Inland Norway University of Applied Sciences

School of Business and Social Sciences (HHS)

Hervé N. Mugemana

Master thesis

Evaluating the impact of sampling frequency on volatility forecast accuracy.

Master of Science in business administration – major digital management and business analytics

2024

Samtykker til tilgjengeliggjøring i digitalt arkiv Brage JA ☒ NEI ☐

Table of contents

List of figures

List of tables

Abstract

This master's thesis, *"Evaluating the impact of sampling frequency on volatility forecast accuracy"* aims to answer the main problem statement of "how do varying sampling frequencies influence the accuracy of volatility forecasts?" The thesis is driven by the idea that better prediction accuracy could result from the increased data accessibility brought about by digital progress. Some research, like the one conducted by Chan et al. (2010), indicates that higher sampling frequencies may not considerably improve forecast precision. Ewald et al. (2023) discovered that increased sampling frequencies resulted in enhanced forecasting precision compared to others. The research goal is to determine if higher sampling rates truly improve the accuracy of forecasts and if the resulting time and computational demands are justified by the increase. Several RealGARCH models using various sampling frequencies are used to assess the relationship between sampling frequency and forecasting accuracy, with data from Brent crude oil futures.

Examination of the in-sample results revealed a link between increased sampling rates and better model alignment, as evidenced by reduced AIC and BIC values and elevated log-likelihood values as sampling frequency declined. This indicates that increasing sampling frequencies may boost the precision of the model. The out-of-sample assessment showed a different situation; the connection between sampling frequency and forecasting precision was not easy to understand. Analysis of visual and regression data indicated that increased sampling rates do not always lead to lower forecast errors. The findings indicated that errors decreased when the sampling frequency was lowered. This was not in line with the belief that increasing the frequency of sampling would lead to more precise predictions. Statistical regression models revealed that only a small percentage of the variations in forecast errors could be attributed to changes in sampling frequency. Both linear and polynomial regression models yielded comparable results, indicating that sampling frequency has minimal effects on forecasting error metrics based on adjusted R^2 values. Negative correlation coefficients between sampling frequency and error metrics (MSE and MAE) indicated a small enhancement in forecast accuracy with decreased sampling

frequency, contradicting initial assumptions. This was backed by substantial p-values, suggesting an actual, albeit small, statistical correlation.

The results differ from the common view in the literature that increasing the frequency of sampling results in improved forecasts. On the contrary, the thesis proposes that there might be a case where lower frequency t enhances precision. The connection between sampling frequency and forecast accuracy seems intricate and is affected by a variety of factors, such as the model's nature and the data set's characteristics. The evaluation pointed out possible problems with the data's reliability and the model's suitability, potentially impacting the findings' generalizability. Certain data points did not align with anticipated price levels, and the model did not get better with higher complexity, possibly because of overfitting or inadequate model specification for dealing with the detailed data.

This thesis highlights the importance of carefully weighing the pros and cons of frequent data collection when predicting volatility. It paves the way for additional studies on the most effective sampling frequencies for diverse markets and asset categories, prompting more extensive testing in different contexts to gain a deeper insight into financial market dynamics by, for example, adding more variables to the models. The results could be particularly beneficial for non-professional traders and researchers who are dedicated to improving the precision of financial models.

1. Introduction

1.1 Academic background

One significant advancement in empirical finance is the creation and prediction of volatility, according to Danıelsson, (2011). This is significant mainly because it can capture fluctuations in trading prices over time, which is essential for risk evaluation, portfolio oversight, derivative valuation, and market supervision. The accurate forecasting of volatility is, therefore, not just of theoretical interest, but it is also of importance for market participants, risk managers, and policymakers who rely on the forecasts for devising hedging strategies, optimising portfolios, and making informed regulatory decisions (Karasan, 2021).

The volatility of financial assets has mainly been modelled by using GARCH models, whose application is considered to have a significant impact on financial methodology (Francq & Zakoian, 2010). These models capture the dynamic patterns observed in financial data and are widely used in financial econometrics and time series analysis. Since Engle's (1982) groundbreaking paper on ARCH models and the further development of GARCH models (Bollerslev, 1986), there has been a significant emphasis on studying volatility and advancing methodologies for measuring, modelling, and forecasting it. Among the advanced methodologies are approaches like the use of volatility measures suited for high-frequency data(Andersen & Bollerslev, 1998; Barndorff-Nielsen et al., 2008a; Barndorff-Nielsen & Shephard, 2004) and models that integrate realised volatility(Hansen et al., 2012). These advancements have largely resulted from observations of the nature of volatility in financial asset returns, or so-called stylised facts, which remain consistent across different assets, asset classes, time periods, and countries (Andersen et al., 2010). Many of the consequent studies in the field have been geared toward finding improved models. Many of the subsequent studies in the field have focused on finding better models. This is evident in the volume of scholarly papers that propose novel and refined models. The performance of 330 GARCH-type models is analysed by Lunde & Hansen (2005), showing the extent of model development. Besides innovations within the GARCH framework, other models like stochastic volatility models (Harvey & Shephard, 1996) and the HAR-RV model (Corsi, 2008) have also been introduced. Although the mentioned alternative approaches are increasingly prevalent in the literature, GARCH models and their variations are still used in many papers. Among the variations of the GARCH model is the log-linear RealGARCH model, which is a GARCH model that includes a realised volatility measure. Models applying realised volatility have been proven to outperform regular GARCH models when using high-frequency data for modelling and forecasting volatility (Hansen et al., 2012; Zhang et al., 2019).

Developments in forecasting volatility have been focused on more than just model innovation. The digital age has ushered in a period marked by the increased availability of high-frequency data. This abundance of data can be harnessed to improve our understanding of market and economic factors, which also opens up new possibilities in modelling and forecasting volatility. Recent studies have applied the concept of realised volatility to what appears to be a trend moving from the strictly parametric approaches that were previously prevalent. Behind this switch lies the realisation that realised volatility, defined as a cumulative summation of squared returns over consecutive, small, and fixed time intervals, is a more accurate measure for volatility when applying high-frequency data (T. Andersen et al., 2001).

It has been suggested that the frequency at which data is sampled can impact the characteristics and performance of predictive models (Merton, 1980). Luong & Dokuchaev (2016) also highlight the presence of a connection between the volatility of a financial instrument and the frequency at which it is sampled. Some find that increasing sampling frequency leads to better forecasting ability (Ewald et al., 2023). At the same time, others find that increasing sampling frequency has no considerable effect on model performance (Chan et al., 2010). Which is in contrast to the statistical principle that more information is generally preferred to less (Aït-Sahalia et al., 2005).

Part of this contrast is brought on by the complexity of volatility, which is influenced by a myriad of factors ranging from market microstructural effects to global geopolitical events. The challenge of modelling volatility increases when dealing with high-frequency data. Although increased granularity provides a richer information set, it also brings forth challenges such as microstructure noise, data quality issues, and a need for computational intensity (Aït-Sahalia et al., 2005). This ultimately sets a limit to how high the sampling frequency can go. One approach to dealing with these challenges has been to determine an optimal frequency at which to sample the data used to calculate realized volatility. But others, while applying different methods to mitigate the negative effects of microstructure noise, have pushed the commonly applied 5 min frequency to higher levels. What remains less explored is the actual accuracy gain from changing sampling frequencies.

Chan et al. (2010) assess the influence of different factors on the accuracy of volatility forecasting. The factors examined include sampling frequency, different measures of realized volatility, different forecasting horizons, and different types of models. In their findings, sampling frequencies are deemed to have little influence on the model's performance. If their findings are true, it makes one wonder what the point of all these increased samplings is.

1.2 Motivation

As part of my master studies in business analytics and digital management, the course on predictive analytics opened my interest in the challenges of predicting stock exchange movements. Given the vital role of volatility within the financial sector, improving the accuracy of its forecasts is of utmost importance for both private and institutional investors. The motivation for this thesis comes from the continued innovation within volatility forecasting, where technological improvements have given, us access to more data. The question is whether using this data leads to better forecasts or not. It is clear from the literature that increased sampling leads to improved measurement of volatility; what isn't clear is whether this also translates to improved forecasting accuracy. Many studies that deal with forecasting volatility focus on developing new models. While such studies offer interesting insights into the measurement, modelling and forecasting of volatility, more attempts have yet to be made to assess the role of the different steps involved in the forecasting process. Additionally, the dynamic nature of volatility and the differences in markets and assets dictate that findings from one setting cannot freely be assumed to be relevant in another without first being tested. Given these points, there are opportunities to test previous findings in other settings and continually update the literature on the effect of sampling frequency on volatility forecasts.

1.3 Problem Statement

One of the strategies given by Furseth & Everett (2022) on how to find workable master thesis projects is the strategy of replicating previous research in another time or another context. This thesis builds on the research by Chan et al. (2010). The core focus is therefore to contribute to the literature on forecasting volatility when using high-frequency data. Since the market microstructure factors influencing realised volatility, such as bid–ask bounce and non-trading, differ between different markets and different assets, there is a need to evaluate the prevalent findings using different assets and different models. This is supported by (Brownlees & Gallo, 2010), who point out that results from a study do not automatically extend to other stocks or classes of assets. Similarly, (Ewald et al., 2023) express that their findings are only relevant for the context used. Such observations suggest that applying different stocks or asset classes to previous work contributes to theory development by either extending the scope of relevance or yielding new insights.

The work by Chan et al. (2010), empirically tests how the choice of the sampling frequency, the realised volatility (RV) measure, the forecasting horizon, and the time-series model affect the quality of volatility forecasting. This thesis differs from Chan et al. (2010) in the application of a different type of model, which is the log-linear RealGARCH. This model is deemed to be better at handling high-frequency data than the models evaluated by Chan et al. (2010)(Hansen et al., 2012). Additionally, Chan et al. (2010) compares time series forecasts with implied volatility, but this study stops short of making such comparisons; only realised volatility is used. The emphasis in this thesis is laid on evaluating how different sampling frequencies affect forecasting accuracy and whether, like Chan et al. (2010) conclude, sampling frequency has a minimal effect on forecasts. Chan et al. (2010) uses 4 sampling frequencies: 30 seconds, 1 minute, 3 minutes, and 5 minutes. In this thesis, the evaluation is done by estimating several RealGARCH models using returns and realised volatilities from 54 sampling frequencies ranging from 1-94 minutes. While holding the model's specifications constant, the parameters and realised volatility will solely be dependent on the sampling frequency. Furthermore, a different data set and asset class are used in this study, with data from ICE Brent oil futures being applied.

My goal is not to prove the superiority of one model over another, so no such model comparisons will be carried out. Instead, I am focused on assessing the changes in forecasting accuracy when different sampling frequencies are used. Relying solely on a 5-minute sampling frequency, as many studies do, can result in lost information and potentially inaccurate forecasts. On the other hand, higher sampling frequencies involve increased costs in terms of time and computational power. The practical significance of this study is to determine if higher frequency sampling is crucial for those using log-linear RealGARCH models or, to an extent, dealing with crude oil assets. If the influence of sampling frequency is limited, there would be no need to incur such costs. Therefore, it is essential to investigate how different sampling frequencies influence volatility forecasts, ultimately addressing the core research question:

How do varying sampling frequencies influence the accuracy of volatility forecasts?

Given that the literature seems to advocate for increased sampling frequency, is it the case that higher sampling frequencies are associated with increased accuracy? First, I'll look at the relationship between sampling frequency and forecast accuracy. It can be said that the relationship between sampling frequency and forecasting accuracy, if it exists, may have a cause-and-effect aspect. Although correlations may exist, they do not imply absolute causation; other aspects may be responsible for this relationship. Further, it is also interesting to explore the magnitude of the forecasting accuracy gained or lost from changing sampling frequencies. In this study, this is viewed as the magnitude aspect of the effect that sampling frequency has on forecasting accuracy. The research question to be explored on this aspect is: how big is the change in forecasting accuracy gained or lost from changing sampling frequency?

1.4 Research context.

This thesis uses high-frequency transaction-level data from Brent crude oil futures contracts traded at the Intercontinental Exchange (ICE) spanning from 2004 to 2021. North Sea crude oil market Brent, which is a North Sea crude, is an important benchmark for most internationally traded crudes (long 2002, Fattouh& Imsirovic, 2020) The following passage outlines some of the characteristics of oil and its trading.

The trading characteristics of oil are significantly influenced by its physical nature, despite the presence of highly standardised paper trading instruments (Long, 1995). Oil, as a commodity, is subject to the logistical challenges of transportation, processing, and storage as it moves from the producer to the consumer. This causes price fluctuations due to mismatches in the location and timing of oil availability, which is in contrast to the instantaneous transferability of financial assets(Long, 1995).

Demand for oil, like other primary commodities, is closely tied to the global economic state, with consumption patterns reflecting economic growth rates(Long, 1995). However, demand is subject to seasonal fluctuations, with different fuels peaking in demand at various times of the year.

While short-term price changes have minimal impact on oil consumption, long-term prices significantly influence demand levels.

On the supply side, matching oil supply with demand has become challenging, particularly due to the shift from an integrated oil industry to a scenario where producers aim to maximise output for quick returns. OPEC plays a key role in attempting to stabilize prices by adjusting production levels, but it faces challenges due to the expansion of non-OPEC production and the seasonal nature of oil demand. Despite these efforts, the oil market has experienced periods of significant price volatility(Long, 1995). Oil prices are very sensitive to geopolitical shocks and events. Wars, trading disputes, pandemics, etc. have been shown to have an impact on oil prices. This sensitivity is evident from the recent events during the COVID-19 pandemic and the 2008 financial crisis, both of which are visible in the data set.

The structure of the oil market has evolved with the introduction of standardised trading instruments like futures and forward contracts, which facilitate liquidity and price transparency. These instruments, along with swaps and options, have transformed market operations, extending trading horizons and allowing for more effective price risk management over longer periods(Long, 1995).

In ICE Brent Crude futures, the contract with the nearest expiration date, commonly known as the front-month contract, is the most actively traded and hence the most liquid. Additionally, the price fluctuations of this particular contract are the ones most commonly reported in news outlets around the globe(Ewald et al., 2023). The focus of this thesis will therefore be on this contract.

1.5 An explanation of terms

ACF – Autocorrelation function

AIC - Akaike Information Criterion

ARCH - Autoregressive conditional heteroskedasticity.

BIC - Bayesian Information Criterion

GARCH - Generalized Autoregressive Conditional Heteroskedasticity

HAR - Heterogeneous autoregressive

I.I.D.: independent and identically distributed.

ICE- Intercontinental Exchange

LLH: Logarithmic Likelihood

MAE: Mean Absolute Error

MSE: Mean Squared Error

OPEC: The Organization of the Petroleum Exporting Countries

R^2 : Coefficient of determination

RV: Realised volatility

1.6 This thesis's outline.

Section 2 starts by presenting related studies that review the role of sampling frequency in forecasting accuracy. As sampling frequency on its own is not significant for forecasting, the section continues to present the data-driven research framework. This study investigates the effects of sampling frequency within this framework. Other topics in Section 2 include the steps undertaken to find relevant literature, which is followed by a review of the literature on the different components that are used in forecasting volatility from a RealGARCH perspective. Section 3 deals with methods. Here the research design is outlined, followed by the challenges encountered while taking on this task. Consequent topics in this section include choice of data collection and analysis, data cleaning and preparation, model estimation, and model evaluation. Section 4 encompasses the results and discussion; these are divided into an in-sample part and an out-of-sample part. The final section of this thesis is the section that gives a conclusion as well as potential avenues for future research.

2. Literature Review

2.1 The impact of sampling frequency on volatility forecasting

Since the ARCH model's inception, most likely even before, sampling frequency has been a part of forecasting volatility since at least the inception of the ARCH model. These models were able to forecast the conditional variance based on daily, weekly, or monthly returns, etc. The move toward high-frequency data and the intra-daily range of volatility opened the discourse on the optimal level of sampling for calculating the realised. There are some similarities between this discourse and general statistical theory. In statistical theory, it is often the case that the population of interest is too large to be studied in its entirety. In order to be able to make inferences about this population, representative samples are used instead. Similarly, in volatility forecasting, where the population is represented by the continuous time series, samples of the price level are taken at different intervals and used to make inferences about the whole time series (McCrorie, 2009). In statistical theory, larger samples are often associated with improved accuracy of the inferences made about populations. Transferred to the volatility forecasting field, higher sampling frequencies, which are related to increasing the sample size, should also lead to improvements in the accuracy of forecasts. However, statistical learning involves a central tenet, which is a compromise between variance and bias. The issue of noise at higher sampling frequencies is an issue that needs to be addressed in statistical learning in general (James, 2013). Similar challenges are also present in modelling and forecasting volatility. The results of choosing a higher sampling frequency are that these will be permeated by microstructure noise, whereas reducing the sampling frequency may lead to a loss of vital insights. This section presents the different aspects of sampling frequency achieved in different studies and those that mention its role in forecasting accuracy.

Andersen et al.'s (2000) research on currency trading data highlights the instability of realized volatility at extremely high sampling frequencies, such as intervals of 5 and 10 seconds. Consequently, Aït-Sahalia et al. (2005) suggested that more reliable estimations of realised volatility can be obtained from a 5-minute sampling frequency. On the other hand, some studies (Oomen, 2006) using IBM transaction data find that the optimal sampling frequency can be reduced to 12 seconds from 2,5 minutes, albeit after incorporating an error correction scheme to reduce the microstructure noise. In another study, Hansen & Huang (2016) calculated six realised volatility measures that differ in terms of sampling frequency, ranging from 15 seconds to 20 minutes. They find that the best combination of realised measures and daily returns in the RealEGARCH framework is obtained using sampling at a frequency of 2 minutes. These studies show that there is variety in the suggested sampling frequency and that catering for microstructure noise leads to an even lower sampling frequency. However, the effect of these reductions is not clearly addressed in all of them.

Amongst the research papers that look at the role of sampling frequency in forecasting volatility is the one by Chan et al. (2010) where they explore the factors that influence the accuracy of volatility forecasts, with a specific focus on the impact of sampling frequency, the realised volatility measure, the forecasting horizon, and the used models. Different measures of realised volatility are evaluated: (1) Intraday Volatility, which measures volatility within a single trading day and captures high-frequency market movements; (2) Total Volatility measures the sum of squared returns over a specific period and provides a comprehensive view of realised volatility; this can be classified as the unconditional volatility mentioned above. (3) Scaled Total Volatility is obtained by dividing total volatility by the square root of the number of periods, allowing for comparison across different forecasting horizons. (4) Close-to-Close Volatility measures volatility between consecutive closing prices. The inclusion of non-trading hours is found to significantly influence the distribution of realized volatility and forecasting performance. Therefore, the choice of realized volatility is more important than the sampling frequency used to calculate it. As a result of their findings, Chan et al. (2010) proceeded to apply a sampling frequency of 5 minutes to their study. Considering that it is an established statistical principle that when all other things are held equal, more data is preferred to less (Aït-Sahalia et al., 2005). In other words, sampling at a frequency of 2.5 minutes or 2 minutes, as achieved by Oomen (2006) or Hansen & Haung (2006), should be preferred to the 5-minute frequency employed by Chan et al. (2010). However, the later finds very little improvement in increasing the sampling frequency from 5 minutes to 30 seconds.

In a more recent study, (Ewald et al., 2023) reviewed sample frequency robustness and accuracy in forecasting the value at risk for Brent Crude Oil Futures. Accurate estimates of volatility are crucial for VaR forecasting, and high-frequency data combined with realised volatility are assessed to provide precise volatility estimates. The sampling frequency used to calculate realized volatility has been shown to have an impact on the model's performance, with higher frequencies generally yielding better results.

To arrive at sampling frequency, a complete forecasting process must be undertaken. Sampling frequency is only a small part of this process. The remainder of this section will review the other crucial parts of this process.

2.2 Research framework.

The process of forecasting volatility involves different steps, among which are the collection of data, estimation of model parameters, forecasting, and evaluation. Given the latent nature of volatility, an important part of the parameter estimation step is the selection of a measure for the volatility proxy. There are a multitude of volatility measures; some examples are conditional volatility(R. F. Engle, 1982), realised volatility(T. Andersen & Bollerslev, 1998) and realised kernels (Barndorff-Nielsen et al., 2008b). Regardless of the volatility measure one chooses consideration of the frequency at which it is calculated must also be made. Considerations of sampling frequency have been part of volatility forecasting since at least its inception. For example, in the ARCH framework, choosing the sampling frequency used to calculate returns involved making a choice between data collected at daily, weekly, or monthly intervals. The use of higher sampling frequencies has only become possible after technical and theoretical developments, with the latter being the invention of suitable measures like realised volatility, as in Andersen & Bollerslev (1998) and the former being improvements in computational ability and availability of data. The literature on volatility forecasting is vast and encompasses a wide range of approaches. Providing an overview of the whole field is beyond the scope of this study. However, if an attempt were to be made to group these approaches into a single framework, the data-driven research framework (Zou & Xu, 2023) illustrated in Figure 1. would be suitable.

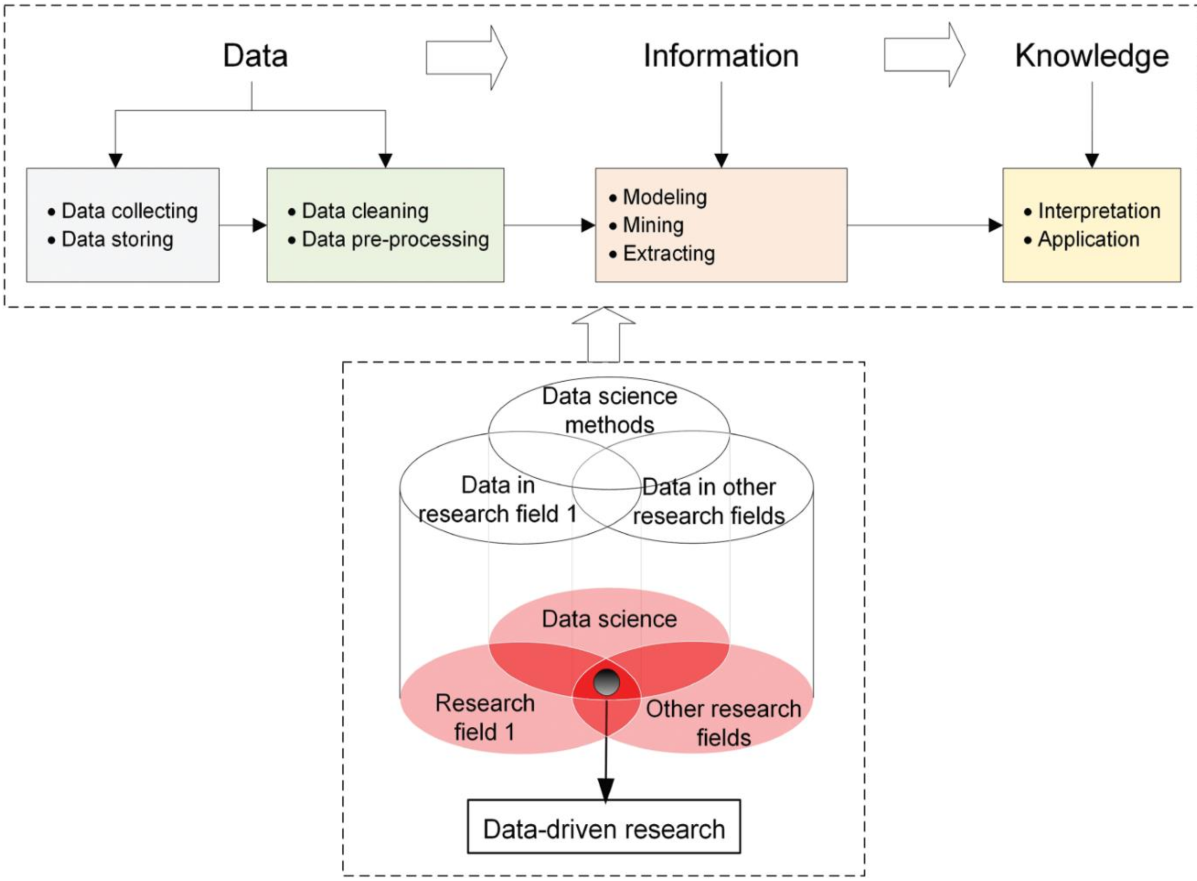


Figure 1.The data-driven research framework.

As illustrated, the data-driven research framework includes data collection, cleaning and preparation, modelling, and interpretation. Research on volatility forecasting generally falls within such a framework. This thesis examines the effects of sampling frequency on forecasting accuracy. Different steps must be taken before one is able to undertake such an evaluation, which falls under the interpretation point within the given framework above. Before continuing on to these steps, I explain how the search for literature was undertaken.

2.3 Searching for relevant literature

The aim at this stage consisted of identifying central literature and condensing the main ideas from the articles that were examined. This was then followed by identifying important terms and findings related to the frequency of sampling and accuracy of volatility forecasts.

To find the main theoretical sources, I conducted a search for articles. Access to reading materials such as books and magazines is available through sources like Google Scholar, ScienceDirect, the Wiley Online Library, and the Oria database. Search words used in locating the most relevant literature were “sampling frequency”, “volatility forecasting”, “high frequency”.

In this study, I conducted a review of the literature relevant to my thesis topic. My main reliance is on academic journals and books, not only for their literal content but also for utilizing their reference sections to find other important sources. I have also depended on specific blogs and news pieces.

The citation management tool Zotero was used to manage all references. Most of the citations are automatically imported from the sources. In specific instances where importation was not feasible, I added the references manually.

2.4 Modelling volatility

The roots of modelling the volatility of financial assets can be traced as far back as the early 1900s to the work of Louis Bachelier, whose process would later be called Brownian motion (Mandelbrot 1963). A key building block of Bachelier’s process was the assumption that the volatility of an asset remained constant regardless of the asset’s price level.

Volatility was often presented as a constant parameter in finance textbooks and research papers in the 1970s (Degiannakis & Floros, 2015). Modelling volatility based on this assumption of constant variance was further adopted into various models; among them are models based on ordinary least squares that are commonly used due to their ability to estimate how changes in one variable affect other variables (Engle 2001), and the Black and Scholes model (1973), a model that was developed to the implied volatility model. This thesis does not explore developments in this implied volatility branch.

The assumption of constant volatility was proven to be incompatible with empiric reality by, among others, Mandelbrot (1963) who illustrated volatility’s clustering effect. After the clustering phenomenon was established, it became clear that large price changes in a financial asset are likely to be followed by more large price changes (of either increase or decrease), and small price changes are likely to be followed by more small price changes. This meant that returns were not independent and identically distributed (i.i.d.)(Danıelsson, 2011). If the returns had been i.i.d., the size of a price change would not be dependent on the size of the earlier change. There was therefore a degree of autocorrelation between changes in returns at a given interval and those at earlier intervals. Models that assume constant volatility are hence incapable of incorporating this clustering phenomenon.

One of the early attempts to model volatility that challenged the assumption of constant variance was proposed by (R. F. Engle, 1982) through the ARCH model. What Engle proposed was a new way to model changes in returns. He suggested that the variability in these didn’t stay the same. Instead, it changes, and these changes depend on what had happened previously. The observed value of volatility at a given interval was the product of a random error and the square root of the conditional variance at that time.

The ARCH model assumed that the mean changes over time were zero. Further, the changes from one interval to the next did not follow a simple, predictable pattern. However, the size of the changes could be predicted based on recent past changes. Hence, the variance was serially uncorrelated but conditional. Making the ARCH model suitable for modelling the clustering phenomenon(R. F. Engle, 1982)

The ARCH model’s ability to model volatility clustering was a better reflection of reality than earlier models. The model also had its shortcomings. Among these was the requirement for lengthy lag structures to effectively capture the patterns and persistence of volatility in financial time series(Bollerslev, 1986). Such lengthy structures could result in negative variance parameter estimates, which is an undesirable outcome since the ARCH processes are based on positive coefficients(Bollerslev, 1986). Given these limitations, there was a clear practical need to extend the ARCH model to allow for both longer memory (i.e., the ability to account for volatility persistence over a more extended period) and a more flexible lag structure. This would enhance the model's ability to reflect the complexity and dynamics of financial time series data more accurately.

Tim Bollerslev's (1986) ARCH model expanded by incorporating an autoregressive structure into the variance equation, leading to the GARCH(p,q) model. As a result, the GARCH model can combine recent error terms with past volatility to predict current volatility, allowing for both short-term shocks and long-term volatility persistence. Which was an improvement to the ARCH model. Numerous methods for forecasting volatility are available, though only a few are commonly used. Choosing a model involves balancing factors like ease of use and reliability(Danıelsson, 2011). The GARCH model and its hybrid innovations that combine different features, such as incorporating realised volatility, are prominent in the volatility literature.

Other models have been developed; Corsi (2009) presents the HAR model. Though these models proclaim important improvements in their forecasting abilities, the GARCH-related models are most prominent in the literature.

2.5 Volatility measures.

The following section presents some of the volatility measures that are believed to provide better volatility forecasts. Central to this thesis is the realised volatility concept, which has emerged as an alternative measure for evaluating volatility forecasts. The focus is given to its evolution and why this measure was chosen.

Volatility is latent and not directly observable; this attribute poses unique challenges in measurement and forecasting(Danıelsson, 2011). Unlike other statistical forecasting fields, in which model accuracy is assessed by comparing predictions with direct observations, volatility remains unobservable even after the fact. The latent nature of volatility means that it must be forecast by a statistical model, a process that inevitably entails making some assumptions. As stated under the ARCH model, it is generally assumed that the mean return is zero. This simplification is due to the significantly smaller size of the daily return mean compared to the volatility. Thus, the mean is often disregarded in volatility forecasting (Danıelsson, 2011). According to Danielsson (2011) the volatility concept can be divided into two concepts: unconditional and conditional volatility. Unconditional volatility refers to the measure of volatility over a long period of time. It's the average volatility that an asset experiences over this extended timeframe, without considering variations that may occur at specific points within that period. Conditional volatility, on the other hand, is the expected volatility at a specific time, given the asset's historical return movements up to that time. It's a period-specific measure, reflecting the idea that volatility is not constant but can be influenced by recent events (Danıelsson, 2011).

Depending on the period of interest, whether past, current, or future, (T. G. Andersen et al., 2010) specify three volatility terms: notional volatility, instantaneous volatility, and expected volatility. Instantaneous volatility refers to the actual volatility at a given instant. Such a measure is dependent on the existence of continuous data records. Data is, however, sampled at discrete intervals, so the accuracy of instantaneous volatility measurement is limited. Notional volatility, in contrast, is based on the quadratic variation, which is used to measure the total variance of a process over a specific period. In the case of volatility, this entails summing up squared returns at finer intervals. This approximation becomes increasingly accurate as the partitions become smaller. Expected volatility refers to the anticipation of future volatility based on present information. Unlike notional volatility, which can be determined through observed data without the need for a predictive model, expected volatility necessitates forecasting future return variations using specific models. However, expected notional volatility can be used to forecast future cumulative squared return changes and is a key factor in determining expected return volatility(Andersen et al., 2010).

Realized volatility is an empirical measure reflecting what has occurred and is calculated using actual past data. Its calculation involves adding up the squared returns at high-frequency intervals within each period. For instance, in a 24-hour trading market, the daily realised volatility, using five-minute returns, is calculated by summing each day's 288 squared five-minute returns (Danıelsson, 2011).

Among the first to mention the concept of as-realized volatility (RV), is Robert Merton (1980). According to Merton (1980), the precision of a variance measure was influenced by the frequency at which the data was collected. This was illustrated by comparing the variability of estimates based on annual data to those derived from daily data over an identical period. Estimates from the annual data exhibited a variability nineteen times greater than those from the annual data. It was further stated that as the observation frequency increased to infinity for any fixed interval, the variance rate could be estimated without error (Merton, 1980). However, there were practical limitations to making this sampling increase. Opting for increasingly shorter observation periods introduced additional errors that outweighed the benefits of such an approach (Merton, 1980).

Others concluded that realized volatility was a better measure of ex-post volatility. The GARCH model had improved conditional volatility modelling and forecasting. In spite of the model's highly significant in-sample parameter estimates, it was deemed to explain little of the variability in the squared returns that were used as a proxy for ex-post volatility (T. Andersen & Bollerslev, 1998). Andersen & Bollerslev (1998) argued that using high-frequency data to calculate the sum of intra-daily returns at very short intervals is a measure of ex-post volatility. The realised volatility based on these intra-daily returns used to assess the ex-post-performance of the ARCH-type models gives better model performance.

The realised volatility measure was part of the progressive movement of the volatility literature toward the use of higher frequency data(T. Andersen et al., 2000). It was also part of the ongoing debate between parametric and nonparametric volatility measurement (T. G. Andersen et al., 2010). Through this perspective, the realised volatility measure was eventually further presented as a viable alternative for volatility forecasting.

Parametric volatility measures, such as those within the GARCH family of models, necessitate the estimation of model parameters as specified earlier in this text. The forecasting of volatility in these cases involves the estimation of coefficients (parameters) that need to be specified based on historical returns (Danıelsson, 2011). In their specification, these parameters primarily rely on daily squared returns. This approach has later been found to offer a weak estimation of the current level of volatility (Lyócsa et al., 2021; Zhang et al., 2019).

According (Bollerslev et al., 1994) the beginning of the shift towards nonparametric methods was a natural progression caused by the variety of ARCH models. The result was the use of various nonparametric techniques, like kernels and Fourier series, to model relationships in financial data (Bollerslev et al., 1994). However, the estimation of these nonparametric methods was challenging (Bollerslev et al., 1994), and consequently, nonparametric methods were judged to perform poorly, so Poon &. Granger (2003) chose to exclude them from their overview of different ways used in volatility forecasting.

On the other hand, research has found nonparametric models advantageous because they don't assume a fixed model structure, relying only on some general assumption of smoothness conditions. These assumptions focus on the continuity and gradual change of the data values, rather than abrupt jumps. This makes them especially beneficial when there's limited information available or when flexibility is needed regarding the underlying model(Zhao, 2008).Realized volatility, which is based on intraday return observations, offers a direct and relatively unbiased way to measure return volatility with uncorrelated errors T. G. Andersen et al. (2010). This approach enables the creation of time-series models for observed volatility. This method simplifies the process by avoiding the need to model complex intraday volatility patterns, while still leveraging the detailed information available in high-frequency data for understanding longer-term volatility trends. However, it's noted that this nonparametric approach to measuring volatility might be less statistically efficient compared to using a well-specified parametric volatility model.

The choice between a parametric approach and a nonparametric approach to volatility measurement is thus a choice between a simple, nonbiased realised measure or a complex, likely biased model estimation. This choice is not necessary since realised variance measures were directly introduced into the dynamic volatility specifications in GARCH models (R. Engle, 2002) .The more accurate estimation and the forecasting gains associated with the inclusion of realised measures into the modelling frameworks inspired a rapidly growing research activity in this area.

In an ideal market scenario without any frictions, realised volatility is considered the optimal measure of volatility (Degiannakis & Floros, 2015). In cases where the market experiences abrupt price shifts and trading microstructure noise, different measures that are better suited can be adopted. For instance, (Hansen & Horel, 2009) proposed a measure-based Markov chain theory that accommodated more of the microstructure noise, eliminating the need to omit data points to satisfy certain noise assumptions. On the other hand, Barndorff-Nielsen & Shephard (2004) introduced power and bipower variation measures. These were considered stable in the face of unpredictable and sudden price changes and were able to distinguish between the effects of overall market instability and occasional, major price spikes. Furthermore, they are able to maintain accuracy even when faced with sudden changes. Another measure was given by (Barndorff-Nielsen et al., 2008a), where they discovered that using kernel functions to smooth the data reduced the influence of market microstructure noise, resulting in a more resilient volatility measure. Realized kernels proved to be a more precise measurement of volatility when working with high-frequency data. Even though the realized kernel provided a stronger measure when dealing with microstructure noise, creating it was a significantly more complex task (Barndorff-Nielsen et al., 2008b). All these measures are potential alternative volatility measures to realised volatility.

Although the alternate measure of volatility is used when dealing with microstructure noise or abrupt price changes, there are some who consider the realised volatility to be the leading example of a high-frequency-based measure of volatility (Hansen & Horel, 2009). The central position that realised volatility has in the literature on high-frequency volatility modelling both in parametric and non-parametric approaches, and subsequent integration within the GARCH framework, makes it the measure chosen for this thesis.

Despite the challenges of using high-frequency data and realizing volatility, both have continued to be used in volatility forecasting. Additional improvements to volatility models can be found in the combination of GARCH and realised volatility models, which have been found to outperform either model used independently (Zhang et al., 2019).Such a measure, realised volatility, which was initially mainly applied as an improved way of assessing model performance, was later incorporated into GARCH models, giving improved performance in high-frequency settings(Hansen et al., 2012). This approach will be applied in this thesis as it is an improvement of the models applied by Chan et al. (2010). Combinations of this kind are deemed better suited to handle high-frequency data (Hansen et al., 2012), and thus better suited to assess the effect sampling frequency

2.6 Incorporation of realized volatility in GARCH models.

Volatility exhibits a range of distinct traits, often referred to as stylized facts, which must be considered in any forecasting endeavour. Following the recognition of these stylized facts and the development of the GARCH model, a lot of research on volatility forecasting has centred on enhancing the foundational model or devising alternative approaches. Lunde & Hansen (2005) present a list of 330 GARCH-type volatility models, showing the extent of development that has been done in the pursuit of improved volatility forecasts. This section tackles the introduction of realised measures into the GARCH model framework.

Although the GARCH (1,1) is among the simplest and most robust of the volatility models, the model can be extended and modified in many ways to accommodate different requirements and mitigate some of its shortcomings(R. F. Engle, 2001). Zivot (2008) demonstrates that the basic GARCH model, particularly the standard variance equation it uses, generally works well for analysing financial data over time. However, this basic model isn't best suited for every situation. Sometimes, it needs to be adjusted or expanded to better understand and predict the patterns and behaviours of specific financial data series.

Different methods have been proposed to incorporate realized volatility measures into modelling and forecasting processes. One proposition was to estimate a GARCH model that includes a realised volatility measure in the GARCH equation, known as a GARCH-X model (R. Engle, 2002). This was further developed into the more complete model referred to as the RealGARCH. RealGARCH (Hansen et al., 2012) was viewed as providing an improved and efficient framework for volatility modelling. The main and most important feature of RealGARCH is that it is a joint modelling of conditional and realised volatility measures, which makes possible the projection of multiple-horizon volatility forecasts. There are a multitude of models within the RealGARCH framework, for example, the refined RealGARCH; log-linear RealGARCH in Hansen et al. (2012), the RealEGARCH in (Hansen & Huang, 2016) and the most recent GARCH@CARR in (Xie et al., 2019).

Hansen et al. (2012) show that both linear RealGARCH and log-linear RealGARCH perform better than the standard GARCH and EGARCH models that only use returns to estimate volatility. Using the high-low daily price range as a realised measure, Xie et al. (2019) find that the GARCH@CARR also outperforms the return-based GARCH and EGARCH models. Although there remains no clear determination of the effects of sampling frequency on these models' performance, there's evidence that RealGARCH-type models provide the possibility of improving volatility forecasts. Furthermore, recent literature indicates a rising interest in using RealGARCH models. (Huang et al., 2017) obtain an analytical approximation formula for option pricing under RealEGARCH. Contino & Gerlach (2017) use the log-linear RealGARCH to forecast tail risk. (Banulescu-Radu et al., 2017) use the RealEGARCH model to investigate the volatility during the financial crisis. Wu et al. (2020) propose using RealEGARCH with skewness and kurtosis to forecast VaR.

The question of which variant is more efficient is addressed by Xie & Yu (2020) who compared three models within the RealGARCH framework. They find that, in line with the principle of parsimony, which states that simpler models usually provide better forecasts than more complex ones, the simpler GARCH@CARR outperforms the other two, while the log-linear RealGARCH outperforms the RealEGARCH.

Although the GARCH@CARR model is considered most efficient, the volatility measure it applies is derived from the daily price range, which is the difference between the logarithm of the day's highest price and the logarithm of the day's lowest price. Such a measure of realized volatility will not make it possible to assess the effect of different sampling frequencies, which is the aim of this thesis. Therefore, the next efficient model, the log-linear RealGARCH, will be applied.

In the studies mentioned above, different sampling frequencies are employed to calculate realised volatility. Wu et al. (2020), Contino & Gerlach (2017) use a 5-minute frequency, while Xie & Yu (2020) use a daily range. Hansen& Huang (2016) on the other hand, construct six realised volatility measures that differ in terms of the sampling frequency, ranging from 15 seconds to 20 minutes. As mentioned in the next segment, some have managed to sample at lower frequencies than the prevalent 5-minute range.

2.7 Summary

Thus, one can make the following summation: the measurement and modelling of volatility are vital for multiple financial applications. Different models and measurements have been developed that improve modelling and forecasting. However, the validity of these approaches and findings for different scenarios remains a scarcely explored topic, as most new developments tend to introduce novel models, new measurements of volatility, new scenarios, or a combination of them. These approaches are dominant compared to those that test the models and measurements on other datasets. Furthermore, while many studies use the 5-minute sampling frequency as an optimal way to manage eventual microstructure noise, there is evidence from others that in some cases this can be lowered, but few examine what they gain from this increase.

Chan et al. (2010) looks at this gain and conclude that the sampling frequency has little effect on the forecasting accuracy compared to other factors like the inclusion or exclusion of non-trading hours in the calculation of the realized variance. Meanwhile, others find that, generally, model performance increases when the sampling frequency is increased. More specifically, Ewald et al. (2023) find that higher sampling frequencies yield better model performance.

3. Methods

A research method has been defined by Vilhelm Auberts as "*a procedure, a means to solve problems and gain new knowledge. Any means that serves this purpose belongs in the arsenal of methods.* ((Furseth & Everett, 2022 ,137), The translation is my own). This section provides an overview of the steps taken to evaluate the effect of sampling frequency on forecasting accuracy. Starting with laying out the research design, then proceeding to data cleaning and preparation, involving the construction of different data sets based on the frequency at which they are sampled. Subsequently, the model is fitted to the corresponding dataset while keeping the model specifications constant to ensure that differences in forecasting performance are due to the data frequency, not model specification changes. The result of this forecasting exercise will then be evaluated, with the aim of assessing the effect of sampling frequency on forecasting accuracy.

3.1 Research design.

The relationship between sampling frequency and forecast accuracy is not quite clear. There are those, like Chan et al. (2010), who find the effect minimal in comparison to the effect of including or excluding overnight trading hours. However, they use only a couple of sampling frequencies to make a definitive conclusion. On the other hand, others find that increasing the sampling frequency, while it requires careful monitoring of the microstructure noise, does contribute to increased forecasting accuracy (Ewald et al., 2023). The latter indicates a correlation between the sampling frequency used and forecasting accuracy. Additionally, in statistical theory, the assumption is often that the bigger the sample, the better the inferences made from it. A similar view is adopted in this thesis; the research design adheres to a causal research design as it aims to investigate the effect of sampling frequencies on the precision of volatility forecasts.

According to Walliman (2022), causal design is one of two classes within correlation research design, with the other being association, which denotes instances where there is some kind of influence on the other but not a cause of change. The relationship between two concepts can be non-existent (no correlation), positive (increases in one cause increase in other, decrease results in decrease), or negative (increase in one lead to decrease in other, or vice versa). Following the results in (Ewald et al., 2023), the indication is that the relationship between the sampling frequency and forecasting accuracy is positive. On the other hand, the results from

In research situations where the goal is to determine differences or correlations between two or more variables, a popular approach is to use statistical methods to confirm the existence of proposed relationships. This involves hypothesis development and testing, which are commonly used to assess the correlation between variables in theoretical models(Zou & Xu, 2023). Approaches of this kind usually distinguish between independent variables and dependent variables. The independent is viewed as having an effect on the value of the dependent. This thesis investigates how sampling frequency affects volatility forecasting accuracy. In this case, the forecasting accuracy is the dependent variable, and the sampling frequency is the independent one. If this view holds true, a model whose parameters are estimated using higher frequencies will ultimately have better forecasting accuracy than one estimated using lower frequencies. To evaluate this, I built 54 models using the same model specifications, with the only differences being the sampling frequencies of the returns and the realized volatility used in estimating the models. After having estimated the models and run one-rolling, one-day-ahead forecasts for each sampling frequency, I proceeded to compare the forecasting accuracy of these models. This assessment provides the answer to whether the sampling frequency has an effect on accuracy. This approach shares similarities to that employed by (Lunde & Hansen, 2005). Although they compare the performance of different models on the same data set, my approach is to compare the same model on the same data set but with different sampling frequencies.

More complex and systematic analytical methods are becoming more popular because of the increasing complexity of research topics(Zou & Xu, 2023). Zou & Xu (2023) present different techniques for analysing quantitative data, among them the method of data-driven research. Data-driven research refers to the emphasis on utilizing data and data science as tools for conducting research. Central to this approach is the progression from data to information to knowledge. The combination of abundant data and sophisticated analytical methods has led to the emergence of novel interdisciplinary research. Through data-driven research, scientists integrate data science with the distinctive features and requirements of their field of study to facilitate the transition from data to knowledge. As alluded to under the theory part, in the context of forecasting volatility, in order to arrive at an evaluation of the effect that sampling frequency has on forecasting accuracy, other steps have to be accomplished. These steps include cleaning, processing the data ready for modelling, fitting the model, forecasting volatility, and then evaluating the results from the forecasts. The explained process is quite similar to the different steps given by the research-driven framework. Hereafter, in this methods section, I will proceed to present the steps taken under the data-driven research framework.

3.2 Challenges

In order to evaluate the effect of different sampling frequencies, I decided to create different data sets based on sampling frequency. These are intervals of 1-750 respectively. These sampling frequencies were then to be used in the estimation of models whose forecasting accuracy would subsequently be compared to each other, while the model specifications were held constant. While estimating the model, it became apparent that some frequencies would not converge; only 54 models below 95 minutes could be used, and everything after 95 minutes failed to converge. Failures in achieving convergence may result from various factors, such as the use of unsuitable solver algorithms when fitting the model, missing values, non-stationary data, or not normally distributed data, among others. I conducted tests of stationarity and distribution in addition to checking for missing values and outliers to confirm that the returns utilized in defining the models and calculating realised volatility meet these criteria. I also decided to utilize the "hybrid" solver as advised for the “rugarch” package when fitting models. This algorithm, which serves to fit the model parameters, is assessed to prevent many potential convergence issues that may arise with other solvers (Ghalanos & Kley, 2023). Despite choosing to do so, the fitting process could not invert the Hessian matrix at multiple sampling frequencies. If a Hessian matrix cannot be inverted, it can be quite challenging or impossible to fix while still using the selected model and data, since the desired inverse simply does not exist(Gill & King, 2004). Most recommend actions like reconsidering the model, making changes to it, conducting the analysis again, or collecting additional data. Most, if not all, of these potential fixes were unapplicable. In order to keep the models similar and have sampling frequency as the only difference amongst them, the solver algorithm had to remain the same. Getting more data was also not possible, though I could extend the number of trading hours included. I therefore decided to drop these sampling frequencies.

3.3 The method chosen for data collection and analysis.

Data is considered the fundamental and central component of any study(Zou & Xu, 2023). The types of data that can be collected in a research setting can be classified into primary and secondary data, respectively. The distinguishing factor for which of these categories the data belongs is the gatherer's closeness to the source of information. Primary data is gathered directly from the original source using questionnaires, surveys, interviews, or observations. Secondary data, on the other hand, often stems from data that has already been collected by others. Such data may include financial reports, budget reports, etc.

In this thesis, the use of primary data is not easily applicable. The likelihood that a primary source, which is often a person, can transmit 16 years' worth of trade data at 1 minute's frequency is not existent. Therefore, secondary data is used. Using secondary data adds requirements for the user to check the accuracy and reliability of the sources from which the data is collected before using them. It is vital for the user to check the relevance of the data for the study in addition to accuracy, credibility, and reliability aspects related to how the data were collected, processed, and stored(Walliman, 2022). These steps are increasingly relevant as there are reports of increased academic scams related to poor vetting of secondary data(Cookson, 2023).

The data utilised in the thesis is trading data for Brent crude oil from ICE. The University of Applied Sciences in Inland, Norway, has been afforded access to this high-frequency data, which was sampled at a 1-minute frequency. Despite the lack of oversite on the collection, cleaning, and storage processes of the data before I was able to access it, the assumption in this study is that since the data is collected and initially handled by the university, it fulfils the accuracy, credibility, and reliability requirements. Although the data showed some unusual characteristics that make its reliability doubtful. Specifically, there were some instances where the price of oil was registered as equal to zero and others where it was lower than the actual price posted online.

When it comes to the relevance requirement, the choice of using this data rests on two things: 1) the will to do the research in another setting different from those commonly used. A lot of the past studies within the volatility forecasting field have used data from currency exchange rates (T. Andersen & Bollerslev, 1998; Chaboud et al., 2010) or stock market assets and indexes like the Hang Seng Index(Chan et al., 2010) or S&P 500(Buncic & Gisler, 2017; Gulay & Emec, 2018; Huang et al., 2017). High-frequency data analysis requires that the data have enough transactions to be sampled at high frequency. If an asset has very few transactions taking place, then there's no point sampling at high frequency since the price will not change often enough to potentially lead to multiple returns equal to zero. Oil is a highly traded commodity because it is the most crucial trade commodity globally. (Ewald et al., 2023). Additionally, the data ranges from January 2004 to October 2021 and provides price data sampled at 1-minute. The expanse of the data set and the liquidity of oil make this dataset relevant for this study.

For data analysis, the data-driven research framework has several analytical approaches, among which are descriptive analytics, which is a method of statistical analysis used for interpreting data and gaining an understanding of patterns within it. Descriptive analytics involves applying statistical calculations to give an overview of the data. Further, data visualisations also provide insight into exploratory data analysis. This method is primarily used to clean and prepare data. It is also applied to evaluate whether the cleaned data is suitable for modelling and meets GARCH requirements (centrality and normal distribution). Another method of analysis within the data-driven framework is time-series analysis. In general terms, this method is usually utilized to provide an understanding of patterns or trends and to track fluctuations in a time series. After this information is extracted from the data, it may be useful for making forecasts. Time series analysis is the central analytical method used in the thesis.

3.4 Data cleaning and preparation

The cleaning of data for meaningful analysis is often a necessary requirement, depending on the aims of the study. Such data cleaning involves choosing relevant variables, handling duplicate values, missing values, outliers, etc.

The data set initially has 21 variables. I start by evaluating the information carried by these variables and its relevance to the study I want to conduct. In a time series analysis based on the price of oil during the period of 2004–2021, only two variables are relevant: the timestamp and the recorded prices. I therefore proceed to trim the data and remain with only two variables.

The trading hours on the ICE are from 00.00 to 22.00. When examining the data, it became evident that some trading hours had missing values. These missing values are mainly concentrated between the 01.00 and 09.00. In statistical learning techniques, it is vital to find a viable way of handling missing values to prevent possible problems with model fitting. One method to address this issue is to remove rows with missing data and analyse the remaining complete rows. This can also be seen as excessive and may not be practical, depending on the percentage that is missing. Another option for dealing with missing values is to replace them with the average of the available data points (James, 2013). To deal with the missing values, I decided to focus on the trading hours between 09:30-22:00. This approach is also compatible with Chan et al. (2010) who compared realised volatility calculated with the inclusion of non-trading against that calculated without. They point out that their decision to exclude non-trading hours had a bigger impact on forecasting accuracy than sampling frequency did. Through the exclusion of these hours with missing values, which coincide with the non-trading hours excluded in Chan et al (2010) allows to focus solely on the effect of sampling frequency on forecasting accuracy. The drawback to this is that the characteristics of the underlying asset are not fully represented. Another change made to prepare the data pertains to dealing with what can be termed extreme values. As is visible from the price evolution plots provided in Figure 2, there was an extreme change in oil prices during May 2017, where prices dropped to zero 39 times during that period and 3 others around 25 dollars for only 1 minute at a time.

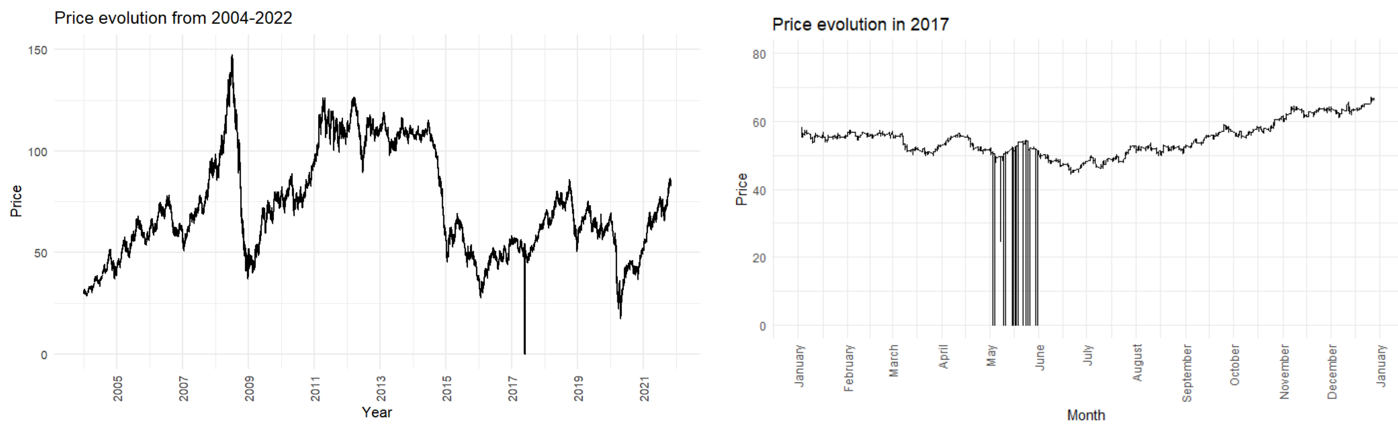


Figure 2. Prices before cleaning.

From historical data on ICE Brent crude oil, there appears to be no such occurrence of zero oil prices during this period, and the price never drops below 45 dollars. (*ICE Brent Crude Oil Front Month price information - FT.com*, u.å.). As these data points are during trading hours and not to close to each other, simply removing them or replacing them by the mean wouldn't be the optimal solution in this case. Linear regression is a better solution because it replaces the missing values at levels that fit better with the adjacent values.

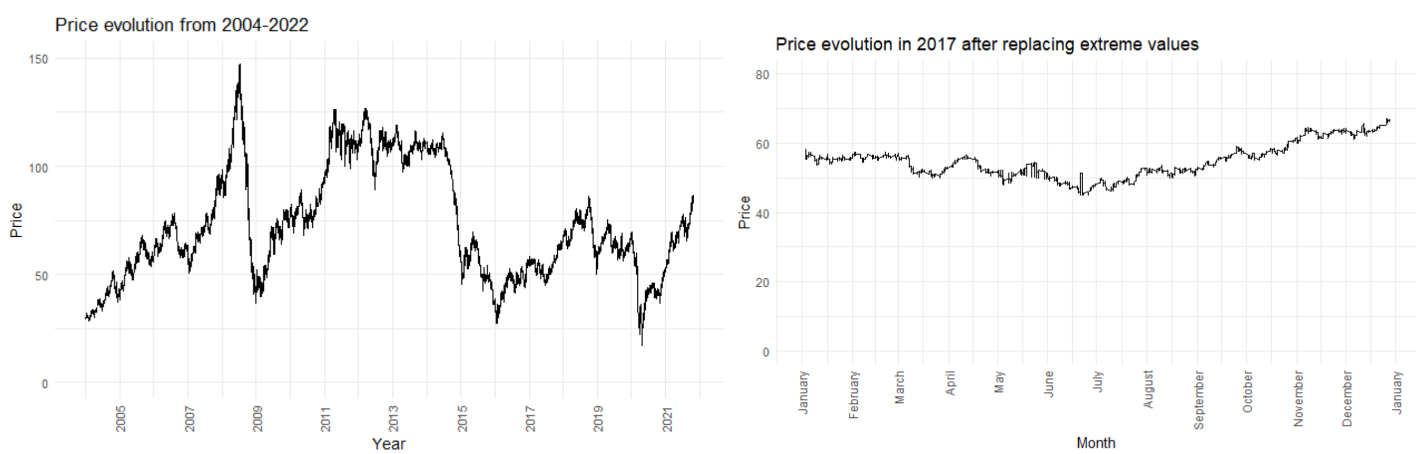


Figure 3. Prices after cleaning.

Having constructed a continuous time series with no missing values, the next step is to create different data sets based on the sampling frequency. The data is originally sampled at 1-minute intervals; to create a data set of, for example, a 3-minute frequency, the price is extracted at three-minute intervals. This creates a data set with missing values that must be omitted. Similar sampling is then undertaken for all frequencies, leading to data sets of different lengths.

The next step is to calculate returns. When analysing financial data, returns are usually the primary focus instead of prices. One of the main explanations given for this preference is that returns have statistical properties that make them more manageable in time series modelling, such as stationarity (Danıelsson, 2011). The literature distinguishes between two types of returns, simple and log returns. Simple returns are the difference between the price of an asset at one period in time compared to a previous period, for example given 1 minute sampling, returns at minute t can be calculated by finding the difference in price at time t and the previous minute given as $t - 1$. One benefit using of log returns is that they possess a time-additive characteristic which lets us present the log returns of multiple periods as the sum of one period. To illustrate, given the intraday log returns of a financial asset, one can calculate the daily returns by adding up the intraday returns. This characteristic makes it easier to calculate and examine returns across different time periods.

In mathematical terms, logarithmic returns are expressed as shown under

$$r_t = log\left(\frac{P_t}{P_{\{t-1\}}}\right) = log(P_t) - log\left(P_{\{t-1\}}\right)$$

Creating a measure of realised volatility is done by adding up the squared log returns from the different sampling frequency dataset(T. Andersen et al., 2000).

$$RVar_t = \sum_{i=1}^{M} r_{t,i}^2$$

This summation is conducted each day. For instance, given that the data has been restricted to 12,5 hours a day, the daily volatility calculated from five-minute returns is determined by adding up the squared returns of 150 five-minute intervals within each day. This summation makes the created datasets eventually have the same length, creating daily returns for 4549 trading days for every sampling frequency. With the returns and the realised volatility for each sampling frequency calculated, the models can now be estimated.

3.5 Model estimation.

The "rugarch" package by (Ghalanos, 2023)is used for estimating the models. This package provides different solutions that are specifically designed for both univariate and multivariate GARCH modelling. In this package, all the methods, from estimation to filtering, forecasting, and simulation, have been included. The model designated as the realGARCH model in the package is the equivalent of the loglinear realised GARCH model by Hansen et al. (2012. The realized GARCH model with a log-linear specification is defined by the GARCH and measurement equations shown below.

$$\log\ h_t = \omega + \sum_{i=1}^{p} \beta_i \log\ h_{t-i} + \sum_{j=1}^{q} \gamma_j \log\ x_{t-j}$$

$$\log\ x_t = \xi + \phi \log\ h_t + \tau(z_t) + u_t$$

In statistical learning, choosing the right model is a balancing act between models that may be too simple to not be able to explain the variety in the data, thus underfitting, and models that may be too complex and capture more than the information of interest, resulting in overfitting. The model specifications in this thesis are kept simple to avoid overfitting. This parsimonious approach is behind the choice of the realized GARCH model, where the GARCH component consists of GARCH (1,1). On the other hand, the simplicity of this model, given the complexity of time series data, could lead to instances of underfitting. Adding ARMA (0,0) to the GARCH (2,2) model enhances its complexity. Furthermore, the realGARCH model includes realized volatility measures in the variance equation, adding extra parameters. More intricate models, which have a greater number of parameters to calculate, are naturally more susceptible to overfitting. Overfitting happens when a model includes noise in the data as though it were a real pattern, resulting in underwhelming forecasting results on new data(James, 2013). Having a higher number of parameters in a model increases the likelihood of overfitting to the in-sample data, capturing noise that does not transfer well to new, unseen data. Nevertheless, the GARCH (2,2) and ARMA (0,0) models are quite parsimonious as they focus on achieving more with fewer parameters, which can reduce overfitting to some degree.

A common practice when training models is to divide the data into two such that the model can be trained on part of the data and the rest of the data can be used to test the model's performance. The distribution, when splitting the data, often involves allocating 80% of it for training and the remaining 20% for testing. The data contains a total of 4549 trading days. By using 3549 days to train the model and keeping out 1000 for out-of sample testing, a split of approximately 78% for the training data and 22% for the testing data. The models' performance at predicting the values in the testing data determines their out-of-sample performance, or, in other terms, their forecasting accuracy. As a result, this accuracy will be attributed to the underlying sampling frequency, since all other things are held equal.

Dividing data into training and test sets is a crucial element in assessing the accuracy of prediction models. Two popular methods to accomplish this are: rolling forecasts (moving window) and expanding windows. In the rolling forecast, the model is trained by advancing through the training dataset while consistently adapting to a set quantity of the latest observations. When new data is obtained, the oldest data point is removed from the training set, and the new data point is included, keeping the training set size constant. This technique is especially valuable for time series data in which recent trends or patterns hold more significance for predicting future outcomes(Montgomery et al., 2008). The other method is the expanding window method, where the training dataset begins with a small size and gradually increases by incorporating additional observations as it progresses. This implies that with each new prediction, the model gains knowledge from additional data. The expanding window method is applicable in cases where gathering additional data as time progresses can lead to a deeper comprehension of the fundamental patterns or when predicting long-term trends is crucial. In situations where the price changes rapidly, a moving window is more appropriate. I find that there is no huge difference in the outcomes of each model and decide to use the simpler expanding window method, which takes less computational power and time.

3.6 Model evaluation.

Model evaluation can be performed through three evaluations: evaluation of model adequacy, also known as a diagnostic check; model parsimony; and error criteria(Bonakdari & Zeynoddin, 2022). These assessments are undertaken to evaluate and quantify the effect of sampling frequency on forecasting accuracy.

Model adequacy evaluation assesses the models' residuals. The aim is to check whether they are normally distributed, independent, and not affected by periodicity, thus fit for modelling (Bonakdari & Zeynoddin, 2022). The inspection of residual plots is undertaken in order to confirm that the estimates are unbiased. A test that can be used to determine residual independence is the ACF (auto-correlation function). ACF is a function that calculates the auto-correlation between a series and its different lags. Components of a time series may include trend, seasonality, cyclicality, and residual. When determining correlations, ACF considers all of these elements. The different ACF plots provided in Appendix Section 2 indicate that all the models have their residuals within 95% significance intervals apart from the first lag, thus showing independence in residuals and the adequacy of the developed model.

One of the objectives of modelling is to create the best-suited models with the simplest structure. Such parsimonious models tend to provide more precise predictions due to their lower likelihood of overfitting the initial dataset. All the models in this study have the same number of parameters and are complexity-wise similar. The only difference amongst them is the sampling frequency used to calculate intraday returns and the realized volatility used in their estimation. The Akaike information criterion can be used for model selection and model structure simplicity evaluation. In addition to model parameters, the AIC also includes model residual variance in its calculation. Hence, the metric is applied to assess the effect of sampling frequency on model parsimony. Like AIC, the Bayesian information criterion is also used in model selection. For both metrics, the model with the lowest value is seen as the most simplistic(Bonakdari & Zeynoddin, 2022).

Measuring the model's performance is done by finding the difference between the model's forecasted values and the actual values in the test data set. The differences are often expressed using two metrics, the MSE and MAE. The mean squared error is calculated by squaring the errors before adding them up; this method penalizes bigger errors more than smaller ones. It's mathematically expressed as shown under

$$MSE = \frac{1}{n}\sum_{\{i=1\}}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

The mean absolute error is calculated by summing up the absolute values of the errors, which provides an indication of the average error size. The mathematical equation is given below.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | Y_i - \hat{Y}_i |$$

Smaller values suggest the model's predictions are closer to the actual data, indicating a higher level of accuracy in the model's performance(Bonakdari & Zeynoddin, 2022) are used in the next section where the statistical results of the forecasting exercise are presented.

4. Results and discussion.

Based on the data that is split into a training and testing set, forecast accuracy can be assessed for both in- and out-of-sample performance from these two sets. In this section, I will present and examine the forecasting findings within and outside the sample.

4.1 In sample results

Insight into how changes in model fitting are affected by sampling frequency can be gained by understanding the correlation between these metrics and sampling frequency. A graph displaying the AIC and BIC metrics at different sampling frequencies is shown in figure 4. The sampling frequency is seen to have an effect on both AIC and BIC. Observably, lower AIC and BIC values are seen with increased sampling frequencies and these increase as the frequencies recedes.

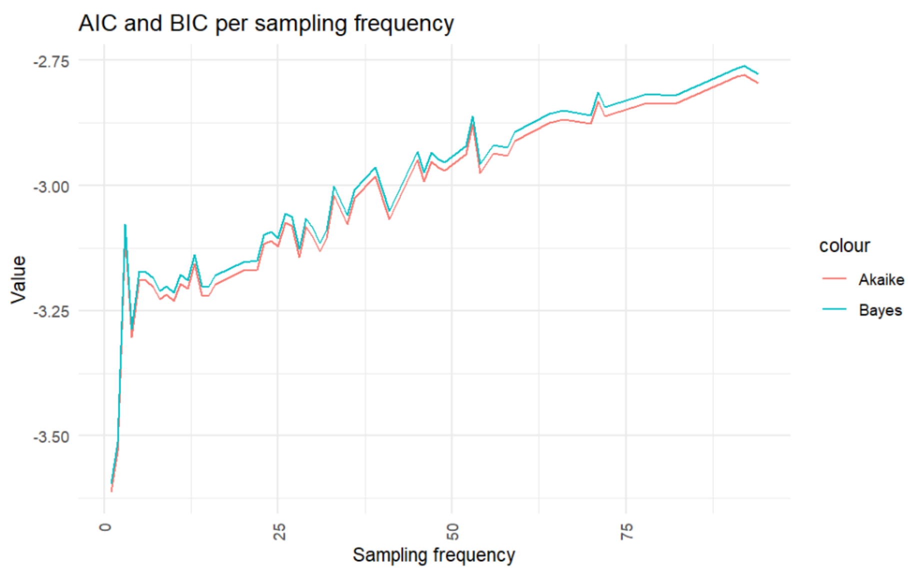


Figure 4. Changes in Akaike and Bayes Criteria

From the diagram that displays the log-likelihood values, it can be seen that the log likelihood goes up as the sampling frequency increases. A greater log-likelihood typically suggests a model that more accurately captures the data. This indicates that higher frequencies could be gathering additional data, thereby increasing the model's fit.

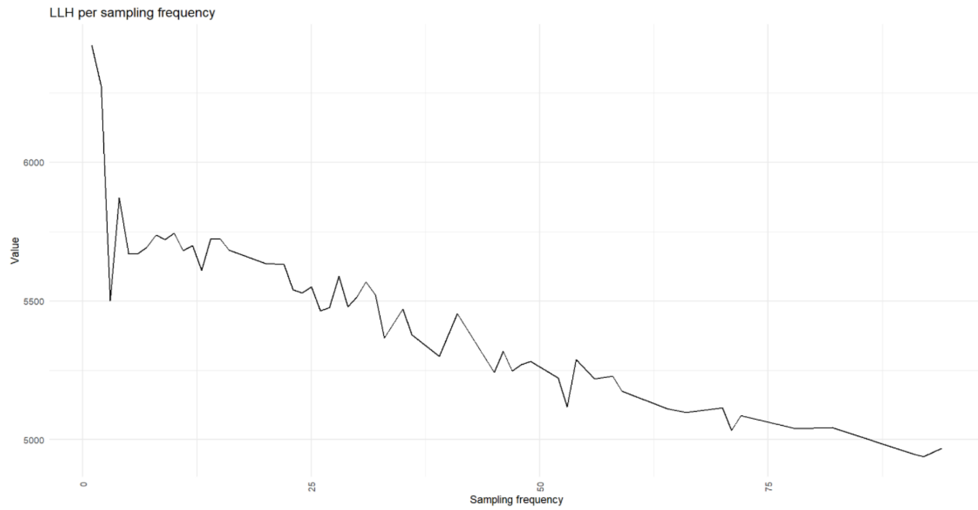


Figure 5. Changes in loglikelihood.

The table below contains the Spearman correlation coefficients for AIC, BIC, LogLikelihood, and sampling frequency. The findings indicate a significant relationship between sampling frequency and both AIC and BIC metrics. The loglikelihood results show a clear inverse relationship. These signals suggest that a higher sampling frequency can enhance the accuracy of the models. This aligns with research showing that higher sampling frequencies improve the accuracy of measuring notional volatility.

Table 1.Correlation coefficients of fit and sampling frequency.

Measure	Correlation coefficients
AIC	0.9464507
BIC	0.9464373
Loglikelihood	-0.9464339

4.2 Out of sample results

The main focus of assessing the effect of sampling frequency is the influence it has on a model's out-of-sample accuracy. Out-of-sample performance is based on the model's ability to forecast values in data that were not included during the model fitting process. This performance is then taken as an indication of the model's performance on subsequent data sets.

Having only one observation for each sampling frequency makes it difficult to perform traditional statistical tests like ANOVA, which require multiple observations per group to analyse the effect of sampling frequency on MSE and MAE. However, I am able to examine how the MSE and the MAE evolve with changing sampling frequencies through visual and regression analysis.

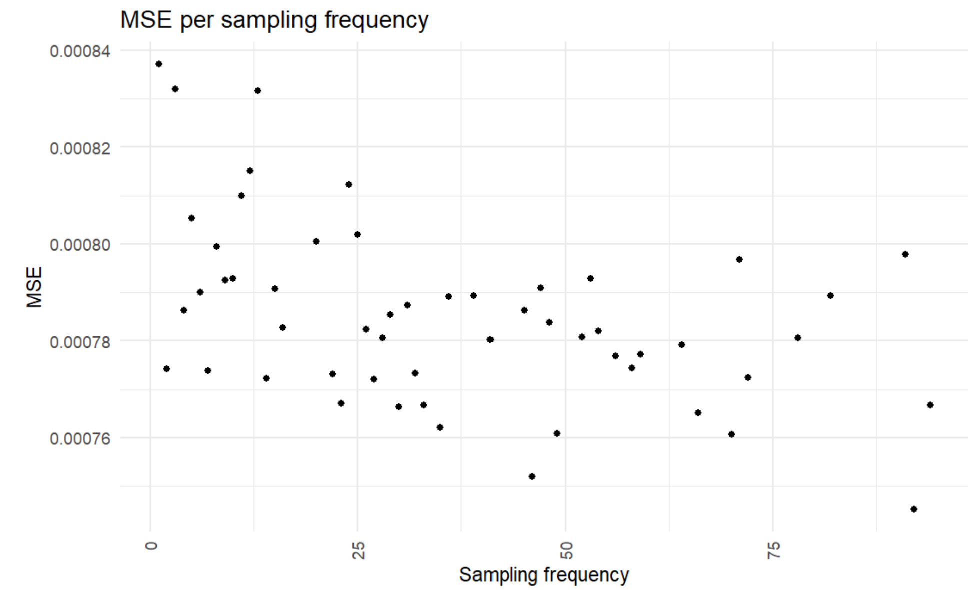


Figure 6.Scattergram of MSE

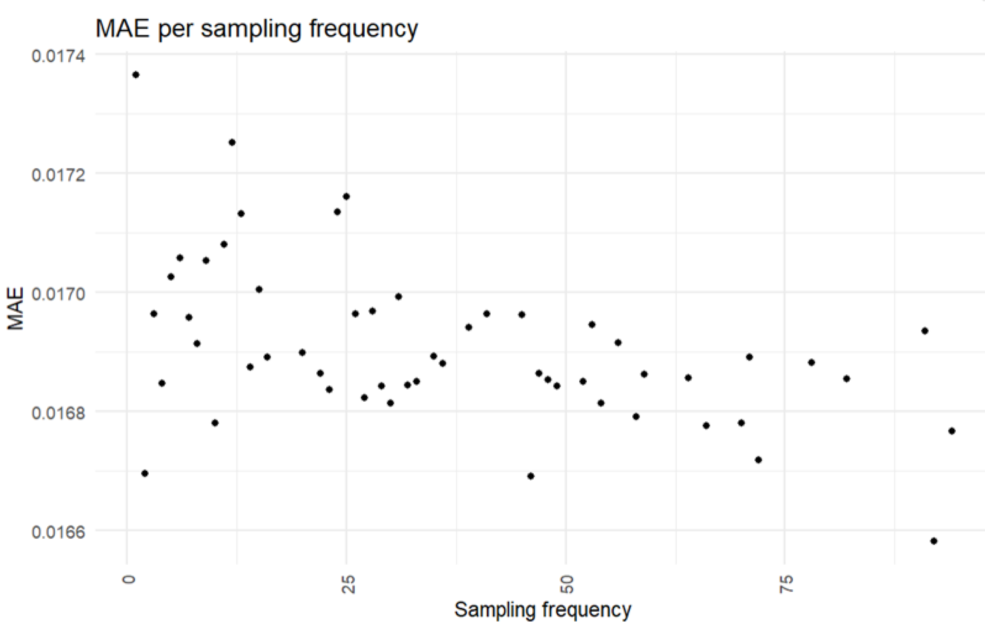


Figure 7.Scattergram of MAE.

The scattergrams above illustrate the variation of MSE and MAE values at different sampling frequencies. When observing visual representations, it seems that at higher sampling rates (smaller data intervals), both the MSE and MAE values vary without showing a definite connection between changes in sampling rate and changes in forecasting error. As the frequency decreases, there seems to be a developing trend where the error measure declines. This observation is better illustrated in figure 8 below with the help of a trendline that shows tendencies in the data.

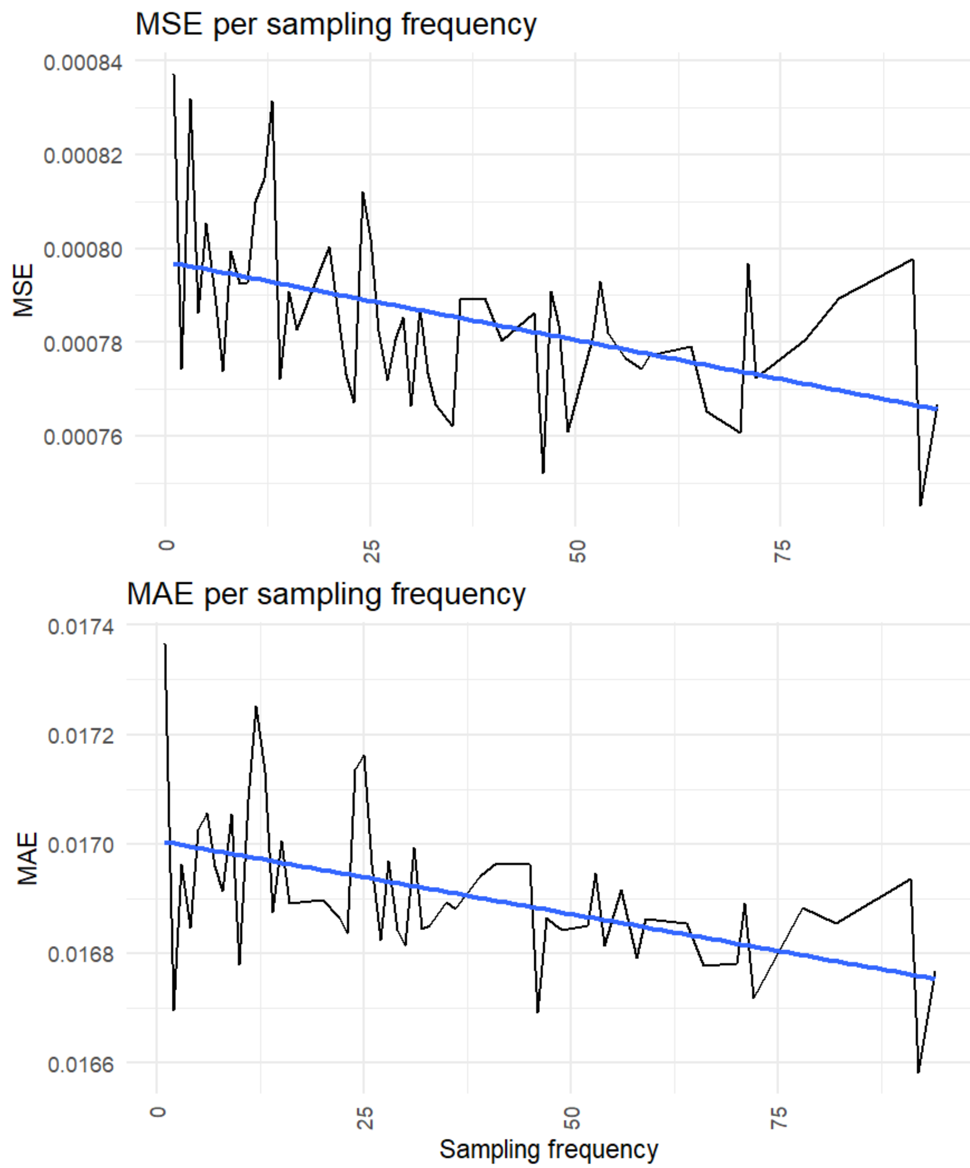


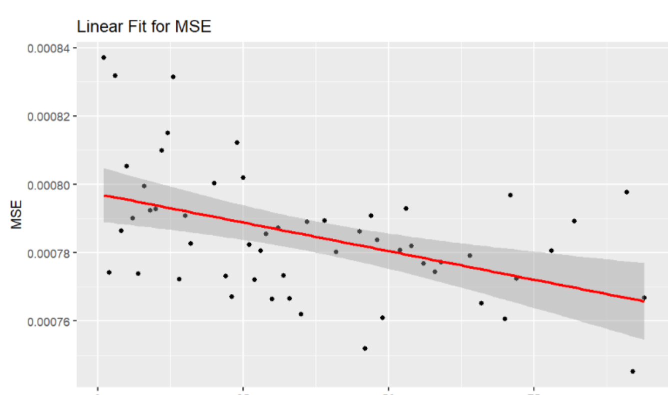
Figure 8.Trendlines of MSE and MAE.

The pattern shows that as the sampling frequency decreases, the model errors decrease. The visual illustrations can be further supported by correlation coefficient calculations between the sampling frequency and the two measures. These are provided below in Table 2. The negative signs of the correlation coefficients indicate that when the sampling frequency decreases (longer intervals between samples), there is a slight decrease in both MSE and MAE. With the p-value being lower than 0.05, the true correlation is not equal to zero in a 95% confidence interval.

Table 2.Correlation coefficients and p-values of MSE and MAE

Measure	Correlation coefficient	p-value
MSE	-0.46	0.0004432
MAE	-0.50	0.0001027

Using regression techniques to model the relationship between sampling frequency and error metrics allows for further analysis(Walliman, 2022). The goal here is to determine whether there is a statistically significant linear trend in the changes in MSE and MAE with sampling frequency. It also serves to quantify this relationship and examine the strength of the correlations. I generate linear and second-degree polynomial regression models for MSE and MAE as a function of sampling frequency. The polynomial regression models are created in cases where the relationship between sampling frequency and the MSE or MAE is not linear.



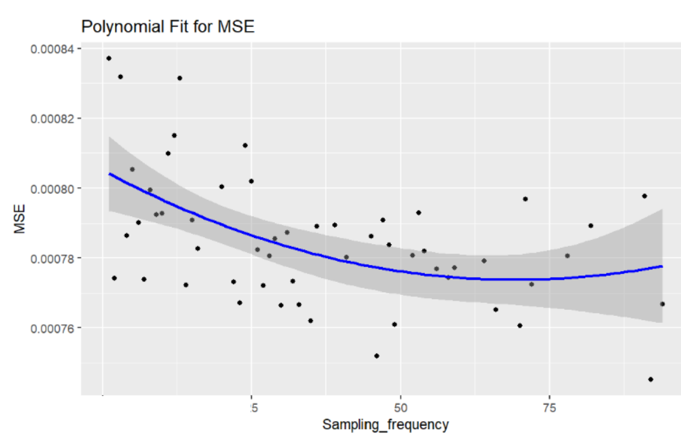


Figure 9.Linear and Polynomial fit of MSE.

Table 3.Adjusted Pearson's R for MSE regression models.

Regression model	Adjusted R^2
Linear model MSE	0.1979
Polynomial model MSE	0.2403

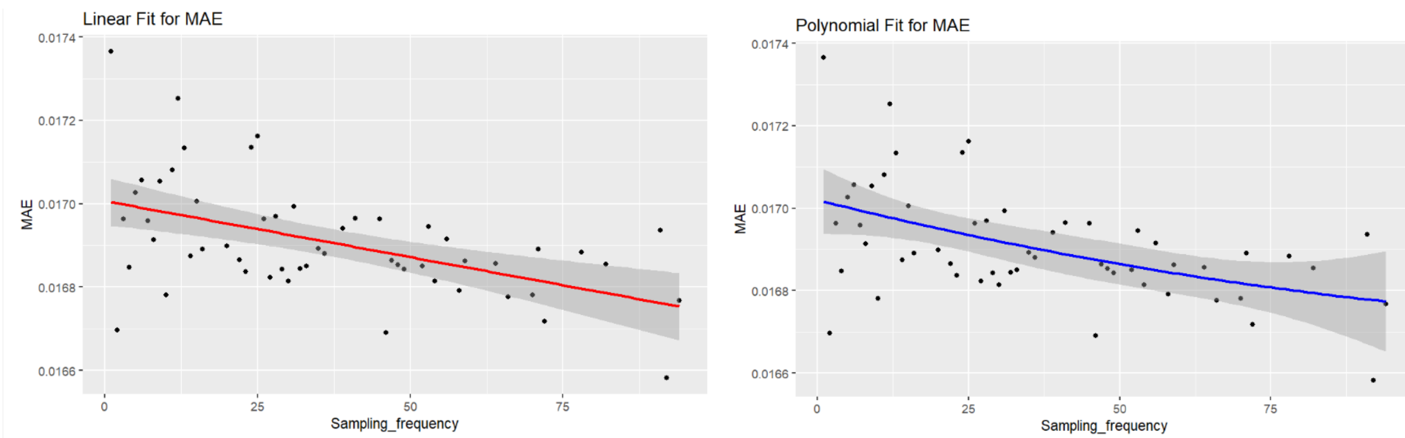


Figure 10.Linear and Polynomial fit of MAE.

Table 4.Adjusted Pearson's R for MAE regression models.

Regression model	Adjusted R^2
Linear model MAE	0.2395
Polynomial model MAE	0.2277

The adjusted R^2 values of the linear regression are around 0.20 for MSE and 0.24 for MAE. These indicate that a modest portion of the variability in these metrics is explained by changes in sampling frequency. The adjusted R^2 values for the polynomial regression are also quite similar to the linear regression values. Further analysis is therefore focused on the linear regression.

The graphs for both MSE and MAE suggest that there is a decrease in the error metrics as the sampling frequency decreases. This aligns with the negative correlation coefficients found earlier. Further evidence is provided by looking at the summary of linear regression, where MSE and MAE are predicted using sampling frequency. Table… below shows the coefficients for sampling frequency. The -3.348e-07 for MSE and -2.689e-06 for MAE indicate that for each unit decrease in sampling frequency, MSE and MAE decrease by this amount, respectively. The effect sampling frequency has on accuracy is better expressed by looking at the coefficients' values. As both P values are considerably low (below a standard alpha level of 0.05), I discard the null hypothesis that the coefficient is zero. This implies that there is a statistically significant correlation between the sampling frequency and both error metrics.

Table 5.Summary linear regression MSE and MAE.

Regression model	**Coefficient sampling frequency**	**Standard Error**	**t-static**	**p-value**
Linear regression MSE	-3.348e-07	8.925e-08	-3.752	0.000443
Linear regression MAE	-2.689e-06	6.393e-07	-4.206	0.000103

The results here appear to indicate that as the sampling frequency was reduced, the model was able to predict the volatility more accurately. This result is different from what the literature indicates, where most point to increased forecasting accuracy as the sampling frequency increases. It should be pointed out that the model does not converge after 95 minutes. It is therefore unclear whether these results hold for all sampling frequencies in this data set or only the interval between 1-95 minutes. The 54 models that were used in this study represent only 7,2% of a potential 750 minutes between 09.30- 22.00. There could be a minimum point that is not included in this study, after which the error metrics increase.

Other possible causes of this surprising result could be related to the model or the data. Raising the sampling rate generally results in a higher number of collected data points. According to the bias-variance trade-off (James, 2013), more data points can lead to various outcomes. Increased data generally offers a more thorough understanding of the phenomenon being studied, potentially improving a model's learning capabilities and decreasing errors in certain situations. Nevertheless, increased data points may also lead to an amplified presence of noise or variability, particularly when the extra data does not aid in clarifying the fundamental trends. If the model complexity is not properly addressed, it could potentially lead to an increase in error metrics. If the model lacks complexity to deal with greater variability or patterns from higher sampling frequencies, MSE and MAE could rise, as the model cannot capture more subtle variations in the data. It is also possible that highly intricate models can cause overfitting, where the model picks up on the noise in the training data instead of the true signal. This usually leads to reduced errors on the training data but increased errors on unseen or test data. The patterns observed in the in-sample results and those in the out-of-sample results may be the result of overfitting that gives a good fit in-sample and increases error on the test data. I have tried to apply more complexity to the model but haven't found any improvement. The other issue is the reliability of the data, which unexpectedly included prices of oil equal to zero and some instances lower than what the source of the data has published online(*ICE Brent Crude Oil Front Month price information - FT.com*, u.å.)

Chan et al. (2010) points out that the gain in accuracy from increasing the sampling frequency from five minutes to 30 seconds is minimal. I look at the percentage change between MSE and MEA for the 54 models. A summary of the percentage changes is given in table… below. The data shows that the conclusion by Chan et al. (2010) could have some value based on the average change, which is only 0.117% for MSE and even much smaller for MAE. Some intervals do have large changes, but its challenging for research to find out exactly which interval gives a worthwhile improvement. Furthermore, the cumulative percentage change of all 54 models is only around 6% for MSE and around -3% for MAE.

Table 6. Percentage change summary.

Measure	Mean	Median	Min	Max
% Change MSE	-0.117	-0.160	-7.510	7.450
% Change MAE	-0.06019	-0.00500	-3.83000	1.79000

5. Conclusions

I set out to explore the effect of sampling frequency on forecasting accuracy. The effect on forecasting accuracy was explained to have a cause-and-effect aspect where changes in sampling frequency would either increase or decrease forecasting accuracy. In addition to this, the effect had a magnitude aspect where the gain or loss would have a significant size to justify the incurred computational and time costs. The data research framework, which encompasses a fair share of the previous research in the field, was used.

The cause-and-effect aspect of the relationship between sampling frequency and forecasting accuracy was assessed in-sample and out-of-sample. Regarding the in-sample assessment, I was able to find evidence that sampling frequency affected how well the model fit. The relationship in this case was that the higher the sampling frequency, the better the model fit. Regarding the out-of-sample assessment, evidence was found that sampling frequency had an effect on forecasting accuracy, albeit in a surprising way. The relationship found in this study is that increasing sampling frequency negatively affects modelling accuracy.

As for the magnitude aspect of the changes in sampling frequency, I find that the average changes in MSE or MEA are so small that it would be difficult to foresee a significant gain from changes in sampling frequency.

However, doubts related to the reliability of the data and uncertainty about the models reduce the possibility of drawing absolute conclusions about the overall effect of sampling frequency and forecasting accuracy in this setting. The results and analysis presented are restricted to sampling frequencies ranging from one minute to 95 minutes, as well as the dataset used.

5.1 Implications and further research

Although there are limitations to the scope of the results and conclusions in this study, I am able to conclude that there is potential for improving the forecasting ability. An area of interest would be to redo this study using a different model and perhaps construct a 24-hour whole-day variance.

The results of the regression analysis showed that sampling frequency accounted for around 20-25% of the variability in the error metrics. From the illustration of the data research method in Figure 1. it is also clear that there is an opening for the inclusion of other research fields. In this study, little input from other fields is used. This opens an opportunity for future research to include other variables that influence, for example, the price of oil and hence play an underlying role in the level of volatility. Such fields could include but not limited to, international politics (modelling of OPEC decisions), security situations in specific geographical regions, etc.

This thesis is most likely of little value to professional practitioners (Jacobson, 2020). However, the knowledge of how sampling frequency affects the volatility forecasts of different assets could prove valuable to nonprofessional day traders.

References

Aït-Sahalia, Y., Mykland, P. A., & Zhang, L. (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *The Review of Financial Studies*, *18*(2), 351–416. JSTOR.

Andersen, T., & Bollerslev, T. (1998). Answering the Skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905.

Andersen, T., Bollerslev, T., Diebold, F., & Labys, P. (2001). *Modeling and Forecasting Realized Volatility* (w8160; s. w8160). National Bureau of Economic Research. https://doi.org/10.3386/w8160

Andersen, T., Bollerslev, T., Diebold, F. X., & Labys, P. (2000). Great Realizations. *RISK*, *13*, 105–108.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2010). CHAPTER 2—Parametric and Nonparametric Volatility Measurement. I Y. AÏT-SAHALIA & L. P. HANSEN (Red.), *Handbook of Financial Econometrics: Tools and Techniques* (Bd. 1, s. 67–137). North-Holland. https://doi.org/10.1016/B978-0-444-50897-3.50005-5

Banulescu-Radu, D., Hansen, P. R., Huang, Z., & Matei, M. (2017). *Volatility during the financial crisis through the lens of high-frequency data: A realized GARCH approach*.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2008a). Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. *Econometrica*, *76*(6), 1481–1536. https://doi.org/10.3982/ECTA6495

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2008b). Realised Kernels in Practice: Trades and Quotes. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1138418

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, *2*(1), 1–37. https://doi.org/10.1093/jjfinec/nbh001

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Bollerslev, T., Engle, R., & Nelson, D. (1994). ARCH models. I R. Engle & D. McFadden (Red.), *Handbook of Econometrics,: Bd. IV* (s. 2959–3038). North-Holland.

Bonakdari, H., & Zeynoddin, M. (2022). Chapter 5—Goodness-of-fit & precision criteria. I H. Bonakdari & M. Zeynoddin (Red.), *Stochastic Modeling* (s. 187–264). Elsevier. https://doi.org/10.1016/B978-0-323-91748-3.00003-3

Brownlees, C. T., & Gallo, G. M. (2010). Comparison of Volatility Measures: A Risk Management Perspective. *Journal of Financial Econometrics*, *8*(1), 29–56. http://dx.doi.org/10.1093/jjfinec/nbp009

Buncic, D., & Gisler, K. I. M. (2017). Jumps and leverage play a role in forecasting volatility in international equity markets. *Journal of International Money and Finance*, *79*, 1–19. https://doi.org/10.1016/j.jimonfin.2017.09.001

Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., & Loretan, M. (2010). Observation frequency and estimation of integrated volatility in deep and liquid financial markets. *Journal of Empirical Finance*, *17*(2), 212–240. https://doi.org/10.1016/j.jempfin.2009.09.003

Chan, W. H., Cheng, X., & Fung, J. K. W. (2010). Forecasting volatility: Roles of sampling frequency and forecasting horizon. *Journal of Futures Markets*, *30*(12), 1167–1191. https://doi.org/10.1002/fut.20476

Contino, C., & Gerlach, R. H. (2017). Bayesian tail-risk forecasting using realized GARCH. *Applied Stochastic Models in Business and Industry*, *33*(2), 213–236. https://doi.org/10.1002/asmb.2237

Cookson, C. (2023, mai 11). *Study reveals scale of 'science scam' in academic publishing*. https://www.ft.com/content/76abf920-effb-4d66-8fb2-3ff842150297

Corsi, F. (2008). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001

Danıelsson, J. (2011). *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab.* (Wiley).

Degiannakis, S., & Floros, C. (2015). *Modeling and Forecasting High-Frequency Financial Data*. The Editor(s) (if applicable) and the Author(s).

Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, *17*(5), 425–446. Scopus. https://doi.org/10.1002/jae.683

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), 987–1007. JSTOR. https://doi.org/10.2307/1912773

Engle, R. F. (2001). GARCH 101: An Introduction to the Use of Arch/Garch Models in Applied Econometrics. *Risk Management*. https://api.semanticscholar.org/CorpusID:154189571

Ewald, C., Hadina, J., Haugom, E., Lien, G., Størdal, S., & Yahya, M. (2023). Sample frequency robustness and accuracy in forecasting Value-at-Risk for Brent Crude Oil futures. *Finance Research Letters*, *58*, 103916. https://doi.org/10.1016/j.frl.2023.103916

Francq, C., & Zakoian, J.-M. (2010). *GARCH models: Structure, statistical inference, and financial applications*. John Wiley & Sons Ltd.

Furseth, I., & Everett, E. L. (2022). *Masteroppgaven, Hvordan begynne—Og fullføre* (3. utg.). Universitetsforlaget.

Ghalanos, A. (2023). *Introduction to the Rugarch package. (Version 1.4-3)*.

Ghalanos, A., & Kley, T. (2023). *rugarch: Univariate GARCH Models*.

Gill, J., & King, G. (2004). What to Do When Your Hessian Is Not Invertible: Alternatives to Model Respecification in Nonlinear Estimation. *Sociological Methods & Research*, *33*(1), 54–87. https://doi.org/10.1177/0049124103262681

Gulay, E., & Emec, H. (2018). Comparison of forecasting performances: Do normalization and variance stabilization methods beat GARCH(1,1)-type models? Empirical evidence from the stock markets. *Journal of Forecasting*, *37*(2), 133–150.

Hansen, P. R., & Horel, G. (2009). Quadratic Variation by Markov Chains. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1367519

Hansen, P. R., & Huang, Z. (2016). Exponential GARCH Modeling With Realized Measures of Volatility. *Journal of Business & Economic Statistics*, *34*(2), 269–287.

Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906. https://doi.org/10.1002/jae.1234

Harvey, A. C., & Shephard, N. (1996). Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. *Journal of Business & Economic Statistics*, *14*(4), 429–434. https://doi.org/10.1080/07350015.1996.10524672

Huang, Z., Wang, T., & Hansen, P. R. (2017). Option Pricing with the Realized GARCH Model: An Analytical Approximation Approach. *Journal of Futures Markets*, *37*(4), 328–358. https://doi.org/10.1002/fut.21821

ICE Brent Crude Oil Front Month price information—FT.com. (u.å.). Hentet le 27. mars 2024, fra https ://markets.ft.com/data/commodities/tearsheet/summary ?c=Brent Crude Oil

Jacobson, H. (2020, november 25). *The Realized Volatility Puzzle*. Medium. https://medium.com/swlh/the-realized-volatility-puzzle-588a74ab3896

James, G. (2013). *An Introduction to Statistical Learning: With Applications in R* (First Edition). Springer.

Karasan, A. (2021). *Machine Learning for Financial Risk Management with Python*. O'Reilly.

Long, D. (1995). *Oil trading manual*. Woodhead.

Lunde, A., & Hansen, P. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, *20*(7), 873–889.

Luong, C., & Dokuchaev, N. (2016). Modeling Dependency Of Volatility On Sampling Frequency Via Delay Equations. *Annals of Financial Economics*, *11*(02), 1650007. https://doi.org/10.1142/S201049521650007X

Lyócsa, Š., Molnár, P., & Výrost, T. (2021). Stock market volatility forecasting: Do we need high-frequency data? *International Journal of Forecasting*, *37*. https://doi.org/10.1016/j.ijforecast.2020.12.001

McCrorie, J. R. (2009). Estimating Continuous-Time Models on the Basis of Discrete Data Via an Exact Discrete Analog. *Econometric Theory*, *25*, 1120–1137.

Merton, R. C. (1980). An exploratory investigation on estimating the market's expected return. *Journal of Financial Economics*, *8*(4), 323–361. https://doi.org/10.1016/0304-405X(80)90007-0

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. Wiley-Interscience.

Oomen, R. C. A. (2006). Properties of Realized Variance under Alternative Sampling Schemes. *Journal of Business & Economic Statistics*, *24*(2), 219–237. JSTOR.

Poon, S.-H. & Clive W. J. Granger. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, *41*(2), 478–539. JSTOR.

Walliman, N. (2022). *Research Methods: The Basics* (Third edition). Routledge Taylor & Francis Group.

Wu, X., Xia, M., & Zhang, H. (2020). Forecasting VaR using realized EGARCH model with skewness and kurtosis. *Finance Research Letters*, *32*, 101090. https://doi.org/10.1016/j.frl.2019.01.002

Xie, H., Qi, N., & Wang, S. (2019). A new variant of RealGARCH for volatility modeling. *Finance Research Letters*, *28*, 438–443. https://doi.org/10.1016/j.frl.2018.06.015

Xie, H., & Yu, C. (2020). Realized GARCH models: Simpler is better. *Finance Research Letters*, *33*, 101221. https://doi.org/10.1016/j.frl.2019.06.019

Zhang, Y., Ma, F., Wang, T., & Liu, L. (2019). Out-of-sample volatility prediction: A new mixed-frequency approach. *Journal of Forecasting*, *38*(7), 669–680. https://doi.org/10.1002/for.2590

Zhao, Z. (2008). Parametric and nonparametric models and methods in financial econometrics. *Statistics Surveys*, *2*(none), 1–42. https://doi.org/10.1214/08-SS034

Zivot, E. (2008). *Practical Issues in the Analysis of Univariate GARCH Models* (Working Papers UWEC-2008-03-FC). University of Washington, Department of Economics. https://EconPapers.repec.org/RePEc:udb:wpaper:uwec-2008-03-fc

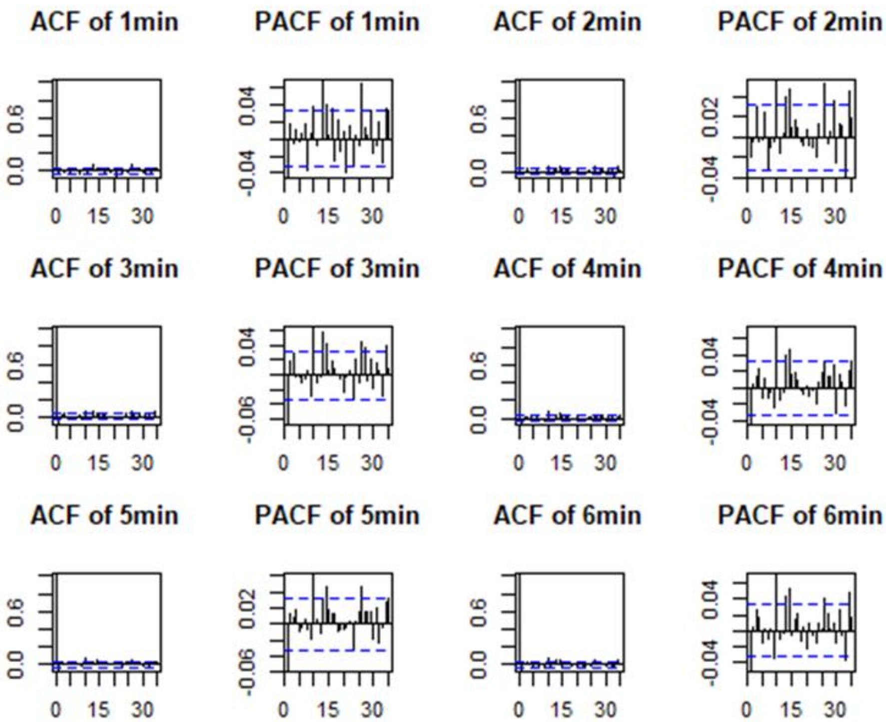
Zou, P. X. W., & Xu, X. (2023). *Research methodology and strategy: theory and practice*. Wiley Blackwell.

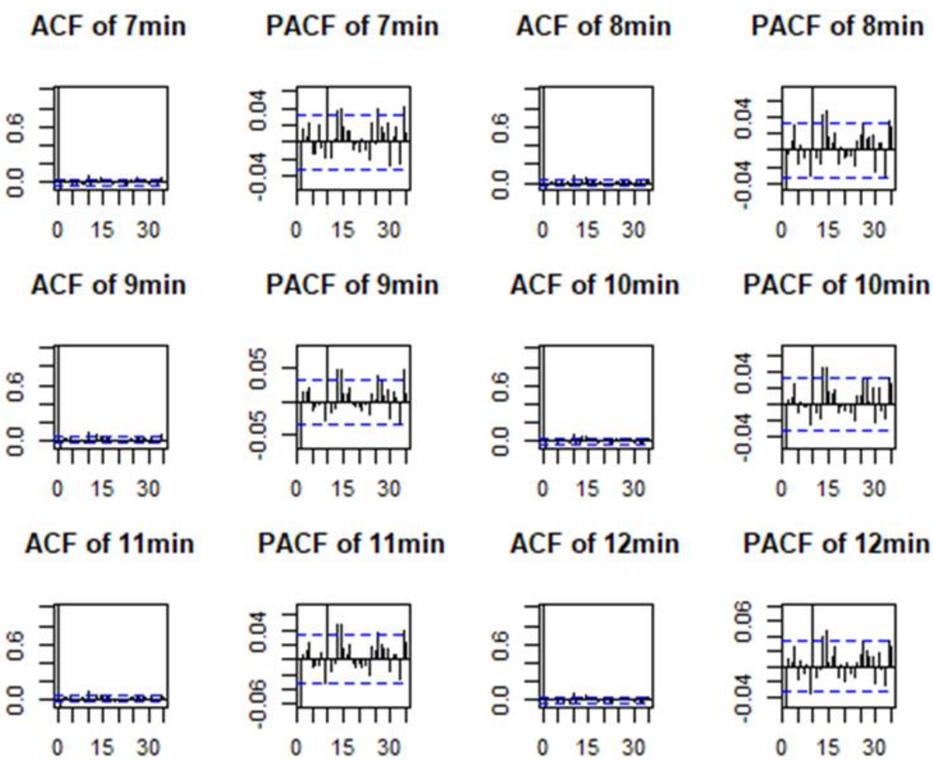
Appendix

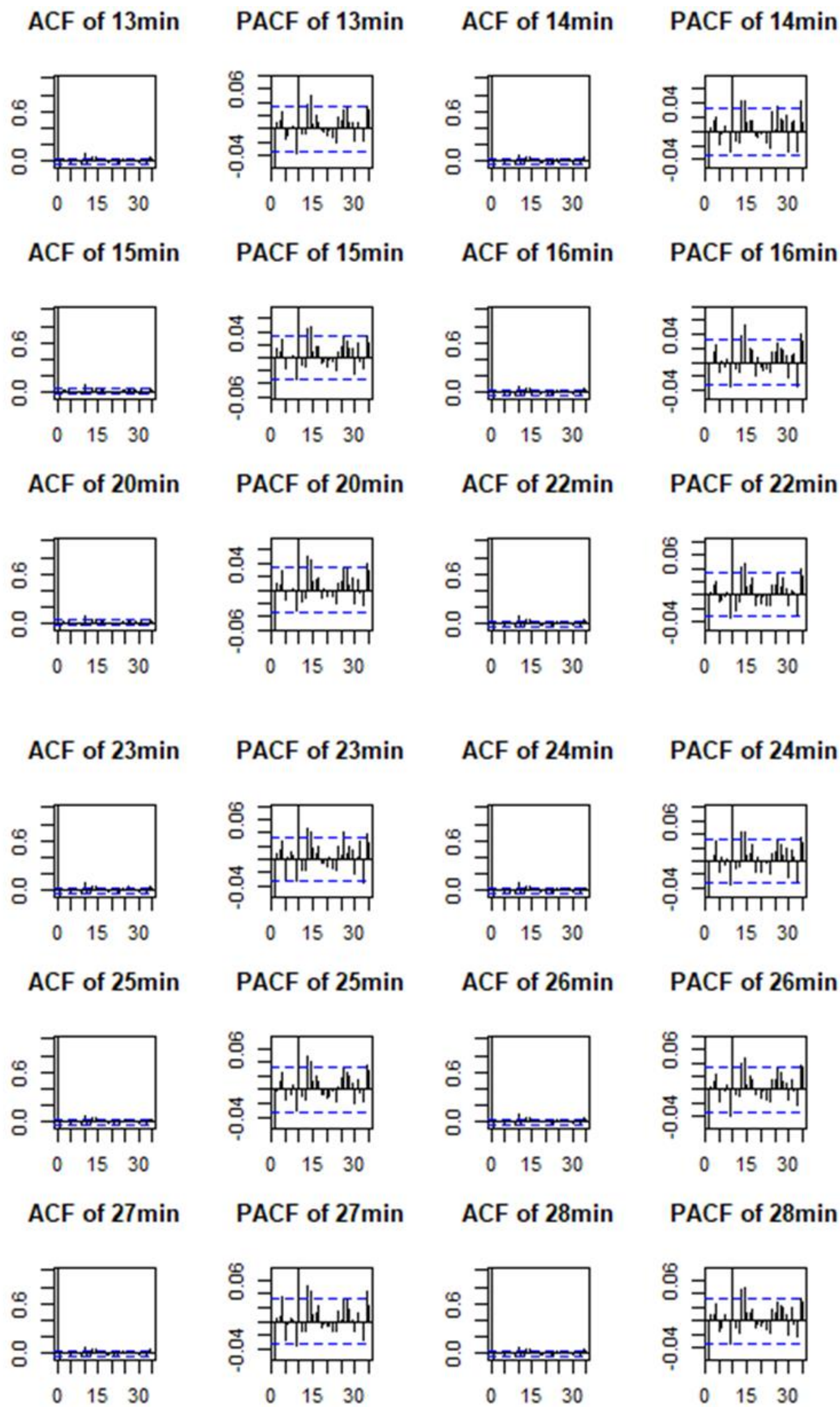
1.Link to the dataset

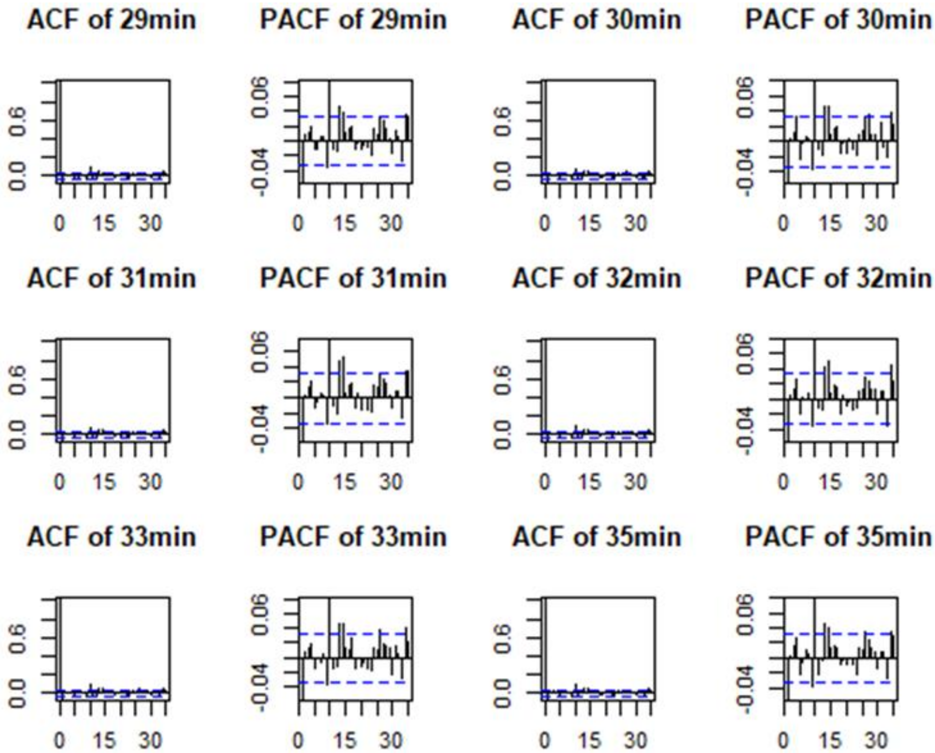
https://www.dropbox.com/scl/fi/l6vb3tmqkk00z36kfg61s/pt_1min_long.csv?rlkey=mj86i8varyfggji5kg7qeadyo&dl=0

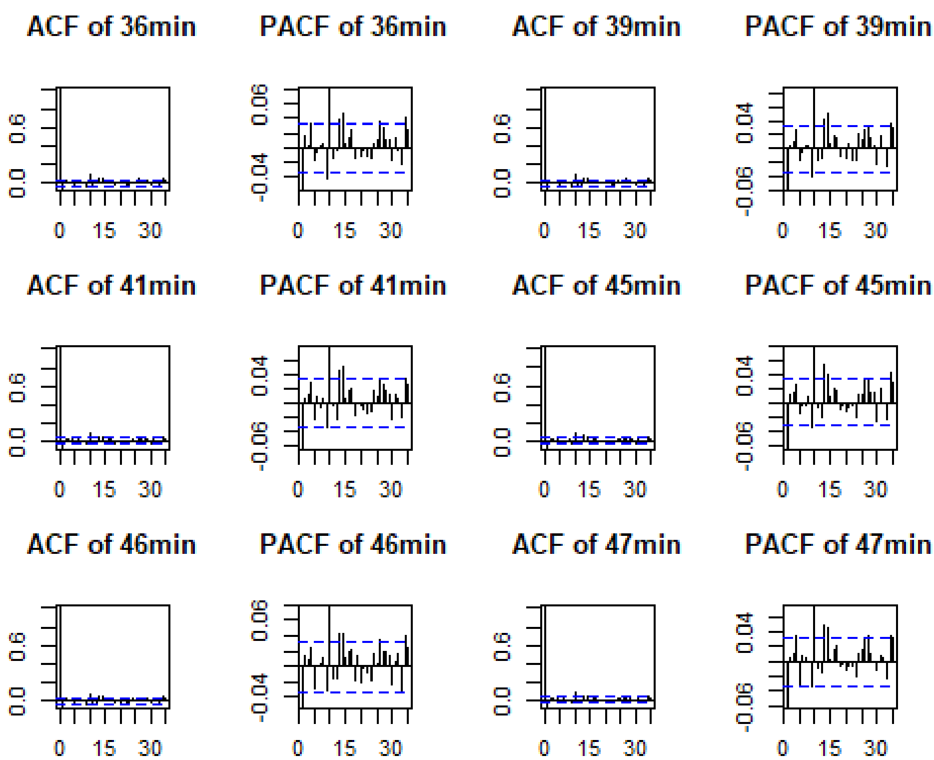
2. ACF and PACF plots

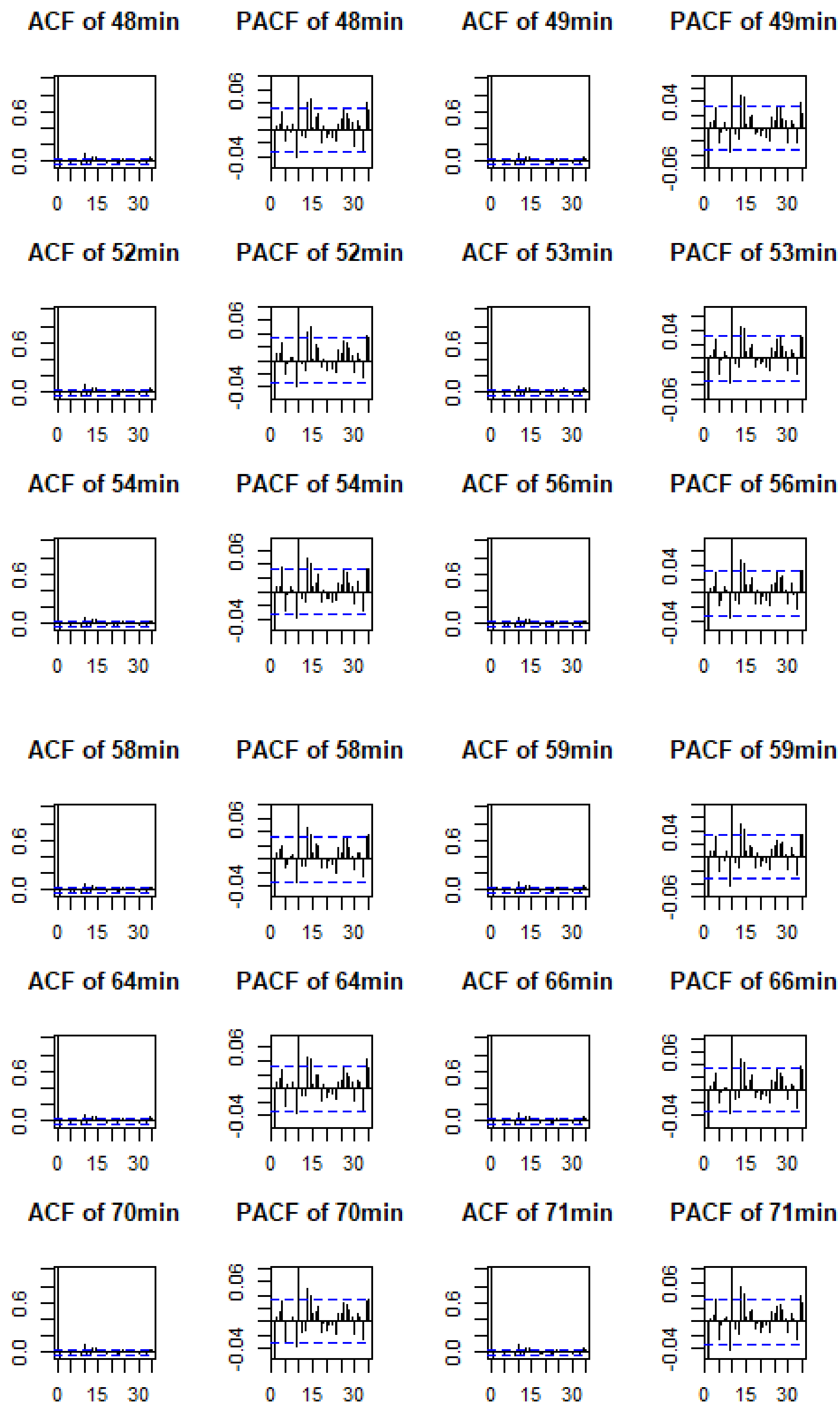


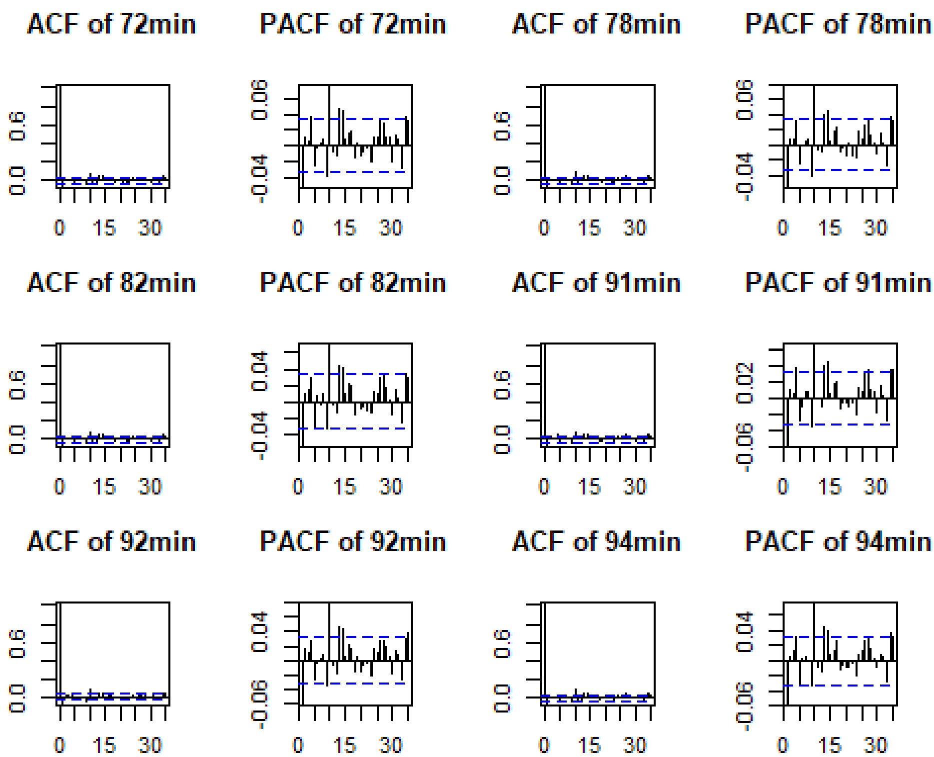



ACF of 13min
PACF of 13min
ACF of 14min
PACF of 14min
ACF of 15min
PACF of 15min
ACF of 16min
PACF of 16min
ACF of 20min
PACF of 20min
ACF of 22min
PACF of 22min
ACF of 23min
PACF of 23min
ACF of 24min
PACF of 24min
ACF of 25min
PACF of 25min
ACF of 26min
PACF of 26min
ACF of 27min
PACF of 27min
ACF of 28min
PACF of 28min


ACF of 29min
PACF of 29min
ACF of 30min
PACF of 30min
ACF of 31min
PACF of 31min
ACF of 32min
PACF of 32min
ACF of 33min
PACF of 33min
ACF of 35min
PACF of 35min


ACF of 36min
PACF of 36min
ACF of 39min
PACF of 39min
ACF of 41min
PACF of 41min
ACF of 45min
PACF of 45min
ACF of 46min
PACF of 46min
ACF of 47min
PACF of 47min


ACF of 48min
PACF of 48min
ACF of 49min
PACF of 49min
ACF of 52min
PACF of 52min
ACF of 53min
PACF of 53min
ACF of 54min
PACF of 54min
ACF of 56min
PACF of 56min
ACF of 58min
PACF of 58min
ACF of 59min
PACF of 59min
ACF of 64min
PACF of 64min
ACF of 66min
PACF of 66min
ACF of 70min
PACF of 70min
ACF of 71min
PACF of 71min



3. MSE &MAE

Sampling frequency	MSE	MAE
1	0,000837114	0,0173654
2	0,000850077	0,01713219
3	0,000831878	0,01696378
4	0,000735021	0,01680133
5	0,000837049	0,01699371
6	0,000837296	0,01687282
7	0,000748954	0,01670987
8	0,000785903	0,01693543
9	0,000879811	0,01693912
10	0,000839196	0,01705439
11	0,000735257	0,01674398
12	0,000756774	0,01685283
13	0,000814861	0,01723923
14	0,000748844	0,01681634
15	0,000840273	0,01708052

16	0,000786699	0,01689655
20	0,000742448	0,01686042
22	0,000731823	0,01674151
23	0,000794962	0,01707588
24	0,000820319	0,01706748
25	0,000895523	0,01716658
26	0,000801111	0,01707314
27	0,000895559	0,01705687
28	0,000732965	0,01678631
29	0,000796471	0,01692102
30	0,000844093	0,01702324
31	0,000791837	0,01691658
32	0,000786125	0,0170246
33	0,00071273	0,01673273
35	0,000713567	0,01669651
36	0,000787504	0,01683633
39	0,000792323	0,01705795
41	0,000780797	0,01690891
45	0,000906077	0,01719232
46	0,000769247	0,0169314
47	0,000793908	0,01691666
48	0,00080751	0,01693422
49	0,000765753	0,01681905
52	0,000784727	0,01684822
53	0,000789748	0,0168946
54	0,000898471	0,01714267
56	0,000783936	0,01687245
58	0,000788463	0,01697469
59	0,000782992	0,01690619
64	0,000782412	0,01689451
66	0,000707088	0,01661104
70	0,000711984	0,01662304
71	0,000778221	0,01682899
72	0,000785511	0,01678617
78	0,000775274	0,01688978
82	0,000781778	0,01684781
91	0,000782564	0,01691845
92	0,000768194	0,01677449
94	0,0007573	0,01671145

4.Percentage changes in MSE and MAE

Sampling_frequency	MSE change	MSE_cumulative	MAE change	MAE cumulative
1	0	0	0	0
2	1,55	1,55	-1,34	-1,34
3	-2,14	-0,59	-0,98	-2,33
4	-11,64	-12,24	-0,96	-3,28
5	13,88	1,65	1,15	-2,14
6	0,03	1,67	-0,71	-2,85
7	-10,55	-8,88	-0,97	-3,82
8	4,93	-3,94	1,35	-2,47
9	11,95	8,01	0,02	-2,44
10	-4,62	3,39	0,68	-1,76
11	-12,39	-9	-1,82	-3,58
12	2,93	-6,07	0,65	-2,93
13	7,68	1,61	2,29	-0,64
14	-8,1	-6,49	-2,45	-3,09
15	12,21	5,71	1,57	-1,52
16	-6,38	-0,66	-1,08	-2,6
20	-5,62	-6,29	-0,21	-2,81
22	-1,43	-7,72	-0,71	-3,52
23	8,63	0,91	2	-1,52
24	3,19	4,1	-0,05	-1,57
25	9,17	13,27	0,58	-0,99
26	-10,54	2,73	-0,54	-1,53
27	11,79	14,51	-0,1	-1,63
28	-18,16	-3,64	-1,59	-3,22
29	8,66	5,02	0,8	-2,41
30	5,98	11	0,6	-1,81
31	-6,19	4,81	-0,63	-2,44
32	-0,72	4,09	0,64	-1,8
33	-9,34	-5,25	-1,71	-3,51
35	0,12	-5,13	-0,22	-3,73
36	10,36	5,23	0,84	-2,89
39	0,61	5,85	1,32	-1,57
41	-1,45	4,39	-0,87	-2,45
45	16,05	20,44	1,68	-0,77
46	-15,1	5,33	-1,52	-2,29
47	3,21	8,54	-0,09	-2,38
48	1,71	10,25	0,1	-2,27
49	-5,17	5,08	-0,68	-2,95

52	2,48	7,56	0,17	-2,78
53	0,64	8,2	0,28	-2,5
54	13,77	21,97	1,47	-1,04
56	-12,75	9,22	-1,58	-2,61
58	0,58	9,8	0,61	-2,01
59	-0,69	9,1	-0,4	-2,41
64	-0,07	9,03	-0,07	-2,48
66	-9,63	-0,6	-1,68	-4,16
70	0,69	0,09	0,07	-4,08
71	9,3	9,4	1,24	-2,85
72	0,94	10,33	-0,25	-3,1
78	-1,3	9,03	0,62	-2,48
82	0,84	9,87	-0,25	-2,73
91	0,1	9,97	0,42	-2,31
92	-1,84	8,13	-0,85	-3,16
94	-1,42	6,72	-0,38	-3,54