Functional Asset Returns

Rebolledo Díaz , Julio Cesar

Functional Asset Returns

Julio Rebolledo Diaz

Durham University Business School

Durham University

A thesis submitted for the degree of

Doctor in Business Administration, DBA

2023

Abstract

Discovering the price of a financial asset is a dynamic and complex process. Based on available literature and empirical evidence there is not singular approach of achieving such a task easily. Despite current advances in technology and in access to data, a general argument in favor of approaching this problem is centered on the information available at each moment of time for an individual financial asset. Accordingly, it seems coherent to use density functions as a reference for studying relevant aspects of asset prices and subsequently equity returns. Forecasts of density functions is an active approach in decision theory and economics. The direction of my dissertation is related to the employment of density forecasting apply to asset prices.

Density forecasting may also may also of the interest for the management research areas since it provides more information than predictions produced considering only point and interval forecasts as these last frameworks yield limited sets of information.

Finding the correct true price density of a financial asset is as crucial as the characterization of it. This has implications for investors and managers in terms of both, risk management and value creation.

Despite, the acute of the underlying assumption made regarding the

properties of the statistical distribution of the future asset fluctuation over time, finding the correct (true) price of a financial asset is as crucial as it is the characterization of it.

Regarding the literature that addresses alternative methods to forecasts asset prices there are some papers that consider that one direct solution for modelling purposes will to assume that the time evolution of the asset price can be described by a random event over time. In this case, the method of choice to produce forecasts will be centered on the idea that the expected asset price will be a discrete process. An alternative discussion may be, to consider that the process that describes the path of the asset price is related to continuous fluctuations in space in which a diffusion process will have a central role. In general, my approach will be is centered on some equties traded on the S&P500. My approach is to forecast the density function of these values using functional time series relying on is principal component analysis (PCA).

Declaration

I, Julio Rebolledo Diaz, hereby declare that this is entirely my own work unless referenced to the contrary in the text. No part of this thesis has previously been submitted else where for any other degree or qualification in this or any other university.

Acknowledgements

First of all, my acknowledgments and compliments are addressed to my family, without whose support few activities in my life would have been possible.

I wish to recognize the support provided by my supervisor at Durham University, Professor Julian Williams, who always was and has been a constant and growing source of motivation and improvement for identifying topics of interest for the development of my research. I would also like to acknowledge the support provided by the Chilean Government, the University of Chile and the Chilean Central Bank, who have also contributed significantly to the fulfillment of this academic achievement.

I would also like to mention the people I met while living in the UK, including those I met and worked with while living in Bath, as their insights helped develop my academic background by motivating and pushing me to complete my finance studies.. Finally, I would like to end by saying that the achievement of this academic degree has also been motivated by my genuine interest in the financial area, which I obtain in my recurring academic and non-academic activities related to finance.

Contents

List of Figures

Chapter 1

Nomenclature

Symbol	Description
Ω	Information Set of a Probability Space
$\mathcal{F}$	Outcome Set of a Probability Space
P	Probability Functions of the Probability Space
$P_1 = P_0 x(1 + E(r_i))$	Price Formation Process
P_1	Future Price of an Asset
P_0	Current Price of an Asset
$(1 + E(r_i))$	Processes explaining the change on a "i" return
$E(r_i)$	Expected Return of an "i"
ADR	American Depository Receipt or Depository Receipt
h	Time Step
$\Pi(t) = S(t) + \Delta F(t)$	A portfolio
$r_t = \mu + \sum_{i=1}^{p} a_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j}$	Auto-regressive Process
$dP_i(t) = P_i(t)[b_i df + \sum_{j=1}^{d} \sigma_{ij}(t) dW_j(t)]$	Brownian Process
$\overline{x} = \frac{Cm + \sum_{i=1}^{n} x_i}{C+n}$	Bayesian Averaging
$\hat{y}_{t+t,t} = (\hat{y}_{t+t,1}, \hat{y}_{t+t,2}, ..., \hat{y}_{t+t,N})'$	Forecast Combination Model
$C = (\hat{y}_{t+t,t}; \omega_c) \epsilon R^c \subset R^n$	A Lower Dimensional Measure
$\hat{y}^c_{t+h,t} = C(\hat{y}^c_{t+h,t}; \omega_c)$	Point Forecast
$g(\hat{y}_{t+h,t}; w_{t+h,t}) = \frac{1}{N} \sum_{j=1}^{n} \hat{y}_{t+h,t}$	Equal Weight Model
$e^c_{t+h,t} = y_{t+h} - g(\hat{y}_{t+h,t}; w_{t+h,t})$	Loss Function
$w^*_{t+h,t} \epsilon W^c$	Optimal Combination
$\{\varepsilon^1_{t+h\|t}\}^T_{t_0}, \{\varepsilon^2_{t+h\|t}\}^T_{t_0,}$	Series of Errors
$\mathrm{L}_q(\varepsilon^i_{t+h\|t}) = (\varepsilon^i_{t+h\|t})^2, i = 1, 2$	Quadratic Loss
$\mathrm{L}_L(\varepsilon^i_{t+h\|t}) = \beta^i [exp^{(-\alpha^i x_{t+h\|t})} + \alpha^i x_{t+h\|t} - 1]$	Linex Loss
$d_t = L\left(\varepsilon_t^{h,1}\right) - L\left(\varepsilon_t^{h,2}\right)$	Diebold-Mariano Test
$\Pi(t + h)$	Density
$\mathbb{E}_Q[\Pi(t + h)] = \Pi(t) \exp(\int_0^h r(s)\mathrm{d}s)$	Risk Neutral Portfolio
$F(t, t + h) = \tilde{S}(t + h) = S(t) \exp(\int_0^h r(s)\mathrm{d}s)$	Forward Price
$g(\tilde{S}(t + h))$	Density Evaluation

Chapter 2

INTRODUCTION

2.1 Introduction

Returns are information sets used to provide support approaching corporate, investment and economic interests within the finance area. There are different models and frameworks available in the financial literature to asses this variable. These models provide predictions for those interested in making academic or non-academic inferences.

The objective of my dissertation is to explore on elements of interest related to return forecasts. For this, throughout this document I will introduce specific considerations about this variable that are discussed in papers interested on forecasts combinations and density forecasts. In addition, I will produce an empirical chapter based on the principal components analysis (PCA) framework where functional times series are central to make inferences considering the role of the factors involved in prediction.

As density forecasts will provide description for the full range of probabilities that the future values of a variable can take, my approach is to produce density forecasts elaborated within the functional space framework. Lebedev [1997] describes that in a mathematical context, a space is formed by a set of elements which includes numbers, functions, subsets, etc. Probability paces and Hilbert spaces, are two types of spaces used to make inferences. On one hand, there is the probability space which is used to form an idea about random variables, and is described by the information set Ω, by the outcome set $\mathcal{F}$ and by the probability function P which is the space where probabilities are assigned to each of the outcomes available in $\mathcal{F}$. On the other hand, there is the Hilbert space used to let functions to be the elements of the space Dette et al. [2020]. Therefore, relying on Hilbert space framework as a functional space, I am using functional principal component analysis to forecast a density function to explore returns.

Papers dealing with financial forecasting show that common approaches for single forecasts are used to study the effects of projections under structural breaks, the evaluation of predictive skills, to comparison forecast errors distribution, for volatility forecasting, copulas, and more. In addition to this there are emerging frameworks such as such forecast combinations that are also present in the literature. Producing single or combined forecasts is important as under both approaches expectations are formed and actions will be undertake base on such projections.

The inspiration of the forecast combinations framework is based on the role of information where the principle for having two or more forecasts forming a new model is considered in the literature to be a valid alternative compared to a prediction based from a single model. Although, the forecast combinations problem will be described with more detail later in this document, the central element with it is that two single forecasts will be combined through a loss function. Clemen [1989] suggested that based on the diversification principle, forecasts combinations can be accepted. Additionally he proposed that forecast combinations should be preferred to one single point forecast as the combined model has more informative ability. In summary, he indicates that single sets of information used to produce single projections have limited forecasting power.

Initiatives to improve the quality of forecasting also stand as a permanent challenge for the economic and finance areas. Therefore, the justification for incorporating a revision on the forecast combination literature is to provide support in favor of an alternative framework used to deal with such expectation. Currently the discussion regarding forecast combinations also include a consideration about the appropriate weights used in the combination and whether an equal weighting combination is a true benchmark. The general approach to determine the best combination is based on the following process. The observed stock return is forecasted and then benchmark to real observation. After this difference, between the forecast and the observed return, a variable known as the forecast error is computed, and after this error is estimated a loss function will be chosen to minimize the difference between the forecasted and the observed return. This process is an optimization procedure which operationalizes the productions of the weights that lead to the optimal combination. In this process the independent variables are the two (or more) forecasts involved in the combination problem.

There have been develops about common frameworks and approaches available to describe and to study both, stock markets and management interest such as organisations dynamics and risk managment. The existing empirical work indicates that the evidence that has been produced based on limited dependant variables, panel data regressions and multivariate modelling has limited sucess. And, as I mentioned earlier as central idea of this document is to explore forecasting asset returns and the density forecasts of these variables I will use the principal component analysis framework to extract factors and to forecast the density relying on the functional time series approach. To achieve this objective I will collect asset prices of elements of the S&P 500 index. Using this data I will process the prices in high frecuency to compute returns. I will then take 5 minutes series of monthly blocks of returns from which the methodology used will produce forecast densities. Regarding the density of data, measured by the number of transactions, I will conduct my analysis based on elements of the S&P 500 index that are traded on the NYSE platform, because they comply with two important conditions a) is the larger equity market therefore continuity in prices is more likely to be true and, b) this stocks are covered widely by market participants therefore market imperfections are less like to be present in the data.

As stated above my work is on the expected return. This variable represents the future fluctuation of an asset value. The expected return is interesting from various perspectives. For example in finance, in the branch of financial economics, this variable provides an expectation for investors future wealth. Current literature shows large amount of work that study returns, and additionally is indicated in papers, that the financial research is divided on models between models to determine equilibrium and on models where mathematical applications are used to study the time evolution and variables that have no direct economic interpretation. Hence in general, there is on one hand the financial economics approach and in on the other hand there is the financial econometrics approach to study the expected return.

In both cases there is a general form where the expected return is used, and is described by equation (2.1), where $P_{1,i}$ represents the future price of the asset, $P_{0,i}$ is the current price of an asset and $(1 + E(r_i))$ is the processes explaining the change in the asset price where $E(r_i)$ is the expected return of an "i" asset.

$$P_{1,i} = P_{0,i}(1 + E(r_i)) \tag{2.1}$$

Equation (2.1) is interesting as it describes a phenomenon of the price formation process. Several models are available to study the expected returns such as The Market Model, The Capital Asset Pricing Model (CAPM) and the Asset Price Theory (APT), these are the most common frameworks found in the literature. It is important to mention that these frameworks have led to a growing literature dedicated to explain the empirical puzzles that these models have reported.

Regarding the rest of this dissertation, the notion of return and expected return are introduced, the relationship between finance and the interests of the area of management that studies organizations and their ability to survive are discussed. Also, this document discusess aspects related to point forecast, forecast combinations, forecast breakdowns and finally, computes and evaluates density forecasts under the functional analysis framework.

Within a standard asset pricing framework, the returns are normally decomposed as follows:

$$r_{i,t} = \underset{t-1}{\mathbb{E}}(r)_{i,t} + \epsilon_{i,t}$$

where $r_{i,t}$ is the realised return and $\mathbb{E}_{t-1}(r)_{i,t}$ is the expected return conditioned on all prior information, usually referred to by Ω_{t-1}. The pricing error is $\epsilon_{i,t}$ and it is to this pricing error that the majority of focus is applied.

A simple question is how to forecast the statistical properties of the pricing error $\epsilon_{i,t}$ and hence understand the properties of $r_{i,t}$. However, as I will show in the remainder of this thesis, ii will be more straightforward to directly mode the density function of $r_{i,t}$ non-parametrically. The reason is straight forward, in a standard framework both $\mathbb{E}_{t-1}(r)_{i,t}$ and $\epsilon_{i,t}$ are uncertain. In a normal econometric specification, the joint density of $p(r_{i,t})$ is assumed to be a product of two independent distributions. However, in a fully non-parametric setting, it is not possible to completely disentangle the joint structure of $\epsilon_{i,t}$ from $r_{i,t}$. In a joint set-up this means that the copula that connects these two marginals is a unit square at maximum entropy. As this cannot be assured across the entire range of returns, then it is simpler to directly model $r_{i,t}$ and I will do this for the remainder of the thesis.

Chapter 3

ASSET RETURNS

3.1 Approaching Asset Returns

This section is dedicated to introduce the observed relationship found in the financial literature relative to various frameworks discussing how to approach asset returns. Asset prices are of central interest for all market participants, academics and practitioners. Current approaches interested in pricing an asset use the price fluctuation, or the so called return, to produce forecast rather than using quoted closing prices from regular trading days of financial assets. Therefore the return is a price transformation that allow studying the future price of an asset from the perspective of expected returns. Consequently, the dynamics of an asset price can be approached considering sets of information based on transformed prices.

In one hand is the economic interest for the expected return as it provides a measure for the expectation that investors have about their future wealth, and on the other hand is the financial interest for the expected return which based on current literature this variable is used broadly to make inferences and to make financial decisions. Financial economics and financial econometrics are two direct frameworks used to assess expected returns. Some papers show that returns are used in general solutions to describe financial puzzles and there are other papers that show that returns can be use to produce forecasts. Thus, the estimation of the return is an important line of research in the field of finance. Merton [1980] describes the expected return as a number frequently required for the solution of many investment and corporate problems. In both, theoretical and empirical financial analysis, the expected return is rather used as a tool employed to distribute financial resources as it helps in specifying the opportunity cost of the investor's choice. Additionally is used to asses, describe, and forecast many other financial problems such as the prediction of density functions which are used then to describe the probabilities of occurrence of the price of some specific asset of interest. One particular element of the expected return that is important in my research proposal lies on the idea that relates to what type of information will be used to form expectations about this value, as dealing with the return, means also to recognise in advance the character of the hypothesis under consideration. This inherently implicates leaving information out of the testing, either past or forward looking data. Hence, this is variable that has a great impact on the decision-making process faced by an investor who has a series of risky decisions from which to choose.

Observed return provide a history of the past fluctuations with the aim to find a measure capable to quantify profit and losses over a period of time. Accordingly it offers important information used to produce models that forecast future asset fluctuations from which expected returns can be derived. In addition, they can be used to reflect the state of investments decisions. Campbell et al. [1998] provide huge evidence in favour of the return as a variable that represent a measure of investment opportunities faced by economical agents, and that also have interesting statistical properties that allow making inferences about their behaviour.

Over time, available evidence on the financial literature shows many successful and unsuccessful attempts to explore alternatives to find good measures for the expected return. However, if the succeed attempt is not a relevant matter, but rather is the effort to identify a robust theoretical framework, an interesting approach is one that relates the role of the correlation. The central point is not related to how this measure of relationship is obtained, as the correlation can be derived for example directly from the variance-covariance matrix. There are papers that show that the experience about this measure of relationship does play a determinant part in early and current efforts to estimate of the expected return. This form of co-movement stands as a natural filter to select frameworks to undertake modelling potential solutions. In the past, early propositions on estimating expected returns such as those of Bachelier (Bachelier [2011]) and Charles Dow (Cottle [1960]), both used this coefficient as a base line of their propositions. Bachelier, used it to price stock options. He relied on physic principles used in quantum mechanics to describe the random behaviour of the stock price, therefore, based on the premise of no correlation among observed returns he proposed that the behaviour of particles suspended on gas (or liquid) were useful to discover the value of the option. He leveraged his proposition on a stochastic process broadly used today in financial econometrics analysis. Later in time, Charles Dow et. al., proposed a framework also based on the correlation, The Dow Theory. He claim that stock prices follow trends, therefore, asset fluctuations were correlated. Since then, several articles have appear discussing both approaches. One example is Cowles 3rd [1933] who using past information rejected the capacity of past asset prices in explaining future fluctuations. Using data from 1902 to 1934 found that the alternative of finding expected returns using historical information was ineffective. He proposed then that simple strategies formed by buy-and-hold portfolios would earn more return than Dow's approach. Another example of the use of correlation is provided by Brown et al. [1998].He revisited Dow's proposal using the same data set employed by Cowles and found evidence of strategies that yielded positive risk-adjusted returns suggesting that timing allows to achieve high Sharpe ratios and positive alphas. In addition to the correlation coefficient used to study return, other statistical properties observed are part of potential solutions on estimating expected returns. Other existing approaches designed to estimate the expected returns rely on frameworks such as, financial econometric tools, firm's valuation, firm's behavioural finance among other frameworks where a common between them is that they are all financial based approaches.

A different alternative to make inferences for returns is covered by Taylor [1986] and Andersen et al. [2001]. They describe that returns follow not a gaussian shape but a leptokurtic distribution. This stylised fact is characterised by a fat tail where the central part of the shape is well described by a Levy or a Pareto distribution. Goldenberg and Schmidt [1996] deal with the properties of models under different distributions frameworks and explain that if prices follow random walks market efficiency has to be interpreted and characterised under the hypothesis that prices follow martingale processes. On the other hand if these processes are under the assumption of continuous time models dedicated to deal with the estimation of the returns should be framed under the constrains considered within diffusion processes and look up for solutions relative to stochastic differential equations, or equivalently to Itô´s approach. Hence, some approaches are more appropriate depending on the problem under investigation and the experience described in the literature that point outs that financial econometric tools can produce conclusive analysis regarding the expected return also. Another alternative to deal with the expected return is to treat the problem from a different angle and aim to explain the variable based on the risk premium framework that is inserted into the CAPM considerations as regarding the expected return this has to be approached differently. An important point to consider in the CAPM framework is that it is originally based on the market model, which is conceived as a theoretical framework based on statistical factors that are related to the problems described by the authors mentioned just before. So, going back to the CAPM framework Derrig and Orr [2004] describe that to measure equity risk premiums one must consider the impact of geometric vs. arithmetic mean, short vs. long investment horizon, short vs. long-run expectation, unconditional vs. conditional, and real vs. nominal returns as the formation of the market return affects the estimation of the premium and leads to different conclusions.

For example, a case of an appropriate model for the expected return can depend upon information available hence if a time series of observed prices were available to estimate the variable an auto-regressive process such as equation 3.1 would be useful to find a model.

$$r_t = \mu_t + \sum_{i=1}^{p} a_i r_{t-i} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j} \tag{3.1}$$

Where μ_t is the average return in time t, a_i and b_j are parameters, r_{t-i} is the lagged return, ε_{t-j} is the lagged residual and ε_t is the residual in time t.

Although, dependency among asset returns proves to be quite a controversial topic the presence of price fluctuations clustering can be observed at least at some extent in the stock market with periods of large movements and periods with smaller fluctuations, thus, this provides sufficient evidence to support the Series [1970] approach.

Alternatively to the autoregressive processes described by equation 3.1 in which the asset fluctuation and the residual are autocorrelated, asset prices models from which expected returns can be estimated and help describing financial markets, the modelling upon the assumption that asset prices fluctuations are independent and evolve following a d-dimensional Brownian motion can be used. For example equation 3.2 is used to estimate the expected return of an asset using the CAPM framework if the asset follow a martingale processes such as the one described by Karatzas [1997], then W(t)=(W1(t),..., Wd(t))′ where W_t denotes a standard Wiener process with $W_t\sim$ N(0,t), dP_i is the a financial asset price fluctuation, P_i is the appropriate probability and σ_{ij} is the volatility of the financial asset price.

$$dP_i(t) = P_i(t)[b_i df + \sum_{j=1}^{d} \sigma_{ij}(t) dW_j(t)], \qquad P_i(0) = p_i \in (0, \infty), i = 1, ..., n. \tag{3.2}$$

The Bayesian averaging approach described by equation 3.3 also can be used to calculate the expected return where the Bayesian average $\overline{x}$ represents the expected value of the return. To compute $\overline{x}$ it is required the prior mean m and a constant C, where C is an assigned value that is proportional to the typical data set size. The value is larger when the expected variation between data sets (within the larger population) is small. It is smaller when the data sets are expected to vary substantially from one another. The general mathematical expression is the following.

$$\overline{x} = \frac{Cm + \sum_{i=1}^{n} x_i}{C + n} \tag{3.3}$$

Another alternative approach to deal with the estimation of this variable is described in the literature dedicated to deal with expected return relating the cash flows approach. This framework stands as another potential solution for the estimation of expected returns as the net preset value of the cash flows may provide a reasonable measure to estimate asset prices. The origin of this type of models is attributed to Fisher and Williams. Fisher [1930] suggested that the investment decision problem is a concept related to the time value of money and employing a discounting factor similar to the expected return can be used as a tool to quantify investor's choices over time. In the other hand, Williams [1938] extending Fisher's proposal proposed the concept of "intrinsic value" in which he highlighted that financial assets must have an intrinsic value produced by their ability to generate future cash flows. Currently the cash flows approach provide important information for market participants as it recognises the limitations of alternative models produced under restrictions related to realised prices, real data, and investor psychology. The downside of this approach is related to current research which indicates that bias consensus is observed as analysts covering companies may have incentives to increase target prices.

An element of distinction, perhaps a separation aspect on the estimation of the expected return is the consensus about the financial problem having a different scope than the economic problem. A brief summary about the economic theory would point out that the central interest of it, is that economics would be in charge of the allocation of limited resources consumed over time by agents, individuals and firms, depending on the case of the study. On the demand side of the economic context agents face the decision of trading off present consumption against future consumption, hence, an agent has to sacrifice current savings for future ones. To resolve this problem, market participants (the agents) are defined initially as being rational decision makers, henceforward, the dynamic of economic theory development follows initially onto these two conditions, a) agents always preferred more than less and b) the consumption will face a decreasing rate. On the supply side there is the firm theory where it is stated that the expected return will be similar to the term that stands for the expected payoff to be received by the factors involved in the production of goods and services, and it takes the form of wages, rents and interests.

Therefore, under the economic theory context, it is important to differentiate between the economic and the financial approach to deal with this variable because whereas economic resources are limited, financial resources are not because money can always be printed. The common element shared between financial theory and economic theory is that in finance we deal with resources consumed over time and also the central idea is to allocate them efficiently. Financial resources can be considered to be unlimited as they are constantly increasing. For example, they grow, in nominal terms and in market capitalisation terms. Thus, the importance of expected returns is their role played in the economic finance area, but not necessarily in other areas of finance such as asset pricing. Another possibility to use the expected return in finance occurs when there is valuation of firms involved, In this case, this variable is required to discount cash flows and terminal values. In summary the return is used in the asset price area of finance, and is closely link to the expected return due to expectation of an investor to receive in the future a different price than the initial price involved in the original investment choice.

3.2 Estimation of Expected Returns

It is central to note that the terms returns and expected returns, are used in the same context across this document. The objective of my research proposal is to implement a method to provide density forecasts for this variable. Having clarified that, this section covers literature in relation with the estimation of the variable just described. For this purpose, I rely on equilibrium models, behavioural finance, financial econometrics, cash flows models and forecasts combinations. Most of the topics just described have been mentioned before in the above prior section however, in this part of the document I will introduce benefits of using forecast combinations as they provide an interesting approach that holds even for the rational expectations theory. Although the forecast combinations framework is not central for forecasting the density function I find that this theoretical framework provides an alternative to estimate returns. However I prefer density functions because I could extract stylised facts from the data.

The justification for the advances in the financial econometrics and in the behavioural finance literature is related to the use of information used to estimate expected returns as the availability of new technology such as high frequency trading (HFT) have let more market participants to interact directly in trading activities.

As the classic equilibrium proposition proofs having limited capacity of explanation on the estimation of expected returns current research tests the linearity assumed within the classic approach and tests non-parametric structures to find patterns that explain conditional returns, and in addition concentrates in testing volatility, in order to develop models to price derivatives, and it is starting to find the distribution of the tails as a interesting area of study. However, there is much more effort to do in terms of developing theory and models capable to capture the structure of true returns as the essential problem in modelling expected returns is that today's approach is that is yet too simple to capture the full array of governing variables that drive asset price fluctuations across different markets [1].

3.2.1 Expected Returns and Classics Models

The literature on the use of equilibrium models to estimate the returns, highlight that this framework is an approach that relies on the central idea that for handling the problem of estimating returns timing solutions are relevant. In principle, this implies hat the expectation is to determine a fundamental quantity to use as the "correct" value for returns in the long run. Some papers provide evidence pointing out that two different frameworks can be used to estimate this

[1] Alan Greenspan. Financial Times, March 16, 2008

variable. On the one hand we have the approach that explores linear functions linking risk and reward, on the other hand there is the approach that use real data collected primarily from the macro variables used to follow the dynamics of the economy. Sharpe [1964], Lintner [1965] and Black et al. [1972] argue in favour of models that explain expected returns from the perspective of the theoretical linear relation existing between risk and returns. Ross [1976] predicted returns and proposes that the expected returns are function of market indices as they stand as a reasonable set of information from which one could estimate using simply functions to account assets prices. This framework is addressing the systematic risk is useful to calibrate the relative importance of the factor within the structure of the model. An example of this approach is the CAPM, and another example for this framework is the ATP approach. Noticeably, under current technology available to test this approach one can state that the downside of the results provided by the CAPM and the ATP models is the assumption about the systematic risk which is expected to remain constant over time or at the least for a significant period of time. Another disadvantage of this type of models is the amount of historical of data i.e. observed prices, risk free rates, inflation rates, etc required to estimate the factors that have an affect on the estimated return. These type models require long sets of information as the purpose is to find all possible significant factors affecting returns and this creates a difficulty in terms of the dynamics of the returns because these are static models and the modern approach aims to deal with real time signalling,Patton and Timmermann [2007].

On the other hand, analogue to the CAPM and ATP models, there are the alternatives for the estimation of the expected returns with the use of economic data. The argument in favour of this type of equilibrium models is that it would be preferable to estimate expected returns based on the idea that asset prices have an economic value which depends on the performance of the economy, and not solely on their relationship with financial markets. The principle of this approach is similar to the one used in the micro economic theory of the firm in which the price of the products are calculated using the factors involved in the production of the asset. Therefore, if the economy is expected to perform well, then all the firms in this economy will enjoy good performance, thereby, asset prices will experience increment in their value. Lucas Jr [1978], Cox et al. [1985], and Fama [1990] examined asset prices determination including the effect of real economic information. Lucas Jr [1978] developed a model setting as framework asset prices fluctuating stochastically in response to changes in the productivity within the real sector. He concluded that the equilibrium value of the asset fluctuate along with the economy and the expected return could be determined based on the productivity of the economy. Cox et al. [1985] complemented Lucas's contribution by augmenting his model to an inter-temporal one in order to estimate expected asset prices. Fama [1990] controls for real expected returns employing fundamental variables currently used in the standard valuation model. He develops his model combining shocks and expected cash flows, time-varying expected returns and shocks to discounting rates and finds evidence of strong relationship between production and returns.

This literature also point out the tendency of the approach to depend on the real sector dynamics to determine the value of the asset price. So, there are considerations to be made regarding a different updating dynamics observed in the economy compared to the updating process observed in financial markets. Consequently, the conclusions drawn from these expectations are used by central banks and their monetary leaders and can have an effect whose impacts vary in the financial market. Therefore, when investors have positive perceptions regarding of the future performance of the economy one could expect that financial assets would discount this favourable information hence investors will move ahead and purchase assets before the real sector actually perform in the predicted direction. In summary, it is reasonable to expect that prices of financial assets will follow different dynamics than the goods listed in the real economy like raw resources, commodities and other elements related to the production of goods and services as these prices do not follow the typical production on the goods markets. This type of equilibrium models theoretically can explain long terms returns. A modern approach finds more attractive the study of the impact of monetary policy rather than the production output, Bernanke and Gertler [2001].

3.2.2 Role of Financial Econometrics

Financial econometrics provide quantitative tools to estimate expected returns. This area of finance sets a notoriously different framework for quantitative analysis in which conditional and unconditional distributions, martingale, diffusion processes, recursive and rolling processes play determinant roles in the estimation of the return. Some papers address this approach by testing if asset prices show for example excess of volatility towards to what was predicted by the efficient markets model. Early propositions in the financial econometrics literature concentrate on the estimation of this variable by averaging returns. However, as the violations of the statistical assumptions were observed, such as the stationary condition and the time independence of returns. This framework provided evidence about the limited power of the historical average in explaining the expected returns. Therefore, early approaches provide sufficient incentive to invest time in studying other forms for modelling risk to link with returns. Engle [1982] described volatility using autoregressive models.Bollerslev [1986] augments Engle's models and includes the lagged forecast errors in the ARCH models as part of the specification. Taylor [1986] undertakes financial modelling returns time series using Markov processes and Harvey et al. [1994] and incorporated multivariate stochastic models to evaluate volatility. There is great and incremental extension in this line of research.

Naturally, the standard approach of testing linear functions used to link interest rates to compute excess of return and risk premium based on the measure of risk to expected returns has had proved to be an unstable relation, thus, giving reason for using non-linear, and nonparametric frameworks appear interesting. For example, Pesaran and Timmermann [1992] explored nonparametric setups and developed tests used in sign forecasting. Fama and French [1988] fund evidence of a non-linear component explaining mean reversion. Engle and Patton [2004] specify that volatility models should be able to forecast volatility to asses the magnitude of returns. In addition, Engle [2004] affirms that today's unsolved problem is the multivariate extension of models dedicated to understand better the dynamics of the information available in the tail of the asset price distribution. Clements et al. [2004] compare non-linear models with ARMA models and found that achieving success by the use of non-linearity approaches is conditioned on computational availability to process the data set. Likewise, they argue that more efforts regarding this quantitative framework should produce better result in forecasting variables such as expected returns. Another example is provided by

Maheu and McCurdy [2007] where specification for time-varying risk premiums are included in the equation. They use non parametric methods to model the variance of the premium which is additionally conditional to the realised volatility and are able to forecast structures. As result in their proposition they are able to forecast a structure which is based on exponential smoothing.

Campbell [2000] reviews components of valuation models by testing the performance of the discounting factor used in asset prices. He indicates that the discount factor has stochastic behaviour where most models, traditionally developed to price financial instruments, were conditioning the structure on the absence of arbitrage ignoring the stochastic movement of the discount factor. This had huge impact on some of the assumptions used by markets participants because they accepted that they might have to offset one of the properties of arbitrage used in the determination of prices when markets were relatively integrated.

Although financial econometrics is a modern quantitative approach to estimate expected returns the outcome produced from this approach is not the holy grail of finance as not always provide the only solution, however, it is accepted as benchmark solution. It is often decouple from the financial problem leading to concentrate on the properties of the original objective which is the estimation of the average expected return. Alternatively to time series data based on historical data set Cowles 3rd [1933], Williams [1938], Gordon [1962], Fama [1995], Samuelson [1973], Merton [1980], Eugene and French [1992], Black [1993], and Elton [1999] disputed employing realized asset prices arguing that historical fluctuations of asset prices were limited as to explain the problem of the estimation of the return due to the random behaviour observed in the stock price. One example is found in the early proposition stated by Cowles 3rd [1933], Fama [1995] and

Samuelson [1973] who presented empirical evidence in opposition to use historical prices as independent variables explaining expected return.Samuelson [1973] show that, observed fluctuations of stocks prices and fluctuations on stock prices adjusted by dividends, were not correlated with the future fluctuation of the same asset. He found that the observed dependence was not statistically significant. Elton [1999] suggested to estimate expected returns based on the history of variables such as treasury bills, notes and bonds. McQueen and Roley [1993] found that fundamental macroeconomic news had little effect on stock prices. He showed that after allowing for different stages of the business cycle, a stronger relationship between stock prices and news was evident. In addition to stock prices, also he examined the effect of real activity news on proxies for expected cash flows and equity discount rates and found that when the economy was strong the stock market responds negatively to news about higher real economic activity. Thus he proposed that this negative relation was caused by the larger increase in discount factor relative to expected cash flows.

3.2.3 Behavioural Finance and Expected Returns

Behavioural finance describe investor attitudes towards financial markets conditions and market imperfections. This literature suggests that market participants have certain psychological biases that affect through the short term decisions financial market preferences for financial asset.

This approach studies the expected return by incorporating concepts related to irrational investment behaviour such as over-reaction and-under reaction to news. Black [1993], Shiller [2003] and Lo [2005] argue in favour of behavioral finance as they find this framework to asset pricing being one with great potential in explaining phenomena such as financial bubbles i.e. in the evolution of prices of cryptocurrencies that currently reflect this investor behaviour. A strongest argument in favour of adopting this approach to estimate returns is that as the financial markets are constituted by investors and it is their behaviour what drives current market states, therefore, observed and expected prices must be reflecting more than just firm fundamentals or states of the economy. Precisely, asset prices contain expectations held by investors which are incorporated into the value of the asset. Conversely, Fama and French [1988] argue that this branch of finance contributes to the science only by collecting anomalies. In particular, behavioral finance emphasises the idea that markets are informationally inefficient and investors act irrationally. Black [1993] points out that it is good practice to incorporate this idea of investor psychology into approaches used to estimate expected returns. Shiller [2003] perceived behavioral finance like a crucial framework in order to test the efficient markets theory. Lo [2005] proposes a framework that reconciles market efficiency with behavioral alternatives by applying the principles found in the artificial intelligence setup where concepts such as evolution, competition, adaptation, and natural selection are used to model financial interactions. Finally, Fama [1998] discusses how overreaction and under-reaction of investors decision sensitive to news can be incorporated into the framework cover by event studies. He finds that the efficient markets theory holds, even in the long run. Therefore, anomalies will tend to disappear and prices will be capable to reflect all available information.

3.2.4 Cash Flows and Expected Return

In the previous sections it has been stated that although it limitations financial econometrics provides powerful tools to quantify expected returns, all the testing is based on estimation theory thus the construction of models is on the basis of vectors of realised asset prices. Capinski [2007] describes the positive value of current mathematical models, however. He points out the challenge that this represents in terms of the computational technology as the applications are on the basis of complex probabilistic and optimisation theories. There is empirical evidence demonstrating the capacity of past information in explaining asset prices future fluctuations. As an alternative to the model success of this econometric approach, there are arguments in favour of investing in models which do not employ only such information. Samuelson [Samuelson] pointed out that in the use of observed prices averages to find expected returns implies to acknowledge a random component in the asset price, thus, the best model for the security should be yesterdays closing price plus a drift which hopefully might be contain in the variance of the asset. Hence, it seems rewarding to invest exploring for more alternatives to estimate expected returns.

One alternative to find expected values, avoiding observed asset prices, could be to use discounted cash flows. Williams [1938] introduced the notion of intrinsic value stating that stocks prices have an implicit value which derives from futures dividends. His model treats stock as a bond which pays coupons, thus, he discounts dividends allowing the present value of the payoffs to be the expect return. Gordon [1962] rewrote the formula adding an extra component, a growing rate, making the method yet more attractive for market participants, academics and non-academics. However, despite that the motivation of their approach stands as an good alternative to price assets, in reality proofs to perform in the same direction as realised prices behave.

Black [1993] finds accounting ratios such as book equity and scaled cash flows are helpful to estimate expected returns, however, criticises the lack of theory developed to describe the effect of this variable and he links the capacity of ratios to explain expected returns to noisy signals, thus one possible explanation for this variable to obtain good results can be that when volatility is high the fundamentals of firms are captured by financial ratios Also, Fama [1998] found evidence that size, and book-to-market-equity ratios can explain stock returns. Furthermore, he view expected returns as average market returns and document the disadvantage of working with linear functions as to explain expected values arguing that the linear relation between systematic risk and average returns is not consistent. For example, the relation disappears in the period of 1963-1990 and is weak during the period of 1941-1990. Conversely, he show an alternative method to explain expected stock return based on financial ratios such as E/P, Book-to-market equity, and leverage level. Black [1993] neatly describe in his article an interesting number of arguments in favour and against traditional models developed to estimate expected returns on the basis of data and theory.

In addition, to the usual rations used to develop all this financial analysis there is another variable which currently is broadly employ by market participants, academics and non-academics, "target prices". Brav and Lehavy [2003] find that target prices provide valuable information to market participants helping to estimate expected returns. Additionally, markets participants react to this forward looking information assessing the potential value of this new information not yet contained in realised prices. Fernandez et al. [2001] and Asquith et al. [2005] estimate target prices as the product of foretasted earnings and financial ratios such as the earnings yield. There is academic literature supporting the use analyst estimates to build expectation looking for alternatives to realised returns and macroeconomic factors. Brav and Lehavy [2003] explain that expectations build on the basis of analyst estimates do not require that rational investors constrain and in addition they do not reject the ability of learning form market participants. Ang and Peterson [1985] examine the relation between expected return and dividend yield in estimating the expected return and use analyst estimates as alternative to realised returns, Botosan and Plumlee [2005] manipulate target prices to estimate expected returns and find a positive relation between the market beta and the cost of equity.

Likewise previous alternatives available to find expected returns, analyst estimates or target prices both, derived from cash flows, do not escape the criticisms. Chen et al. [2005] dedicate his article to find evidence of miss-pricing due to private information overweighting. They show that analysts have private and public information and they privilege their own information against public information when they project earnings. The possible two explanations for bias projections are (1) analyst cover firms that they know, and (2) analyst know, base on their results, that they can influence the trade and therefore commissions. Moreover, Easton and Sommers [2007], find empirical evidence that estimates of returns have upward bias, and this is also found by Bernhardt, Bernhardt et al. [2006]. They analyse the effect of herding and suggests that there is bias behaviour among analysts.

3.3 Forecast Combinations

For the estimation of the return any of the the frameworks introduced in the previous section can be used. Nevertheless, the proposition made by Clemen [1989] which is a combined variable can provide an alternative solution. In this section I provide some intuition about the forecasting combination problem, which is an alternative framework that offers a benchmark in the estimation of returns. I will also highlight the value of combining a forecast, I will introduce a general form of the combination and finally I will show the role of the loss function which is the general restriction of the optimisation problem involved in the combination. Finally, I will show briefly a problem related to the so-called deformation effect and how this issue can affect returns when they are integrated, or combined.

Clemen [1989] is in favour of combining forecasts simply because this framework increase forecast accuracy. Recent articles focuses on applying the same econometric approach to the combination of probabilities and probability distributions. Some papers suggest that combination of forecasts is preferable to single forecasts as a different forecasts framework because this approach is capable to capture different aspects of the information available. For diversification purposes combining forecasts is also accepted in the literature. Another argument in favour of this method is that individuals forecast may be affected by structural breaks caused, for example, by a changes in the rating of the the firm, therefore and accordingly some individual forecasts can adapt more rapidly to changes in the data than others. Another argument in favour of forecast combination is that individual forecasting models may be subject to misspecification bias of unknown form due to the true data generating process. In addition another argument for combination of forecast is that the underlying forecasts may conditioned for different loss functions.

The information set at time t is denoted as F_t and comprises an N-vector of forecasts such as express in the following notation 3.4 by $\hat{y}_{t+t,t}$;

$$\hat{y}_{t+t,t} = (\hat{y}_{t+t,1}, \hat{y}_{t+t,2}, ..., \hat{y}_{t+t,N})' \tag{3.4}$$

where the lower dimensional measure "C" takes the following 3.5 form;

$$C = (\hat{y}_{t+t,t}; \omega_c) \epsilon R^c \subset R^n \tag{3.5}$$

where "C" is the aggregator that reduces the information from high dimensional vector forecast to lower dimensional forecast, and ω_c are the parameters (weights) associated with the combination.

The initial evaluation of the combination deals with simple point forecast. Potentially an alternative of future research includes density forecast. Thus, the point forecast expression of the combination could be described by 3.6

$$\hat{y}^c_{t+h,t} = C(\hat{y}^c_{t+h,t}; \omega_c) \tag{3.6}$$

where the parameter ω_c could be time-varying. A special case of the combination $g(\hat{y}_{t+h,t}; w_{t+h,t})$, is 3.7 when the weights are equal:

$$g(\hat{y}_{t+h,t}; w_{t+h,t}) = \frac{1}{N} \sum_{j=1}^{n} \hat{y}_{t+h,t} \tag{3.7}$$

3.4 Specification of the Loss Function

Timmerman [2005] finds that forecast combination represents a good solution for forecast purposes and simply for diversification purposes this method is valid, and in addition he argues that a simply loss function L(et+h) such as the following which is only depended on the forecast error from the combination, thus

$$e^c_{t+h,t} = y_{t+h} - g(\hat{y}_{t+h,t}; w_{t+h,t}) \tag{3.8}$$

where the parameters of the optimal combination follows;

$$w^*_{t+h,t} \epsilon W^c \tag{3.9}$$

,

Defining an appropriate loss function across each instantaneous point in the functional range of the asset prices, would as first glance appear to be the primary goal of the exercise. It is relatively straightforward to see that as the objective value changes then the degree of importance also changes.

However, this is a almost certainly a market participant function and hence the range of values of these functions will need to be extracted for data, given the forward looking conditional density. Our objective is therefor to identify a function that connects the distribution of prior returns to the forward return non-parametrically. It is this approach that I will move onto in the next chapter.

Chapter 4

FORECASTING

4.1 Introduction to Forecasting

Discovering the future price of a financial asset is a complex process. In current literature there is empirical evidence pointing out to various alternatives to approach this objective. One alternative, broadly used is the Arrow-Debreu framework. This is an approach used to study the dynamics that an asset follow under the probability theory framework. For this a probability space triple is considered. Therefore a set of data Ω, a set of outcomes $\mathcal{F}$ and a set of probabilities P linked to each outcome is necessary. Under this setting prices financial securities can be modeled as continuous random processes, therefore plausible states prices can be extracted from the density function. This can be achieved as the price of the asset can be decomposed as the weighted sum of all its state prices. Hence, all the state prices should provide the information required to form a density. The Arrow-Debreu security is an important concept in finance as it is used as a framework to price financial assets. This approach suggests that the security price has unique states. This is the state price of the security. At each state the security, agrees or promises to deliver a previously determined payoff. Thus, there is a single payoff for each singular state.This payoff will occur at previously specified period in time. The state price vector, being a collection of data, represents the state price of all possible states.

For my research interest, it is central to forecast the future path of a previously observed asset price. Therefore, this dissertation is focus on asset price forecasting via density function forecasting. The scope of my work is to explore on observed stock prices that are part of the S&P 500 index. This is data that will help in forming expectations about the future value of some of the asset that are part of this market. Density forecasting can be used to forecast derivatives prices , equities & foreign exchange returns, yields on bonds and/or macroeconomic time series.

Forecasting density functions is an active aspect in decision theory and economics, the direction of my dissertation is centered on the asset pricing does not preclude the posibility od using density forecasting as help to other management research areas as it provides greater range of information compared to point and interval forecasts.

Finding the correct true asset distribution of the financial asset is cruccial for investors but its characterization relies on very specific assumptions regarding the underlying distribution. Relying on assumptions has importnat implications for investors and managers in terms of both, risk management and value creation. Greenspan [2004] points out that under uncertainty conditions the probability distribution of outcomes is unknown, and under risky conditions the range of outcomes is delimited by a known probability distribution. Granger and Poon

[2002] propose that volatility and risk are not the same. Risk acts as a measure of uncertainty, therefore restricts investment decisions and the creation of portfolios as decision makers set levels of risk threshold at which they will react, similar to what a reaction functions would do. In evaluating investment risks of a given holding period volatility forecasts for asset prices is a useful risk management tool.Borovička et al. [2016], indicate that the investor risk aversion element is expressed by stochastic discount factors that include compensations for risk exposures.

Literature provides different approaches to do financial analysis that can be related to the forecasting of asset prices. One approach could be to start from a direct solution assuming a random event over time therefore producing the forecast would be centered on the idea that the expected asset price can be described as discrete process. On the other hand, the alternative is to consider that the process that describes best the path of the asset price if the future is a continuous fluctuations in a space thus diffusion processes would be more appropriate.

Therefore providing empirical evidence relating forecasts produced for a security, or for an index of securities is important and technically challenging undertaking. For example, for the forecast of the physical process, only a single realized path could be observed. However, multiple evaluations of different time intervals of this path can provide evidence regarding of the effectiveness of a model generating density forecasts. Many models can provide statistically similar evaluations. White [2000] deals with data snooping and finds that several competitive models can be produced using the same data set.

Forecasting expected returns relates conditional and unconditional distributions, martingale, diffusion, recursive and rolling processes. All these considerations play a determinant role in the estimation of this variable. Recent research also includes the study of the excess of volatility relative to what would be predicted by the efficient markets model.

One example, is the following consider a risk neutral processes. Under this approach those pricing processes formed by the price of a delta neutral portfolio of hedging instruments and their underlying security, make the complexity of the problem harder. Because it requieres volatility forecasting the and the conditional distribution of the conditional undelying. This is can be due to a. the existiance of several ways to evaluate the variability around the reference numéraire b. the time frame used for the evaluation of the diffusion and the frequency of portfolio re-balancing, as these are endogenous to the forecast exercise. Hence, under the same example, consider the forecast of the risk neutral density function for a stock index. In this case we might have a depository receipt (ADR) for the index and a continuum of potential derivatives contracts such as options and futures. For a given time step h, $\Pi(t) = S(t) + \Delta F(t)$ is a delta hedge portfolio. The objective is to determine the density of $\Pi(t+h)$. The risk neutral portfolio for t to $t+h$ is the solution to $\mathbb{E}_Q[\Pi(t+h)] = \Pi(t)\exp(\int_0^h r(s)\mathrm{d}s)$, where $r(t)$ is the current return.

The forward price is there denoted by $F(t, t+h) = \tilde{S}(t+h) = S(t)\exp(\int_0^h r(s)\mathrm{d}s)$.

One approach could be to start from asset market equilibrium and the oberved price could be the solution for this equilibrium, and following a secuance of random events over time the idea is that the expected price would retrun to its original state. The alternative is to consider is in the future is continuos fluctuation

Hence for an evaluation of a density $g(\tilde{S}(t+h))$, we have some choices in the definition of h, in particular as $h \to 0$, then the replication strategies approach the continuous limit. Similarly, there are a number of choices in terms of the types of hedging instruments one might use to determine forward price, not all will agree not consensus

. Furthermore, the implied forward rate is a factor in the determination of the density function.

4.2 Forecasting economic and financial data

Forecasting economic and financial time series such as macroeconomic variables, financial asset returns, interest rates, exchange rates and derivative prices, is an important area of financial research. Considerations in regard to the stability of the predictions and about the intrinsic and extrinsic degree of predictive capacity of forecast models is central to discuss this topic.

Some papers on the financial econometric literature forecasts indicate that expected returns were reviewed through first-order conditions. Engle [1982] described volatility relying on autoregressive models. Bollerslev [1986] enhance Engle's approach by including ARCH affects in the formulation of the modelling. Taylor [1986] used Markov processes, and Harvey et al. [1994] employed multivariate structures to study the stochastic behaviour of risk. These examples in the literature document how the preliminary efforts gave way to the current interest in the topics of interest related to forecasting.

The simple question is whether a time series variable can be forecasted with a relatively narrow prediction error given the right model or if it is inherently uncertain across some interval structure. The relationship between model structure and predictive power is broad, prior work such as Martens et al. [2002], Liyan and Chenli [2002], Pong et al. [2004] Koopman et al. [2005], Clement and Tse [2005] Kumar and Thenmozhi [2014], Dolinar [2014], Haugom et al. [2014] and Sen and Ma [2015] has focused on the 'in-model' or 'between-model' comparison of the first two moments of the time series of interest between model by comparing the first two moments of the time series of interest of the forecast error

. Canonically, for financial and economic applications, the Box-Jenkins approach, see Makridakis and Hibon [1997] for a comprehensive summary, has dominated the literature. Here some form of auto-regressive model transmits shocks through time via the conditional mean. Extensions to the ARCH-GARCH realm extended this framework into conditional second moments. The resulting effectiveness of this type of approach has been reviewed in a vast array of research across numerous settings, in economics and finance, an in-exhaustive set of examples can be found in Ahmed [2001], Fleming et al. [2001], Martens [2001], Hueng and McDonald [2005] Pérez-Rodríguez et al. [2005], Petkova [2006], Balaban et al. [2006], Albuquerque et al. [2008] Choudhry and Wu [2009], Lu and Perron [2010], Dhamija and Bhalla [2010], Zhu and Galbraith [2011], Kim et al. [2011], Westerlund and Narayan [2012] and Weiß and Supper [2013] amongst others. The performance, of this type of model is mixed and subject to the choice of the evaluation criteria. A number of regression based approaches have used the Neyman-Pearson metholodgy to evaluate forecast accuracy.

Typically, this is in the form of coefficients of determination or chi-squared restriction tests. Examples of this approach can be found across a range of forecasting scenarios in Guo and Savickas [2008], Bollapragada et al. [2009], Choi et al. [2010], Verma [2011], Benavides and Capistrán [2012], Tripathy and Rahman [2013], Siburg et al. [2015] and Wang et al. [2016]. In each case a parametric or semi-parametric forecasting model is specified and a set of forecasting evaluation metrics, usually with respect to the quadratic variation about some stable or unstable mean are established. The use of the R-squared (or R^2) is controversial. Forecasting accuracy can be broken down into three components the inherent uncertainty in the underlying stochastic process driving the variables of interest, the model specification structural breaks and estimation errors and the inherent error in underpinning modelling framework selection.

Discussion of the interplay of first two types of uncertainty that is the model framework is correct; however, estimation is conducted with error and this error structure maybe correlated to the inherent uncertainty posed by the stochastic process driving the observed time series is discussed in Pesaran and Timmermann [2002], Ferson et al. [2003], Almeida and Vicente [2008], Frijns and Margaritis [2008], Roodposhti and Amirhosseini [2011], Singh and Ahmad [2011] and Zakamulin [2013]. Richer approaches involving model averaging and uncertainty in model selection have been discussed more recently in Jordan et al. [2014], Wenjun and Bangsheng [2017], Vortelinos and Gkillas [2018] and Faria and Verona [2018].

In this context would be relevant to test test for the quality of the projection, detecting the forecast breakdowns, studying in-sample and out-of sample forecasts, and looking at theoretical benchmarks that consider normality, linear, non-linear, asymmetric and symmetric conditions. White [2000] observed that many models are admissible in-sample and Timmermann [2018] proposed alternatives to divide the sample into individual subsets, in-sample and out-of-sample, to deal with the forecast evaluation.

The process of producing forecasts starts by collecting the information that will be used to perform the analysis. The process of producing predictions then continues with the selection of candidate models, and it ends with the diagnostic tests used to evaluate the forecasts. The data can be either, observed (realized) or forward-looking such as are end-of-day prices.

Forecasting involves the consideration of available frameworks like in-and out-of samples exercises, one-step-ahead point forecasting, linear and non-linear models, interval forecasting and density forecasting. The literature also extends over elements involving multi-step-ahead point forecasting based on iterated one-step forecasts, direct multi-step forecasts, interval and density forecasts. It is also a subject of interest dealing with forecast breakdowns in the mean, and in the variance-covariances. there are cases of multiple breaks and structural changes, and different tests estimation models studing the breaks. However of you have different models you can alivate the study by the use of forecast combinations is also an important consideration for the research undertaken in forecasting.

Pesaran and Timmermann [1992] explored non-parametric setups and developed sign tests to study the performance of the forecasts. Fama and French [1988] provided evidence for the a non-linear component capable to explain mean reversion. Engle and Patton [2004] detailed how volatility models could be used to forecast the magnitude of returns. Engle [2004] suggested that currently an unsolved problem is related to the multivariate extension of models seeking to find the dynamic behavior of long tails. Clements et al. [2004] compare non-linear models with ARMA models and found that the success of non-linearity depended on the availability of technical advances related to computational capabilities.

Maheu and McCurdy [2009] developed a nonparametric method used to study the variance of the term premium that is additional to the realized volatility. They proposed a forecast setup based on exponential smoothing. Campbell [2000] tested the behavior of the discounting factor used to price assets. Harvey et al. [2016] listed different factors of risk related to stocks forecasting.

In recent years forecasting based on point estimates has been displaced by a growing concern relating to more informative forecasts, such as interval and density forecasts. Density forecasts and point forecasts provide different outcomes in order to form expectations about the future. Density forecasts suggest that for a random variable at some future time the forecast of the density would offer a description of the full range of probabilities that the possible future values of the variable will take. The intermediate assumption between both approaches is the prediction interval approach. Under this framework the expectation is that the probability of the value of the random variable will fall within a stated range. One central consideration regarding interval forecast is the adding point of standard deviations of the point forecast in a symetric maner. Definitely, is more limited than the density forecast however it is preferable to work under this type of assumption than the point forecast approach. By enlarge interval forecast are considered to be symmetric bands, therefore, the assumption of normal or gaussian distribution are applied to this approach. Anscombe [1967], suggests that as that the normal distribution is "too good not to be true"' and using student's t-distribution can be considered to be a good alternative. He points out that is the variance of the forecast error is a key element in the forecast.

Dealing with point forecasting requires on the one hand to consider the rules used to rank the estimates and on the other hand the uses and applications of this approach. Gneiting [2011] proposes that the scoring functions that can be used to compared and assessed point forecasts are; the squared error (SE), the absolute error (AE), the absolute percentage error (APE) and the relative error (RE). Elliott and Timmermann [2016] provide arguments about this forecasting problem. They propose that the point forecasting analysis is similar to the statistical problem of estimating a parameter of the conditional probability distribution of the outcome. Regarding applications of the point forecast method Duffee [2018] employed this approach to explore the relationship between nominal bond yields and inflation. He looked for the role of future inflation on the nominal yield in the context of dynamic equilibrium macroeconomic models using quarterly data in order to asses the impact of the different shocks to nominal yields.

Regarding approaching forecast based on interval forecasting Granger et al. [1989] indicate that interval forecasting would generally require the estimation of upper and lower limits associated with a predefined probability. Chatfield [1993] pointed out that there is not a general approach to estimate prediction intervals, and the that the only exception for this non general rule of thumb is to proceed by fitting a probability model with some variance of forecast errors that can be properly evaluated. Christoffersen [1998] proposed a different approach described as the conditional coverage criterion. Under this assumption he approached the evaluation of a sequence of interval forecasts and suggested that this was an estimation that can be achieved with good levels of confidence. He pointed out that relaxing the previous knowledge of the underlying model the loss of confidence was not too costly for the estimation. Giacomini and Rossi [2009] follow Christoffersen [1998] approach and augmented it by introducing a framework in which quantile forecasts were exploited successfully in the context of risk management.

4.3 Forecast Evaluation

Since Diebold et al. [1991], forecasting has drifted from comparing models to compare forecasts. This change in the approach to produce forecasts has impacted the center of interest hence current literature on forecasting focusess on the choice of the loss function and therefore of the model. In this lineGiacomini and Rossi [2010],Bollerslev et al. [2016] and Timmermann [2018] described how unstable forecasts are due to low signal-to-noise ratio, persistent predictor bias, overfiting and model instability.

Relative to forecast evaluation Giacomini and Rossi [2009] evaluated forecasts considering the loss of the forecasting ability by splitting the data into two subsets end use exchange rates series of prices. To asses this approach the Diebold-Mariano test compare the forecasts fro each of these two subsets and the residuals obtained from each sample were compare. The Diebold and Mariano [2002] test suggest that under the null hypothesis there is a Diebold-Mariano statistic which has an asymptotic standard normal distribution. The relevance of this statistic is important due to the long run variance used by the the test.

Clements and Taylor [2003], provide a different approach for interval forecast evaluation which is particularly useful in the presence of periodic heteroscedasticity. He suggest that his proposal was capable of detecting inadequacies in interval forecasts generated by traditional methods of modeling high frequency asset returns such as in the GARCH model approach. Anjum and Malik [2020] observed that when volatility is clustered, a VaR model that ignores mean-volatility relationship, may yield the correct unconditional coverage but may give poor conditional coverage. In other words, an interval forecast that does not account for higher-order dynamics may give correct results on average (i.e. having correct unconditional coverage) but may give incorrect conditional coverage in a given time period because the outliers may be clustered.

For financial markets, the very nature of the accuracy of forecasting is inherently linked to the absence of arbitrage argument and it is to this construction that I will turn to next.

4.4 Derivatives, Arbitrage and Forecasting

Derivatives prices are considered to be forward looking. Forecasting density function using these prices is promoted in the literature as a beneficial alternative to form expectations. By using the concept proposed by the Arrow-Debreu security and following the Black and Scholes framework risk neutral density functions can be forecast. Therefore, the application of this forecast is very interesting as the resulting density function can be perceived as an ideal or well behave benchmark for the future prices derive from the probabilities extracted from the foretasted density.

Derivatives are financial assets widely traded, and in their nature they are characterized by being perceived, for academics and non-academic market participants, as contracts. The terms stated in the contract in principle are used to price the derivative. Accordingly, derivatives data contributes to my research because markets participants use this contract for speculation, arbitrate and/or hedging strategies. The previous point is confirmed by the Black and Scholes framework which is a standard approach to study derivatives. This model uses the specifications explicit in the contract to price the derivative. For example, one of the features specified in contract is the expiration date. This provides to the investor a specific time frame to use the contract, because after the expiration date the contract would have zero value. Hence, one knows that only before the expiration date the price of the contract would fluctuate. Another characteristic of the contract refers to the underlying asset on which the derivative has been produced. The settlement value of the derivative is a function of an underlying security. And, the value of the underlying asset is subject to change due to systematic and non-systematic information.

For the case of derivatives, the approach is centered on values of financial asset prices that have not yet occurred. Therefore, to forecast density functions using derivatives prices consists testing based on observed option prices which effectively are considered forward looking data. The relevance of this type of data sets is that compared to the commonly used end-of-day data of the asset prices, options data provide an approach that account for information not disclosed nor incorporated yet in the asset price.

Arbitrage, is a market condition used to price options. Involves the principle of the law of one price. By conviction such market condition will disappear rapidly and generally this will happen because in the presence of two different prices for the same asset, these different values will converge to each other. Therefore, a no arbitrage condition should prevail in the dynamics of the price update.

Part of the literature suggests that such a discrepancy would exist exclusively due to the presence of transaction costs, and common sense or the law of one price also suggests the same. Under the Arrow-Debreu securities framework, a unique state price vector does have to exist, or else there would be arbitrage, thus there is arbitrage if the vector of state prices is overdetermined. If there is arbitrage, linearity of the neoclassical problem implies that any candidate optimum can be dominated by adding the arbitrage Dybvig and Ross [2003].

An arbitrage condition is an assumption that is arguable because it has been observed. Arbitrage opportunities provide an argument to support such a discrepancy. Hence, the arbitrage condition is available if there is an arbitrage opportunity, Lépinette [2019] . This condition allows to develop strategies which generally involve the risk free rate. Hence, and although this may be arguable for different reasons , my approach starts off with the a no arbitrage condition .

Arbitrage free is a setting used in financial modelling for purposes of testing as let risk-free excess returns and market anomalies out of the range of outcomes. Among practitioners or non-academic this point of view it is arguably a solid starting point for testing a methodology under review, yet it offers a clean benchmark. In my proposal this assumption allows me to observe the volatility that results from the derivative contract.

Granger and Poon [2002] point out that to price a derivative it is required to know the volatility of the underlying asset until the contract matures, the tenor. Additionally, they show the advantages and disadvantages from a series of models used to evaluate volatility which have been categorized between time series models and option ISD (implied standard deviation). Such a separation is important because the second category of models is relevant in my proposal. Broadie and Jain [2008] describe that stochastic volatility is used to price options since it has the capacity to fit skews and smiles, while simultaneously covering considerations related to the risk involved in the greeks.

Prior work such as Britten-Jones and Neuberger [2000], Claessen and Mittnik [2002], Ferris et al. [2003], Cairney and Swisher [2004], Panigirtzoglou and Skiadopoulos [2004] and Bernales and Guidolin [2014] have looked at the risk neutral densities of assets derived from options prices and attempted to reconcile this mean adjusted density with the observed dispersion of asset returns. Further work by Battalio and Schultz [2006], Mixon [2007], Becker et al. [2007], Mo and Wu [2007], De La Bruslerie and Deffains-Crapsky [2008], Muzzioli [2010], Taylor et al. [2010], Wayne et al. [2010], Shackleton et al. [2010], Chalamandaris and Tsekrekos [2010], Constantinides et al. [2011] and Chalamandaris and Tsekrekos [2011], have looked exclusively at the second moments pricing the option and the degree of predictability of these measures of dispersion. Higher moments, above two, have been analyzed in Brown and Robinson [2002], Kim and Lee [2013], An et al. [2014], Yun [2014], Guo and Qiu [2014], Lin and Lu [2015] and Fricke et al. [2018]. The evidence on the effectiveness of these approaches is mixed. With a broad variety of results indicating that central tendency is hard to measure with any consistency, but volatility appears to have more predictability and a narrower set of modelling frameworks to review.

Equity markets price valuations on future accumulated dividends and capital appreciation, currency markets have a very different connection to their forecasts as these are typically the valuation of future interest rates and inflation, I will address these next.

4.5 Exchange Rates Forecasts

Empirical studies dealing with exchange rates forecasts show how difficult it is to exceed predictions associated to the ex-ante best individual forecasting model.

There is some evidence in favor of the in-sample forecasting capacity although consensus in academic literature is not quite certain to whether such results are consistent out-of-sample. For example, Meese and Rogoff [1983]find that economic and monetary aggregates are not capable of producing good forecasts of future changes in the spot rate, Rossi [2006] shed light on the nature of the lack of forecast ability of models based on economic variables concluding that studying parameter instability contributes to produce better forecasts instead of those generated by random walk predictions.The exchange rate disconnection puzzle which studies why economic aggregates are not capable of producing good forecasts it is also well documented in the literature. For example, Bacchetta and Van Wincoop [2006] look at the exchange rate determination puzzle and by analyzing order flow confirm that fundamentals can explain little of exchange rate movements in the short and medium run, an addition they suggest that over longer horizons the exchange rate follows closely the fundamentals and finally the propose that the exchange rate is related to order flow over both short and long Rime et al. [2010] argue in favor of the link between order flow and macroeconomic information as the first reflects heterogeneous beliefs about macroeconomic fundamentals

Another issue related to exchange rates forecasting are the statistical tests applied to the data used to make such projections. To date, the evidence suggests that the tests applied to the time-series of the exchange rates under study used to explore and understand the data can play an important role in order to produce good forecasts, especially now with recent findings suggesting that the nature of the true generating process of the data can be non-linear. For example, Sweeney [2006] detected that systematic movements of nominal rates for the G-10 countries towards stable long-run equilibrium are in fact non-linear. Thus given the unit root test applied looking at the stationary condition necessary to accept the mean reversion, attention needs to be paid to the generating data process. Sarno [2001], looking at the behavior of US public debt, finds that the US debt-GDP ratio can be described with a non linearly mean-reverting stochastic process and proposes the typical ADF unit root test cannot pick the non-linearity of the process, this finding can also be apply to exchange rates knowledge.

In addition to the tests applied to the data looking for assistance in producing good forecasting models it is also common knowledge in the literature that some other considerations need to be taken into account to make forecasts. Crownover et al. [1996] and Lothian [1997] agree that one important consideration relates the evidence in current research showing that the short-term dynamic of the exchange rate follows a I(1) process and the long-term exchange rate equilibrium can be described reasonably well by I(0) process. The implications of specifying I(0) or I(1) processes lead the observer to different model specifications; thus while the I(1) forecasts look to predict the short-term dynamic of the exchange rate the I(0) assumption aims to study the projection from the mean reversion perspective.

4.6 Forecasting Considerations

Perhaps one the first consideration dealing with forecast combinations is related to obtaining spurious results. There is certainly great deal of interest in exploring the effects that statistical assumptions such normality and the stationary condition could have on the forecast ability and on the stability of the forecast. This is not trivial since the obvious principle of forecasting is to predict. Therefore, due to the lack of exact rules used to make or produce forecasts, it is possible to anticipate that there may be incentives to relax statistical assumptions in favor of obtaining reasonable predictions. An example of this is the model-free principle. Politis [2015] describes that this assumption allows to start the forecast process without a model.

A second aspect or consideration is based on the Diebold [2012]. In this article he noted that the Diebold-Mariano (DM) test was developed to compare forecasts and not intended for comparing models. In the same article it is clarified that the essence of the DM approach is to assume intentionally that the forecast errors are primitive, to then be able to make assumptions directly about those forecast errors. In the model-free framework, for simplicity, there are no a priori models to consider. This last point is fundamental in the bases of the DM tests since these have as logic to avoid the settings associated with the Model-Free Prediction Principle.

To date, the quality and quantity of information related to the prices of financial assets causes the models used to make predictions to rapidly lose predictive ability. This implies that new models are required. As a consequence of this there will be other potentially successful models. This could explain the existence of competing models. Thus, regarding a model in general is expected that it does not produce spurious results and that at the same time it makes good predictions. Both elements of doubt are important in the analysis of the forecasts. And this duality can be approached from the data processing perspective. Whitel, White [2000],deals with data snooping, Hansen and Timmerman (2018) address sample size. White studies a phenomenon of the data set, from which several competitive models can be produced among themselves, Hansen and Timmermann [2012] discuss how to split the sample into two individual set, in-sample and out-of-sample, both asses to explore forecast evaluation.

4.6.1 Forecasting Steps

Step One

Engle and Brown [1986] described that model selection is a procedure used to explore the forecast accuracy. Rao et al. [2001] described that in regression analysis and times series analysis some interesting possibilities available are the following: model selection based on hypothesis testing, based on the prediction errors, based on information criteria, based on cross-validation and on bootstrap methods, and finally based on bayesian approaches. Mehdiyev et al. [2016] approach this problem by looking at different accuracy measures such as root mean squared error, relative absolute error, average accuracy rate, f-score, f-measure and f-value among others. However, if the approach of choosing a model is based on an information criteria, then, the model selection test would be dedicated to investigate the stability of the model over time.

Some of the relevant information criteria described in the literature are Theil Information Criterion (Adjusted R2), Akaike's Information Criterion (AIC) and Bayes-Schwarz Information Criterion (BIC).

- R-squared (R^2)

$$1 - \frac{SS_{eer}}{SS_{tot}} \tag{4.1}$$

  where, SS_{eer} the sum of squares error and stands for the variance of the forecast errors and SS_{tot} is the sum of squares error and it stands for total variance of the data.

- (AIC)
$$log(MSE) + \frac{2K}{n} \tag{4.2}$$
where, MSE stands for the mean squared error, K represents the number of parameters of the model and n stants for the number of observations of the the data.

- (BIC)
$$log(MSE) + \frac{log(n)K}{n} \tag{4.3}$$

Step 2: Choosing Data Generating Processes

The initial point in all inference approaches is to begin by collecting information about a population. The data collected is a sample. This sample will provide observations, and the investigation, the analysis and the testing over this data set will provide conclusions about the population. For my research proposal I will illustrate an application of tests for factor construction from the 100 most actively traded constituents of the S&P 500 index. I will then use the number of principal components selected by the test in a cross sectional model for all traded stocks in the data file available. The hypothesis is that the pair-wise correlation of the resulting residuals of these principal components are viable pricing factors. This application has the objective to estimate returns used to forecast orthogonal density functions of the S&P 500 index.

My proposal considers to take advantage of properties available from these frameworks to compare results for the principal components mentioned as the bootstrap and asymptotic approximation can be used as alternatives to make inferences about the S&P 500 index as for other financial asset as well. The justification to use this approach is provided by Dovonon et al. [2022] where is pointed point out that this tests have been investigated by an extensive Monte Carlo simulation study where several data generating processes have been considered as well as different sampling frequencies and small and large cross-section dimensions.

In this document I will also provide alternative examples relating exchange rates as the set of information. In this case I have had explore forecasts dedicated to analyze some considerations regarding the forecasts such as forecast breakdowns. Regarding this approach, the distribution follow by exchange rates cab be controversial as including features related to empirical distributions such as FXs have to take in consideration relevant in information contained in the tails specially if they are considered "fat tails" as it suggested by Tu and Zhou [2004] because this could improve the quality of the conclusions.

In general, the central incentive to use data generation processes is the expectation about the results attained from this approach. As the data is expected to be similar to the observed results produced by rational analysis derived, for example from an equilibrium model, such the bootstrap and asymptotic methodology would allow me to study some empirical implications involved in the forecast. Summarizing, with the data a generating process that I am proposing I will obtain simulated distributions of the data that allow to forecast density distributions.

Some other approaches used to make predictions about exchange rates are available in Meese and Rogoff [1983]and Cheung et al. [2004] where they look at structural models such the flexible-price monetary policy (Frenkel-Bilson), the sticky-price monetary (Dornbush-Frankel) and the Hooper-Morton model, and compare results with the random model. Also in this line Neely and Sarno [2002] test the forecasting power of flexible monetary policy fundamentals.

Step 3: Choosing Diagnosytic Tests

The diagnosis is used traditionally to distinguish between two types of events, "signals" and "noise." However, due to progress in both, computational power and literature, forecasting currently also is concerned about structural beaks. Since Demski and Feltham [1972] the advances that have been made in terms of the forecast evaluation shows an interest dynamic. That is to say, it is possible to permanently find new contributions regarding this issue. In their article, they propose that a forecasting method can be interpreted as an informative two stage process where in the first stage signals are produced and the second stage the signals are processed.

In-sample and out-of-sample forecast analysis are both used to evaluate models. Although out-of-sample tests are certainly better standard then just in-sample testing, Clements and Hendry [2005] supported this idea, however, as they evaluated non-statonary processes they pointed out that analysts must account of misleading results. They studied economics models and found that although out-of-sample forecast can have reasonable/good performance they can also be counter intuitive in terms of the underlying economic theory. This is a strong indication that even is a model is out-of-sample accepted there is always more possibilities nested in other methodologies not being considered at the time.

For Giacomini and White [2006] and Giacomini and Rossi [2009] out-of-sample tests are used for evaluating and selecting among different predictions dedicated to address both, economics and finance problems. However, for me it is important to implement diagnostics on financial variables, which could be, for example, the exchange rate.

In this sense, assessing the out of sample forecasting performance of financial data is relevant for my research design. An important part of the current literature is based on the diagnostic tests suggested by the Diebold and Mariano [2002] where they study the dependence of relative performance of a certain information set at a given point in time. Giacomini and White [2006] condition the test on a set of covariates, enabling a test for possible variation of relative performance over time.

The structure of out-of-sample tests is described by Tashman [2000] . He mentions that issued related to the rolling window are relevant. Also differentiates that updating is different tan recalibrating. In addition, he talks about the diversity of models referring to the reality of competitors that the best model could have. Another point in his article is related to pooled averages to calculate the forecast error and the stability of the error. Finally he describes that the technology involved in the forecasting meaning that the quality of the software package is relevant.

Regarding forecasts breaks Li et al. [2017] deal with a forecast model that deals with equity premium, inflation, exchange rates and the Treasury bill interest rates. They used a non-linear non-Gaussian state space framework and find that allowing the parameter being flexible they can r gains in forecasting accuracy.

4.6.2 Forecast Combinations and Forecast Breakdowns

Forecast combinations from parametric modesl have been the subject of considerable attention in recent research. Since Clement (1989) this area of research has gained general acceptance based on encouraging results reported in the literature, such as Pesaran and Pick [2010], and Granger et al. [2006]. Both of these papers report that combinations of different forecasts typically result in more accurate projections than single forecasts. While Granger et al. [2006] suggest that equal weighting seems natural for the mix of the combination they also finds it interesting to explore on the idea of working with an optimal weight combinations different than the the following case. Consider the resulting forecast from the same variable from two models, $\lambda 1 = \lambda 2 = 0,5$ where stand for the weight in the mixing, an alternative could be $\lambda 1 \neq \lambda 2$ where λ also stands for a particular weight in the mixing however their argument for not equal weighting is based on the idea that individual forecasts, whose parameters are estimated recursively, could be affected by the choiche of the weight itself if the combination of the weights differs from case where equal. Pesaran and Pick [2010] study forecast combinations and look at averaging forecasts over different estimation windows. They finds this approach being capable of generating forecasts reasonably robust to structural breaks.

In the econometric literature a structural break is defined as a change in the parameters of the system. Ericsson et al. [1998] indicates that the structural break occurs when the parameters of a conditional distribution are time variant. The importance of this issue relates to the properties of the framework proposed by Giacomini and Rossi [2009] where they show that forecast breakdowns can be caused by instabilities in the data generating process itself (this relates to the properties of their forecast breakdown test which is used in this paper). The implication of the presence of breaks is also discussed in Granger et al. [2006]. They suggests that this issue needs to be taken into account, arguing that the instability in the data generating process can affect the estimation of the weights as it may cause under performance relative to that of the best individual forecasting.

4.6.3 Detecting Forecast Breakdowns

Giacomini and Rossi [2009] introduced forecasts breakdowns as a formalization of the deterioration of the forecasting performance in the out-of-sample forecast. Their suggestion is very much along the lines of a large literature looking at forecasting exchange rates in which the performance of the forecast in-sample cannot be reproduced out-of-sample. They propose a recursive scheme, which is adopted their this paper, to compare forecasts in-sample with forecasts out-of-sample where they assess the surprise loss defined as the difference between the out-of-sample loss with the average in-sample loss. In this paper they explore on this idea but consider the Diebold-Mariano tests to compare the forecasts.

In ranking forecast the Diebold-Mariano test states that if $\{yt\}$ denotes the series to be forecast and $y^1_{t+h|t}$ and $y^2_{t+h|t}$ are two competing forecasts of y_{t+h} based on Ω_t then the forecast errors from the two competing models can be described by equations (4.4) and (4.5),

$$\varepsilon^1_{t+h|t} = y_{t+h} - y^1_{t+h|t} \tag{4.4}$$

$$\varepsilon^2_{t+h|t} = y_{t+h} - y^2_{t+h|t} \tag{4.5}$$

Furthermre h-steps forecasts are computed from t,T with $t = t_0, ..., T$ and the series of errors are shown bellow,

$$\{\varepsilon^1_{t+h|t}\}^T_{t_0,}, \{\varepsilon^2_{t+h|t}\}^T_{t_0,} \tag{4.6}$$

The accuracy of the forecasts are measured using the *Linex (L)* loss and the *quadratic(q)* loss functions. Then they are evaluated by computing following forms:

$$\mathrm{L}_q(\varepsilon^i_{t+h|t}) = (\varepsilon^i_{t+h|t})^2, i = 1, 2 \tag{4.7}$$

$$\mathrm{L}_L(\varepsilon^i_{t+h|t})^1 = \beta^i[exp^{(-\alpha^i x_{t+h|t})} + \alpha^i x_{t+h|t} - 1], i = 1, 2 \tag{4.8}$$

In order to find which model produces more accurate forecasts the tests sets the null and the alternative hypothesis respectively as,

$$H_0 : \mathrm{E}[L(\varepsilon^1_{t+h|t})] = \mathrm{E}[L(\varepsilon^2_{t+h|t})]$$

and,

$$H_1 : \mathrm{E}[L(\varepsilon^1_{t+h|t})] \neq \mathrm{E}[L(\varepsilon^2_{t+h|t})]$$

With the null of equal predictive accuracy given by;

$$H_0 : \mathrm{E}[d_t] = 0$$

$$d_t = L\left(\varepsilon^{h,1}_t\right) - L\left(\varepsilon^{h,2}_t\right) \tag{4.9}$$

[1] In Christoffersen and Diebold (1997) the optimal h-step-ahead predictor under linex loss solves $min_{\hat{y}} E_t\{b[exp(a(y_{t+h} - \hat{y}_{t+h})) - a(y_{t+h} - \hat{y}_{t+h}) - 1]\}$

The Diebold-Mariano test statistic under null hypothesis follows asymptotically the standard normal distribution. Because the long run variance of the forevcast error is used by the computation of the statistic in the test. Harvey et al. [1997] note that this test, based on the loss differential, is likely to produce good comparison of competing forecasts. Thus the loss differential is given by,

The is rejected with a significance level of 5% if Λ_τ >1.96 (4.13) , where the composition of the estatistic is,

$$\Lambda_\tau = \varsigma_\tau^{-1} \mu_\tau \tag{4.10}$$

with,

$$\varsigma_\tau^2 = 2 \sum_{j=0}^{\tau} \operatorname{cov}\left(d_t, d_{t-j}\right) \tag{4.11}$$

and,

$$\mu_\tau = \tau^{-1} \sum_{t=1}^{\tau} d_t \tag{4.12}$$

$$\Lambda_\tau \sim \mathrm{N}\left(0, 1\right) \tag{4.13}$$

where, ς_τ^2 (4.11) is the variance of the two residuals know as the loss-differencial and μ_τ (4.12) is the average loss-differential.

4.6.4 Preliminaries and Notation

Althoug the DM is the most pupular crteria i would like to introduce an alternative measure for forecast evaluation. Let $f^c(x)$ be a density function, such that $f^h : \mathbb{R} \to \mathbb{R}_+$ and is constrained such that $\int f^c(z)dz = 1$. Our objective of interest is to compare an ex-ante forecasted density $c = a$ with and ex-post realized density $c = p$. Let $\tau \in \mathbb{N}_+$, be a discrete time index and $t \in [0, T]$ be a continuous time index, such that $t_\tau - t_{\tau-1} = h$ is a constant and $t_\tau > t_{\tau-1}$, $\forall t_\tau \in [0, T]$. This indexation permits a simple switching between continuous time and discrete time and identifies a specific stratification approach to any given price process.

4.6.5 The Admissible Loss Function

For our purposes we consider a pair of density functions in tuple $[f^a(x), f^p(x)]_\tau$. Let $\mathscr{L}(f^a(x), f^p(x))$ be a loss function of the form $\mathscr{L} = \int_Z L(f^a(z), f^p(z))dz$, that is we define some cumulant $L_q(.)$ that determines a norm for the metric space between $f^a(x)$ and $f^p(x)$. This space could be represented as a simply polynomial function difference, akin to a Euclidean norm, or a spectral norm, such as a Laplace or Fourier distance. In the simple case of comparing two forecasts the a simple case is $L_2(f^a(z), f^p(z)) = (f^a(x) - f^p(x))^2$, a Kullback-Leibler style loss function would be as described in (4.14):

$$L_K(f^a(z), f^p(z)) = \log(f^a(x)) - \log(f^p(x)) \tag{4.14}$$

4.6.6 The error structure of the functional

For simplicity opf notation, let $G(z) = L_q(f^a(z), f^p(z))$ be the compound function describing the difference between density functions $f^a(z)$ and $f^p(z)$. I denote the infinite and d order Taylor expansions of $G(z)$ as follows in (4.15):

$$G^*(z) = \sum_{n=1}^{\infty} \frac{G^{(n)}(z_0)}{n!}(z - z_0), \quad \text{and } G^{*,d}(z) = \sum_{n=1}^{d} \frac{G^{(n)}(z_0)}{n!}(z - z_0) + R(z) \quad (4.15)$$

Where the remainder term $R(z)$ is case specific to the chosen loss function.

The expansion of $G(z)$ allows the scholar to evaluate a number of loss function types in terms of the derivatives of the expansion and the number of trailing non-zero terms. For example, a quadratic loss function that has two non-zero derivatives and no remainder, as the choosen polinomial would be exactly approximated by itself whilst the chooice if the hyperbolic function may result in a large remainder.

Chapter 5

FUNCTIONAL RETURN DENSITY

5.1 Introduction

As previously noticed in the above chapter, forecasting is an area of research that deals with varying volatility and co-variances, macroeconomic variables, asset returns, interest rates, exchange rates and derivatives, therefore is crucial for economic, financial and management decision making. Considerations regarding the stability of the predictions, about the predictive capacity and the certain possibility of using combined forecasts, currently are central topics are for my research. Additionally, is important testing the quality of the projections, detecting forecasts breakdowns, studying in-sample (data snooping biases) and out-of sample forecasts, also is relevant looking at theoretical benchmarks extracted from cases build under normality, linear, non-linear, asymmetric and symmetric conditions.

The process of studying forecasts begins with the definition of the data that will be used, where the information can be observed or realized, or it can be generated by a certain process of data generation. Then it continues with the selection of a candidate models that will be extracted from a set of forecast models. Finally, once the competitors are obtained, it is necessary to choose appropriate diagnostic tests to evaluate the predictions.

To select the candidate(s) to be used in the forecast, a single or combined model approach can be used. This implicates the initial node that will determine whether the whole research will be about dealing with single or combined forecasts is not trivial. As mentioned earlier, in the forecast combination the literature it is suggested that combining forecasts improves accuracy over individual forecasts. Clemen [1989] indicated that this is due to the principle of diversification as combined forecasts are preferable to individual forecasts. Hibon and Evgeniou [2005] noted that the advantage of combining forecasts is not that the best possible combinations will perform better than the best single possible forecasts, but the combined forecast is expected to be less risky, therefore, in practice to combine forecasts should be a better approach than just finding conclusions from an individual forecast.

5.2 Related Asset Management Problem

The asset management problem link with this document it's related to common interests between organization ecology and finance. Both, organization ecology and financial research are interested about why some firms fail and why others do not. Whereas financial papers approach this interest from the corporate finance perspective like capital structures and dividend policy, among other aspects, the organizational ecology literature approach this problem looking at aspects such as age dependence, age inertia, density dependence, density delay, propagation strategies and niche operations related to resource partitioning and innovation. Age and inertia, are found in current financial literature therefore there is evidence in favour for using these variables to study the relationship and the dynamics regarding organizations. Some other similar approaches to deal with organization ecology and financial research are of interest to equity markets and organisations dynamics. Statistical frameworks used to connect both areas are principal component analysis, limited dependant variables models, panel data regression and multivariate modelling to study organisational dynamics.

In relation with the scope of the organizational theory Hannan and Freeman [1989] describe this area of management as being dependant on ecological levels and population ecology. While, population ecology is dealings with dynamic changes across a defined set of organizations, the analysis of the ecology levels explores birth and death rates, interaction between populations and communities of populations that share similar characteristics. Baum and Amburgey [2002] point out that organizational ecology explains how social, economic and political conditions affect resources and diversity of organizations and consequently take in these changes into account and their variability over time requires the demographic tools available in order to study the effects of organization-level characteristics on rates of organizational change and failure.

As age is a big predictor of organizational future longevity study of age dependence constitute an important field of enquire to estabish the effect of organizational aging on firm failure. Thornhill and Amit [2003] relate failure to bankruptcy and employ principal components approach to analize the data. Age inertia looks at how organizational age influences the ability of firms to adapt. Literature on organizational age dynamics and organizational ecology suggests that old organizations generally have a higher level of reproducibility. Hannan and Freeman [1984] support this idea and suggest that old firms have larger formal and hierarchical structure and bigger size, implying that organizational inertia increases monotonically with age. Morgan [2006] propose that organizations as organisms and self-directed entities are able to adapt to the environment and adjust their strategies to counteract the inertia that age promotes. Density dependence, look at how the number of organizations in the population (firms) affected by increasing competition. According to density dependence theory, population density controls the population-level processes of legitimation and competition. Lomi and Larsen [2001] suggest that density delay deals with the competitive conditions that outline the observed change in the organizational population. Propagation strategies according to Péli and Masuch [1997] study the behaviour of first movers and their efficiency in terms of production. The evidence shows that first-order predicate logic, and game theory are frameworks used to model this aspect. Niche strategies look at what affects the decision of specialist and generalist firms. Hannan and Freeman [1984] study the variety of organizations looking at new forms of organization. Greve et al. [2006]study the resource partitioning theory and suggest that as large firms or generalists concentrate, marginal resources should be more available for specialists. One direct implication of resource partitioning is that if the creation of opportunities is significant the anti-mass movement of the production can explain the partitioning. And finally, Innovation, deals with the determinants of an organization's propensity to innovate with focus on organizational structure.

Age, size and inertia are similarities that organizational ecology and financial research have in common to study the future of firms. In the broad sense age is important because is a good predictor of organizational longevity, consequently it should be a good predictor of the continuity of the firm operations. That forecast (the continuity of the firm operations), must be very significant knowledge for firm officers and for investors as it will allow them to anticipate degrees of firm failure that can lead to bankruptcy, thus it will introduce tactical and strategic actions on behalf of high officers in order to prevent the full discontinuity of the firm operations. A firm failure represents a firm change in the status and is the relationship to be predicted.

Spiess and Affleck-Graves [1995] study the long run performance in stock returns employing controlling variables such as firm's age and size and show that these variables are used as standard practice in financial research. Faccio et al. [2011] compare large shareholder diversification and corporate risk using a sample of public European firms and find that firm riskiness declines with age. You and Zhang [2009] researched the role of information uncertainty in price continuation anomalies and cross-sectional variations in stock returns and finds that greater information uncertainty leads to relatively lower future stock returns following bad news and relatively higher future returns following good news suggesting that uncertainty delays the flow of information into stock prices using firm´s age as control variable. Amburgey et al. [1990] look at organizational change because contrary to inertia theory predictions, and under a particular definition of change, they found that discontinuous environmental change was not associated with an increased probability of organizational change. They begin their article describing that change has a variety of definitions thus they suggest that the scope of the research requires clear definitions.

Fama and French [2012] look at size, value, and momentum to study stock returns. Momentum rather than being a control variable in their case represents an independent variable used to explain and forecast the asset price fluctuation. This form of inertia measures the speed of the change in the stock price and leads to estimate the length of periods where price fluctuates. Thus, studying the dynamic of the oscillator, meaning its stability, is crucial to the understanding and interpretation of price behaviour. Another example of the use of this idea is provided by Jegadeesh et al. [2022] where they used momentum to compare the performance of stocks. They look at stocks that accumulated positive and negative return over different time frames providing them with a framowrk to discuss the role of transactions cost in financial strategies.

5.3 Functional Time Series

The contribution of this thesis is in forecasting the density function of expected returns rather than forecasting the moments of their distributions. Alternatives to this approach are univariate time series and multivariate time series of pint forecatss. Under the univariate approach one just orders the observations in synk with time and make inferences about the population. Whilst under the multivariate approach when information comes in diverse time intervals to proceed with the estimation a single operational time interval has to be chosen and this can potentially cause loss information. Alternative the functional time series approach lets the estimation start from a model free framework in which every observation turns into a function which describes the structure of the data and one can forecast the density based on this setup. Therefore, the multivariate approach has to deal with high-dimensional time series analysis, the univariate framework deals with observed data and the functional time series approach deals in making inferences about the structure of the time evolution of the data using a function by relating the past observations in mostly non linear functions. At such, the approach to produce density forecast occurs in a functional space.

5.3.1 Preliminary Notation

Let $p(x, t+h|\mathcal{X}(t), \mathcal{Z}(t), \mathcal{P}(t))$ be a density function at timestep $t+h$. Where $\mathcal{Y}(t)$ and $\mathcal{Z}(t)$ are functional histories of some variables $x(t)$ and $y(t)$, with $\mathcal{P}(t)$ being the history of density functions. We assume that $p(.)$ has the normal properties of a density function, that is $\int_{-\infty}^{\infty} p(x, t+h|\mathcal{X}(t), \mathcal{Z}(t), \mathcal{P}(t))\mathrm{d}x = 1$ and $p(x, t+h|\mathcal{X}(t), \mathcal{Z}(t), \mathcal{P}(t)) \geq 0 \quad \forall x \in (-\infty, \infty)$.

A related problem is understanding the stability of the cumulative density function. This is denoted by $\mathscr{P}(a, x) = \int_{-\infty}^{a} p(x, t+h|\mathcal{X}(t), \mathcal{Z}(t), \mathcal{P}(t))\mathrm{d}x$. Let $P^*(a, x)$ be an approximation of $P(a, x)$ such that the norm $\int_{-\infty}^{\infty} |\mathscr{P}^*(a, x) - \mathscr{P}(a, x)|^2 \mathrm{d}a \leq \varsigma^2$, where ς is an arbitrary threshold. An obvious example of this approach can be given by the realized empirical cumulative density f unction, defined as follows, let $\boldsymbol{x}$ be a vector of data with elements x_i, indexed by $i \in \{1, \ldots, I\}$

$$\mathscr{P}^*(a, x) = I^{-1} \sum_{i=1}^{I} \mathbb{1}_{x_i < a}$$

EXPLAIN EQATION

for some interval of $a \in \mathbb{R}$. Setting $P_t(x, a, b)$ to be the evaluation of $\mathscr{P}^*(a, x)$ for some arbitrary time step $t \in \{1, \ldots, T\}$ then a first pass approximation is to construct an expectation of $\mathbb{E}[\mathscr{P}^*(a,x)]$ non parametrically.

5.3.2 Simple Example of Autoregressive Density Forecasting

The first example, considers a block model for monthly forward densities, using an AR(p) set up. Let $\boldsymbol{x}_t$ be the vector of monthly five minute returns such that $x_{j,t} = \ln S_{j,t,T} - \ln S_{j-1,t,T} \equiv s_{j,t,T} - s_{j-1,t,T}$, hence in this case h is a monthly calendar window and $j \in \{1,\ldots,J\}$ is the five minute business time (therefore excluding evenings, weekends and holidays).

Example data set: Apple Inc. Monthly Five Minute Returns

As a starting example of forecasting density functions I shall start with high frequency equity returns. First I, compute for irregular business time returns the equivalent five minute return. In this case I have best-bid best offer tick-by-tick data for Apple Inc. As noted above this is given by the set of prices $\boldsymbol{S}_{t,T}$, where t is the start of the month and T is the end calendar date. Within this month, there is the business time, this is the time when the US stock market is trading Apple Inc. stock. Business time is indexed by $S_{j,t,T}$, with the natural logarithm of $S_{j,t,T}$ denoted by $s_{j,t,T}$. We also have the calendar time ticktimes in the vector $\boldsymbol{\tau}_{t,T}$ with each time index set to $\tau_{j,t,T}$. Five minute equivalent returns are computed by $\Delta_5 \times (s_{j,t,T} - s_{j-1,t,T})/(\tau_{j,t,T} - \tau_{j-1,t,T})$. Typically, $\tau_{j,t,T}$ is measured in fractions of a day or year. Δ_5 is a multiplier, such that when τ is measured in fractions of a day, the multiplier is $5/(24 \times 60)$, if τ is in annual fractions then it is $5/(24 \times 60 \times 365)$,

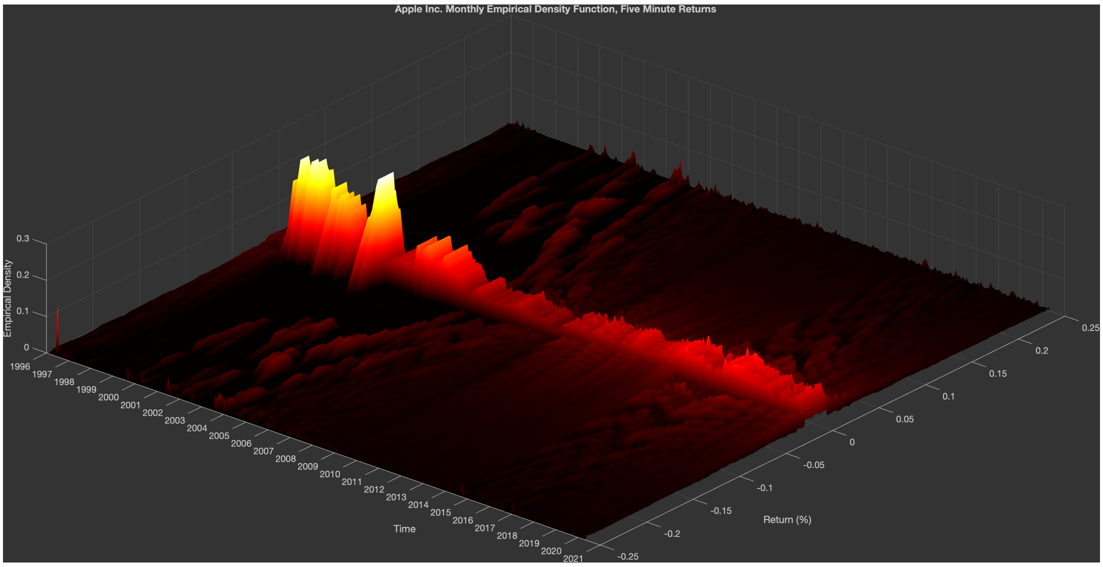

Figure 5.1: Calendar Month Five Minute Business time Return Densities For Apple Inc.

note that we are not working in trading time so the year is 365 not 252 days, as we are not computing a volatility where the free diffusion is only during the calendar time. I then apply the non-parametric approach to cumulative densities outlined previously and recover the probability density via numerical differentiation of the smoothed density function

Fig. 5.1 displays the density surface for Apple Inc. for five minute returns for each calendar month from January 1, 1996 to April 30, 2020. The griding and collection of the data is consistent with Dovonon et al. [2021] who use mixed calendar time scales to analyze moments and comoments.

The contribution of this thesis is to asses approaches realated to combine quantile and functional forecasting to determine the predictability of returns and density functions.

5.3.3 Risk of the indexing strategies

The risk-neutral density function for an underlying security is a probability density function for which the current price of the security is equal to the discounted expectation of its future prices, Monteiro et al. [2008]. Under this framework it is assumed that the change detected in the price an the option is due to the change of the underlying security. This is know as the delta in the greeks.

Hence the risk of the indexing strategy is related to the estimation of the transition matrix. The next section introduces this risk. Determining the transition matrix of a discrete Markov process from sequential forecasts of smoothed density functions is an important element of many problems in decision theory and economics. Recent theoretical results have demonstrated that the Perron-Frobenius eigenfunction of a Markov risk neutral state price transition matrix has an interesting economic interpretation and could permit the extraction of physical forward pricing densities from options markets. Yet, the application to actual market prices is challenging. For instance, even at the intraday frequency, option market panels contain substantial gaps and can contain unpredictable levels of noise across strike prices and tenors, making meanigful computations rather challenging.

5.3.4 Building a functional residual

Similarly to a scalar residual, a functional time series can have a stochastic residual functional. There are a number of specifications of this type, most of them have scalar parametric functions. Consider a process of the form:

$$\boldsymbol{\theta}_t^* = F(\boldsymbol{\vartheta}_{t-1}^*, \boldsymbol{\vartheta}_{t-2}^*, \ldots, \boldsymbol{\vartheta}_{t-p}^*) + \boldsymbol{\epsilon}_t^* \tag{5.1}$$

with the residual component being of a parametric form with $\int_{-\infty}^{\infty} p(x, \vartheta_t) - p(x, \vartheta_t^*)^2 dx$, being the minimal distance between the ex-post and ex-ante density functions $p(x, \vartheta_t)$ and $p(x, \vartheta_t^*)$. This is the typical functional approach, where $F(.)$ is a linear or non-linear function and the functional process is driven by this scalar-vector process. However, in my case I am interested in extending the problem set to a fully generalized autoregressive functional density data generating set-up.

$$p(x, t, t+h) = G(p(x, t-h, t), p(x, t-2h, t-h), \ldots, p(x, t-qh, t-h)) \times E(x, t, t+h) \tag{5.2}$$

Where $E(x)$ is a random functional.Whilst this set-up looks complicated, in a bayesian set up the analytical construction is relatively straight forward. Similarly to the parametric form. In this case to I seek to minimise the geometric distance between $p(x, t, t+h)$ and $G(p(x, t-h, t), p(x, t-2h, t-h), \ldots, p(x, t-qh, t-h))$.

5.3.5 Simulation Studies

The usefulness of a fully non-parametric functional forecasting approach is that it can approximate a large number of parametric and non-parametric data generating processes. I will start with specifying a parametric data generating process with a specific shape to the transition density function and test the ability of the non-parametric procedure to estimate that density function against a correctly specified estimator for the true underlying process.

Let $S(t)$ be the time t instance of a stock index, such that $S \in \mathbb{R}_+$. Set the holding return from t to $t+h$ to be $R(t,t+h) = \log S(t+h) - \log S(t)$. The instantaneous return as $h \to 0$, admits the following Griglionis decomposition:

$$\mathrm{d}S(t) = (r(t) - q(t))\mathrm{d}t + \sqrt{V(t)}\mathrm{d}W^S(t) + \mathrm{d}N(t)J(t) \tag{5.3}$$

with the following volatility equation:

$$\mathrm{d}V(t) = \kappa(\theta - V(t))\mathrm{d}t + \sigma\sqrt{V(t)}\mathrm{d}W^V(t) \tag{5.4}$$

with $W^i(t+h) - W^i(t) \sim \mathcal{N}(0,h)$ for $i \in \{S,V\}$ and $< W^S(t), W^V(t) >= \rho$. The jump component is a compound Poisson point process where $\mathbb{P}(N(t-h) - N(t) = n) = (1/n!)(\lambda h)^n \exp(-\lambda h)$ and $J(t) \sim \mu_{\mathcal{J}}, \sigma_{\mathcal{J}}$.

The transition density function for this type of process is well understood and can be used to estimate the underlying process via standard parametric maximum likelihood estimation.

5.3.6 The Simulation

Even at the intraday frequency option market data there is substantial gaps in realized financial asset prices. These issue offers unpredictable levels of noise across strike prices and tenors from which forecasts are made.

The simulation process allows to fill such gaps. The simulation analysis used in my proposal follows Ross [2015] as it provides consistency of the algorithm that will produce a transition matrix. This matrix allows to identify and to obtain the probabilities required to perform the forecasts that my proposal is after.

My methodology pursues to forecast risky financial asset returns. Under the general framework in my proposal the algorithm used for this objective estate prices are required. These prices, as it was mentioned before, will move in the future and the transition matrix, which holds the state prices, gather the possibilities that the path of the fluctuation of the asset can hit.

The simulation process considers that the forecast horizon is $h_{i,j} = 1$. Therefore, the short-term forecast is produced only for for the next period. This implies that there is no martingale component in the stochastic discount factor. This is an important assumption as if a martingale process would take effect in the discount factor, this last element would not be linear.

After several repeated trials in my proposal I start with a physical probability transition matrix $\mathbf{P}$ with 13 states. Therefore, it is presume that the correct number of states is known a-priori, hence the simulation conditions are assumed to be correctly specified.

5.3.7 Evaluation Criteria

My evaluation criteria will be based around a pseudo out-of-sample function R^2, with the following formulation:

$$R^2 = \frac{\sum_j \int_{-\infty}^{\infty} (\ln p(x, t + jh, t + (j-1)h) - \mathbb{E}[\ln p(x, t + jh, t + (j-1)h)|\Omega_t])^2 dx}{\int_{-\infty}^{\infty} (\ln p(x, t + jh, t + (j-1)h) - \mathbb{E}[\ln \bar{p}(x)])^2 dx} \tag{5.5}$$

where $\bar{p}(x)$ is the unconditional density function of x. Notice here we have two integral errors of the log-likelihood, hence we can think of it as the functional R^2. To compare the effectiveness of the parametric and non-parametric techniques I will utilize a numerical quadrature approach to estimating the error structure.

Quadratic loss over the functional allows the econometrician to estimate an R^2 and then compare this to the benchmark from the maximum likelihood estimation from the correctly specified data generating model (in this case a jump diffusion stochastic volatility model).

It is relatively straightforward to show from the Feller condition (explicar) that as the jump volatility σ_J and the volatility of volatility σ increase then the degree of variation in the transition density from t to $t+h$ at some arbitrary time scale will increase at least quadratically. The mean reversion θ in volatility is also important, but in practical usage in stocks and FX, this is normally held to be close to 5.

Table 5.1 presents a preliminary set of results for the ratio of the integrated function R^2 for the simulated dataset versus the non-parametric estimator. Hence my initial evaluation looks at three sets of experiments. First, I hold σ_J constant at 2% and vary σ for small (1%), medium (5%) and high (10%) levels of volatility. Next I hold σ at 5% and vary σ_J, again with low, medium and high discontinuous volatility's of (0.5,3 and 5%). For each of the cases I generate one month of simulated trading data at 30 seconds, 1 minute, 5 minute and 15 minute frequencies. I then compare the ratio of the out of sample functional R^2 for the transition density from an estimated parametric model using the standard transition densities.

The denominator is the non-parametric estimator hence when the ratio is less than one, the non-parametric estimator dominates the parametric estimator. We can see that for all of the estimates the functional density from the correctly specified parametric model is better (as would be expected), however, the non-parametric model, for higher frequencies is within 10% of the parametric model.

σ	2%	2%	2%	1%	3%	5%
σ_J	1%	5%	10%	5%	5%	5%
30 seconds	1.1089	1.2371	1.0744	1.0822	1.1571	1.1865
1 minute	1.2371	1.3478	1.3759	1.3759	1.3680	1.3975
5 minutes	1.2117	1.1924	1.3871	1.3319	1.3924	1.3989
15 minute	1.2802	1.3944	1.3852	1.3979	1.3565	1.3983

Table 5.1: Comparative functional R^2 with change in time frequency and noise structure. Continuous Volatility of Volatility σ (annual) and discontinuous Volatility σ_J, results from 1,000 repeated simulations.

A more rigorous test for both models will be to estimate the parametric and non-parametric models on a model generated from data that is not of the same specification as the parametric model. As these simulations take many hours on the HPC facilities, I will be conducting this as the last part of my work for a single stock listed in the NYSE.

My next chapter it is to construct the difference between the actual densityes an their fprediction for one period ahead and evaluate the performance of the functional approach acrross a wide range of returns, this would allow me to establish the predictive preformance of this method not only near the central tendendy but also as predictor of extreme values.

Chapter 6

Empirical Examples

6.1 Introduction

Even though financial asset prices fluctuate over time often as a reaction caused by a random event, they can reflect a combination of investors' risk aversion and the probability distributions used to assess risk Borovička et al. [2016]. In dynamic models, investors' risk aversion is expressed by stochastic discount factors that include compensations for risk exposures. Asset valuation becomes a matter of randomly discounting payoffs under different states of nature and weighing them according to the agent's probability structure about future prices.

Forecasting is an area of research that deals with varying volatility and covariances betweeen macroeconomic variables, asset returns, interest rates, exchange rates and derivatives, therefore it is crucial for economic and financial management. Considerations regarding the stability of the predictions, about the predictive capacity and the possibility of using combined forecasts are currently central topics in financial reserach. It is important testing the quality of the projections, detecting forecasts breakdowns, stuying in-sample (data snooping biases) and out-of sample forecasts, and looking at the theorical benchmarks extracted from cases build under normality, linear or non-linear, asymetric and symmetric model specifications.

The process of studying forecasts begins with the definition of the data that will be used, where the information can be observed or realized, or it can be generated by a certain process of data generation. Then it continues with the selection of candidate models that come from a set of forecasting models. Finally, once the competitors are obtained, it is necessary to choose the diagnostic tests to evaluate the predictions.

To select the candidate (s) to be used for forecasting, a single or combined model approach can be used. This choice is important becasue it will determine whether the whole forecasting effort will be about dealing with single or combined forecasts. The literature suggests in favor of the forecast combination because combining forecasts improves accuracy relative to individual forecasts. Clemen (1989) indicates that this is based on the principle of diversification and combined forecasts are preferable to individual forecasts. Hibon and Evgeniou (2005) noted that the advantage of combining forecasts is not that the best possible combinations perform better than the best individual forecasts, but that it is less risky in practice to combine forecasts than to select amongs individual forecasting method.

The possibility of developing the density function of asset prices is based on the functional relationship between their derivatative price and the corrsponfing probability. Or it

Density estimation is the estimation of the probability density function of an unobserved random variable **X**, and is based on observed data, Kadir and Brady [2005] suggests that this analysis generally fall into one of the following three categories, parametric, non-parametric and semiparametric.

A random variable **X** is defined by having certain cumulative distribution function **CDF**, or $F(x) = P(X \leq x)$, Parametric techniques are suitable where a particular form of function can be assumed due to some application specific reasons. For example, Rician and Rayleigh functions are often used in ultrasound signal processing applications.

Of the non-parameteric techniques probably the simplest and most widely used method is the histogram. Its limitations are, principally the requirement to define the number of bins, the arbitrary bin boundaries and the block like nature of the resulting PDF estimate have lead to the development of a number of alternative methods. Parzen windowing avoids arbitrary bin assignments and leads to smoother PDFs, however, a suitable kernel shape must be chossen and size must be chosen. It has been noted that conventionally this choice has been somewhat arbitrary and largely driven by aesthetics [8], although some work has been done on defining systematic methods for selecting kernel sizes; see [1, 10, 7]. Other, more exotic methods, such as Wavelet density estimators [2] have also been been proposed. Semi-parametric techniques such as Gaussian Mixture Models offer a useful compromise between these two approaches whereby the superposition of a number of parametric densities are used to approximate the underlying density.

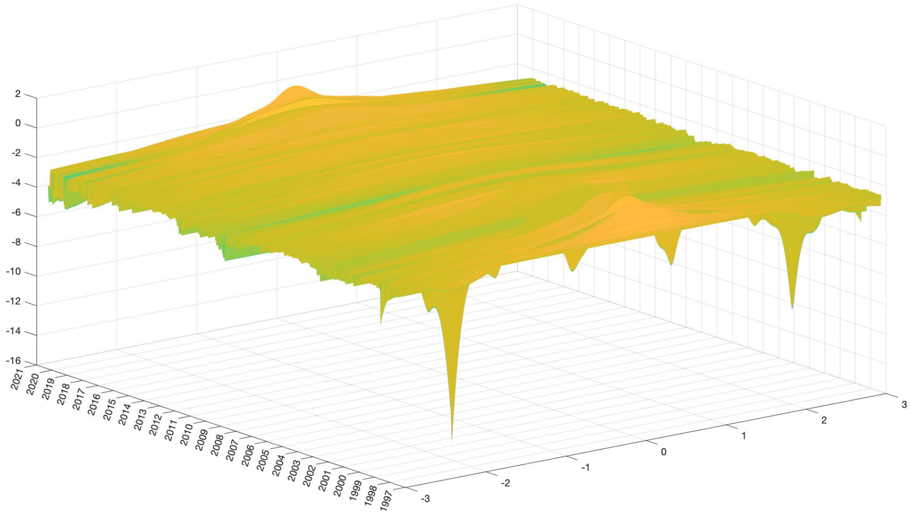

Figure 6.1: Calendar Month Five Minute Business time Return Densities For Apple Inc. (in natural log)

6.2 Empirical Illustrations and Case Study

In this illustration I will outline the framework using the S&P 500 index cross section. The full data set is given in Appendix A. The set up uses the protocol above with a linear framework.

The evaluation engine then matches exactly the best forecast density available and returns the optimal functional over the time frame for an individual asset (in this case Apple Inc). Finally, I then apply the loss function approach to determine the point of maximal uncertainty in the forecast (hence the realised stability). This can, of course, be applied to any asset in the cross section, using the first component as a conditioning density.

Fig. 6.1 presents the realized return density for Apple Inc, by month for five minute returns (annualized). Fig. 6.2 then presents the forecasted density func-

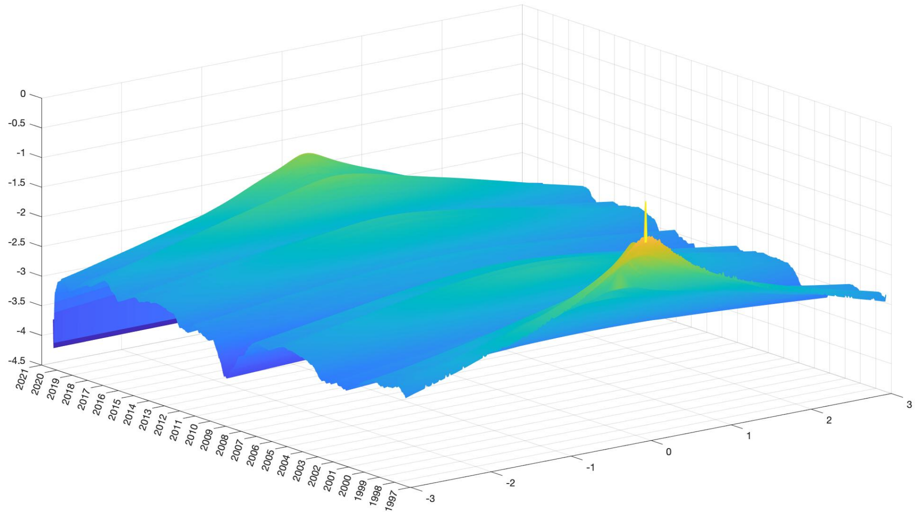

Figure 6.2: Calendar Month Five Minute Business time Forecasted Return Densities For Apple Inc. (in natural log)

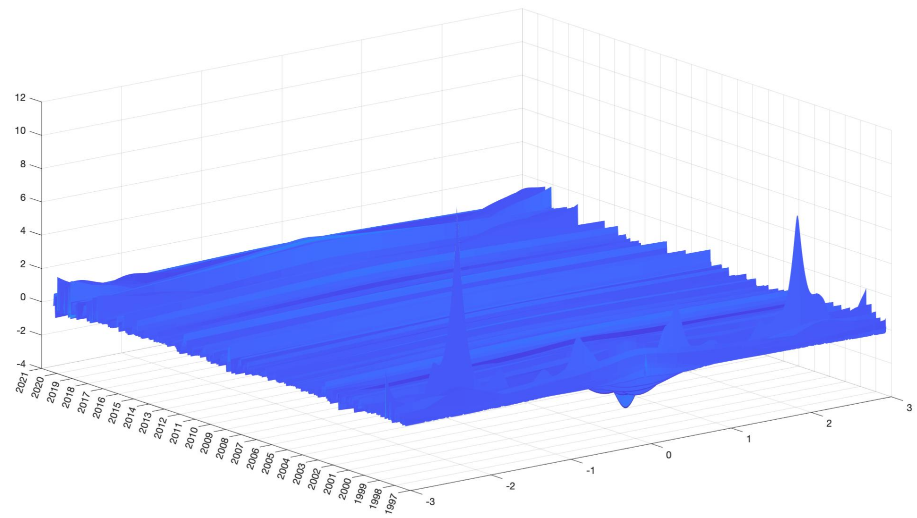

Figure 6.3: Calendar Month Five Minute Business time Kullback-Liebler Distances For Apple Inc.

tion for Apple using the algorithm set out in Chapter 5. Finally in Fig. 6.3 I present the Kullback-Leibler divergence of the realized and forecasted density

Figure 6.4: Calendar Month Five Minute Business time Functional Diebold-Mariano for both tales of the distribution and the central probability mass

functions.

6.2.1 Analysis

Of importance here are two components, first how the forecasted density captures the realized density over time and second how the forecasted density captures the realized density across the range of returns. The objective of any non-parametric forecasting approach is to not only predict the central probability mass, but also to evaluate the tails of the distribution. As annualized equivalents the range of values in Fig. 6.1 runs from -300% to +300%, hence a wide range of extrema. Observing the surface in Fig. 6.3 we can see that whilst the central probability mass is well determined, the tails are not always fully captures. The algorithm has significant *burn in* time in the initial period with significant spikes in both the tails of the distribution.

In Fig. 6.4 I present my functional version of the Diebold-Mariano test for the two tails and the central probability mass for my functional approach as compared to a standard parametric approach in this case standard asymmetric GARCH. When the line is in the negative area the loss function for the functional approach is larger than the GARCH model. Whereas in the positive area the relative loss favours the functional approach over GARCH. We can see that the shape of the central mass and both tales both favour the functional approach. However, it is not completely even, with the central mass heavily favoured relative to the tales.

Chapter 7

Conclusions

7.1 Summary and remarks

This thesis has outlined a menu of combined parametric and non-parametric forecasting approaches to high and low frequency return series. The thesis has outlined how the combination of parametric and non-parametric methods can be successfully used to deliver near model free forecasts of conditional density.

Why is this important? Point estimates with a parametric confidence interval are common in most areas of economics, finance and management. However, the construction of these confidence intervals is normally a function of the presumed underlying stochastic process driving future innovations. Hence, when undertaking the classical *Box-Jenkins* style approach to forecasting the data generating process (DGP) is presumed and then parameterized. In some cases parameters are structural (such as the lag structure and notions of general adherence to deterministic or stochastic trends) or are to be directly estimated from some estimation procedure such as maximum likelihood, method of moments and it's generalized form.

However, this is often unattractive. Suites of forecasting models are presumed and then need to be combined in some way and then loss functions of various types need to be inferred with certain desirable properties. This often violates the both *naturalness* and *parsimony*, desirable properties of any forecasting method.

Naturalness, is the property that a model setup should reflect some underpinning theoretical structure to some level of abstraction. Hence a model that is completely untethered from the underpinning data is no satisfactory and may only have a limited window of efficacy before the inherent misalignment perturbs the accuracy and results in significant deterioration of forecasting effectiveness (by whatever metric you might wish to apply). In contrast parsimony, is the property that a model should be no more complex than is needed for the task at hand. Often naturalness and parsimony are common properties, although this is not always the case. My approach is to combine the best properties of parametric and non-parametric forecasting to provide an adaptive mechanism that still remains centered on the core application, asset pricing.

In my approach I consider both the information set and the structure of how that information is transformed into a forward looking density despite some time-inhomogenous properties. I use high frequency and low frequency data simultaneously to build index using high frequency principle components, then use the density of this index to forecast future individual asset returns through an adaptive functional. This functional is a mapping operator that transforms a collection of past densities into a future density forecast.

I then illustrate a series of evaluation tools, against functionals that operate on the realized densities versus the forecasted ones and provide alternative summary measures of fit. This extends the Diebold-Mariano framework into the density forecasting arena, a first in the literature.

Finally, I use a large data set of S&P 500 assets to illustrate the operationalisation of my technique as an empirically illustrated example. The code used to build these functionals, will be available as a free library programmed in Matlab and eventually in Python.

References

Ahmed, P. (2001). Forecasting correlation among equity mutual funds. *Journal of banking & finance 25*(6), 1187–1208. 35

Albuquerque, R., E. De Francisco, and L. B. Marques (2008). Marketwide private information in stocks: Forecasting currency returns. *The Journal of Finance 63*(5), 2297–2343. 35

Almeida, C. and J. Vicente (2008). The role of no-arbitrage on forecasting: Lessons from a parametric term structure model. *Journal of Banking & Finance 32*(12), 2695–2705. 36

Amburgey, T. L., D. Kelly, and W. P. Barnett (1990). Resetting the clock: The dynamics of organizational change and failure. In *Academy of Management Proceedings*, Volume 1990, pp. 160–164. Academy of Management Briarcliff Manor, NY 10510. 63

An, B.-J., A. Ang, T. G. Bali, and N. Cakici (2014). The joint cross section of stocks and options. *The Journal of Finance 69*(5), 2279–2337. 44

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of financial economics 61*(1), 43–76. 10

Ang, J. S. and D. R. Peterson (1985). Return, risk, and yield: evidence from ex ante data. *The Journal of Finance 40*(2), 537–548. 26

Anjum, H. and F. Malik (2020). Forecasting risk in the us dollar exchange rate under volatility shifts. *The North American Journal of Economics and Finance 54*, 101257. 40

Anscombe, F. J. (1967). Topics in the investigation of linear relations fitted by the method of least squares. *Journal of the Royal Statistical Society: Series B (Methodological) 29*(1), 1–29. 38

Asquith, P., M. B. Mikhail, and A. S. Au (2005). Information content of equity analyst reports. *Journal of financial economics 75*(2), 245–282. 26

Bacchetta, P. and E. Van Wincoop (2006). Can information heterogeneity explain the exchange rate determination puzzle? *American Economic Review 96*(3), 552–576. 45

Bachelier, L. (2011). Louis bachelier's theory of speculation. In *Louis Bachelier's Theory of Speculation*. Princeton University Press. 9

Balaban, E., A. Bayar, and R. W. Faff (2006). Forecasting stock market volatility: Further international evidence. *The European Journal of Finance 12*(2), 171–188. 35

Battalio, R. and P. Schultz (2006). Options and the bubble. *The Journal of Finance 61*(5), 2071–2102. 44

Baum, J. and T. Amburgey (2002). Organizational ecology. joel ac baum. 61

Becker, R., A. E. Clements, and S. I. White (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking & Finance 31*(8), 2535–2549. 44

Benavides, G. and C. Capistrán (2012). Forecasting exchange rate volatility: The superior performance of conditional combinations of time series and option implied forecasts. *Journal of Empirical Finance 19*(5), 627–639. 36

Bernales, A. and M. Guidolin (2014). Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests. *Journal of Banking & Finance 46*, 326–342. 44

Bernanke, B. S. and M. Gertler (2001). Should central banks respond to movements in asset prices? *american economic review 91*(2), 253–257. 19

Bernhardt, D., M. Campello, and E. Kutsoati (2006). Who herds? *Journal of Financial Economics 80*(3), 657–675. 26

Black, F. (1993). Estimating expected return. *Financial analysts journal 49*(5), 36–38. 21, 22, 23, 25

Black, F., M. C. Jensen, M. Scholes, et al. (1972). The capital asset pricing model: Some empirical tests. 17

Bollapragada, R., I. Savin, and L. Kerbache (2009). Price forecasting and analysis of exchange traded funds. *International Journal of Finance 21*(1). 36

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327. 20, 34

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics 192*(1), 1–18. 40

Borovička, J., L. P. Hansen, and J. A. Scheinkman (2016). Misspecified recovery. *The Journal of Finance 71*(6), 2493–2544. 32, 74

Botosan, C. A. and M. A. Plumlee (2005). Assessing alternative proxies for the expected risk premium. *The accounting review 80*(1), 21–53. 26

Brav, A. and R. Lehavy (2003). An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. *The Journal of Finance 58*(5), 1933–1967. 25, 26

Britten-Jones, M. and A. Neuberger (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance 55*(2), 839–866. 43

Broadie, M. and A. Jain (2008). Pricing and hedging volatility derivatives. *The Journal of Derivatives 15*(3), 7–24. 43

Brown, C. A. and D. M. Robinson (2002). Skewness and kurtosis implied by option prices: A correction. *Journal of Financial Research 25*(2), 279–282. 44

Brown, S. J., W. N. Goetzmann, and A. Kumar (1998). The dow theory: William peter hamilton's track record reconsidered. *The Journal of finance 53*(4), 1311–1333. 10

Cairney, T. and J. Swisher (2004). The role of the options market in the dissemination of private information. *Journal of Business Finance & Accounting 31*(7-8), 1015–1042. 44

Campbell, J. Y. (2000). Asset pricing at the millennium. *The Journal of Finance 55*(4), 1515–1567. 21, 38

Campbell, J. Y., A. W. Lo, A. C. MacKinlay, and R. F. Whitelaw (1998). The econometrics of financial markets. *Macroeconomic Dynamics 2*(4), 559–562. 9

Capinski, M. (2007). A model of credit risk based on cash flow. *Computers & Mathematics with Applications 54*(4), 499–506. 24

Chalamandaris, G. and A. E. Tsekrekos (2010). Predictable dynamics in implied volatility surfaces from otc currency options. *Journal of Banking & Finance 34*(6), 1175–1188. 44

Chalamandaris, G. and A. E. Tsekrekos (2011). How important is the term structure in implied volatility surface modeling? evidence from foreign exchange options. *Journal of International Money and Finance 30*(4), 623–640. 44

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics 11*(2), 121–135. 39

Chen, Q., J. Francis, and W. Jiang (2005). Investor learning about analyst predictive ability. *Journal of Accounting and Economics 39*(1), 3–24. 26

Cheung, Y.-W., M. D. Chinn, and A. I. Garcia Pascual (2004). Empirical exchange rate models of the nineties: Are any fit to survive? 50

Choi, K., W.-C. Yu, and E. Zivot (2010). Long memory versus structural breaks in modeling and forecasting realized volatility. *Journal of International Money and Finance 29*(5), 857–875. 36

Choudhry, T. and H. Wu (2009). Forecasting the weekly time-varying beta of uk firms: Garch models vs. kalman filter method. *The European Journal of Finance 15*(4), 437–444. 35

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841–862. 39

Claessen, H. and S. Mittnik (2002). Forecasting stock market volatility and the informational efficiency of the dax-index options market. *The European Journal of Finance 8*(3), 302–321. 43

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting 5*(4), 559–583. 3, 27, 60

Clement, M. B. and S. Y. Tse (2005). Financial analyst characteristics and herding behavior in forecasting. *The Journal of finance 60*(1), 307–341. 35

Clements, M. P., P. H. Franses, and N. R. Swanson (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting 20*(2), 169–183. 20, 37

Clements, M. P. and D. F. Hendry (2005). Evaluating a model by forecast performance. *Oxford Bulletin of Economics and Statistics 67*, 931–956. 51

Clements, M. P. and N. Taylor (2003). Evaluating interval forecasts of high-frequency financial data. *Journal of Applied Econometrics 18*(4), 445–456. 40

Constantinides, G. M., M. Czerwonko, J. Carsten Jackwerth, and S. Perrakis

(2011). Are options on index futures profitable for risk-averse investors? empirical evidence. *The Journal of Finance 66*(4), 1407–1437. 44

Cottle, S. (1960). Charles h. dow and the dow theory. 9

Cowles 3rd, A. (1933). Can stock market forecasters forecast? *Econometrica: Journal of the Econometric Society*, 309–324. 10, 21

Cox, J. C., J. E. Ingersoll Jr, and S. A. Ross (1985). An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, 363–384. 18

Crownover, C., J. Pippenger, and D. G. Steigerwald (1996). Testing for absolute purchasing power parity. *Journal of International Money and Finance 15*(5), 783–796. 46

De La Bruslerie, H. and C. Deffains-Crapsky (2008). Information asymmetry, contract design and process of negotiation: The stock options awarding case. *Journal of Corporate Finance 14*(2), 73–91. 44

Demski, J. S. and G. A. Feltham (1972). Forecast evaluation. *The Accounting Review 47*(3), 533–548. 51

Derrig, R. A. and E. D. Orr (2004). Equity risk premium: expectations great and small. *North American Actuarial Journal 8*(1), 45–69. 11

Dette, H., K. Kokot, and A. Aue (2020). Functional data analysis in the banach space of continuous functions. *The Annals of Statistics 48*(2), 1168–1192. 2

Dhamija, A. and V. Bhalla (2010). Financial time series forecasting: comparison of neural networks and arch models. *International Research Journal of Finance and Economics 49*, 185–202. 35

Diebold, F. (2012). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests (working paper no. 18391). 47

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & economic statistics 20*(1), 134–144. 40, 52

Diebold, F. X., R. S. Mariano, et al. (1991). *Comparing predictive accuracy I: An asymptotic test*. Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis 40

Dolinar, A. L. (2014). Decomposition as a tool for forecasting stock exchange index: Example of sbi top (slovenianstock exchange index). *Journal of Finance 2*(1), 01–07. 35

Dovonon, P., A. Taamouti, and J. Williams (2021). Testing the eigenvalue structure of spot and integrated covariance. *Journal of Econometrics*. 67

Dovonon, P., A. Taamouti, and J. Williams (2022). Testing the eigenvalue structure of spot and integrated covariance. *Journal of Econometrics 229*(2), 363–395. 50

Duffee, G. R. (2018). Expected inflation and other determinants of treasury yields. *The Journal of Finance 73*(5), 2139–2180. 39

Dybvig, P. and S. A. Ross (2003). Arbitrage, state prices and portfolio theory handbook of the economics of finance. *The Economics of Finance 1*, 605–637. 43

Easton, P. D. and G. A. Sommers (2007). Effect of analysts' optimism on estimates of the expected rate of return implied by earnings forecasts. *Journal of accounting research 45*(5), 983–1015. 26

Elliott, G. and A. Timmermann (2016). Forecasting in economics and finance. *Annual Review of Economics 8*, 81–110. 39

Elton, E. J. (1999). Presidential address: expected return, realized return, and asset pricing tests. *The Journal of Finance 54*(4), 1199–1220. 21, 22

Engle, R. (2004). Risk and volatility: Econometric models and financial practice. *American economic review 94*(3), 405–420. 20, 37

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007. 20, 34

Engle, R. F. and S. J. Brown (1986). Model selection for forecasting. *Applied Mathematics and Computation 20*(3-4), 313–327. 48

Engle, R. F. and A. J. Patton (2004). Impacts of trades in an error-correction model of quote prices. *Journal of Financial Markets 7*(1), 1–25. 20, 37

Ericsson, N. R., D. F. Hendry, and G. E. Mizon (1998). Exogeneity, cointegration, and economic policy analysis. *Journal of Business & Economic Statistics 16*(4), 370–387. 53

Eugene, F. and K. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), 427–465. 21

Faccio, M., M.-T. Marchica, and R. Mura (2011). Large shareholder diversification and corporate risk-taking. *The Review of Financial Studies 24*(11), 3601–3641. 63

Fama, E. F. (1990). Stock returns, expected returns, and real activity. *The journal of finance 45*(4), 1089–1108. 18

Fama, E. F. (1995). Random walks in stock market prices. *Financial analysts journal 51*(1), 75–80. 21

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of financial economics 49*(3), 283–306. 23, 25

Fama, E. F. and K. R. French (1988). Permanent and temporary components of stock prices. *Journal of political Economy 96*(2), 246–273. 20, 23, 37

Fama, E. F. and K. R. French (2012). Size, value, and momentum in international stock returns. *Journal of financial economics 105*(3), 457–472. 64

Faria, G. and F. Verona (2018). Forecasting stock market returns by summing the frequency-decomposed parts. *Journal of Empirical Finance 45*, 228–242. 36

Fernandez, P. et al. (2001). Valuation using multiples. how do analysts reach their conclusions. *IESE Business School 1*, 1–13. 26

Ferris, S., W. Guo, and T. Su (2003). Predicting implied volatility in the commodity futures options markets. *International Journal of Finance and Banking 1*(1), 73–94. 44

Ferson, W. E., S. Sarkissian, and T. T. Simin (2003). Spurious regressions in financial economics? *The Journal of Finance 58*(4), 1393–1413. 36

Fisher, I. (1930). *Theory of interest: as determined by impatience to spend income and opportunity to invest it*. Augustusm Kelly Publishers, Clifton. 13

Fleming, J., C. Kirby, and B. Ostdiek (2001). The economic value of volatility timing. *The Journal of Finance 56*(1), 329–352. 35

Fricke, E., S. Fung, and S. Goktan (2018). Informed versus uninformed investors: Internet searches, options trading, and post-earnings announcement drift. *International Research Journal of Applied Finance 9*(8), 374–393. 44

Frijns, B. and D. Margaritis (2008). Forecasting daily volatility with intraday data. *The European Journal of Finance 14*(6), 523–540. 36

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies 76*(2), 669–705. 39, 40, 51, 53, 54

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25*(4), 595–620. 40

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578. 51, 52

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association 106*(494), 746–762. 39

Goldenberg, D. H. and R. J. Schmidt (1996). On estimating the expected rate of return in diffusion price models with application to estimating the expected return on the market. *Journal of Financial and Quantitative Analysis 31*(4), 605–631. 10

Gordon, M. J. (1962). The savings investment and valuation of a corporation. *The Review of Economics and Statistics*, 37–51. 21, 24

Granger, C. W. and S.-H. Poon (2002). Forecasting volatility in financial markets: A review (revised edition). *Available at SSRN 331800*. 31, 43

Granger, C. W. J., G. Elliott, and A. Timmermann (2006). *Handbook of Economic Forecasting*, Volume 1. Elsevier. 53

Granger, C. W. J., H. White, and M. Kamstra (1989). Interval forecasting: an analysis based upon arch-quantile estimators. *Journal of Econometrics 40*(1), 87–96. 39

Greenspan, A. (2004). Risk and uncertainty in monetary policy. *American Economic Review 94*(2), 33–40. 31

Greve, H. R., J.-E. Pozner, and H. Rao (2006). Vox populi: Resource partitioning, organizational proliferation, and the cultural impact of the insurgent microradio movement. *American Journal of Sociology 112*(3), 802–837. 62

Guo, H. and B. Qiu (2014). Options-implied variance and future stock returns. *Journal of Banking & Finance 44*, 93–113. 44

Guo, H. and R. Savickas (2008). Forecasting foreign exchange rates using idiosyncratic volatility. *Journal of Banking & Finance 32*(7), 1322–1332. 36

Hannan, M. T. and J. Freeman (1984). Structural inertia and organizational change. *American sociological review*, 149–164. 62

Hannan, M. T. and J. Freeman (1989). *Organizational ecology*. Harvard university press. 61

Hansen, P. R. and A. Timmermann (2012). Choice of sample split in out-of-sample forecast evaluation. 47

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *The Review of Economic Studies 61*(2), 247–264. 20, 34

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *The Review of Financial Studies 29*(1), 5–68. 38

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting 13*(2), 281–291. 56

Haugom, E., H. Langeland, P. Molnár, and S. Westgaard (2014). Forecasting volatility of the us oil market. *Journal of Banking & Finance 47*, 1–14. 35

Hibon, M. and T. Evgeniou (2005). To combine or not to combine: selecting among forecasts and their combinations. *International journal of forecasting 21*(1), 15–24. 60

Hueng, C. J. and J. B. McDonald (2005). Forecasting asymmetries in aggregate stock market returns: Evidence from conditional skewness. *Journal of Empirical Finance 12*(5), 666–685. 35

Jegadeesh, N., J. Luo, A. Subrahmanyam, and S. Titman (2022). Momentum and short-term reversals: Theory and evidence. *Nanyang Business School Research Paper* (22-13). 64

Jordan, S. J., A. J. Vivian, and M. E. Wohar (2014). Forecasting returns: new european evidence. *Journal of Empirical Finance 26*, 76–95. 36

Kadir, T. and M. Brady (2005). Non-parametric estimation of probability distributions from sampled signals. Technical report, Technical report, OUEL. 76

Karatzas, I. (1997). Lectures on the mathematics of finance. 12

Kim, N. and J. Lee (2013). No-arbitrage implied volatility functions: Empirical evidence from kospi 200 index options. *Journal of Empirical Finance 21*, 36–53. 44

Kim, Y., G. J. Lobo, and M. Song (2011). Analyst characteristics, timing of forecast revisions, and analyst forecasting ability. *Journal of Banking & Finance 35*(8), 2158–2168. 35

Koopman, S. J., B. Jungbacker, and E. Hol (2005). Forecasting daily variability of the s&p 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance 12*(3), 445–475. 35

Kumar, M. and M. Thenmozhi (2014). Forecasting stock index returns using arima-svm, arima-ann, and arima-random forest hybrid models. *International Journal of Banking, Accounting and Finance 5*(3), 284–308. 35

Lebedev, V. (1997). Functional spaces and problems in the theory of approximation. In *An Introduction to Functional Analysis in Computational Mathematics*, pp. 1–93. Springer. 2

Lépinette, E. (2019). A short introduction to arbitrage theory and pricing in mathematical finance for discrete-time markets with or without friction. 43

Li, Y., P. Perron, and J. Xu (2017). Modelling exchange rate volatility with random level shifts. *Applied Economics 49*(26), 2579–2589. 52

Lin, T.-C. and X. Lu (2015). Why do options prices predict stock returns? evidence from analyst tipping. *Journal of Banking & Finance 52*, 17–28. 44

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The journal of finance 20*(4), 587–615. 17

Liyan, H. and Z. Chenli (2002). A research on forecasting corporate default risk based on fuzzy-stochastic methods [j]. *Journal of Finance 8*. 35

Lo, A. W. (2005). Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. *Journal of investment consulting 7*(2), 21–44. 22, 23

Lomi, A. and E. R. Larsen (2001). *Dynamics of organizations: Computational modeling and organizational theories*, Volume 1. The MIT Press. 62

Lothian, J. R. (1997). Multi-country evidence on the behavior of purchasing power parity under the current float. *Journal of International Money and Finance 16*(1), 19–35. 46

Lu, Y. K. and P. Perron (2010). Modeling and forecasting stock return volatility using a random level shift model. *Journal of Empirical Finance 17*(1), 138–156. 35

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. *Econometrica: journal of the Econometric Society*, 1429–1445. 18

Maheu, J. M. and T. H. McCurdy (2007). Components of market risk and return. *Journal of Financial Econometrics 5*(4), 560–590. 21

Maheu, J. M. and T. H. McCurdy (2009). How useful are historical data for forecasting the long-run equity return distribution? *Journal of Business & Economic Statistics 27*(1), 95–112. 38

Makridakis, S. and M. Hibon (1997). Arma models and the box–jenkins methodology. *Journal of Forecasting 16*(3), 147–163. 35

Martens, M. (2001). Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance 20*(1), 1–23. 35

Martens, M., Y.-C. Chang, and S. J. Taylor (2002). A comparison of seasonal adjustment methods when forecasting intraday volatility. *Journal of Financial Research 25*(2), 283–299. 35

McQueen, G. and V. V. Roley (1993). Stock prices, news, and business conditions. *The review of financial studies 6*(3), 683–707. 22

Meese, R. A. and K. Rogoff (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics 14*(1-2), 3–24. 45, 50

Mehdiyev, N., D. Enke, P. Fettke, and P. Loos (2016). Evaluating forecasting methods by considering different accuracy measures. *Procedia Computer Science 95*, 264–271. 48

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics 8*(4), 323–361. 8, 21

Mixon, S. (2007). The implied volatility term structure of stock index options. *Journal of Empirical Finance 14*(3), 333–354. 44

Mo, H. and L. Wu (2007). International capital asset pricing: Evidence from options. *Journal of Empirical Finance 14*(4), 465–498. 44

Monteiro, A. M., R. H. Tütüncü, and L. N. Vicente (2008). Recovering risk-neutral probability density functions from options prices using cubic splines and ensuring nonnegativity. *European Journal of Operational Research 187*(2), 525–542. 68

Morgan, W. M. (2006). *Measuring performance in nonprofit organizations: A stochastic frontier approach.* Nova Southeastern University. 62

Muzzioli, S. (2010). Option-based forecasts of volatility: an empirical study in the dax-index options market. *The European Journal of Finance 16*(6), 561–586. 44

Neely, C. J. and L. Sarno (2002). How well do monetary fundamentals forecast exchange rates? *Federal Reserve Bank of St. Louis Working Paper Series* (2002-007). 50

Panigirtzoglou, N. and G. Skiadopoulos (2004). A new approach to modeling the dynamics of implied distributions: Theory and evidence from the s&p 500 options. *Journal of Banking & Finance 28*(7), 1499–1520. 44

Patton, A. J. and A. Timmermann (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association 102*(480), 1172–1184. 17

Péli, G. and M. Masuch (1997). The logic of propagation strategies: Axiomatizing a fragment of organizational ecology in first-order logic. *Organization Science 8*(3), 310–331. 62

Pérez-Rodríguez, J. V., S. Torra, and J. Andrada-Félix (2005). Star and ann models: forecasting performance on the spanish "ibex-35" stock index. *Journal of Empirical Finance 12*(3), 490–509. 35

Pesaran, M. and A. Pick (2010). Forecasting combination across estimation windows. Technical report, CESifo Working Paper Series, 2, 22–93. 53

Pesaran, M. H. and A. Timmermann (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics 10*(4), 461–465. 20, 37

Pesaran, M. H. and A. Timmermann (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance 9*(5), 495–510. 36

Petkova, R. (2006). Do the fama–french factors proxy for innovations in predictive variables? *The Journal of Finance 61*(2), 581–612. 35

Politis, D. N. (2015). Model-free prediction in regression. In *Model-Free Prediction and Regression*, pp. 57–80. Springer. 47

Pong, S., M. B. Shackleton, S. J. Taylor, and X. Xu (2004). Forecasting currency volatility: A comparison of implied volatilities and ar (fi) ma models. *Journal of Banking & Finance 28*(10), 2541–2563. 35

Rao, C., Y. Wu, S. Konishi, and R. Mukerjee (2001). On model selection. *Lecture Notes-Monograph Series*, 1–64. 48

Rime, D., L. Sarno, and E. Sojli (2010). Exchange rate forecasting, order flow and macroeconomic information. *Journal of International Economics 80*(1), 72–88. 45

Roodposhti, F. R. and Z. Amirhosseini (2011). Revised capital assets pricing model: an improved model for forecasting risk and return. *Journal of Finance and Accountancy 5*, 1. 36

Ross, S. (2015). The recovery theorem. *The Journal of Finance 70*(2), 615–648. 70

Ross, S. A. (1976). The arbitrage theory of capital asset pricing." qumal cf eccncmic thegry 13341860.. 1978.". *The Current Status of the Capital Asset Pricing Model (CAPM)." qurnal cf Finance 33*, 885–901. 17

Rossi, B. (2006). Are exchange rates really random walks? some evidence robust to parameter instability. *Macroeconomic dynamics 10*(1), 20–38. 45

Samuelson, P. Proof that properly anticipated prices fluctuate randomly'(1965). *Ind. Man Rev. 6*, 41. 24

Samuelson, P. A. (1973). Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science*, 369–374. 21, 22

Sarno, L. (2001). The behavior of us public debt: a nonlinear perspective. *Economics Letters 74*(1), 119–125. 46

Sen, R. and C. Ma (2015). Forecasting density function: application in finance. *Journal of Mathematical Finance 5*(05), 433. 35

Series, B. G. J. G. T. (1970). Analysis: Forecasting and control. 12

Shackleton, M. B., S. J. Taylor, and P. Yu (2010). A multi-horizon comparison of density forecasts for the s&p 500 using index returns and option prices. *Journal of Banking & Finance 34*(11), 2678–2693. 44

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance 19*(3), 425–442. 17

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of economic perspectives 17*(1), 83–104. 22, 23

Siburg, K. F., P. Stoimenov, and G. N. Weiß (2015). Forecasting portfolio-value-at-risk with nonparametric lower tail dependence estimates. *Journal of Banking & Finance 54*, 129–140. 36

Singh, V. K. and N. Ahmad (2011). Forecasting performance of volatility models for pricing s&p cnx nifty index options via black-scholes model. *IUP Journal of Applied Finance 17*(3). 36

Spiess, D. K. and J. Affleck-Graves (1995). Underperformance in long-run stock returns following seasoned equity offerings. *Journal of financial economics 38*(3), 243–267. 63

Sweeney, R. J. (2006). Mean reversion in g-10 nominal exchange rates. *Journal of Financial and Quantitative Analysis 41*(3), 685–708. 45

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting 16*(4), 437–450. 52

Taylor, S. (1986). Modeling financial time series john wiley & sons. *Great Britain*. 10, 20, 34

Taylor, S. J., P. K. Yadav, and Y. Zhang (2010). The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance 34*(4), 871–881. 44

Thornhill, S. and R. Amit (2003). Learning about failure: Bankruptcy, firm age, and the resource-based view. *Organization science 14*(5), 497–509. 61

Timmermann, A. (2018). Forecasting methods in finance. *Annual Review of Financial Economics 10*, 449–479. 36, 40

Tripathy, S. and A. Rahman (2013). Forecasting daily stock volatility using garch model: A comparison between bse and sse. *IUP Journal of Applied Finance 19*(4), 71. 36

Tu, J. and G. Zhou (2004). Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics 72*(2), 385–421. 50

Verma, R. (2011). Testing forecasting power of the conditional relationship between beta and return. *The Journal of Risk Finance 12*(1), 69–77. 36

Vortelinos, D. I. and K. Gkillas (2018). Intraday realised volatility forecasting and announcements. *International Journal of Banking, Accounting and Finance 9*(1), 88–118. 36

Wang, Y., F. Ma, Y. Wei, and C. Wu (2016). Forecasting realized volatility in a changing world: A dynamic model averaging approach. *Journal of Banking & Finance 64*, 136–149. 36

Wayne, W. Y., E. C. Lui, and J. W. Wang (2010). The predictive power of the implied volatility of options traded otc and on exchanges. *Journal of Banking & Finance 34*(1), 1–11. 44

Weiß, G. N. and H. Supper (2013). Forecasting liquidity-adjusted intraday value-at-risk with vine copulas. *Journal of Banking & Finance 37*(9), 3334–3350. 35

Wenjun, L. and X. Bangsheng (2017). A new discovery for analyst forecasting information sources: Audit firms. *Journal of Finance and Economics* (5), 6. 36

Westerlund, J. and P. K. Narayan (2012). Does the choice of estimator matter when forecasting returns? *Journal of Banking & Finance 36*(9), 2632–2640. 35

White, H. (2000). A reality check for data snooping. *Econometrica 68*(5), 1097–1126. 32, 36, 47

Williams, J. (1938). The theory of investment value (harvard: Cambridge, ma). 13, 21

You, H. and X.-j. Zhang (2009). Financial reporting complexity and investor underreaction to 10-k information. *Review of Accounting studies 14*(4), 559–586. 63

Yun, J. (2014). Out-of-sample density forecasts with affine jump diffusion models. *Journal of Banking & Finance 47*, 74–87. 44

Zakamulin, V. (2013). Forecasting the size premium over different time horizons. *Journal of Banking & Finance 37*(3), 1061–1072. 36

Zhu, D. and J. W. Galbraith (2011). Modeling and forecasting expected shortfall with the generalized asymmetric student-t and asymmetric exponential power distributions. *Journal of Empirical Finance 18*(4), 765–778. 35

A Code and Data Summary

Codes and data used in the thesis.

A.1 Empirical Density Estimator

Function that delivers the empirical density function through a choice of smoother. In this case it is a linear smoother.

```
function [xi,yi,dyi]=smooth_empirical_density(x,y,xi)


xmin = min(x);
xmax = max(x);


ind = intersect(find(xi>=xmin),find(xi<=xmax));
N = length(ind);
xxi = xi(ind);

yi_c = interp1(x,y,xxi,'linear');
yi = zeros(size(xi));
yi(ind) = yi_c;
yi(find(xi>xmax)) = 1; %#ok<FNDSB>
yi = smooth(yi,round(N./20));

dyi = diff(yi);
```

A.2 Forecasting Functional

This function constructs the forecasting functional for a dataset of high frequency data.

```
function functionalAnalysis(data);
```

```
nGrid = 2500;
nAssets = 5;
N = length(data);
TS = data(1).GridTimes;
dT = nanmedian(diff(TS').*24*60);
mx = linspace(-5,5,nGrid);

for i=2:N
    tic;
    R = data(i).Returns(:,1:5);
    [T,n] = size(R);
    R = R./repmat(std(R),T,1);
    F = data(i-1).Factor(:,1);
    F = F./std(F);
    adj_dT = dT./60./7./252;
    funF = ksdensity(F,mx);funF = funF';
    funR = zeros(nGrid,nAssets);
    for j=1:5
        funR(:,j) = ksdensity(R(:,j),mx);
    end
    data_out(i).StockReturnDensity = funR;
    data_out(i).FactorReturnDensity = funF;
    data_out(i).StandardDeviationStocks = std(R);
    data_out(i).StandardDeviationFactor = std(F);
    data_out(i).AbscissaRange = mx;
    [L,l] = lossFunction(funR,funF,mx);
    data_out(i).KB_Difference = l;%funR - repmat(log(
       funF'),1,nAssets);
    data_out(i).KB_Function = L;%funR - repmat(log(funF
       '),1,nAssets);
    data_out(i).Date = data(i).Date;
    t = toc;
    disp(['Completed: ',num2str(i),' Time Taken ',
       num2str(t),'.'])
end
data_out(1).StockReturnDensity = NaN;
data_out(1).FactorReturnDensity = NaN;
data_out(1).StandardDeviationStocks = NaN;
data_out(1).StandardDeviationFactor = NaN;
data_out(1).AbscissaRange = NaN;
data_out(1).KB_Difference = NaN;
```

```
y = zeros(N,nAssets);
Y = zeros(N,nGrid);
for i=2:N
    y(i,:) = data_out(i).KB_Difference;X(i)= data_out(i
       ).Date;
    Y(i,:) = data_out(i).KB_Function(:,1);
end

X(1) = [];
y(1,:) = [];
Y(1,:) = [];

sd = datenum('01-Jan-2000');
ed = max(X);
ind = find(X<sd);

X(ind) = [];
Y(ind,:) = [];


figure('position',[5 75 1591 854],'color','w');

s = surf(mx,X,Y);
s.EdgeColor = 'none';datetick('y');
set(gca,'view',[46.3433   22.6078]);
```

A.3 Loss Functional

This function provides the loss functional to build the fit assessment across the range of data.

```
function [L,l] = lossFunction(Y1,Y2,X)


[M,N] = size(Y1);
L = zeros(M,N);
for i=1:N
```

```
    L(:,i) = cumtrapz(X,Y1(:,i) - Y2);
end
l = L(end,:)';
```

A.4 Data Table

The following table outlines the data set used in the empirical example.

Table 1: Thesis Data Sample

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
A	2040477	5837	349.57	273.64
AA	2221518	6081	365.32	237.45
AAPL	3669526	6712	546.71	460.01
ABT	2268216	6081	373.00	276.60
ACE	1441181	4644	310.33	259.63
ACN	1896081	5910	320.82	254.59
ACS	947814	3509	270.10	210.84
ADBE	2618703	6688	391.55	308.08
ADI	1585531	4555	348.08	272.54
ADM	2143481	6080	352.54	225.78
ADSK	2403810	6570	365.87	279.66
ADT	600834	1831	328.14	185.35
AEE	2017015	5587	361.01	261.09
AEP	2152151	6080	353.97	252.83
AES	2095153	5885	356.01	174.98
AET	2036869	5738	354.97	289.60
AFL	2114566	6080	347.79	274.27
AGN	2077095	6020	345.03	298.38
AIG	2257675	6080	371.32	294.13
AIV	1894877	6081	311.60	227.27
AIZ	1638713	5869	279.21	224.83
AKAM	2304276	5719	402.91	307.61
AKS	1952840	6060	322.25	168.53
ALL	2244917	6080	369.22	277.09
ALTR	2345175	6091	385.02	272.21
ALXN	2145565	6570	326.57	272.85
AMAT	2821397	6699	421.16	270.84
AMD	1912969	5247	364.58	198.48
AME	1889937	6077	310.99	241.04
AMG	1879610	5615	334.74	292.68

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
AMGN	2758504	6688	412.45	342.33
AMP	1593161	4455	357.61	307.08
AMT	1946736	5607	347.19	273.44
AMZN	3193488	6368	501.48	433.50
AN	2110813	6019	350.69	235.20
ANF	2067576	5900	350.43	281.87
ANR	1118753	3698	302.52	214.13
APA	2144448	6080	352.70	303.98
APC	2053934	5912	347.41	303.09
APD	2124120	6080	349.36	295.04
APH	1915150	6081	314.94	260.77
APOL	1946639	5785	336.49	250.11
ARG	1522112	5104	298.21	225.94
ATI	2098614	5975	351.23	273.10
AVB	1859245	5484	339.03	287.83
AVGO	1327634	3687	360.08	294.36
AVP	2164158	6014	359.85	200.26
AVY	2084055	6080	342.77	275.17
AXP	2280966	6079	375.22	308.48
AYE	1097592	3354	327.24	214.03
AZO	2060039	6080	338.82	289.26
BA	2274856	6080	374.15	316.39
BAC	2309659	6081	379.81	243.76
BAX	2171172	6079	357.15	274.82
BBBY	2459181	6669	368.74	275.36
BBT	1987627	5980	332.37	237.23
BBY	2190650	6080	360.30	290.19
BCR	1813253	5509	329.14	265.13
BDK	1105923	3543	312.14	237.66
BDX	2134359	6081	350.98	290.93
BEN	2121540	6081	348.88	282.36
BFb	1982935	6080	326.14	258.08
BHI	1949428	5383	362.14	297.84
BJS	1162251	3575	325.10	253.82
BK	2211943	6080	363.80	258.43
BLK	1781945	6038	295.12	257.11
BLL	2019069	6080	332.08	259.88
BMS	1938625	5871	330.20	231.88
BMY	2294662	6079	377.47	263.34
BRCM	2104917	4939	426.18	335.33
BRKb	1693073	5994	282.46	243.98
BSX	2188840	6080	360	193.53
BWA	1961159	6080	322.55	265.67
BXP	1888179	5721	330.04	277.39
C	2271779	6066	374.51	268.55
CAG	2180497	6081	358.57	205.09
CAH	2143968	6080	352.62	283.33

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
CBG	1368851	4388	311.95	224.24
CBS	1567657	4508	347.75	256.35
CCE	1942843	5723	339.47	204.93
CCI	2023245	5484	368.93	284.79
CCL	2147111	6080	353.14	260.02
CEG	1164994	3227	361.01	274.66
CELG	2409271	6562	367.15	312.27
CEPH	1341735	4198	319.61	253.88
CERN	2287061	6616	345.68	277.55
CF	1924500	5612	342.92	278.35
CFN	685796	3602	190.39	114.47
CHK	1974371	6081	324.67	200.27
CHRW	2143100	6094	351.67	284.80
CI	2199521	6080	361.76	305.38
CINF	2184680	6520	335.07	235.54
CL	2231253	6079	367.04	293.85
CLF	1839530	6080	302.55	222.10
CLX	2171045	6081	357.02	288.60
CMA	2144472	6080	352.70	280.83
CMCSA	2428563	6682	363.44	246.79
CMI	1628658	5089	320.03	297.02
CMS	2041274	6080	335.73	186.37
CNX	1836060	5263	348.86	271.37
COF	2126940	6079	349.88	303.22
COG	1882393	6081	309.55	243.37
COH	1587334	4286	370.35	304.17
COL	1931153	5463	353.49	272.94
COP	1760202	5477	321.38	284.20
COST	2529781	6397	395.46	320.10
CPB	2187649	6080	359.81	241.08
CPWR	1782873	5105	349.24	131.45
CRM	1566365	5445	287.67	258.18
CSC	1899394	5320	357.02	280.16
CSCO	3224065	6700	481.20	300.74
CSX	1926547	5493	350.72	255.50
CTAS	2315710	6523	355	265.26
CTL	2072991	6080	340.95	224.66
CTSH	2229101	6055	368.14	297.68
CTXS	2535043	6642	381.66	300.07
CVC	1549302	4153	373.05	240.94
CVG	1826766	5060	361.02	204.90

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
CVH	1128300	3479	324.31	249.74
CVS	2153645	5886	365.89	280.54
D	2135979	6080	351.31	274.18
DAL	2075944	5691	364.77	245.94
DD	2132829	5642	378.02	299.30
DE	2187527	6079	359.84	302.28
DELL	2015701	4840	416.46	254.24
DF	1938649	5978	324.29	174.07
DFS	1608753	5680	283.23	212.33
DG	1848020	5486	336.86	238.10
DGX	1952166	5827	335.02	284.52
DHI	1978307	6080	325.37	231.44
DHR	2066197	6080	339.83	282.27
DIS	2296810	6080	377.76	278.94
DLTR	2333572	6640	351.44	272.97
DNB	1878981	5787	324.69	250.83
DNR	1789199	5750	311.16	180.33
DO	2147971	6080	353.28	280.37
DOV	2105004	6080	346.21	284.37
DOW	2076036	5684	365.24	276.92
DPS	998388	2602	383.70	307.28
DRI	2059474	6081	338.67	260.08
DTE	2096381	6080	344.79	264.39
DUK	2179894	6080	358.53	233.19
DV	1645741	5356	307.27	225.91
DVA	1815901	4898	370.74	298.20
DYN	1716591	4897	350.53	166.42
EBAY	2783791	6256	444.97	333.12
ECL	2062930	6080	339.29	270.09
ED	2140587	6080	352.07	258.50
EFX	2067271	6081	339.95	257.57
EIX	2109407	6055	348.37	253.83
EL	2031938	6080	334.20	271.01
EMC	1903910	5177	367.76	229.84
EMN	2125874	6080	349.65	285.58
EMR	2217214	6080	364.67	294.77
EOG	2048174	6081	336.81	295.32
EQR	1978649	6079	325.48	259.33
EQT	1925753	6079	316.78	259.65
ERTS	1522287	4274	356.17	270.87
ESRX	2211830	6324	349.75	288.96

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
ESS	1794719	6080	295.18	253.38
ESV	2101464	5904	355.93	271.84
ETN	2091651	6081	343.96	291.37
ETR	2093237	6079	344.33	276.78
EW	1826973	5030	363.21	300.05
EXC	1917471	5681	337.52	257.88
EXPD	2193606	6530	335.92	261.29
EXPE	2034736	5169	393.64	304.99
F	2231903	6075	367.39	157.11
FAST	2259506	6632	340.69	265.10
FCX	2092799	6080	344.21	246.18
FDO	1573720	4880	322.48	236.86
FDX	2178218	6081	358.20	313.20
FE	2040899	5623	362.95	259.36
FFIV	2289549	5749	398.25	335.47
FII	1861654	5451	341.52	239.19
FISV	2363549	6527	362.11	288.66
FITB	2375099	6649	357.21	223.17
FLIR	2069682	6525	317.19	219.49
FLR	2088430	6065	344.34	282.43
FLS	1880938	5697	330.16	271.76
FMC	2057789	6080	338.45	275.83
FOSL	2084197	6612	315.21	236.98
FPL	1526980	5134	297.42	201.44
FRX	1390588	3696	376.24	294.64
FTR	574603	2608	220.32	75.43
GAS	1642536	5126	320.43	229.19
GCI	2180581	6081	358.58	226.56
GD	2130823	6081	350.40	300.98
GE	2325726	6080	382.52	234.47
GENZ	1376583	4059	339.14	275.87
GGP	1627547	5448	298.74	185.14
GILD	2638467	6692	394.27	321.13
GIS	2166899	6079	356.45	253.67
GLW	2205813	6081	362.73	216.14
GM	2109282	5708	369.53	259.81
GMCR	1419045	5518	257.16	211.27
GME	1732877	4673	370.82	271.06
GPC	2082753	6081	342.50	263.89

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
GPS	2213385	6081	363.98	237.99
GR	1325039	4142	319.90	232.71
GRMN	2036438	5706	356.89	280.01
GS	2004379	5261	380.98	357.66
GT	1662446	4735	351.09	216.17
GWW	2060499	6081	338.84	286.37
HAL	2271199	6080	373.55	296.96
HANS	931805	4270	218.22	176.49
HAR	1620588	5304	305.54	258.47
HAS	1187549	3378	351.55	226.25
HBAN	2234148	6679	334.50	145.93
HCBK	1491786	4949	301.43	118.96
HCN	1660846	5547	299.41	234.83
HCP	1878390	5973	314.48	225.60
HD	2271196	6079	373.61	297.90
HIG	2171192	6080	357.10	274
HNZ	1522589	4359	349.29	224.23
HON	2200455	6081	361.85	299.89
HOT	1678450	5189	323.46	264.24
HP	2076917	6081	341.54	285.59
HRB	2106876	6081	346.46	215.01
HRL	1952920	6081	321.15	217.05
HRS	1987621	5883	337.85	269.55
HSP	1357020	3684	368.35	277.33
HSY	2139033	6081	351.75	277.51
HUM	2082803	6080	342.56	272.95
IBM	2298859	6080	378.10	344.17
IFF	2091009	6080	343.91	266.52
IGT	2049894	6080	337.15	215.29
INTC	3259077	6712	485.55	321.21
INTU	2511024	6618	379.42	298.90
IP	2266209	6080	372.73	277.43
IPG	2142489	6081	352.32	163.90
IR	2161653	6080	355.53	289.08
IRM	1831510	5267	347.73	237.55
ISRG	1960366	5551	353.15	308.65
ITT	1964119	5451	360.32	289.66
ITW	2168574	6081	356.61	298.57
JBL	2053368	5506	372.93	252.92
JCI	2112705	6081	347.42	267.25
JCP	2185135	6081	359.33	221.50

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
JDSU	1898220	4474	424.27	207.98
JEC	1916851	5997	319.63	264.72
JNJ	2296920	6080	377.78	299.37
JNS	1424064	4664	305.33	164.43
JPM	2267853	6081	372.94	303.64
JWN	1950631	5235	372.61	308.68
K	2162828	6080	355.72	259.30
KBH	2048350	6080	336.89	246.37
KEY	2167379	6081	356.41	190.41
KG	1013006	2699	375.32	203.76
KIM	1886252	6081	310.18	189.43
KLAC	2633101	6634	396.91	327.97
KMB	2228730	6081	366.50	299.28
KMI	1545872	4217	366.58	240.16
KMX	1851269	5813	318.47	254.47
KO	2290887	6080	376.79	258.31
KR	2151941	6080	353.93	218.68
KSS	2133149	6080	350.84	298.27
KSU	2025538	6080	333.14	262.57
L	1669718	5536	301.61	201.11
LEG	2021440	6080	332.47	226.73
LEN	2002687	6079	329.44	270.86
LH	1940136	6079	319.15	262.56
LIFE	1098997	5457	201.39	135.11
LLL	1848569	5300	348.78	302.83
LLTC	2161769	5826	371.05	275.62
LLY	2310786	6081	380	299.87
LM	1999676	6080	328.89	267.39
LMT	2202793	6081	362.24	306.60
LNC	2150317	6080	353.67	281.16
LOW	2184577	6080	359.30	281.03
LRCX	2537020	6665	380.64	310.09
LSI	1895872	5161	367.34	198.93
LTD	1525757	4481	340.49	228.42
LUK	1677733	5607	299.22	202.94
LUV	2209479	6080	363.40	212.28
LXK	1786175	5236	341.13	271.38
MA	1378698	3653	377.41	355.17
MAC	1849187	6081	304.09	248.19
MAR	1655981	4945	334.87	244.08
MAS	2147133	6081	353.08	227.60

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
MBI	2120931	6081	348.78	205.04
MCD	2250045	6080	370.07	290.16
MCHP	2439960	6665	366.08	280.10
MCK	2110387	6080	347.10	286.80
MCO	1870824	5645	331.41	279.10
MDP	1945280	6080	319.94	229
MDT	2238390	6080	368.15	284.64
MEE	962804	3491	275.79	217.77
MET	1913739	5080	376.72	304.96
MHK	1908025	5598	340.84	292.66
MHP	1442341	4339	332.41	258.70
MI	1107732	3857	287.20	187.27
MIL	1356647	4727	286.99	195.68
MKC	1879911	5246	358.35	269.34
MLM	1946616	6081	320.11	271.61
MMC	2180988	6080	358.71	262.43
MMI	726650	3031	239.73	169.41
MMM	2236930	6080	367.91	320.03
MNK	664365	1978	335.87	281.88
MO	2323517	6081	382.09	255.61
MOLX	1570855	4676	335.93	220.58
MON	1657348	4564	363.13	311.45
MRK	2292172	6081	376.94	284.92
MRO	2160427	6080	355.33	256.99
MS	1447941	3931	368.33	287.24
MSFT	3309159	6708	493.31	343.18
MTB	1886130	5635	334.71	297.58
MU	1487777	3985	373.34	245.05
MUR	2001979	6079	329.32	277.95
NAVI	1193194	4759	250.72	116.73
NBL	2044903	6079	336.38	277.63
NE	2165382	6018	359.81	257.51
NEM	2211055	6080	363.66	282
NFX	1832255	5789	316.50	262.17
NI	2020446	6079	332.36	177.63
NKE	2224473	6079	365.92	298.91
NOC	2094598	6081	344.44	294.11
NOV	1573503	4759	330.63	285.87
NOVL	1349640	4078	330.95	111.83
NRG	1672350	4538	368.52	258.63
NSC	2175649	6081	357.77	289.86

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
NSM	1996944	5541	360.39	244
NTAP	2487241	6682	372.23	279.71
NTRS	2332169	6473	360.29	288.89
NU	1448773	4786	302.71	179.99
NUE	2135579	6081	351.18	288.99
NVDA	2684496	5959	450.49	331.34
NVLS	1809544	5019	360.53	271.10
NWL	2131328	6080	350.54	219.40
NYT	2041906	5655	361.07	198.24
ODP	1760912	5182	339.81	145
OI	2002438	6080	329.34	216.29
OKE	1983066	6081	326.10	247.84
OMC	2125336	6080	349.56	289.02
ORLY	2148506	6595	325.77	257.27
OXY	2174665	6081	357.61	287.69
PAYX	2449028	6669	367.22	257.87
PBCT	1975144	6588	299.80	140.45
PBG	929642	2735	339.90	205.03
PBI	2154627	6081	354.32	210.24
PCAR	2356138	6637	355	292.97
PCG	2128396	6081	350	243.48
PCL	1543526	5039	306.31	230.29
PCLN	2112002	5313	397.51	323.76
PCP	1546273	5023	307.83	257.82
PCS	1084949	3250	333.83	194.47
PDCO	2137652	6485	329.63	226.44
PEG	2123829	6080	349.31	249.53
PEP	2228345	5998	371.51	288.42
PETM	1632708	5193	314.40	221.28
PFE	2328477	6081	382.91	240.96
PFG	1789089	5553	322.18	254.46
PG	2314239	6081	380.56	305.21
PGR	2121030	6079	348.91	228.79
PH	2107013	6080	346.54	292.68
PHM	2042253	6080	335.89	228.92
PKI	1933481	5139	376.23	268.28
PLD	1875169	5466	343.06	253.51
PLL	1605514	4938	325.13	232.20
PNC	2197949	6079	361.56	298.26
PNR	1949350	6036	322.95	256.82
PNW	2019549	6080	332.16	253.67

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
POM	1602116	5063	316.43	152.09
PPG	2167029	6080	356.41	298.61
PPL	2067127	6079	340.04	221.37
PSA	1888269	6081	310.51	261.15
PSX	803084	2553	314.56	288.62
PTV	976124	2766	352.90	222.43
PVH	1807982	6079	297.41	253.05
PWR	1868736	5560	336.10	231.20
PX	1990693	5718	348.14	292.25
PXD	1950527	5687	342.98	296.11
Q	1515869	3960	382.79	196.95
QCOM	2849469	6703	425.10	339.11
QLGC	2001183	5604	357.09	216.68
R	2019734	6080	332.19	266.37
RCL	1976250	6081	324.98	260.91
RDC	2076304	5830	356.14	267.21
REGN	2087151	6642	314.23	253.40
RHI	2035511	6081	334.73	252.26
RIG	2180099	6080	358.56	279.15
RL	1902044	5937	320.37	272.72
ROK	2150278	6076	353.89	289.30
ROP	1862049	5874	316.99	266.25
ROST	2319058	6643	349.09	274.18
RRC	1748909	5431	322.02	258.45
RSG	1932858	5467	353.55	235.06
RTN	1930725	5230	369.16	306.94
S	2204212	5975	368.90	162.02
SAI	821032	2490	329.73	159.08
SBUX	2680668	6692	400.57	288.72
SCG	1857179	5751	322.93	231.88
SE	1722290	5513	312.40	205.73
SEE	2050409	6077	337.40	241.50
SHW	2099741	6080	345.35	281.84
SIAL	1744630	5296	329.42	252.98
SII	1205460	3660	329.36	271.17
SJM	1773704	4927	359.99	292.87
SLB	2271445	6081	373.53	323.40
SLE	1439969	4122	349.33	162.36
SNA	1994519	6081	327.99	258.73
SNDK	2043724	5599	365.01	297.89
SO	2179579	6081	358.42	228.99

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
SPG	1982000	6080	325.98	275.50
SPLS	2140893	5978	358.12	192.78
SRCL	2074532	6322	328.14	259.43
SRE	2012253	5469	367.93	293.58
STI	2140876	5995	357.11	278.39
STJ	1749978	5034	347.63	265.84
STR	1614366	5184	311.41	195.30
STT	2150837	6080	353.75	298.24
STZ	1835118	4911	373.67	277.13
SUN	1843540	5564	331.33	252.86
SVU	1907957	5713	333.96	167.09
SWK	2052025	6079	337.55	275.28
SWN	1798725	6081	295.79	213.46
SWY	1654047	4773	346.54	220.82
SYK	2004851	5697	351.91	298.32
SYMC	2399193	6550	366.28	222.16
SYY	2148710	6080	353.40	231.01
T	2255745	6073	371.43	225.39
TAP	1593289	4809	331.31	271.44
TE	1673046	5130	326.12	140.98
TEG	831809	2460	338.13	276.97
TEL	1260966	3968	317.78	263.91
TER	2193190	6080	360.72	234.62
TGT	1956846	5072	385.81	330.57
THC	2141146	6080	352.16	200.41
TIE	971804	3633	267.49	194.16
TIF	2070864	6081	340.54	280.71
TJX	2135203	6081	351.12	262.09
TLAB	1771722	4858	364.70	163.66
TMK	2007666	5910	339.70	264.66
TMO	2106333	6080	346.43	268.16
TRIP	1017153	3039	334.69	278.75
TROW	2278480	6549	347.91	284.69
TRV	1492936	3984	374.73	315.57
TSCO	2079629	6588	315.66	260.19
TSN	1950834	5639	345.95	209.22
TSO	1637192	5403	303.01	242.23
TSS	1772421	5937	298.53	188.92
TWX	1821187	4937	368.88	257.71
TXN	1668418	4495	371.17	292.78
TXT	2151982	6079	354.00	273.03
TYC	1860870	5176	359.51	260.35

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
UA	1305754	3827	341.19	284.01
UHS	1953488	6080	321.29	264.37
UNH	2189587	6080	360.12	300.96
UNM	2133439	6080	350.89	229.98
UNP	2158248	6080	354.97	302.43
UPS	1951671	5128	380.59	328.87
URBN	2109211	6636	317.84	239.23
URI	1905870	5596	340.57	267.73
USB	2124485	5690	373.37	253.85
UTX	2232005	6075	367.40	315.79
V	1687854	4867	346.79	292.98
VAR	1980031	6080	325.66	271.42
VFC	2086784	6080	343.22	281.75
VLO	2006094	6074	330.27	273.52
VMC	1999038	6079	328.84	277.28
VNO	1890700	6080	310.97	261.43
VRSN	2447514	6086	402.15	304.22
VRTX	2333086	6661	350.26	287.53
VTR	1814627	5509	329.39	261.56
VZ	1926918	4967	387.94	282.63
WAG	1687283	4753	354.99	256.49
WAT	2012484	6081	330.94	277.86
WDC	1509382	4599	328.19	221.79
WEC	2008516	6080	330.34	233.86
WFC	2257351	6079	371.33	280.06
WFMI	1235929	4081	302.84	237.67
WFR	1205154	4352	276.91	174.40
WHR	2127939	6079	350.04	300.50
WIN	983338	3343	294.14	132.58
WLP	1540070	4732	325.45	276.11
WM	1966129	5144	382.21	273.12
WMB	2183170	6079	359.13	247.60
WMT	2257067	6079	371.28	290.96
WPI	1332685	3845	346.60	251.80
WY	2190983	6080	360.35	260.01
WYN	1243807	4203	295.93	237.43
X	2129784	6080	350.29	270.35
XEL	1892938	5528	342.42	202.68
XL	1837282	5678	323.57	225.02
XLNX	2671151	6669	400.53	302.17
XOM	1980937	5113	387.43	334.25

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
XRAY	2166171	6446	336.04	243.88
XRX	2265252	6081	372.51	172.57
XTO	997655	3615	275.97	214.11
YHOO	2558951	5861	436.60	307.25
YUM	2060980	5648	364.90	293.13
ZION	2232899	6589	338.88	245.43
ABC	1807644	4676	386.57	320.57
ADS	1721158	4734	363.57	309.18
BIIB	2013304	4690	429.27	374.59
BTU	1675096	4500	372.24	295.23
CNP	1708266	4408	387.53	189.11
COV	800622	2164	369.97	300.87
CVX	1807457	4653	388.44	359.61
DISCA	1707163	4167	409.68	276.12
DISCB	825438	3759	219.58	172.18
DVN	1519726	3901	389.57	353.61
EP	1012625	2614	387.38	218.48
FHN	1564114	4021	388.98	210.07
FTI	1766695	4729	373.58	308.90
GNW	1547212	3995	387.28	173.95
GOOG	2420334	4538	533.34	494.42
HPQ	1755199	4510	389.17	262.17
ICE	1404895	3624	387.66	361.51
JOYG	1056235	2874	367.51	306.33
KFT	1063160	2770	383.81	234.87
MFE	651845	1677	388.69	286.29
MHS	840997	2167	388.09	326.59
MJN	903622	3116	289.99	231.89
MNST	1605548	3838	418.32	336.72
MOS	1508979	3893	387.61	309.25
MWV	1301478	3374	385.73	283.26
NBR	1414483	3633	389.34	265.58
NDAQ	1704069	4327	393.82	289.85
NFLX	2392224	5108	468.32	397.60
PRU	1782469	4608	386.82	344.58
RAI	1272362	3269	389.22	309.53
RF	1749826	4511	387.90	201.78
RHT	1312607	3458	379.58	319.58
SHLD	1552837	3962	391.93	315.45
WYNN	2048467	4971	412.08	361.32
XEC	1671344	4408	379.16	338.87

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
ZMH	1338825	3488	383.83	332.89
ADP	1381458	3323	415.72	327.54
AON	1014779	2608	389.10	334.29
BIG	1333212	3435	388.12	324.09
BMC	480119	1144	419.68	287.48
CA	1240887	3111	398.87	221.51
CME	1371192	3428	399.99	352.14
CMG	1367851	3577	382.40	348.73
DTV	920482	2268	405.85	295.31
ES	1028230	2686	382.81	238.94
ETFC	1566814	3901	401.64	216.44
FIS	1388771	3573	388.68	304.02
FSLR	1773622	3970	446.75	388.54
HES	1366377	3506	389.72	364.26
HOG	1338192	3438	389.23	338.13
HST	1371924	3521	389.64	215.45
IVZ	1261434	3244	388.85	275.01
JAVA	277271	712	389.42	138.20
JNPR	1024662	2630	389.60	247.41
LO	687618	1766	389.36	338.16
M	1262010	3239	389.62	311.39
MAT	1290404	3056	422.25	245.13
MWW	781287	2010	388.70	170.39
MYL	1436118	3304	434.66	291.94
NWSA	1263345	3181	397.15	174.31
NYX	754846	1937	389.69	319.32
PM	1181089	3031	389.66	342.40
RRD	787510	2079	378.79	176.77
SNI	752499	1940	387.88	335.76
STX	1412496	3402	415.19	303.94
TDC	1225942	3155	388.57	318.15
TWC	900725	2322	387.90	341.95
WU	1325573	3405	389.30	192.97
ABBV	712387	1831	389.06	343.64
ACT	261264	772	338.42	321.87
ALLE	617439	1600	385.89	337.64
ANTM	524012	1347	389.02	373.16
BEAM	251240	692	363.06	281.88
CIEN	617016	1585	389.28	278.35
DLPH	814834	2112	385.81	324.43
EA	1040832	2401	433.49	330.68

Ticker	Obs.	Active Days	Obs. Per Data	Updates Per Day
FB	1455790	2316	628.57	514.97
FOXA	795548	1950	407.97	258.53
GOOGL	961708	1726	557.18	509.23
JOY	521891	1342	388.89	305.20
KORS	687523	1772	387.99	350.27
KRFT	322867	828	389.93	284
LB	622066	1600	388.79	339.29
LVLT	590527	1520	388.50	308.43
LYB	928822	2389	388.79	357.13
MDLZ	974535	2162	450.75	284.98
MHFI	289262	745	388.27	350.43
MPC	859484	2209	389.08	356.49
MSI	906893	2333	388.72	335.82
NEE	961305	2468	389.50	354.17
NLSN	897829	2318	387.32	274.67
ORCL	660786	1698	389.15	267.24
PRGO	668722	1724	387.88	356.53
QEP	957557	2462	388.93	275.49
SCHW	991593	2544	389.77	257.32
VIAB	944616	2293	411.95	313.15
WBA	699263	1478	473.11	380.74
WFM	789154	1878	420.20	317.22
WPX	810182	2082	389.13	258.78
XYL	823287	2124	387.61	305.24
ZTS	702710	1810	388.23	314.51