period return forecast. The paper will also consider using state variables for asset class return predictability. The standard historical approach to asset class return predictability is to use all historical data equally to forecast asset class returns. Here we deviate from this and apply non-equal time and state variable conditioned probabilities to historical returns. To draw conclusions about the efficacy of this approach to asset allocation we explicitly compare the HS-FP approach to the classic mean variance optimization approach [135] and a naive equal weighting approach observing the out-of-sample returns and other descriptive statistics. We then assess the robustness of these results and assess the potential of backtest overfitting.

1.3 Research Objective

The aim of this research is to showcase a systematic asset allocation model that incorporates popular market state variables in a mathematically sound manner. Specifically the approach utilizes the HS-FP approach of Meucci's and tests the efficacy of the model in the South African context.

The objectives of this research can be stated as follows:

1. Demonstrate the use of the HS-FP model on South African financial market data for systematic asset allocation purposes;
2. Display a range of explicable state variables that can be used to understand market conditions;
3. Evaluate whether this time and state conditioned risk and return forecasts of asset class return distribution can provide any significant improvement on portfolio performance when comparing the out of sample performance of the HS-FP model to benchmark models.

The success of the research is based on a thorough analysis of the out of sample performance of the HS-FP model. Although the key contribution of this paper is to consider the potential effectiveness of a simple HS-FP implementation in the context of South African markets, the work can be of interest to a more general audience as it links the HS-FP framework to an explicit benchmark based test, and a quantitative reflection on the perils of backtest overfitting as we step into the post-quant world of data-informed asset management aimed at making better decisions and not merely reducing costs. The performance of the HS-FP model relative to benchmark models is primarily assessed based on which strategy provides superior risk adjusted returns however other performance, risk and turnover statistics are also assessed. The benchmark models used here are the equally weighted and the classic mean variance optimization models. The HS-FP model is also interrogated for robustness by varying assumption of transaction costs and training windows and by observing crucial backtest overfit statistics.

1.4 Dissertation Structure

Our aim is to showcase a systematic asset allocation model in the South African context. In Section 2 - 5 we summarize the main academic literature that is key to the development of optimal portfolio choice and asset pricing. In Section 6 we discuss the main focus of the thesis which is detailing the HS-FP model setup of Meucci. In Section 7 we cover the benchmark models used, in Section 8 we outline the implementation which covers the description of the data, data processing steps and the methodology applied, in Section 9 we analyze the out-of-sample results of conditioning by a single state variable, multiple state variables and finally the performance of the HS-FP model against the benchmark models. In Section 10 we discuss whether these results can be considered sufficiently robust to be practically useful in asset management use-cases. Finally, in Section 11 we have the conclusion.

2 Choice and Utility Theory

The terms uncertainty and risk are commonly used interchangeably and are generally understood to have the same meaning. Frank Knight [120] provided a fundamental distinction between the two terms stating that risk refers to a situation where the outcome is unknown, but the distribution or probability of its occurrence is known, while uncertainty refers to a situation where both the outcome and distribution are unknown. Most choices are made somewhere in the middle between known and unknown distributions, so we assume uncertainty and risk to have the same meaning, that is, a situation where the outcome is unknown, but the distribution is known. As will be explained in Section 4 there is uncertainty or risk in the asset pricing process in both the future asset cash flows and how the individual will discount the future cash flows.

Utility theory is the foundational normative theory for choice under uncertainty and is an important concept in economics and mean variance portfolio theory which we will detail in Section 3. Utility theory aims to depict the preferences of an individual by an utility function. An utility function measures an individual's relative preference for different levels of total wealth. Due to expected value being a constraining theory of risk bearing utility theory was originally suggested as an alternative. Pierre de Fermat, Blaise Pascal and Christiaan Huygens developed modern probability theory in the 17th century and proposed that the attractiveness of a gamble offering the payoffs $(x_1, \ldots, x_n)$ with probabilities $(p_1, \ldots, p_n)$ was given by its expected value $\mathbb{E}(\boldsymbol{x}) = \sum x_i p_i$.

Daniel Bernoulli in 1738 challenged the idea that rational decisions were based on purely the expected value by introducing the term 'expected utility'. He used this term to solve the St. Petersburg paradox. The St. Petersburg paradox posed the following situation: "a fair coin will be tossed until a heads appears. If the first heads appears on the n^{th} toss, then the payoff is 2^n ducats. How much should one pay to play this game?" [76]. He showed that if the expected value was calculated the answer would be infinite:

$$\mathbb{E}(\boldsymbol{x}) = \sum_{i=1}^{\infty} \frac{1}{2^n} \cdot 2^n = \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 2^2 + \frac{1}{8} \cdot 2^3 + \ldots = 1 + 1 + 1 + \ldots = \infty, \tag{1}$$

suggesting that an individual would be willing to pay infinite amounts of money to play this game. Herein lies the paradox, as, in reality, it is unlikely that a rational individual would be willing to pay an infinite amount of money to play this game. His conjecture was that the expected value $\mathbb{E}(\boldsymbol{x}) = \sum x_i p_i$ was insufficient for solving this paradox and that instead the expected utility $\mathbb{E}(\boldsymbol{u}) = \sum U(x_i) p_i$ needed to be calculated to solve this paradox. He showed that a rational individual would make decisions based on the expected utility of the game, because this would have a finite answer:

$$\mathbb{E}(\boldsymbol{u}) = \sum_{i=1}^{\infty} \frac{1}{2^n} \cdot U(2^n) = \frac{1}{2} \cdot U(2) + \frac{1}{4} \cdot U(2^2) + \frac{1}{8} \cdot U(2^3) + \ldots < \infty. \tag{2}$$

Bernoulli's utility function was a strictly concave logarithmic function, such that an individual's expected utility is less than its expected value. Bernoulli assumed most individuals are risk averse. This solution to the paradox revolutionized the field of economics by introducing two new ideas. The first was the concept of diminishing marginal utility which states that the individual's utility from wealth $U(\boldsymbol{w})$ is not linearly related to wealth ($\boldsymbol{w}$) but rather increases at a decreasing rate i.e. $U'(\boldsymbol{w})>0$ and $U''(\boldsymbol{w})<0$. The second was that rational decisions regarding risky ventures should be based on expected *utility* of the venture rather than only the expected *value* of the venture, expressing that a gain or loss of money might mean different things to different individuals and is dependent on their wealth.

While Bernoulli's ideas formed the basis of expected utility theory by linking utility to wealth, it was John von Neumann and Oskar Morgenstern [198] who provided the mathematical foundation for Expected Utility Theory (EUT). According to von Neumann and Morgenstern the concept of rationality implies that an individual's utility function satisfies certain axioms of rational choice. These axioms result in a set of assumptions about an individual's preferences and is required before one can construct a utility function. These axioms provide certain choice features such as completeness, transitivity, continuity, and independence which ensure that the utility function is well behaved. Neumann-Morgenstern's EUT states that the rational choice by an individual in the face of uncertainty is to choose the action that maximizes the individual's expected utility. Unlike Bernoulli who assumed most individuals are risk averse in Neumann-Morgenstern's EUT framework an individual can have different attitudes towards risk as will be discussed in Section 2.1.

EUT is the foundational normative theory, and normative theory details how individuals should make decisions in the face of risk. Normative theory is not without its flaws, as many experiments have been performed and documented such as Simons' critique [99] that states that in reality when individuals make decisions they do not obey these rational choice axioms. Simons' critique on the normative theory states that human beings do not possess sufficient cognitive abilities to understand and synthesize all the relevant information to make a rational choice. The identification of these flaws have brought about descriptive theories for decision making.

Descriptive theory studies show how rational or irrational individuals actually make decisions in reality. The most recognized descriptive theories are prospect and regret theory. Prospect theory revolves around the concept of loss aversion which implies that individuals react differently to decisions with potential gains versus potential losses [112]. An individual faced with a risky decision that has potential for gains will be risk averse; whereas that individual while faced with a risky decision that has potential losses the individual, will be risk loving. This deviates from EUT which only considers decisions which maximizes the expected utility. Regret theory on the other hand revolves around the concept of 'anticipated regret', a regret term is incorporated into the utility function [48]. Regret theory always contradicts the transitivity axiom of EUT.

The third school of decision-making theory is prescriptive theory. Although descriptive theory covers how individuals actually make decisions, the theory does not offer any advice on how to make strategic decisions. Prescriptive theory assists to help improve the decisions of individuals. The theory is grounded on helping individuals improve their decision-making process by providing constructive aid to help them make better choices [200]. In prescriptive theory individuals are perceived as not necessarily rational but aspiring to become rational individuals. Prescriptive theory forms part of the operations and management science category. The application of prescriptive theory in reality has been poor due to limited application in reality and hence prescriptive theory has had minimal contribution to decision making process in business and government [33].

Despite its flaws, EUT allows one to set up an utility function for decision making by a rational individual. In reality, each individual has his own utility function and these utility functions need to be identified from infinite possibilities. Once the individual's utility function is determined it serves as a proxy for how the rational individual makes decisions. The rational investor's optimal investment choice is the investment that maximizes the individuals expected utility of wealth [168], hence the choice of the utility function is important in quantitative portfolio management.

In the following sections the basic theory of risk preference is mentioned and the main forms of utility curves and utility functions that are commonly used in investment management are discussed. A brief introduction to indifference curves is also provided. This allows us to establish the theoretical links to the mathematical formulation used in Section 3 and is the basis for the alternative formulations of portfolio optimizations based on utility theory and risk aversion.

2.1 Risk Preference

A common requirement for a utility function is that it represents the investor's attitude towards risk adequately. An investor's risk preference can be categorized as either risk averse, risk neutral or risk loving. A risk averse investor is one who does not like taking risk, preferring an investment with lower risk to one with higher risk at the same expected return. A risk loving investor is one who is attracted to risk, preferring an investment with higher risk to one with lower risk at the same expected return. A risk neutral investor is one who is insensitive to risk, effectively preferring to ignore the risk component in any investment and is purely concerned with the expected return of the different investment choices.

These preferences toward risk are captured in the curvature of an investor's utility function. We illustrate the differing shapes of the utility function below, this illustrates how an investor's attitude to risk can vary, displaying three very different investor preference and approaches to risk at varying levels of wealth.

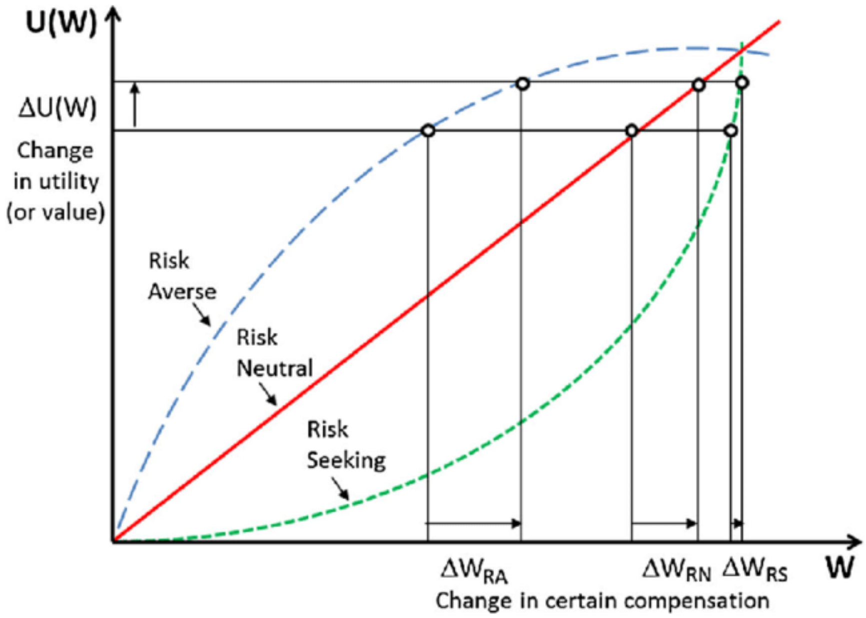

Figure 1: Shows the utility function shapes for different risk preference. The y axis represents the investor's utility and the x axis shows the wealth levels. The figure shows how investor's utility from wealth changes for changes in wealth in different types of investor's namely risk averse, risk neutral and risk loving.

The curvature of the utility function for the three investors in Figure 1 tells us whether they are risk averse, neutral or loving. The blue line represents a risk averse investor with a concave utility function, implying that the investor's preference for risk increases at a decreasing rate with increasing wealth. The red line represents a risk neutral investor with a linear utility function, implying that the investor's preference for risk increases at a constant rate with increasing wealth. The green line represents a risk seeking investor with a convex utility function, implying that the investor's preference for risk increases at an increasing rate with increasing wealth.

2.1.1 Risk Aversion Coefficient

A risk averse investor is said to prefer a sure bet over a gamble [77]. In reality, individuals tend to have different degrees of risk aversion. If the relationship between utility and wealth can be specified through a utility function, then the risk aversion coefficient measures how much utility we gain (or lose) as we add to, or subtract from, our wealth [51].

The risk aversion is determined by the curvature of a utility function, the greater the curvature the greater the degree of risk aversion. A widely used measure to calculate the risk aversion are the Arrow Pratt coefficients of absolute and relative risk aversion.

The Arrow-Pratt measure of absolute risk aversion is defined as:

$$r_A(\boldsymbol{x}) = -\frac{u''(\boldsymbol{x})}{u'(\boldsymbol{x})}. \tag{3}$$

The Arrow-Pratt measure of relative risk aversion is defined as:

$$r_B(\boldsymbol{x}) = -\frac{xu''(\boldsymbol{x})}{u'(\boldsymbol{x})}. \tag{4}$$

Here, the first derivative of the utility function $u'(\boldsymbol{x})$ represents an individual's degree of risk aversion, and the second derivative of the utility function $u''(\boldsymbol{x})$ represents how an individual's risk aversion itself changes as a function of wealth.

Although $u'(\boldsymbol{x})$ measures an individual's risk aversion we need an easy measure that allows us to compare the risk aversion across individuals with different utility functions. To deal with this issue Arrow-Pratt suggested to calculate $u''(\boldsymbol{x})$ and divide it by $u'(\boldsymbol{x})$ to obtain the absolute risk aversion. A risk averse investor has an Arrow-Pratt measure that is positive and decreasing for increasing levels of wealth (Varian, 1999).

The first derivative of $r_A(\boldsymbol{x})$, $r'_A(\boldsymbol{x})$ points to whether individuals have an increasing, constant or decreasing absolute risk aversion. When $r'_A(\boldsymbol{x}) > 0$, this means that as an individual's wealth increases the individual holds less dollars in the risky asset. When $r'_A(\boldsymbol{x}) = 0$ this means that as an individual's wealth increases the individual holds the same dollars in the risky asset. Lastly, when $r'_A(\boldsymbol{x}) < 0$ this means that as an individual's wealth increases the individual holds more dollars in the risky asset.

The relative risk-aversion $r_B(\boldsymbol{x})$ measures the individual's sensitivity to a percentage change in their wealth. The first derivative of $r_B(\boldsymbol{x})$, $r'_B(\boldsymbol{x})$ points to whether an individual has an increasing, constant or decreasing relative risk aversion. When $r'_B(\boldsymbol{x}) > 0$, this means that as an individual's wealth increases the individual decreases the percentage held in the risky asset. When $r'_B(\boldsymbol{x}) = 0$ this means that as an individual's wealth increases the individual keeps the percentage held in the risky asset constant. And lastly, when $r'_B(\boldsymbol{x}) < 0$ this means that as an individual's wealth increases the individual increases the percentage held in the risky asset.

An agreeable and tractable utility function requires certain functional assumptions about the form of the utility function. Hyperbolic Absolute Risk Aversion (HARA) refers to a type of risk aversion that is particularly convenient to model mathematically and to obtain empirical predictions from. Two popular forms of utility functions that are special cases of the Hyperbolic Absolute Risk Aversion (HARA) utility function is the Constant Absolute Risk Aversion (CARA) and the Constant Relative Risk Aversion (CRRA) utility functions. These functions exhibit HARA because their absolute risk aversion is a hyperbolic functions.

CARA implies that as an individual's wealth increases, the individual holds the same dollars in the risky asset. Because the function displays no wealth effect, this function renders poorly to utility modeling. The CRRA utility function, on the other hand, implies that as an individual's wealth increases the individual holds the same percentage in the risky asset, hence the function does display wealth effects. Some examples of CRRA utility function include the power and logarithmic utility functions. In general, the common forms of utility functions are linear, quadratic, exponential and power utility functions .

2.2 Indifference Curve

In economics, an indifference curve is used to visually represent all combinations of goods that provide the same utility for an individual [6]. In portfolio theory, the indifference curve visually represents all combinations of portfolios that provide the same utility for an investor, where an investor's preference is traditionally based on two dimensions, the portfolio's risk and return. A map of indifference curves then represents the infinite set of unique indifference curves for each investor that is derived in conjunction with the investor's unique utility function. The following graph shows three indifference curves for the same investor.

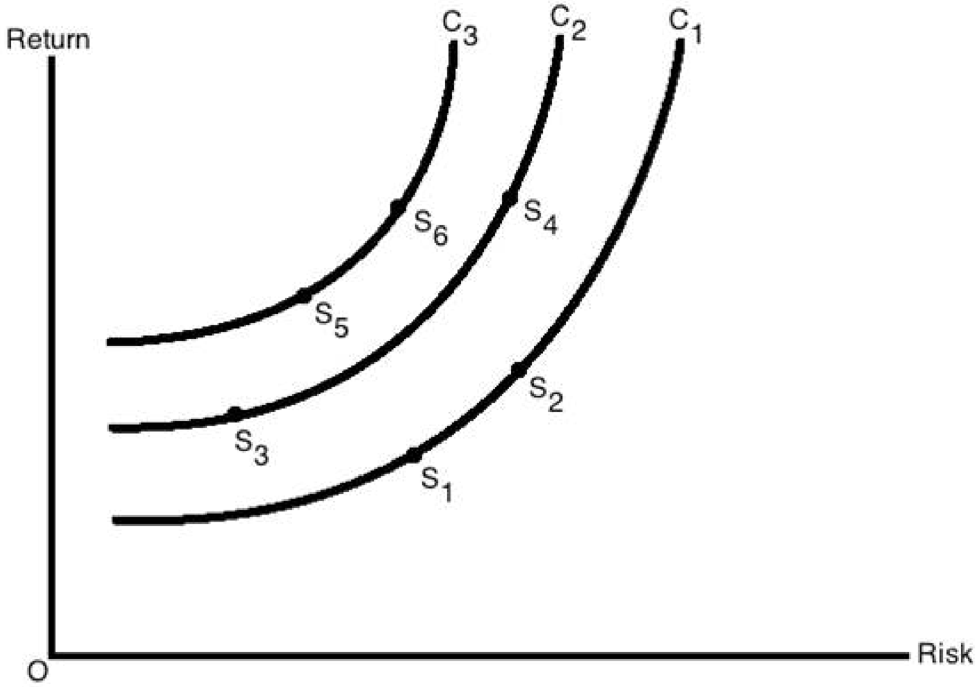

Figure 2: The investor derives different levels of utility on each of the indifference curves (C1, C2, and C3). However, the investor derives the same level of utility on any point say S1 and S2 on the utility curve C1. Additionally, if an investor has to decide between points S4 and S6 an investor would choose point S6 because the investor obtains the same amount of return for a lower level of risk on point S6. Hence, the utility of a risk-averse investor increases as you move leftwards from indifference curve C1 to C3.

As will be discussed in more detail in Section 3, markowitz portfolio theory assumes that all investors have a quadratic utility function; this in turn has implications on the form of the investors' indifference curve. The indifference curve under this assumption also takes a quadratic form of convex shape in the return-risk space. The investors' degree of risk aversion, as proxied by the risk aversion coefficient also has an impact on the form of the indifference curve. The more risk averse an investor, the higher the risk aversion coefficient and hence the steeper the indifference curve.

3 Markowitz Portfolio Theory

Harry Markowitz in 1952 pioneered Markowitz Portfolio Theory (MPT) through an article in the Journal of Finance titled “Portfolio Selection” [135]. The creation of this theory transformed the landscape of portfolio management and 30 years later won Markowitz a Nobel Peace Prize in Economic Science. In this section MPT is detailed in a summarised and mathematical manner highlighting the crucial concepts and portfolios that are necessary for the sections that follow.

MPT is an investment theory that focuses on the selection and construction of optimal portfolios based on emphasizing that risk is an inherent part of higher return. MPT makes use of the Mean Variance Optimization (MVO) framework, MVO provided the first quantitative framework for optimal portfolio selection and construction. Prior to MPT financial literature had treated the interrelationship between an investment’s return and risk in a relaxed manner [66] and this in turn resulted in the selection and construction of optimal portfolios which were highly subjective, with no insight into the return investors could expect.

Additionally, investors did not have a quantitative understanding of how allocating to various low correlated asset classes impacted the overall portfolio’s risk. Markowitz was the first to formally quantify the concepts of risk and expected return in the context of portfolio theory, as well as introduce the modern definition of diversification. MPT asserted the importance of observing risk not in isolation but as the contribution of each asset class risk to the overall risk level of the portfolio. The outcome of MPT is a framework that optimally allocates to various asset classes to either maximize the expected return of a portfolio for a given risk level or minimize the risk of a portfolio for a given expected return. The optimal portfolio selection became grounded in the trade-off between expected return and risk in the portfolio.

By building quantitative models and utilizing historical data, MPT defined concepts such as “expected portfolio return”, and “acceptable levels of portfolio risk”, and showed how to construct an optimal portfolio. The outcome of MPT resulted in the following mathematical conclusions:

1. The expected portfolio return is the weighted average of the expected return of the asset classes.
2. The acceptable level of portfolio risk is a function of the proportion in each asset class, the variance of each asset class return and the covariance between the various asset classes.

Despite the mathematical sophistication of MPT, the overarching principles can be simply explained as discussed by Wilford Sykes in his article [205]. MPT only requires an estimation of three important inputs from an investor: The first, is the expected return of all the assets in the portfolio; the second, is the variance of all the assets in the portfolio; and the third, is the correlation between the various asset classes. Given these three inputs and assuming a normal distribution, investors can easily derive a range of efficient MVO portfolios.

In order to make the MPT framework possible Markowitz set a couple of assumptions regarding the prevailing market conditions and how market participants behave. One of the assumptions is that asset class returns follow a normal distribution. This is an important assumption as it underpins the focus on only the first two moments of the return distribution: the expected return and the standard deviation of returns of the portfolio [102]. If returns are normally distributed then this assumption is appropriate and the return distribution can be fully explained by the two moments, however, if asset class returns are not normally distributed, the predictability of the model is limited. In reality, asset class returns have been seen to not be normally distributed and in the analysis of observed data display higher order moments such as skewness and kurtosis [108].

Another assumption is that a rational investor is risk averse. 'Risk Averse' implies that given two portfolios with the same expected return the investor will pick the portfolio with lower risk. Hence, investors will only allocate to a higher risk asset if they expect the asset will offer higher expected return, and conversely an investor who wants a higher expected return must be willing to accept higher risk. The risk-return trade-off will be the same for all investors, but investors will evaluate the trade-off differently based on their individual risk aversion profile. In terms of utility theory this implies that an individual's utility function is quadratic and displays an Absolute Risk Aversion (ARA). Assuming a utility based on mean and variance of returns and not directly accommodating for an investors logarithmic utility function is cited as being unrealistic [137].

In a book written by Scherer [176] he states that MPT should be used to empirically confirm whether appropriate assumptions are used rather than as a dichotomous theory. Despite the potential flaws of MPT it remains an essential first attempt into the selection and construction of optimal portfolios in a rational and intelligent manner. The following subsections detail all the important aspects of MPT as it relates to the paper. The theory is summarized in a mathematical manner, and all the interesting portfolios are graphically displayed. The Markowitz diversification is explained in Section 3.1, the efficient frontier is explained in Section 3.2, the Two-Fund Separation Theorem with selection of optimal portfolio as tangency portfolio is explained in Section 3.3 and finally some problems with MPT are explained in Section 3.4.

3.1 Diversification

The old adage "don't put all your eggs in one basket" [20] has been around for many years, long before the development of the MPT. However, Markowitz was the first to quantitatively relate diversification to portfolio theory by creating a framework for investors to understand how to approach diversification and the benefits of diversification. Markowitz showed the importance of looking beyond the individual risk and return of an asset class by showing that a portfolio's risk could be reduced by including the benefit of diversification in the portfolio construction process.

Diversification is the reduction of risk in a portfolio by allocating to a variety of asset classes that would each react differently to the same event. This is achieved by combining multiple asset classes that are uncorrelated by assessing the covariance or co-movement of the asset with other asset classes. When these co-movements are considered, a portfolio can be constructed with less risk and the same expected return in comparison to a portfolio that is constructed in the absence of assessing the co-movement of the various asset classes.

Markowitz, in creating his risk and return two-parameter world, defined portfolio risk using the statistical measure variance. The selection of variance as a measure for portfolio risk was not a coincidence as revealed in his interview in [124]. While considering using variance as the risk measure, he needed to first understand how variance was calculated, and in order to understand this he looked to a statistical text titled "Introduction to Mathematical Probability" [194]. The portfolio variance equation revealed that variance of a portfolio was not only dependent on the variance of the individual asset class but also the covariance of each asset class with the other asset classes.

Risk as measured by the variance function is reduced by diversification because the variance measure is a convex function, a convex combination of a portfolio of asset classes will have a lower combined risk than the convex combinations of individual asset class risks. Additionally, it is evident that if asset classes are negatively correlated this would result in a reduction of portfolio risk. Demonstrating this mathematically:

$$\sigma_p = \sqrt{\boldsymbol{\omega}\boldsymbol{\Sigma}\boldsymbol{\omega}} = \sqrt{\sum_i \sum_j \omega_i \omega_j \sigma_i \sigma_j \rho_{ij}} < \sum_i \omega_i \sigma_i \tag{5}$$

when

$$\rho_{ij} < 1$$

Any rational investor who is risk averse will to some degree diversify their portfolio, however, diversification can only help reduce a portfolio's risk as there is no way to completely eliminate all risk. There is always a portion of a portfolio's risk that is due to the exposure of all asset classes to common macroeconomic risk factors that cannot be diversified away even by holding a variety of asset classes that are perfectly uncorrelated. Hence, a portfolio's total risk is made up of two types of risk: systematic risk and unsystematic risk.

The systematic risk is often referred to as market risk and cannot be diversified away since it arises from the common exposure of all asset classes to recessions, interest rates, inflation and other external macroeconomic risk factors. Unsystematic risk is often referred to as the 'diversifiable' risk and is the part of the risk equation that can be reduced or based on some research even eliminated [150]. Importantly, an efficiently diversified portfolio will aim to significantly maximize return while minimizing unsystematic risk. Grinold and Kahn [82] showed that the true benefit of diversification in a portfolio is achieved by combining a fixed number of asset classes that are independent - have low correlation with each other, and is not achieved by including an increasing number of independent asset classes. The derivation of the efficient frontier and some interesting portfolios will follow.

3.2 Efficient Frontier

The efficient frontier is a curve that plots the set of all efficient portfolios in a risk-return framework. In line with MPT, a rational investor is risk averse and will only invest in an efficient portfolio. An efficient portfolio is defined as the portfolio that maximizes the expected return for a given risk level or the portfolio that minimizes risk for a given expected return [135]. In this section we highlight some features of the efficient frontier as well as the mathematical derivation of some interesting portfolios.

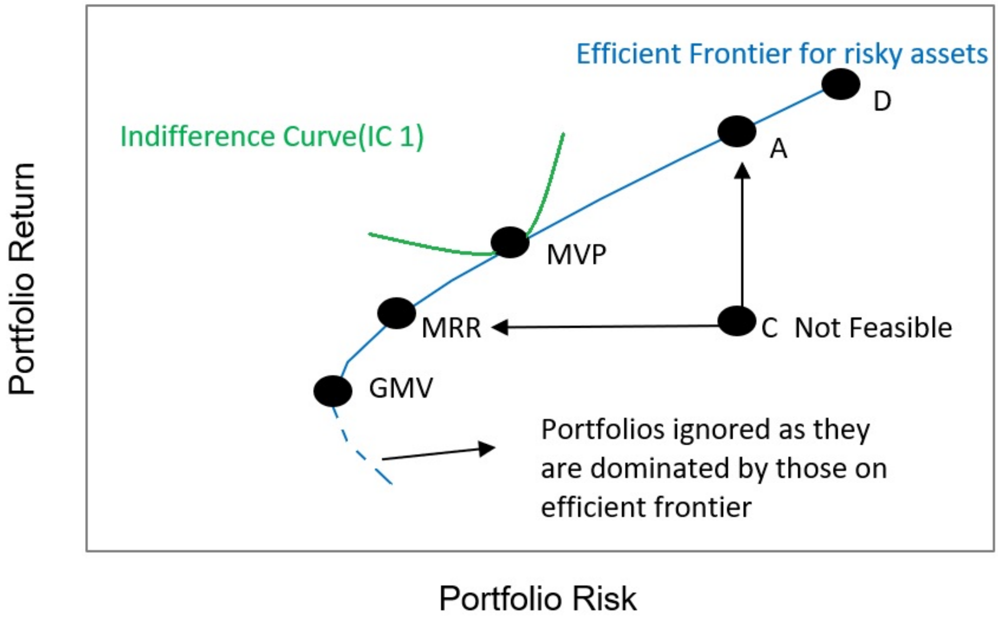

Figure 3: The dots on the efficient frontier represent optimal portfolios for different levels of risk tolerance. The dot labeled GMV represent the Global Minimum Variance portfolio. The dots labeled D, A, MVP, MRR and GMV are all optimal portfolios. The dot labeled MVP is the Mean Variance Portfolio is the optimal portfolio for an investor who has risk tolerance depicted by his highest indifference curve IC1. The dot labeled MRR is the Maximum Risk adjusted Return portfolio. The difference in risk levels between the GMV and any other portfolios represents unsystematic risk

Figure 3 displays the efficient frontier which plots the portfolio return on the y axis and portfolio risk on the x axis. Return is measured by the expected return including any additional earnings, like dividends, over the investment time horizon and risk is measured by standard deviation of returns. Portfolio C falls in an area that consists of all combinations of asset classes that are feasible portfolios but not efficient. A feasible portfolio is a portfolio that an investor can construct given the asset classes available. Portfolio C is referred to as feasible but not efficient because there exists a more efficient portfolio say portfolio A which provides higher return for the same level of risk as portfolio C, or, alternatively there exists a portfolio MRR which provides lower risk for the same return as portfolio C.

The efficient frontier represents the trade-offs between risk and return. This trade-off occurs because moving from left to right on the efficient frontier the optimal portfolio's risk increases but so does the return at the expense of diversification. Portfolio A, MRR and D are mean variance optimal portfolios as they fall on the efficient frontier. In fact all portfolios that lie on the efficient frontier from portfolio D to portfolio GMV on Figure 3 are mean variance optimal portfolios. As all the portfolios on this curve are efficient, the optimal portfolio for an investor is dependent on an investor's preference or utility to the trade-off between risk and return. Portfolio MVP is the optimal mean variance portfolio for a specific investor as it lies on the efficient frontier and is tangent to the investor's highest indifference curve (seen as IC1 in 3). Portfolio GMV is the global minimum variance portfolio which is the portfolio that has the lowest risk on the efficient frontier. Portfolio MRR is the maximum risk-adjusted return portfolio which is the portfolio that lies tangent to a line drawn from the origin and the efficient frontier.

The optimal MVP as explained in [66] is the portfolio that maximizes the value from a utility function as represented by the indifference curve and lies on the efficient frontier. The investor's preferences for risk and return are displayed by the shape of the indifference curve (see Figure 2). Quantifying an investor's utility is difficult to accurately achieve as there is a level of subjectivity needed in deciding what is optimal for an investor. The efficient frontier shown in Figure 3 is only for risky assets and does not include the risk-free asset that will follow in Section 3.3. Below we discuss the mathematical formulation of the mean variance efficient set problem as well as the mathematical derivation of the GMV and MRR portfolios.

The Mean Variance (MV) efficient set problem can be formulated as:

$$\min_{\boldsymbol{\omega}} \left(\frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} \right) \text{ s.t. } \boldsymbol{\omega}^T \boldsymbol{\mu} = \mu_p, \boldsymbol{\omega}^T \mathbf{1} = 1. \tag{6}$$

Here an investor aims to construct a fully-invested portfolio, $\boldsymbol{\omega}$, where risk is minimized subject to matching the investor's return expectation at the end of a single period horizon. The asset returns are assumed to be normally distributed and adequately expressed by the first two moments the expected return $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ which are assumed to be constant through the investment horizon. The first constraint relates to meeting the investors return expectation by forcing the expected portfolio return to be fixed to μ_p. The second constraint ensures the portfolio is fully invested by forcing the total weights of asset classes in the portfolio to sum to 1.

Markowitz showed that the mean variance efficient set of portfolios could be formulated as the portfolio that maximizes the expected return for a given risk level or, equivalently, the portfolio that minimizes risk for a given expected return level [135]. Here we formulate the mean variance efficient set based on the portfolio that minimizes risk for a given expected return level. There is a second formulation based on the portfolio that maximizes the expected return for a given risk level.

The Mean Variance (MV) efficient set problem can also be formulated as:

$$\max_{\boldsymbol{\omega}} \left(\boldsymbol{\omega}^T \boldsymbol{\mu} - \frac{1}{2\gamma} \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} \right) \text{ s.t. } \boldsymbol{\omega}^T \mathbf{1} = 1. \tag{7}$$

The third formulation of the MV efficient set is a special formulation of the risk-return trade-off (quadratic utility) of an investor. Here the investor seeks to maximize the objective function as represented by the utility function. Thus the investor aims to form a fully-invested portfolio to maximize the utility function at the end of a single period subject to the risk aversion parameter γ. A low γ corresponds to risk averse investor, while a high γ corresponds to a risk-taking investor see (Section 2.1.1 for further detail).

The solution for this trade-off quadratic utility optimization problem is [199]:

$$\boldsymbol{\omega}^* = \left(1 - \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma}\right)\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{1^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} + \left(\frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma}\right)\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} \tag{8}$$

This is achieved by optimizing via the Lagrange multiplier method. The proof is in Appendix 13.1. The weight of an optimal portfolio in the efficient set changes based on the risk tolerance or risk aversion parameter of the investor. All three representations of the MV efficient set problem are equivalent [165, 125]. Note that the optimization problem specified above are in terms of returns, rather than absolute levels of wealth.

Suppose an investor is not concerned with the expected return of a portfolio and is concerned with investing in an efficient portfolio that has the least amount of risk. This investor will pick the portfolio from the MV efficient set of portfolios that has the lowest risk, this portfolio is the GMV portfolio and lies at the apex of the efficient frontier.

The Global Minimum Variance (GMV) problem can be formulated as:

$$\min_{\boldsymbol{\omega}} \left(\frac{1}{2}\boldsymbol{\omega}^T \boldsymbol{\Sigma}\boldsymbol{\omega}\right) \text{ s.t. } \boldsymbol{\omega}^T \mathbf{1} = 1. \tag{9}$$

Here an investor aims to construct a fully invested portfolio that minimizes the portfolios risk, similar to the first MV formulation but with no constraint on the investor's return expectation.

The solution for this GMV problem is seen below:

$$w_{GMV} = \frac{\Sigma^{-1}\mathbf{1}}{1^\top \Sigma^{-1}\mathbf{1}} \tag{10}$$

This is again achieved by optimizing the Lagrange multiplier method. The investor will choose the portfolio on the efficient frontier with lowest risk, as seen by GMV point on Figure 3, and, when there is no risk-free asset, unsystematic risk increases by moving along the frontier from the GMV point towards the riskier optimal portfolios.

Finally, suppose an investor is not indifferent to all the portfolios that lie on MV efficient set and is concerned with investing in an efficient portfolio that has the highest expected return for a unit of risk. This is the MRR portfolio and lies tangent to a line drawn from the origin and the efficient frontier.

The Maximum Risk-Adjusted Return (MRR) problem can be formulated as:

$$\max_{\boldsymbol{\omega}} \left(\frac{\boldsymbol{\omega}^T \boldsymbol{\mu}}{\boldsymbol{\omega}^T \boldsymbol{\Sigma}\boldsymbol{\omega}}\right) \text{ s.t. } \boldsymbol{\omega}^T \mathbf{1} = 1. \tag{11}$$

Here an investor aims to construct a fully invested portfolio that provides the highest expected return for a unit of risk. The MRR portfolio is also referred to as the tangent portfolio and maximizes the $SR_0 = \frac{\boldsymbol{\omega}^T \boldsymbol{\mu}}{\boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}}$. The ratio is a version of the reward-to-variability ratio called the Sharpe Ratio in the absence of a risk-free asset.

The solution for this MRR problem is seen below:

$$\boldsymbol{\omega}_{MRR} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} \tag{12}$$

In re-observing the solution for the optimal MV portfolio, we see that equation 8 can be written as $\boldsymbol{\omega}^* = (1-x)\boldsymbol{\omega}_{GMV} + x\boldsymbol{\omega}_{MRR}$ where $x = \frac{\mathbf{1}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma}$. Thus the MV efficient portfolio is a convex combination of two portfolios: the global minimum variance portfolio and the maximum risk adjusted return portfolio. Where each component $\boldsymbol{\omega}_{GMV}$ and $\boldsymbol{\omega}_{MRR}$ has the property that $\boldsymbol{\omega}^T \mathbf{1} = 1$.

In the absence of a risk-free asset, this is known as the two-fund separation theorem which shows an investor can hold any optimal portfolio by holding proportions of any two other optimal portfolios. Instead of determining the allocation across a variety of asset classes, the optimal portfolio decision for any investor can be determined by distributing wealth over two optimal portfolios. The two-fund separation theorem simplifies the focus of the single-period investor to identifying these two optimal funds.

3.3 Tobin's Two-Fund Separation Theorem

The Tobin's Two-Fund Separation Theorem was introduced by James Tobin in 1958 who used Keynesian liquidity preference theory to expand on Markowitz Mean Variance Optimization (MVO) framework and included the risk-free asset. A risk-free asset is one that has no risk, low expected return and is uncorrelated to the risky asset. In his research [192] bonds, which have obligations to pay stated cash amounts at future dates, with no risk of default where used as a proxy for the risk-free asset.

In the presence of a risk-free asset Tobin showed that the efficient frontier becomes a straight line which is referred to as the Capital Market Line (CML). The CML stretches from the risk-free asset and is tangent to the risky asset efficient frontier. An important implication of the theorem is that an investor's optimal portfolio choice is simplified into two steps: first step, identify the portfolio of risky assets that maximizes the Sharpe Ratio; and second step, determine the allocation between the Maximum Sharpe Ratio (MSR) risky portfolio and the risk-free asset. The derivation of the efficient frontier in the presence of risk-free asset is provided in Appendix 13.2.

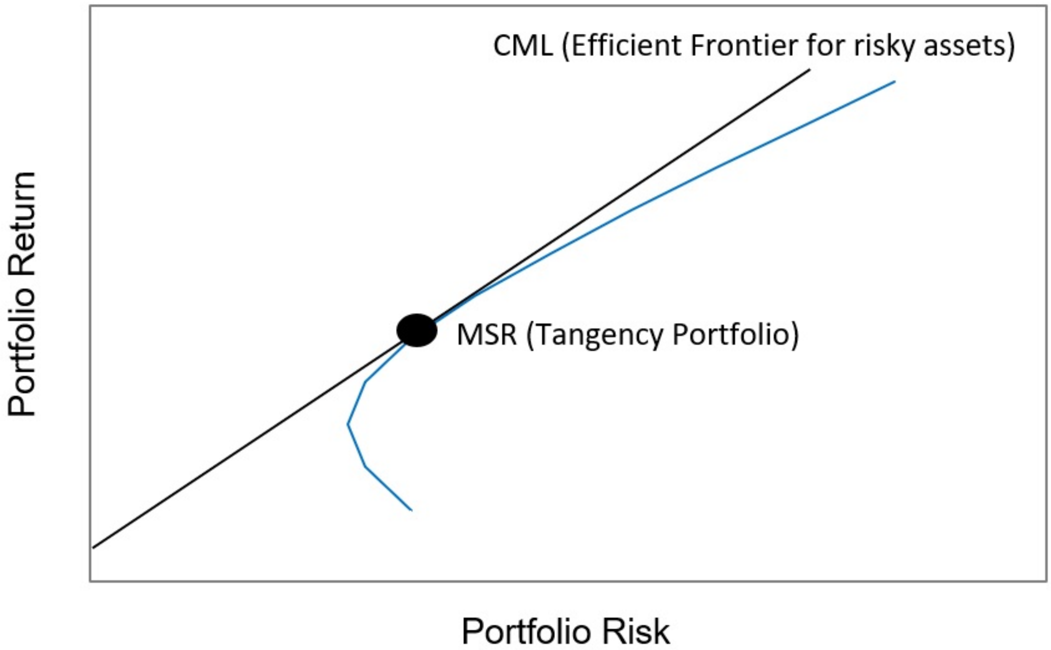

Figure 4: When the risk-free asset is included the efficient frontier becomes a straight line called the Capital Market Line (CML) which is tangent to the risky asset efficient frontier at a single point MSR (Maximum Sharpe Ratio). The MSR dot represents the portfolio which is the optimal mix of risky assets in the mean variance space. All efficient portfolios are a mixture of the MSR tangency portfolio and the risk-free asset.

The theorem also implies that the risk-return profile of any feasible portfolio cannot lie above the CML. The tangency of the CML can be intuitively understood through considering the alternative positions of the CML. If the CML intersected the risky asset efficient frontier, then portfolios would exist on the risky asset efficient frontier that dominates portfolios on the CML, thus contradicting the definition of an efficient frontier. Portfolios that lie above the CML are not possible, thus the CML must touch the risky asset efficient frontier.

An important implication of this result is that if investors "maintain identical expectations regarding the future" ([206] citing [192]) they will hold the same risky MSR portfolio (market portfolio) and this no longer depends on the investors risk tolerance. The theorem states that under a MV framework the optimal portfolio is a linear combination of risky and risk-free asset. The implications of Tobin's two-fund separation theorem for equilibrium prices is derived by [184, 131], and is known as the Capital Asset Pricing Model (CAPM).

Hence investors who live in this two-parameter mean variance world and who only care about risk and expected return, will hold the same risky portfolio. The investors will account for their individual risk tolerance by altering the allocation to the risky asset and risk free asset. All investment decisions are made along the CML line, and the role of risk aversion is confined to the second step while playing no role in the first step of finding the MSR portfolio.

The two-step optimal portfolio allocation process utilizes the same assumptions as the mean-variance framework and is based on the assumption that there is no model uncertainty or model mis-specification this implies that the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are perfectly known [118]. In practice, these parameters are unknown and frequently estimated from historical asset class returns or from thorough analysis of various types of fundamental and macroeconomic data. Hence, in reality, there are errors in the estimation and specification of the two-step optimal portfolio choice due to the sensitivity to estimation error. We will discuss this and other shortcomings in the next section.

3.4 Problems with MPT

In this section, we discuss two of the commonly cited shortcomings of MPT. The first shortcoming we discuss is that MPT does not address estimation risk and delivers efficient portfolios assuming the input parameters are known with certainty. The MVO efficient portfolios have been observed to be highly sensitive to small changes in input parameter values, and hence the model is said to be not robust and "estimation-error maximizers". The second shortcoming is that the application of diversification (the only "free lunch") has come under attack specifically post the global financial crisis.

3.4.1 Sensitivity and Error-Maximization

MPT utilizes a MVO framework to derive optimal portfolios. MVO requires the investor to know the true parameters of the model namely the mean and covariance of the asset classes, however in reality the true parameters of the model are unknown and need to be estimated. Thus, the model's efficacy is partly dependent on the accuracy of these estimates. In practice, these parameters are frequently empirically estimated from historical asset class returns or from thorough analysis of various types of fundamental and macroeconomic data. Hence, there are errors in the estimation and specification of the MVO optimal portfolio.

The optimal MVO portfolio can be highly sensitive to the estimation error. The financial literature written on the sensitivity problem of MVO to estimation error or model uncertainty is expansive; see, [21, 23, 32, 31, 84, 109, 146] to name a few. To touch on some in more detail Best and Grauer [21] state that the main difficulty with MVO when using sample parameters is the extreme positions that exist in the sample-based efficient MVO portfolios. They show how these extreme positions arise due to the sensitive nature of optimal portfolio weights to the estimation of the mean parameter, which implies that sampling error in the mean and covariance parameters in turn have a significant impact on the estimation and selection of optimal portfolio weights.

Similarly, Green and Hollifield [84] identify the lack of diversification in MVO portfolios due to extreme allocations as a serious shortcoming of MPT. Specifically they state that as the number of asset classes increases and sample parameters are used to determine portfolio allocations, these allocations do not approach zero and often involve extreme allocations. The paper also explores the notion of whether mean variance efficient portfolios and "well-diversified" portfolios actually coincide.

The standard MVO framework, ignores this estimation risk and simply treats these estimates as the true parameters in the optimal portfolio formula see equation 7. The optimal portfolio asset allocation can be very sensitive to the estimation error as stated in Kan and Zhou [114]. They show that when optimal portfolios are constructed from sample estimates that are then replaced with another set of estimates that are similar and statistically indistinguishable from the one used initially, then there can be poor performance of the optimal portfolio. Additionally, they show when there is parameter estimation risk the two-fund separation theorem is never optimal and that an investor can benefit by holding a three-fund (perhaps even more) portfolio which consists of a third risky portfolio that helps diversify the estimation error. Specifically, the third portfolio proposed is the sample GMV portfolio - this portfolio reduces the estimation error as it does not need the estimation of the mean parameter.

Extensive research has gone into easing the sensitivity problem of MVO. The financial literature covering this topic has various perspectives on elevating the sensitivity issue from a variety of different angles, including : resampling approach [146]; shrinkage approaches [42, 18, 106, 128, 127, 208]; imposing constraints such as no short-selling [106]; Black-Litterman approach [23]; Bayesian approaches [119, 34, 152, 153, 211]; and worst-case approaches [41, 63, 78, 87, 117, 121, 174, 178]. In the resampling approach Michaud [146] suggests creating multiple alternative efficient frontiers from resampled versions of sample estimates through Monte Carlo methods and then averaging the portfolios from the multiple efficient frontiers to derive the resampled efficient frontier portfolios. This portfolio is different and hence suboptimal in comparison to Markowitz efficient portfolio, but it displays optimal characteristics when averaged over the many possible values of the unknown true mean and covariance. Michaud is also famously quoted for referring to the "MVO portfolio optimization" as "error maximization", attributing the sensitivity issue as the reason that optimization amplifies the effects of estimation error.

The Black Litterman premise [23] uses a Bayesian approach to create a new mixed estimate of expected returns by combining an investors subjective view on expected returns with the market equilibrium return. The investor views are derived by the process of reverse optimization, more specifically the investor specifies their own vector of portfolio weights and then uses reverse optimization to extract their views. The Black Litterman model has been proposed [103] to overcome some of the key problems of MVO such as sensitivity to estimation error, error maximization and extreme portfolio weights. Black-Litterman overcomes these problems by not requiring the investor to estimate the mean but rather specify an investor's assumption regarding how the mean differs from the market's and the degree of confidence in the alternative assumption. The Black-Litterman model has been said [146] to largely mitigate the estimation error-maximization problem of MVO by spreading the estimation errors throughout the vector of expected returns. As a result, the model is argued to create more stable, and robust, mean variance efficient portfolios.

The traditional approach for estimating the covariance matrix involves using the sample covariance matrix, however this approach is said to suffer from the "curse of dimensionality". Significant financial literature has been written to improve the estimation of the covariance matrix. The shrinkage approaches have been proposed as an alternative approach to build robust covariance matrices [42, 18, 106, 128, 127, 208, 136, 202]. These approaches stem from the fundamental principle that in statistical theory there exists a tradeoff between the estimation error and the specification error. Thus, to obtain a better estimator, the large estimation error that exists in the sample covariance matrix must be decreased without increasing the specification error. The combination of their findings propose that the best estimators to manage this tradeoff are the shrinkage estimators and portfolios of estimators.

3.4.2 Failure of Diversification

Diversification is a fundamental concept brought by MPT into the investment decision-making process under risk. Diversification offers the prospect of reducing a portfolio's risk without sacrificing its returns and is often referred to as the only "free lunch" in investing. However, the ability of diversification to protect against significant losses has come under attack in recent times after the Global Financial Crisis (GFC) in 2008. Ironically, it is in a crisis where the benefit of diversification is needed the most, but many of the portfolios that were perceived to be well diversified suffered significant losses as all asset classes moved in the same downward direction.

Fabiozzi et al [67] states MPT is an oversimplification of reality that relies on the properties of diversification in a purely statistical and naive manner. The theory doesn't acknowledge the difficulty in forecasting the essential parameters needed for MVO and does not account for the fact that return distributions display fatter tails in reality. They argue that an analysis of the S&P 500 Index over the past 20 years shows the equity index does not follow a random walk and is actually characterized by the reversal of local trends. To make diversification beneficial, they argue, a different concept of diversification should be considered: diversification of local trends, and hence investors should look at trend reversals and at correlations between local trends.

Defining risk by standard deviation of historical returns has been argued by some as inefficient stating that the notion of risk in MPT needs to be amplified. Risk goes further than standard deviation of historical returns and the existence of anomalies such as low volatility and style premia (momentum, value, size etc.) demonstrate the inefficiencies of this risk measure, so states Mr. Kieselstein who is CIO and Managing Partner of Frankfurt-based quantitative asset manager Quoniam [68]. MPT only considers benign risks, namely, the expected fluctuation in asset values, and is blind to systemic risk and thus oblivious to significant losses, such as those experienced during the GFC that are systemic in nature states ETH Zurich's Professor Sornette [68].

Further, correlation is not a good measure in the context of diversification noted Statman and Scheid [187, 188] who provide two reasons for this. Firstly, correlations are often misunderstood, for example, there is a misperception that a high correlation (e.g 0.8) automatically implies low diversification but it is possible to have significant diversification benefits at high correlation. Secondly, both the correlation between asset class returns and the standard deviation of each asset class' return determine the diversification benefit of a portfolio. Hence, a better diversification measure is return gaps which accounts for both the impact of correlations and standard deviations. Additionally, they stated that when the benefits of diversification were assessed through this return gaps measure, the benefit was actually higher in bear markets than bull markets as the higher standard deviation compensated more towards diversification than the higher correlations detracted from it.

Full sample correlations mislead diversification benefits states Page and Panariello [149]. They argue that the full sample correlation is an average of extremes and misleads investors' perception of diversification. When conditional correlations are observed the asymmetry of correlations is observed, as diversification across risk assets almost completely disappears in market crashes and is more prevalent in market rallies. This undesirable asymmetry is pervasive through many markets and asset classes and is undesirable as it is in market crashes that diversification is needed the most. They propose a robust approach to measure left- and right-tail correlations and argue that to improve diversification benefits investors need to enhance their risk management process by including tools like downside risk and scenario analysis in the optimal portfolio construction and selection step. Additionally, they propose investors look beyond diversification to reduce portfolio risk and also consider overlaying their portfolio with alternative strategies such as hedging and dynamic risk-based strategies that can help reduce the portfolio risk.

Proponents of diversification [5, 104] have argued that diversification was misunderstood and did not fail in GFC. These authors acknowledge that correlations do rise in market crashes and together offer at least three counter arguments to the "diversification is dead" statement. Firstly, Asness, Israelov and Liew [5] states that the critique misses the point- although diversification does fail in short-term systemic panics such as GFC it still provides long-term downside risk protection, arguing long-term returns are more important to wealth creation and destruction than short-term returns and that diversification protects investors from holding concentrated portfolios that can have long-term wealth destruction. Secondly, asset classes do exist such as high quality bonds that provide positive returns in market crashes. Lastly, Ilmanen and Kizer [104] states that utilizing factor building blocks instead of traditional asset class building blocks improve diversification in the long-term as well as particularly in market crashes.

Finally, Koumou [123] reviews all the current research on the failure or misperception of diversification and concludes that investors need to understand the concept of diversification better and choose diversification measures more effectively in relation to portfolio theory. To understand the importance of the choice of diversification measure, they review four core diversification principles: law of large numbers; correlation; capital asset pricing model; and risk parity diversification principles. These four core diversification principles are studied based on understanding their definitions better, reviewing their optimality: in terms of expected utility theory and usefulness for a risk averse investor and, finally, by considering their measurement. To summarize the author proposes the view that diversification is not dead, but it is misunderstood.

4 Asset Pricing Models

Asset pricing theory aims to understand and determine the price of a financial asset in an uncertain world. The cornerstone of contemporary asset pricing is based off the fundamental relationship that the price of an asset equals its discounted future expected cash flows. The key components to determine then become the timing and risk of the future cash flows. The timing of future cash flows are relatively simple to determine, however the risk of future cash flows are more difficult to determine. The risk of future cash flows are also cited to be more important, "for example over the past 50 years U.S. stocks have given a real return of about 9% on average 1% of this can be attributed to interest rates; the remaining 8% is a premium earned for holding risk" [47].

The two different approaches of applying this fundamental pricing relationship are absolute and relative pricing models. Absolute pricing models determine the price of an asset based on its exposure to fundamental sources of macroeconomic risk, while relative asset pricing models determine the price of an asset based on the given price of some other assets. Classic examples of absolute pricing models are the consumption-based and general equilibrium models, while Black-Scholes option pricing model is an example of relative pricing model. Absolute pricing offers generality at the expense of precision while relative pricing offers simplicity and tractability at the expense of limited practical application. Asset pricing models in reality are not used in isolation but rather in combination with other pricing models.

Absolute pricing is commonly used in an academic setting, where it is positively used to explain why prices are the way they are and how prices could change in the future. Whereas relative pricing is used in a less demanding manner in order to learn more about an asset's price given other asset prices, with no answer to how the other asset prices were derived, and what the fundamental risk factors are. There are positive and normative arguments to the application of asset pricing theory however in reality the theory is used to describe both the way the world works and the way the world should work. For example, CAPM and subsequent factor models APT are examples of general equilibrium models, however in application the price of an asset class is determined "relative" to the market or other risk factors assuming these factors are given.

In Section 3 the price of the asset or equivalently expected return is exogenously given, and each investor has to only form beliefs about the moments of the asset class distribution i.e. the mean and covariance. CAPM is considered the first asset pricing theory and takes off from the foundation and assumptions of portfolio theory with a risk-free asset and unlimited short sales. However, in CAPM returns are no longer exogenously given and all investors are assumed to have the same beliefs. Investors have the same belief regarding the probability distribution of all assets, i.e. they agree on the mean and covariance, and these beliefs are aggregated to determine a market equilibrium. There is a linear relationship between the asset's expected return and the covariance or exposure to the market portfolio's return.

4.1 Stochastic Discount Factor

The cornerstone of contemporary asset pricing is the fundamental relationship that the price of an asset equals its discounted future expected cash flows. For example a bond's price is equal to its discounted future coupons and principle; a stock's price is equal to its discounted future dividends and price; an option's price is equal to its discounted future value of another asset. This fundamental relationship can be expressed using the stochastic discount factor approach [173, 185, 88, 47]. Most asset pricing can be summarized mathematically by the following two formulas:

$$p_t = \mathbb{E}[m_{t+1}x_{t+1}]$$

$$m_{t+1} = f(\text{data}, \text{parameters}),$$

where p_t is the price of the asset today (t), $\mathbb{E}$ is the expectation operator, x_{t+1} is the assets future cash flows, f is some function and m_{t+1} is the stochastic discount factor (SDF). This pricing equation shows that the price of an asset today is a function of its expected future cash flows discounted by its SDFs and was introduced by [92, 91, 89]. The SDFs is associated with state space geometry instead of the mean variance geometry [47]. In this section we specify that future uncertainty as mutually exclusive discrete states, $S = \{s_1, s_2, ..., s_n\}$. Hence, the SDFs is a random variable that computes the price today of an asset class by discounting the state-by-state future asset cash flows.

The benefit of the SDFs approach is its simplicity and broad application. Previously each asset class for example bonds, equity, options had its own theory for pricing but using the SDFs approach there is one pricing theory for all different asset class types. All asset pricing models result in alternative specifications of the SDFs which is also referred to as the pricing kernel. The main task in asset pricing theory becomes identifying which SDFs should be used for a specific model and how to estimate the SDFs in that particular model setting. Thus leaving us with an understanding how the SDF is determined in the market to produce prices.

4.2 Arbitrage Pricing Theory

CAPM has been criticized for being too simplistic and restrictive in its determination of asset class returns [35]. Empirical studies have shown that CAPM poorly explains asset class returns and other factors in addition to the market factor drive asset class returns [69, 71, 14, 15, 13, 132, 115, 116]. It is this empirical failure that has resulted in the development of other asset pricing models and the realization that the market is more complex and intertwined than previously acknowledged with most theories being only reasonable local approximations of reality.

Arbitrage Pricing Theory (APT) was primarily introduced by Ross [169, 170]. It is a single period model in which every investor believes there is a linear relationship between the asset's expected return and the loadings to multiple factors. The factor loadings, or betas, are proportional to the asset's return covariance. The covariance represents the systematic risk that investors cannot diversify away. The intuition behind APT is that the unsystematic risk of an asset should not carry any risk premium as it can be diversified away, thus the expected return of an asset should only be related to the asset's covariance with common systematic components or factors.

APT is seen as a generalization of the CAPM model which aggregates all risk factors into a single factor the market factor. The APT model has less restrictive assumptions than the CAPM, it does not assume that asset returns are normally distributed and does not make any assumptions regarding the investors' utility function besides that investors are risk averse. The simplicity of the model increases its ability to be empirically validated. APT does aim to explain asset returns through common systematic factors but is not limited to the clearly specified single market factor of CAPM. APT is a more generalized factor theory that uses a broad set of k factors but the issue remains identifying the value and nature of the k factors. We discuss this issue further in Section 4.4 which details the theory and intuition behind the APT model.

A central assumption in the derivation of APT is that in a well-functioning security market equilibrium prices offer no-arbitrage opportunities. An arbitrage opportunity is a trading strategy that promises a positive future cashflow in some state contingency with no initial net investment [171]. Mathematically, an arbitrage opportunity is a portfolio, $\boldsymbol{\psi}$, such that:

$$\boldsymbol{p}\boldsymbol{\psi} \leq 0 \quad \text{and} \quad \boldsymbol{X}^\top \boldsymbol{\psi} > 0 \tag{13}$$

where $\boldsymbol{p}$ is a (row) vector of asset prices, $\boldsymbol{\psi}$ is the portfolio depicted by the shares held in each asset, $\boldsymbol{X}$ is a matrix of state-dependent cash flows for each asset. The assumption of no-arbitrage results in the exclusion of arbitrage portfolios from the set of possible portfolios. Mathematically, the No-Arbitrage (NA) principle can be displayed mathematically as follows [172]:

$$NA \equiv \{\boldsymbol{\psi} \mid \boldsymbol{p}^\top \boldsymbol{\psi} \leq 0 \texttt{ and } \boldsymbol{X}^\top \boldsymbol{\psi} > 0\} = \varnothing \tag{14}$$

In a well-functioning security market an arbitrage opportunity would not exist because of the Law of One Price (LOP). The LOP states that if two assets have the same cash flows then they should be quoted and trade at the same price in the security market. There is an intertwined relationship between LOP and no-arbitrage where it is the action of arbitrage that corrects any initial mispricing and drives an asset price to reconcile with the LOP.

Dybvig and Ross [61] showed that the absence of arbitrage opportunities brings about a relationship between asset prices and asset cash flows. The formal application of the no-arbitrage principle to asset pricing is founded on the Fundamental Theorem of Asset Pricing (FTAP). The FTAP states that the following three conditions are equivalent:

1. No-arbitrage.
2. The existence of a positive linear pricing rule that prices all assets.
3. The existence of a finite optimal demand for some agent who prefers more to less.

The proof can be found in [171].

The equivalence of the first two conditions states that the no-arbitrage assumption is equivalent to the existence of a unique linear pricing rule for all assets. The linear pricing rule, $\boldsymbol{q}$, is a state-price vector that creates a linear mapping from an assets cash flows to the assets price, such that, it prices the asset when applied to the asset's cash flows:

$$\boldsymbol{p} = \boldsymbol{q}^\top \mathbf{X}$$

The laws and fundamental concepts of asset pricing provide some general conclusions for the SDF. In particular, Chen and Knez [43] shows the LOP is equivalent to the existence of (at least) one SDFs, Ross [172] demonstrates through rewriting the state-price vector as $q_i = \pi_i m_i$ where $m_i = \frac{q_i}{\pi_i}$ that the no-arbitrage condition is equivalent to the existence of a positive SDF, and finally a complete market is equivalent to the existence of a unique SDF.

It is important to note that the no-arbitrage principle limits relative asset mispricing but the FTAP does permit absolute asset mispricing. Thus, asset prices can differ from their fundamental valuation, where actual market prices can provide arbitrage opportunities. However, these arbitrage opportunities disappear quickly as they are identified and exploited by market participant. Regardless of this, one can identify arbitrage opportunities by comparing actual market prices to the prices derived from the pre-specified SDF.

4.2.1 Derivation of APT

All asset pricing models involve discounting future asset cash flows. The central role of the SDF is complicated by the existence of many potentially infinite number of states. Prior to the development of APT, asset pricing theory involved numerous assumptions and required insights regarding the mysterious utility of an individual. APT requires fewer assumptions, is straightforward and offers a practical approach to determining an asset price in different states.

Ross [169] shows that asset prices are driven by its exposure to few relevant state variables or factors and it is only the risk due to the exposure to these factors that is priced or rewarded. Furthermore, the exposure of an asset to non-factor risk is not priced or rewarded and hence can be diversified away. Formally, he showed that any asset's expected return can be expressed as a linear combination of factors:

$$\mathbb{E}[\boldsymbol{r}] = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{f} + \boldsymbol{\varepsilon} \tag{15}$$

where

$$\boldsymbol{r} = \begin{bmatrix} r_1 \\ \vdots \\ r_K \end{bmatrix}, \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_K \end{bmatrix}, \boldsymbol{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_K \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_K \end{bmatrix}$$

and

$$\boldsymbol{B} = \begin{bmatrix} \beta_{11} \cdots \beta_{1K} \\ \vdots \ddots \vdots \\ \beta_{K1} \cdots \beta_{KK} \end{bmatrix}.$$

Here $\mathbb{E}[\boldsymbol{r}]$ is the expected returns of the asset, $\boldsymbol{\alpha}$ is the excess return of the asset, $\boldsymbol{f}$ represents the factor returns , $\boldsymbol{B}$ is the beta or loading to the factor and the last term $\boldsymbol{\varepsilon}$ explained in Roll and Ross [166] is the idiosyncratic risk of the asset. In the framework of factor analysis, the $\boldsymbol{B}$ is the factor loading and is a K x K matrix here. When multiple K factors exist, Ross [169] shows if no-arbitrage opportunities exist then the following is true and APT can be expressed as:

$$\mathbb{E}[\boldsymbol{r}] = r_f + \sum_{k=1}^{K} \beta_k (\mathbb{E}[f_k] - r_f). \tag{16}$$

Where $(\mathbb{E}[f_k] - r_f)$ is the excess return of the k-th factor over the risk free rate, r_f. Because, the idiosyncratic risk of a portfolio reduces as diversification increases, the idiosyncratic or unsystematic risk factor is not priced by investors. Hence, this relationship states that only factors are priced and any asset's return can be replicated by a combination of factors. Roll and Ross [167] show that between three and four factors are appropriate to account for the systematic risk of an asset. It can also be seen that the factor model for asset returns is equivalent to the factor model of the asset's SDF. Thus factors explain both the covariation between assets and cross-section of expected returns.

4.3 Extending APT

The driving force behind APT is the idea that asset prices are driven at any single-period by its exposure to a few relevant factors and that it is only the risk due to the asset's exposure to these factors that is priced. Furthermore, the exposure of an asset to the idiosyncratic non-factor risk is not priced and hence can be diversified away. The original APT theory states that if there is no-arbitrage there is a linear constant relationship between the asset returns and the multiple factors. Thus, the original APT model proposed by Ross was created as a constant single-period linear factor model.

Due to the limitations of the original APT model, two classes of models, the non-linear APT and conditional APT have aroused the attention of researchers. The non-linear APT allows for a non-linear relationship between returns and factors while conditional APT allows for a non-constant or static time-varying factor models of returns. These extensions maintain the favorable features of the original APT model while being consistent with the no-arbitrage principle and are essential for the practical application of the APT in financial markets. In the following section we discuss how APT can be extended to include non-linear cash flows and multi-period horizons.

4.3.1 The Non-Linear APT

A factor model is assessed empirically by comparing the observed returns against the estimated returns using the pre-specified factor model. The efficiency of a factor model is assessed based on how well the pre-specified factors explain the variation of the realized asset returns. The alpha ($\boldsymbol{\alpha}$) in equation 15 is the portion of an asset's return that is not explained by its exposure to the pre-specified factors. If a factor model is well specified and covers all the asset's key drivers of systematic risk factors then the $\boldsymbol{\alpha}$ should be zero.

An empirical statistically significant alpha can occur for a couple of reasons. It can be a signal of a poorly specified factor model as evidenced by a large portion of idiosyncratic risk that cannot be diversified away, here additional factors need to be considered to overcome this issue. Alternatively, in a well specified factor model it can represent a statistical arbitrage opportunity where an investor can derive a positive future cash flows with no initial net investment. Another potential reason for the existence of a significant α is that it could exist due to applying a linear pricing kernel to price a non-linear relationship [12].

Not all assets have linear cash flows, for instance assets such as derivatives have non-linear cash flows. It has been shown by Dreger and Schumacher[60] that applying the linear single factor CAPM model to price derivatives will result in a violation of the no-arbitrage principle. This inflexibility and potential for prediction errors have driven the consequent development of the non-linear APT framework. Bansel et el [12] believe that the linearity assumption of the initial APT model is unnecessarily constraining and display conditions where a flexible pricing kernel that is a non-linear exists and is a function of few factors.

This non-linear APT framework is seen to be consistent with no-arbitrage condition and utilizes semi-parametric techniques where the non-linear pricing kernel is estimated by applying neural networks. In the empirical application Bansel et al [12] compare the performance of the non-linear CAPM model to a linear unconditional CAPM model deducing that the non-linear CAPM model explains the variation of stock returns better than the linear model. The extension of the original APT into non-linear APT models infers the existence of a non-negative non-linear pricing kernel.

Daniel and Titman [52] find that the alpha is non-zero for both a non-time-varying characteristic based model and a Fama-and-French APT model. A similar conclusion is obtained by Gebbie and Wilcox [79] in South Africa. Gebbie and Wilcox [203] believe that the non-zero alpha implies that a non-linear pricing kernel exists and the alpha must be a non-linear function of risk factors. They use a predictive form similar to Ferson and Harvey [73] with time varying versions of characteristic based and Fama-and-French APT model and argue that the pricing kernel is non-linear due to time-variation [203].

4.3.2 The Conditional APT

Conditional asset pricing models have gained prominence due to an abundance of evidence stating the existence of time-varying betas and time-varying risk premia. Additionally, research such as [62, 90] show that the conditional CAPM can hold perfectly even when the unconditional CAPM has grave pricing errors. A wide range of research has investigated conditional linear factor models [81, 74, 27, 94, 182, 38]. This extension of the APT provides allowance to implement a multi-period pricing model. In the original APT the beta or factor loadings remain constant and is applicable for a single period horizon.

The original APT uses an unconditional stochastic discount factor to assess the riskiness of future cashflows. Numerous research pieces have highlighted this as an unrealistic assumption [139, 93, 72, 107, 129]. The basic argument being that future cashflows will vary because of the impact of varying business cycles. The conditional APT calculates the riskiness of future cash flows by accounting for a time varying stochastic discount factor. The conditional APT model applies a similar logic to the single period case where at any relevant time period say $T = 1, \dots, N$ a stochastic discount factor is specified that maps the future cashflows to a no-arbitrage price at the previous time period.

So, for example a conditional stochastic discount factor exists that maps the assets cash flows at N to no-arbitrage price at N-1, then at N-1 another conditional discount factor maps the assets cash flows at N-1 to a no-arbitrage price at N-2. Continuing in this manner at each future horizon means a conditional discount factor can be specified such that at any single horizon sub-period a no-arbitrage price exists. Since at each single horizon sub-period there is a no-arbitrage opportunity this also implies at the overall multi-period horizon no-arbitrage opportunities exist. Hence, a conditional

APT model can be obtained by setting the conditional pricing kernel as a linear combination of state-dependent factors.

In general conditional models that allow for time-varying betas and risk premia in a single factor or multiple factor context have been researched to perform significantly better than their unconditional counterparts. However, there is also evidence called the Lewellen-Nagel critique [130] that shows that the conditional CAPM's explanatory power is not significantly better than unconditional CAPM to warrant all the effort that is required in its setup. There is also a debate regarding which between a conditional linear or unconditional non-linear factor model, is empirically more successful, with [197, 80] concluding that the non-linear factor model performs better than the conditional factor model while; [113] counters these findings.

4.4 Specification of Factors

APT states that the asset's return is only related to its exposure to systematic factor risks that cannot be diversified away. APT does not specify the nature or number of factors. To apply APT empirically, these factors need to be specified and it involves a prior phase where the factors are identified based on their significance and where respective risk premiums are estimated. In practice multi-factor models are grouped into three categories: statistical, macroeconomic and fundamental. Numerous studies have taken place in each of these categories to specify the factor choices and model setups but there is no standardized set of factors or model setup.

The three categories of factor models can either utilize an explicit or implicit factor specification approach. An explicit or exogenous approach pre-specifies the factors by choosing the most significant factors in advance. The benefit of this approach is that factors can be easily interpreted but the negative is that there could be dislocations in the explanatory ability of the factors with the evolution of returns through time. An implicit approach or endogenous approach derives the factors directly from the historical asset returns. The benefit of this approach is that factors are derived directly from the returns and evolve with returns through time; however, the negative is that these factors cannot be easily interpreted.

The three categories of factor models are not necessarily inconsistent. Conner [49] states that the three categories are simply restatements or rotations of one another when there is no estimation error and large volumes or abundance of data. Therefore they are not in conflict and can hold true simultaneously. He also finds that the explanatory power of statistical and fundamental factor models is superior to macroeconomic factor models but acknowledges that explanatory power is only one criterion in measuring the relative worth of the three types of factor models. Below we briefly discuss each of the three categories of factor models in more detail.

4.4.1 Statistical Factor Models

Statistical factor models use historical and cross-sectional data to identify the factors and betas. The statistical factors are derived from the covariance of asset returns and not dependent on any external data sources. Hence, statistical factor models use an endogenous approach to specify factors. In the initial tests of APT, statistical factor models where first used to identify the number of explanatory factors. The statistical factor models use Principal Component Analysis (PCA) and Maximum-Likelihood

(ML) statistical methods to derive the persistent factors from returns. These methods can quickly summarize large volumes of data but the disadvantage of this approach is interpreting what risk the factors actually represent and the economic intuition behind the factors.

4.4.2 Macroeconomic Factor Models

Macroeconomic factor models use historical macroeconomic data to specify the factors and estimate the betas. The factors are based on economic time series data and generally covers the core macroeconomic variables that impact asset class returns. Macroeconomic factor models use an exogenous approach to specify factors. Common macroeconomic variables used are gross domestic product, inflation, interest rates and money supply amongst others. Macroeconomic factor models tend to be the simplest and easiest to interpret of the factor models. However, a disadvantage is that the main macroeconomic factors that drive the pervasive shocks need to be identified and pre-specified, therefore there can be periods of time when other risk factors that have not been included drive returns.

4.4.3 Fundamental Factor Models

Fundamental factor models use observable company characteristics to specify the factors and estimate the betas. These fundamental factors are pre-specified hence this model uses an exogenous approach to specify factors. The most famous fundamental factor model was introduced by Fama and French [70] which originally consisted of three factors namely High-Minus-Low (HML), Small-Minus-Big (SMB) and market. Another popular model is MSCI Barra risk model which was pioneered by Bar Rosenberg and consists of fundamental factors such as price to earnings, dividend yield, earnings growth, as well as quantitative factors such as momentum and volatility. A similar analysis was performed in South Africa by Wilcox and Gebbie [204]. These models grew in popularity due to the finding of empirical evidence that these company characteristics explained a significant portion of asset returns. Fundamental factor models are also referred to as characteristic based factor models.

5 Law of Large Numbers

In this Section we review the Law of Large Numbers (LLN) theorem and show how the theorem enables the estimation of the classic historical distribution. We then demonstrate how this distribution can be generalized into the flexible probabilities paradigm which is required to build the HS-FP framework discussed in Section 6. Finally, we discuss the relationship between the LLN and Ergodic theorem and the applicability of ergodic systems in financial literature.

The law of large numbers is one of the most important, well established and powerful theorems in statistics. It was first introduced in the 16th century by Cardano and later proved in the 18th century by Bernoulli [181]. The theorem states that the sample mean of a large number of independent identically distributed (i.i.d) random variables moves closer to the expected value of the random variable as the number of random variables increases.

Formally, the LLN states for an i.i.d distributed random variables $X_1, X_2, ..., X_n$ the sample mean, denoted by $\bar{X}$, is defined as:

$$\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n} \tag{17}$$

In terms of probabilities, the theorem states that the relative frequencies of possible outcomes in a probabilistic process will move closer to their respective probabilities, as the process is repeated, a large number of times. It is important to note that the theorem has been applied in many different fields such as economics, insurance and finance for risk management purposes [1, 4, 111, 175, 193]. In this paper we focus on the use of LLN theorem in financial literature.

5.1 Derivation of HS-FP Framework using LLN

The LLN theorem is important in the derivation of the HS-FP framework which we discuss in Section 6. The objective of HS-FP framework is to derive the distribution of the invariants $\boldsymbol{\epsilon}_t$ as denoted by the cumulative density function (cdf) $F_{\boldsymbol{\epsilon}}$ so that inferences can be made about its statistical features.

However, in reality we do not know the true distribution $F_{\boldsymbol{\epsilon}}$ and hence it needs to be estimated. Using a specific formulation of the LLN, the Glivenko-Cantelli theorem, it can be seen that the historical distribution of the time series of invariants $\boldsymbol{\epsilon}_t$ becomes more similar to the true distribution of invariants $F_{\boldsymbol{\epsilon}}$ as the number of observations increases.

5.1.1 Classic Historical Estimation

Given a set of $\bar{t}$ historical time series of the invariants $\{\epsilon_1, \ldots, \epsilon_{\bar{t}}\}$ we can then define the historical probability density function (pdf) as follows:

$$\hat{f}_{\boldsymbol{\epsilon}}^{\,Hist}(\boldsymbol{x}) \equiv \frac{1}{\bar{t}} \sum_{t=1}^{\bar{t}} \delta^{(\boldsymbol{\epsilon}_t)}(\boldsymbol{x}), \tag{18}$$

where $\delta^{(r)}$ denotes Dirac delta centered in r. Here the Dirac delta $\delta^{(r)}(\boldsymbol{x})$ is a generalized function or distribution that "peaks" at r when $r = x$, is zero wherever r $\neq$ x and whose integral is equal to one. The historical *pdf* displays a set of "peaks", each centered in the respective observation of $\boldsymbol{\epsilon}_t$, where each observation carries a uniform weight equal to $p_t \equiv \frac{1}{t}$.

Equivalently, we can build the integral of the historical pdf or the historical cdf:

$$\hat{F}_{\boldsymbol{\epsilon}}{}^{\text{Hist}}(\boldsymbol{x}) \equiv \frac{1}{\bar{t}} \sum_{t=1}^{\bar{t}} \mathbf{I}_{\boldsymbol{\epsilon}_t \leq \boldsymbol{x}}. \tag{19}$$

A variable that follows the historical distribution represented by equations 18 or 19 can only output a discrete set of outcomes specifically the historical observed values $\boldsymbol{\epsilon}_t$. For a random variable $\boldsymbol{X}_t$, at any general future time $t > \bar{t}$ there is no relationship between the historical observations $\{x_1, \ldots, x_{\bar{t}}\}$ and the true unknown distribution $F_{\boldsymbol{X}_t}$.

However, using a specific type of the LLN namely the Gilvenko-Cantelli theorem, if $\boldsymbol{X}_t$ is an independent and identically distributed random variable across time as is the case with the invariants $\boldsymbol{\epsilon}_t$, then the historical distribution $\hat{F}_{\boldsymbol{\epsilon}}{}^{Hist}$ approximates the true unknown distribution $F_{\boldsymbol{\epsilon}}$ of the invariants $\boldsymbol{\epsilon}_t$. This approximation improves as the number of observations increases and approaches infinity as seen below:

$$\lim_{\bar{t} \to \infty} \hat{F}_{\boldsymbol{\epsilon}}{}^{Hist} = F_{\boldsymbol{\epsilon}} \tag{20}$$

or

$$\sup_{\boldsymbol{x}} \mid \hat{F}_{\boldsymbol{\epsilon}}{}^{Hist}(\boldsymbol{x}) - F_{\boldsymbol{\epsilon}}(\boldsymbol{x}) \mid \to 0. \tag{21}$$

The proof can be seen in Appendix 13.3. This is referred to as classic historical estimation due to the probabilities having a uniform weight of $p_t \equiv \frac{1}{t}$. In the next section we discuss how this can be extended to flexible probabilities that are not uniform weighted.

5.1.2 Generalization to Flexible Probabilities

In equation 18 each observation of invariant has an equal probability $p_t \equiv \frac{1}{t}$. An important feature that needs to be possessed by the invariants in order to estimate its unknown distribution is that it must be i.i.d. However, in reality it is difficult to confirm whether the realizations of invariants are perfectly i.i.d. Thus, it is prudent to apply non-equal probabilities weight $\boldsymbol{p}_t$ to the invariants $\boldsymbol{\epsilon}_t$.

This unequal probability weight that is applied to invariants could be conditioned by time - based on the time that an event occurred or conditioned by state - based on the state of the market at the point in time. For example at time t these probabilities could be specified based on giving higher weighting in the probabilities to the most recent observations at t or where state variables as proxied by macroeconomic variables are most similar to market conditions at t.

If we replace the equal probabilities with the more general flexible probabilities $\{\boldsymbol{p}_t\}_{t=1}^{\bar{t}}$ we obtain the historical flexible probabilities distribution as follows:

$$\boldsymbol{\epsilon} \sim \left\{ \boldsymbol{\epsilon}_t \equiv \begin{pmatrix} \epsilon_{1,t} \\ \cdot \\ \epsilon_{n,t} \\ \cdot \\ \epsilon_{\bar{n},t} \end{pmatrix}, \boldsymbol{p}_t \right\}_{t=1}^{\bar{t}} \tag{22}$$

Hence, equation 18 with flexible probabilities can be re-written as the historical with flexible probabilities (HS-FP) pdf:

$$\hat{f}_{\boldsymbol{\epsilon}}^{\,\textit{HS-FP}}(\boldsymbol{x}) \equiv \sum_{t=1}^{\bar{t}} \boldsymbol{p}_t \delta^{(\boldsymbol{\epsilon}_t)}(\boldsymbol{x}). \tag{23}$$

Equivalently, equation 19 with flexible probabilities can be re-written as the historical with flexible probabilities (HS-FP) cdf:

$$\hat{F}_{\boldsymbol{\epsilon}}^{\,\mathrm{HS-FP}}(\boldsymbol{x}) \equiv \sum_{t=1}^{\bar{t}} \boldsymbol{p}_t \mathbf{I}_{\boldsymbol{\epsilon}_t \leq \boldsymbol{x}}. \tag{24}$$

Finally, if the effective number of scenarios (see Section 6.2.2 for more details regarding ens) covered by $\{\boldsymbol{p}_t\}_{t=1}^{\bar{t}}$ approaches its maximum, then the flexible probabilities will approach uniformity, $ens(\boldsymbol{p}) \approx \bar{t}$, and we can again extend the Glivenko-Cantelli theorem to obtain the below:

$$\lim_{ens(\boldsymbol{p}) \to \bar{t} \to \infty} \hat{F}_{\boldsymbol{\epsilon}}^{\,\textit{HS-FP}} = F_{\boldsymbol{\epsilon}}. \tag{25}$$

This conclusion is dependent on the condition that the flexible probabilities are set independently of the invariants.

5.2 LLN and Ergodic Theorem

The ergodic theorem simplistically states that an ergodic system is one where the ensemble average (state domain average) and the time average (time domain average), are the same. In essence, the ergodic theorem assumes that the future is basically a statistical shadow of the past [55]. There are two versions of the LLN: the weak LLN and the strong LLN which have subtle differences. The weak LLN states that the mean of a large sample has a high probability of converging to the expected value, while the strong LLN states that the mean of a large sample will almost surely converge to the expected value. The weak LLN is associated with weak convergence while the strong LLN is associated with strong convergence. The strong LLN theorem can be seen as a special case of the point-wise ergodic theorem.

It is the relationship between LLN and ergodic theorem that equates the optimal choice derived from Markowitz MVO to the Kelly criterion [189]. The random variable in this context is an investor's utility from wealth. MVO optimizes the expected utility of wealth over the state domain of all possible outcomes over a single period horizon and their probabilities. Kelly criterion optimizes the expected utility of wealth over the trajectory of the investor's wealth as it evolves over the time domain. According to the ergodic theorem, when a long enough time has passed and all possible outcomes have been observed, the optimal choice from the world of economics (MVO) will coincide with the world of signal processing (Kelly criterion).

The ergodic theorem applied to finance or economics assumes that the financial market or economy follows an ergodic stochastic process where the future can be predicted from the past. The adoption of ergodic theorem in economics was largely due to Samuelson who argued that "the "ergodic hypothesis [axiom]" is a necessary foundation if economics is a hard science" as stated by Davidson in [56]. There has been significant debate regarding the appropriateness and the application of ergodic hypothesis in finance and economics [155]: "Whether economic reality is an ergodic process after suitable transformation is a deep issue" [98]; there have also been comparisons performed using financial market data regarding whether the random walk model or ergodic theorem based model is more appropriate in financial markets, the finding here is that the random walk model is more appropriate [26]. MVO has also been criticized for its naive application of ensemble averages to calculate "expected" returns in the real world [157]. The ergodicity problem in economics is further discussed in [156].

6 HS-FP Model Setup

Portfolio and risk management decisions are anchored on collecting and distilling the relevant financial market data into useful aid that assists with investment decisions. The relevant financial market data forms the basis for the estimation of important portfolio characteristics such as the return and risk estimates which are essential in the generation of optimal portfolios. Hence, it is crucial to have a flexible, systematic and robust estimation process which incorporates all relevant information regarding the market into the return and risk prospects of asset classes. The standard historical approach creates estimations by using the predictive ability of the empirical distribution of asset class returns. Simulation is another approach of creating estimations by specifying assumptions upfront regarding the asset class and generating multiple future paths for the asset class distribution.

Some of the benefits of using the standard historical approach is that it is a non-parametric approach that is simple to implement and represents the actual asset class history so that assumptions on asset class distribution, dependence structure and other features need not be explicitly made. Hence, the standard historical approach is popular in practice as discussed in [19, 138, 154]. However, the disadvantage of using the standard historical approach is that it presupposes that the past or historical experience will some how match the future experience, this may not be the case. It is also commonly criticized for its lack of "conditionality" which results in a delayed response to underlying market risk changes [158, 19, 37]. Here, we follow a data-informed approach that uses historical data to inform investment decisions.

The Historical Simulation with Flexible Probabilities (HS-FP) model introduced by Meucci [142, 144, 145] leverages all the benefits of the standard historical data approach and provides the added benefit of "conditionality". This conditionality is achieved by applying non-equal probabilities to historical data that is conditioned by: 1) the time it occurred; 2)the state of the market when it occurred; or 3) a combination of both. The state of the market is measured by observable state variables such as financial market ratios, volatility measures, or macroeconomic growth measures. The HS-FP does not necessitate the application of equally weighted probabilities to historical data like the standard historical data approach. The different approaches to deriving the flexible probability weights will be covered in Section 6.1.

The HS-FP model is a simple non-parametric approach that unlike simulation methods does not require an extensive consideration of the assumptions used and possible paths that can occur, hence return and risk forecasts can be quickly estimated. Additionally, as the probabilities assigned to historical data is consistent across the asset class returns it can be easily applied in a robust manner to high-dimensional problems. The premise behind Meucci's HS-FP model is that market participants typically would like the flexibility of considering historical scenarios that are more recent and include additional information regarding the market. In Section 8 we implement a specific scenario which is that the market participants want to stress the most recent information from both a time and state of the market perspective. Naturally, the framework can be extended for a broader range of scenario testing applications than the one chosen in Section 8.

The HS-FP framework allows the estimation of time-varying risk and return statistics for a variety of portfolio and risk management applications. This section explains the HS-FP model setup. We detail the HS-FP framework developed by Meucci [142, 144, 145] and specifically how this framework can be used in the estimation step of the investment process. In the following subsections we outline the involved techniques and how these techniques assist in forming the HS-FP framework. We start by explaining how to generate flexible probabilities that are conditioned by time and/or a single state variable (as seen in Section 6.1), and then explain how to generate flexible probabilities that are conditioned by time and/or multiple state variables in (as seen in Section 6.2).

6.1 Flexible Probabilities

In Meucci's "the Prayer" [143] he outlines ten sequential steps that are essential for investment professionals whether portfolio managers, traders or risk managers across all asset classes to consider in their investment decision making process. We discuss the first two steps of "the Prayer" in this section in more detail. The first step, and the core of risk modeling, is the "quest for invariance". This step involves deriving from the relevant market data the appropriate invariant. This first step can be broken down into two sub-steps, firstly identifying the series of risk drivers and secondly deriving the series of invariants from the risk drivers.

The risk drivers for any specific asset class are the set of random variables $\boldsymbol{Y}_t$ that possess the following two properties: 1) the risk driver entirely describes the asset's price at any given time t, and 2) the risk driver follows a stochastic process that is homogenous across time. It is important to note that the risk drivers do not necessarily have to be the asset's price, and are not *i.i.d.* The invariants $\boldsymbol{\epsilon}_{t \leftarrow t+1}$ are then the shocks that guide the stochastic process of the risk drivers from time t to the future time horizon $t + 1$. The appropriate invariants need to possess the following two properties: they are i.i.d across time and they only become available at future time horizon $t + 1$.

An important characteristic that a candidate invariant needs to possess is that it displays patterns that repeats itself independently and identically across time. Hence, determining whether a variable is i.i.d across time becomes essential in deciding whether it can be considered as an invariant for the asset class. There are a couple of tests that can be performed to assess whether a variable is i.i.d across time. A popular and simple test for invariance used by Meucci [140] is plotting a scatter plot of the variable against its own lags, if the plot or specifically the location-dispersion ellipsoid is a circle, then the variable is a good candidate for an invariant. It is important to note that in reality many financial time series are not perfectly i.i.d, a reason for this is they display time-varying conditional distributions and are heteroscedastic.

Once the invariants have been specified, the second step of "the Prayer" involves estimating the distribution of the invariants. Here, the distribution is specified by its probability density function $f_{\boldsymbol{\epsilon}}$. The "estimation" step is the process of fitting a distribution $f_{\boldsymbol{\epsilon}}$ to the empirical time series of the invariants $\boldsymbol{\epsilon}_{t \rightarrow t+1}$ and any other available information I_t at the current time t. Basic estimation approaches only focus on fitting a distribution to the empirical time series of the invariants, however other complex estimation approaches such as maximum likelihood, subjective Bayesian priors, and so on, can process other available information I_t. The simplest approach to estimating the distribution of invariants is the non-parametric empirical distribution of the invariant, however the distribution of invariants can also be estimated using parametric distributions such as multivariate normal, elliptical and other distributions.

The standard empirical estimation approach is anchored on the LLN and assigns equal probability weights to all historical observations of the invariants. This standard historical approach often fails to account for the time-varying aspect of financial time series and to be applicable practically, it needs to react faster to risk changes. An example of providing this time-varying conditionality is by applying higher probability weights to historical observations when volatility was higher, thus accounting for the heteroscedasticity or volatility clustering of financial time series. The flexible probability approach put forward by Meucci becomes useful in dealing with this problem.

It is crucial to note that irrespective of the advanced estimation techniques chosen to model the joint distribution of the invariants the estimation step is prone to estimation risk. Estimation risk is the uncertainty due to employing an estimate of the invariants distribution instead of the true unknown distribution of the invariants. This estimation risk starts as an issue in step 2 and filters through to the many other steps in “the Prayer” and is an inevitable risk in financial modeling. As we have now introduced the major theoretical concepts behind the HS-FP framework we can formally specify the HS-FP framework and detail the setup of the flexible probabilities in the following paragraphs.

The HS-FP model commences at a generic time t, with a set of $\bar{n}$ joint market invariants $\boldsymbol{\varepsilon}_t \equiv (\varepsilon_{1,t}, \ldots, \varepsilon_{\bar{n},t})$ that are approximately independently and identically distributed (i.i.d) across time, and considers a historical time series of the $\bar{n}$ market invariants $\{\boldsymbol{\epsilon}_t \equiv (\epsilon_{1,t}, \ldots, \epsilon_{\bar{n},t})\}_{t=1}^{\bar{t}}$ where t=1 is the first and t=$\bar{t}$ is the most recent historical observation. If the number of historical observations in the time series of market invariants is large then using the LLN, then the distribution of the future market invariant $\boldsymbol{\varepsilon}_{\bar{t}+1}$ can be estimated non-parametrically from its historical time series $\{\boldsymbol{\epsilon}_t \equiv (\epsilon_{1,t}, \ldots, \epsilon_{\bar{n},t})\}_{t=1}^{\bar{t}}$ [195].

The historical non-parametric approach to estimating future invariant distributions at current time $\bar{t}$, uses historical observations of the market invariant as forward looking scenarios with probabilities $\{\boldsymbol{p}_t\}_{t=1}^{\bar{t}}$assigned to each of these historical scenarios such that the future distribution of market invariant is estimated as:

$$\boldsymbol{\varepsilon}_{\bar{t}+1} \equiv \begin{pmatrix} \varepsilon_{1,\bar{t}+1} \\ \vdots \\ \varepsilon_{\bar{n},\bar{t}+1} \end{pmatrix} \sim \left\{ \boldsymbol{\epsilon}_t \equiv \begin{pmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{\bar{n},t} \end{pmatrix}, p_t \right\}_{t=1}^{\bar{t}} . \tag{26}$$

The standard empirical distribution, by definition, applies equal probabilities to all historical observations. Applying equal probabilities to historical observations is equivalent to assuming all historical observations are equally important. Practitioners tend to rely more on recent observations, and potentially, on additional market information.

This led to Meucci's alternatively specified flexible probabilities which are time and state conditioned. Meucci estimates the future distribution of market invariant non-parametrically [145] by using the historical observations of the market invariant as forward looking scenarios, and applying unequal flexible probabilities to these historical scenarios. The flexible probability in this paper is based on how similar the current state of the market is to the state of the market when the historical realization occurred. The state of the market can be measured by any observable state variables such as Consumer Price Index (CPI) and can arise from macroeconomic, financial, risk and other categories.

Here we set the $\bar{n}$ asset class returns as invariants, and consider a historical time series of $\bar{n}$ joint asset class returns $\{\boldsymbol{\varepsilon}_{\bar{n},\bar{t}} = \boldsymbol{r}_{\bar{n},\bar{t}}\}_{t=1}^{\bar{t}}$. Using m state variables, we obtain a historical time series of state variables $\{\boldsymbol{z}_{m,t}\}_{t=1}^{\bar{t}}$ and, relying on recent observations we derive time and state conditioned flexible probabilities $\boldsymbol{p}_t^{\text{HFP}}$. This flexible probability is then applied to historical asset class returns to estimate the future asset class return distribution:

$$\boldsymbol{r}_{\bar{t}+1} \equiv \begin{pmatrix} r_{1,\bar{t}+1} \\ \vdots \\ r_{\bar{n},\bar{t}+1} \end{pmatrix} \sim \left\{ \boldsymbol{r}_t \equiv \begin{pmatrix} r_{1,t} \\ \vdots \\ r_{\bar{n},t} \end{pmatrix}, \boldsymbol{p}_t^{\text{HFP}} \right\}_{t=1}^{\bar{t}}. \tag{27}$$

The application of flexible probabilities in practice generally takes the form of either time conditioning, state conditioning or a combination of both time and state conditioning. In the following sub-sections we detail each of these different approaches of deriving flexible probabilities separately before detailing how the time and state conditioning flexible probabilities can be blended together through Meucci's flexible probabilities approach.

6.1.1 Time Conditioned Probabilities

Time conditioned probabilities weight historical returns based on when in time they occurred. In reality, financial market data is generally not independent nor identically distributed where equally weighting all historical returns may not accurately reflect the current or future market conditions. Investment professionals frequently tend to trust and weight more recent information than those from far back in time. We consider two methods of deriving time conditioned probabilities, namely: 1) rolling window, and 2) exponentially decayed probabilities.

6.1.1.1 Rolling Window Probabilities

The simplest way to derive flexible probabilities that are time conditioned is the rolling window approach. The rolling window approach is frequently used by investment practitioners as it offers a simple and easy manner of deriving time-varying estimates. The method applies equal probability weights to all historical returns that fall within a rolling window of length λ, and applies zero probability weights to all historical returns outside this window, such that:

$$\boldsymbol{p}_t \mid \bar{t} \equiv \boldsymbol{p}_t^{roll} \propto \begin{cases} 1 & \text{if } t > \bar{t} - \lambda \\ 0 & \text{otherwise} \end{cases}. \tag{28}$$

Here the symbol $\propto$ means that the probabilities are re-scaled to sum to 1. A benefit of the rolling window approach is that it easily allows for time-varying estimation. This approach is widely used in the investment industry to give greater importance to recent data than older data. There are also problems with this approach, firstly, it reacts slowly to changes in risk, and secondly, it is quite an abrupt approach where historical returns will either receive full weight or zero weight depending on whether they fall within the targeted window.

6.1.1.2 Exponentially Decayed Probabilities

Another way of deriving flexible probabilities that are time conditioned is the exponentially decayed approach. This approach is more responsive to risk changes than the rolling window approach and offers a less abrupt manner for time conditioning. This smoother approach to deriving time conditioned flexible probabilities makes use of the exponential decay function:

$$\boldsymbol{p}_t \mid \bar{t} \equiv \boldsymbol{p}_t \propto e^{-\frac{\ln 2}{\tau}(\bar{t}-t)}. \tag{29}$$

Here τ is the half-life decay of the exponential function which represents the amount of time it would take for the probability to decay and become half the value of the highest value of $\bar{t}$. The higher the half-life, the higher the weight to recent data. In contrast to rolling window approach, the exponential decaying approach is more responsive to changing risk and offers a smoother profile where historical returns that are more recent are given a higher probability weight. Time conditioning based on exponential decayed probabilities was first introduced by Bourdoukh et al [163].

Algorithm 1 Exponential decay

function EXPONENTIAL DECAY($\bar{t}$,τ_{HL},t^*)

1. Input:
   (a) Length of exponential decay probability, $\bar{t}$
   (b) Half-life parameter, τ_{HL}
   (c) Target time parameter, t^* (optional) with default $t^* = \bar{t}$
2. Output:
   (a) exponential decay probability, $\{\boldsymbol{p}_t \mid \tau_{HL}\}_{t=1}^{\bar{t}}$
3. Compute probability, $\{\boldsymbol{p}_t \mid \tau_{HL}\}_{t=1}^{\bar{t}} = \left\{ e^{-\frac{\ln 2}{\tau_{HL}} \mid \bar{t}-t \mid} \right\}_{t=1}^{\bar{t}}$
4. Rescale, $\{\boldsymbol{p}_t \mid \tau_{HL}\}_{t=1}^{\bar{t}} = \{(\boldsymbol{p}_t \mid \tau_{HL}) / \sum_{s=1}^{t}(\boldsymbol{p}_s \mid \tau_{HL})\}_{t=1}^{\bar{t}}$

end function

In Algorithm 1, the Exponential Decay Algorithm calculates the set of $\bar{t}$ exponential decayed probabilities $\boldsymbol{p}_t|\tau_{HL}$ from specified half-life parameter τ_{HL} and target time t^*.

In Figure 5 we display an example of applying time conditioned probability through the rolling window and the exponential decayed approach. Each panel shows the probability assigned to all historical observations from the first historical observation at the leftmost point to the latest historical observation at the rightmost point on Figure 5. The rolling window approach is shown on top of Figure 5 and applies a zero-probability weight to all historical observations that fall outside the target window and applies equal probability weights to the most recent $\lambda = 60$ months. The exponential decayed approach is shown at the bottom of Figure 5 and can be seen to be smoother option where the most recent τ= 24 months historical observations are given a higher probability weight.

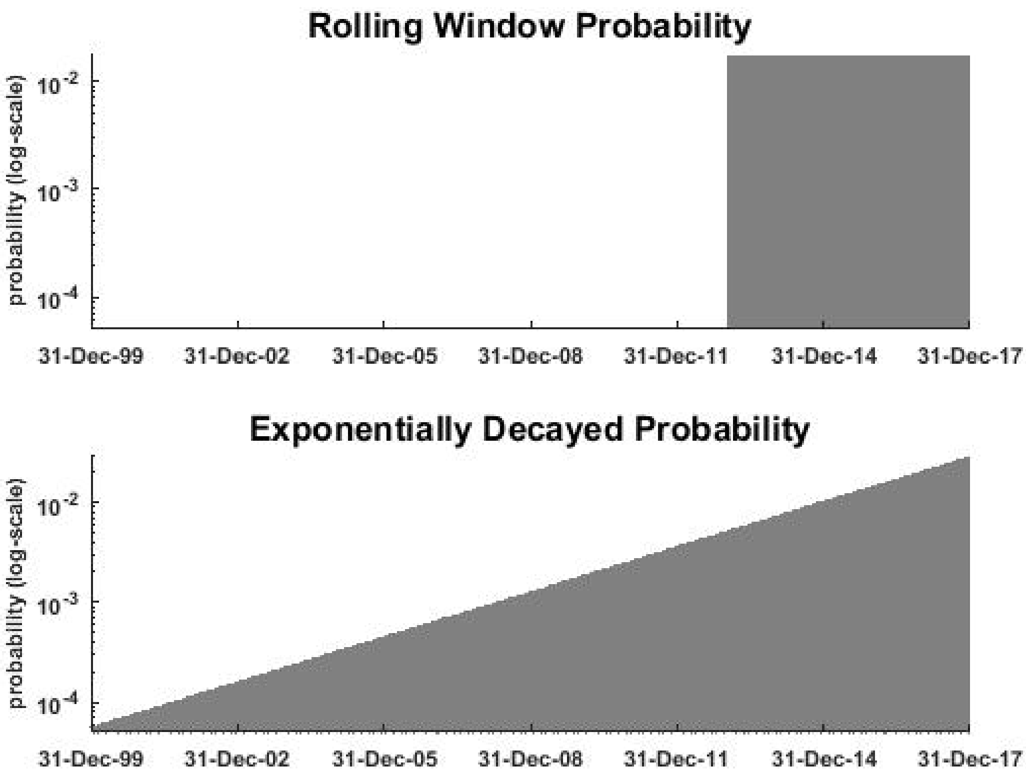

Figure 5: Shows an example of how to obtain time conditioned probabilities using: 1) the rolling window approach where $\lambda = 60$ months, and 2) an exponential decay approach where τ= 24 months. The grey region represents the weight or probability applied relative to the corresponding calendar date.

6.1.2 State Conditioned Probabilities

State conditioned probabilities weight historical returns based on when market conditions display particularly desired characteristics. State conditioning is an important lens to observe historical data, Chow, Jacquier, Kritzman, and Lowry [45] have argued that time conditioning can be irrelevant especially if no events of interest occurs over the historical analysis period, they propose the use of event-measured observations.

The state conditioned probabilities can be used to give a greater weight to history that is most similar to market conditions prevailing today or for stress testing any specific event or market conditions that is of interest. State conditioned probabilities are based on the identification of proxies for the market conditions (the proxy for market conditions are referred to as state variables or indicators), and single or multiple state indicators can be used.

These state indicators can be internal or external state indicators. Internal state indicators use the internal characteristics of an asset class as state indicators, for example the different return percentiles of an asset class. External state indicators use external market indicators that are not directly related to the asset class, for example inflation, GDP or volatility levels. We consider two methods of deriving state conditioned probabilities using a single state variable in this section, namely: 1) crisp, and 2) kernel conditioned probabilities.

6.1.2.1 Crisp Conditioned Probabilities

The simplest way to state condition is by using crisp conditioned probabilities. In this approach historical returns are given an equal weight if the state variable historically falls within a specified α range of the target state variable z^* and zero weight to all other returns. The crisp probabilities can be written as:

$$\boldsymbol{p}_t \mid z^* \equiv \boldsymbol{p}_t^{\text{crisp}} \propto \begin{cases} 1 & \text{if } \boldsymbol{z}_t \in \mathcal{R}(z^*) \\ 0 & \text{otherwise} \end{cases} . \tag{30}$$

The α range is specified such that the percentage probability is symmetrical on both sides of the target state value. However, if the α range strikes the upper or lower bound of the historical range of $\boldsymbol{z}$, then the upper or lower bound of the target range is set to the upper or lower bound of the historical range of $\boldsymbol{z}$.

The benefits and problems with this are approach is similar to the rolling window approach. The crisp state conditioned probabilities provide an easy simple approach to calculating state-varying estimations. However, the approach suffers from robustness issues where the number of observations used in the estimation can be drastically reduced based on the selected indicator range. This results in an abrupt approach where historical returns will either receive full weight or zero weight, depending on whether they fall within or outside the range of the targeted state.

Algorithm 2 Crisp Window

function CRISP($\boldsymbol{z}_t$, z_t^*,α)

1. Input:
   (a) state variable, $\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}$
   (b) target values, $\{z_t^*\}_{t=1}^{\bar{t}}$
   (c) Leeway, α
2. Output:
   (a) crisp probabilities, $\{\boldsymbol{p} \mid \mathcal{R}(z_k^*)\}_{k=1}^{\bar{k}}$
   (b) lower bound, $\{\underline{z}_k\}_{k=1}^{\bar{k}}$
   (c) upper bound, $\{\bar{z}_k\}_{k=1}^{\bar{k}}$
3. sort state variable sample, $\{\boldsymbol{z}_t^{\text{sort}}\} = \text{sort}(\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}})$ and calculate sorted state variables cdf,
$$\{F_{\boldsymbol{z}}(\boldsymbol{z}_t^{\text{sort}})\}_{t=1}^{\bar{t}} = \text{cdf}\left(\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}, \left\{\frac{1}{t}\right\}_{t=1}^{\bar{t}}\right)$$
4. evaluated at the target values, $\{F_{\boldsymbol{z}}(z_k^*)\}_{k=1}^{\bar{k}}$
5. lower and upper quantiles of state variables,
$$z_{\min} = \text{quantile}\left(\frac{\alpha}{2}, \{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}\right)$$
,
$$z_{\max} = \text{quantile}\left(1 - \frac{\alpha}{2}, \{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}\right)$$

for $k = 1, \ldots, k$ **do**
  if $P\{\boldsymbol{Z}_t > z_k^* < \alpha/2\}$ or $P\{\boldsymbol{Z}_t \leq z_k^* < \alpha/2\}$ **then**
    flatten cdf
  end if
  if $F_{\boldsymbol{z}}(z_k^*) \geq 1 - \alpha/2$ **then**
    $F_{\boldsymbol{z}}(z_k^*) = 1 - \alpha/2$
  end if
  if $F_{\boldsymbol{z}}(z_k^*) \leq \alpha/2$ **then**
    $F_{\boldsymbol{z}}(z_k^*) = \alpha/2$
  end if
  if $z_k^* \leq z_{\min}$ **then**
    $\underline{z}_k = \min(\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}})$
    $\bar{z}_k = \texttt{smooth\_quantile}(F_{\boldsymbol{z}}(z_{k^*}) + \alpha/2, \{\boldsymbol{z}_t\}_{t=1}^{\bar{t}})$
  end if
  if $z_k^* \geq z_{\max}$ **then**
    $\bar{z}_k = \texttt{smooth\_quantile}(F_{\boldsymbol{z}}(z_{k^*}) - \alpha/2, \{\boldsymbol{z}_t\}_{t=1}^{\bar{t}})$
    $\bar{z}_k = \max(\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}})$
  else
    $\bar{z}_k = \texttt{smooth\_quantile}(F_{\boldsymbol{z}}(z_{k^*}) - \alpha/2, \{\boldsymbol{z}_t\}_{t=1}^{\bar{t}})$
    $\bar{z}_k = \texttt{smooth\_quantile}(F_{\boldsymbol{z}}(z_{k^*}) + \alpha/2, \{\boldsymbol{z}_t\}_{t=1}^{\bar{t}})$
  end if
  $\{\boldsymbol{p}_t \mid \mathcal{R}(z_k^*)\}_{t=1}^{\bar{t}} = \left\{\mathbf{1}_{\underline{z}_k \leq \boldsymbol{z}_t \leq \bar{z}_k}\right\}_{t=1}^{\bar{t}}$
  $\{\boldsymbol{p}_t \mid \mathcal{R}(z_k^*)\}_{t=1}^{\bar{t}} = \{\{(\boldsymbol{p}_t \mid \mathcal{R}(z_k^*)\}_{k=1}^{\bar{k}})/\sum_{s=1}^{\bar{t}}(\boldsymbol{p}_s \mid \mathcal{R}(z_k^*)\}_{t=1}^{\bar{t}}$
end for
end function

In Algorithm 2, the Crisp Algorithm calculates the crisp state conditioned probabilities based on the time series of state variable $\boldsymbol{Z}_t$ as well as outputting the lower $\underline{z}_k$ and upper $\bar{z}_k$ bounds.

6.1.2.2 Kernel Conditioned Probabilities

The smoother method to time condition is Kernel conditioning. In this approach historical returns are given a weight based on the distance between the state variable at the given point in time and the target state variable. The distance is measured using either an exponential or a Gaussian kernel; we can then write the probabilities as:

$$\boldsymbol{p}_t \mid z^* \equiv \boldsymbol{p}_t^{\text{ker}} \propto e^{\frac{-\mid \boldsymbol{z}_t - z^* \mid^\gamma}{h}}. \tag{31}$$

Here the constant h is the bandwidth of the kernel and determines the smoothness of the kernel and hence the smoothness of the probabilities. The constant γ indicates the type of kernel chosen, where $\gamma = 1$ corresponds to an exponential kernel and $\gamma = 2$ corresponds to a Gaussian kernel. The kernel conditioned probabilities are suggested by Meucci in [145] to reduce the robustness issue prevalent in crisp state conditioning by allowing the use of all data in the estimation procedure.

Algorithm 3 Kernel Probability

---

function KERNEL PROBABILITY($\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}$,$z^*$,$h$,$\gamma$)

1. Input:
   (a) State variable, $\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}$
   (b) Target value for state variable, z^*
   (c) Bandwidth parameter, h
   (d) Kernel type parameter, γ
2. Output:
   (a) Kernel probability, $\{\boldsymbol{p}_t \mid z^*\}_{t=1}^{\bar{t}}$
3. Compute probability, $\{\boldsymbol{p}_t \mid z^*\}_{t=1}^{\bar{t}} = \left\{ e^{\frac{-\mid \boldsymbol{z}_t - z^* \mid^\gamma}{h}} \right\}_{t=1}^{\bar{t}}$
4. Rescale, $\{\boldsymbol{p}_t \mid z^*\}_{t=1}^{\bar{t}} = \{(\boldsymbol{p}_t \mid z^*)/\sum_{s=1}^{t}(\boldsymbol{p}_s \mid z^*)\}_{t=1}^{\bar{t}}$

end function

In Algorithm 3, the Kernel Probability Algorithm calculates the set of $\bar{t}$ kernel probabilities $\boldsymbol{p}_t \mid z^*$ for specified bandwidth parameter h and kernel type parameter γ for a given state variable $\boldsymbol{Z}_t$ with target state variable z^*.

In Figure 6 we show how CPI can be used as a state variable and use the latest CPI state value as the desired market condition we want to stress. We smooth, score as well as standardize the raw CPI data into the market state variable and condition by latest CPI market state value.

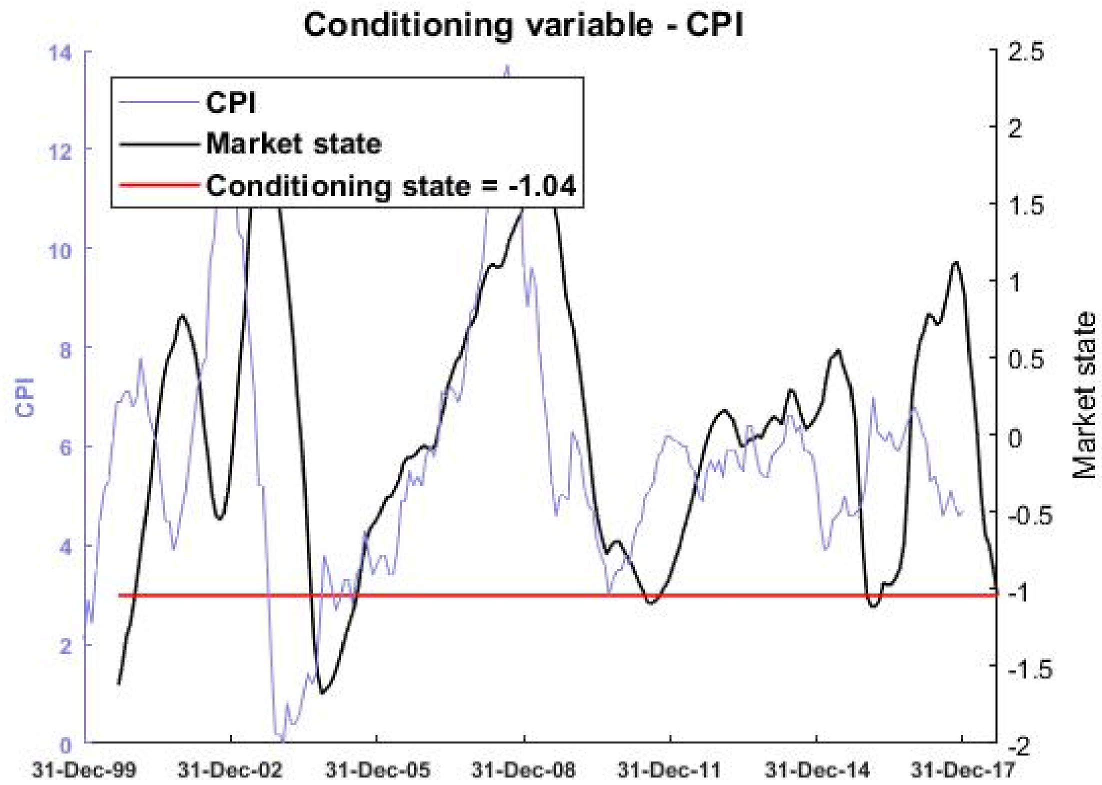

Figure 6: Here we consider month-on-month changes in inflation, CPI (Table 3), as a state variable. We plot the raw CPI on y_1 axis which we then smooth, score and finally standardize into the market state variable which we plot on y_2 axis, and condition by target state $CPI = -1.04$, which is the latest CPI market state value as shown by the red line.

In Figure 7 we display an example of applying state conditioned probability through the crisp or the kernel conditioned probabilities. Each panel shows the probability assigned to all historical observations from the first historical observation at the leftmost point to the latest historical observation at the rightmost point on Figure 7. The crisp approach is shown on top of Figure 7 and applies a zero probability weight to all historical observations that fall outside the α range and applies equal probability weights to all the observations that fall with α range. The kernel conditioned approach is shown at the bottom of Figure 7 and can be seen to be smoother than the crisp option where observations that are closer to the current level of CPI are gradually given higher probability weighting.

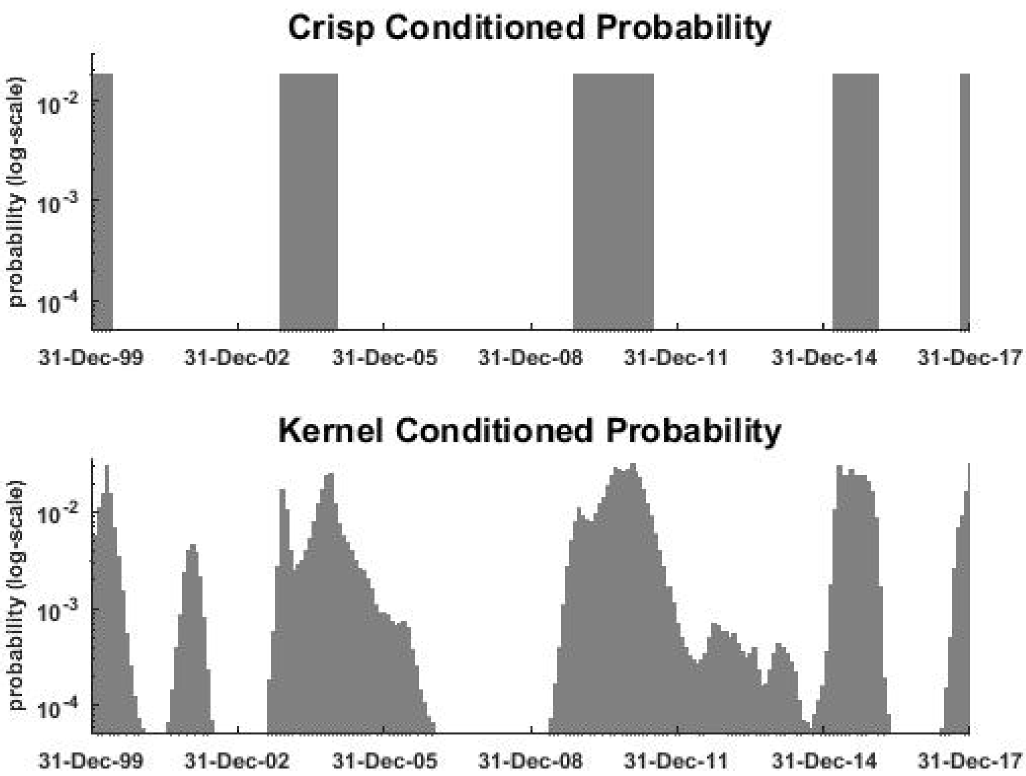

Figure 7: Shows an example of how to obtain state conditioned probabilities using: 1) crisp conditioning probability, and 2) kernel conditioning probability with h = 0.5 and $\gamma = 2$. The state variable used to condition is CPI and the target state is CPI=-1.04. The grey region represents the weight or probability applied relative to the corresponding calendar date.

6.1.3 Time and State Conditioned Probabilities

As mentioned in the previous subsections we can derive flexible probabilities that are conditioned by time or conditioned by state. But, both time and state conditioning are important, so it would be useful to derive flexible probabilities that are conditioned by both time and state. For example, we can give higher weight to historical returns that are both more recent and have similar characteristics to current market conditions.

Meucci [145] demonstrates that the HS-FP approach [141] can be used to derive time and state conditioned flexible probabilities by utilizing some of the aforementioned techniques. This approach delivers a forward looking or posterior distribution that reflects most recent and desired market conditions with least distortion to prior distribution. However, we need to be able to quantify how different one distribution is from another.

6.1.3.1 Entropy-Based Conditioning Procedure

It is through the relative entropy criterion that we can quantify how different one distribution is from another, and by minimizing the relative entropy we obtain a forward looking distribution that is as close as possible to the prior distribution. Meucci [145] applies the following steps to ensure that the distribution is both time and state conditioned.

First, he specifies a state variable $\boldsymbol{z}$ with a target state value for this state variable z^* and sets the initial flexible probabilities as the crisp state conditioned probabilities:

$$\boldsymbol{p}_t \mid z^* \equiv \boldsymbol{p}_t^{\text{crisp}} \propto \begin{cases} 1 & \text{if } \boldsymbol{z}_t \in \mathcal{R}(z^*) \\ \\ 0 & \text{otherwise} \end{cases}. \tag{32}$$

Second, he calculates the crisp mean and standard deviation of the target state variable as below:

$$\mu \mid z^* = \sum_{t=1}^{\bar{t}} \boldsymbol{z}_t \boldsymbol{p}_t^{\text{crisp}} \text{ and } \sigma \mid z^* = \sqrt{\sum_{t=1}^{\bar{t}} \boldsymbol{z}_t^2 \boldsymbol{p}_t^{\text{crisp}} - (\mu \mid z^*)^2} \tag{33}$$

Third, the views are expressed. The views are statements on target state variable that distort the prior distribution in the least spurious way. The view is that the yet-to-be determined flexible probabilities $\boldsymbol{p}_t$ must match the moments of the crisp state conditioned probability:

$$\nu \mid z^* \equiv \begin{cases} \sum_{t=1}^{\bar{t}} \boldsymbol{z}_t \boldsymbol{p}_t = \mu \mid z^* \\ \\ \sum_{t=1}^{\bar{t}} \boldsymbol{z}_t^2 \boldsymbol{p}_t \leq (\mu \mid z^*)^2 + (\sigma \mid z^*)^2. \end{cases} \tag{34}$$

Finally, Meucci sets prior distribution as the exponentially decayed time conditioned probabilities

$$\boldsymbol{p}_t \mid \bar{t} \equiv \boldsymbol{p}_t \propto e^{-\frac{\ln 2}{\tau}(\bar{t}-t)} \tag{35}$$

and calculates the posterior distribution that satisfies views by minimizing the relative entropy or Kullback-Leibler divergence [126] between the exponential decayed time conditioned probability prior in equation 35 and flexible probability distributions while satisfying crisp state moment views in equation 34

$$\mathbf{p} \mid z^* \equiv \underset{\mathbf{p} \in \nu \mid z^*}{\operatorname{argmin}} \, \epsilon(\mathbf{p}, \mathbf{p}^{\exp}) \tag{36}$$

where the relative entropy in discrete form is given by :

$$\epsilon(\mathbf{p}, \mathbf{p}^{\exp}) = \sum_{t=1}^{\bar{t}} \boldsymbol{p}_t \ln\left(\frac{\boldsymbol{p}_t}{\boldsymbol{p}_t^{\exp}}\right). \tag{37}$$

The final flexible probability as explained in Meucci [145] is a combination of the exponential decay prior equation 29 and kernel conditioning with optimal bandwidth and center. As a result of setting an inequality view on the second moment in equation 34 the kernel estimator is allowed to switch freely between the best exponential kernel and best Gaussian kernel.

To summarize, the relative entropy criterion is proposed as a mechanism to quantify the uncertainty within a distribution and to quantify how different one distribution is from another. The aim is to obtain a time and state conditioned probability by distorting the prior distribution of time conditioned probability in the least intrusive manner such that the new distribution moments match the moments of the state conditioned distribution. The least intrusive manner is where the relative entropy is minimized between the prior and posterior distributions. We detail how the joint time and state conditioned probability and the minimum relative entropy criterion is setup in the Algorithms 4 and 5 that follow.

Algorithm 4 Time and State Probability

function TIME AND STATE PROBABILITY($\boldsymbol{z}_t$, z^*_t,α)

1. Input:
   (a) State variable, $\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}$
   (b) Target values, $\{z^*_t\}_{t=1}^{\bar{t}}$
   (c) Leeway, α
2. Output:
   (a) Time and state probability, $\tilde{\boldsymbol{p}}$
3. Compute crisp probabilities, $\{\boldsymbol{p}^{\text{crisp}(k)}\}_{k=1}^{\bar{k}} = \texttt{crisp\_fp}(\{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}, \{z^*_k\}_{k=1}^{\bar{k}}, \alpha)$

for $k = 1, \ldots, \bar{k}$ **do**

$m \mid z^*_k = \sum_{t=1}^{\bar{t}} \boldsymbol{z}_t \boldsymbol{p}_t^{\text{crisp}(k)}$

$s^2 \mid z^*_k = \sum_{t=1}^{\bar{t}} \boldsymbol{z}_t^2 \boldsymbol{p}_t^{\text{crisp}(k)} - (m \mid z^*_k)^2$

$\{\boldsymbol{a}_t^{\text{ineq}}\}_{t=1}^{\bar{t}} = \{\boldsymbol{z}_t^2\}_{t=1}^{\bar{t}}$

$b^{\text{ineq}} = (m \mid z^*_k)^2 + s^2 \mid z^*_k$

$\{\boldsymbol{a}_t^{\text{eq}}\}_{t=1}^{\bar{t}} = \{\boldsymbol{z}_t\}_{t=1}^{\bar{t}}$

$b^{\text{eq}} = m \mid z^*_k$

$\boldsymbol{p} \mid (z^*_k) = \texttt{min\_rel\_entropy\_sp}(\boldsymbol{p} \mid \tau_{HL}, \boldsymbol{a}^{\text{ineq}}, b^{\text{ineq}}, \boldsymbol{a}^{\text{eq}}, b^{\text{eq}})$

end for

end function

In Algorithm 4, the Time and State Probability Algorithm calculates the joint time and state conditioned flexible probabilities specified in equation 36. Specifically it calculates the posterior set of flexible probabilities by combining the crisp state conditioned probabilities with the set of prior time conditioned probabilities through the minimum relative entropy criterion. Usually the prior time conditioned probabilities are specified as the exponential decayed probabilities.

Algorithm 5 Minimum Relative Entropy

function MINIMUM RELATIVE ENTROPY($\boldsymbol{prior}$,a^{ineq},b^{ineq},a^{eq},b^{eq})

1. Input:
   (a) prior probability, $\underline{\boldsymbol{p}}$
   (b) inequality constraints, a^{ineq},b^{ineq}
   (c) equality constraints, a^{eq},b^{eq}
2. Output:
   (a) posterior probability, $\bar{\boldsymbol{p}}$
3. concatenate the inequality and equality constraints, $v = \begin{pmatrix} a^{ineq} \\ a^{eq} \end{pmatrix}$, $\mu = \begin{pmatrix} b^{\text{ineq}} \\ b^{\text{eq}} \end{pmatrix}$
4. normalize the constraints to avoid numerical problem,
   (a) $\hat{m}_v = \text{mean}(\{v_{\cdot j}\}_{j=1}^{\bar{j}})$
   (b) $\hat{s}_v = \text{std}(\{v_{\cdot j}\}_{j=1}^{\bar{j}})$
   (c) $\{v_{\cdot j}\}_{j=1}^{\bar{j}} = \{(v_{\cdot j} - \hat{m}_v)./\hat{s}_v\}_{j=1}^{\bar{j}}$
   (d) $\mu = (\mu - \hat{m}_v)./\hat{s}_v$
5. for probability,
   (a) $p(l) = \left\{ \underline{p}^{(j)} \times e^{l \times v_{\cdot j}} / \sum_{k=1}^{\bar{j}} \underline{p}^{(k)} \times e^{l \times v_{\cdot k}} \right\}$
   (b) $h(l) = \ln \sum_{j=1}^{\bar{j}} \underline{p}^{(j)} \times e^{l \times (v_{\cdot j} - \mu)}$
   (c) $\nabla_l h(l) = \sum_{j=1}^{\bar{j}} [p(l)]_j \times (v_{\cdot j} - \mu)$
   (d) $\nabla_l^2 h(l) = \sum_{j=1}^{\bar{j}} [p(l)]_j \times (v_{\cdot j} - \mu - \nabla_l h(l)) \times (v_{\cdot j} - \mu - \nabla_l h(l))$
6. compute,
   (a) $a^{ineq} = \begin{pmatrix} \mathcal{J} & 0 \\ 0 & 0 \end{pmatrix}$ and $b^{\text{ineq}} = 0$,
   (b) $\text{grad} = \nabla_l h(l)$,
   (c) $\text{hessian} = \nabla_l^2 h(l)$,
   (d) $\theta = \text{minimize}\left(h(l), a^{ineq}, b^{\text{ineq}}, \text{grad}, \text{hessian} \right)$
   (e) $\bar{\boldsymbol{p}} = p(\theta)$

end function

In Algorithm 5, the Minimum Relative Entropy Algorithm calculates the posterior probability $\bar{\boldsymbol{p}}$ by solving for the optimization problem specified in equation 36. In the case that both equality and inequality are not specified the algorithm returns prior probability $\underline{\boldsymbol{p}}$.

In Figure 8 we display an example of applying time, state and joint time and state conditioned probability using the exponential decayed, crisp and entropy pooling approach respectively. Each of the first three panels show the probability assigned to all historical observations from the first historical observation at the leftmost point to the latest historical observation at the rightmost point on Figure 8. In Figure 8 the exponential decayed time conditioned approach is shown first, followed by the crisp state conditioned approach, thereafter the time and state conditioned entropy pooling approach, finally the scatter of the historical monthly returns of the Resources 20 Index is shaded based on the probabilities derived from the time and state conditioned entropy pooling approach. In the second and third panel the state indicator that is used to condition returns is the CPI state variable and target state value as specified in Figure 6.

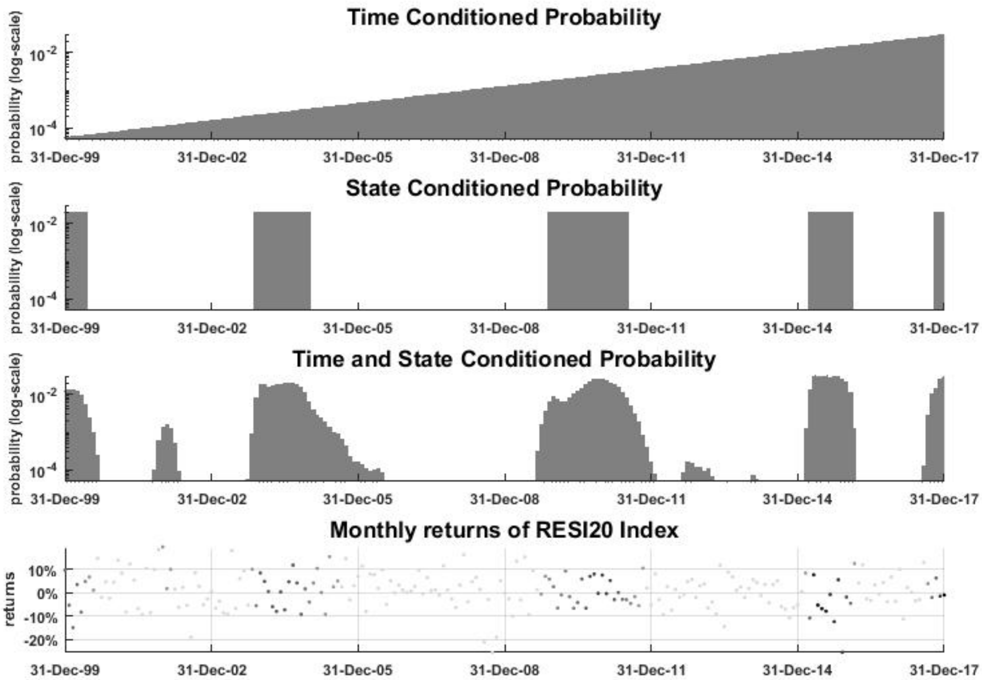

Figure 8: The first three panels show an example of obtaining time, state, time and state conditioned probabilities using a single state variable. The state variable used to condition is CPI and the target state is CPI=-1.04. The grey region represents the weight used relative to the corresponding calendar date.The first panel shows time conditioning by exponential decayed probability where τ= 24 months, the second panel shows state conditioning by crisp conditioning, the third panel shows time and state conditioning using entropy pooling approach. Finally the last panel shows how the Resource 20 Index returns (see Table 1) can be conditioned by time and state using entropy pooling approach.

6.2 Combining Multiple State Variables

In practice, market participants would prefer to combine information from multiple different state variable. This could enhance diversification benefits by not concentrating all views of the market condition in a single state variable. It has also been cited by Bates and Granger[16] that there is an improvement of forecast performance when a combination of forecasts are used together rather than using each of the individual forecasts in isolation. However, in order to do so we need to calculate the time and state conditioned probability for each state variable and decide how much to weight each probability:

$$\mathbf{p}_{comb} = w_1 \mathbf{p} \mid z_1^* + \ldots + w_{\bar{q}} \mathbf{p} \mid z_{\bar{q}}^*. \tag{38}$$

We consider two combination methodologies equal weighting and degree of conditioning and correlation weighting.

6.2.1 Equal Weighting

The simplest approach to weighting the probabilities of the different state variables is equal weighting: $EQ_q = \frac{1}{q}$, where $q = 1, \ldots, \bar{q}$ is the number of state variables. The equal weighting approach is straight forward and has been argued by many as a hard combination method to beat Timmermann [190] and Rapach, Strauss, and Zhou [160].

6.2.2 Degree of Conditioning and Correlation Weighting

In this approach the flexible probability from each state variable is weighted based on the Degree of Conditioning and Correlation (DCC) between the different state variables [144]. If a given state variables is lowly correlated with the other state variables, or imposes a smaller degree of conditioning on many historical scenarios, then this state variable will receive a higher weighting.

To measure the degree of conditioning from a given state variable the effective number of scenarios introduced by Meucci in [144] is used:

$$\mathcal{T} = e^{-\sum_{t=1}^{\bar{t}} \boldsymbol{p}_t \ln \boldsymbol{p}_t}. \tag{39}$$

When all historical scenarios are equally weighted, $p_t = \frac{1}{t}$, then the degree of conditioning is minimal and the effective number of scenarios is maximal at $\mathcal{T} = \bar{t}$. On the other hand, if only one historical scenario is assigned all probability, then the degree of conditioning is maximal, and the effective number of scenarios is minimal with $\mathcal{T} = 1$.

To measure the degree of correlation between a given state variable and the rest of the state variables the following steps are followed. Firstly the Bhattacharyya coefficient [22] is used to measure the degree of correlation between the different state variables for any pair of flexible probabilities $(\boldsymbol{p}_q, \boldsymbol{p}_r)$ as:

$$\boldsymbol{b}_{q,r} \equiv \sum_{t=1}^{\bar{t}} \sqrt{p_{t,q} p_{t,r}} \tag{40}$$

then the Hellinger distance [95] is calculated as

$$d_{q,r} \equiv \sqrt{1 - b_{q,r}}. \tag{41}$$

Finally, the diversity index [186], $\mathcal{D}_q$ is used to summarize the degree of similarity between the probabilities from a given state variable and the probabilities from the others state variables. The diversity index is basically the average of the Hellinger distances [96] between the given set of probabilities and the remaining probabilities:

$$\mathcal{D}_q = \frac{1}{\bar{q} - 1} \sum_{r \neq q} d_{q,r}. \tag{42}$$

The final weighting for the flexible probability from each state variable is then calculated as

$$\boldsymbol{DCC}_q = \frac{\mathcal{T}_q \mathcal{D}_q}{\sum_{r=1}^{\bar{q}} \mathcal{T}_r \mathcal{D}_r} \tag{43}$$

where $q = 1, \ldots, \bar{q}$. The $\boldsymbol{p}_{comb}$ is the time and state conditioned probability derived from weighting each of the multiple state variables by DCC approach. This is also referred to as the ensemble flexible probability $\boldsymbol{p}_{ensemble}$.

Algorithm 6 Effective Number of Scenarios

function EFFECTIVE NUMBER OF SCENARIOS($\boldsymbol{p}^{\exp}$, `type_ent`,γ)

1. Input:
   (a) exponential entropy, $\boldsymbol{p}^{\exp}$
   (b) type of entropy, `type_ent`
   (c) parameter of exponential entropy, γ
2. Output:
   (a) effective number of scenarios, ens(p)
3. **if** `type_ent` $=$ exp *or* none **then**
4. ens$(\boldsymbol{p}) = e^{-\sum_{t=1}^{\bar{t}} \boldsymbol{p}_t \ln \boldsymbol{p}_t}$
5. **else**
6. $\text{ens}_\gamma(\boldsymbol{p}) = [\sum_{t=1}^{\bar{t}} (\boldsymbol{p}_t)^\gamma]^{-1/(\gamma-1)}$
7. **end if**
8. **end function**

In Algorithm 6, the Effective Number of Scenarios Algorithm calculates the effective number of scenarios in a set of probabilities by using different types of functions.

In Figure 9 we display an example of applying time and state conditioned probabilities using state variables individually and in combination. The state variables used are CPI and GDP with latest value of each specified as the target state value. The first two panels of Figure 9 overlays the time and state conditioned probability with the state variable used to condition the probability. The first panel uses CPI as the state variable and the latest CPI state value to condition the probability. The second panel uses GDP as the state variable and the latest GDP state value to condition the probability. The third panel shows the time and state conditioned ensemble probability which uses ensemble weighting formula shown in equation 43 to derive the combined weighting to each of the probabilities displayed in panel 1 and panel 2 of Figure 9. Finally, in the last panel the historical monthly return series of the Resources 20 Index is shaded based on the probabilities derived from the time and state conditioned ensemble probability displayed in the third panel.

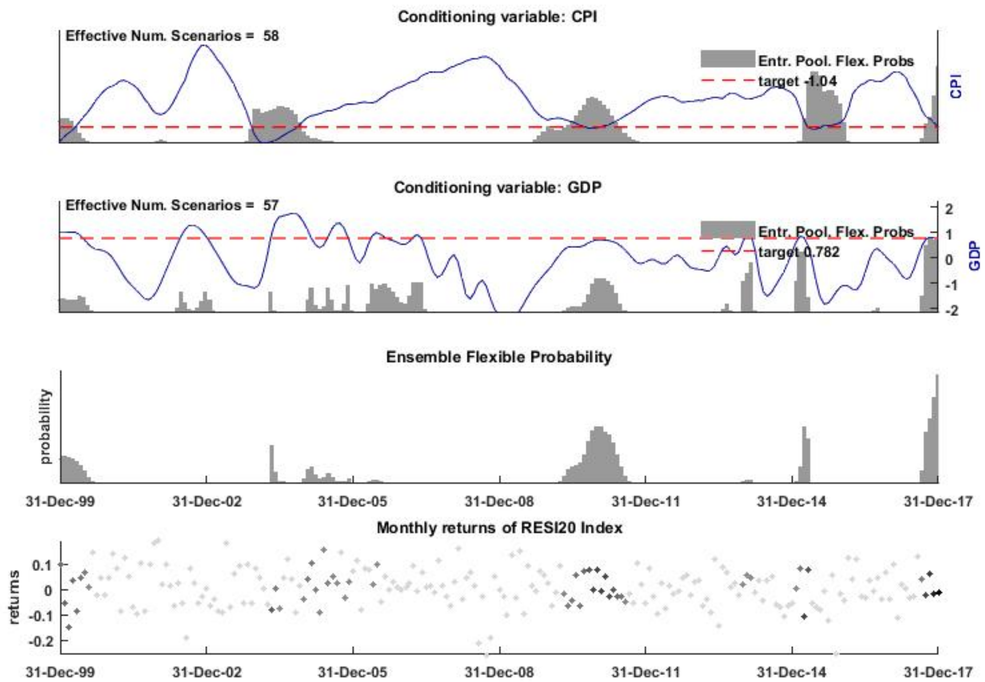

Figure 9: Shows an example of how to obtain time and state conditioned probabilities using multiple state variables CPI and GDP. In the first panel we condition only by the latest market state of CPI state variable and overlay the time and state conditioned probability. In the second panel we condition only by the latest market state of GDP state variable (see Table 3) and overlay the time and state conditioned probability. In the third panel we show the ensemble probability when conditioning by both time and multiple state variables. Finally in the last panel we show the Resources 20 Index (see Table 1) conditioned by both time and the multiple state variables CPI and GDP.

7 Benchmark Models

“Performance evaluation cannot be conducted in a vacuum. By its nature, performance evaluation is a relative concept” [105]. To assess any investment strategy successfully, it needs to be compared relative to an appropriate benchmark model. The benchmarking exercise is an important component in assessing the performance, risk and application of a potential investment strategy in a practical sense. Timmerman and Blake discuss the importance of benchmarking for institutional investors [24]. They argue that the assignment of benchmarks are of crucial importance to investors as it drives the investment behavior of the fund managers and hence the investment outcome experienced by the investor. The benchmark model should also be chosen to align with the risk appetite and loss tolerance of client objectives [7].

In the asset management industry, there are commonly three approaches used to benchmark multi-asset portfolios. The first approach uses a combination of indices that proxy various asset classes that are then weighted based on a client specified asset allocation. The client specified asset allocation generally takes a long term or strategic asset allocation format. The strategic asset allocation could be either: an equally weighted approach, based on mean variance optimization using long term historical average return and risk estimates; or based on a fundamental subjective view of the optimal long term asset allocation. The second approach compares the performance of the portfolio to a group of external peers that ideally have a similar investable universe and mandate constraints. The third approach specifies a benchmark with an inflation beating target for the portfolio, in South Africa the high equity balanced category generally has a CPI + 5 % target while the low equity balanced category generally has CPI + 2 % target.

In this paper we utilize the first approach of benchmarking where we proxy asset classes with indices and consider two weighting allocations namely: 1) equal weighted, and, 2) mean variance optimization weighted. The HS-FP performance, risk, allocation and other important statistics are then compared to the equal weighted and mean variance optimization weighted benchmark models. In Section 9.5 we also compare the performance and risk of HS-FP portfolios with the multi-asset peer group performance and risk as seen in Table 7. We acknowledge that this is not a representative benchmark as the investable universes are not similar, the HS-FP investable universe does not include global asset classes whereas the funds in the peer group could contain global allocations. The equally weighted and mean variance optimization benchmark models are favored as it contains the same investable universe opportunity set to the HS-FP model. In the following sections we will detail the equal weighted and mean variance optimization weighted benchmark models in more detail.

7.1 Equal Weighted (EW)

Equal weighted or $1/n$ portfolios is a naive and straight forward approach to use as a benchmark, and, this approach has been widely used as a benchmark by investment practitioners [17, 207]. Block and French [25] states that the equal weighted benchmark is an appropriate benchmark for many equity portfolio managers in United States as the investment process followed by them is generally similar to an equally weighting approach. Portfolios that have used equal weighted indices as a benchmark have been argued to be more diversified than those that have used market cap indices as their benchmark. However, equal weighted indices tend to upweight the smaller illiquid companies in comparison to market cap indices, and in highly concentrated equity markets when large portfolios have equal weighted benchmarks it has been seen that the buying and selling of the smaller companies can have a negative impact on the prices that are executed by the fund [64]. The equal weighted strategy has been cited in numerous academic literature to be efficient out-of-sample [57] and hard to beat [190] and [160], supporting the usage of equal weighting as a benchmark model in the paper.

7.2 Mean Variance Optimization (MVO)

As elaborated in Section 3 the mean variance optimization framework formed the first quantitative framework for optimal portfolio selection and construction. In this paper the classic Mean Variance Optimization (MVO) benchmark model is used as it assumes that the predicted mean of the next period's return distribution is taken as the current historical mean return, this is equivalent to the HS-FP model with equal weighted probabilities assigned to historical returns. This benchmark model is particularly useful in this paper as at its core the HS-FP model assesses whether a joint time and state conditioned probability can add value over and beyond the standard equal weighted probability. In practice, many pension funds and mutual funds derive strategic asset allocations for their investment portfolios using the classic MVO. Additionally, many academic papers [201, 28, 83] have argued that most of the traditional return predictors fail to outperform the historical average, thus supporting the usage of classic mean variance optimization portfolio as a benchmark model in this paper.

8 Implementation

In this section we detail all the relevant parts that enable the implementation of the HS-FP model. We discuss the data that is used in the analysis as well as outline the methodology utilized. The section is split into two subsections namely: 1) Data and 2) Methodology. In the data subsection we discuss the asset class and state variable data that is used in the paper, and in the methodology subsection we detail how this asset class and state variable data is then used to deliver each asset class's next period forecasts that are then backtested and results shown in Section 9.

8.1 Data

In this section we detail the data used in the paper. The data includes both the asset class data as well as state variables used. We use monthly data for both the five asset classes and the ten state variables. The asset classes cover the three broad equity sectors, bonds and currency. The state variables belong to macroeconomic, financial and risk indicator categories. Where a state variable is not available in monthly frequency we convert this variable into monthly frequency. We obtain the data for both the asset classes and state variables for the past approximately 19 years *i.e.* Feb 1998 to Dec 2017. All the data is sourced from Bloomberg. The next subsections will provide some explanatory analysis on both the asset class and state variable data.

8.1.1 Data on Asset Classes

In this section the asset classes considered in this paper are introduced. Table 1 below lists the names of the asset classes, proxies and codes used to extract from Bloomberg and Table 2 shows the summary return, risk and Sharpe ratio of the various asset classes where the risk-free rate is calculated to be 8.74% over the period Feb 1998 to Dec 2017.

Asset Class	Code	Proxy
Resource	RESI20 Index	FTSE/JSE Africa Resource 10 Index
Industrial	INDI25 Index	FTSE/JSE Africa Industrial 25 Index
Financial	FINI15 Index	FTSE/JSE Africa Financial 15 Index
All Bond	ALBTR Index	ALBI Total Return Index
USDZAR Currency	USDZAR Curncy	United State Dollar/South African Rand Cross

Table 1: Asset Classes: The list of asset classes is given with their Bloomberg codes and the proxies used to represent these asset classes.

In the explanatory analysis we first observe a summary table containing the performance, risk and risk adjusted returns of the various asset classes. We look at the annualized performance of the various asset classes, we proxy risk and risk adjusted returns by the annualized volatility and Sharpe ratio respectively. The five asset classes namely: resources, industrials, financials, bonds and currency are listed in Table 2 alongside the summary statistics mentioned before. The risk free rate is calculated to be 8.74% during the analysis period.

Asset Class	Ann. Return	Ann. Volatility	Ann. Sharpe Ratio
Industrial	13.4%	19.1%	0.25
Financial	6.6%	21.4%	-0.10
Resources	10.6%	28.3%	0.07
Bonds	11.0%	8.6%	0.27
Currency	-4.5%	16.2%	-0.82

Table 2: Indicative Asset Class Performance: The annualised returns, annualised volatilities, and finally, annualised Sharpe ratios are provided as summary statistics for the asset classes over the period February 1998 until December 2017.

In Table 2 we see that the industrial index has the highest annualized return, lowest risk and highest risk adjusted return out of the three equity sectors during the analysis period. Whereas the financial index has the lowest annualized return and lowest risk adjusted return from the three equity sectors. The bond asset class has the highest risk adjusted return from the five asset classes and the currency asset class has the lowest return and risk adjusted return of the asset classes. Observing the summary table an investor would have been best off from a return perspective from having a higher allocation to industrial and lowest allocation to currency asset classes. The summary table shows the long term statistics, there could be interesting insights gained about the varying performance, risk and risk-adjusted returns of the asset classes in different historical periods when the rolling 1 year statistics are observed as seen in Figure 10, 11 and 12.

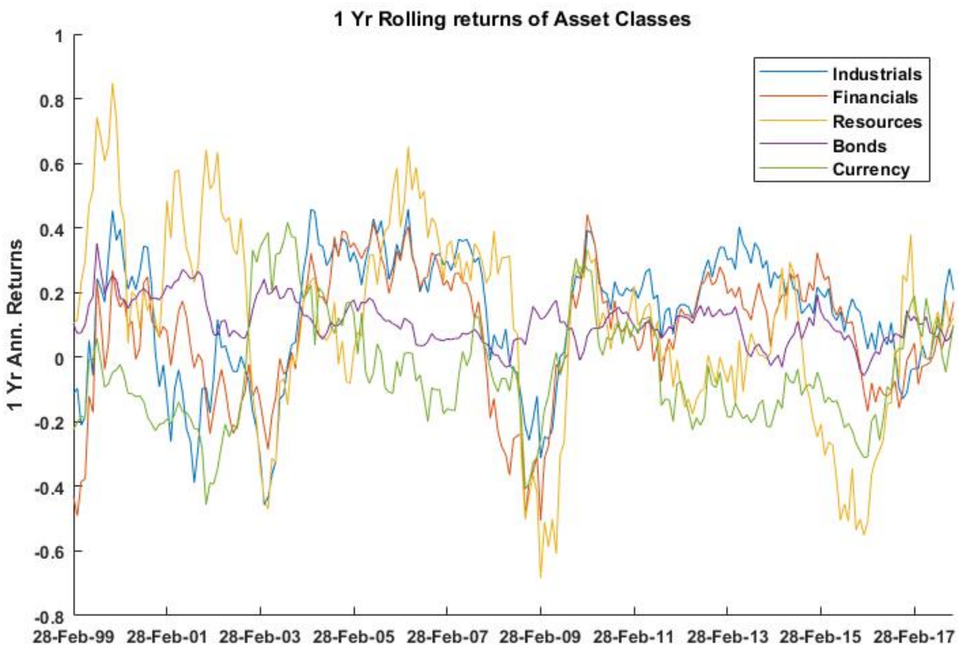

Figure 10: Rolling 1 Year return for Asset Classes: A comparison of the 1 Year rolling return for the industrial, financial, resource, bonds and currency asset classes over the period February 1999 until Dec 2017.

To understand the time-varying return behavior of the various asset classes, we plot the rolling 1 year annualised returns of the asset classes in Figure 10. Although over the entire analysis period the industrial index has the highest annualised return (see Table 2), it can be observed that when rolling 1 year returns are assessed, that no single asset class has consistently outperformed or underperformed the other asset classes at every single period of time historically. For example, the resource index recorded good performance in the earlier periods 1998-2002 however struggled in 2002-2003, while currency experienced poor performance in 1998-2003 and recorded good performance in 2003-2004. The resource index generally struggles in recessionary periods such as the dotcom bubble (2002-2003) and the global financial crisis (2007-2008). The stable performance of the bond index through time can be observed in 10, the bond index offers most comfort to investors during crisis periods, recording best relative performance to other asset classes in the dotcom bubble and global financial crisis. More recently in the 2012-2018 period the industrial and financial indices recorded the better performance while the resource and currency indices struggled in comparison to the other asset classes.

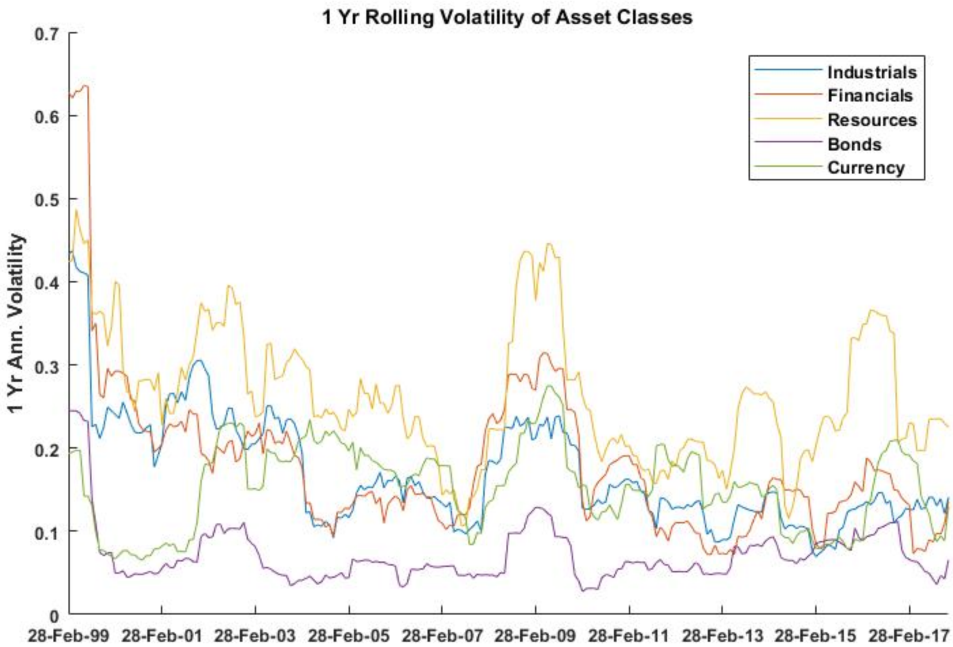

Figure 11: Rolling 1 Year Volatility for Asset Classes: A comparison of the 1 Year rolling volatility for the industrial, financial, resource, bonds and currency asset classes over the period February 1999 until Dec 2017.

To understand the time-varying volatility behavior of the various asset classes, we plot the rolling 1 year annualised volatility of the asset classes in Figure 11. It is interesting to note that over the entire analysis period the resource index has the highest annualised volatility (see Table 2) and when rolling 1 year volatility is observed the resources index also generally has the highest volatility during historical time-varying periods. This is understandable as the resource index tends to be cyclical, related to commodity cycles and sentiment pertaining to future economic growth. The volatility of resource index spikes during recessionary periods such as the dotcom bubble (2002-

2003) and the global financial crisis (2007-2008) due to negative sentiment surrounding future growth. The bond index has the lowest volatility over the entire analysis period and generally maintains this characteristic through time. This is understandable as the bond index is mostly compromised of government bonds with low risk association as it is unlikely that the government would default on future bond payments.

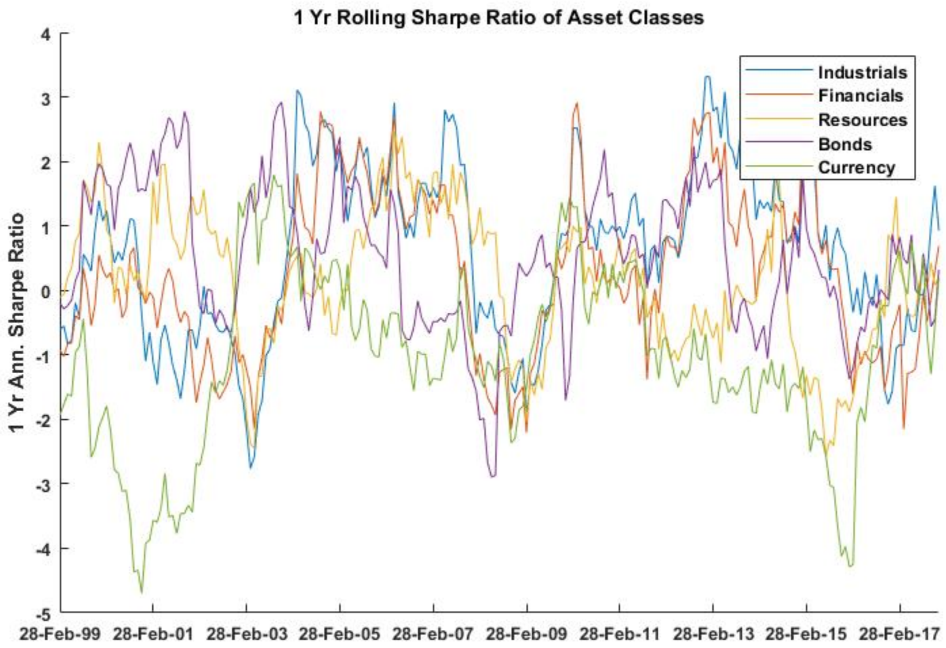

Figure 12: Rolling 1 Year Sharpe Ratios for Asset Classes: A comparison of the 1 Year rolling Sharpe ratio for the industrial, financial, resource, bonds and currency asset classes over the period February 1999 until Dec 2017.

Finally, to understand the time-varying risk-adjusted return behavior of the various asset classes, we plot the rolling 1 year Sharpe ratio of the asset classes in Figure 12. Although over the entire analysis period the bond index followed by industrial index has the highest Sharpe ratio (see Table 2), it can be observed that when rolling 1 year Sharpe ratios are assessed that no single asset class has consistently recorded the highest or lowest Sharpe ratio at every single period of time historically. For example, the bond index recorded highest Sharpe ratios in the earlier periods 1998-2002 but struggled in 2003, while currency experienced poor Sharpe Ratios in 1998-2002 and recorded good Sharpe ratios in 2003. More recently in the 2012-2018 period the industrial and financial indices have recorded better Sharpe ratios while the resource and currency indices have recorded poorer Sharpe ratios relative to the other asset classes.

Correlation Matrix	Resources	Industrials	Financial	Bond	Currency
Resources	1.00	0.52	0.38	0.09	0.04
Industrials	0.52	1.00	0.75	0.31	0.19
Financial	0.38	0.75	1.00	0.56	0.34
Bond	0.09	0.31	0.56	1.00	0.44
Currency	0.04	0.19	0.34	0.44	1.00

Figure 13: Correlation Matrix: A full sample correlation matrix for the industrial, financial, resource, bonds and currency asset classes over the period February 1999 until Dec 2017.

Additionally, we observe the correlation structure of the various asset classes over the entire analysis period in Figure 13. Correlation is an important statistic to observe as it is often used to assess the diversification benefit from the various asset classes and is an essential metric in deriving the optimal asset allocation. In the long term the financial and industrial indices have the highest positive correlation of 0.75, this is understandable as these two sector indices do fall part of broader equity asset class. The bond and financial indices also have a high positive correlation of 0.56, both of these asset classes have interest rate dependencies so this relationship is understandable. The resource index has a low correlation to bonds and currency. The lowest correlation of 0.04 is between the bonds and currency indices. Currency has in general the lowest average correlation to all other asset classes over the analysis period. It should be noted that these are long term correlation statistics and as demonstrated by Chin in [39] the correlations between asset classes tend to have time varying differences over time.

8.1.2 Data on State Variables

In this section we list the state signals, these state signals are processed to form state variables. We use 10 state variables, these range from macroeconomic, risk and trend based categories. We explore the various state variables based on grouping them into two broad categories namely 1) macroeconomic and 2) financial, risk and other indicators. The macroeconomic indicators consist of growth, inflation, liquidity and exchange rate indicators. The financial indicators consist of risk, trend as well as other useful indicators. The choice of these variables were largely selected based on : the existence of academic literature supporting the usage of these variables, if these variables were commonly used by investment practitioners and if historical data was readily available for each of these state variables for the specified analysis period.

State Signal	Freq.	Code(s) used
SA Real Gross Domestic Product	Quarterly	SAGDPANN Index
SA Domestic Leading Indicator	Monthly	OEZAKLAP Index
SA Consumer Price Index	Monthly	SACPIYOY Index
SA Money Supply	Monthly	SAMYM3Y Index
SA Equity Momentum	Monthly	JALSH Index
JP Morgan EMBI Index	Monthly	JPEIGLBL Index
USD ZAR Currency	Monthly	USDZAR Curncy
USA PMI	Monthly	NAPMPMI Index
USA VIX Index	Monthly	VIX Index
Equity Risk Premium (USA-SA)	Monthly	SPX Index, US0003M Index, JALSH Index, JIBA3M Index

Table 3: State Variable Signals: The list of state variables used as investment signals, along with their reporting frequency and Bloomberg codes are provided; these include variables from both domestic and global econometric candidate signals and cover macroeconomic, risk and trend based variables. The list is indicative and not exhaustive. The sampling frequencies are Monthly (M) and quarterly (Q).

All state signal data was sourced from Bloomberg. The Bloomberg codes listed above were used to pull data for state signal 1-4 and 6-9. For state signal 5, the 12 month rolling return of the FTSE JSE All Share Index was used as a proxy for SA Equity Momentum. For state signal 10, the differential between the (S&P 500 less 3 month LIBOR) and (FTSE JSE All Share Index less 3 month JIBAR) was used as a proxy of equity risk premium differential. All state signals except for GDP have monthly frequency hence cubic-spline interpolator was used to convert quarterly GDP data into monthly data. In order to account for the delay in receiving data we lag: GDP data by 3 months, the OECD Domestic Leading Indicator by 6 months, CPI by 1 month and Money Supply by 1 month. It should be noted that the paper uses a couple of state variables sourced from the US or relative to the US i.e. the US PMI, US VIX, Equity Risk Premium (ERP) differential and Currency relative to the US. This is because the US is a major driver of equilibrium relations in eleven emerging Asia-Pacific stock markets [53] which suggests the leading role that the United States plays in other emerging market countries and potentially South Africa.

8.1.2.1 Macroeconomic State Variables

We cover a couple of indicators that are related to economic growth such as Gross Domestic Product (GDP), Domestic Leading Indicator (DLI) and Purchasing Manager Index (PMI). In academic literature such as [122] and [65] GDP and Inflation are commonly used together to understand the economic state of a country. Investment practitioners such as Munro and Silberman [147] use these two variables together to classify the current regime of the South African economy. The SARB uses the long run trend of GDP to identify the business cycle [196], and [100] examines the effect of macroeconomic variables including GDP on the South African equity market and found South Africa's equity market index is positively related to the change of the South African real GDP. The paper also considers the South African OECD Domestic Leading Indicator, which is widely used by practitioners and aims to give a sense of future GDP growth by approximately 6 months, Dreger and Schumacher [59] explain that the composite leading indicators have significant in-sample relationship in describing the course of the economy. Finally, the US ISM Manufacturing PMI Composite Index is a survey-based indicator that is widely followed by investors; this indicator has been investigated by many [148, 54, 46] and could have some return predictability.

We also cover other important macroeconomic variables such as Consumer Price Index (CPI), Money Supply and USD ZAR Exchange rate (USDZAR Currency), as these variables are also essential in understanding the state of the economy. Inflation has been cited by both local and global research to have asset class return predictability. Locally, Gupta and Modise [86] examined the effect of macroeconomic variables including inflation on the South African equity market and found that the South African inflation rate can strongly forecast the out-of-sample South African equity market returns by 6 months. In the United States Boucher [29] explains that inflation has substantial out-of-sample forecasting abilities for real US stock returns. Money supply has been shown to display out-of-sample equity return predictability [58], locally Gupta and Modise [86] shows that money supply has in-sample predictive power in the short run. The relationship between stock returns and exchange rates was analyzed by Wu in [209] who showed that a relationship exists between the stock returns and exchange rates due to the real interest rate disturbance and inflationary disturbance.

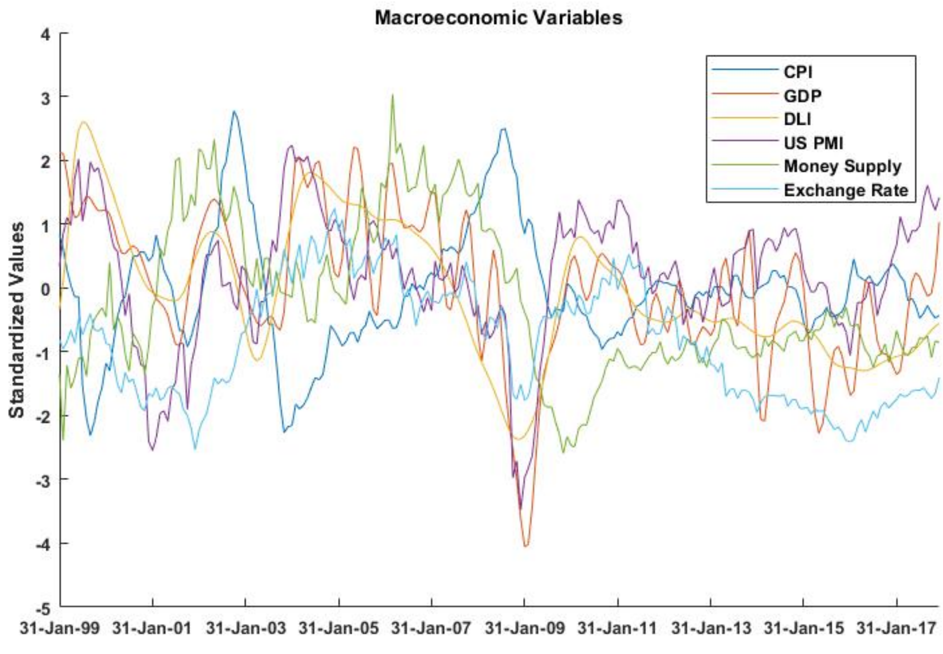

Figure 14: Standardized values for macroeconomic state variables over the period 31 Jan 1999 to 31 Dec 2017.

In Figure 14 we show the standardized values for the macroeconomic state variables used in this paper. It can be seen that some of these variables share common dynamics. The GDP, DLI and US PMI indicators are the most correlated over the observed period and CPI has the lowest correlation to all other macroeconomic indicators with the exception of money supply. Majority of the macroeconomic indicators signal stress in crisis period as seen by the weakening of standardized scores during and prior to the global financial crisis.

8.1.2.2 Financial, Risk and other State Variables

In the financial, risk and other state variables category we observe the Volatility Implied Index (VIX), SA Equity Momentum (SAM), Equity Risk Premium (ERP) and JP Morgan EMBI Index (JPMEMBI). Qadan et al. in [159] show that the VIX plays a role in the relationship between idiosyncratic volatility and stock returns such that when VIX decreases there is a decrease in risk aversion resulting in improved stock returns. There is significant predictability from past returns and momentum investments earn abnormal returns as documented in [101] and [191]. Jostova [110] showed that emerging market spreads could be used to predict emerging sovereign debt market returns. In [161] the authors show that lagged U.S. returns significantly predict returns in numerous non-U.S. countries substantiating the usage of equity risk premium differential relative to the US and showcasing the leading role that United States plays in return predictability of other countries.

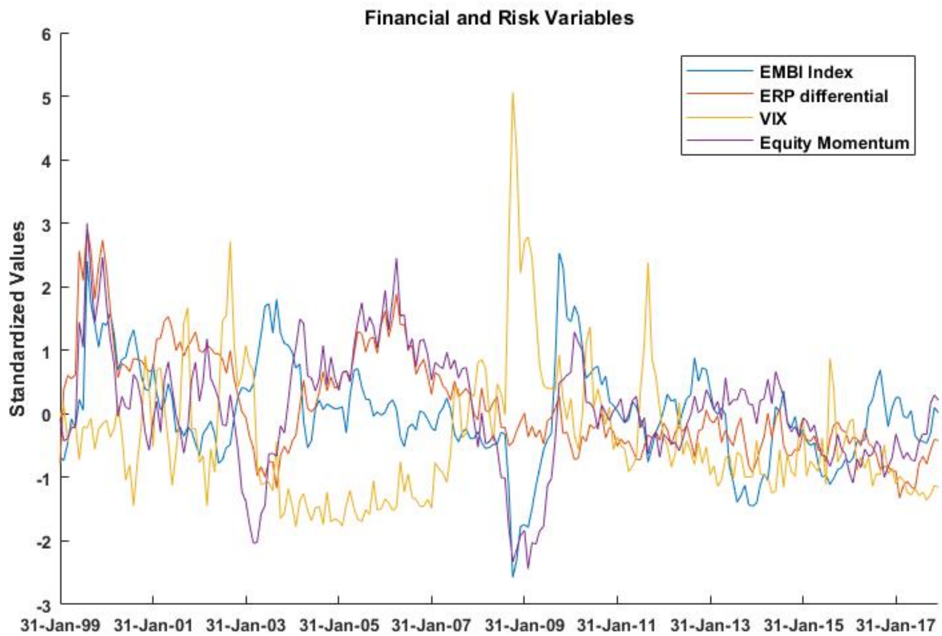

Figure 15: Standardized values for financial, risk and other state variables over the period 31 Jan 1999 to 31 Dec 2017.

In Figure 15 we show the standardized values for the financial, risk and other state variables used in this paper. We can see that many of these variables share common dynamics. The ERP and Momentum indicators are the most correlated over the observed period which is understandable considering the equity commonality of the two indicators. The VIX and equity momentum indicators have the lowest correlation amongst these state variables. During the GFC we see the spike in VIX corresponding to the increased volatility and uncertainty surrounding this period, the EMBI and equity momentum indicators see severe weakening during this GFC period signaling the stress signaled by market variables in crisis periods.

8.2 Methodology

In this section we detail the steps required to convert the data in the previous subsection into the required format for the HS-FP model. We discuss how the HS-FP model is used to forecast asset class return and risk, how we select the optimal portfolio and finally how the backtesting is setup to calculate the out-of-sample returns.

8.2.1 Data Preparation

First, we need to determine the historical time series of invariants $\boldsymbol{e}_t$. The asset class index levels in Table 1 are retrieved and the logarithmic return series $\boldsymbol{r}_t$ is calculated. The historical time series of invariants $\boldsymbol{e}_t$ is estimated by $\boldsymbol{r}_t$ as seen below:

$$\{\boldsymbol{\varepsilon}_{\bar{n},\bar{t}} = \boldsymbol{r}_{\bar{n},\bar{t}} = \ln\big(\boldsymbol{v}_{\bar{n},\bar{t}}/\boldsymbol{v}_{\bar{n},\bar{t}-1}\big)\}_{t=1}^{\bar{t}}. \tag{44}$$

Second, we need to determine the historical time series of state variables $\boldsymbol{z}_{m,t}$. In Table 3 we provide a list of m state signals $\{\boldsymbol{S}_{m,t} \equiv (S_{1,t}, \ldots, S_{m,t})\}_{t=1}^{t=\bar{t}}$ used in this paper. State signals are typically noisy so we need to smooth and score them to obtain a historical time series of state variable $\{\boldsymbol{z}_{m,t} \equiv (z_{1,t}, \ldots, z_{m,t})\}_{t=1}^{t=\bar{t}}$ that we can compare and combine easily.

Starting with signal time series $\boldsymbol{S}_{m,t}$ we first smooth this series to obtain $\boldsymbol{z}_{m,t}^{\text{smooth}}$ by calculating the exponential weighted moving average where we set the fast half life and slow half life parameters at 3 and 12 months respectively in Section 9.

$$\boldsymbol{z}_{m,t}^{\text{smooth}} = \text{smooth}(\boldsymbol{S}_{m,t}) \equiv \text{EWMA}_w^{\tau_{HL}}(\boldsymbol{S}_{m,t}). \tag{45}$$

Then in order to compare how this smoothed signal has evolved over time, we standardize the state signal by calculating the z-score of $\boldsymbol{z}_{m,t}^{\text{smooth}}$. This then forms the state variable time series $\boldsymbol{z}_{m,t}$ required in HS-FP model.

$$\boldsymbol{z}_{m,t} = \text{score}(\boldsymbol{z}_{m,t}^{\text{smooth}}) \equiv \frac{\boldsymbol{z}_{m,t}^{\text{smooth}} - \mu(\boldsymbol{z}_{m,t}^{\text{smooth}})}{\sigma(\boldsymbol{z}_{m,t}^{smooth})}. \tag{46}$$

In Section 9 of the paper we use monthly data spanning approximately 19 years from which we use 5 years as an initial training window which grows with the addition of new information each subsequent month. The target state variable z_t^* at each monthly update of the model is set as the most recent value of the state variable. The leeway parameter which is the +/- range from the current value of the state variable is set at 0.1 throughout Section 9.

8.2.2 Forecasting Return and Covariance

We show how to use the HS-FP model to forecast the expected return $\mathbb{E}[\boldsymbol{R}]$ and covariance $\mathbb{C}\text{ov}[\boldsymbol{R}]$ of asset classes listed in Table 1.

The HS-FP model assigns probabilities to historical returns over the growing window each month conditioned by time and either single or multiple state variables. This flexible probability adjusted return and covariance then forms each asset class predicted mean and covariance for the next period's return distribution.

At time t having obtained $\boldsymbol{r_t}$, $\boldsymbol{z}_t$ and specified the target state variable as z_t^*, we then use the HS-FP model detailed in Section 6 to obtain the time and state conditioned flexible probabilities $\boldsymbol{p}_t^{\text{HFP}}$ for each asset class. Here the estimated $\mathbb{E}[\boldsymbol{R}]$ and covariance $\mathbb{C}\text{ov}[\boldsymbol{R}]$ for each asset class is given by:

$$\begin{aligned} \mathbb{E}[\boldsymbol{R}] &= \sum_{t=1}^{\bar{t}} \boldsymbol{p}_t^{\text{HFP}} \boldsymbol{r}_{\bar{t}}, \quad \text{and} \\ \mathbb{C}\text{ov}[\boldsymbol{R}] &= \sum_{t=1}^{\bar{t}} \boldsymbol{p}_t^{\text{HFP}} \boldsymbol{r}_t \boldsymbol{r}_t' - \left[\sum_{t=1}^{\bar{t}} \boldsymbol{p}_t^{\text{HFP}} \boldsymbol{r}_t\right] \left[\sum_{t=1}^{\bar{t}} \boldsymbol{p}_t^{\text{HFP}} \boldsymbol{r}_t\right]'. \end{aligned} \tag{47}$$

8.2.3 Portfolio Optimization and Backtesting

Having obtained the forward looking distribution for the next period's asset class return distribution we then need to select the optimal asset allocation. We use the Markowitz framework [135] and select the maximum Sharpe ratio [183] as the optimal portfolio. The only constraints used in the portfolio optimization step is that the portfolio weight needs to be non-negative and to sum to 1 (long only constraint). No upper or lower bounds or group bounds were specified for asset class weights and no turnover and tracking error constraints were specified.

In this paper we use monthly data spanning approximately 19 years from which we use 5 years i.e. Feb 1998 to Feb 2003 as an initial training window which grows with the addition of new information each subsequent month and hence the out-of-sample returns span from March 2003 to Dec 2017 which is about 14 years. Estimates are updated monthly upon the arrival of new data. The forecast horizon is fixed at one month and the rebalancing frequency is half yearly.

The backtesting process is as follows: the initial training window is used to estimate the first $\mathbb{E}[R]$ and $\mathbb{C}\text{ov}[R]$ for each asset class, then the efficient frontier is plotted and we select the optimal portfolio as the portfolio from the efficient frontier with maximum Sharpe ratio, the optimal portfolio is fixed and tested on the out-of-sample returns and after 6 months the process is repeated.

9 Results

In Meucci [142, 144, 145] the theoretical framework of the HS-FP model is introduced and a simple illustration of the implementation of the HS-FP model is showcased. The papers do not explicitly compare the empirical results of the HS-FP model to potential alternative benchmark models by observing performance on an in-sample or out-of-sample format, nor does it suggest a set of potential state variables that could be used to condition asset class returns. One of the key contributions of this research is to provide insight into the empirical results of the HS-FP model in comparison to alternative benchmark models, introduce a set of state variables that could be used, and assess whether the HS-FP model can be useful in conditioning asset class returns and deriving time varying results.

In order to understand the results of the HS-FP model we report on a number of summary statistics, observe the optimal portfolio weights over time and plot cumulative and relative rolling return graphs. The analysis of the results is based on an out-of-sample performance analysis, meaning we evaluate the performance of the HS-FP and benchmark models on a period of data which is different from the period of data we used to identify the optimal asset allocation. The statistics we report are popular in the investment industry and consist of the annualized gross returns, annualized volatility, Sharpe ratios, maximum draw down, average monthly turnover and conditional value-at-risk. We will discuss some of these statistics further in Section 9.1.

In Section 9.2 we observe the results when conditioning by each of the 10 state variables individually. In Section 9.3 we then combine all the state variables and compare the results of the different combination methodologies namely: the equal weighting (EQ) and weighting by Degree of Conditioning and Correlation (DCC). In Section 9.4 we select and assign the best performing strategy amongst the different combination methodologies as the HS-FP model going forward. Finally, in Section 9.4, the performance of the HS-FP model is compared to benchmark models namely the classic MVO and EW. The risk free rate is 7.25% over the analysis period. The out-of-sample returns in Section 9 assumes zero transaction costs and makes no adjustment for indirect costs related to portfolio rebalancing.

9.1 Performance Measures

To evaluate the performance of the various strategies we assess the annualised gross return, different measures of annualised risk namely: volatility, maximum draw down and conditional value-at-risk, we also assess the risk adjusted returns based on the popular Sharpe ratio and finally compare the average monthly turnover. These statistics are widely used in the investment industry and provide a general but not exhaustive insight into the different dimensions of the results. In this section we only observe gross returns however in Section 10.1 we also assess the net returns of the HS-FP model relative to benchmark models at varied transaction costs.

In addition to the annualised volatility we assess two other risk metrics namely the Maximum Draw Down (MDD) and Conditional Value at Risk (CVaR). The MDD adds an important perspective to risk assessment as it does not focus on the smoothness of the returns like the volatility measure but rather focuses on the capital preservation of the strategy. The MDD is widely studied in numerous academic literature and is a popular statistic utilised in the investment industry [40]. The MDD calculates the maximum peak-to-trough percentage decline of a strategy over a specified period of time. This measure can be applied on absolute returns or benchmark relative returns. Here the measure is applied on absolute returns. The formula for the maximum draw down is:

$$\text{MDD}(S_T) = \frac{(P - L)}{L}, \tag{48}$$

where P is the highest or peak value of the strategy before its biggest drop and L is the lowest or trough value of the strategy before a new high is achieved.

The CVaR measures the risk over a specified time horizon based on the expected losses of a strategy in an extreme situation. It is derived from the VaR measure which calculates the level of worst-case loss but unlike the VaR , the CVaR calculates the average of losses or expected losses when the loss exceeds a specified tolerance level. The calculation is performed by focusing on the exceeded tail portion of the strategy distribution. The formula for conditional value at risk is:

$$\text{CVaR} = \frac{1}{1 - tol} \int_{-1}^{VaR} rp(r)dr, \tag{49}$$

where $p(r)dr$ is the probability density of obtaining a return with a value r, tol is the cut-off tolerance level and VaR is the specified VaR level. The CVaR is also referred to as the expected shortfall. The CVaR was introduced by Rockafellar and Uryasev [164] and is a vital risk measure that rose in popularity especially post market extremes such as the global financial crisis, where it was used by investors to understand, measure and allocate risk in their portfolios.

The Sharpe Ratio (SR) is a risk adjusted return measure, that was introduced by Sharpe in [183]. The measure enforced the importance of observing the merits of a strategy not purely on returns but also adjusting for the risk associated with the strategy. The SR is a popular statistic in the investment industry, widely used in academic literature [177] and the portfolio that satisfies the maximum SR criterion is often selected as the optimal asset allocation. The formula for the Sharpe ratio is defined as:

$$\text{SR} = \frac{(\mu_p - r_f)}{\sigma_p}, \tag{50}$$

where μ_p is the average return of the strategy, r_f is the risk free rate of return and σ_p is the volatility or standard deviation of the strategy's return. The SR disclosed in this section is a point estimate of the true Sharpe ratio. This can naturally vary in different testing periods and out of sample periods. In Section 10 we calculate the probabilistic sharpe ratio and deflated sharpe ratio in order to assess the robustness of the point estimate Sharpe ratio.

The turnover simplistically measures how much a portfolio has changed over a period of time. The turnover ratio is calculated by dividing the total purchases or sales by the total market value of the portfolio. The formula for turnover is:

$$\text{Turnover} = \sum_{i=1}^{N} \left| w_{i,t} - w_{i,t^-} \right|, \tag{51}$$

here we measure Turnover as the difference in weights of a portfolio at the beginning of the month ($w_{i,t}$) and the end of the month w_{i,t^-}. It is important to assess the impact of transaction costs on the performance of the backtested strategies to assess the applicability of the strategy in reality. In this section we observe the gross returns of the strategies, however the net returns are also assessed in Section 10.1. The net return nr_t for the strategy is given by the formula:

$$1 + nr_t = (1 + r_t)\left(1 - c \times \text{Turnover}\right), \tag{52}$$

where c represents the transaction costs as a proportion of the turnover. The transaction costs applied in Section 10.1 is 0, 20, 30 and 50 bps.

9.2 Comparing Results of Single State Variables

In this subsection we report the results where each of the state variables are used individually to forecast all asset classes next period return distribution. Looking at these performance, risk, risk adjusted return and turnover statistics in Table 4 we see varied results. When annualised gross returns are observed the best performers are the: SA Inflation, US PMI and ERP differential state variables. In terms of the risk statistics: US VIX, EMBI and ERP differential state variables have the lowest annualised volatility, whilst VIX, USDZAR currency and ERP differential state variables have the lowest MDD and VIX, ERP differential and US PMI state variables have the lowest CVaR.

In terms of turnover, the VIX, GDP and Momentum state variables have the lowest average monthly turnover. Finally in terms of risk adjusted returns: US PMI, ERP differential and SA Inflation are the state variables that have the highest Sharpe ratio. Different state variables do well in the different summary statistic categories. There is no single state variable that outperforms all other state variables in every category. This suggests the benefit of using multiple state variables to condition asset class return distributions. This can help reduce the noise of individual state variables and potentially mitigate the risk of over-fitting and model misspecification and create a more diversified approach of incorporating information from multiple sources.

Although there are varied results in terms of best performing state variable in the different summary statistic. There are some interesting insights. From a risk perspective, the US VIX and ERP differential are the only state variables that perform well across all the different risk measures. Additionally, the ERP differential is the only state variable that ranks in the top three position across all the risk, return and risk adjusted return categories. In terms of poor performers, Momentum is the only state variable that ranks in the bottom three across all the return, risk and risk adjusted return categories and Money Supply also performs poorly from a risk perspective, ranking in the bottom three across the different risk measures.

State Variables	Ann. Return	Ann. Volatility	Sharpe Ratio	Max. Drawdown	CVaR	Turnover
SA Inflation	9.19	10.25	0.18	23.78	7.64	6.64
SA GDP	6.03	8.39	-0.14	28.48	8.17	4.63
SA DLI	8.33	8.54	0.12	15.06	6.73	6.30
EMBI	7.03	6.84	-0.03	10.54	4.75	6.42
USDZAR Currency	8.58	7.47	0.17	9.53	4.55	5.32
ERP differential	8.87	6.88	0.23	10.07	4.01	5.01
US PMI	9.03	7.05	0.25	12.53	4.04	5.86
VIX	7.96	6.66	0.10	7.44	3.47	4.07
Money Supply	7.92	10.05	0.06	28.48	8.21	5.43
Momentum	5.73	9.37	-0.16	29.46	9.21	4.88

Table 4: Single State Variable Summary Statistics I: The state variables listed in Table 3 are used individually as the only state variable to condition the HS-FP model; popular summary statistics such as gross annualised return, annualised volatility, annualised Sharpe ratio, maximum draw down, CVaR and turnover (see Sec 9.1 for details regarding the calculation of these ratios) are shown for the March 2003 - Dec 2017 out of sample period.

9.3 Comparing Results of Multiple State Variables

In this subsection we report the results where all state variables mentioned in the previous sections are used to forecast all asset classes next period return distribution. We allocate to these multiple state variables based on two combination methods EQ and DCC.

Combination method	Ann. Return	Ann. Volatility	Sharpe Ratio	Max. Drawdown	CVaR	Turnover
EQ	8.26%	6.08%	0.16	5.80%	3.58%	2.60
DCC	11.16%	12.12%	0.32	21.20%	9.84%	10.93

Table 5: Multiple State Variable Summary Statistics I: The state variables listed in Table 3 are used in combination to condition the HS-FP model; two methods to combine information from each state variable are assessed. Equal weighting (EQ) (see Section 6.2.1) where the conditioned probability from each state variable is equally weighted to give final probability and DCC (see Section 6.2.2) where the conditioned probability from each state variable is weighted by Degree of Conditioning and Correlation (DCC) between the different state variables. Popular summary statistics such as gross annualised return, annualised volatility, annualised Sharpe ratio, maximum draw down, CVaR and turnover are shown for the March 2003 - Dec 2017 out of sample period.

Firstly, we note that based on the results in Table 5, combining multiple state variables improves return, risk, risk adjusted return and turnover statistics in comparison to single state variable approach as seen in Table 4. For example, the SA Inflation state variable records the highest annualised gross return 9.19% from the single state variable but the DCC approach of combining multiple state variables records a higher 11.16%. Looking at risk and turnover in Table 5 the EQ approach has a lower volatility and turnover than any single state variable. Finally, when risk adjusted returns are observed, the Sharpe ratio of DCC is higher than any single state variable. These results highlight the benefit of using multiple state variables for improved diversification and returns.

When comparing the combination methodologies, the best performing combination method in terms of annualised gross returns is weighting by DCC. However EQ has better risk and turnover statistics. Finally in terms of risk adjusted returns, the Sharpe ratio is the highest for the combination method DCC. Due to this, we refer to time and conditioning by multiple state variables where the weighting to multiple state variables are derived by DCC as the HS-FP model going forward.

9.4 Comparing Out-of-Sample HS-FP Results to Benchmark Models

In this section we compare results of HS-FP to benchmark models. We compare allocations over time, summary performance, risk and turnover statistics, as well as rolling return profiles to observe if there are consistent differences between HS-FP approach and the benchmark models.

9.4.1 Comparison of Allocations over Time

Comparing the asset allocations of HS-FP model to the classic MVO and EW benchmark models over time, the following insights are observed. The HS-FP and naturally EW approach does allocate at least once to all the asset classes, whilst the classic MVO never allocates to financial and currency asset classes. The least favorite asset classes on average for the HS-FP approach are also the financial and currency asset classes. Hence both HS-FP and classic MVO are underweight financial and currency asset classes relative to EW approach.

The fixed income position is the largest allocation for both HS-FP and classic MVO on average over time. The classic MVO has a considerably larger allocation to fixed income on average than HS-FP. Hence HS-FP and classic MVO is overweight fixed income relative to EW approach. During the Global Financial Crisis (2007-2008) the HS-FP approach had a greater overweight to fixed income and industrial asset classes and greater underweight to financial and currency asset classes relative to EW approach. Meanwhile the HS-FP approach was overweight industrial and underweight fixed income (despite this asset class being HS-FP's biggest average allocation during this period) relative to classic MVO.

Currently as at Dec 2017, the HS-FP model holds no position in resources, currency and fixed income asset classes and hence is overweight financial and underweight fixed income relative to classic MVO, and is overweight financial and industrial asset classes relative to EW approach. The HS-FP model has a more erratic asset allocation profile in comparison to the classic MVO and naturally EW models which is not unexpected considering this approach doesn't follow a historical expected mean or a heuristic approach to asset allocation, and asset class distributions are conditioned by time and state variables.

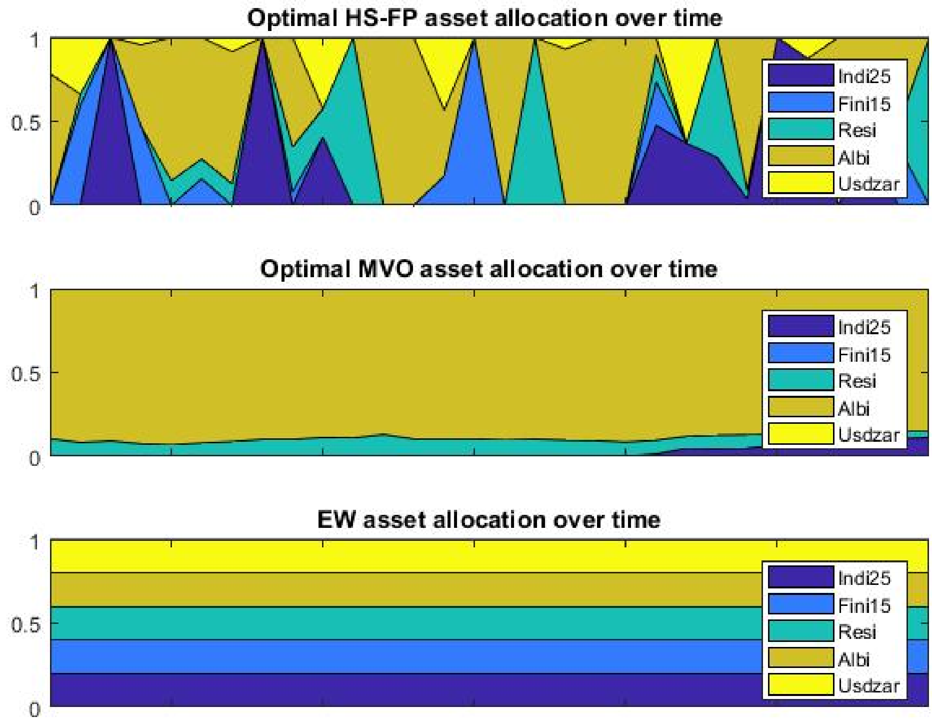

Figure 16: Optimal Asset Allocation: The out of sample optimal allocation to each asset class over time is shown for the three portfolio construction methods considered; the Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimization (MVO) (see Section 7.2) and Equal Weighting (EW) (see Section 7.1) respectively.

9.4.2 Comparison of Out-of-Sample Statistics

In this subsection we compare the out-of-sample performance, risk, risk adjusted return and turnover statistics of HS-FP model to the benchmark models.

Approaches	Ann. Return	Ann. Volatility	Sharpe Ratio	Max. Drawdown	CVaR	Turnover
HS-FP	11.16%	12.12%	0.32	21.20%	9.84%	10.93
MVO	8.27%	6.26%	0.16	10.63%	4.01%	0.14
EW	8.13%	10.91%	0.08	34.82%	7.98%	0.57

Table 6: Asset Allocation Models Summary Statistics I: Popular summary statistics for the three portfolio construction methodologies covered in this paper are considered, Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimal (MVO) (see Section 7.2) and Equally Weighted (EW) (see Section 7.1) respectively. The gross annualised return, annualised volatility, annualised Sharpe ratio, maximum draw down, CVaR and turnover summary statistics are provided for the out of sample period of March 2003 until December 2017.

The best performing allocation model is the HS-FP approach in terms of annualised gross returns. However, the HS-FP model performs worse overall in terms of risk, it has higher volatility and CVaR than both benchmark models and higher maximum draw down than classic MVO. It also has a higher turnover than both benchmark models which is not surprising considering its erratic optimal asset allocation as shown in Figure 16. Finally in terms of risk adjusted returns the Sharpe ratio is the highest for the HS-FP approach and is 2 and 4 times greater than the Sharpe ratio of classic MVO and EW respectively. From these results and based on the parameter assumptions it could be deduced that over the entire analysis period the HS-FP model has better risk adjusted returns than the benchmark models.

9.4.3 Comparison of Out-of-Sample Rolling Returns

In Subsection 9.4 we see HS-FP model has better risk adjusted returns than the benchmark models. This was based on summary stats over the entire period, there could be periods of time historically where the HS-FP model significantly under/over performed the benchmark models. In this section we look closer into the rolling returns of the HS-FP and benchmark models to see whether there are notable periods with significant deviations from the long term experience. We do this by observing cumulative returns (see Figure 17) and annualised returns (see Figure 18) relative to benchmark models.

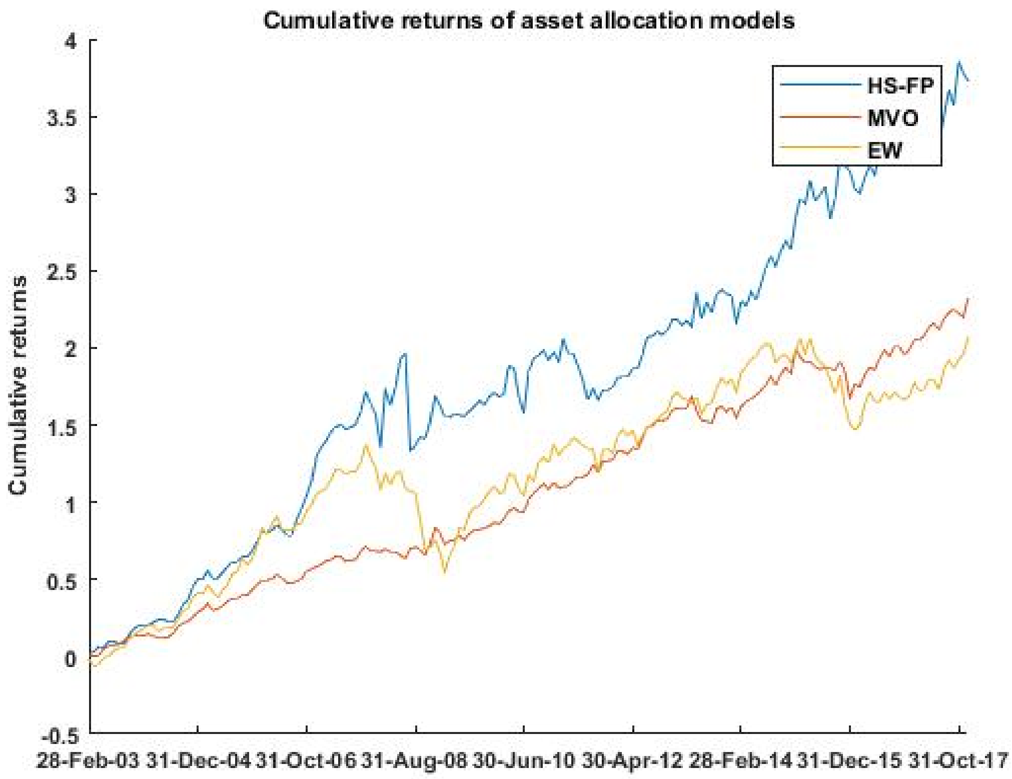

Figure 17: Cumulative Return for Asset Allocation Models: A comparison of the out of sample cumulative returns for the Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimal (MVO) (see Section 7.2) and Equally Weighted (EW) (see Section 7.1) asset allocation models.

In Figure 17 the out of sample cumulative returns of the HS-FP model is plotted alongside the benchmark models over the entire out of sample period. Based on the cumulative graph it appears that on a return basis the HS-FP model is far superior to both the classic MVO and EW benchmark models. The EW benchmark model performs the worst on a cumulative basis over the out of sample analysis period. It is important to note that the series of cumulative returns displays the average out of sample returns and could be prone to few periods of strong performance based purely on luck rather than the skill of the strategy. Hence, we also look at the rolling 1 year relative returns of the HS-FP model relative to MVO and EW benchmark model respectively as in Figure 18.

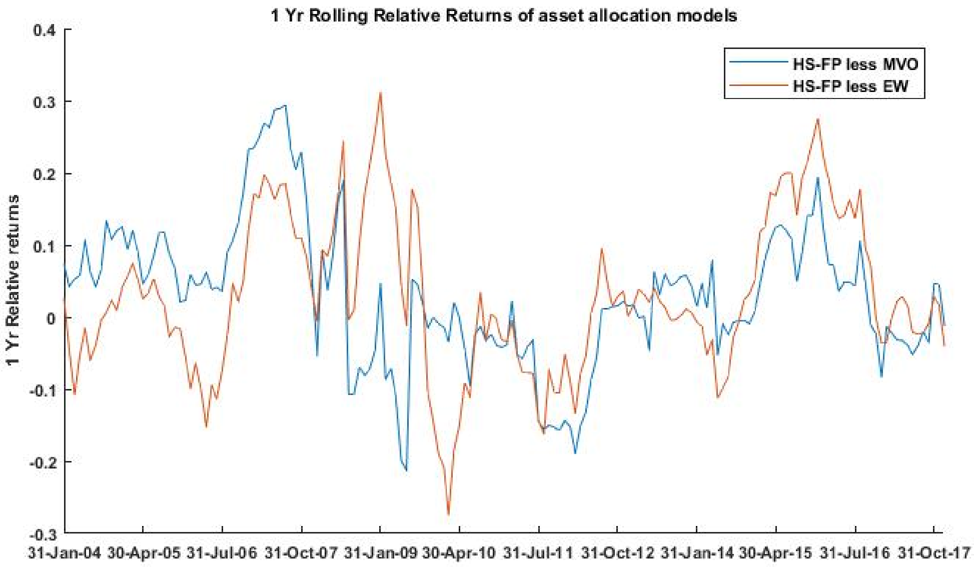

Figure 18: Rolling Relative Returns for Asset Allocation Models: A comparison of the 1 Year out of sample relative returns for the Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimal (MVO) (see Section 7.2) and Equally Weighted (EW) (see Section 7.1) asset allocation models.

When the relative return graph is considered, we see periods where the HS-FP model significantly under and outperformed benchmark models. When performance of HS-FP model is compared to classic MVO benchmark model we see that from 2003-2008 HS-FP outperformed the benchmark, then from 2008-2012 HS-FP struggled against benchmark and bounced back up recently from 2012-2017. In comparison to EW benchmark model, HS-FP model struggled initially from 2003-2006, outperformed from 2006-2009, then underperformed for the next 3 years and strongly outperformed from 2014 till recently.

During the credit crisis (2007-2008), the HS-FP model recorded an annualised return of 8.1% outperforming both the classic MVO and EW benchmark models which recorded 7.7% and −7.5% respectively. The major contributor of this out-performance was the HS-FP model having a greater overweight to asset classes that performed better during the Global Financial Crisis (GFC) in comparison to the benchmark models. In terms of risk, the HS-FP model was more volatile during this period recording an annualised volatility of 23% while classic MVO and EW recorded lower volatilities of

7% and 14% respectively.

The HS-FP model had the highest Sharpe ratio during the GFC, recording an annualised risk adjusted return of -0.08 while classic MVO and EW models recorded lower Sharpe ratios of -0.32 and -1.28 respectively. The risk free rate during this period was 10%. This relative performance suggests the cyclical nature of HS-FP model and flags the potential of backtest overfitting due to starting period assumption and hence under estimation of out-of-sample generalization errors.

9.5 Comparison with Multi-Asset Funds

It is also important to consider how the HS-FP and benchmark models performed in comparison to South African balanced funds. We consider the average performance of balanced funds from the ASISA South African (SA) Multi Asset(MA) Flexible category and the ASISA Global Multi Asset(MA) Flexible category over the past 5 years (Jan 2013-Dec 2017). The risk free rate over this period was 6.35%.

Approaches	Ann. Return	Ann. Volatility	Sharpe Ratio
HS-FP	8.23%	9.98%	0.19
MVO	5.07%	7.04%	-0.18
EW	2.72%	9.09%	-0.40
SA MA Flexible	10.16%	6.72%	0.57
Global MA Flexible	14.64%	12.92%	0.54

Table 7: Summary Statistics for Asset Allocation Models relative to Balanced Funds: Popular summary statistics such as annualised return, annualised volatility and annualised Sharpe ratio for the Historical Simulation with Flexible Probabilities (HS-FP) (see Section 6), Mean Variance Optimal (MVO) (see Section 7.2) and Equally Weighted (EW) (see Section 7.1) asset allocation models are compared to South African and Global Multi-Asset Flexible Category funds over the period Jan 2013-Dec 2017.

In Table 7 we observe that over the past 5 years, the HS-FP model recorded an annualised performance of 8.23% and underperformed both the average of the SA MA Flexible category and Global MA Flexible category which recorded 10.16% and 14.64% respectively. In terms of risk, the volatility of HS-FP model is higher than that of the SA MA Flexible category but lower than that of the Global MA Flexible category. Finally in terms of the Sharpe ratio, the HS-FP has a lower Sharpe ratio than both SA MA Flexible category and Global MA Flexible category.

A potential reason for this underperformance could be the lack of offshore asset classes in the asset class universe for the HS-FP and benchmark models. This suggests that there could have been a benefit to performance, risk and diversification for the HS-FP model and benchmark models from including offshore assets. Over the past 5 years, the HS-FP model maintains its outperformance relative to the classic MVO and EW benchmark models.

10 Robustness of Out-of-Sample Results

The results discussed so far in Section 9 is one potential historical path based on the initial assumption and parameter steps applied. In this section we analyse the robustness of the out-of-sample returns discussed in the previous section. We follow the backtesting protocol suggested by Arnott, Harvey and Markowitz in [3], discuss assumptions and parameter steps that may lead to overfitting and hence under estimating the out-of-sample generalization errors.

We consider measures that assess performance inflation due to non-normality of asset class returns and due to selection bias under multiple testing. We also assess the legitimacy and consistency of backtest result when assumptions are varied. We calculate the probabilistic sharpe ratio [10] and deflated sharpe ratio [11] as introduced by Bailey and Lopez. Additionally, we calculate and discuss the probability of backtest overfitting using the methodology suggested by Bailey et al [8] in this section.

10.1 Transaction Cost

The empirical results that were discussed in Section 9 are based on the assumption of zero transaction costs. In reality to practically implement an asset allocation strategy like this we would use ETF's or futures contracts. In order to trade these financial instruments there would be a transaction cost incurred. Hence, an analysis of the cost impact is essential in assessing the robustness of the HS-FP model's results. In this section we analyse the impact of varying the transaction cost assumption on the HS-FP models performance in comparison to the benchmark models.

In Table 8 we see that the conclusions drawn in Section 9 do not drastically change. The HS-FP model still has the highest Sharpe ratio relative to benchmark models even at a transaction cost set as high as 50 bps. The Sharpe ratio of the HS-FP model does however reduce from 0.32 in the case of 0 bps transaction cost to 0.26 in the case of 50 bps transaction cost. The HS-FP model maintains its higher annualised return relative to the benchmark models. The return does however reduce from 11.1% in the case of 0 bps transaction cost to 10.5% in the case of 50 bps transaction cost.

The HS-FP model maintains its relatively higher turnover and generally poorer risk statistics than the benchmark models. The turnover statistics do not change as the calculation is not impacted by the changes in the transaction cost applied. The risk statistics of the HS-FP model do increase slightly across the different risk measures. We see that when transaction cost is increased from 0 to 50 bps we see an increase from 12.1% to 12.2%, 21.2% to 21.7%, and 9.8% to 10.0% for the volatility, MDD and CVAR risk measures respectively.

From Table 8 we can summarize that when transaction costs are varied, we see that the HS-FP model maintains its higher annualised return, poorer risk statistics and higher Sharpe ratio relative to the benchmark models as seen previously in Section 9. It is important to note that the primary evaluation of whether the HS-FP model adds performance benefit relative to benchmark models used in this paper is based on achieving higher risk adjusted returns. Hence, we can conclude that when transaction costs are varied, the HS-FP model does successfully pass its primary objective which is to provide superior risk adjusted returns (as seen by the higher SR) relative to benchmark models.

Allocation Model	Ann. Return	Ann. Vol.	SR	MDD	CVaR	Turnover
Transaction cost = 0						
HS-FP	11.1%	12.1%	0.32	21.2%	9.8%	10.9%
MVO	8.2%	6.3%	0.16	10.6%	4.0%	0.1%
EW	8.1%	10.9%	0.08	34.8%	7.9%	0.5%
Transaction cost = 20 bps						
HS-FP	10.9%	12.1%	0.30	21.4%	9.9%	10.9%
MVO	8.2%	6.3%	0.16	10.6%	4.0%	0.1%
EW	8.1%	10.9%	0.08	34.8%	7.9%	0.5%
Transaction cost = 30 bps						
HS-FP	10.7%	12.1%	0.28	21.5%	10.0%	10.9%
MVO	8.2%	6.3%	0.16	10.6%	4.0%	0.1%
EW	8.1%	10.9%	0.07	34.8%	7.9%	0.5%
Transaction Cost = 50 bps						
HS-FP	10.5%	12.2%	0.26	21.7%	10.0%	10.9%
MVO	8.2%	6.3%	0.16	10.6%	4.0%	0.1%
EW	8.1%	10.9%	0.07	34.8%	7.9%	0.5%

Table 8: Summary Statistics for Transaction Cost = 0, 20, 30 and 50 bps: Summary statistics such as annualised return, annualised volatility, annualised Sharpe ratio, maximum drawdown, conditional value at risk and average monthly turnover are shown for the three asset allocation models (see Section 6, Section 7.2 and Section 7.1); when the transaction cost is set to 0, 20, 30 and 50 bps. The out of sample performance is observed over the March 2003 - Dec 2017 period. Here we see that when transaction costs are increased it does reduce the HS-FP's Sharpe ratio relative to benchmark models, however it remains the highest Sharpe ratio, even at a transaction cost as high as 50 bps.

10.2 Training Window

In Section 9 we used 5 years of data (Feb 1998 - Feb 2003) as the initial training period or window. We made this assumption so that we could sufficiently assess the out-of-sample performance of the HS-FP model during the Global Financial Crisis (GFC) market turmoil. The training window assumption has a crucial impact on the out-of-sample results obtained, as we would cross-validate the results by training the HS-FP model on a part of the data and then validating its results versus benchmark models on another part of the data. In this section we analyse the impact of varying the training window assumption on the HS-FP models performance in comparison to the benchmark models.

In Table 9 we do see changes in the conclusions drawn in Section 9 when the training window is varied. The HS-FP model does maintain a higher Sharpe ratio relative to benchmark models at training window = 8 and 10 years, however at a training window of 12 years it has a lower Sharpe ratio relative to MVO benchmark. The HS-FP model maintains a higher Sharpe ratio relative to EW benchmark at all the varied training window assumptions. Additionally, it is seen that the HS-FP model has the best relative SR in comparison to benchmark models when training window is set to 8 years, here HS-FP records 0.14 SR versus MVO -0.08 SR and EW -0.18 SR, and worst relative SR in comparison to benchmark models when training window is set to 12 years, here HS-FP records 0.14 SR versus MVO 0.22 SR and EW -0.06 SR.

At varied training windows the HS-FP model maintains its higher turnover and generally poorer risk statistics relative to benchmark models. The turnover for the HS-FP model is the highest for training window = 8 years and lowest for training window = 10 and 12 years. The risk statistics of the HS-FP model are generally the worst relative to benchmark models at training window of 8 years and best relative to benchmark models at training window of 12 years. The risk statistics for HS-FP model in isolation is the best at training window = 8 years and worst at training window = 12 years, specifically, the CVaR and volatility rises from 12.1% and 9.8% respectively at training window = 5 years to 13.2% and 11.1 % respectively at training window = 8 years; the CVaR and volatility falls from 12.1% and 9.8% respectively at training window = 5 years to 10.4% and 5.9 % at training window = 12 years.

From Table 9 we can summarize that when training window assumption is varied the HS-FP model can have poorer Sharpe ratios at specific training windows (training window = 12 years). Additionally, it maintains its poorer risk statistics in aggregate and higher turnover relative to the benchmark models as previously seen in Section sec:analysis. In terms of its primary objective of achieving higher SR relative to benchmark models, when training windows are varied, the HS-FP model does not successfully achieve this target at specific training window assumptions. Hence, this result indicates a potential reason to doubt the significance of the HS-FP's Sharpe ratio relative to benchmark models as seen in Section 9.

Allocation Model	Ann. Return	Ann. Vol.	SR	MDD	CVaR	Turnover
Training window = 5 years						
HS-FP	11.1%	12.1%	0.32	21.2%	9.8%	10.9%
MVO	8.2%	6.3%	0.16	10.6%	4.0%	0.1%
EW	8.1%	10.9%	0.08	34.8%	7.9%	0.5%
Training window = 8 years						
HS-FP	8.9%	13.0%	0.14	21.2%	10.9%	11.1%
MVO	6.9%	6.5%	-0.08	10.6%	4.1%	0.1%
EW	4.9%	11.0%	-0.18	34.8%	8.3%	0.5%
Training window = 10 years						
HS-FP	7.9%	13.2%	0.09	21.2%	11.1%	10.3%
MVO	7.1%	6.8%	0.07	10.6%	4.4%	0.1%
EW	4.5%	11.2%	-0.19	29.6%	8.6%	0.5%
Training window = 12 years						
HS-FP	7.6%	10.4%	0.14	13.0%	5.9%	10.3%
MVO	7.5%	6.6%	0.22	10.6%	4.4%	0.1%
EW	5.5%	9.7%	-0.06	19.5%	6.1%	0.5%

Table 9: Summary Statistics for Training Window = 5, 8, 10 and 12 years: Summary statistics such as annualised return, annualised volatility, annualised Sharpe ratio, maximum drawdown, conditional value at risk and average monthly turnover are shown for the three asset allocation models (see Section 6, Section 7.2 and Section 7.1); when the initial training window is set to 5, 8, 10 and 12 years. The out of sample returns are observed over the March 2003 - Dec 2017 period. Here we see that when training windows are altered it does impact the Sharpe ratio of HS-FP relative to benchmark models and when a training window of 12 years is used the HS-FP does underperform the MVO benchmark model.

10.3 Probabilistic Sharpe Ratio

The Probabilistic Sharpe Ratio (PSR) [8, 134] is an uncertainty-adjusted investment skill metric that measures the probability that a strategy exceeds a benchmark threshold in the presence of non-normal returns. We consider the PSR for the following reasons: to assess whether the HS-FP model consistently outperforms the MVO and EW benchmark models once corrected for performance inflation due to non-normality of asset class returns. The PSR is the probability that the true SR is above a given rejection threshold. The rejection threshold is set in this paper as the Sharpe ratio of the Benchmark models specified in Section 7. The PSR takes into account the sample length, skewness and kurtosis of the returns' distribution.

$$\widehat{PSR} \equiv \widehat{PSR}[\widehat{SR}_*] = Z\left[\frac{(\widehat{SR} - \widehat{SR}_*)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3 \widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right] \tag{53}$$

where:

T is the number of observed returns
$\hat{\gamma}_3$ is the skewness of the returns

$\hat{\gamma}_4$ is the kurtosis of the returns
$Z[.]$ is the CDF of the standard Normal distribution.

In Section 9, specifically Table 6, we see that the HS-FP point estimate Sharpe ratio is 2 times and 4 times greater than the classic MVO and EW benchmark models respectively. Based on this result it could be tempting to believe that the HS-FP model's Sharpe ratio is significantly greater than the benchmark models. However, when the PSR is calculated it is observed that the HS-FP model outperforms EW benchmark at 20 percent significance level and outperforms the MVO benchmark at 30 percent significance level as seen in Table 10. Additionally, at 5 percent significance level the HS-FP performance is not statistically significant relative to both benchmark model.

PSR Matrix	HS-FP	MVO	EW
HS-FP	0.50	0.72	0.81
MVO	0.27	0.50	0.62
EW	0.17	0.38	0.50

Table 10: PSR Matrix: The table represents the PSR matrix, the rows of the table represent the candidate strategy and the columns represent the benchmark strategy. The table should be interpreted as the probability that the candidate strategy SR exceeds the benchmark strategy SR.

Finally, when the Minimum Track Record Length (MTRL) is calculated to see what the appropriate track record length should be to reject null hypothesis that: the HS-FP's Sharpe ratio displays no skill beyond MVO and EW Sharpe ratio thresholds. At a 95 percent confidence level, we see that the MTRL is calculated as 1404 and 624 monthly observations or 117 and 52 years of track record is required to reject the null hypothesis for MVO and EW benchmark respectively. So the 179 months of out-of-sample performance we have currently, is too short and does not meet the MTRL required. Hence, we cannot reject the null hypothesis that HS-FP's Sharpe ratio is below the benchmark models threshold SR's at a 95% confidence level.

To summarize, we observe that the PSR implies that the Sharpe ratio of the HS-FP model is not statistically significant relative to benchmark models, additionally, the MTRL implies that the number of months or years of out-of-sample returns calculated in Section 9 is too short to claim significance. These results indicate reasons to doubt the significance of the HS-FP's Sharpe ratio relative to benchmark models as seen in Section 9. However, it is important to note that in this subsection, we assess the significance of the single out-of-sample historical path we calculated in Section 9 based on the initial assumption and parameter steps applied. In the next subsection 10.4 we will assess whether the results of the HS-FP model constitutes a legitimate empirical finding which holds true under varied assumptions and parameters.

10.4 Probability of Backtest Overfitting

The probability of backtest overfitting is the non-null probability that a strategy with optimal performance In Sample (IS) ranks below the median Out-Of-Sample (OOS) [8, 134]. We consider the probability of backtest overfitting in this paper to assess whether the results shown in Section 9 constitute a legitimate empirical finding which holds true under varied assumptions and parameters.

We estimate the probability of backtest overfitting through Combinatorially Symmetric Cross Validation (CSCV). Similar to approach applied by Bailey and Lopez [8] the following overfit statistics are calculated:

1. Probability of backtest overfitting which is the probability that the model configuration selected as optimal IS will underperform the median of the N model configurations OOS.
2. Performance degradation which determines to what extent greater performance IS leads to lower performance OOS.
3. Probability of loss which is the probability that the model selected as optimal IS will deliver a loss OOS.
4. Stochastic dominance which determines whether the procedure used to select a strategy IS is preferable to randomly choosing one model configuration among the N alternatives.

We analyse the PBO through CSCV procedure by varying the following five parameters: leeway, rebalancing frequency and state variable data transformation parameters i.e. prior half life, fast half-life and slow half-life. In the main results shown in Section 9 we assumed a leeway = 0.1, rebalancing frequency = semi-annually, prior half life = 5 years, fast half-life = 3 months and slow half-life = 12 months. In this section these parameters will be varied.

In the probability of backtest overfitting analysis we explore the following specific values for these parameters:

1. leeway= 0.1, 0.2, 0.3,
2. rebalancing frequency= 1, 2, 3, ..., 6,..., 12 (monthly, bi-monthly, quarterly,.., half-yearly,..., annually),
3. prior half life = 5, 6, 7, 8 years,
4. fast half life = 3, 6, 9, 12 months and
5. slow half life = 12, 18, 24, 36 months.

The combination of all these parameters results in a 5 dimensional mesh of 2304 elements. We estimate the PBO using the CSCV procedure and set the threshold Sharpe Ratio (SR) as 0.

Figure 19 plots the distribution of logits, the logits is the logarithm of odds and in our analysis the odds represents the odds that the optimal strategy chosen IS happens to underperform OOS. When the distribution of logits is centered in a significantly positive value with left tail marginally covering the region of negative logits, this implies backtest results are conducive to good OOS results.

In our analysis as seen in Figure 19, the distribution of logits is centered around a positive value with left tail marginally covering negative logits thus implying the backtest results are conducive to good OOS results. Using the logits we calculate PBO of approximately 1.5%. The PBO of 1.5% implies that there is a low probability that the backtest from the HS-FP framework was overfitted.

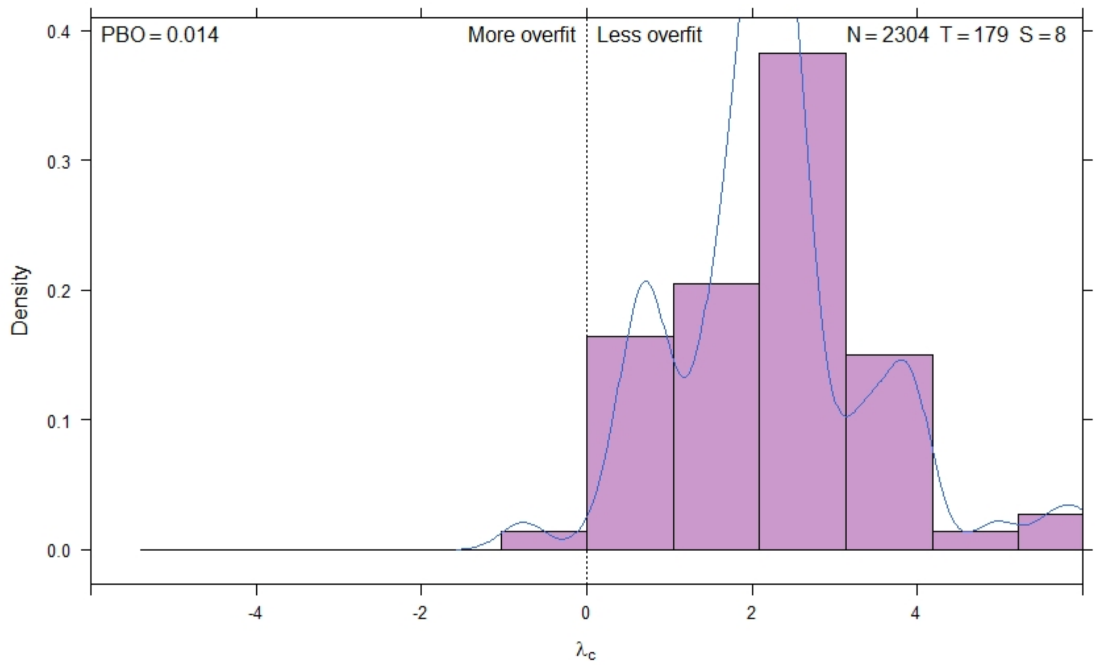

Figure 19: Shows the distribution of Logits and Probability of Backtest Overfitting (PBO) for the HS-FP model, the Logits is the logarithm of odds, and in our analysis the odds represents the odds that the optimal strategy chosen IS will underperform OOS. The Logits are then used to calculate the PBO. When the distribution of Logits is centered in a significantly positive value with left tail marginally covering the region of negative logits this implies the backtest results are conducive to good OOS results.

In Figure 20 we show the pairs of IS and OOS SR for each of the combination subsets of the optimal model configuration. The regression line that goes through the pairs of SR IS and SR OOS as seen in Figure 20 has a negative slope of -0.88, this implies that the higher the SR is IS the lower the SR is OOS. This strong negative relationship indicates that seeking the optimal performance at some point does becomes detrimental. Figure 20 also shows that approximately 10% of HS-FP SR OOS are negative despite all SR IS being positive and ranging approximately between 0.07 and 0.32.

The negative slope of the regression line is common in most practical model applications. This is due to the compensation effect as explained by Bailey et al. in [9]. The reason for this negative relationship can be intuitively understood as the effect of overfitting a model to historical scenarios such that it is often rendered unfit to future predictions. It is important to note that the PBO framework of Bailey and Lopez [8] can be applied to any performance metric we have used the Sharpe ratio metric as it is a key aspect to the way we select the optimal strategy in this paper.

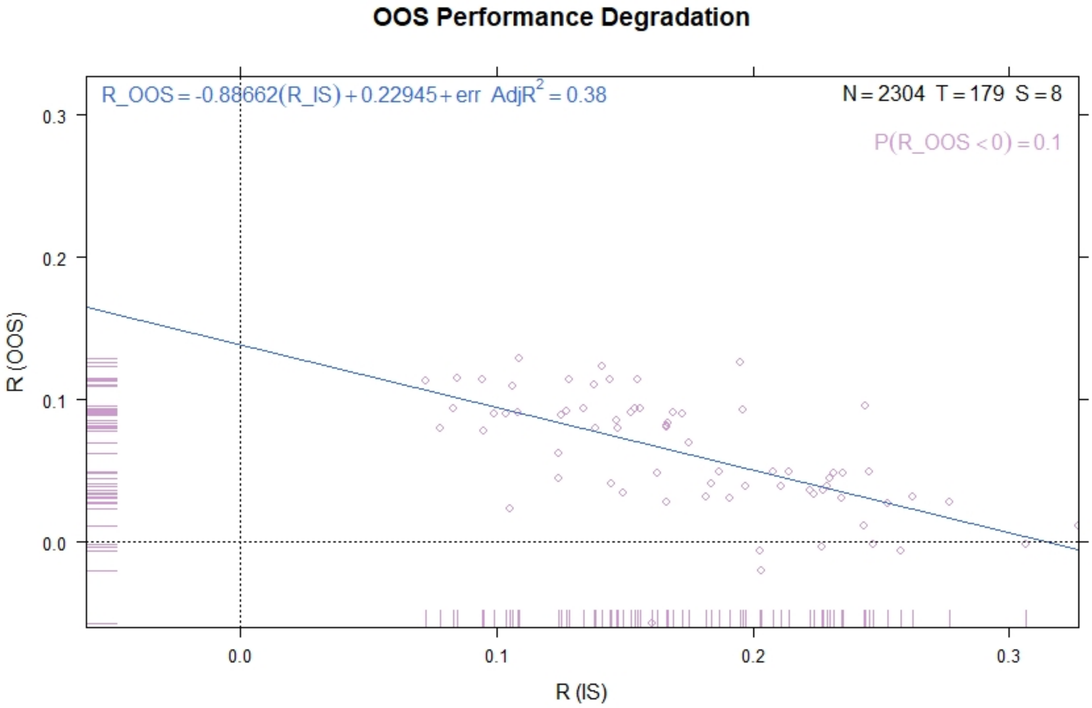

Figure 20: Shows the pairs of (SR IS,SR OOS) for optimal model configuration or the performance degradation associated with HS-FP model backtest and the OOS probability of loss.

Finally, we look at the stochastic dominance, to determine whether the distribution of the best performing OOS SR across all combinations stochastically dominates the distribution of all OOS SR. If this dominance is not seen then it would present strong evidence that selecting by best SR does not provide consistently better OOS results than a random strategy selection criterion. In our analysis as seen in Figure 21 we see that selecting by best SR did add value, since the distribution of OOS performance for the best performing OOS SR dominates the overall distribution of OOS SR.

This can be seen by the fact that for every level of OOS SR the proportion of optimized model configurations is lesser than the proportion of non-optimized model configuration, thus the probabilistic mass of the optimized model is shifted to the right of the non-optimized model. From a second-order dominance perspective as seen in Figure 21 on y2-axis it implies that the distribution does not dominate. However, as mentioned in [8] first-order stochastic dominance is a sufficient condition for second-order stochastic dominance.

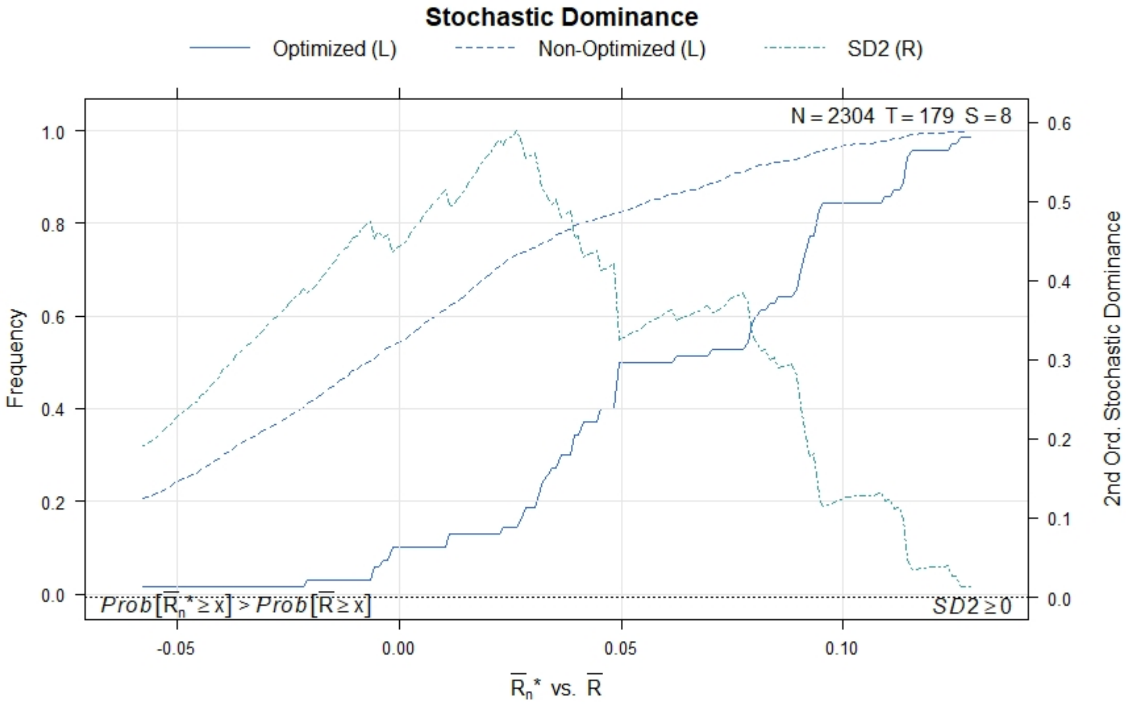

Figure 21: Shows the stochastic dominance of HS-FP with the cumulative distribution function of best OOS SR across all combinations (optimized) and OOS SR (non-optimized), as well as the second order stochastic dominance (SD2) for every OOS SR.

10.5 Deflated Sharpe Ratio

Performance inflation can occur due to non-normality of asset class returns and due to selection bias under multiple testing. The Deflated Sharpe Ratio (DSR) is a uncertainty-adjusted investment skill metric that measures the probability that a strategy exceeds a benchmark threshold in the presence of non-normal returns and selection bias under multiple testing. Performance inflation due to selection bias under multiple testing is important to assess as all historical backtests in the field of finance is not truly OOS as the researcher can tweak the inputs of a backtest to obtain an ideal OOS backtest. The DSR can be interpreted as $PSR[SR_0]$ where the benchmark SR_0 is no longer user-defined but is estimated as seen in equation 55.

The formula for the DSR as seen in Prado and Lewis [134] is:

$$\widehat{DSR} \equiv \widehat{PSR}[\widehat{SR}_0] = Z\left[\frac{(\widehat{SR}-\widehat{SR}_0)\sqrt{T-1}}{\sqrt{1-\hat{\gamma}_3\widehat{SR}+\frac{\hat{\gamma}_4-1}{4}\widehat{SR}^2}}\right] \tag{54}$$

where:

$$\widehat{SR}_0 = \sqrt{V\left[\{\widehat{SR}_n\}\right]}\left((1-\gamma)Z^{-1}\left[1-\frac{1}{K}\right]+\gamma Z^{-1}[1-\frac{1}{K}e^{-1}]\right) \tag{55}$$

also where:

$\widehat{SR}_0$ is a user-defined benchmark level

$V\left[\left\{\widehat{SR}_n\right\}\right]$ is the variance across the trials' estimated SR

K is the number of independent trials

Z[.] is the CDF of the standard Normal distribution

γ is the Euler-Mascheroni constant.

In Section 10.3 we assess the probability of performance inflation due to only non-normality of asset class returns. In this section we assess the probability of performance inflation due to non-normality of asset class returns and selection bias under multiple testing. We calculate the DSR to assess the legitimacy of the backtest results shown in Section 9. The DSR is a PSR where the rejection threshold is adjusted to reflect the multiplicity of trials.

We calculate the DSR using the method suggested in the paper [134]. This method involves using optimal K-means cluster analysis to reduce the number of total trials (N) to approximate the number of independent trials (K), deriving optimal weights for each of the strategies within the independent trials by using inverse variance optimization, then calculating the aggregate Sharpe ratio for each of the independent trials and finally calculating the variance of these K Sharpe ratios. More details regarding this calculation can be seen in Appendix 13.4.

In our analysis, aided by the number of total trials obtained in Section 10.4 we have N=2304, from the recursive cluster analysis the number of independent trials is calculated to be K=9, the length of our observations is T=179 months as seen in Section 9, the observed SR^* equals 0.09 and the user defined benchmark SR adjusted for multiplicity of trials $\hat{SR}$ equals 0.11, this results in a DSR of 0.40. This DSR implies there is 40% chance that the true SR associated with this strategy is greater than zero. Hence, the DSR also implies that there is performance inflation in the results displayed in Section 9.

11 Conclusions

The aim of this research is to implement and assess the performance of a systematic asset allocation model that incorporates the most recent asset class return behavior with a popular set of market state variables in a mathematically sound manner. The performance of the strategy is assessed based on whether this time and state conditioned risk and return forecast of asset class return distribution provides any significant improvement on portfolio performance when comparing the out-of-sample performance of the strategy to alternative benchmark models. In the next sections we summarise the findings of this research.

11.1 A Systematic Asset Allocation Framework

The framework for systematic asset allocation uses Meucci's Historical Simulation with Flexible Probabilities Model (see Section 6). This is a non-parametric approach to asset allocation which considers future asset behavior to be conditional on most recent asset behavior and different market state variables. The framework then outputs a forward looking distribution that is consistent with this investor view but remains as close as possible to the prior distribution. The framework uses the concept of flexible probabilities to achieve time varying asset class distributions that are conditioned by most recent time information and the state of the market.

The framework comprises of two main steps: the first step is prediction and the second step is portfolio construction. In the first step, the user must specify the horizon of recent time information and state variable data that should be used by the HS-FP model to systematically output the next period mean and covariance predictions. The mean and covariance predictions are informed through the adjustment of historical asset class returns and covariances with flexible probabilities that incorporate time and state conditioning. In the second step, the optimal portfolio is constructed using traditional portfolio construction theory and user specified portfolio constraints.

The key benefits of this framework are: first, it allows for time-varying expected returns and covariances, second, this framework allows one to combine multiple state variables which can help reduce the noise of individual state variables and potentially mitigate the risk of over-fitting and model misspecification. Here we have considered state variables ranging from macroeconomic, financial, trend and risk based indicators (see Table 3, Section 9.2 and Section 9.3). This framework allows investors the flexibility to stress test any view, here we specifically tested the view that investors trust more recent information and prefer specific popular market state variables in making their investment decisions.

Although the key contribution of this paper is to consider the potential effectiveness of a simple HS-FP implementation in the context of South African markets, the work can be of interest to a more general audience as it links the HS-FP framework to an explicit benchmark based test, and a quantitative reflection on the perils of backtest overfitting (see Section 10). In recent years backtests have been flagged for poor out-of-sample robustness, to assess the robustness of our results we assess the performance of the framework under varied transaction cost assumptions (see Section 10.1) and training window assumptions (see Section 10.2), as well as observing powerful backtest overfitting statistics such as the probabilistic sharpe ratio (see Section 10.3), probability of backtest overfitting (see Section 10.4) and deflated sharpe ratio (see Section 10.5).

11.2 Implementation and Performance

In Meucci's papers [142, 144, 145] the theoretical framework of HS-FP model is introduced and a basic implementation of the HS-FP model is showcased. His work does not explicitly compare the empirical results of the HS-FP model to alternative benchmark models, nor does he suggest a set of potential state variables that could be useful in conditioning asset class distributions. Here we implement a single configuration of the HS-FP framework to assess the out-of-sample performance benefit of the HS-FP model against explicit benchmark models. The benefit of the HS-FP framework is assessed through achieving better point estimate Sharpe ratios than benchmark models and successfully passing extensive robustness tests.

The single configuration of HS-FP model out-of-sample performance is observed over the period March 2003 to December 2017 on a monthly basis. The outcome of this implementation shows that the HS-FP model that combines all state variables results in a point estimate Sharpe ratio that exceeds both the classic MVO and EW benchmark models (see Section 9.4.2, Table 6). These results continue to hold true when the transaction costs are varied, even as high as 50 bps (see Section 10.1, Table 8). However, these results are inconsistent when some robustness checks are assessed, specifically, when the training windows are varied (see Section 10.2, Table 9), the PSR and DSR measures imply that the point estimate Sharpe ratio is inflated (see Table 10, Section 10.3 and Section 10.5), and the track record length fails to be above the minimum track record length required for statistical significance (see Section 10.3).

Can this type of idea deliver in a real way? We think it can, and we think that we have provided some evidence to support the claim in the context of the South African market use case. Can this protect against bad events? It is almost certain that it cannot protect against all bad events. However, time-scales are everything in finance, what we have shown is that the strategy did relatively well through the Global Financial Crises on an out-of-sample basis (see Section 9.4.3), even though there is an increase in portfolio turnover. Are there any concerns with the implementation of HS-FP framework? Yes, the strategy does not pass some crucial robustness checks and does have higher turnover than benchmark models historically. Additionally, the impact of turnover in this implementation does not factor in the indirect costs or market impact of the rebalancing act.

There does appear to be a risk benefit to the higher levels of turnover experienced by the strategy after costs (see Table 9) and over the recent past. However this risk benefit is from a volatility perspective, and not in terms of the other risk stats MDD nor CVaR (See Table 9). The HS-FP does show a low probability that the backtested returns from the HS-FP framework was overfitted (see Section 10.4); but this does not provide compelling evidence that this approach can be naively used for asset management decision making, neither smart beta type indexation nor even fully automated quasi-active management, at least in its current implementation. It is however, conceivable that a relatively shorter and more adaptive time-horizon, where more cross-sectional return and state variable data is quantitatively included into the investment decision making process of the HS-FP framework, could have an effective risk-return advantage.

11.3 Approach in Practice

Due to this approach being highly sensitive to the data inputed, in practice, for this approach to be implemented successfully in the investment industry a thorough analysis needs to be done to identify useful data that is both appropriate and reliable in predicting asset class returns. In South Africa similar to most countries globally there remains a shortage of reliable data. It can be argued that reliable South African financial and macroeconomic data barely exists for the past ten years.

It can also be argued that the future would look different to the past and that the world changes quickly as strategic agents adapt. This can, not only make historic data stale, but it increases the importance of cross-sectional data aggregation and modeling methods? In the absence of reliable and appropriate data this approach becomes futile. As it is the quality not quantity of data that is important. In identifying useful data, alternative datasets need to be considered as traditional datasets that arise from structured data is already highly commoditized. In practice alternative datasets such as social media feeds, data from satellites, transaction data and so on should be tested for usefulness in predicting asset class returns.

Additionally quantitative approaches are often stated by investment professionals to be black boxes hence to implement this approach practically would require an up-skilling of individuals to understand the techniques behind this approach thoroughly. This is a more difficult task as majority of individuals in asset management industry are from an accounting background utilizing fundamental valuation primarily. There is also skepticism regarding quantitative models post GFC and this lack of trust has narrowed the consideration of quantitative models in the asset management industry as discussed in Paskaramoorthy, Gebbie and Zyl [151].

Finally the usefulness of this approach is crucially dependent on whether the results are a true representation of reality and whether there is backtest overfitting and inflation of performance. As shown in Section 10.4 the approach does have a low probability of backtest overfitting. However these results are inconsistent when the training windows are varied, the PSR and DSR measures imply that the point estimate Sharpe ratio is inflated and the track record length fails to be above the minimum track record length required for statistical significance. All of the previously mentioned failures imply that this approach has to be further interrogated and is currently pre-mature to argue that it will be worthwhile in the asset management industry.

11.4 Future Work

In future research, we hope to incorporate transaction costs directly into the portfolio construction step. The current portfolio construction step, could however, account for costs by adjusting the objective function. Additionally, we could investigate the performance of the HS-FP model across a broader range of asset classes and regions, and consider a broader set of state variables that could increase the breadth of information used in order to mitigate window dependencies, improve the forward looking adaptability and predictive power, while controlling the rebalancing variance of the model implementation.

In future research, we could also empirically test the risk benefit of the HS-FP framework against explicit risk based benchmark models such as GARCH, RiskMetrics and other risk models assessing which risk approach better assesses and estimates the time varying risk property of asset classes. Additionally we could use cluster analysis to identify the true asset class clusters, allocate to various asset class clusters based on risk parity weighting scheme and potentially build scenarios that can be used as training sets for a machine learning approach that maps the time and state conditioned flexible probabilities to portfolio controls directly. Papers such as Hendricks [97] and Paskaramoorthy [151] already attempt to incorporate machine learning techniques and an integrated and online framework into the investment workflow.

12 References

[1] Al-Najjar, N. Aggregation and the law of large numbers in large economies. *Games and Economic Behavior 47*, 1 (2004), 1–35.

[2] Arnott, R., and Germeten, J. Systematic asset allocation. *Financial Analysts Journal - FINANC ANAL J 39* (11 1983), 31–38.

[3] Arnott, R., Harvey, C., and Markowitz, H. A backtesting protocol in the era of machine learning. *SSRN Electronic Journal* (01 2018).

[4] Arrow, K. J., and Radner, R. Allocation of resources in large teams. *Econometrica 47*, 2 (March 1979), 361–385.

[5] Asness, C., Israelov, R., and Liew, J. International diversification works (eventually). *Financial Analysts Journal 67* (06 2011).

[6] Awk, R. Microeconomics: Theory and applications. *New York:John Wiley and Sons, Inc.* (1976), 447.

[7] Bai-Marrow, A., and Radia, S. Benchmarks and indices. *CFA Society United Kingdom.*

[8] Bailey, D., Borwein, J., Lopez de Prado, M., and Zhu, Q. The probability of backtest overfitting. *Journal of Computational Finance 20* (04 2017), 39–69.

[9] Bailey, D., Borwein, J. J., Lopez de Prado, M., and Zhu, Q. Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. *Notices of the American Mathematical Society 61* (05 2014), 458.

[10] Bailey, D., and Lopez de Prado, M. The sharpe ratio efficient frontier. *The Journal of Risk 15* (12 2012), 3–44.

[11] Bailey, D., and Lopez de Prado, M. The deflated sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality. *The Journal of Portfolio Management 40* (09 2014), 94–107.

[12] Bansal, R., and Viswanathan, S. No arbitrage and arbitrage pricing: A new approach. *Journal of Finance 48*, 4 (1993), 1231–62.

[13] Banz, R. W. The relationship between return and market value of common stocks. *Journal of Financial Economics 9*, 1 (1981), 3–18.

[14] Basu, S. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance 32*, 3 (1977), 663–82.

[15] Basu, S. The relationship between earnings' yield, market value and return for nyse common stocks: Further evidence. *Journal of Financial Economics 12*, 1 (1983), 129–156.

[16] Bates, J., and Granger, C. The combination of forecasts. *OR 20* (12 1969), 451–468.

[17] Benartzi, S., and Thaler, R. Naive diversification strategics in retirement saving plans. *Amer. Econ. Rev 91* (01 2001), 78–98.

[18] Bengtsson, C., and Holst, J. On portfolio selection: Improved covariance matrix estimation for swedish asset returns.

[19] Berkowitz, J., and O'Brien, J. How accurate are value-at-risk models at commercial banks? *Finance and Economics Discussion Series 2001* (07 2001), 1–27.

[20] Bernstein, W. J. *The Investor's Manifesto: Preparing for Prosperity, Armageddon, and Everything in Between.* 2009.

[21] Best, M., and Grauer, R. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies 4* (02 1991), 315–42.

[22] Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calc. Math. Soc. 35* (11 1942).

[23] Black, F., and Litterman, R. Global portfolio optimization. *Financial Analysts Journal* (09 1992).

[24] Blake, D., and Timmermann, A. *Performance benchmarks for institutional investors.* 12 2002, pp. 108–141.

[25] Block, S., and French, D. The effect of portfolio weighting on investment performance evaluation: The case of actively managed mutual funds. *Journal of Economics and Finance 26* (01 2000), 16–30.

[26] Bélyácz, I., and Zsolt, N. Random walk, ergodicity versus predictability – the case of the budapest stock exchange. *Advances in Economics and Business 4* (2016), 165–181.

[27] Bollerslev, T., Engle, R., and Wooldridge, J. A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*, 1 (1988), 116–31.

[28] Bossaerts, P., and Hillion, P. Implementing statistical criteria to select return forecasting models: What do we learn? *Review of Financial Studies 12* (02 1999), 405–28.

[29] Boucher, C. Stock prices, inflation and stock returns predictability. *SSRN Electronic Journal 27* (12 2004).

[30] Brinson, G., Singer, B., and Beebower, G. Determinants of portfolio performance ii: An update. *Financial Analysts Journal - FINANC ANAL J 47* (05 1991), 40–48.

[31] Britten-Jones, M. The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance 54*, 2 (1999), 655–671.

[32] Broadie, M. Computing efficient frontiers using estimated parameters. *Ann. Oper. Res. 45*, 1–4 (Dec. 1993), 21–58.

[33] Brown, R. Toward a prescriptive science and technology of decision aiding. *Annals of Operations Research 19* (05 1989), 467–483.

[34] Brown, S. The Effect of Estimation Risk on Capital Market Equilibrium. *Journal of Financial and Quantitative Analysis 14*, 2 (June 1979), 215–220.

[35] Burmeister, E., Roll, R., and Ross, S. A practitioner's guide to factor models. *The Research Foundation of the Institute of Chartered Financial Analysts* (1994).

[36] Campbell, J. Asset pricing at the millennium. *Journal of Finance 55* (02 2000), 1515–1567.

[37] Campbell, S. A review of backtesting and backtesting procedures. Finance and Economics Discussion Series 2005-21, Board of Governors of the Federal Reserve System (U.S.), 2005.

[38] Carhart, M., Krail, R., Stevens, R., and Welch, K. Testing the conditional capm.

[39] Carlton, C. Correlations have personality too an analysis of correlations between assets. *Price Asset Management* (2013).

[40] Casati, A. *The Statistics of the Maximum Drawdown in Financial Time Series.* 01 2013, pp. 347–363.

[41] Ceria, S., and Stubbs, R. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management 7* (04 2006).

[42] Chan, L. K. C., Karceski, J., and Lakonishok, J. On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies 12*, 5 (1999), 937–74.

[43] Chen, Z., and Knez, P. J. Portfolio performance measurement: Theory and applications. *Review of Financial Studies 9*, 2 (1996), 511–55.

[44] Chou, Y.-L. A new proof for a strong law of large numbers of kolmogorov's type via weak convergence.

[45] Chow, G., Jacquier, E., Kritzman, M., and Lowry, K. Optimal portfolios in good times and bad. *Financial Analysts Journal 55* (05 1999), 65–73.

[46] Christiansen, C., Nygaard Eriksen, J., and Vinther Møller, S. Forecasting us recessions: The role of sentiment. *Journal of Banking and Finance 49* (12 2014), 459–468.

[47] Cochrane, J. H. Asset pricing.

[48] Connolly, T., and Zeelenberg, M. Regret in decision making. *Current Directions in Psychological Science - CURR DIRECTIONS PSYCHOL SCI 11* (12 2002), 212–216.

[49] Connor, G. The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal 51* (05 1995), 42–46.

[50] Connor, G. Book review. active portfolio management: A quantitative approach to providing superior returns and controlling risk, 2nd edition. rc grinold, rn kahn. *Review of Financial Studies - REV FINANC STUD 13* (10 2000), 1153–1156.

[51] Damodaran, A. Strategic risk taking: A framework for risk management.

[52] Daniel, K., and Titman, S. Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance 52*, 1 (1997), 1–33.

[53] Darrat, A., and Zhong, M. Permanent and transitory driving forces in the asian-pacific stock markets. *Financial Review 37* (02 2002), 35 – 51.

[54] Dasgupta, S., and Lahiri, K. On the use of dispersion measures from napm surveys in business cycle forecasting. *Journal of Forecasting 12* (04 1993), 239 – 253.

[55] Davidson, P. Can future systemic financial risks be quantified? ergodic vs nonergodic stochastic processes. *Revista de Economia Política 29* (10 2009).

[56] Davidson, P. Samuelson and the keynes/post keynesian revolution: The evidence showing who killed cock robin.

[57] Demiguel, V., Garlappi, L., and Uppal, R. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies 22* (05 2009).

[58] Dou, Y., Gallagher, D., Schneider, D., and Walter, T. Out-of-sample stock return predictability in australia. *Australian Journal of Management 37* (12 2012), 461–479.

[59] Dreger, C., and Schumacher, C. Out-of-sample performance of leading indicators for the german business cycle: Single vs. combined forecasts. *Journal of Business Cycle Measurement and Analysis 2005* (06 2005), 3–3.

[60] Dybvig, P., and Ingersoll, J. Mean-variance theory in complete markets. *The Journal of Business 55*, 2 (1982), 233–51.

[61] Dybvig, P., and Ross, S. *Arbitrage.* 01 2018, pp. 377–388.

[62] Dybvig, P., and Stephen, R. Differential information and performance measurement using a security market line. *Journal of Finance 40* (1985), 383—400.

[63] Erdo˘, E., Goldfarb, D., and Iyengar, G. Robust active portfolio management. *The Journal of Computational Finance 11* (12 2006).

[64] Evans, J., Orszag, M., and Piggott, J. *Pension fund governance: A global perspective on financial regulation.* 01 2008.

[65] Eychenne, K., and Roncalli, T. Strategic asset allocation. *SSRN Electronic Journal* (03 2011).

[66] Fabozzi, F., Kolm, P., Pachamanova, D., and Focardi, S. *Robust Portfolio Optimization and Management.* 05 2007.

[67] Fabozzi, F. J., Focardi, S., and Jonas, C. Investment management: A science to teach or an art to learn. *CFA Institute Research Foundation Monograph* (05 2014).

[68] Fabozzi, F. J., Focardi, S., and Jonas, C. Investment management: A science to teach or an art to learn. *CFA Institute Research Foundation Monograph* (5 2014).

[69] Fama, E., and French, K. The cross-section of expected stock returns. *Journal of Finance 47*, 2 (1992), 427–65.

[70] Fama, E., and French, K. Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*, 1 (1996), 55–84.

[71] Fama, E. F., and French, K. R. The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives 18*, 3 (September 2004), 25–46.

[72] Ferson, W., and Harvey, C. The variation of economic risk premiums. *Journal of Political Economy 99*, 2 (1991), 385–415.

[73] Ferson, W., and Harvey, C. Conditional variables and cross-section of stock returns. *Journal of Finance 54* (08 1999), 1325–1360.

[74] Ferson, W. E., Kandel, S., and Stambaugh, R. Tests of asset pricing with time-varying expected risk premiums and market betas. *Journal of Finance 42*, 2 (1987), 201–20.

[75] Flavin, T., and Wickens, M. Macroeconomic influences on optimal asset allocation. *Review of Financial Economics 12* (02 2003), 207–231.

[76] Fonseca, G. Bernoulli and the st. petersburg paradox.

[77] French, D., Dumas, B., and Allaz, B. Financial securities - market equilibrium and pricing methods. *The Journal of Finance 51* (09 1996), 1565.

[78] Garlappi, L., Uppal, R., and Wang, T. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies 20*, 1 (2007), 41–81.

[79] Gebbie, T., and Wilcox, D. Evidence of characteristic cross-sectional pricing of stock returns in south africa. *Unpublished manuscript* (2007).

[80] Ghysels, E. On stable factor structures in the pricing of risk: Do time varying betas help or hurt? *Journal of Finance 53* (12 2002), 549–574.

[81] Gibbons, M. R., and Ferson, W. Testing asset pricing models with changing expectations and an unobservable market portfolio. *Journal of Financial Economics 14*, 2 (1985), 217–236.

[82] Goetzmann, W., Grinold, R., and Kahn, R. Active portfolio management. *The Journal of Finance 51* (09 1996), 1559.

[83] Goyal, A., and Welch, I. Predicting the equity premium with dividend ratios. *Management Science 49* (03 2002).

[84] Green, R., and Hollifield, B. When will mean-variance efficient portfolios be well diversified? *Journal of Finance 47*, 5 (1992), 1785–809.

[85] Guidolin, M., and Timmermann, A. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control 31* (01 2008), 3503–3544.

[86] Gupta, R., and Modise, M. Macroeconomic variables and south african stock return predictability. *Economic Modelling 30* (04 2011).

[87] Halldórsson, B. An interior-point method for a class of saddle-point problems. *Journal of Optimization Theory and Applications 116* (03 2003), 559–590.

[88] Hansen, L., and Jagannathan, R. Restrictions on intertemporal marginal rates of substitution implied by asset returns.

[89] Hansen, L., and Jagannathan, R. Implications of security market data for models of dynamic economies. *Journal of Political Economy 99*, 2 (1991), 225–62.

[90] Hansen, L., and Richard, S. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica 55* (05 1987).

[91] Hansen, L. P., and Richard, S. F. The Role of Conditioning Information in Deducing Testable. *Econometrica 55*, 3 (May 1987), 587–613.

[92] Harrison, J. M., and Kreps, D. M. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory 20*, 3 (June 1979), 381–408.

[93] Harvey, C. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics 24*, 2 (1989), 289–317.

[94] Harvey, C. R. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics 24*, 2 (1989), 289–317.

[95] Hellinger, E. Neue begr??ndung der theorie quadratischer formen von unendlichvielen ver??nderlichen. *Journal Fur Die Reine Und Angewandte Mathematik - J REINE ANGEW MATH 1909* (01 1909), 210–271.

[96] Hellinger, E. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik 136* (1909), 210–271.

[97] Hendricks, D. Using real-time cluster configurations of streaming asynchronous features as online state descriptors in financial markets. *Pattern Recognition Letters 97* (03 2016).

[98] Hendry, D. *Dynamic Econometrics*. Oxford University Press, 1995.

[99] Herbert, S. Administrative behavior: A study of decision-making processes in administrative organizations. *4th edn. New York: Basic Books* (1997).

[100] Hsing, Y. The stock market and macroeconomic variables in a brics country and policy implications. *International Journal of Economics and Financial Issues 1* (01 2011), 12–18.

[101] Huang, D., Jiang, F., Tu, J., and Zhou, G. Forecasting stock returns in good and bad times: The role of market states. *27th Australasian Finance and Banking Conference 2014 Paper; Asian Finance Association (AsianFA) 2016 Conference* (07 2017).

[102] Hull, J. Risk management and financial institution.

[103] Idzorek, T. A step-by-step guide to the black-litterman model: Incorporating user-specified confidence levels.

[104] Ilmanen, A., and Kizer, J. The death of diversification has been greatly exaggerated. *The Journal of Portfolio Management 38* (04 2012), 15–27.

[105] Institute, C. *CFA Program Curriculum 2017 Level III Volumes 1 - 6*. 2017.

[106] JAGANNATHAN, R., and MA, T. Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance 58* (01 2003), 1651–1684.

[107] Jagannathan, R., and Wang, Z. The conditional capm and the cross-section of expected returns. *Journal of Finance 51*, 1 (1996), 3–53.

[108] Jarque, C. M., and Bera, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters 6*, 3 (1980), 255–259.

[109] Jorion, P. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis 21* (09 1986), 279–292.

[110] Jostova, G. Predictability in emerging sovereign debt markets. *The Journal of Business 79* (03 2006), 527–566.

[111] Judd, K. The law of large numbers with a continuum of iid random variables. *Journal of Economic Theory 35*, 1 (1985), 19–25.

[112] Kahneman, D. Thinking, fast and slow. *Allen Lane, Penguin Books Ltd* (11 2011).

[113] Kan, R., and Wang, K. Does the nonlinear apt outperform the conditional capm?

[114] Kan, R., and Zhou, G. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis 42*, 3 (2007), 621–656.

[115] Keim, D. Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics 12*, 1 (1983), 13–32.

[116] Keim, D. B. Dividend yields and stock returns: Implications of abnormal January returns. *Journal of Financial Economics 14*, 3 (September 1985), 473–489.

[117] Khodadadi, A., and Zangari, P. Optimisation and quantitative investment management. *Journal of Asset Management 7* (07 2006), 83–92.

[118] Kim, S.-J., and Boyd, S. Two-fund separation under model mis-specification.

[119] Klein, R. W., and Bawa, V. S. The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics 3*, 3 (June 1976), 215–231.

[120] Knight, F. Risk, uncertainty and profit. *Boston, New York, Houghton Mifflin Company* (1921).

[121] Koenig, M. Robust asset allocation. *Annals of Operations Research 132* (01 2004), 157–187.

[122] Kollár, M. A sketch of macro based asset allocation. *International Journal of Economic Sciences 2(3)* (2013), 101–120.

[123] Koumou, N. B. G. Diversification and portfolio theory: A review.

[124] Kritzman, M., and Markowitz, H. An interview with nobel laureate harry m. markowitz. *Financial Analysts Journal 73* (09 2017), 1–6.

[125] Krokhmal, P., Palmquist, J., and Uryasev, S. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk 4* (05 2003).

[126] Kullback, S., and Leibler, R. On information and sufficiency. *Annals of Mathematical Statistics 22* (03 1951), 79–86.

[127] Ledoit, O., and Wolf, M. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management 30* (07 2003).

[128] Ledoit, O., and Wolf, M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance 10*, 5 (2003), 603–621.

[129] Lettau, M., and Ludvigson, S. Resurrecting the (c)capm: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy 109*, 6 (2001), 1238–1287.

[130] Lewellen, J., and Nagel, S. The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics 82*, 2 (2006), 289–314.

[131] Lintner, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics 51*, 2 (1969), 222–24.

[132] Litzenberger, R. H., and Ramaswamy, K. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics 7*, 2 (1979), 163–195.

[133] Lopez de Prado, M. Building diversified portfolios that outperform out of sample:. *The Journal of Portfolio Management 42* (07 2016), 59–69.

[134] Lopez de Prado, M., and Lewis, M. Detection of false investment strategies using unsupervised learning methods. *Quantitative Finance 19* (07 2019), 1–11.

[135] Markowitz, H. Portfolio selection. *The journal of finance 7(1)* (03 1952), 77–91.

[136] Matoti, L., Gebbie, T., and Wilcox, D. Building a statistical linear factor model and a global minimum variance portfolio using estimated covariance matrices. *UCT* (06 2009).

[137] McDonnell, P. J., and ebrary, I. Optimal portfolio modeling : models to maximize return and control risk in excel and r + cd-rom.

[138] Mehta, A., Neukirchen, M., Pfetsch, S., and Poppensieker, T. Managing market risk: today and tomorrow. *Technical report, McKinsey Working Papers on Risk* (05 2012).

[139] Merton, R. Theory of rational option pricing. *Bell Journal of Economics 4*, 1 (1973), 141–183.

[140] Meucci, A. Risk and asset allocation. *Springer* (2005).

[141] Meucci, A. Fully flexible views: Theory and practice. *Risk Magazine 21* (2008).

[142] Meucci, A. Historical scenarios with fully flexible probabilities. *SSRN Electronic Journal* (10 2010).

[143] Meucci, A. 'the prayer' ten-step checklist for advanced risk and portfolio management. *SSRN Electronic Journal* (02 2011).

[144] Meucci, A. Effective number of scenarios in fully flexible probabilities.

[145] Meucci, A. Estimation and stress-testing via time- and market-conditional flexible probabilities.

[146] Michaud, R. The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal - FINANC ANAL J 45* (01 1989), 31–42.

[147] Munro, B., and Silberman, K. Optimal asset allocation in different economic environments. *Cadiz Securities Quantitative Research Report* (2008).

[148] Nazmi, N. Leading economic indicators: New approaches and forecasting record : Review of lahiri k. and g. moore (eds.), 1991, (cambridge university press, cambridge). *International Journal of Forecasting 10* (02 1994), 382–385.

[149] Page, S., and Panariello, R. When diversification fails. *Financial Analysts Journal 74* (05 2018), 1–14.

[150] Pascal, F., and Payne, R. Corporate finance. *University of London Press* (2009).

[151] Paskaramoorthy, A., Zyl, T., and Gebbie, T. A framework for online investment algorithms.

[152] Pastor, L. Portfolio selection and asset pricing models. *Journal of Finance 55*, 1 (2000), 179–223.

[153] Pastor, L., and Stambaugh, R. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics 56*, 3 (2000), 335–381.

[154] Perignon, C., and Smith, D. The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking and Finance 34*, 2 (2010), 362–377.

[155] Peters, O. The ergodicity problem in economics. *Nature Physics 15* (12 2019), 1216–1221.

[156] Peters, O. The ergodicity problem in economics. *Nature Physics 15* (12 2019), 1216–1221.

[157] Poitras, G., and Heaney, J. Classical ergodicity and modern portfolio theory. *Chinese Journal of Mathematics 2015* (08 2015), 1–17.

[158] Pritsker, M. The hidden dangers of historical simulation. *Journal of Banking and Finance 30*, 2 (2006), 561–582.

[159] Qadan, M., Kliger, D., and Chen, N. Idiosyncratic volatility, the vix and stock returns. *The North American Journal of Economics and Finance 47* (06 2018).

[160] Rapach, D., Strauss, J., and Zhou, G. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies 23* (04 2009), 821–862.

[161] Rapach, D., Strauss, J., and Zhou, G. International stock return predictability: What is the role of the united states? *Capital Markets: Market Efficiency eJournal 68* (03 2010).

[162] Rapach, D., and Zhou, G. Forecasting stock returns. *Handbook of Economic Forecasting 2* (12 2013), 327–383.

[163] Richardson, M., Boudoukh, J., and Whitelaw, R. The best of both worlds: A hybrid approach to calculating value at risk. *RISK 11* (11 1997).

[164] Rockafellar, R., and Uryasev, S. Optimization of conditional value-at-risk. *Journal of Risk 2* (2000), 21–41.

[165] Rockafellar, R., and Uryasev, S. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance 26* (07 2002), 1443–1471.

[166] Roll, R., and Ross, S. An empirical investigation of the arbitrage pricing theory. *Journal of Finance 35*, 5 (1980), 1073–1103.

[167] Roll, R., and Ross, S. The arbitrage pricing theory approach to strategic portfolio planning. *Financial Analysts Journal* (05 1984), 14–26.

[168] Rose, M. Reward management. *Kogan Page* (2014).

[169] Ross, S. The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*, 3 (1976), 341–360.

[170] Ross, S. Return, risk and arbitrage. *Risk and Return in Finance Vol. I* (01 1977).

[171] Ross, S. Neoclassical finance. *Princeton University Press* (2009).

[172] Ross, S. The arbitrage theory of capital asset pricing. *HANDBOOK OF THE FUNDAMENTALS OF FINANCIAL DECISION MAKING: Part I* (2013), 11–30.

[173] Rubinstein, M. The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics 7*, 2 (1976), 407–425.

[174] Rustem, B., Becker, R., and Marty, W. Robust min–max portfolio strategies for rival forecast and risk scenarios. *Journal of Economic Dynamics and Control 24* (10 2000), 1591–1621.

[175] Samuelson, P. Risk and uncertainty: A fallacy of large numbers. *Scientia 98* (01 1963).

[176] Scherer, B. Portfolio construction & risk budgeting.

[177] Schmid, F., and Schmidt, R. *Statistical Inference for Sharpe Ratio.* 11 2009.

[178] Schöttle, K., and Werner, R. Towards reliable efficient frontiers. *Journal of Asset Management 7* (07 2006), 128–141.

[179] Sebastian, A., and Gebbie, T. Systematic asset allocation using flexible views for south african markets.

[180] Sebastian, A., and Gebbie, T. Systematic asset allocation using flexible views for south african markets.

[181] Seneta, E. A tricentenary history of the law of large numbers. *Bernoulli 19*, 4 (09 2013), 1088–1121.

[182] Shanken, J. Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics 45*, 1-2 (1990), 99–120.

[183] Sharpe, W. The sharpe ratio. *J Portf Manag 21* (01 1994), 49–58.

[184] Sharpe, W. F. The sharpe ratio. *The Journal of Portfolio Management 21*, 1 (1994), 49–58.

[185] Shiller, R. Consumption, asset markets and macroeconomic fluctuations. *Carnegie-Rochester Conference Series on Public Policy 17*, 1 (January 1982), 203–238.

[186] Sokal, R., and Michener, C. A statistical method of evaluating systematic relationships. *The University of Kansas Science Bulletin 38* (01 1958), 1409–1438.

[187] Statman, M., and Scheid, J. Global diversification. *SSRN Electronic Journal* (10 2004).

[188] Statman, M., and Scheid, J. Correlation, return gaps and the benefits of diversification. *The Journal of Portfolio Management 34* (11 2007).

[189] Stein, R. Red-blooded risk: The secret history of wall street, by aaron brown. *Quantitative Finance - QUANT FINANC 12* (08 2012), 1193–1195.

[190] Timmermann, A. Chapter 4 forecast combinations. *Handbook of Economic Forecasting 1* (12 2006), 135–196.

[191] Titman, S., and Jegadeesh, N. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48* (02 1993), 65–91.

[192] Tobin, J. Liquidity preference as behavior towards risk. *Review of Economic Studies 25*, 2 (1958), 65–86.

[193] Uhlig, H. A law of large numbers for large economies. *Economic Theory 8*, 1 (1996), 41–50.

[194] Uspensky, J. Introduction to mathematical probability. *McGraw-Hill Book Company* (1937).

[195] Vapnik, V., and Chervonenkis, A. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications 16* (01 1971), 264–280.

[196] Venter, J. A brief history of business cycle analysis in south africa.

[197] Viswanathan, S., Bansal, R., and Hsieh, D. A new approach to international arbitrage pricing. *Journal of Finance 48* (02 1993), 1719–47.

[198] Von Neumann, J., and Morgenstern, O. Theory of games and economic behavior. *Princeton University Press* (1944), xvii, 625 p. :.

[199] Wai, L. Theory and methodology of tactical asset allocation. *The Frank J. Fabozzi Series* (08 2000).

[200] Weber, E., and Coskunoglu, O. Descriptive and prescriptive mòdels of decisionmaking: Implications for the development of decision aids. *Systems, Man and Cybernetics, IEEE Transactions on 20* (04 1990), 310 – 317.

[201] Welch, I., and Goyal, A. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21* (02 2008), 1455–1508.

[202] Wilcox, D., and Gebbie, T. On the analysis of cross-correlations in south african market data. *Physica A Statistical and Theoretical Physics 344* (12 2004), 294–298.

[203] Wilcox, D., and Gebbie, T. On pricing kernels, information and risk. *Investment Analysts Journal 44* (01 2015), 1–19.

[204] Wilcox, D., and Tim, G. On pricing kernels, information and risk. *Investment Analysts Journal 44*, 1 (2015), 1–19.

[205] Wilford, D. S. True Markowitz or assumptions we break and why it matters. *Review of Financial Economics 21*, 3 (2012), 93–101.

[206] William, M. A historical overview of research in finance. *Journal of Finance 39*, 2 (1996), 323–346.

[207] Windcli, H., and Boyle, P. The 1/n pension investment puzzle. *N. Am. Actuar. J. 8* (12 2003).

[208] Wolf, M. Resampling vs. shrinkage for benchmarked managers. *SSRN Electronic Journal* (02 2006).

[209] Wu, Y. Stock prices and exchange rates in vec model—the case of singapore in the 1990s. *Journal of Economics and Finance 24* (04 2012), 260–274.

[210] Yang, Z. Pointwise convergence and ascoli theorems for nearness spaces. *Applied general topology, ISSN 1576-9402, Vol. 10, Nº. 1, 2009, pags. 49-68 10* (04 2009).

[211] Zellner, A., and Chetty, V. Prediction and decision problems in regression models from the bayesian point of view. *Journal of The American Statistical Association - J AMER STATIST ASSN 60* (06 1965), 608–616.

13 Appendix A

13.1 Derivation of MV Frontier

The investor faces the following constrained optimization problem:

$$\max_{\boldsymbol{w}} \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \tag{56}$$

subject to the full investment constraint,

$$\boldsymbol{w}^\top \mathbf{1} = 1 \tag{57}$$

To find the optimum, form the Lagrangian:

$$L = \boldsymbol{w}^\top \mu - \frac{\gamma}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} - \lambda(\boldsymbol{w}^\top \mathbf{1} - 1) \tag{58}$$

where λ is a Lagrangian multiplier for the full-investment constraint. The first-order conditions are:

$$\nabla_{\boldsymbol{w}} L = \boldsymbol{\mu} - \gamma \boldsymbol{\Sigma} \boldsymbol{w} - \lambda \mathbf{1} = 0 \implies \boldsymbol{w}^* = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \lambda \mathbf{1}) \tag{59}$$

and

$$\frac{\partial L}{\partial \lambda} = -(\boldsymbol{w}^\top \mathbf{1} - 1) = 0 \implies \boldsymbol{w}^\top \mathbf{1} = 1 \tag{60}$$

Substituting 59 into 60, we obtain the following expression for λ,

$$\lambda = \frac{\mathbf{1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} - \frac{\gamma}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} \tag{61}$$

Substituting 61 into 59, we arrive at the solution for the optimal portfolio:

$$\boldsymbol{w}^* = \left(1 - \frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma}\right)\frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} + \left(\frac{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\gamma}\right)\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} \tag{62}$$

13.2 Derivation of MV Frontier with Risk-Free Asset

Suppose the investible set includes the risk-free asset, with expected return r_f and standard deviation equal to zero. Let w_0 be the portfolio weight invested in the risk-free asset, such that $w_0 = (1 - \boldsymbol{w}^\top 1)$.

Then, the investor's constrained optimization problem becomes:

$$\min_{\boldsymbol{w}} \frac{\gamma}{2}\boldsymbol{w}\boldsymbol{\Sigma}\boldsymbol{w} \text{ subject to } \boldsymbol{w}^\top \boldsymbol{\mu} + (1 - \boldsymbol{w}^\top \mathbf{1})r_f = \mu_p \tag{63}$$

The optimization problem is solved by forming the Lagragian:

$$L = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} + \lambda\big(\mu_p - \boldsymbol{w}^\top \boldsymbol{\mu} - (1 - \boldsymbol{w}^\top \mathbf{1})r_f\big) \tag{64}$$

Then, the first-order conditions are:

$$\nabla_{\boldsymbol{w}} L = \boldsymbol{\Sigma w} - \lambda(\boldsymbol{\mu} - \mathbf{1} r_f) = 0 \tag{65}$$

and

$$\frac{\partial L}{\partial \lambda} = \mu_p - \boldsymbol{w}^\top \boldsymbol{\mu} - (1 - \boldsymbol{w}^\top \mathbf{1}) r_f = 0 \tag{66}$$

Solving for λ yields,

$$\lambda = \frac{\mu_p - r_f}{(\boldsymbol{\mu} - \mathbf{1} r_f)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{1} r_f)} \tag{67}$$

Substituting 67 into 65 results in the solution for the optimal portfolio in the risky assets:

$$\boldsymbol{w}^* = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f) \left(\frac{\mu_p - r_f}{(\boldsymbol{\mu} - \mathbf{1} r_f)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \mathbf{1} r_f)} \right) \tag{68}$$

13.3 Glivenko-Cantelli Theorem

Let us assume for simplicity $\bar{i} = 1$, then the same arguments can be easily generalized to multivariate case.

Law of large numbers

A preliminary result is the law of large numbers (see Section 5), which states the $\{\epsilon_1, \ldots, \epsilon_{\bar{t}}\}$ are observations of a sequence $\{\boldsymbol{\varepsilon}_t\}_t$ of i.i.d random variables with finite expectation $\mathbb{E}\{\boldsymbol{\varepsilon}_t\} = m$ and finite variance $\mathbb{V}\{\boldsymbol{\varepsilon}_t\} = s^2$, then the sample mean converges to the true expectation

$$\lim_{t \to \infty} \frac{1}{\bar{t}} \sum_{t=1}^{\bar{t}} \boldsymbol{\epsilon}_t = m. \tag{69}$$

Pointwise convergence

We prove here that the historical cdf $\hat{F}_{\boldsymbol{\varepsilon}}^{Hist}$ converges to the true cdf $F_{\boldsymbol{\varepsilon}}$ pointwise. Pointwise convergence means that for any $x_0 \in \mathbb{R}$,

$$\lim_{\bar{t} \to +\infty} \hat{F}_{\boldsymbol{\varepsilon}}^{Hist}(x_0) = F_{\boldsymbol{\varepsilon}}(x_0)[210]. \tag{70}$$

First we notice that the true unknown cdf $F_{\boldsymbol{\varepsilon}}(\boldsymbol{x})$ can be rewritten as follows

$$F_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \mathbb{P}\{\boldsymbol{\varepsilon}_t \leq \boldsymbol{x}\} = \mathbb{E}\{\mathbf{1}_{\boldsymbol{\varepsilon}_t \leq \boldsymbol{x}}\} \tag{71}$$

where the last identity follows from $\mathbb{E}\{\mathbf{1}_{\mathbf{A}}\} = \mathbb{P}(\mathbf{A})$. Then, since for any fixed $x_0 \in \mathbb{R}$ the random variables $\mathbf{1}_{\boldsymbol{\varepsilon}_t \leq x_0}$ are still i.i.d, we have

$$\lim_{\bar{t}\to+\infty} \hat{F}_{\boldsymbol{\varepsilon}}^{Hist}(x_0) = \lim_{\bar{t}\to+\infty} \frac{1}{\bar{t}} \sum_{t=1}^{\bar{t}} \mathbf{1}_{\boldsymbol{\varepsilon}_t \leq x_0} = \mathbb{E}\{\mathbf{1}_{\boldsymbol{\varepsilon}_t \leq x_0}\} = F_{\boldsymbol{\varepsilon}}(x_0) \tag{72}$$

Hence the historical cdf $\hat{F}_{\boldsymbol{\varepsilon}}^{Hist}$ converges pointwise to the true, unknown distribution $F_{\boldsymbol{\varepsilon}}$ as $\bar{t} \to \infty$.

Uniform convergence

The Glivenko-Cantelli statement [44] is much stronger, since the convergence of the historical cdf $\hat{F}_{\boldsymbol{\varepsilon}}^{Hist}$ to the true cdf $F_{\boldsymbol{\varepsilon}}$ is not only pointwise, but also uniform. To see why this is important, consider a normal cdf $F_{\mu,\sigma^2}^{N}(\boldsymbol{x})$ and the sequence of functions $\{g_t(\boldsymbol{x})\}_{t=1}^{\infty}$ with

$$g_t(\boldsymbol{x}) = \begin{cases} F_{\mu,\sigma^2}^{N}(\boldsymbol{x}) & \boldsymbol{x} \notin (t-1, t) \\ F_{\mu,\sigma^2}^{N}(\boldsymbol{x}) + 2 & \boldsymbol{x} \in (t-1, t) \end{cases} \tag{73}$$

As $t \to \infty$ you can see that for any specific x_0, $g_t(x_0) \to F_{\mu,\sigma^2}^{N}(x_0)$, since for all $t > x_0$, $g_t(x_0) = F_{\mu,\sigma^2}^{N}(x_0)$. However no matter how large t becomes, $g_t(\boldsymbol{x})$ as a function is substantially different from $F_{\mu,\sigma^2}^{N}(\boldsymbol{x})$. Importantly, it is not a cdf, since it will always have some values greater than 1. To quantify the difference, we can look at the total area between the functions using integrals. Indeed

$$\int_{-\infty}^{\infty} \left|g_t(\boldsymbol{x}) - F_{\mu,\sigma^2}^{N}(\boldsymbol{x})\right| dx = \int_{t-1}^{t} \left|F_{\mu,\sigma^2}^{N}(\boldsymbol{x}) + 2 - F_{\mu,\sigma^2}^{N}(\boldsymbol{x})\right| dx = 2 \tag{74}$$

for all t. Uniform convergence, as in the Glivenko-Cantelli theorem, guarantees that this problem does not occur.

13.4 Deflated Sharpe Ratio Calculation

1. Using K-means cluster analysis calculate the optimal number of independent trials (K) as shown by Prado and Lewis [134].
2. Derive the optimal weights for each of K clusters using the minimum variance allocation from Prado [133]:

   For each cluster k, where $k = 1, ..., K$ with components C_k, the minimum variance weightings for each component is calculated using the below equation:

   $$\boldsymbol{w}_k = \frac{\boldsymbol{\Sigma}_k^{-1} 1}{1' \boldsymbol{\Sigma}_k^{-1} 1} \tag{75}$$

   where:

   $\boldsymbol{\Sigma}_k$ is the covariance matrix for the components of C_k.

Then the cluster return series is calculated by applying these weights to the return series:

$$\boldsymbol{S}_{k,t} = \sum_{i \in C_k} \boldsymbol{w}_{k,i} \boldsymbol{r}_{i,t} \tag{76}$$

where $\boldsymbol{r}_{i,t}$ is the return series for component i.

3. Annualize the Cluster Sharpe Ratios:

$$a\widehat{\boldsymbol{S}R}_k = \frac{\mathrm{E}\left[\left\{\boldsymbol{S}_{k,t}\right\}\right] \mathit{Frequency}_k}{\sqrt{V\left[\left\{\boldsymbol{S}_{k,t}\right\}\right] \mathit{Frequency}_k}} = \widehat{\boldsymbol{S}R}_k \sqrt{\mathit{Frequency}_k}. \tag{77}$$

4. Estimate the Variance of Cluster Sharpe Ratios:

$$\mathrm{E}\left[V\left[\{\widehat{\boldsymbol{S}R}_k\}\right]\right] = \frac{V\left[\{a\widehat{\boldsymbol{S}R}_k\}\right]}{\mathit{Frequency}_k}. \tag{78}$$

5. Calculate the Benchmark SR using equation 54.
6. Calculate DSR Using equation 55.

14 Appendix B

14.1 Code patterns

https://github.com/PonniAnn/HS-FPcode/tree/main/Code

14.1.1 Workflow for HS-FP Framework

```
%% Script for HS-FP Framework

close all;
clear all;
clc;

%% 1. Retrieving data from Bloomberg (Index Levels for Sectors ...
    and State variable)
javaaddpath('C:\blp\DAPI\blpapi3.jar')
c=blp;
s = {'RESI20 Index','INDI25 Index','FINI15 Index','SACPIYOY ...
    Index','SAGDPANN Index'};
f = {'LAST_PRICE'};
fromdate = '2/28/1998';
todate = '12/31/2017';
period = 'monthly';
[data,sec] = history(c,s,f,fromdate,todate,period);
formatOut = 'mmmm-dd-yyyy';
time=length(data{1}(:,1));

%Index levels of 3 Sectors (RESI20, INDI25,FINI15)
Res.Date=data{1}(:,1)';
Indi.Date=data{2}(:,1)';
Fini.Date=data{3}(:,1)';
Res.Price_close=data{1}(:,2)';
Indi.Price_close=data{2}(:,2)';
Fini.Price_close=data{3}(:,2)';

%State signal data (CPI and GDP)
CPI.Date=data{1}(:,1)';
CPI.value=data{4}(:,2)';
GDPq.value=data{5}(:,2)';
x=1:3:time;
xx=1:time;
GDP.Date=data{1}(:,1)';
GDP.value = spline(x,GDPq.value,xx); %converting qoq GDP into ...
    monthly data

%% 2. Compute the time series for the ...
    Sectors(RESI20,INDI25,FINI15 Return) and State variable ...
    (CPI and GDP)

```

```
% Parameters for smoothing and scoring
alpha = 0.1; % range around target z
tauHL_prior = 5*12; % prior half life 8 years
tauHL_smoo = 3; % fast half-life time 3 month
tauHL_scor = 1*12; % slow half-life time 1 yr
t_start = 12; % time starting point must be atleast 12 months

% Parameters for rebalancing
years=5;
months=12;
rebalance=1;
Reb_freq=6;
t_end=years*months:rebalance:length(data{1});
tc=0.0/100;

for roll_end=1:length(t_end)

% Returns of 3 Sectors (RESI20, INDI25,FINI15)
RESI20_rets = diff(log(Res.Price_close));
RESI20_rets = RESI20_rets(t_start:t_end(roll_end)-1);
INDI25_rets = diff(log(Indi.Price_close));
INDI25_rets = INDI25_rets(t_start:t_end(roll_end)-1);
FIN15_rets = diff(log(Fini.Price_close));
FIN15_rets = FIN15_rets(t_start:t_end(roll_end)-1);

%% State Variables
% CPI State Signal
c_CPI = CPI.value;
c_CPI = c_CPI(t_start+1:t_end(roll_end));
dr1 = CPI.value;

% GDP State Signal
c_GDP = GDP.value;
c_GDP = c_GDP(t_start+1:t_end(roll_end));
dr2 = GDP.value;

% Merging datasets
[date, i_epsi, i_z] = intersect(Res.Date,CPI.Date);
t_ = length(date(2:t_end(roll_end)));
date_z = date(t_start+1:t_end(roll_end));
times = 1:t_;

% Smoothing State Variable
state_var1=dr1;
state_var2=dr2;

z1 = zeros(1,t_);
z2 = zeros(1,t_);

for t=1:t_
    p_smoo_t = exp(-log(2)/tauHL_smoo*(repmat(t,1,t)-times(1:t)));
    gamma_t = sum(p_smoo_t);
```

```
    z1(t) = sum(p_smoo_t.*state_var1(1:t))/gamma_t;
    z2(t) = sum(p_smoo_t.*state_var2(1:t))/gamma_t;
end

% scoring state variables
mu_hat_1 = zeros(1,t_);
mu2_hat_1 = zeros(1,t_);
sd_hat_1 = zeros(1,t_);

mu_hat_2 = zeros(1,t_);
mu2_hat_2 = zeros(1,t_);
sd_hat_2 = zeros(1,t_);

for t=1:t_
    p_scor_t = exp(-log(2)/tauHL_scor*(repmat(t,1,t)-times(1:t)));
    gamma_scor_t = sum(p_scor_t);

    mu_hat_1(t) = sum(p_scor_t.*z1(1:t))/gamma_scor_t;
    mu2_hat_1(t) = sum(p_scor_t.*(z1(1:t)).^2)/gamma_scor_t;
    sd_hat_1(t) = sqrt(mu2_hat_1(t)-(mu_hat_1(t))^2);

    mu_hat_2(t) = sum(p_scor_t.*z2(1:t))/gamma_scor_t;
    mu2_hat_2(t) = sum(p_scor_t.*(z2(1:t)).^2)/gamma_scor_t;
    sd_hat_2(t) = sqrt(mu2_hat_2(t)-(mu_hat_2(t))^2);
end
z1 = (z1 - mu_hat_1)./sd_hat_1;
z1 = z1(t_start:end);

z2 = (z2 - mu_hat_2)./sd_hat_2;
z2 = z2(t_start:end);

%% 3. Compute the Flexible Probabilities conditioned via ...
    Minimum Relative Entropy and Ensemble Method for Mutliple ...
    State Variables
% Exponential Probabilitiy
% Prior is same for both z1 and z2
prior = exp(-log(2)/tauHL_prior*(t_:-1:t_start));
prior = prior/sum(prior);

% setting conditioning variable z1 = CPI
cond_1 = struct('Series',[],'TargetValue',[],'Leeway',[]);
cond_1.Series = z1;
z_star_1 = z1(end); %Target value is latest point
cond_1.TargetValue = z_star_1;
cond_1.Leeway = alpha;
p_1 = Conditional_FP_001(cond_1, prior);

% setting conditioning variable z2 = GDP
cond_2 = struct('Series',[],'TargetValue',[],'Leeway',[]);
cond_2.Series = z2;
z_star_2 = z2(end); %Target value is latest point
cond_2.TargetValue = z_star_2;
```

```
cond_2.Leeway = alpha;
p_2 = Conditional_FP_001(cond_2, prior);

% effective number of scenariois
ens_1 = EffectiveScenarios(p_1);
ens_2 = EffectiveScenarios(p_2);

% diversity indicator
b_12=sum(sqrt(p_1.*p_2));

% hellinger distance
h_12=sqrt(1-b_12);

d1=h_12;
d2=d1;

% weight for each state variable using effective number of ...
    scenarios
weight_state_var(roll_end,:)=[d1*ens_1 d2*ens_2];
weight_state_var(roll_end,:)=weight_state_var(roll_end,:)./...
    sum(weight_state_var(roll_end,:));

% equal weight for each state variable
weight_state_var_1=repmat(0.5,size(weight_state_var));

% equal weighted probability
opt_prob_1=weight_state_var_1(roll_end,1)*p_1...
    +weight_state_var_1(roll_end,2)*p_2;
opt_prob_1=opt_prob_1./sum(opt_prob_1);
p_3=opt_prob_1;

% DCC or ensemble probability
opt_prob=exp(weight_state_var(roll_end,1)*log(p_1)...
    +weight_state_var(roll_end,2)*log(p_2));
opt_prob=opt_prob./sum(opt_prob);
p_entropy=opt_prob;

%% 4. HS-FP model setup
Sector_returns=[INDI25_rets',FIN15_rets',RESI20_rets']';
Sectors_name={'Indi25','Fini15','Resi'};
num_of_eff_ports=10;
Sects_ret=Sector_returns(:,years*months-months:end)';

% 4.1 HS-FP conditioned by CPI(z1)
[m1, C1]=FPmeancov_001(Sector_returns,p_1);
meanvarFFP_1 = Portfolio('AssetList',Sectors_name,'assetmean', ...
    m1, 'assetcovar', C1,'lowerbudget', 1, 'upperbudget', 1, ...
    'lowerbound', 0);
pmvarFFP_wgt_1{roll_end} = estimateFrontier(meanvarFFP_1, ...
    num_of_eff_ports);

% 4.2 HS-FP conditioned by GDP(z2)
```

```
[m2, C2]=FPmeancov_001(Sector_returns,p_2);
meanvarFFP_2 = Portfolio('AssetList',Sectors_name,'assetmean', ...
    m2, 'assetcovar', C2,'lowerbudget', 1, 'upperbudget', 1, ...
    'lowerbound', 0);
pmvarFFP_wgt_2{roll_end} = estimateFrontier(meanvarFFP_2, ...
    num_of_eff_ports);

% 4.3 HS-FP conditioned by CPI and GDP equally weighted
[m3, C3]=FPmeancov_001(Sector_returns,p_3);
meanvarFFP_3 = Portfolio('AssetList',Sectors_name,'assetmean', ...
    m3, 'assetcovar', C3,'lowerbudget', 1, 'upperbudget', 1, ...
    'lowerbound', 0);
pmvarFFP_wgt_3{roll_end} = estimateFrontier(meanvarFFP_3, ...
    num_of_eff_ports);

% 4.4 HS-FP conditioned by CPI and GDP DCC weighted
[m, C]=FPmeancov_001(Sector_returns,p_entropy);
meanvarFFP = Portfolio('AssetList',Sectors_name,'assetmean', m, ...
    'assetcovar', C,'lowerbudget', 1, 'upperbudget', 1, ...
    'lowerbound', 0);
pmvarFFP_wgt{roll_end} = estimateFrontier(meanvarFFP, ...
    num_of_eff_ports);

% 4.5 Selecting Maximum Sharpe Ratio HS-FP Portfolio
pmvarFFP_MSR_wgt_1{roll_end} = ...
    estimateMaxSharpeRatio(meanvarFFP_1);
pmvarFFP_MSR_wgt_2{roll_end} = ...
    estimateMaxSharpeRatio(meanvarFFP_2);
pmvarFFP_MSR_wgt_3{roll_end} = ...
    estimateMaxSharpeRatio(meanvarFFP_3);
pmvarFFP_MSR_wgt{roll_end} = estimateMaxSharpeRatio(meanvarFFP);


%% 5. Benchmark Models Setup
% 5.1 Classic Mean Variance Optimization
meanvar=Portfolio;
meanvar=meanvar.setAssetList(Sectors_name);
meanvar = meanvar.setDefaultConstraints;
meanvar = meanvar.estimateAssetMoments(Sector_returns');
pmvar_wgt{roll_end}=estimateFrontier(meanvar,num_of_eff_ports);
pmvar_MSR_wgt{roll_end} = estimateMaxSharpeRatio(meanvar);
end

% 5.2 Equal Weighted
[num_assets time]=size(Sector_returns);
equal_wgt=repmat(1/num_assets,roll_end,num_assets);

%% 6. Optimal monthly portfolios through time

%optimal weights for HS-FP CPI conditioned
MVO_FFP_MSR_OT_1=cell2mat(pmvarFFP_MSR_wgt_1)';

```

```
%optimal weights for HS-FP GDP conditioned
MVO_FFP_MSR_OT_2=cell2mat(pmvarFFP_MSR_wgt_2)';

%optimal weights for HS-FP CPI and GDP conditioned (Equal Weighted)
MVO_FFP_MSR_OT_3=cell2mat(pmvarFFP_MSR_wgt_3)';

%optimal weights for HS-FP CPI and GDP conditioned (DCC Weighted)
MVO_FFP_MSR_OT=cell2mat(pmvarFFP_MSR_wgt)';

%optimal weights for classic MVO model
MVO_MSR_OT=cell2mat(pmvar_MSR_wgt)';

%optimal weights for EW benchmark model
EW_OT=equal_wgt;

%% 7. Plot optimal weights through time for HS-FP models and ...
    Benchmark models

figure()
subplot(6,1,1);
area(MVO_FFP_MSR_OT_1(1:Reb_freq:end,:));
axis([1,30,0,1]);
title('\bfOptimal HS-FP Weights with CPI state variable only ...
    through time');
legend(Sectors_name)
set(gca,'XTickLabel',{' '})

subplot(6,1,2);
area(MVO_FFP_MSR_OT_2(1:Reb_freq:end,:));
axis([1,30,0,1]);
title('\bfOptimal HS-FP Weights with GDP state variable only ...
    through time');
legend(Sectors_name)
set(gca,'XTickLabel',{' '})

subplot(6,1,3);
area(MVO_FFP_MSR_OT_3(1:Reb_freq:end,:));
axis([1,30,0,1]);
title('\bfOptimal HS-FP with EW Weights to CPI and GDP through ...
    time');
legend(Sectors_name)
set(gca,'XTickLabel',{' '})

subplot(6,1,4);
area(MVO_FFP_MSR_OT(1:Reb_freq:end,:));
axis([1,30,0,1]);
title('\bfOptimal HS-FP with DCC Weights to CPI and GDP through ...
    time');
legend(Sectors_name)
set(gca,'XTickLabel',{' '})

subplot(6,1,5);
```

```
area(MVO_MSR_OT(1:Reb_freq:end,:));
axis([1,30,0,1]);
title('\bfOptimal MVO Benchmark Weights through time');
legend(Sectors_name)
set(gca,'XTickLabel',{' '})

subplot(6,1,6);
area(EW_OT(1:Reb_freq:end,:));
axis([1,30,0,1]);
title('\bfOptimal EW Benchmark Weights through time');
legend(Sectors_name)
set(gca,'XTickLabel',{' '})
```

14.1.2 Conditional HS-FP Function

```
function Fprob=Conditional_FP_001(Conditioner,prior)
%Flexible probabilities conditioned via entropy pooling

% INPUT
% Conditioner :[struct] with fields
% Series      :[vector] (1 x t_) time series of the conditioner
% TargetValue :[vector] (1 x k_) target values for the conditioner
% Leeway:     :[scalar] (alpha) probability contained in the ...
    range, which is symmetric around the target value.
% prior       :[vector] (1 x t_) prior set of probabilities

% OUTPUT
% Fprob       :[vector] (k_ x t_) conditional flexible ...
    probabilities for each of the k_ target values

%% Code
z=Conditioner.Series;
zz=Conditioner.TargetValue;
alpha=Conditioner.Leeway;

t_=length(z);
k_=length(zz);

%Crisp Probability
p=Crisp_Prob_001(Conditioner);
p(p==0)=10^-20;

for i=1:length(zz)
    p(i,:)=p(i,:)./sum(p(i,:));
end

Fprob=nan(k_,t_);

%Flexible Probability
for i=1:k_
```

```
mu=p(i,:)*z';
s2=p(i,:)*(z'.^2)-mu^2;

%Setting up constraints for min entropy
a=(z.^2); %?
b=(mu^2)+s2;
aeq=[z;ones(1,t_)];
beq=[mu;1];

%FP
Fprob(i,:) = MinRelEntFP_001(prior,a,b,aeq,beq);
end
end
```

14.1.3 Crisp Probability Function

```
function [crisp_prob, z_lb, z_ub]=Crisp_Prob_001(Conditioner)
%Conditioning via crisp Flexible Probabilities

% INPUT
% Conditioner :[struct] with fields
% Series      :[vector] (1 x t_) time series of the conditioner
% TargetValue :[vector] (1 x k_) target values for the conditioner
% Leeway:     :[scalar] (alpha) probability contained in the ...
    range, which is symmetric around the target value.

% OUTPUT
% crisp_prob  :[matrix] (k_ x t_) crisp probabilities for each ...
    of the k_ target values
% z_lb        :[vector] (k_ x 1) range lower bound for each of ...
    the k_ target values
% z_ub        :[vector] (k_ x 1) range upper bound for each of ...
    the k_ target values

%% Code
Z=Conditioner.Series;
zz=Conditioner.TargetValue;
alpha=Conditioner.Leeway;
t_=length(Z);

z=unique(sort(Z,'ascend')');
ecdf_z=unique(HFPcdf_001(z,Z,repmat(1/t_,1,t_)));
cdf_zz=interp1(z,ecdf_z,zz,'linear','extrap');

%upper and lower quantiles of z
zmin=quantile(z,alpha/2);
zmax=quantile(z,1-(alpha/2));

z_lb=zeros(length(zz),1);
z_ub=zeros(length(zz),1);
```

```
p=zeros(length(zz),t_);
pp=zeros(length(zz),t_);

for i=1:length(zz)
cdf_zz(cdf_zz≥1-alpha/2)=1-alpha/2;
cdf_zz(cdf_zz≤alpha/2)=alpha/2;

z=zz(i);
if z≤zmin
    z_lb(i)=min(Z);
    z_ub(i)=quantile(Z,(cdf_zz(i)+(alpha/2)));
elseif z≥zmax
    z_lb(i)=quantile(Z,(cdf_zz(i)-(alpha/2)));
    z_ub(i)=max(Z);
else
    z_lb(i)=quantile(Z,(cdf_zz(i)-(alpha/2)));
    z_ub(i)=quantile(Z,(cdf_zz(i)+(alpha/2)));
end
% crisp probabilities
for t=1:t_
    if Z(t)≤z_ub(i) && Z(t)≥z_lb(i)
        pp(i,t)=1;
    else
        pp(i,t)=0;
    end
end
p(i,:)=pp(i,:)./(sum(pp(i,:)));
end

crisp_prob=p;
end
```

14.1.4 HS-FP CDF Function

```
function F = HFPcdf_001(x,epsi,p)
% This function computes the Historical Flexible Probabilities cdf

% INPUTS
% x     :[vector] (1 x k_) points
% epsi  :[vector] (1 x t_) scenarios
% p     :[vector] (1 x t_) Flexible Probabilities

% OUTPUT
% F     :[vector] (k_ x 1) cdf values

%% Code
k_ = length(x);
F=nan(k_,1);
for k=1:k_
    F(k)=sum(p.*(epsi≤x(k)));
```

```
end
end
```

14.1.5 Minimum Relative Entropy Function

```
function [p_pos, lg_] = MinRelEntFP_001(p_pri, v_ineq, mu_ineq, ...
    v_eq, mu_eq, options)
% This function computes the FP-updated distribution according ...
    to linear
% constraints on Flexible Probabilities via Entropy Pooling

%  INPUTS
% p_pri   : [vector] (1 x j_) Flexible probabilities (prior)
% v_ineq  : [matrix] (l_ x j_) pick matrix for combinations in ...
    inequality views
% mu_ineq : [vector] (l_ x 1) vector that quantifies inequality ...
    views
% v_eq    : [matrix] (m_ x j_) pick matrix for combinations in ...
    equality views
% mu_eq   : [vector] (m_ x 1) vector that quantifies equality views
% options : optimization options created with "optimoptions"

% OUTPUTS
% p_pos   : [vector] (1 x j_) Flexible probabilities (posterior)
% lg_     : [vector] ((l_ + m_) x 1) optimal parameters of dual ...
    Lagrangian

%% Code

k_ = size(v_ineq, 1);
l_ = size(v_eq, 1);

if nargin < 6  isempty(options);
    if isempty(v_ineq)
        options = optimset('GradObj', 'on', 'Hessian','on', ...
            'MaxIter', 10^6, 'TolFun', 1e-16, 'MaxFunEvals', ...
                10^6, 'Display', 'off');
    else

            options = optimset('GradObj', 'on', 'Hessian', ...
                'user-Supplied', ...
            'HessFcn', @(lv, lambda) mHessianFun(lv, lambda, ...
                p_pri, v_ineq, v_eq, k_, l_), ...
            'MaxIter', 10^6, 'TolFun', 1e-16, 'MaxFunEvals', ...
                10^6, 'Display', 'off');

    end
end

%% Code
```

```
lv0 = zeros(k_ + l_, 1); % initial point

if ¬k_ % equality constraints
    v_ = fminunc(@(v) mDualLagrangian_eq(v, p_pri, v_eq, mu_eq, ...
        l_), lv0, options);
    p_pos = exp(log(p_pri) - 1 - v_' * v_eq);
    lg_ = v_;
else % inequality and equality constraints
    % specify constraints l ≥0
    alpha = -eye(k_ + l_);
    alpha(k_ + 1 :  end, :) = [];
    beta = zeros(k_, 1);
    lg_ = fmincon(@(lv) mDualLagrangian(lv, p_pri, v_ineq, ...
        mu_ineq, v_eq , mu_eq, k_), lv0, alpha, beta, [], [], ...
        [], [], [], options);
    l_ = lg_(1 : k_);
    v_ = lg_(k_ + 1 :  end);
    p_pos = exp(log(p_pri) - 1 - l_' * v_ineq - v_' * v_eq);
end

end

function [mh, mgrad, mHess] = mDualLagrangian_eq(v, p_pri, a, ...
    b, l_)
% opposite dual Lagrangian for equality constraints
vt = v';
at = a';

p = exp(log(p_pri) - 1 - vt * a);
p = max(p, 10^(-32));
pt = p';

% dual Lagrangian
h = (log(p) - log(p_pri)) * pt + vt * (a * pt - b);

mh = - h; % value

if nargout > 1
    mgrad = b - a * pt; % gradient
end

if nargout > 2
    mHess = (a .* (ones(l_, 1) * p )) * at; % Hessian: a * ...
        diag(p) * a'
end

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y=ineqcons(x,A,b)
y=-A*x+b;
end
```

```
function [mh, mgrad] = mDualLagrangian(lv, p_pri, c, d, a , b, k_)
% opposite dual Lagrangian for inequality and equality constraints
l = lv(1 : k_);
v = lv(k_ + 1 :  end);
lt = l';
vt = v';

p = exp(log(p_pri) - 1 - lt * c - vt * a);
p = max(p, 10^(-32));
pt = p';

% dual Lagrangian
h = (log(p) - log(p_pri)) * pt + lt * (c * pt - d) + vt * (a * ...
    pt - b);

mh = - h; % value

if nargout > 1
    mgrad = [d; b] - [c; a] * pt; % gradient
end

end

function mgrad = mDualLagrangiangrad(lv, p_pri, c, d, a , b, k_)
% gradient opposite dual Lagrangian for inequality and equality ...
    constraints
l = lv(1 : k_);
v = lv(k_ + 1 :  end);
lt = l';
vt = v';

p = exp(log(p_pri) - 1 - lt * c - vt * a);
p = max(p, 10^(-32));
pt = p';

mgrad = [d; b] - [c; a] * pt; % gradient

end

function mHess = mHessianFun(lv, lambda, p_pri, c, a, k_, l_)
% Hessian of opposite dual Lagrangian for inequality and ...
    equality constraints
l = lv(1 : k_);
v = lv(k_ + 1 :  end);
lt = l';
vt = v';
ct = c';
at = a';

p = exp(log(p_pri) - 1 - lt * c - vt * a);
p = max(p, 10^(-32));
```

```

mHess = ([c; a] .* (ones(k_ + l_, 1) * p )) * [ct, at]; % ... Hessian: [c; a] * diag(p) * [c; a]'
end

function mHess = mHessianFun1(lv,p_pri, c, a, k_, l_)
% Hessian of opposite dual Lagrangian for inequality and ... equality constraints
l = lv(1 : k_);
v = lv(k_ + 1 :  end);
lt = l';
vt = v';
ct = c';
at = a';

p = exp(log(p_pri) - 1 - lt * c - vt * a);
p = max(p, 10^(-32));

mHess = ([c; a] .* (ones(k_ + l_, 1) * p )) * [ct, at]; % ... Hessian: [c; a] * diag(p) * [c; a]'
end
```

14.1.6 Effective Number of Scenarios Function

```
function ens = EffectiveScenarios(p, Type)
% This function computes the Effective Number of Scenarios of ... Flexible
% Probabilities via different types of functions

%  INPUTS
%   p       : [vector] (1 x t_) vector of Flexible Probabilities
%   Type    : [struct] type of function: 'ExpEntropy', ... 'GenExpEntropy'

%  OUTPUTS
%   ens     : [scalar] Effective Number of Scenarios

% NOTE:
% The exponential of the entropy is set as default, otherwise
% Specify Type.ExpEntropy.on = true to use the exponential of ... the entropy
% or
% Specify Type.GenExpEntropy.on = true and supply the scalar
% Type.ExpEntropy.g to use the generalized exponential of the ... entropy


if nargin < 2  isempty(Type); Type.ExpEntropy.on = true; end
if ¬isfield(Type, 'ExpEntropy')  isempty(Type.ExpEntropy); ... Type.ExpEntropy.on = false; end
```

```
if ¬isfield(Type, 'GenExpEntropy')  ...
    isempty(Type.GenExpEntropy); Type.GenExpEntropy.on = false; end

%% Code

if Type.ExpEntropy.on
    p(p==0)=10^-250; %avoid log(0) in ens computation
    ens = exp(-p * log(p'));
elseif Type.GenExpEntropy.on
    ens = sum(p .^ Type.GenExpEntropy.g) ^ (-1 / ...
        (Type.GenExpEntropy.g - 1));
end
```

14.1.7 HS-FP Mean Covariance

```
function [m, s2]=FPmeancov_001(x,p)
% This function computes the mean and covariance matrix of a ...
    Flexible Probabilities distribution

% INPUT
% x    :[matrix] (i_ x t_) scenarios
% p    :[vector] ( 1 x t_) Flexible Probabilities

% OUTPUT
% m    :[vector] (i_ x 1) mean
% s2   :[matrix] (i_ x i_) covariance matrix

%% Code
if size(p,2)==1; p=p'; end
[i_,t_]=size(x);
m = x*p'; % mean
X_cent = x - repmat(m,1,t_);
s2 =(X_cent.*repmat(p,i_,1))*X_cent'; % covariance matrix
s2 = (s2 + s2')/2; % eliminate numerical error and make ...
    covariance symmetric
```