Modelling and forecasting oil market volatility: A regime switching GARCH MIDAS approach

Menli Tirkishova

School of Computing and Mathematics

Keele University

A thesis submitted for the degree of

Doctor of Philosophy

October 2023

Abstract

In this thesis, we consider a regime switching GARCH MIDAS model with Student-t innovations. By allowing the error term to be non-Gaussian, we want to see how effective it is in describing the volatility displayed in financial time series. For the long-term volatility component, as natural explanatory variables we also consider realised volatility (RV) calculated as absolute returns. This extension is particularly important since regime switching GARCH MIDAS models have been partially implemented by very few authors where they assumed innovations are normally distributed and calculated RV as squared returns.

In addition, by Monte Carlo simulation and a real data application we provide evidence to support our proposed model setting RS GARCH MIDAS-t with RV. We consider the misspecification in terms of; (a) not considering regime switching, (b) misspecifying the error term, (c) omitting the long-term volatility component, (d) all three combined. The simulation results confirms the importance of correctly specifying volatility models. We show that when models are misspecified, the bias of parameter estimates increase while the ability of regime switching process to correctly identify regimes deteriorate quite considerably. Moreover, the long-term volatility in misspecified models leads to an overestimation.

The validity of our proposed model is then explored through a real data application. Empirical analysis of West Texas Intermediate crude oil returns show that regime switching models outperform single-regime models. Specifications with Student-t innovations are superior to their Gaussian counterparts in terms of higher log-likelihood value, lower model selection criteria, and ability to better identify the volatility regimes. This provides strong within-sample estimation evidence in favour of our non-Gaussian assumption. We also find that production has a significant positive effect on crude oil volatility while demand has insignificant negative effect. For out-of-sample forecasting evaluation, while considering models with long-term volatility component, we show that under loss functions models with t innovations are favoured over those with a normal innovation, while RS GARCH MIDAS-t with RV outperforms other models.

Declaration

I declare that the work contained within this thesis, submitted for the degree of Doctor of Philosophy at Keele University, has been solely composed by me, Menli Tirkishova, and has not been submitted, in part or as a whole, for any other degree. Except where otherwise stated by reference or acknowledgement the entirety of the work presented is entirely my own.

Acknowledgments

First and foremost I would like to express my great gratitude to my supervisor Dr. Jie Cheng for her invaluable advice, stimulating encouragement, continuous support and guidance throughout my PhD. I would also like to thank all the members in the School of Computing and Mathematics for welcoming me and making my study and life in UK a wonderful time.

Some special words of gratitude go to my parents and my little sister for their unconditional love, support and encouragement. They have always been a major source of support when things would get a bit discouraging. Without them, this day would not have been possible.

And last but not least, I would like to thank my friends who supported me during hard times, cheered me on, and celebrated each accomplishment. Thank you guys for always being there for me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Accurately modelling and forecasting volatility is of significant importance for investors, traders, policy makers and anyone else involved in the financial markets. Since the introduction of Markov switching model by Hamilton (1989, 1990), regime switching models have been extensively applied in research. In particular, RS GARCH models have been widely used in financial time series forecasting because of their ability to identify and account for the impact of different changes in volatility and capture many conditional volatility characteristics (Hamilton and Susmel, 1994; Dueker, 1997; Klaassen, 2002; Fong and See, 2002; Gunay, 2015; Herrera et al., 2018; Haas and Liu, 2018; Zhang et al., 2019).

Mixed data sampling (MIDAS) regression scheme introduced by Ghysels et al. (2007) have attracted attention in the literature because of their ability to capture complicated volatility dynamics and inclusion of data from different frequencies in the same mode. This makes it possible to combine the high-frequency return data with macroeconomic data that are only observed at lower frequencies. Using a GARCH MIDAS model within MIDAS framework proposed by Engle et al. (2013) we can describe volatility in two components where one component is linked to the short-term volatility, similar to GARCH setting, and the other component describes long-term volatility driven by realised volatility or macroeconomic variables. Furthermore, to capture the potential regime changes we introduce regime switching framework into GARCH MIDAS model, since according to Lamoureux and Lastrapes (1990) and Cai (1994), the unconditional volatility may change across different regimes due to structural breaks in the volatility.

The regime switching model of Hamilton (1989), was derived with the assumption that the error term follows a normal distribution. However, significant evidence suggests that most asset return distributions exhibit a fat-tailed behaviour (Bollerslev, 1987), meaning they have more of the distribution in the tails than would a normal distribution with the same mean and variance. Hence, the idea of applying an appropriate distribution to accommodate excess kurtosis became essential. A commonly used alternatives of Gaussian innovations are fat-tailed distributions such as Student-t, General Error Distribution (GED), and their skewed versions. Furthermore, in recent years, more and more researchers introduce regime switching to more general classes of non-Gaussian models. For some examples see Gunay (2015), Haas and Liu (2018) and Herrera et al. (2018).

In this thesis we consider a regime switching GARCH MIDAS model with non-Gaussian innovation setting. This is particularly important since RS GARCH MIDAS models have been partially implemented by few authors where they all assume innovations are normally distributed. Furthermore, we try to see how this model performs under different innovations, especially under fat tail distribution.

While consistency and asymptotic theory for the quasi maximum likelihood estimator (QMLE) for RS GARCH (Xie, 2009; Bauwens et al., 2010) and for a special case of GARCH MIDAS (Wang and Ghysels, 2015) with realised volatility as the explanatory variable have been established, to the best of our knowledge, it is not yet available for RS GARCH MIDAS models especially if the explanatory variables are considered. Therefore, we first evaluate the finite-sample performance of QMLE in a Monte Carlo simulation to show that QMLE is unbiased and the asymptotic standard errors are valid. Furthermore we demonstrate the importance and consequences of potential misspecification. We consider the misspecification in terms of not considering regime switching, misspecifying the error term, omitting the long-term volatility component, or all three combined. We find that GARCH models have shortcomings when there are regime shifts in the volatility process causing upward biased estimate in the degree of volatility persistence. In addition, we find that when the models are misspecified, the bias of parameter estimates increase while the ability of regime switching process to correctly identify regimes deteriorate quite considerably. Consistent with Conrad and Kleen (2020) we also find that changing the innovation in single component GARCH and GARCH MIDAS models hardly affects the parameter estimates. However, omission of regime switching, causes an overestimation of the long-term component in GARCH MIDAS models. These simulation results confirms the importance of correctly specifying volatility models.

Finally, the validity of our proposed model, RS GARCH MIDAS with Student-t innovations is assessed through a real data application. We apply this model to West Texas Intermediate crude oil returns for the period between 1986 and 2020 (34 years in total). We compare 6 models, GARCH, RS GARCH, GARCH MIDAS, RS GARCH MIDAS, Endogenous RS GARCH and Endogenous RS GARCH MIDAS. Furthermore, we consider the MIDAS component with realised volatility, where it is calculates as absolute returns, and two macroeconomic variables, demand and production. For regime-switching models, we only allow one parameter to switch between a low and high volatility regime, following Marcucci (2005) who showed that the differences between the parameters α and β are likely to be insignificant. The within-sample estimation demonstrates that models with t innovation easily beats their Gaussian counterparts and RS GARCH MIDAS-t with RV outperforms all other models in terms of the highest log-likelihood value and lowest model selection criteria. Moreover, we compare our results with the findings of Pan et al. (2017). We find that production has a significantly positive impact on long-term crude oil volatility while demand has an insignificantly negative effect, meaning that demand has a negligible influence on the monthly component of crude oil daily volatility. These results contradict to findings of Pan et al. (2017) but align with the conclusions drawn by Wei et al. (2017) who also found that the demand factor is insignificant, meanwhile, the supply factor is positive and statistically significant in oil volatility modelling.

For out-of-sample forecast evaluation we estimate all models on a rolling window approach and compare the RS GARCH MIDAS-t where the long-term component is driven by RV with competitor models. Furthermore, following Patton (2006) we evaluate all models using MSE and QLIKE loss functions and compare the forecasts by implementing equal predictive ability test introduced in Diebold and Mariano (1995). When considering models with long-term volatility component, we find that neither GARCH MIDAS with RV nor RS GARCH MIDAS models with macroeconomic variables can beat RS GARCH MIDAS-t with RV only. This findings coincides with Conrad and Loch (2015) who showed that almost all specifications based on macroeconomic variables performed worse than their benchmark model, GARCH MIDAS with RV. On the contrary, this contradicts to Pan et al. (2017). A possible explanation for this is that we use absolute returns as a realised volatility measure, while they considered squared returns. Furthermore, we follow an approach developed by Klaassen (2002) to overcome path dependence problem whereas Pan et al. (2017) implement Gray (1996) approach.

To summarise, this thesis contributes to the current literature in several ways. We begin by providing evidence to support the chosen model setting, RS GARCH MIDAS-t, through Monte Carlo simulation. We conduct a systematic investigation into how different components capture volatility, particularly focusing on regime-switching, short-term, and long-term volatility components and how each component affects each other. By dissecting the volatility dynamics, we contribute to a deeper understanding of the interplay between various components and their impact on the overall volatility structure. The simulation results demonstrate the importance of correctly specifying volatility models. By conducting the simulation, we establish a solid foundation for the subsequent empirical analysis.

Next, we conduct an empirical analysis specifically focused on West Texas Intermediate (WTI) crude oil returns. This analysis emphasizes the significance of correctly specifying the error distribution within the volatility models. The results obtained from the empirical analysis indicate that models incorporating Student-t innovations outperform those with Gaussian counterparts in terms of higher log-likelihood value, lower model selection criteria and regime identification. The filter probabilities show that regime switching models with t innovations are able to capture the major events in the market. For instance, the filter probabilities obtained from RS GARCH MIDAS-

t show that the crude oil market switches to high-volatility regime during the 1990 Gulf War, 1996 backwardation in the oil market, during the Asian financial crisis in 1997, after the terrorist attack in 2001, during the U.S. invasion of Iraq in 2003, the financial crisis of 2007-2008 and so on. Thus, this thesis presents strong within-sample estimation evidence in favour of the non-Gaussian assumption.

Lastly, taking into consideration the long-term volatility component we evaluate the performance of different models for out-of-sample forecasting. Two loss functions, namely MSE and QLIKE, are utilized for the evaluation. The findings reveal that the GARCH MIDAS and RS GARCH MIDAS models with Student-t innovations outperform those with normal innovation, again supporting the superiority of non-Gaussian assumptions. Additionally, our proposed model, RS GARCH MIDAS-t demonstrates better forecasting performance compared to other models that incorporate low-frequency component.

Overall, this thesis makes significant contributions to the literature by providing empirical evidence for the importance of correctly specifying volatility models, highlighting the superiority of non-Gaussian assumptions, and examining the dynamics of different volatility components. These contributions advance the existing literature and contribute to a more comprehensive understanding of volatility modelling in the context of WTI crude oil volatility. To the best of our knowledge this is the first attempt to estimate and predict the volatility of crude oil prices using RS GARCH MIDAS approach with Student-t innovations and RV as absolute returns.

The rest of this thesis is organised as follows: Chapter 2 is the literature review where the characteristics of volatility are given and the importance of these characteristics are discussed. The main focus of this chapter is on the exploration of different volatility estimation and forecasting models in the literature that are applied to volatility with different innovations. Starting with GARCH model more advanced GARCH type models are introduced. Moreover, GARCH type models with different distribution of the innovations are discussed. In Chapter 3 various model specifications are described along with their estimation procedures, properties and out-of-sample forecasting steps. Chapter 4 presents our simulation study, where we first evaluate a finite-sample performance of QMLE in a Monte Carlo simulation and demonstrate the importance and consequences of potential misspecification. In Chapter 5 the validity of our proposed model is then explored through a real data application. The within and out-of-sample performances are compared between Normal and Student-t innovations. Model selection criterion and likelihood ratio tests are computed while for out-of-sample evaluation we calculate loss function and implement equal predictive ability test. Concluding remarks as well as ideas for future research are provided in Chapter 6.

The definitions of terms are given in Appendix A and supplementary materials for Chapter 4 and Chapter 5 are given in Appendix B. To facilitate the replication of our results we provide R codes for data simulation, estimation and forecasting in Appendix C.

Chapter 2

Literature Review

2.1 Characteristics of volatility

Over the past few decades modelling and forecasting crude oil price volatility has become an important subject in research area (Ma et al., 2019) due to the high levels of price volatility observed in oil market prices (Ural, 2016). In addition, the accurate estimation and forecasting of market volatility are essential for various financial applications, including risk management, portfolio optimization, option pricing, and trading strategies. The ability to capture and anticipate changes in volatility patterns can provide valuable insights for investors and financial institutions, allowing them to make informed decisions in a rapidly evolving market environment. However, the estimation and forecasting of market volatility pose significant challenges due to a number of volatility characteristics present in many time series returns.

For instance, a well-known phenomenon called volatility clustering, observed by Mandelbrot (1963), states that large changes in the asset prices tend to be succeeded by other large changes, likewise small changes were often followed by other small changes. The presence of volatility clustering has also been documented by Chou (1988) and Schwert (1988) among others. Furthermore, the existence of volatility clustering suggests extreme volatility persistence, therefore, it is important to detect and analyse the causes of such clustering. Another stylized fact of volatility is that volatility may be stationary, since it evolves over time by varying within some fixed range and does not diverge to infinity (Tsay, 2005).

The leverage effect, highlighted in Poon and Granger (2003), is another significant factor impacting volatility. It describes the phenomenon where volatility tends to rise more prominently following a significant price decline, as opposed to an equal increase. As a result, it is highly unlikely that positive and negative shocks will have the same impact on volatility. Moreover, there exists a negative correlation between returns and conditional volatility, known as asymmetric volatility (Engle and Ng, 1993; Zakoian, 1994). These characteristics play a crucial role in the development of volatility models. While various models, such as diffusion and stochastic volatility models, have been developed to estimate volatility, well known and frequently applied models are the heteroscedastic models.

2.2 GARCH models

The first model that provides a systematic framework for volatility modelling is the Autoregressive Conditional Heteroscedasticity (ARCH) model by Engle (1982). This model was developed to improve econometric models by replacing the assumption that the variance of the error term is constant, also known as homoscedasticity. In addition, ARCH model is able to capture the persistence and volatility clustering, however it only considers past returns to adequately describe the volatility process of an asset return. Furthermore, it was found that a rather long lag structure for the conditional variance is required to capture the long memory present in the data.

By extending the idea of ARCH, Bollerslev (1986) proposed a useful extension known as Generalised ARCH (GARCH) model. The GARCH model adds an autoregressive component to the ARCH process by letting the conditional variance be dependent on both past innovations and its lags. GARCH type models are especially useful when the aim is to analyse and forecast volatility. The advantages of GARCH model is that it can capture some behaviours in financial time series such as fat tails, excess kurtosis and volatility clustering. For instance, Bollerslev (1986) showed that GARCH model is an extremely useful tool for describing the influence of volatility clustering and its relationship to fat-tails in return distributions. On the other hand, a limitation of GARCH model is that it allows the variance to be affected only by the square of the lagged innovation, thus completely disregarding the sign of that innovation, hence failing to capture the leverage effect which is a phenomena where the tendency for volatility to increase more after a large fall in prices rather than after an increase on the same amount.

Over the course of time, ARCH and GARCH models have been extended numerously to consider leverage effects, asymmetric volatility and nonlinearity. For instance, Nelson (1991) argued that the non-negativity constraints in the GARCH model are too restrictive and hence developed an exponential GARCH model (EGARCH) where it captures asymmetry in volatility induced by big positive and negative returns. He showed that this model significantly outperforms their counterparts that do not accommodate the asymmetry. Furthermore, EGARCH model uses a conditional variance equation in logarithmic form which allows the parameter restriction in GARCH models to relax. Glosten et al. (1993) proposed GJR-GARCH to take into account the negative and positive shocks caused by leverage phenomenon. Another volatility model commonly used to handle leverage effects is the threshold GARCH (TGARCH) model proposed by Zakoian (1994). The GJR-GARCH and TGARCH models are very similar to each other where they both use an indicator function to allow more reaction to negative shocks. Other variations include Integrated GARCH (IGARCH) which restricts the persistence parameter to be equal to one. Mikosch and Stărică (2004) showed that applying GARCH model to a sample displaying structural changes in the unconditional volatility creates an IGARCH effect. Moreover, in ARCH and GARCH models the conditional variance is a linear function of past squared residuals and lagged conditional variances. Asymmetric GARCH models can therefore be regarded as a non-linear GARCH specifications.

2.3 Regime switching GARCH models

The accumulated evidence from empirical research suggests that the volatility of financial markets display some type of persistence that cannot be appropriately captured by GARCH model and its variations (Lamoureux and Lastrapes, 1990; Engle and Mustafa, 1992). In particular, these models usually indicate high persistence in the conditional volatility (Hamilton and Susmel, 1994; Gray, 1996). It has also been shown that observed high-volatility persistence can be due to neglected non-linearities such as level shifts (Lamoureux and Lastrapes, 1990) and neglecting such potential non-linearities can lead to poor forecasts (Hamilton and Susmel, 1994). Furthermore, Klaassen (2002) also showed that GARCH forecasts are too high in volatile periods, and the reason for such excessive forecasts may be high persistence.

Lamoureux and Lastrapes (1990) and Kim and Kon (1999) among others found that the introduction of different regimes into the GARCH process reduces the persistence parameters in the GARCH process. Furthermore, Gray (1996) has argued that GARCH models may be misspecified since they are not flexible enough, thus ignoring possible structural changes in volatility processes. This prompted the shift toward regime switching models.

Markov switching model also known as regime switching (RS) model governed by Markov-chain was first introduced in 1989 by Hamilton (1989) which was further analysed by Kim (1994) and since then continued to gain popularity especially in financial time series. The Markov switching model is a regime switching model that incorporates the assumption that unobserved states are determined by a Markov chain. In the simplest case, known as a two-regime switching model, the economy is characterized by two distinct states, "calm and turbulent" (Ho et al., 2004). The two states differ in that the turbulent state has more volatile and higher index value than the calm state. The economy switches from one state to the other according to a constant or time-varying transition probability. We will discuss the time-varying transition probability in the Section 2.11.

RS models have been extended and applied to a broad range of fields in the literature. For instance, Schwert (1988) considered a model in which returns may have high or low variance and the switches between these two states are determined by a two-state Markov process. Hamilton and Susmel (1994) introduced an ARCH model with regime switching parameters in order to account for sudden changes in the conditional variance. They found that implementing a regime switching model improves fit and forecasting accuracy by distinguishing between high- and low-volatility regimes where high-volatility regimes are loosely associated with recessions. Cai (1994) applied similar idea by incorporating both the regime switching and the ARCH model. He noted that the volatility of treasure bills becomes much less persistent when allowing for regime switching in some parameters.

Later, a class of regime switching GARCH models named RS GARCH model was presented by Gray (1996) where he concluded that the RS GARCH model outperforms the no regime models in forecasting performance and reduces the persistence in volatility. However, the limitation of Gray's variant is that it is unable to compute the multi-period-ahead volatility forecasts needed for a detailed forecasting analysis. Therefore, Klaassen (2002) proposed a modified version of Gray's model where the multi-period-ahead forecasts can be obtained through a convenient first-order recursive procedure. Another approach namely Markov switching GARCH by using ARCH (∞) specification was developed by Haas et al. (2004). They found significantly improved forecasts in exchange rate volatility as the regimes are able to account for the changes in unconditional variance. However, they also found that in some cases the transition probabilities are not significant meaning that necessarily not all volatility series have a regime switching structure.

2.4 Volatility component models

Volatility changes can be caused by many factors such as inflation, deflation, extreme events like war, financial crisis, changes in government policy or market disruptions. These factors can also lead to structural breaks in the data. For instance, regular modification in government policies increase internal and external uncertainties which lead to increase in stock market volatility. In addition, Hamilton and Susmel (1994) argued that volatility seems to change or switch between different regimes due to many existing factors such as extreme events and business cycles.

Volatility component models have attracted attention in the literature because of their ability to capture complicated volatility dynamics and also handle structural breaks in asset return volatility (Wang and Ghysels, 2015; Andreou and Ghysels, 2002). Engle et al. (1999) developed a volatility component model where the conditional variance is specified as the sum of two components; long-run (trend) component and short-run (transitory) component and applied this to S&P 500 and NIKKEI 225. They showed that this new specification beats traditional GARCH model. By introducing a regression scheme, namely mixed data sampling (MIDAS) which allows inclusion of data from different frequencies into the model, Ghysels et al. (2007) made it possible to combine high frequency data with macroeconomic data that are only observed at a lower frequency.

Later in 2008, Engle and Rangel (2008) proposed a new model namely multiplicative two component Spline-GARCH which consists of gradually changing deterministic and a short-term GARCH component. By extending Spline-GARCH model, Engle et al. (2013) developed GARCH MIDAS model where GARCH term uses standard GARCH(1,1) to model short-term (high-frequency) volatility while MIDAS term is used to describe the long-term (low-frequency) process. The difference between Spline-GARCH and GARCH MIDAS is the specification of long-term component. Spline-GARCH describes low-frequency volatility in a nonparametric way, meaning that long-run variance is time varying. This specification allows the model to be more flexible, however it loses its mean reverting property. Whereas, in GARCH MIDAS the long-term component facilitates a direct inclusion of low-frequency macroeconomic data and this is an advantage of this model since it allows to directly examine the macroeconomic variable's impact on the volatility. Using commodity futures Nguyen and Walther (2020) conducted an empirical study where they fitted both Spline-GARCH and GARCH MIDAS and found that disentangling high- and low-volatilities produced better in-sample fit for both models.

The GARCH MIDAS model has become a popular model to investigate the relationships between aggregate volatility and macroeconomic variables (Conrad et al., 2014; Conrad and Loch, 2015; Wei et al., 2017; Pan et al., 2017; Conrad et al., 2018; Conrad and Kleen, 2020). Conrad and Loch (2015) applied GARCH MIDAS to a wide set of macroeconomic variables related to U.S., whereas Pan et al. (2017) applied RS GARCH MIDAS model to oil price volatility in terms of supply and demand characteristics. Using GARCH-MIDAS model, Wei et al. (2017) investigated which determinant helps to make the most accurate daily volatility forecasts across three different forecasting horizons.

Based on a simulation study of Conrad and Kleen (2020), if correctly specified, the GARCH MIDAS model beats two competitor models which are RS GARCH and nested GARCH. This argument is based on an out-of-sample forecast evaluation employing QLIKE loss function. Majority of the above mentioned papers find a significant effect of the macroeconomic variables on the volatility series under consideration. Moreover, these papers find that the volatility forecast at longer horizons are improved substantially over the basic GARCH model. Also, current literature primarily finds that including realised volatility as macroeconomic variable or any other single macroeconomic variable provides useful information in forecasting volatility (Takahashi et al., 2021).

2.5 Regime switching GARCH MIDAS models

Despite the inclusion of a long-term macroeconomic component, Engle et al. (2013) found that the full sample models are not immune to structural breaks when using production and inflation as macroeconomic factors. Therefore, they had to split their sample into sub-samples to improve fit. The fact that the structural breaks are not accounted for is a significant limitation of GARCH MIDAS model. This overcome this shortcoming Pan et al. (2017) developed a RS GARCH MIDAS model where he set the short-term volatility component to change between two states. Their in-sample estimation results showed that the level of macroeconomic variables has a significantly adverse effect on the oil volatility, whereas the out-of-sample results showed that two-regime GARCH MIDAS model can significantly beat the single-regime GARCH MIDAS model and additional macroeconomic variables can significantly increase the predictive performance of RS GARCH MIDAS model. While Pan et al. (2017) kept the long-term component non-switching, Ma et al. (2021) modified long-term volatility process by introducing regime switching structure. Furthermore, they set the transition probabilities to be time-varying and are driven by global economic policy uncertainty (GEPU). Their out-of-sample findings indicate that RS GARCH MIDAS model with time-varying transition probabilities and GEPU outperforms other models. The latest research comes from Wang et al. (2022) where they extended Pan et al. (2017) and Ma et al. (2021) models to incorporate Markov-regime switching into both short- and long-term components, separately or in combination. They found that for in-sample estimation Full RS GARCH MIDAS model, where both short- and long-term components switch, can best capture the volatility of renewable energy stock market. As for out-of-sample forecasting their model generated more accurate volatility forecasts in short-term while for long-term forecasting Ma et al. (2021) model showed the best prediction. Thus, these studies suggest that it is necessary to adopt an appropriate model that can include short-term, long-term or both terms in regime switching to improve the model's forecasting accuracy (Pan et al., 2017; Ma et al., 2021; Wang et al., 2022).

2.6 The stochastic volatility models

In addition to the GARCH class models, the literature offers various alternative volatility estimation models that can be applied to modelling volatility. One such model is the stochastic volatility (SV) models, where the evolution of volatility in a time series is described by introducing a stochastic innovation to the conditional variance equation. This additional stochastic innovation could then be used to explain the unexpected shocks to the volatility process. Notable examples include Melino and Turnbull (1990), Taylor (1994) and Vo (2009).

In recent years, SV models have been extended to allow for long-memory in volatility. This extension is driven by the observation that the autocorrelation function of the squared or absolute value series of asset returns often exhibits slow decay, even when the return series has no serial correlation (Ding et al., 1993). For example, Abbara and Zevallos (2023) proposed a new method to estimate a univariate long-memory stochastic volatility (LMSV) model by formulating the LMSV model in a state-space representation with non-Gaussian perturbations in the observation equation. Furthermore, Kalimipalli and Susmel (2004) introduces the regime switching into stochastic volatility framework (RSV) to explain the behaviour of short-term interest rates whereas Vo (2009) used RSV in an attempt to explain the behaviour of crude oil prices to forecast their volatility.

One of the limitations of SV models is the problems that arise as a consequence of the intractability of the likelihood function which prohibits its direct evaluation (Hafner and Preminger, 2010). Moreover, according to Mazzeu et al. (2019) the long-memory stochastic volatility models are not an effective alternative due to the complexity regarding their estimation and because they are based on fractional integrated roots. Therefore, in this thesis we consider GARCH type models since the likelihood function is easier to handle than continuous-time models.

2.7 Estimation of model parameters

GARCH models can be estimated via maximum likelihood estimation (MLE). Hamilton (1989) used filter probabilities to evaluate the log-likelihood function and applied a modified Newton-Raphson method to obtain MLE. Expectation-maximization (EM) algorithm is another method to estimate the parameters of a switching regression. EM algorithm maximises the expectation of the log-likelihood function based on the complete data that we observe. Hamilton (1990) used EM algorithm for Markov switching autoregressive models while Kim (1994) gave Markov-switching models a state-space representation together with its filtering and smoothing algorithm. Another method is Markov Chain Monte Carlo (MCMC) method was developed by Kim and Nelson (1998) to fit the Markov switching models which is based on Bayesian inference.

For regime switching models various methods of estimating the parameters have been developed. The computation of the likelihood function of RS GARCH model is not feasible due to the well known problem of path dependence. Path dependence occurs because the current conditional variance depends on the entire sequence of past regimes due to the recursive nature of the GARCH process. When Hamilton and Susmel (1994) introduced an ARCH model with Markov-switching parameters they used ARCH specification instead of GARCH to avoid this problem. Since regimes are not observable when computing the sample likelihood it is required to integrate over all possible paths, which becomes unfeasible since the number of possible regime paths grow exponentially with time. To bypass this problem a variety of approaches have been proposed, see Gray (1996), Dueker (1997), Klaassen (2002). Gray (1996) proposed a method where the path dependence problem is removed by aggregating the conditional variance from the regimes at each step. While integrating out the unobserved regimes Gray's approach uses only a part of available information, while Klaassen (2002) proposed to use all available information. In this thesis Klaassen's path-independent RS GARCH framework that permits the estimation of all model parameters using quasi maximum likelihood estimation (QMLE) will be used. The consistency and asymptotic theory for the quasi maximum likelihood estimator (QMLE) for RS GARCH was established by Xie (2009), whereas the asymptotic properties for the RS GARCH model with finite number of regimes using MCMC estimation have been studied by Bauwens et al. (2010). The non-bayesian estimation of RS GARCH model was studied by Francq et al. (2005) where they proposed to estimate the model by generalised method of moments.

2.8 Crude oil price volatility

Among academicians and policy makers, oil price volatility has become an important topic of discussions for many decades since it has serious implications on the economy of most oil producing nations. During 1980's and 1990's oil prices remained relatively low and stable, however, since 2003, prices started to experience a steady upward trend. Since 2005, this upward increase became more rapid and in July 2008 oil prices reached unprecendented highs to only fall dramatically by the end of the same year. However, since the end of 2008 oil prices have rose sharply until mid-June 2014 when the petroleum prices began to fall worldwide and that drop continued through the end of January 2015. A historic drop occurred on April 2020, when the prices of WTI crude oil dropped by almost 300% and was trading at around negative $37 per barrel.

The fluctuations in the crude oil prices has originated from an imbalance between supply and demand which resulted from events such as wars, geopolitical tensions, changes in political regimes, economic crisis and cost of inflation. Modelling and forecasting crude oil price volatility is crucial in many financial and investment applications since oil prices like many other commodity prices have been volatile and characterised by uncertainties. Furthermore, empirical studies suggest that crude oil time series, also exhibit fat-tail distributions, asymmetry and volatility clustering (Fan et al., 2008).

Even though regime switching models are effective in capturing potential state transition and non-linearity in crude oil price volatility, for forecasting this is inconclusive. For instance, in Sévi (2014) and Zhang et al. (2019) claim that RS GARCH model performs quite well in-sample, but the out-of-sample results show weak limited significance. On the contrary, Hillebrand (2005) claimed that when the model shows high persistence, the forecasts suffer, hence if this persistence is reduced by properly accounting for structural changes one can expect an improvement in forecasting performance. Fong and See (2002) showed that out-of-sample tests indicate that the RS model performs noticeably better than non-switching models when applied to crude oil futures.

2.9 Macroeconomic variables

Many scholars have investigated the impact of macroeconomic variables on volatility forecasts and the underlying mechanisms that drive their influence (Barsky and Kilian, 2004; Asgharian et al., 2013; Conrad et al., 2014; Pan et al., 2017; Conrad et al., 2018; Conrad and Kleen, 2020; Fang et al., 2020; Yu and Huang, 2021; Ma et al., 2021; Chuang and Yang, 2022). And because it is difficult to point out which factors have the dominant effect on the oil prices, numerous determinants such as macroeconomic uncertainty, financial market uncertainty, world uncertainty index, national economic policy uncertainty, (EPU), global economic policy uncertainty (GEPU) indices, along traditional determinants, such as global oil demand, supply, and speculation, were examined for their capacity to predict crude oil price volatility. For instance, while investigating the effect of uncertainty on crude oil returns, Aloui et al. (2016) showed that EPU indices significantly increase crude oil returns, however only during certain periods. This was also supported by Ma et al. (2019) who found that EPU has a positive and significant impact on the crude oil return volatility, especially the U.S. EPU index which has the best forecasting power for crude oil return volatility over the long-term. The effects of macroeconomic uncertainty and financial market uncertainty indices were examined by Chuang and Yang (2022) who concluded that both indices have positive impacts, whereas default yield spread has negative impacts on the crude oil market volatility.

However, if all possible factors are added into the existing model, it may affect the forecasting results because of over-fitting problems. Therefore, of the various determinants, global oil demand and supply have traditionally been considered the most powerful. For instance, Dees et al. (2007) suggest that crude oil prices are mainly influenced by oil supply. Hamilton (2009) observed that oil price fluctuation in 2007-2008 were due to excess demand at the period of declined world production. Barsky and Kilian (2004) and Hamilton (2009) argued that most of crude oil fluctuations experiences in the past were caused by political events rather than market forces. Furthermore, Barsky and Kilian (2004) splits oil price shocks into supply shocks, demand shocks, and precautionary demand shocks, and points out that the oil price rise until mid-2008 was mainly driven by growth in aggregate demand. According to Baumeister and Kilian (2015), the crude oil price fall in 2014 was due to a production that exceeded the demand. Recent pandemic outbreak, COVID-19, has badly affected all sectors by bringing a series of chain reactions such as surge in unemployment, a drop in oil prices and a decline in a stock markets. Since the prices of crude oil have significant influence on the global economy, COVID-19 led to an unexpectedly sharp drop in demand for oil. This drop in demand is essentially caused by the quarantine restrictions of countries which lead to a drop in consumption.

Overall, the oil price fluctuations are found to be susceptible to several factors, described above and including speculative trading, economic outlook, volatility of the real exchange rate and the political stability of major oil exporting countries. However, in the long run, supply and demand are still the primary and main influencing factors of crude oil price fluctuations (Zhao, 2022; Le et al., 2023), hence we use production and demand levels as our macroeconomic variables affecting oil prices.

2.10 Alternative error distributions

When GARCH model was first introduced for modelling financial time series, the main assumption was that the errors are normally distributed, however, empirical evidence suggests that the financial time series is rarely normal but are leptokurtic and often skewed (Bollerslev, 1987). Due to this fact, normal distribution was found to be inappropriate in capturing the tail behaviour of the series. Most often a GARCH model with a non-normal error distribution is required to fully capture the observed fat-tailed behaviour displayed in returns. Therefore, Bollerslev, 1987 proposed to use Student-t distribution in order to capture the long tail behaviour of return series. Fernández and Steel (1998) extended the Student-t distribution by allowing for skewness and proposed a skewed Student-t. Another non-normal distribution is generalised error distribution (GED) proposed by Nelson (1991). These distributions, known as fat-tailed error distributions, are used to describe the characteristics of financial data.

For GARCH model estimation, the QMLE can accommodate for fat-tail situation through their specification which will still generate consistent estimates. However, for regime switching models if the regimes are leptokurtic instead of normal the use of within-regime normality seriously affects the identification of regimes (Klaassen, 2002; Ardia, 2009). Hence an appropriate distribution to capture the excess kurtosis of the time series for RS GARCH models is important.

There exists a great deal of literature on the model distribution comparison, see Gunay (2015), Haas and Liu (2018), Herrera et al. (2018), Hung et al. (2018). Gunay (2015), for instance, found that RS GARCH-t model outperforms all other GARCH models in estimating volatility by accurately capturing the most stylized facts of time series. Haas and Liu (2018) showed that in both in- and out-of-sample estimation, regime-switching models with Student-t innovations dominate their Gaussian counterparts. Even though both models showed fairly persistent regimes, the persistence was more pronounced with Student-t innovations (Haas and Liu, 2018). Furthermore, they concluded that Gaussian specification turns out to suffer from its inability to correctly track regime switching process. Herrera et al. (2018) evaluated out-of-sample forecasting performances of different volatility models including GARCH and RS GARCH of oil returns and showed that due to the extremely high kurtosis in the oil return volatility RS GARCH-t did better job at forecasting and also yielded more accurate long-term forecasts of the spot WTI return volatility. Using three GARCH type models under four different distributions, Hung et al. (2018) found that EGARCH(1,1) model with Student-t distribution provides the most accurate forecast. Therefore, the choice of a distribution is central since specifying the most suitable error distribution in volatility modelling could yield a model which has higher estimation and forecasting ability.

2.11 Endogeneity in regime switching

Although regime switching models have been proven to be quite useful in a wide range of contexts it has some limitations. Firstly, these models assume that the underlying finite state markov chain choosing a state of regime is completely independent from all other parts of the model, thus implying that the future transition between states is completely determined by the current state and does not rely on the observed time series. Secondly, they cannot accommodate non-stationarity in the transition probability, i.e., the markov chain determining the state of regime is assumed to be strictly stationary. All of the regime switching models discussed up to now are assumed to be exogenous, i.e., constant regime switching transition probabilities. However, studies including Kim et al. (2008), Kim (2004, 2009), reported evidence of endogeneity in regime changes. Furthermore, Ma et al. (2021) built a Markov regime switching GARCH MIDAS model with time varying probabilities and showed that it is better than the RS GARCH MIDAS model with constant probabilities. Inference via maximum likelihood estimation is possible with slight modifications to existing Hamilton's recursive filters and the genetic algorithm of Kim (1994).

To the best of our knowledge this is the first attempt to estimate and predict the volatility of crude oil prices using RS GARCH MIDAS approach with Student-t innovations. This extension is particularly important as a regime switching GARCH MIDAS models have been partially implemented by Pan et al. (2017), Ma et al. (2021), Wang et al. (2022) where they only consider Gaussian innovations. Furthermore, there is also a need to study the effect of misspecifying the RS GARCH MIDAS distributional assumption during estimation.

Chapter 3

Methodology

This chapter delves into the specifications, estimation, and forecasting techniques of various models. We start by exploring the GARCH model, followed by discussions on RS GARCH, GARCH MIDAS, RS GARCH MIDAS, and Endogenous RS GARCH models. Finally, we conclude with an examination of forecast evaluation methods.

3.1 The GARCH model

Over the past few years, GARCH models have been extensively used for analysing, modelling and forecasting volatility of a time series because of their ability to capture some behaviours in financial time series such volatility clustering and volatility persistence. The GARCH model proposed by Bollerslev (1986) is an extension of Engle (1982) ARCH model. The ARCH model expresses the conditional variance as a function of past squared returns whereas the GARCH model expresses it as a function of past squared returns and past variances.

Let r_t be a return of a time series. A general GARCH(p, q) model is given by:

$$\begin{aligned} r_t &= \mu_t + a_t, \qquad a_t = \sqrt{h_t}\varepsilon_t, \\ h_t &= \omega + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \quad \forall t \in \mathbb{Z}, \end{aligned} \tag{3.1}$$

where μ_t represents the conditional mean of the return process of r_t and is defined as $\mu_t = E[r_t|\mathcal{F}_{t-1}]$ with $\mathcal{F}_{t-1}$ representing the information set available up to time $t-1$.

The innovation at time t is a_t and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and unit variance. To ensure the conditional variance, h_t, is positive as well as stationary, we set $\omega > 0$, $\alpha_i \geq 0$ for $i = 1, ..., p$, $\beta_j \geq 0$ for $j = 1, ..., q$ and

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{3.2}$$

In Equation (3.1) and Equation (3.2), α_i represents how volatility reacts to new information whereas β_j represents persistence of the volatility. The larger β_j indicates persistent volatility since shocks to the conditional volatility take long time to die out, whereas the larger α_i means larger response of h_t to new information.

Moreover, p and q are the orders of the GARCH model and it may depend on the frequency of the return series. For example, daily returns of a market index often show some minor serial correlations, but monthly returns of the index may not contain any significant serial correlation. In the case where $q = 0$, the model simplifies to Engle's ARCH model, where the conditional variance is given as a function of past squared returns (Engle, 1982). On the other hand, when both p and q are equal to 1, the model represented by Equation (3.1) becomes a GARCH(1,1) model, which is the simplest case of GARCH(p, q) and has the following form:

$$\begin{aligned} a_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= \omega + \alpha a_{t-1}^2 + \beta h_{t-1}, \end{aligned} \tag{3.3}$$

with $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$ (Bollerslev, 1986).

As to the distribution of innovations, $\{\varepsilon_t\}$ is often assumed to follow a standard normal distribution. Although normal GARCH models are able to capture some of the non-linearity displayed in volatility, empirical evidence shows the usefulness of the non-normal distributions such as Student-t, generalised error distribution (GED), skewed Student-t distribution, skewed generalised error distribution (GED) and normal inverse Gaussian (NIG) when dealing with fat tails of the conditional errors, extreme observations, skewness and outliers in returns. Bollerslev (1987) suggested replacing the assumption of conditional Normality with that of conditional Student-t distribution in estimation of volatility since Gaussian GARCH model could not explain the leptokurtosis displayed in returns. The GED distribution proposed by Nelson (1991) was used in estimation of EGARCH models. Skewed Student-t distribution was proposed by Hansen (1994) and Fernández and Steel (1998), whereas the skewed GED was first used by Theodossiou (1998) to capture the skewedness of GED. These non-normal distributions have been proven to describe the volatility better than normal distribution in GARCH models (Bollerslev, 1987; Hansen, 1994).

There have been studies that compared the performance of different error distributions in financial modelling and forecasting, including the Student-t, skewed Student-t, and GED and skewed GED. Empirical evidence has suggested that the Student-t distribution can often provide better fit and forecasting performance compared to alternative distributions. For instance, Wilhelmsson (2006) conducted a comparative study of nine different error distributions, including the Student-t distribution, in GARCH modeling. The results showed that the model estimated with the Student-t distribution exhibited superior performance compared to other distributions considered. Liu and Morley (2009) found that the model with the Student-t distribution slightly outperformed the one with the GED distribution in terms of forecasting accuracy. Similarly, Mattera et al. (2018) compared six alternative distributions in GARCH modeling and forecasting. They found that the model with the Student-t distribution consistently outperformed the other distributions in terms of in-sample evaluation.

These studies provide empirical support for the effectiveness of the Student-t distribution in financial modelling and forecasting. Therefore, in this thesis as an alternative to Gaussian we consider standardised Student-t distribution.

3.1.1 Model estimation

Maximum likelihood estimation (MLE) method is typically used to estimate GARCH parameters. However, due to the complexity of the maximum likelihood function for GARCH$(1, 1)$ models, one generally uses an approximating function

$$f(\boldsymbol{\theta}|x_0,\ldots,x_n) = f(x_0,\ldots,x_n|\boldsymbol{\theta}) = f(x_n|\mathcal{F}_{n-1})f(x_{n-1}|\mathcal{F}_{n-2})\ldots f(x_1|\mathcal{F}_0)f(x_0|\boldsymbol{\theta}) \tag{3.4}$$

where $\boldsymbol{\theta}$ is a vector of parameters to be estimated, $f(x_0,\ldots x_n|\boldsymbol{\theta})$ is the joint probability distribution of $\{r_0,\ldots r_n\}$ in a GARCH model with parameters $\boldsymbol{\theta}$. To derive the likelihood the distribution of $\{\varepsilon_t\}$ also needs to be specified. It is important to note that even when $\{\varepsilon_t\}$ is assumed to be i.i.d. and Normally distributed since the distribution of h_t is unknown the distribution of a_t is also not known. A standard way in estimating GARCH models when deriving the likelihood is to make use of Equation (3.4) by assuming $\{\varepsilon_t\}$ is Gaussian and then remove this assumption later. This is called Quasi-maximum likelihood estimation (QMLE). The general idea of the QMLE is to construct a likelihood function conditionally on random or fixed initial values and based on the sample data and its assumed distribution, then by maximising this likelihood function the parameter estimates can be obtained. Using an iterative algorithm this method provides the parameters that are most likely to describe the observed data. Furthermore, this method provides consistent and asymptotically normal estimators for strictly stationary GARCH models (Francq and Zakoian, 2019). The asymptotic properties of QMLEs for the GARCH(1,1) models have been studied by Lee and Hansen (1994) and Lumsdaine (1996) whereas the general GARCH(p,q) models were studied by Horv et al. (2003) and Francq and Zakoian (2019).

3.1.1.1 GARCH with Gaussian innovations

Let the observations $\{a_1,...,a_n\}$ represent a realisation of length n of the unique non-anticipative strictly stationary solution a_t of a GARCH(p,q) model. The parameters are $\boldsymbol{\theta} = (\omega,\alpha_1,...,\alpha_p,\beta_1,...,\beta_q)'$, while the true value of the parameters $\boldsymbol{\theta}_0 = (\omega_0,\alpha_{01},...,\alpha_{0p},\beta_{01},...,\beta_{0q})'$ are unknown. Here no assumptions on the distribution of these variables are made. Instead we consider Gaussian quasi-likelihood, which conditional on initial values coincides with the likelihood when the $\{\varepsilon_t\}$ are distributed as standard Normal. The initial values, i.e. $a_0,...,a_{1-p}$ are unavailable and $\tilde{\sigma_0},...,\tilde{\sigma}_{1-q}$ are unobservable. For a given value of $\boldsymbol{\theta}$, under the second-order stationarity assumption, a reasonable choice for the unknown initial values is the unconditional variance:

$$a_0^2 = ... = a_{1-p}^2 = \sigma_0^2 = ... = \sigma_{1-q}^2 = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}. \tag{3.5}$$

However, such initial values are not always suitable, for instance for IGARCH model, resulting in Equation (3.5) being undefined. Hence a suitable initial values are set to:

$$a_0^2 = ... = a_{1-p}^2 = \tilde{\sigma}_0^2 = ... = \tilde{\sigma}_{1-q}^2 = \omega, \tag{3.6}$$

or

$$a_0^2 = ... = a_{1-p}^2 = \tilde{\sigma}_0^2 = ... = \tilde{\sigma}_{1-q}^2 = a_1^2. \tag{3.7}$$

Given initial values, the conditional Gaussian quasi-likelihood function, denoted as $\mathcal{L}_n(\boldsymbol{\theta})$, is given by

$$\mathcal{L}_n(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}; a_1, ..., a_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t}} \exp\Big(- \frac{a_t^2}{2\tilde{\sigma}_t} \Big), \tag{3.8}$$

where $\tilde{\sigma}_t$ are defined recursively, for $t \geq 1$ by

$$\tilde{\sigma}_t = \tilde{\sigma}_t(\boldsymbol{\theta}) = \omega + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{t-j}. \tag{3.9}$$

A QMLE of $\boldsymbol{\theta}$ is then defined as any measurable solution $\hat{\boldsymbol{\theta}}_n$ of:

$$\hat{\boldsymbol{\theta}}_n = \text{argmax} \mathcal{L}_n(\boldsymbol{\theta}). \tag{3.10}$$

Taking the logarithms, maximising the likelihood is equivalent to minimising with respect to $\boldsymbol{\theta}$:

$$\tilde{\mathbf{I}}_n(\boldsymbol{\theta})^1 - \frac{1}{n} \sum_{t=1}^{n} \log\left(\tilde{\sigma}_t\right) + \frac{a_t^2}{\tilde{\sigma}_t}, \tag{3.11}$$

[1]Computation of $\tilde{\mathbf{I}}_n(\boldsymbol{\theta})$ involves number of operations of order n^2, whereas Francq and Zakoian (2019) propose a method which involves a number of order n. More details can be found in Francq and Zakoian (2019), Ch.7

and $\tilde{\sigma_t}^2$ is defined in Equation (3.9). A QMLE is thus a measurable solution $\hat{\boldsymbol{\theta}}_n$ of:

$$\hat{\boldsymbol{\theta}}_n = \text{argmin}\tilde{\mathbf{I}}_n(\boldsymbol{\theta}). \tag{3.12}$$

Consider the simplest case of GARCH(1,1) model with normally distributed innovations, the log QML function is a function of $\boldsymbol{\theta}$ (with fixed $a_1, ..., a_n$) is given by:

$$\log \mathcal{L}_n(\boldsymbol{\theta}) = -\frac{1}{2}\left[n\log(2\pi) + \sum_{t=1}^{n}\left(\log\,(\tilde{\sigma}_t) + \frac{a_t^2}{\tilde{\sigma}_t}\right)\right], \tag{3.13}$$

where $\tilde{\sigma}_t$ is given in Equation (3.9) with $p = q = 1$. The likelihood function in Equation (3.13) is then maximised to obtain QML estimates of the parameters in the conditional variance equation in addition to any parameters in the density function. In this case the parameters are $\boldsymbol{\theta} = (\omega, \alpha, \beta)$. The choice of initial values are not important for the asymptotic properties of the QMLE (Francq and Zakoian, 2019).

3.1.1.2 GARCH with Student-t innovations

The Student-t distribution has an additional parameter v to describe the degrees of freedom which controls the fatness of the tails. In addition, Student-t distribution can accommodate the excess kurtosis of the innovations (Bollerslev, 1987). The kurtosis of a random variable that follows a Student-t distribution is higher for lower v. When $v \to \infty$ the Student-t distribution converges to Normal distribution whereas a lower value of v indicates fatter tails. The constraint $v > 2$ is imposed to ensure that the second order moment exist. Let t_v denote a Student-t distribution with v degrees of freedom. Then:

$$\text{Var}(t_v) = \frac{v}{v-2}, \qquad v > 2. \tag{3.14}$$

The probability density function of ε_t can be written as:

$$f(\varepsilon_t|v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{(v-2)\pi}\Gamma(\frac{v}{2})}\left[1 + \frac{\varepsilon_t^2}{(v-2)}\right]^{-\frac{v+1}{2}}, \quad v > 2, \tag{3.15}$$

where $\varepsilon_t = t_v/\sqrt{v/v-2}$ and $\Gamma(\cdot)$ is the usual Gamma function.

Following the same procedure as for the normal distribution and using $a_t = \sqrt{h_t}\varepsilon_t$, the conditional log-likelihood becomes:

$$\ell_n(\boldsymbol{\theta}) = \log\mathcal{L}_n(\boldsymbol{\theta}) = n\log\left[\frac{\Gamma(\frac{v+1}{2})}{\sqrt{(v-2)\pi}\Gamma(\frac{v}{2})}\right] - \frac{1}{2}\sum_{t=1}^{n}\log(\tilde{\sigma}_t) - \frac{(1+v)}{2}\sum_{t=1}^{n}\log\left[1+\frac{a_t^2}{\tilde{\sigma}_t(v-2)}\right]. \tag{3.16}$$

The unknown parameters set is $\boldsymbol{\theta} = (\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q, v)'$. For the simple GARCH-(1,1) model given in Equation (3.3) with Student-t innovations, the quasi log-likelihood function is given by:

$$\log\,\mathcal{L}_n(\boldsymbol{\theta}|a) = n\log\left[\frac{\Gamma(\frac{v+1}{2})}{\sqrt{(v-2)\pi}\Gamma(\frac{v}{2})}\right] - \frac{1}{2}\sum_{t=1}^{n}\log(\omega+\alpha a_{t-1}^2+\beta\hat{h}_{t-1}) - \frac{(1+v)}{2}\sum_{t=1}^{n}\log\left[1+\frac{a_t^2}{(\omega+\alpha a_{t-1}^2+\beta\hat{h}_{t-1})(v-2)}\right]. \tag{3.17}$$

In this case the parameter set is $\boldsymbol{\theta} = (\omega, \alpha, \beta, v)$.

3.1.2 Out-of-sample forecasting

For the GARCH$(1,1)$ model in Equation (3.3), suppose that we have information up to time t and are interested in forecasting h_{t+l}, where $l \geq 1$. The positive integer l is the forecast horizon and t is the forecast origin. Thus for one-step-ahead forecast we obtain:

$$h_{t+1} = \omega + \alpha a_t^2 + \beta h_t, \tag{3.18}$$

where it is assumed that a_t and h_t are known at time index t. Similarly for $l = 2$ we get:

$$\begin{aligned} h_{t+2} &= \omega + \alpha E[a_{t+1}^2|\mathcal{F}_t] + \beta h_{t+1} \\ &= \omega + (\alpha+\beta)h_{t+1} \end{aligned} \tag{3.19}$$

For $l = 3$:

$$\begin{aligned} h_{t+3} &= \omega + \alpha E[a_{t+2}^2|\mathcal{F}_t] + \beta h_{t+2} \\ &= \omega + (\alpha+\beta)h_{t+2} \\ &= \omega + (\alpha+\beta)(\omega + (\alpha+\beta)h_{t+1}) \\ &= \omega(1+\alpha+\beta) + (\alpha+\beta)^2(h_{t+1}). \end{aligned} \tag{3.20}$$

For $l = 4$:

$$
\begin{aligned}
h_{t+4} &= \omega + \alpha E[a_{t+3}^2 | \mathcal{F}_t] + \beta h_{t+3} \\
&= \omega + (\alpha + \beta) h_{t+3} \\
&= \omega + (\alpha + \beta)(\omega + (\alpha + \beta) h_{t+2}) \\
&= \omega(1 + (\alpha + \beta) + (\alpha + \beta)^2) + (\alpha + \beta)^2 (h_{t+1}).
\end{aligned}
\tag{3.21}
$$

In general, the l-step-ahead forecast of the conditional variance, for $l \geq 2$ is:

$$
\begin{aligned}
h_{t+l} =& \omega + (\alpha + \beta) h_{t+l-1} \\
=& [\omega \sum_{i=0}^{l-1} (\alpha + \beta)^i] + (\alpha + \beta)^{l-1} (\alpha a_t^2 + \beta h_t) \\
=& \omega(1 + (\alpha + \beta) + (\alpha + \beta)^2 + \ldots + (\alpha + \beta)^{l-2}) + (\alpha + \beta)^{l-1} h_{t+1} \\
=& \frac{\omega[1 - (\alpha + \beta)^{l-1}]}{1 - (\alpha + \beta)} + (\alpha + \beta)^{l-1} h_{t+1}.
\end{aligned}
\tag{3.22}
$$

As the forecast horizon tends to infinity, provided that $\mathrm{Var}(a_t)$ exists, the multi-step-ahead volatility forecasts of GARCH(1,1) converge to the unconditional variance of the process, i.e., as $l \to \infty$ we have:

$$
h_{t+l} = \omega(1 + (\alpha + \beta) + (\alpha + \beta)^2 + \ldots) \to \frac{\omega}{1 - (\alpha + \beta)} \tag{3.23}
$$

provided that $\alpha + \beta < 1$. The forecasting steps for GARCH(1,1) model can easily be generalised to GARCH(p, q) model.

3.2 The RS GARCH model

Although GARCH model is able to capture the volatility displayed in time series, when the returns experience a short period of high volatility the GARCH model tends to overestimate the persistence which leads to an overestimation of the conditional volatility in a period after the shock (Lamoureux and Lastrapes, 1990). Mikosch and Stărică (2004) showed that the high persistence can be explained by level shifts in the unconditional variance. This was also supported by Lamoureux and Lastrapes (1990) who demonstrated that by introducing deterministic shifts in the variance reduced the degree of volatility persistence. This suggests that to prevent overestimation of this persistence we allow for structural breaks by merging GARCH with regime switching model first introduced by Hamilton (1989). The general idea behind the Markov switching GARCH (RS GARCH) model is to reduce the long GARCH persistence by switching from one variance structure to another. Hamilton and Susmel (1994) were among the first scholars to discuss the RS GARCH model. A distinguishing feature of regime switching models is the possibility for some, or for all parameters to switch throughout the whole sample period. This switch is governed by a state variable s_t based on a Markov process. An advantage of RS GARCH models is its ability to deal with fat-tails (Haas et al., 2004).

Let r_t be the return of time series. The general regime switching GARCH(p, q) (RS GARCH(p, q)) model is represented as follows:

$$\begin{aligned} r_t = \mu_t + a_t, \qquad a_t = \sqrt{h_{s_t,t}}\varepsilon_t, \\ h_{s_t,t} = \omega_{s_t} + \sum_{i=1}^{p} \alpha_{s_t,i} a_{t-i}^2 + \sum_{j=1}^{q} \beta_{s_t,j} h_{s_t,t-j}, \end{aligned} \tag{3.24}$$

where μ_t is the conditional mean of the return process r_t defined as $\mu_t = E[r_t|\mathcal{F}_{t-1}]$ and $h_{s_t,t}$ is the conditional variance $\text{Var}(r_t|\mathcal{F}_{t-1})$ with $\mathcal{F}_{t-1}$ representing the information set available up to time $t-1$. As usual, the $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean equal to zero and unit variance. The regime switching process, s_t, indicates the states of the market and the necessary conditions to ensure the conditional variance remains positive in each regime are same as in GARCH(p, q) model given in Equation (3.2).

The state variable, s_t, in Equation (3.24) evolves according to a first-order Markov chain with the following transition probabilities:

$$Pr(s_t = j|s_{t-1} = i) = p_{ij}, \tag{3.25}$$

which indicates the probability of switching from state i at time $t-1$ to state j at time t. By grouping together these probabilities we obtain a transition matrix P which is given by:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix}.$$

The simplest case of RS GARCH(p, q) model, namely RS GARCH(1,1) model is when $p = q = 1$, and only considering two states can be written as follows:

$$\begin{aligned} r_t &= \mu_t + a_t, \qquad a_t = \sqrt{h_t}\varepsilon_t, \\ h_{s_t,t} &= \begin{cases} \omega_1 + \alpha_1 a_{t-1}^2 + \beta_1 h_{1,t-1} & \text{when} \quad s_t = 1, \\ \omega_2 + \alpha_2 a_{t-1}^2 + \beta_2 h_{2,t-1} & \text{when} \quad s_t = 2, \end{cases} \end{aligned} \tag{3.26}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and unit variance. Furthermore, $h_{s_t,t}$ is the conditional volatility in state s_t at time t, where s_t is assumed to be a stationary, irreducible Markov process with discrete state space $\{1, 2\}$ and transition probabilities defined in Equation (3.25). To ensure the conditional volatility is positive as well as stationary we set $\omega_{1,2} > 0$, $\alpha_{1,2} \geq 0$, $\beta_{1,2} \geq 0$ and for process to be covariance stationary we set $\alpha_{1,2} + \beta_{1,2} < 1$.

A two-state transition matrix is given by:

$$\begin{aligned} P &= \begin{bmatrix} Pr(s_t = 1 \mid s_{t-1} = 1) & Pr(s_t = 2 \mid s_{t-1} = 1) \\ Pr(s_t = 1 \mid s_{t-1} = 2) & Pr(s_t = 2 \mid s_{t-1} = 2) \end{bmatrix} \\ &= \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \end{aligned} \tag{3.27}$$

This matrix governs the random behaviour of the state variable and it only contains two parameters (p_{11}, p_{22}) which determine the persistence of each regime. If p_{11} and p_{22} are large then the model has highly persistent regimes and the transition between regimes will be limited.

Most available RS GARCH models are constructed based on the assumption that the innovations follow a Gaussian distribution. Since regime switching can account for large unconditional kurtosis, the non-normal distributions might be considered not necessary. However, empirical evidence suggests that by introducing Student-t innovations, estimation and volatility forecasting performance of RS GARCH models are significantly improved (Gunay, 2015; Haas and Liu, 2018; Herrera et al., 2018) because it can capture the observed fat-tailed behaviour displayed in returns. Furthermore, Klaassen (2002) showed that Student-t innovations improves the stability of regimes.

Furthermore, it should be noted that the single-regime GARCH is nested in RS GARCH where the regime specific parameters $(\omega_i, \alpha_i, \beta_i)$ are equal across the two regimes. A difficulty with this model is that it may not be easy to estimate the parameters because the state variables are unobservable and transition probabilities are unknown. Hence statistical inference is drawn based only on observed a_t.

3.2.1 Model estimation

Here we focus on the model in Equation (3.26) and discuss its quasi-maximum likelihood estimation. The consistency and asymptotic theory for the quasi maximum likelihood estimator (QMLE) for RS GARCH can be found in Xie (2009).

A two-state RS GARCH(1,1) model with Gaussian innovations can be written as follows:

$$\begin{aligned} r_t = \mu + a_t, \qquad a_t = \sqrt{h_t}\varepsilon_t, \qquad \varepsilon_t \overset{\text{i.i.d}}{\sim} N(0,1), \\ h_{s_t,t} = \omega_{s_t} + \alpha_{s_t} a_{t-1}^2 + \beta_{s_t} h_{t-1}, \end{aligned} \tag{3.28}$$

where $s_t = \{1, 2\}$, a_t is the error at time t, and $\{\varepsilon_t\}$ is an i.i.d. sequence with mean zero and unit variance. The vector of parameters is $\boldsymbol{\theta} = (\omega_1, \omega_2, \alpha_1, \alpha_2, \beta_1, \beta_2, p_{11}, p_{22})'$.

The GARCH specification in Equation (3.28) implies that conditional variance at time t depends not only on s_t but also indirectly on $\{s_{t-1}, s_{t-2}, ...\}$. That is to say that $h_{s_t,t}$ at time t is depended on the whole unobserved regime path $s_t = \{s_t, s_{t-1}, ..., s_{t-n}\}$. The conditional variance is therefore given as:

$$h_{s_t,t} = \mathrm{Var}[a_t | s_t, \mathcal{F}_{t-1}]. \tag{3.29}$$

This makes the model more complex and the estimation procedure becomes intractable since the number of possible paths of the regime process grows exponentially as t increases. The likelihood function has to be constructed by integrating out all possible paths since regimes are unobserved (Klaassen, 2002). This makes the estimation of log-likelihood infeasible and leads to a path dependence problem. Furthermore, Hamilton and Susmel (1994) showed that estimating path dependence is challenging because the exact computation of the likelihood is infeasible in practice. Consequently some authors proposed to use a modified versions of RS GARCH to avoid this problem. For example, Cai (1994) and Hamilton and Susmel (1994) reduced the RS GARCH model to an RS-ARCH model by leaving out the GARCH term. With their specification, the conditional variance only depended on the current regime, which allowed them to bypass this path dependence issue. However, this solution is not very feasible since the persistence parameter β is discarded. To tackle the path dependence, Gray (1996) proposed to use conditional expectation of the lagged conditional variance $E_{t-2}\{h_{s_t,t-1}\}$ instead of lagged conditional variance $h_{s_t,t-1}$ in Equation (3.26) thus by summing over all possible regimes it makes it possible to construct the conditional variance by integrating out the dependence on the entire regime. Hence there will be no need to consider all possible values of $(s_t, ..., s_1)$. Gray's approach can be written as follows:

$$h_{s_t,t-1} = E_{t-2}\{h_{s_t,t-1}\} = Pr(s_{t-1} = 1|\mathcal{F}_{t-2})h_{1,t-1} + Pr(s_{t-1} = 2|\mathcal{F}_{t-2})h_{2,t-1}. \quad (3.30)$$

Furthermore, Gray showed that by applying this method, the likelihood function can be evaluated in a first order recursive way. A modification of Gray's approach was suggested by Klaassen (2002) where they proposed to use the conditional expected value $E_{t-1}\{h_{t-1}|s_t\}$ instead of $E_{t-2}\{h_{s_t,t-1}\}$. This means that when integrating out the previous regime s_{t-1}, Klaassen's specification uses the information up to time $t-1$ with regime s_t, whereas Gray's approach only considers information up to time $t-2$ with regime s_{t-1}. Moreover, Klaassen (2002) has argued that if regimes are highly persistent then current regime provides valuable information about past regime and thus it should be included in the probability calculation.

Klaassen's specification for the conditional variance can be written as:

$$h_{s_t,t} = \omega_{s_t} + \alpha_{s_t} a_{t-1}^2 + \beta_{s_t} E_{t-1}\left\{h_{s_t,t-1}|s_t\right\}. \tag{3.31}$$

where the expected conditional variance is calculated as:

$$\begin{aligned} E_{t-1}\left\{h_{s_t,t-1}|s_t\right\} &= E_{t-1}\left\{h_{s_t,t-1}|s_t = i\right\} \\ &= \sum_{j=1}^{2} \tilde{p}_{ji,t-1}\left[\mu^2 + h_{s_t,t-1}\right] - \left[\sum_{j=1}^{2} \tilde{p}_{ji,t-1}\mu\right]^2, \end{aligned} \tag{3.32}$$

with the probabilities $\tilde{p}_{ji,t-1}$ given by:

$$\begin{aligned} \tilde{p}_{ji,t-1} &= Pr(s_{t-1} = j|s_t = i, \mathcal{F}_{t-1}) \\ &= \frac{Pr(s_t = i|s_{t-1} = j)Pr(s_{t-1} = j|\mathcal{F}_{t-1})}{Pr(s_t = i|\mathcal{F}_{t-1})} \\ &= \frac{p_{ji}p_{j,t-1}}{p_{i,t}}, \end{aligned} \tag{3.33}$$

where $i, j = 1, 2$ denotes the two regimes and s_t is the regime variable. The aggregated conditional variance is a weighted average of the regime dependent conditional variances weighted with the probability of being in the specific regime conditioned on all observed information. A notable feature of Equation (3.32) is that both $h_{1,t-1}$ and $h_{2,t-1}$ have been used to form h_{t-1} making it path independent. Thus, this model can be computed without considering all possible values of $(s_t, ..., s_1)$. Since Klaassen's approach contains information up to time $t-1$, we will be using this approach. Moreover, comparing the RS ARCH model of Hamilton and Susmel (1994) and Cai (1994), the RS GARCH model of Klaassen (2002) allows all the GARCH parameters to switch and does not impose any constraints on these parameters thus offering much more flexibility. Another advantage of Klaassen's method compared to Gray's approach is that it provides a straightforward expression for the multi-step ahead volatility forecasts that can be computed recursively as in standard GARCH model (Marcucci, 2005).

Let $\mathcal{F}_t = \{a_t, ..., a_1\}$ denote a collection of all observed variables up to time t, which represents the information set we have at time t. Then $\mathcal{F}_T$ denotes the information set based on the full sample. The density of returns conditioned on all observed history is defined as $f(a_t|\mathcal{F}_{t-1};\boldsymbol{\theta})$. Since a_t is set to switch between different two regimes we can rewrite the conditional density as a weighted average of the joint density functions, weighted with the probability of being in regime j at time t. This takes the following form:

$$\begin{aligned} f(a_t|\mathcal{F}_{t-1};\boldsymbol{\theta}) &= \sum_{j=1}^{2} f(a_t, s_t = j|\mathcal{F}_{t-1};\boldsymbol{\theta}) \\ &= \sum_{j=1}^{2} f(a_t|s_t = j, \mathcal{F}_{t-1};\boldsymbol{\theta}) Pr(s_t = j|\mathcal{F}_{t-1};\boldsymbol{\theta}) \\ &= \sum_{j=1}^{2} f(a_t|s_t = j, \mathcal{F}_{t-1};\boldsymbol{\theta}) \times p_{j,t}, \end{aligned} \tag{3.34}$$

where the density of a_t conditional on $\mathcal{F}_{t-1}$ and random variable s_t with the assumption of Normality is then given by:

$$f(a_t|s_t = j, \mathcal{F}_{t-1};\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi h_{j,t}}} \exp\Big(-\frac{a_t^2}{2h_{j,t}}\Big). \tag{3.35}$$

In Equation (3.34), $p_{j,t}$ are the predicted probabilities and it is the probability of being in state j at time t given all available information up to time $t-1$. Since the states s_t are unobservable, Hamilton (1989) introduced an iterative algorithm, namely Hamilton's filter, to draw probabilistic inference about this unobservable state given observations on a_t. Applying Hamilton's filter, the predicted (prior) probabilities can be written as:

$$Pr(s_t = j|\mathcal{F}_{t-1};\boldsymbol{\theta}) = \sum_{i=1}^{2} Pr(s_t = j|s_{t-1} = i) Pr(s_{t-1} = i|\mathcal{F}_{t-1};\boldsymbol{\theta}), \tag{3.36}$$

where $Pr(s_t = j|s_{t-1} = i)$ are the transition probabilities given in Equation (3.25).

The probability of being in regime 1 at time t given all observable information up to time $t-1$ is derived as:

$$\begin{aligned} Pr(s_t = 1|\mathcal{F}_{t-1};\boldsymbol{\theta}) &= Pr(s_t = 1|s_{t-1} = 1)Pr(s_{t-1} = 1|\mathcal{F}_{t-1};\boldsymbol{\theta}) \\ &\quad + Pr(s_t = 1|s_{t-1} = 2)Pr(s_{t-1} = 2|\mathcal{F}_{t-1}) \\ &= p_{11}Pr(s_{t-1} = 1|\mathcal{F}_{t-1};\boldsymbol{\theta}) + (1-p_{22})Pr(s_{t-1} = 2|\mathcal{F}_{t-1};\boldsymbol{\theta}) \\ &= p_{11}Pr(s_{t-1} = 1|\mathcal{F}_{t-1};\boldsymbol{\theta}) + (1-p_{22})(1 - Pr(s_{t-1} = 1|\mathcal{F}_{t-1};\boldsymbol{\theta})), \end{aligned} \tag{3.37}$$

where p_{11} and p_{22} are the transition probability parameters to be estimated. Similarly, one can obtain the probability of being in regime 2 at time t given all observable information up to time $t-1$.

By applying the Bayes theorem and the law of total probability, the posterior probabilities also known as filtered probabilities can be written as:

$$\begin{aligned} Pr(s_{t-1} = i|\mathcal{F}_{t-1};\boldsymbol{\theta}) &\equiv Pr(s_{t-1} = i|a_{t-1}, \mathcal{F}_{t-2};\boldsymbol{\theta}) \\ &= \frac{f(a_{t-1}|s_{t-1} = i, \mathcal{F}_{t-2};\boldsymbol{\theta})Pr(s_{t-1} = i|\mathcal{F}_{t-2};\boldsymbol{\theta})}{\sum_{i=1}^{2} f(a_{t-1}|s_{t-1} = i, \mathcal{F}_{t-2};\boldsymbol{\theta})Pr(s_{t-1} = i|\mathcal{F}_{t-2};\boldsymbol{\theta})}. \end{aligned} \tag{3.38}$$

By substituting the Equation (3.38) in Equation (3.36), the probability if being in regime j at time t can be calculated. Since probabilities only depend on constants, p_{11}, p_{22}, last periods regime probability and densities, the regime probability are called a first-order recursive process. The filtered probabilities $Pr(s_{t-1} = i|\mathcal{F}_{t-1};\boldsymbol{\theta})$ for $t = t+1$ are obtained recursively from Hamilton's filter[2].

Having specified the dynamics of regime switching probabilities, the sample quasi-log-likelihood function can be calculated as:

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\theta}) &= \sum_{t=1}^{n} \log[f(a_t|s_t = j, \mathcal{F}_{t-1};\boldsymbol{\theta})] \\ &= \sum_{t=1}^{n} \log\left[p_{1,t}f(a_t|s_t = 1, \mathcal{F}_{t-1};\boldsymbol{\theta}) + p_{2,t}f(a_t|s_t = 2, \mathcal{F}_{t-1};\boldsymbol{\theta})\right], \end{aligned} \tag{3.39}$$

[2] see Hamilton (2020)[Chapter 22] for more details.

or written completely

$$\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log \left[p_{1,t} \frac{1}{\sqrt{2\pi h_{s_t,t}}} \exp\left(-\frac{a_t^2}{2h_{s_t,t}}\right) + (1-p_{1,t}) \frac{1}{\sqrt{2\pi h_{s_t,t}}} \exp\left(-\frac{a_t^2}{2h_{s_t,t}}\right) \right]. \tag{3.40}$$

In Equation (3.40), as discussed previously, $h_{s_t,t}$ suffers from path dependence problem. To overcome this, Klaassen's approach described in Equation (3.31) to Equation (3.33) is implemented. By maximising the log-likelihood function in Equation (3.39) the QML estimates for RS GARCH(1,1) model are obtained. The detailed flow-chart for recursive calculation of Hamilton's filter probabilities and the log-likelihood value for RS GARCH(1,1) model with two-regimes is given in Figure 3.1.

In some applications, it is more appropriate to assume that $\{\varepsilon_t\}$ follows a heavy-tailed distribution such as a standardised Student-t distribution. The parameter set of RS GARCH(1,1) model with standardised Student-t innovations is
$\boldsymbol{\theta} = (\omega_1, \omega_2, \alpha_1, \alpha_2, \beta_1, \beta_2, p_{11}, p_{22}, v)$, where v represents degrees of freedom. The calculation of quasi-log-likelihood function is similar to Equation (3.39) with the conditional density of returns changed from Normal to Student-t density function.

Figure 3.1: Flow-chart for recursive calculation of Hamilton's filter probabilities for RS GARCH.

$$Pr(s_{t-1} = i \mid a_{t-1}, ..., a_1;\ \boldsymbol{\theta})$$

$$\downarrow$$

$$Pr(s_t = j,\ s_{t-1} = i \mid a_{t-1}, ..., a_1;\ \boldsymbol{\theta}) = Pr(s_{t-1} = i \mid a_{t-1}, ..., a_1; \boldsymbol{\theta}) \times Pr(s_t = j \mid s_{t-1} = i)$$

$$\downarrow$$

$$Pr(s_t = j \mid a_{t-1}, ..., a_1; \boldsymbol{\theta}) = \sum_{i=1}^{2} Pr(s_t = j,\ s_{t-1} = i \mid a_{t-1}, ..., a_1;\ \boldsymbol{\theta})$$

$$\downarrow$$

$$a_t = r_t, \quad h_{s_t,t} = \omega_{s_t} + \alpha_{s_t} a_{t-1}^2 + \beta_{s_t} E_{t-1}\{h_{t-1} \mid s_t\}$$

$$E_{t-1}\{h_{t-1} \mid s_t = i\} = \sum_{j=1}^{2} \tilde{p}_{ji,t-1} \times h_{s_t,t-1}$$

$$\downarrow$$

$$Pr(s_{t+1} = i \mid a_t) = \sum_{j=1}^{2} Pr(s_t = j \mid a_t, ..., a_1;\ \boldsymbol{\theta}) \times\ Pr(s_{t+1} = i \mid s_t = j)$$

$$\downarrow$$

$$\tilde{p}_{ji,t} = Pr(s_t = j \mid s_{t+1} = i,\ a_t) = \frac{p_{ji} Pr(s_t = j \mid a_{t-1})}{Pr(s_{t+1} = i \mid a_t)} = \frac{p_{ji} p_{j,t}}{p_{i,t+1}}$$

$$\downarrow$$

$$f(a_t,\ s_t = j \mid a_{t-1}, ..., a_1;\ \boldsymbol{\theta}) = f(r_t \mid s_t = j, a_{t-1}, ..., a_1;\ \boldsymbol{\theta}) \times Pr(s_t = j \mid a_{t-1}, ..., a_1; \boldsymbol{\theta})$$

$$\downarrow$$

$$f(r_t \mid a_{t-1}, ..., a_1;\ \boldsymbol{\theta}) = \sum_{j=1}^{2} f(a_t,\ s_t = j \mid a_{t-1}, ..., a_1;\ \boldsymbol{\theta})$$

$$\downarrow$$

$$Pr(s_t = j \mid a_t, ..., a_1;\ \boldsymbol{\theta}) = \frac{f(a_t,\ s_t = j \mid a_{t-1}, ..., a_1;\ \boldsymbol{\theta})}{f(a_t \mid a_{t-1}, ..., a_1;\ \boldsymbol{\theta})}$$

$$t = t + 1$$

Notes: This flow-chart provides recursive calculation of Hamilton's filter probabilities when model is two-state RS GARCH(1,1). The $\tilde{p}_{ji,t-1}$ can be obtained from Equation (3.33).

3.2.2 Out-of-sample forecasting

For the RS GARCH (1,1) model one-step-ahead volatility is calculated as the weighted average of volatility forecasts in two regimes:

$$h_{t+1} = Pr(s_{t+1} = 1|\mathcal{F}_t)h_{t+1}^{(1)} + P(s_{t+1} = 2|\mathcal{F}_t)h_{t+1}^{(2)}, \tag{3.41}$$

where

$$\begin{aligned} Pr(s_{t+1} = 1|\mathcal{F}_t) &= p_{11}Pr(s_t = 1|\mathcal{F}_t) + p_{21}Pr(s_t = 2|\mathcal{F}_t), \\ Pr(s_{t+1} = 2|\mathcal{F}_t) &= p_{12}Pr(s_t = 1|\mathcal{F}_t) + p_{22}Pr(s_t = 2|\mathcal{F}_t). \end{aligned} \tag{3.42}$$

The $Pr(s_t = 1|\mathcal{F}_t)$ and $Pr(s_t = 2|\mathcal{F}_t)$ are the filter probabilities obtained using Hamilton's filter at time t and p_{11}, p_{22} are the transition probabilities estimated from the model.

As for further horizons, following Klaassen (2002), we need to sum the actual volatility during the k periods to forecast the k-step-ahead volatility at time t as:

$$h_{t,t+k} = \sum_{r=1}^{k} h_{t,t+r} = \sum_{r=1}^{k}\sum_{i=1}^{2} Pr(s_{t+r} = i|\mathcal{F}_t)h_{(s_{t+r}=i),t,t+r}, \tag{3.43}$$

where $h_{t,t+k}$ denotes the volatility forecasts at time t for the next k steps and $h_{(s_{t+r}=i),t,t+r}$ can be computed recursively as:

$$h_{(s_{t+r}=i),t,t+r} = \omega_{s_t} + (\alpha_{s_t} + \beta_{s_t})E\{h_{(s_{t+r=i}),t+r-1}|s_{t+r=i}\}, \tag{3.44}$$

is the r-step ahead volatility forecast of regime i at time t.

The above specification provides with a straightforward expression for multi-step ahead forecast that can be obtained recursively in a similar manner to GARCH model.

3.3 The GARCH MIDAS model

In this subsection the GARCH MIDAS model is introduced and its estimation methods with theoretical properties are discussed.

Although GARCH and RS GARCH models are able to capture the volatility, a number of studies showed usefulness of the macroeconomic and financial variables in modelling and forecasting volatility (Engle et al., 2013; Conrad and Loch, 2015; Conrad et al., 2018). The GARCH MIDAS model proposed by Engle et al. (2013) can directly incorporate these variables to model volatility. They showed that the inflation and industrial production rates are useful for forecasting long-horizon US stock return volatility. Moreover, Conrad and Loch (2015) also concluded that GARCH MIDAS is advantageous since it allows the incorporation of macroeconomic variables in low frequencies to forecast daily asset volatility.

A GARCH MIDAS model is a multiplicative two component model that describes the conditional variance. This model decomposes conditional volatility into two components, where the high-frequency, i.e. short-term component is modelled as a mean reverting unit daily GARCH process while the low-frequency, i.e. long-term component is determined by realised volatility and/or some macroeconomic variables. Let i be a day within a period t, whereby the index $t = 1, ..., T$ refers to a certain period such as a week or month. The return $r_{i,t}$ can be modelled as:

$$r_{i,t} = \mu + \sqrt{\tau_t \, h_{i,t}} \, \varepsilon_{i,t}, \qquad \forall i = 1, \ldots, N_t. \tag{3.45}$$

where μ is return mean, $\{\varepsilon_{i,t}\}$ is an i.i.d. innovation process with mean equal to zero and unit variance, and N_t is the number of trading days included in t. The expected return is assumed to be constant, i.e. $E(r_{i,t}|\mathcal{F}_{i-1,t}) = \mu$ for all i and t, where $\mathcal{F}_{i-1,t}$ is the information set in period t. The distribution of $\varepsilon_{i,t}$ is similar to what we discussed in GARCH model. Different distributions can be used while here we consider Normal and Student-t innovations. In Equation (3.45), $\delta^2_{i,t} = \tau_t h_{i,t}$ is the total conditional variance and the volatility is decomposed into two parts where τ_t captures the long-term volatility and $h_{i,t}$ describes the short-term fluctuations. The role of τ_t is to describe smooth movements in the conditional variance. We assume that $\tau_{i,t}$ is fixed for all i in period t, whereas $h_{i,t}$ varies daily, hence we can rewrite τ_t removing subscript i. Alternatively, one can specify τ_t to change daily similar to the short-term volatility component. However, the findings of Engle et al. (2008) show that both specifications of τ_t yield similar empirical fit, hence we assume that τ_t changes only once every N_t

days. The short-term volatility component is intended to describe day-to-day volatility and is assumed to follow a standard GARCH(1,1) process:

$$h_{i,t} = (1-\alpha-\beta) + \alpha\frac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta h_{i-1,t} \tag{3.46}$$

where it is assumed that $\alpha \geq 0, \beta \geq 0$ and $\alpha + \beta < 1$ to ensure that $h_{i,t}$ remains positive. The feature that distinguishes GARCH MIDAS model in Equation (3.46) from standard GARCH model is $\frac{(r_{i-1,t}-\mu)^2}{\tau_t}$. When $\tau_t = 1$, GARCH MIDAS model in Equation (3.46) is simplified to GARCH(1,1) model.

The MIDAS approach proposed by Ghysels et al. (2007) tackles the problems involved in using the data sampled at different frequencies within the same model. The innovation of Engle et al. (2013) lies in the definition of τ_t in the MIDAS term. The first specification of the τ_t for GARCH MIDAS model dates back to work of Schwert (1988) and others, where they measure long-run volatility by realised volatility over a different horizons. In particular, they consider monthly RV$_t$. However, τ_t here is specified is a weighted sum of K lags of smoothed RV over a long horizon and takes the following form:

$$\begin{aligned} \log\tau_t &= m + \theta\sum_{i=1}^{K}\phi_i(\kappa)RV_{t-i}, \\ RV_t &= \sum_{i=1}^{N_t} r_{i,t}^2, \end{aligned} \tag{3.47}$$

where m is constant, K is the number of lagged months, quarters or annuals, θ measures the impact of lags of r_t and RV stands for realised volatility. It is also important to note that the τ_t is predetermined $E_{t-1}[(r_{i,t}-\mu)^2] = \tau_t E_{t-1}(h_{i,t}) = \tau_t$ assuming that $E_{t-1}(h_{i,t}) = 1$. Furthermore, it can be modified to involve the macroeconomic variables along with RV to see the impact of these variables on the long-term volatility:

$$\log\tau_t = m + \left(\theta_1\sum_{i=1}^{K}\phi_i(\kappa_1)RV_{t-i} + \theta_2\sum_{i=1}^{K}\phi_i(\kappa_2)X_{t-i}\right). \tag{3.48}$$

The variable X_t refers to one of the macroeconomic variables. Taking logarithm ensures non-negativity of the long-term component even when the explanatory variables take negative values. When $\theta_1 = 0$, the realised volatility has no influence on long-term volatility, similarly when $\theta_2 = 0$ the macroeconomic variable has no effect on volatility. Moreover, when $\theta_1 = \theta_2 = 0$, both realised variance and macroeconomic factors have no impact on volatility, meaning that all volatility is captured by the short-term component and the model becomes GARCH(1,1) model with $\tau_t = 1$. In this case, standard GARCH model is nested in the GARCH MIDAS model.

The $\phi_i(\kappa)$ in Equation (3.47) are the weighting schemes which are based on the exponential or the beta-weight specification. The most commonly used weighting schemes as suggested by Engle et al. (2013) are:

$$\phi_i(\kappa) = \begin{cases} \dfrac{\left(\frac{i}{K}\right)^{\kappa_1-1}\left(1-\frac{i}{K}\right)^{\kappa_2-1}}{\sum\limits_{j=1}^{K}\left(\frac{j}{K}\right)^{\kappa_1-1}\left(1-\frac{j}{K}\right)^{\kappa_2-1}} & \text{Beta weights,} \\[2em] \dfrac{\kappa^i}{\sum\limits_{j=1}^{K}(\kappa)^j} & \text{Exponential weights.} \end{cases} \tag{3.49}$$

These specifications involve two parameters $\kappa \equiv (\kappa_1, \kappa_2)$, however the Beta weighting scheme can be restricted to a one parameter by setting $\kappa_1 = 1$. Such single parameter specifications are often used in practice (Ghysels et al., 2007; Pan et al., 2017) and existing literatures have shown that both methods yield similar results (Engle et al., 2008). Furthermore, the weights capture the effect of past fundamental information on return volatility and can be freely estimated or fixed before estimation. Following work of Pan et al. (2017) we use a single parameter beta weighted specification which can be written as:

$$\phi_i(\kappa_d) = \frac{(1-\frac{i}{K+1})^{\kappa_d-1}}{\sum\limits_{j=1}^{K}(1-\dfrac{j}{K+1})^{\kappa_d-1}}, \qquad d = 1, 2, \tag{3.50}$$

where the weights $\phi_i(\kappa_d)$ sum up to 1. The formulation of the weighting function requires only two parameters κ_1, κ_2 and guarantees that all weights are non-negative. For the choice of K, Conrad and Loch (2015) showed that as long as the selected K

is large enough, the estimation results are robust with respect to the specific choice of the maximum number of lags included.

3.3.1 Model estimation

The estimation of the GARCH MIDAS model is done via QMLE following the identification of long-term component in Equation (3.47) and specification of innovations. The estimation steps are similar to GARCH model since the difference between these two models is the long-term volatility parameter τ_t in conditional variance. Consistency and asymptotic normality of the QML estimator for GARCH MIDAS model was established by Wang and Ghysels (2015) but only for a special case with RV as the explanatory variable.

To write the likelihood of the GARCH MIDAS model a distribution for the i.i.d. variables $\{\varepsilon_t\}$ must be specified as well as the specification for the long-term component. For the GARCH MIDAS model with rolling window RV the parameter set is $\boldsymbol{\theta} = (\alpha, \beta, m, \theta, \kappa)'$ whereas with a rolling window RV and a macroeconomic variable the parameter set is $\boldsymbol{\theta} = (\alpha, \beta, m, \theta_1, \theta_2, \kappa_1, \kappa_2)'$ The conditional Gaussian quasi-likelihood is given by:

$$\mathcal{L}_n(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}; r_1, ..., r_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi \tilde{h}_{i,t} \tilde{\tau}_t}} \exp\Big(- \frac{(r_t - \mu)^2}{2\tilde{h}_{i,t}\tilde{\tau}_t} \Big), \tag{3.51}$$

where $\tilde{h}_{i,t}$ and $\tilde{\tau}_t$ are defined recursively, for $t \geq 1$ by

$$\begin{aligned} \tilde{h}_{i,t} &= \tilde{h}_{i,t}(\boldsymbol{\theta}) = (1 - \alpha - \beta) + \frac{\alpha(r_{i-1,t} - \mu)^2}{\tilde{\tau}_{t-1}} + \beta \tilde{h}_{i-1,t}, \\ \tilde{\tau}_t &= \tilde{\tau}_t(\boldsymbol{\theta}) = m + \theta \sum_{i=1}^{K} \phi_i(\kappa) RV_{t-i}, \end{aligned} \tag{3.52}$$

The likelihood function:

$$\ell_T(\boldsymbol{\theta}) = \log \mathcal{L}_T(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \log(2\pi) + \log(\tilde{\tau}_t \tilde{h}_{i,t}) + \frac{(r_{i,t} - \mu)^2}{(\tilde{\tau}_t \tilde{h}_{i,t})}, \tag{3.53}$$

is then maximised to obtain parameter estimates.

If the long-term component of a GARCH MIDAS model with Student-t innovations is described by RV and a macroeconomic variable, then the parameters set is $\boldsymbol{\theta} = (\alpha, \beta, m, \theta_1, \theta_2, \kappa_1, \kappa_2, v)'$. The constraint $v > 2$ is imposed to ensure that the second order moment exist. QMLE for GARCH MIDAS Student-t is obtained similarly as for GARCH MIDAS Normal distribution.

3.3.2 Out-of-sample forecasting

While standard GARCH models accurately forecast the short-term return volatility (Andersen and Bollerslev, 1998), empirical evidence suggests that using models such as GARCH MIDAS model which include explanatory variables have been successful at forecasting longer horizons (Engle et al., 2013; Conrad and Loch, 2015).

To forecast the return volatility using the identified GARCH MIDAS model with macroeconomic variables assume that the volatility is estimated on the last day of period t. The one-step ahead volatility forecast for GARCH MIDAS can be written as:

$$\begin{aligned}
\sigma^2_{i,t+1} &= h_{i,t+1} \times \tau_{t+1}, \\
h_{i,t+1} &= (1-\alpha-\beta) + \alpha \frac{(r_{i,t}-\mu)^2}{\tau_{t+1}} + \beta h_{i,t}, \\
\log \tau_{t+1} &= m + \theta \sum_{i=1}^{K} \phi_i(\kappa) RV_{t-i},
\end{aligned} \tag{3.54}$$

where τ_{t+1} is pre-determined at point t and it is assumed that τ_{t+1} is constant across all days within period t and thus changes only when the month changes.

For further horizons, the short-term forecasts $h_{i,t}$ can be obtained iteratively from GARCH part given sub-sample parameter estimates. We use the estimated τ_t from Equation (3.47) and Equation (3.48) as the prediction of the long-term variance. Since at the beginning of period t, the long-term volatility component is predetermined with respect to $\mathcal{F}_{t-1}$, the volatility forecast for a specific day i in period t is given by:

$$E[h_{i,t}\tau_t\varepsilon^2_{i,t}|\mathcal{F}_{t-1}] = \tau_t E[h_{i,t}|\mathcal{F}_{t-1}], \tag{3.55}$$

where

$$E[h_{i,t}|\mathcal{F}_{t-1}] = 1 + (\alpha+\beta)^{i-1}(h_{1,t}-1), \tag{3.56}$$

and $\mathcal{F}_{t-1}$ is the information set available in period $t-1$. As the forecast horizon tends to infinity, $E[h_{i,t}|\mathcal{F}_{t-1}]$ converges to unity, i.e. to its (constant) unconditional variance of given in Equation (3.23) hence in the long-run the GARCH MIDAS forecasts are entirely driven by the long-term components. The volatility forecast forecast for period t is then given by:

$$E[h_{i,t}\tau_t\varepsilon_{i,t}^2|\mathcal{F}_{t-1}] = \tau_t\left(N_{(t)} + (g_{1,t}-1)\frac{1-(\alpha+\beta)^{N_t}}{1-\alpha-\beta}\right). \tag{3.57}$$

For a more than one-period-ahead prediction, we need to forecast the τ_t itself. For longer horizons Conrad and Loch (2015) assumed that the long-term component is $\hat{\tau}_{t+s|t-1} = \hat{\tau}_{t|t-1}$ for $s > 0$ which means that it remains at the level of the one-step prediction. Daily volatility forecasts are then calculated as the product of GARCH and the long-term component forecasts.

3.4 The RS GARCH MIDAS model

It is reasonable to assume that the conditional volatility of a return may switch across regimes because of structural breaks (Cai, 1994). Hence combining Markov-switching with GARCH MIDAS, a RS GARCH MIDAS model is developed. Pan et al. (2017) applied a RS GARCH MIDAS model to forecast crude oil volatility by allowing the short-term volatility component to switch. In their paper they assumed that the innovations follow a Normal distribution. However, since returns exhibit heavy tails, to capture the fat tails situation we consider a RS GARCH MIDAS model with Student-t distribution to see whether this model is effective in describing volatility displayed in financial time series.

In order to allow for regime switches in the short-term volatility component it is assumed the $h_{i,t}$ in Equation (3.46), depends on the latent and unobservable state $s_{i,t}$. This implies that RS GARCH MIDAS model now takes the form:

$$h_{s_{i,t}} = \omega_{s_{i,t}} + \alpha_{s_{i,t}}\frac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta_{s_{i,t}}\bar{h}_{s_{i-1,t}}, \tag{3.58}$$

where the $s_{i,t}$ indicates the states of the market and is governed by a first order ergodic homogeneous Markov Chain with the transition probability matrix

$$Pr(s_{i,t} = m | s_{i-1,t} = l) = p_{lm}, \tag{3.59}$$

where p_{lm}[3] implies that the current state $s_{i,t}$ depends only on the prior state $s_{i-1,t}$. The general transition matrix P is given by:

$$P = \begin{bmatrix} p_{11} & \dots & p_{1K} \\ \vdots & \ddots & \vdots \\ p_{K1} & \dots & p_{KK} \end{bmatrix}.$$

The simplest case of RS GARCH MIDAS model with two-states can be written as follows:

$$\begin{aligned} r_{i,t} &= \mu + \sqrt{h_{i,t}\tau_t}\varepsilon_{i,t} \\ h_{i,t} &= \begin{cases} \omega_1 + \alpha_1 \frac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta_1 h_{i-1,t} & \text{when} \quad s_t = 1, \\ \omega_2 + \alpha_2 \frac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta_2 h_{i-1,t} & \text{when} \quad s_t = 2, \end{cases} \end{aligned} \tag{3.60}$$

where $\{\varepsilon_{i,t}\}$ is a sequence of i.i.d. random variables with mean zero and unit variance. The necessary conditions to ensure that conditional variance remains positive in each regime as well as the two-state transition matrix are same as in RS GARCH$(1,1)$ model given Equation (3.27).

Coming to the long-term component we assume that τ_t is fixed for all i in period t, because of two reasons. First, Engle et al. (2008) and Wang et al. (2022) showed that daily and monthly switching τ_t yield similar empirical fit. Second, we did consider incorporating RS in the long-term component, however, the results showed insignificance of the long-term component switching, which also coincides with work of Ma et al. (2021) who showed that the in-sample estimation results for the transition probabilities in the long-term component are small and not statistically significant. Therefore, τ_t with RV can be written as in Equation (3.47) and τ_t with RV and macroeconomic variable, X_t, takes the form in Equation (3.48). The parameter weighting scheme is exactly same as in GARCH MIDAS defined in Equation (3.50). The conditional distribution

[3] The transition probability becomes $Pr(s_{i,t} = m | s_{N_{t-1},t-1} = l)$ on the first day of the new month.

is assumed to follow either standard normal or standardised Student-t distribution.

3.4.1 Model estimation

Here we focus on the model in Equation (3.60). The parameters can be estimated using the QMLE with the filtering probabilities obtained by Hamilton's filter. We present the estimation procedure of this model in detail. To the best of our knowledge, asymptotic results for the RS GARCH MIDAS model are not yet available. However, we evaluate the performance of QMLE using a Monte Carlo simulation.

As in RS GARCH model RS GARCH MIDAS model also suffers from the issue of path dependence due to a presence of $h_{i-1,t}$ in the short-term volatility process in Equation (3.60). By adopting Klaassen (2002) approach and following similar steps as in RS GARCH, we can rewrite $h_{i-1,t}$ as:

$$h_{s_t,i,t} = \omega_{s_t,t} + \alpha_{s_t,t}\frac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta_{s_t,t}E_{i-1,t}\left\{h_{i-1,t}|s_{i,t}\right\}, \tag{3.61}$$

where the expected conditional variance is calculated as:

$$E_{i-1,t}[h_{i-1,t}|s_{i,t}] = E_{i-1,t}[h_{i-1,t}|s_{i,t}=i] = \sum_{j=1}^{2}\tilde{p}_{ji,(i-1,t)} \times h_{s_t,t-1}, \tag{3.62}$$

with the probabilities $\tilde{p}_{ji,(i-1,t)}$ given by:

$$\begin{aligned}\tilde{p}_{ji,(i-1,t)} &= Pr(s_{i-1,t}=j|s_{i,t}=i,\mathcal{F}_{i-1,t})\\ &= \frac{Pr(s_{i,t}=i|s_{i-1,t}=j)Pr(s_{i-1,t}=j|\mathcal{F}_{i-1,t})}{Pr(s_{i,t}=i|\mathcal{F}_{i-1,t})}\\ &= \frac{p_{ji}p_{j,(i-1,t)}}{p_{i,t}},\end{aligned} \tag{3.63}$$

where $i,j=1,2$ and $s_t=\{1,2\}$.

The log-likelihood function can be derived in a similar way to RS GARCH model where we slightly modify it to include the long-term component. For a two-state RS GARCH MIDAS model in Equation (3.60) the density of returns conditional on $\mathcal{F}_{i-1,t}$ is given by:

$$
\begin{aligned}
f(r_{i,t}|\mathcal{F}_{i-1,t};\boldsymbol{\theta}) &= \sum_{j=1}^{2} f(r_{i,t}, s_t = j|\mathcal{F}_{i-1,t};\boldsymbol{\theta}) \\
&= \sum_{j=1}^{2} f(r_{i,t}|s_t = j, \mathcal{F}_{i-1,t};\boldsymbol{\theta}) Pr(s_t = j|\mathcal{F}_{i-1,t};\boldsymbol{\theta}).
\end{aligned}
\tag{3.64}
$$

Under the assumption of Gaussian innovations:

$$
f(r_{i,t}|s_t = j, \mathcal{F}_{i-1,t};\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi h_{i,t}\tau_t}} \exp\Big(-\frac{(r_{i,t}-\mu)^2}{2h_{i,t}\tau_t}\Big). \tag{3.65}
$$

Under the assumption of Student-t innovations::

$$
f(r_{i,t}|s_t = j, \mathcal{F}_{i-1,t};\boldsymbol{\theta}) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{(v-2)\pi h_{i,t}\tau_t}\Gamma(\frac{v}{2})} \left[1 + \frac{(r_{i,t}-\mu)^2}{(v-2)h_{i,t}\tau_t}\right]^{-\frac{v+1}{2}}. \tag{3.66}
$$

Applying Hamilton's filter, the predicted probabilities are written as:

$$
Pr(s_t = j|\mathcal{F}_{i-1,t};\boldsymbol{\theta}) = \sum_{i=1}^{2} Pr(s_t = j|s_{t-1} = i) Pr(s_{t-1} = i|\mathcal{F}_{i-1,t};\boldsymbol{\theta}), \tag{3.67}
$$

where $Pr(s_t = j|s_{t-1} = i)$ are the transition probabilities. Applying the Bayes theorem and the law of total probability, the filtered probabilities can be recursively computed as:

$$
\begin{aligned}
Pr(s_{t-1} = i|\mathcal{F}_{i-1,t};\boldsymbol{\theta}) &\equiv Pr(s_{t-1} = i|t_{i-1,t}, \mathcal{F}_{i-2,t};\boldsymbol{\theta}) \\
&= \frac{f(r_{i-1,t}|s_{t-1} = i, \mathcal{F}_{i-2,t};\boldsymbol{\theta}) Pr(s_{t-1} = i|\mathcal{F}_{i-2,t};\boldsymbol{\theta})}{\sum_{i=1}^{2} f(r_{i-1,t}|s_{t-1} = i, \mathcal{F}_{i-2,t};\boldsymbol{\theta}) Pr(s_{t-1} = i|\mathcal{F}_{i-2,t};\boldsymbol{\theta})}.
\end{aligned}
\tag{3.68}
$$

Hence, the log-likelihood function is obtained as:

$$
\begin{aligned}
\mathcal{L}_n(\boldsymbol{\theta}) &= \sum_{t=1}^{n} \log[f(r_{i,t}|s_t = j, \mathcal{F}_{i-1,t};\boldsymbol{\theta})] \\
&= \sum_{t=1}^{n} \log\left[p_{1,t} f(r_{i,t}|s_t = 1, \mathcal{F}_{i-1,t};\boldsymbol{\theta}) + p_{2,t} f(r_{i,t}|s_t = 2, \mathcal{F}_{i-1,t};\boldsymbol{\theta})\right],
\end{aligned}
\tag{3.69}
$$

where the unknown parameter set is $\boldsymbol{\theta} = (\omega_0, \omega_1, \alpha, \beta, m, \theta_0, \theta_1, \kappa_0, \kappa_1, p_{00}, p_{11})$. The QML estimates are obtained by maximising the log-likelihood in Equation (3.69) where the conditional densities for Gaussian innovations are specified in Equation (3.65) and for Student-t in Equation (3.66). Similar to RS GARCH model, to avoid the path dependence problem caused by $h_{i-1,t}$, Klaassen's approach, Equations (3.61) - (3.63) are implemented. The iterative steps to obtain Hamilton's filter probabilities are given in more details in Figure 3.2.

3.4.2 Out-of-sample forecasting

The main step for generating out-of-sample volatility forecasts for RS GARCH MIDAS is the calculation of the short-term volatility component, which is done in a similar way to RS GARCH model. The forecasting process for the long-term volatility component is the same as in single-regime GARCH MIDAS and it is predetermined by Equation (3.47) and Equation (3.48). We calculate one-step-ahead short-term volatility forecast as the weighted average of volatility forecast in two regimes:

$$\begin{aligned} E[h_{i+1,t}|\mathcal{F}_{i,t}] &= h_{i+1,t} \\ &= P(s_{i+1,t}=1|\mathcal{F}_{i,t}) \times h_{1,i+1,t} + P(s_{i+1,t}=2|\mathcal{F}_{i,t}) \times h_{2,i+1,t}, \end{aligned} \tag{3.70}$$

where

$$\begin{aligned} P(s_{i+1,t}=1|\mathcal{F}_{i,t}) &= p_{11} \times Pr(s_{i,t}=1|\mathcal{F}_{i,t}) + p_{21} \times Pr(s_{i,t}=2|\mathcal{F}_{i,t}), \\ P(s_{i+1,t}=2|\mathcal{F}_{i,t}) &= p_{12} \times Pr(s_{i,t}=1|\mathcal{F}_{i,t}) + p_{22} \times P(s_{i,t}=2|\mathcal{F}_{i,t}). \end{aligned} \tag{3.71}$$

The forecast of total volatility is defined as the product of short-term and long-term volatility forecasts components. The l-step ahead forecast within the same period t is written as:

$$\begin{aligned} E[h_{i+l,t}\tau_t|\mathcal{F}_{i,t}] &= \tau_t E[h_{i+l,t}|\mathcal{F}_{i,t}] \\ &= \tau_t \left[Pr(s_{i+l,t}=1|\mathcal{F}_{i,t}) E[h_{1,i+l,t}|\mathcal{F}_{i,t}] + Pr(s_{i+l,t}=2|\mathcal{F}_{i,t}) E[h_{2,i+l,t}|\mathcal{F}_{i,t}] \right] \end{aligned} . \tag{3.72}$$

For longer horizons it is common to assume that $\tau_{t+l|t} = \tau_{t+1|t}$ for $s > 1$, see Conrad and Kleen (2020).

Figure 3.2: Flow-chart for recursive calculation of Hamilton's filter probabilities for RS GARCH MIDAS.

$$Pr(s_{i-1,t} = i \mid \mathcal{F}_{i-1,t};\ \Theta)$$

$$Pr(s_{i,t} = j,\ s_{i-1,t} = i \mid \mathcal{F}_{i-1,t};\ \Theta) = Pr(s_{i-1,t} = i \mid \mathcal{F}_{i-1,t}; \Theta) \times Pr(s_{i,t} = j \mid s_{i-1,t} = i)$$

$$Pr(s_{i,t} = j \mid \mathcal{F}_{i-1,t}; \Theta) = \sum_{s_{i-1,t}=1}^{2} Pr(s_{i,t} = j,\ s_{i-1,t} = i \mid \mathcal{F}_{i-1,t};\ \Theta)$$

$$r_{i,t} = \mu + \sqrt{\tau_t\ h_{i,t}} \quad \text{where} \quad \tau_t = m + \exp\left(\theta_0 \sum_{i=1}^{K} \phi_i(\kappa_0) RV_{t-i} + \theta_1 \sum_{i=1}^{K} \phi_i(\kappa_1) X_{t-i}\right)$$

$$h_{i,t}^{(j)} = \omega^{(j)} + \frac{\alpha(r_{i-1,t}-\mu)^2}{\tau_t} + \beta E_{i-1,t}\left\{h_{i-1,t}^{(j)} \mid s_{i,t}\right\}$$

$$E_{i-1,t}\left\{h_{i-1,t}^{(j)} \mid s_{i,t}\right\} = \sum_{m=1}^{2} \tilde{p}_{ij,(i-1,t)} \times h_{i-1,t}^{(m)}$$

$$Pr(s_{i+1,t} = j \mid r_{i-1,t}) = \sum_{s_{i,t}=1}^{2} p(s_{i,t} = i \mid \mathcal{F}_{i-1,t};\ \Theta) \times\ Pr(s_{i+1,t} = j \mid s_{i,t} = i)$$

$$\tilde{p}_{ij,\ i,t} = Pr(s_{i,t} = i \mid s_{i+1,t} = j,\ r_{i-1,t}) = \frac{p_{ij} Pr(s_{i,t}=i \mid r_{i-1,t})}{Pr(s_{i+1,t}=j \mid r_{i-1,t})} = \frac{p_{ij} p_{i,t}}{p_{j,t+1}}$$

$$f(r_{i,t},\ s_{i,t} = j \mid \mathcal{F}_{i-1,t};\ \Theta) = f(r_{i,t} \mid s_{i,t} = j, \mathcal{F}_{i-1,t};\ \Theta) \times Pr(s_{i,t} = j \mid r_{i-1,t},\ \mathcal{F}_{i-1,t})$$

$$f(r_{i,t} \mid \mathcal{F}_{i-1,t};\ \Theta) = \sum_{s_{i,t}=1}^{2} f(r_{i,t},\ s_{i,t} = j \mid \mathcal{F}_{i-1,t};\ \Theta)$$

$$Pr(s_{i,t} = j \mid r_{i,t}; \Theta) = \frac{f(r_{i,t},\ s_{i,t}=j \mid \mathcal{F}_{i-1,t};\ \Theta)}{f(r_{i,t} \mid \mathcal{F}_{i-1,t};\ \Theta)}$$

$i, t = i+1, t$

Notes: This flow-chart shows how to derive Hamilton's filter probabilities and conditional log-likelihood. The $\phi_i(\kappa_0)$ and $\phi_i(\kappa_1)$ is the weighting scheme given in Equation (3.50). The R Code to implement this flow chart is given in Appendix B.

3.5 The Endogenous RS GARCH models

In all of the above mentioned regime switching models, the evolution of s_t is independent of r_{t-1}, meaning that p_{11}, p_{22} are constants and the switching of regimes is exogenous. However, this assumption sometimes needs to be modified to enrich the dynamics of the model by introducing endogenous regime switching. Moreover, studies including Kim (2004), Choi (2009) and Kim et al. (2008) reported evidence of endogeneity in regime changes. Hence two extra models namely Endogenous RS GARCH(1,1), denoted as Endo RS GARCH(1,1), and Endogenous RS GARCH MIDAS, denoted as Endo RS GARCH MIDAS are considered. Following, Choi (2009) we specify p_{11} and p_{22} as a function of r_{t-1}, so that the transition probabilities are time-varying and are driven by past observation. The specification of this model is similar to RS GARCH model in Equation (3.24) and RS GARCH MIDAS in Equation (3.58) with the only difference in transition probabilities.

Consider the following specification for the endogenous transition probabilities:

$$p_{11} = \frac{\exp(c_1 + \gamma_1 \ r_{t-1})}{1 + \exp(c_1 + \gamma_1 \ r_{t-1})}, \qquad p_{22} = \frac{\exp(c_2 + \gamma_2 \ r_{t-1})}{1 + \exp(c_2 + \gamma_2 \ r_{t-1})}. \tag{3.73}$$

Notably, the time-varying probabilities in Equation (3.73) can be reduced to constant transition probabilities when the parameters γ_1 and γ_2 are set to equal to zero. An advantage of modelling endogeneity this way is that we can directly specify the time-varying transition probabilities as function of lagged values of r_t.

The parameters can be estimated using QMLE with a slight modification in filtering probabilities obtained by Hamilton's filter and the genetic algorithm of Kim (1994). For Endo RS GARCH(1,1) with two-regimes the conditional density is similar to Equation (3.34). The predicted probabilities can be computed as in Equation (3.36) where $Pr(s_t = j|s_{t-1} = i)$ is given in Equation (3.73), while the filtered probabilities are computed recursively using Hamilton's filter as in Equation (3.38). Then the log-likelihood function is similar to Equation (3.39) for Endo RS GARCH and Equation (3.69) for Endo RS GARCH MIDAS. The parameter set is: $\boldsymbol{\theta} = \{\omega_{s_t}, \alpha_{s_t}, \beta_{s_t}, c_1, c_2, \gamma_1, \gamma_2\}$, while the parameter set for Endo-RS GARCH MIDAS will have five extra parameters $\{m, \theta_1, \theta_2, \kappa_1, \kappa_2\}$ and v for Student-t distribution.

3.6 Forecast evaluation

In this section we are interested in evaluating the performances of the forecasting models discussed up to now. The out-of-sample predictive ability is important in the field of forecasting volatility, hence various types of measures are employed to evaluate the volatility prediction of a specific model.

When estimating and forecasting volatility models, we face a problem that the true volatility is unobserved, hence we need a good volatility proxy for it. The volatility proxy aims to represent the true, unobservable, volatility process which the generated forecasts will be evaluated against. Various measures have been suggested as proxy variables for volatility in financial markets, some of which are the squared and absolute returns, realized volatility, i.e., the sum of squared intraday returns. For example, see Hansen and Lunde (2005); Ghysels et al. (2006); Triacca (2007); Giles (2008); Patton (2006); Pan et al. (2017). In the present context, our aim is to compare the relative predictive accuracy of various models, and if the loss function is quadratic, the use of squared returns ensures that we actually obtain the correct ranking of models (Awartani and Corradi, 2005).

One way to evaluate how well the model fits the dataset is to use loss functions. The types of loss function are: Mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), root mean absolute error (RMAE), Gaussian quasi-likelihood (QLIKE) and etc. While there are many loss functions in the literature, Patton (2006) proved that only certain loss functions are robust to noise in the volatility proxy. Here we only consider two commonly used loss functions MSE and QLIKE, because they are more robust to the imperfect volatility proxies (Patton, 2006), which means that using a volatility proxy yields the same ranking as using the unobserved volatility. This also coincides with the findings of Conrad and Kleen (2020). Furthermore, Patton (2006) showed that QLIKE loss function provides consistent ranking because of the lower impact of the most extreme observations in the sample.

The MSE and QLIKE is defined as:

$$\begin{aligned} MSE &= \frac{1}{T}\sum_{i=1}^{T}(\sigma_i^2 - \hat{\sigma}_i^2)^2, \\ QLIKE &= \frac{1}{T}\sum_{i=1}^{T}\left(\log(\hat{\sigma}_i^2) + \frac{\sigma_i^2}{\hat{\sigma}_i^2}\right), \end{aligned} \tag{3.74}$$

where σ_i^2 is the volatility proxy for the out-of-sample forecasting, $\hat{\sigma}_i^2$ is the forecast value of volatility obtained by the proposed models and T is the total number of volatility forecasts. The forecast model with the lowest MSE and QLIKE loss coefficient is preferred. Nevertheless, the loss functions do not tell us whether the differences between the forecasts are statistically significant. To make meaningful inferences about predictive ability of the forecasts we need suitable significance tests.

In addition, researchers are often interested in determining which of the two competing forecasting models predicts the volatility best. Such tests are known as relative forecast performance tests. Examples of the relative forecast performance tests include Diebold and Mariano (1995) and West (1996). Diebold-Mariano (DM) test provides information about whether the difference in forecast performance is statistically significant or not, i.e., it compares if the two forecasts have equal predictive accuracy. For this test, let $\hat{r}_i$ and $\hat{r}_j$ denote two sequences of forecasts of r_i, generated from two competing models i and j. Next, let b_i and c_j be the residuals for the two forecasts calculated as:

$$\begin{aligned} b_i &= r_i - \hat{r}_i, \\ c_i &= r_i - \hat{r}_j. \end{aligned} \tag{3.75}$$

The loss differential d_i is then obtained by:

$$d_i = b_i - c_i, \tag{3.76}$$

to calculate the relative performance between the models. Under the null hypothesis of equal predictive accuracy $E(d_i) = 0$ for all t. The only assumption required by DM test is for the loss differential to be covariance stationary. The test statistic for DM test can then be formulated as:

$$DM = \frac{\bar{d}}{\sqrt{\frac{[\gamma_0 + 2\sum_{k=1}^{h-1} \gamma_k]}{n}}}, \tag{3.77}$$

where γ_k is given as:

$$\gamma_k = \frac{1}{n} \sum_{i=k+1}^{n} (d_i - \bar{d})(d_{i-k} - \bar{d}), \tag{3.78}$$

and $\bar{d}$ is calculated as:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i, \tag{3.79}$$

where γ_k is the auto-covariance at lag k and $DM \sim N(0,1)$. The null hypothesis of the DM test is that the two forecasts have the same level of accuracy.

Chapter 4

Simulation

In this chapter we evaluate the quasi-maximum likelihood estimator of RS GARCH MIDAS model in a Monte Carlo simulation. The consistency and asymptotic theory for the QMLE for RS GARCH, (Xie, 2009; Bauwens et al., 2010) and for a special case of GARCH MIDAS (Wang and Ghysels, 2015) with realised volatility as the explanatory variable have been established. However, this is not yet available for regime switching GARCH MIDAS models. Therefore, we first evaluate the finite-sample performance of QMLE in a Monte Carlo simulation to show that QMLE is unbiased and the asymptotic standard errors are generally applicable.

Since correctly specifying and modelling volatility has important implications, we also examine the potential impact of misspecification. Therefore, we compare the QML estimates of the misspecified models against the QML estimates of correctly specified model to see how the misspecification affects the results. We assume the misspecification in terms of: (a) not considering regime switching, (b) misspecifying the error term, (c) omitting the long-term volatility component, (d) a combination of all three.

4.1 Data generating process

Since our empirical application focuses on the RS GARCH MIDAS model with Student-t innovations, we simulate 1,000 replications from the following data generating process

(GDP):

$$r_{i,t} = \sqrt{h_{i,t}\tau_t}\varepsilon_{i,t}, \tag{4.1}$$

where $\{\varepsilon_{i,t}\}$ are assumed to be i.i.d. and follow a standardised Student-t distribution with 6.6087 degrees of freedom. We generate daily data for the period of 340 months where we assume that each month consists of $N_t = 22$ days. This gives us a total of 7480 daily observations, which is approximately the same length as our empirical dataset. To simulate the regimes we first generate a random number from uniform distribution between 0 and 1. If the random number is less than the specified probability, we set $s_t = 0$, otherwise 1.

From simulated daily returns the monthly realised volatility is calculated as the sum of absolute daily returns, $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$, as a measure of the monthly variance. The reason behind using absolute returns rather than squared returns is discussed in Chapter 5. The parameters of the short-term volatility, i.e., the GARCH specification, $h_{i,t}$, are set to be $\omega_1 = 0.1785$, $\omega_2 = 0.4044$, $\alpha = 0.0944$ and $\beta = 0.7241$. In addition, the state variables are assumed to be $p_{11} = 0.9961$ and $p_{22} = 0.9986$.

Now, considering the MIDAS part, we set the long-term volatility component, τ_t, to be fluctuating at a monthly frequency. We also assume that the long-term component is driven by the dynamics of realised volatility without any additional macroeconomic variables. The specification for the τ_t is set to be exactly same as in Equation (3.47) where $m = 0$, $\theta = 0.0262$ and we choose the MIDAS weights to be specified according to the Beta weighting scheme in Equation (3.50) with $\kappa = 4.4037$ and $K = 8$. Setting $\kappa = 4.4037$ implies a monotonically decaying weights, meaning that the lower weights are given to the most recent observations. The choice of $K = 8$ months as a lag length, follows our empirical application where we estimated the model with various lag lengths beginning from 6 months to 12 month and stopped at 8 month when the improvement in the log-likelihood was negligible.

These parameters are also reported in the first column of Table 4.1. The reason we set these parameters to be equal to the values given is because they are the QML estimates from our empirical application, Table 5.3, which we discuss in more details in the next chapter. Furthermore, it should be noted that taking lags in model estimates leads to a loss of observations, as a result both the short and long-term volatilities start from month 9. Moreover, we set the first 8 months of the RV to be equal to the realised volatility obtained from our empirical application of WTI crude oil return data.

4.2 Parameter estimates

In this section, using the full sample we obtain QML estimates of the model parameters and investigate the effect of model misspecification. Tables 4.1 and 4.2 reports the average of the QML estimates and average bias across 1,000 Monte Carlo simulations, respectively. Based on these replication we also calculate the average asymptotic standard errors. The asymptotic standard errors of the estimates are obtained using the Hessian matrix. It is possible to estimate the standard errors of parameters by calculating the square root of the inverse of the diagonal. The equation for obtaining these errors is written as:

$$\text{s.e. } (\hat{\boldsymbol{\theta}}) = \sqrt{diag\left(\frac{\hat{J}^{-1}\hat{I}\hat{J}^{-1}}{T}\right)}, \tag{4.2}$$

where J and I are the expected Hessian and Information matrix respectively and *diag* denotes the diagonal elements of the matrix.

The average parameter estimates along with their corresponding standard errors in brackets, and empirical standard deviations in curly brackets are reported in Table 4.1. For a better observation we also provide average bias of the estimates in Table 4.2. Let GARCH-N and GARCH-t denote GARCH(1,1) model where N and t indicate normal or Student-t distributions, respectively. Other models are denoted similarly.

Starting with Table 4.2 panel B, since this is where we correctly specified the density we observe that the average bias for most of the parameters is close to zero except for κ, v and β in GARCH and RS GARCH along with ω_2 in RS GARCH. For these parameters, except for v which displays a negative bias in two models, we clearly see a positive bias. Conrad and Kleen (2020) also observed that κ is positively biased in their simulation.

Furthermore, by comparing empirical standard deviation with asymptotic standard errors we can see that nearly all of the asymptotic standard errors in the true model are close to the empirical standard deviation of the estimated parameters except for the κ parameter, where it appears that the asymptotic standard error is too big. On the other hand, in the misspecified RS GARCH-t model some of the asymptotic standard errors appear to be too small compared to the empirical standard deviation. Overall, most of the asymptotic standard errors are quite close to the empirical standard deviation, hence we can say that the performance of asymptotic standard errors is satisfying.

Next, we investigate the effect of model misspecification, where long-term component and regime switching is omitted and look at the bias of single-component GARCH and GARCH MIDAS models. From a quick inspection, we observe that the average bias for GARCH and GARCH MIDAS models in both innovations are quite similar to each other, except for the κ in GARCH MIDAS. Moreover, it appears that in GARCH models parameter β is positively biased. Andreou and Ghysels (2002) and Mikosch and Stărică (2004) showed that an upward bias in the degree of persistence in GARCH models can be caused by failure to accommodate structural changes in the model. The parameter κ in GARCH MIDAS models is also positively biased. Additionally, from Table 4.1 we can see that θ is nearly twice as large as the true parameter. These results indicate that changing the innovation in single component GARCH and GARCH MIDAS models hardly affects the parameters. These findings are also consistent with work of Conrad and Kleen (2020). On the other hand, omitting regime switching in GARCH MIDAS models in both innovations causes θ to be larger than the true parameter. This large value of θ causes the scale of long-term volatility component to be large as well. A possible explanation for such effect, is that a larger θ in low-frequency component makes it possible to capture extreme observations. The effect of a large scale of long-term component will be shown at the end of the Section.

Furthermore, according to $\alpha + \beta$ in Table 4.1 the volatility persistence in GARCH models is close to 1 whereas it is much lower in GARCH MIDAS. The reason for this might be that accounting for long-term volatility can reduce the persistence in the short-term volatility component.

Now, by replacing the Student-t errors with normal innovations we misspecify the

Table 4.1: Monte Carlo simulation parameter estimates.

		Panel A: Gaussian				Panel B: Student-t			
Par	DGP	GARCH	RS GARCH	GARCH MIDAS	RS GARCH MIDAS	GARCH	RS GARCH	GARCH MIDAS	RS GARCH MIDAS
ω_1	0.1785	-	0.1670	-	0.0907	-	0.2023	-	0.2014
			(0.1314)		(0.0866)		(0.0599)		(0.0449)
			{0.2447}		{0.1207}		{0.1481}		{0.0453}
ω_2	0.4044	-	4.1245	-	1.8923	-	0.6115	-	0.4193
			(0.7178)		(0.3595)		(0.1988)		(0.1031)
			{3.5815}		{0.9815}		{0.2876}		{0.1182}
θ	0.0262	-	-	0.0424	0.0394	-	-	0.0425	0.0245
				(0.0008)	(0.0027)			(0.0011)	(0.0037)
				{0.0008}	{0.0057}			{0.0009}	{0.0037}
α	0.0944	0.0392	0.0598	0.1044	0.0460	0.0400	0.0794	0.1068	0.0965
		(0.0025)	(0.0125)	(0.0107)	(0.0184)	(0.0034)	(0.0128)	(0.0131)	(0.0139)
		{0.0061}	{0.0264}	{0.0163}	{0.0229}	{0.0054}	{0.0180}	{0.0138}	{0.0132}
β	0.7241	0.9591	0.7647	0.7660	0.7146	0.9580	0.8322	0.7661	0.7368
		(0.0025)	(0.0479)	(0.0272)	(0.0753)	(0.0035)	(0.0328)	(0.0324)	(0.0477)
		{0.0062}	{0.1057}	{0.0466}	{0.1311}	{0.0056}	{0.0578}	{0.0376}	{0.0453}
p_{11}	0.9961	-	0.9556	-	0.9019	-	0.9928	-	0.9946
			(0.0076)		(0.0212)		(0.0048)		(0.0022)
			{0.0339}		{0.0298}		{0.0176}		{0.0031}
p_{22}	0.9986	-	0.7782	-	0.5195	-	0.9934	-	0.9980
			(0.0551)		(0.1107)		(0.0019)		(0.0007)
			{0.2172}		{0.1664}		{0.0207}		{0.0011}
κ	4.4037	-	-	4.6133	5.9962	-	-	4.6656	4.6734
				(0.6812)	(1.2267)			(0.8326)	(1.5511)
				{1.1060}	{1.5899}			{0.9567}	{0.5389}
v	6.6087	-	-	-	-	6.8347	6.2521	6.2701	6.6196
						(0.3981)	(0.4388)	(0.4251)	(0.4344)
						{0.4858}	{0.4503}	{0.4662}	{0.3909}

Notes: This table reports the average of parameter estimates, corresponding asymptotic standard errors, given in brackets, empirical standard deviations of parameter estimates, given in curly brackets, across 1,000 Monte Carlo simulations. In panel A and B the innovations are normally and Student-t distributed, respectively. The density is correctly specified in panel B and the true model is the RS GARCH MIDAS-t. The first column shows the true values of coefficients in the DGP. The high- and low-volatility regimes are given by p_{22} and p_{11} respectively.

Table 4.2: Average bias of parameter estimates.

	Panel A				Panel B			
	GARCH	RS GARCH	GARCH MIDAS	RS GARCH MIDAS	GARCH	RS GARCH	GARCH MIDAS	RS GARCH MIDAS
ω_1	-	0.0175	-	-0.0878	-	0.0238	-	0.0229
ω_2	-	3.7201	-	1.4879	-	0.2071	-	0.0149
θ	-	-	0.0162	0.0132	-	-	0.0163	-0.0017
α	-0.0552	-0.0552	0.0100	-0.0484	-0.0544	-0.0150	0.0124	0.0001
β	0.2350	0.0406	0.0419	-0.0095	0.2339	0.1081	0.0420	0.0127
p_{11}	-	-0.0405	-	-0.0942	-	-0.0033	-	-0.0015
p_{22}	-	-0.2204	-	-0.4791	-	-0.0052	-	-0.0006
κ	-	-	0.2096	1.5925	-	-	0.2619	0.2697
v	-	-	-	-	0.2260	-0.3566	-0.3386	0.0109

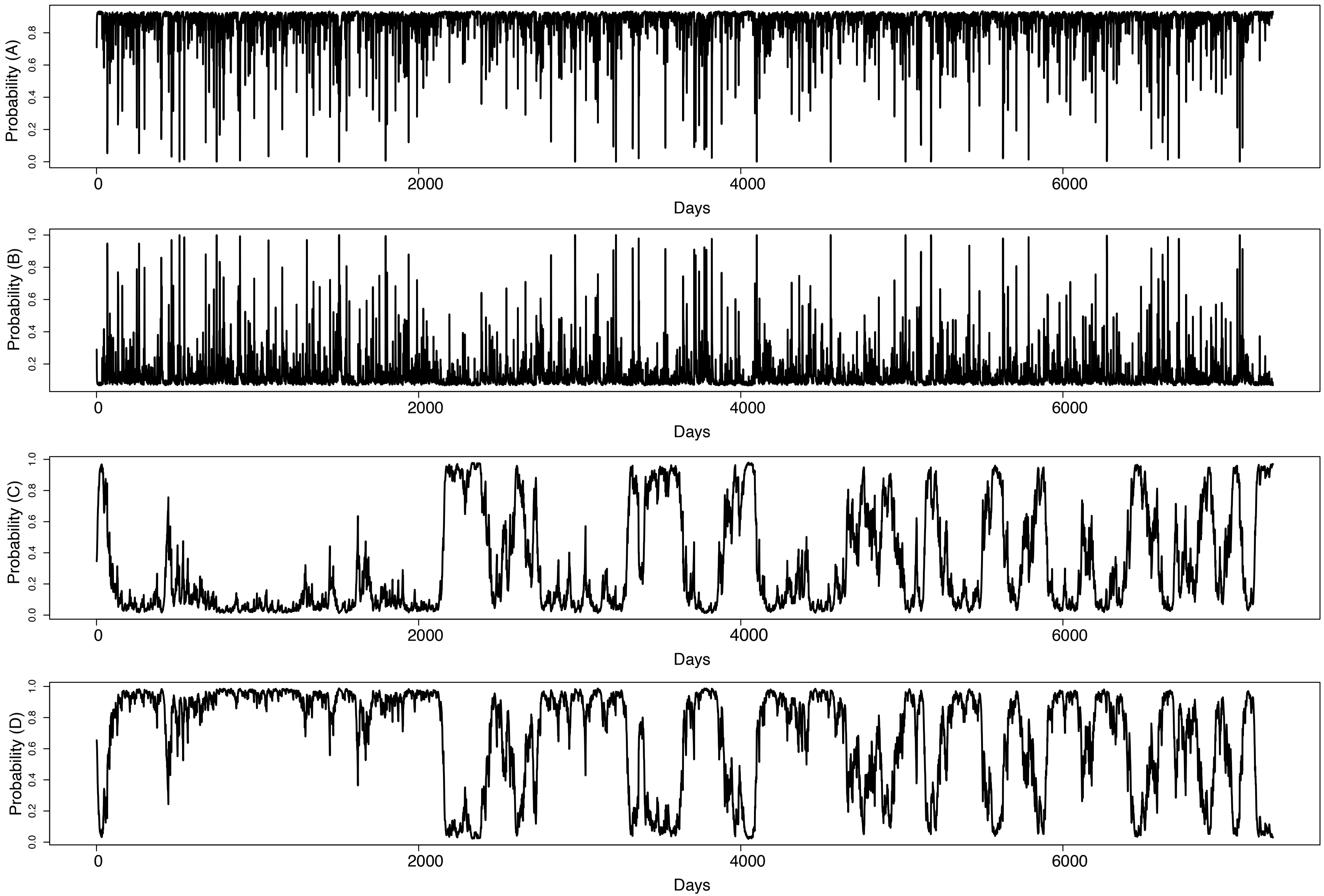
Notes: This table reports the average bias of parameter estimates across 1,000 Monte Carlo simulations. In panel A and B the innovations are normally and Student-t distributed, respectively. The density is correctly specified in panel B and the true model is the RS GARCH MIDAS-t. The high- and low-volatility regimes are given by p_{22} and p_{11} respectively.

density and investigate the consequences of this misspecification. For RS GARCH-N and RS GARCH MIDAS-N models, in Table 4.2 Panel A, ω_2 parameter is substantially positively biased whereas the corresponding transition probability is negatively biased. This is expected, because larger values of ω are required to compensate for the more extreme and frequent observations. Moreover, the regime persistence is more pronounced with Student-t innovations where both p_{11} and p_{22} are rather close to unity. By calculating the unconditional volatility[1] for each of the models we find that the state 2 always has higher volatility than state 1, in both innovations. The higher ω_2 parameters in regime switching models in Table 4.1 is also an indication for this. Therefore, it is reasonable for us to assume that regime 1 is the low-volatility regime while regime 2 is the high-volatility regime.

Next, to visualise how this misspecification affects the identification of transition probabilities in regime switching models we present the filter probabilities of RS GARCH MIDAS-N and RS GARCH MIDAS-t for a single simulation in Figure 4.1. Figure 4.1 (A) and (B), shows the filter probabilities of being in low-volatility and high-volatility regimes for RS GARCH MIDAS-N, whereas (C) and (D) shows the filter probabilities of being in low and high-volatility regime for RS GARCH MIDAS-t, respectively. Since all simulations have similar behaviour, we do not present all results. In Figure 4.1, (A) and (B) we can see that the Gaussian specification suffers from its inability to correctly identify the regime switching process. Similar results were shown by Haas and Liu (2018). This is because the probability for RS GARCH MIDAS-N to stay in high-volatility regime is very low at 0.5195, whereas for low-volatility regime it is 0.9019, indicating that it is more likely to stay in low-volatility regime most of the time as seen in plot (A). Furthermore, from plots (C) and (D) we can clearly see that RS GARCH MIDAS-t can distinguish between low- and high-volatility regimes. Similarly, RS GARCH model with Gaussian innovations also suffers from inability to correctly identify the regimes, whereas RS GARCH-t can identify them. The plots of filter probabilities for RS GARCH-N and RS GARCH-t can be found in Appendix C.1. This shows that the model under correctly specified innovations can capture more extreme observations than misspecified models. Moreover, plotting the filter probabilities

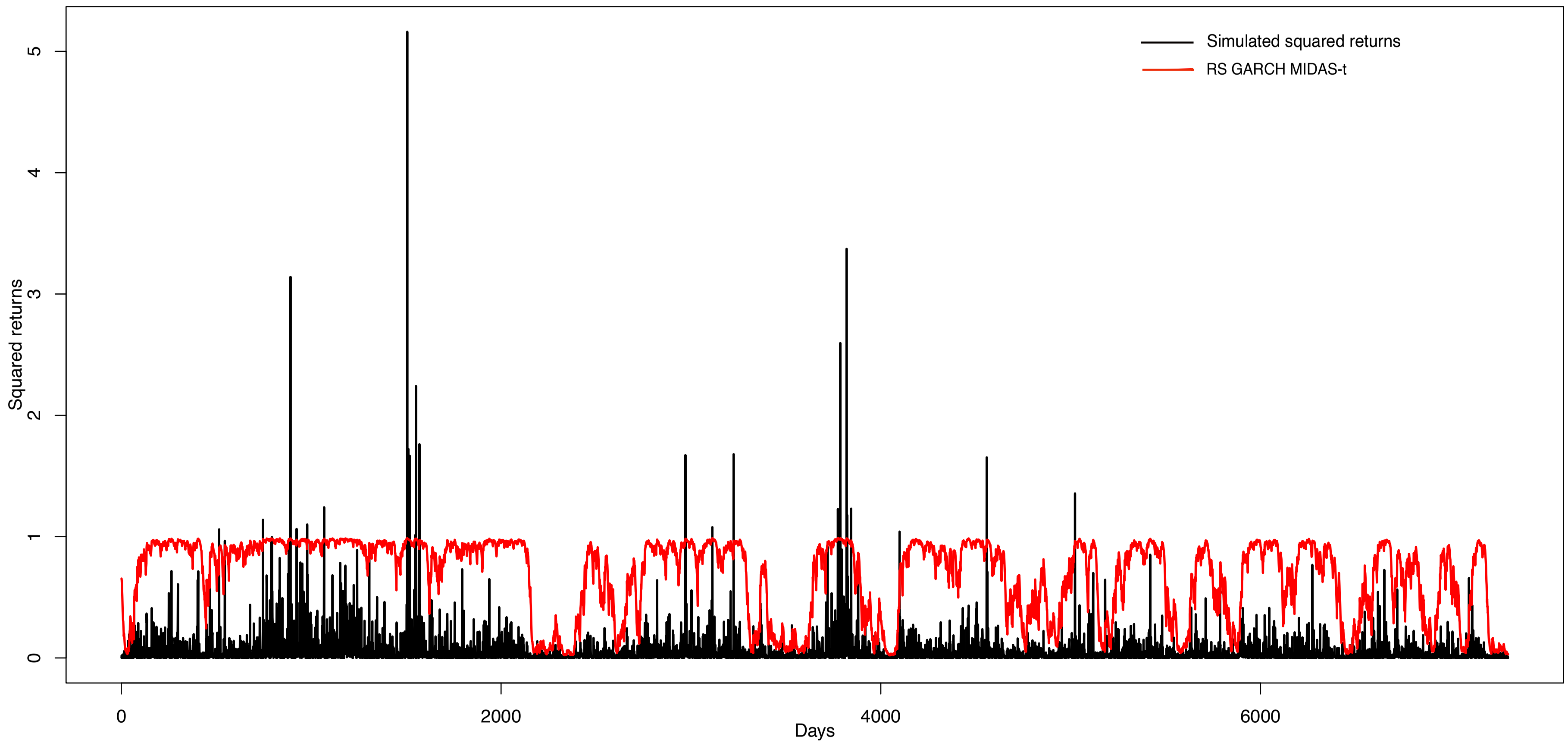
[1] Calculated as $\frac{\omega_{s_t}}{(1-\alpha-\beta)}$.

Figure 4.1: Filter probabilities for RS GARCH MIDAS models.



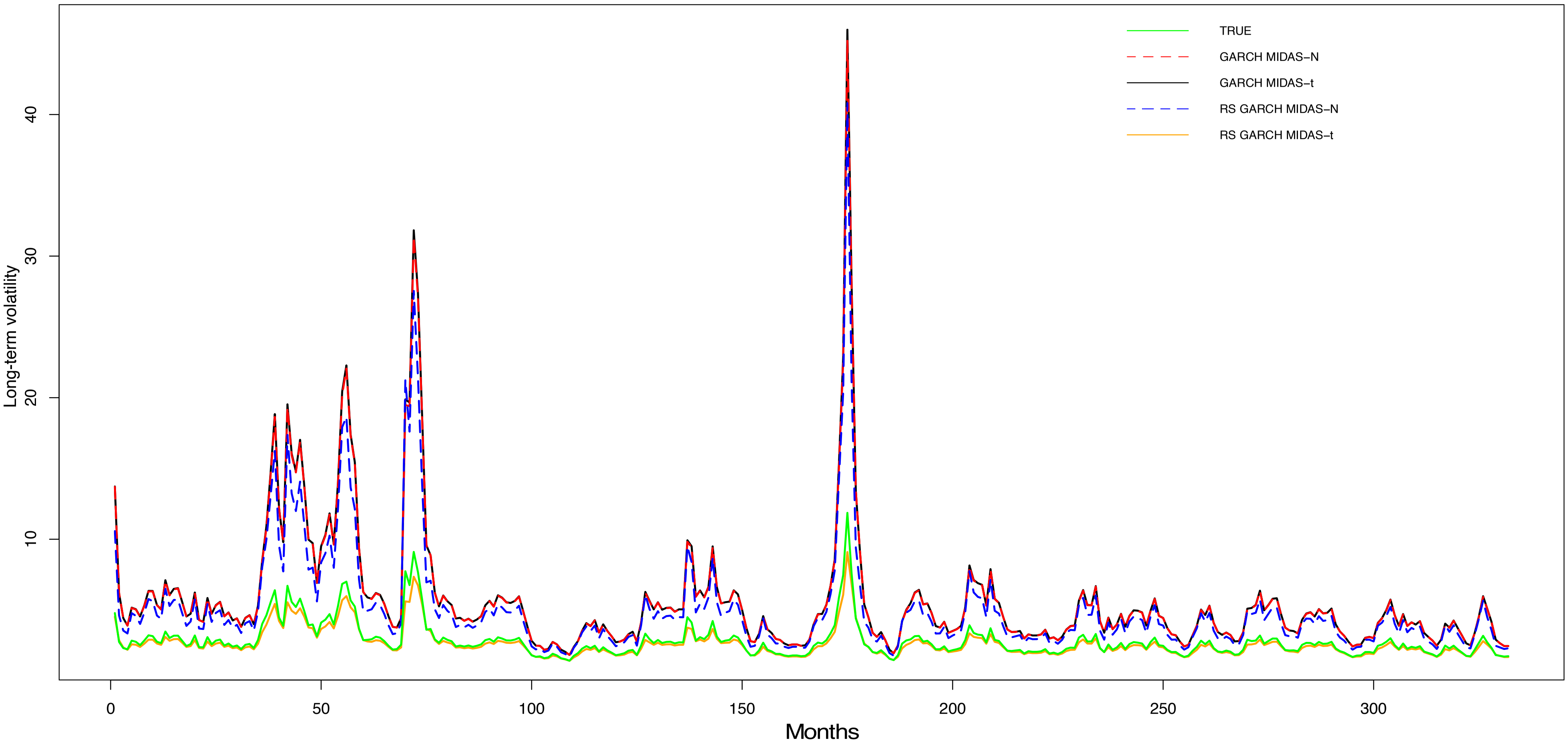
Notes: This Figure shows filter probabilities for RS GARCH MIDAS models. The (A) and (B) are the filter probabilities of being in low-volatility regime and high-volatility regime, respectively, according to RS GARCH MIDAS-N, whereas (C) and (D) are the filter probabilities of being in low-volatility regime and high-volatility regime, respectively, according to RS GARCH MIDAS-t.

Figure 4.2: Filter probabilities for RS GARCH MIDAS-t.



Notes: This Figure shows filter probabilities of being in high-volatility regime for RS GARCH MIDAS-t model, plotted on simulated squared returns. The squared returns are divided by 100.

Figure 4.3: Long-term volatility component.



Notes: This Figure shows the estimated long-term volatility, τ_t, by 4 different models: GARCH MIDAS-N, GARCH MIDAS-t, RS GARCH MIDAS-N and RS GARCH MIDAS-t. Green line indicates the true simulated τ_t while orange line is the true model.

with squared returns in Figure 4.2 indicates that the high-volatility regimes given in red, for RS GARCH MIDAS-t coincides with the time simulated returns experienced volatility changes.

As discussed before, the innovation change in GARCH MIDAS models hardly affects the parameters, and the plot of long-term volatility in Figure 4.3 also shows that there is barely any difference. On the other hand, by introducing regime switching we compare the long-term volatility in RS GARCH MIDAS models. It clearly shows that misspecifying the error term causes overestimation in long-term component, probably because of larger value for θ. In addition, the degree of persistence in the long-term component in RS GARCH MIDAS-t, measured by $\varrho = p_{11} + p_{22} - 1$, is very high at 0.9926, and in RS GARCH MIDAS-N it is very low at 0.4214.

In summary, under our simulation setting, we first evaluated the performance of the QMLE and afterwards analysed the effect of model misspecification in terms of omitting the long-term volatility component, not considering regime switching and ignoring fat-tails. We found that the correctly specified model, RS GARCH MIDAS-t, has much smaller bias compared to misspecified models and the asymptotic standard errors are close to the empirical standard deviation of the estimated parameters except for the κ parameter. Interestingly, GARCH and GARCH MIDAS models were not affected by misspecified innovations, however misspecification in terms of not considering regime switching affected GARCH MIDAS models, in both innovations, causing θ parameter to be large, leading to an over-estimation in the long-term volatility component.

More importantly, the misspecified innovations in regime switching models caused regime identification problem. To elaborate, the Gaussian specification in RS GARCH MIDAS model could not correctly identify high and low-volatility regimes, which lead to larger ω_2 parameter estimates and more frequent regime changes due to lower p_{22} parameter. In addition, this misspecification also caused an over-estimation of long-term component.

Hence we show importance of correctly specified model, in our case, the volatility model should contains long, short volatility component and regime switching with fat-tailed error distribution. We also conclude that, models with correctly specified innovations, namely Student-t innovations dominate their Gaussian counterparts for in-sample estimation in terms of smaller bias, asymptotic standard errors close to the empirical standard deviation and ability to correctly identify the low and high-volatility regimes. Some additional evidence is given in our empirical application in the next chapter.

Chapter 5

Application

5.1 West Texas Intermediate crude oil

Pan et al. (2017) fitted RS GARCH-MIDAS model to the daily spot price data of West Texas Intermediate (WTI) from 1986 to 2015. They suggested that two-regime GARCH MIDAS model can significantly outperform its no regime counterpart in forecasting oil volatility out-of-sample. Here we consider slightly longer series and use a fat-tailed distribution to see if there are any significant improvements in estimating and forecasting crude oil volatility.

The daily spot prices of WTI crude oil, $ per barrel, are obtained from Energy Information Administration (EIA) website[1]. The sample period is from January 1, 1986 to July 27, 2020. The total number of observations is 8709. The data is divided into two subsets. The in-sample which covers periods from January 2, 1986 to December 31, 2015 (30 years) resulting in 7567 observations is used for estimation purposes while the remaining period from January 1, 2016 to July 27, 2020 (4 years) with 1142 observations is used for forecasting evaluation.

Volatility changes can be caused by many factors and it is difficult to point out which specific factors have the dominant effect on the oil prices. Many scholars have investigated the impact of various macroeconomic variables, such as macroeconomic uncertainty, national economic policy uncertainty (EPU), global economic policy uncertainty (GEPU), along traditional determinants, such as global oil demand, supply,

[1] https://www.eia.gov/dnav/pet/hist/RWTCD.htm

and speculation, were examined for their capacity to predict crude oil price volatility. Overall, the oil price fluctuation is found to be susceptible to several factors, however, the oil market's demand and supply shocks remain the major drivers (Zhao, 2022; Le et al., 2023). Furthermore, the GEPU index wasn't considered because the monthly dataset is only available from 1997 onward, while the WTI crude oil prices start from January 1986, on the other hand, we did consider the U.S. EPU index in our analysis, however, the results obtained from incorporating the US EPU in our models were statistically insignificant. Therefore, as a fundamental factor, production and demand levels are considered as main factors affecting oil prices. Further discussion can be found in Section 2.9.

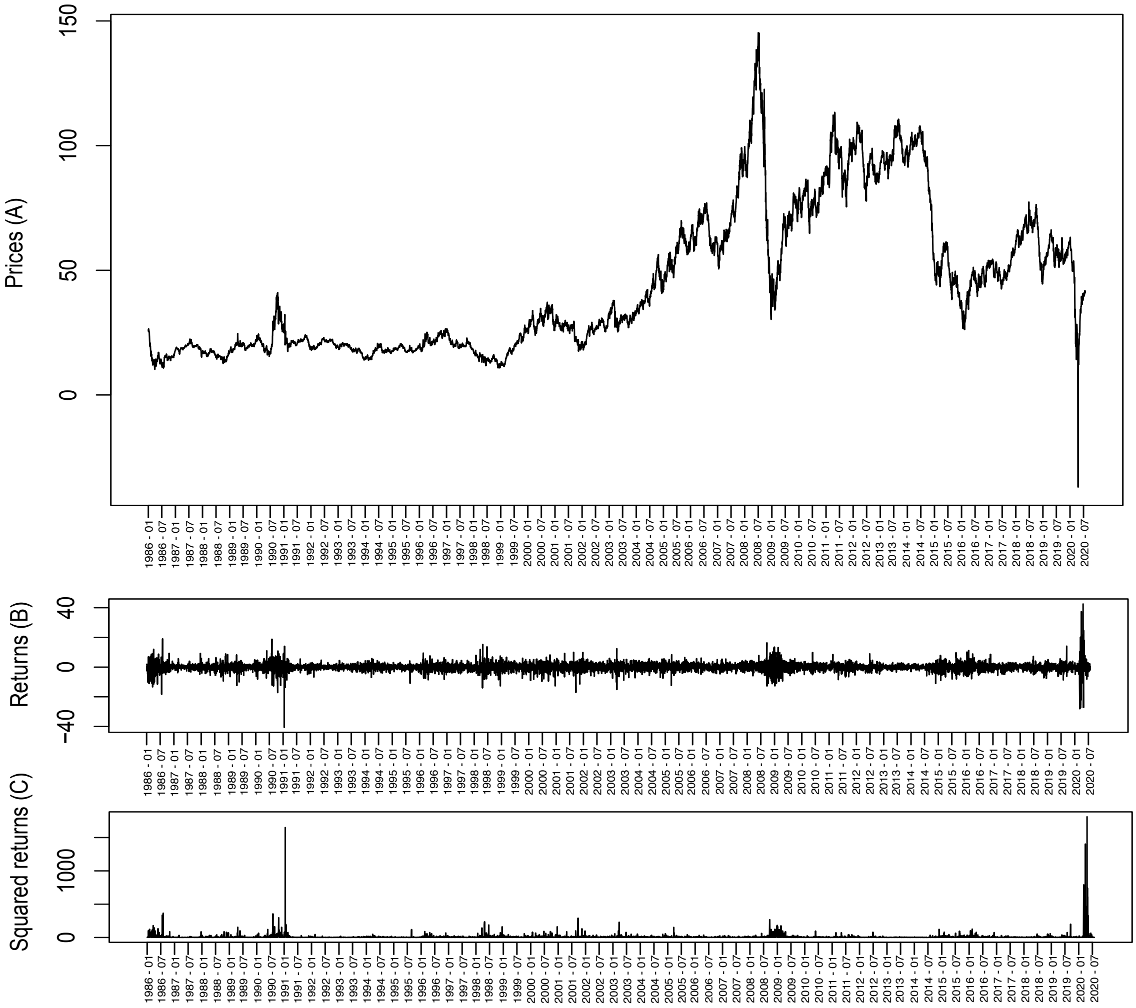
The monthly data reflecting oil production and demand levels are selected for the period January 1986 to July 2020. Global oil production, obtained from EIA is used as a proxy of world oil supply where the data is given in number of barrels produced per month. Following recent market studies by Baumeister and Kilian (2015), Pan et al. (2017), index of Kilian (2009)[2] is used as the signal for oil demand. Kilian (2009) developed a structural VAR model to explain the global crude oil price fluctuation where the crude oil price was decomposed into three components: crude oil supply shock, shocks to the global demand for all industrial commodities and the demand shock to the global crude oil market.

The daily oil prices are converted to log returns using the first order difference of log prices $r_t = 100[\log(p_t) - \log(p_{t-1})]$, where p_t denotes crude oil price at time t. The reason that the log returns are considered instead of simple returns is the additivity property of log returns which is not seen in simple returns. The graphical representations of daily WTI crude oil prices (A), returns (B) and squared returns (C) are shown in Figure 5.1. From the plot it is observed that crude oil prices experience high uncertainty over time. As discussed in Section 3.6 daily squared returns is used as volatility proxy.

In Figure 5.1 (C) two largest volatilities occur around 1990-1991 and 2020. The oil price increase, in plot Prices (A) in 1990-1991 was caused by Iraq invading Kuwait leading to the first Gulf war, resulting in higher prices due to decrease of oil produc-

[2]https://www.dallasfed.org/research/igrea

Figure 5.1: Daily crude oil prices, returns and volatilities



Notes: This figure shows the graphical representations of WTI crude oil prices, given in dollars per barrel in (A), returns in (B) and squared returns used as volatility proxy in (C).

tion. Whereas in 2020, a historic drop, plot (A), occurred in April due to COVID-19, when the prices of WTI crude oil dropped by almost 300% and was trading at around negative $37 per barrel. This drop in demand is essentially caused by the quarantine restrictions in countries which lead to a drop in consumption. Furthermore, in (A), the increase in crude oil prices between 2003 - 2008 is driven by economic boom in emerging countries followed by a quick drop caused by global economic recession. In 2014, crude oil prices fell sharply following a production that exceeded the demand (Baumeister and Kilian, 2015). This indicates that some supply and demand shocks can lead to large fluctuations in crude oil markets. Therefore it is vital to estimate and forecast crude oil price volatility using appropriate models with appropriate macroe-

Table 5.1: Descriptive statistics of daily WTI crude oil returns and monthly macroeconomic variables.

	WTI	Production	Demand
Mean	0.0049	0.1020	-0.1961
Var	6.3840	1.1438	243.9191
Min	-40.6396	-7.0825	-100.1827
Max	19.1507	4.5266	69.4562
Skewness	-0.7194	-1.2432	-0.9112
Kurtosis	17.2860	12.8387	10.5614
JB ($\times 10^4$)	6.5042***	0.1567***	0.0921***
	(0.0000)	(0.0000)	(0.0000)
ADF(5)	-38.4600***	-9.8045***	-8.4187***
	(0.0100)	(0.0000)	(0.0100)
Q(5)	35.4960***	10.1960*	45.0490***
	(0.0000)	(0.0699)	(0.0000)
$Q^2(5)$	461.6090***	7.0149	106.2500***
	(0.0000)	(0.2195)	(0.0000)
ARCH(5)	333.0900***	5.9610	86.7260***
	(0.0000)	(0.3100)	(0.0000)

Notes: Jarque-Bera (JB), augmented Dickey-Fuller (ADF), Ljung-Box, Q(5), and ARCH are the statistics testing for normal distribution, stationarity, serial correlation and heteroskedastic effects respectively. The corresponding p-values are given in brackets. WTI, production and demand are unitless.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

conomic variables. To include these periods in our estimation we select the period from January 1986 to December 2015 as the within-sample period. The graphical representations of production and demand levels can be found in Appendix C.2 and C.3.

Descriptive statistics of the crude oil returns, production and demand levels for the within-sample period are given in Table 5.1. The full dataset descriptive statistics are reported in Appendix C.1. As Table 5.1 shows, the WTI crude oil has a daily average return of 0.0049 with a variance of 6.3840. The returns and macroeconomic variables exhibit negative skewness indicating that the data's are slightly skewed to the left with production being the lowest, which confirms the existence of asymmetry. This might be due to asymmetric tendencies sometimes observed in financial time series such as leverage effects. Furthermore, the negative skewness and positive excess kurtosis imply that the returns are not normally distributed indicating that fat-tailed distributions might be necessary to describe these variables. In addition, this is confirmed by the Jarque and Bera (1980) (JB) test, which rejects the null hypothesis of normally distributed returns at any level of significance for all three datasets.

Thus, a non normal distribution, such as Student-t might provide a better fit to the data for it can capture fat-tails. The augmented Dickey-Fuller (ADF) unit-root test yields a p-value of less than 0.01 for all data series considered indicating that WTI, production and demand returns are stationary. Furthermore, we use Engle's ARCH test to assess the significance of ARCH effects present in the series. The ARCH test confirms the presence of heteroscedasticity and therefore a significant ARCH effects in the crude oil returns and demand but not in production. This can also be seen in autocorrelation function (ACF) plots of the three datasets given in Appendix C.4. Finally, the Ljung-Box statistics, Q(5) and $Q^2(5)$ of serial correlation suggests the significant autocorrelations in both returns and squared returns in WTI and demand. The $Q^2(5)$ statistic for production is not statistically significant at any significance level and thus indicates strong evidence for the null hypothesis whereas Q(5) statistic shows that production returns are statistically significant at 10%. This can also be seen in the ACF plot for production shown in Appendix C.4.

5.2 Within-sample estimation evaluation

In this section we discuss the estimation results of models with different specifications.

We begin by estimating simple GARCH(1,1) model presented in Equation (3.3). Since mean of WTI crude oil returns are close to zero, we assume that $\mu = 0$ in all model specifications. Furthermore, for a fair comparison of GARCH model with GARCH MIDAS we set $\omega = (1 - \alpha - \beta)$. Let GARCH-$N$ and GARCH-t denote GARCH(1,1) model where N and t indicate normal or Student-t distributions, respectively.

Next, since macroeconomic variables are sampled at lower frequencies we use a MIDAS approach to link these variables to the long-term component. The GARCH MIDAS model given in Equation (3.45) allows us to incorporate the short- and long-term volatility components (Equations (3.46), (3.47), (3.48)) directly into the model to detect the effects of these variables on oil price volatility. Following work of Pan et al. (2017), in long-term component we assume $m = 0$. Similar to GARCH model, denote GARCH MIDAS-N for normal and GARCH MIDAS-t for Student-t innovations.

Based on the empirical evidence that the volatility of financial markets display some type of persistence that cannot be appropriately captured by GARCH model and its variations (Lamoureux and Lastrapes, 1990; Engle and Mustafa, 1992) we introduce regime switching to reduce the persistence parameters and to take into account the role of structural breaks. The models considered are RS GARCH(1,1) given in Equation (3.26) and RS GARCH MIDAS in Equation (3.58). Similarly, denote RS GARCH-N, RS GARCH-t, RS GARCH MIDAS-N, RS GARCH MIDAS-t for RS GARCH(1,1) and RS GARCH MIDAS with Gaussian and Student-t innovations.

To be consistent with GARCH MIDAS model, we also set $m = 0$ in RS GARCH MIDAS. Furthermore, we only allow the short-term volatility component to switch between two regimes. Additionally, in this short-term component it is assumed that only ω_{s_t} is allowed to switch regimes as Marcucci (2005) and Wang et al. (2022) showed that the differences of parameters α and β between two regimes are likely to be insignificant. Furthermore, according to Guérin and Marcellino (2013) and Asgharian et al. (2013) the identification and convergence problem caused by increment of the number of parameters will happen, hence inclusion of several macroeconomic variable in the long-term is also not considered.

In models with long-term volatility component the realised volatility, where it is calculated as sum of squared returns, is mainly used as the natural explanatory variable (Engle et al., 2013), however, since absolute returns could also capture fluctuations in future return volatility Ding et al. (1993), Ghysels et al. (2006) and Taylor (2008) explored the advantages of calculating RV as absolute value of returns. For example, Ding et al. (1993), provided evidence that the low-frequency components of volatility might be more effectively captured by using absolute returns instead of squared returns. Similarly, Forsberg and Ghysels (2006) showed that regressors involving volatility measures based on absolute returns are superior in predicting future volatility. Therefore in this thesis, absolute returns[3] calculated as:

$$RV_t = \sum_{i=1}^{N_t} |r_{i,t}|, \tag{5.1}$$

are considered as natural explanatory variables. As for the lag choice, we estimated the model for different values of K and stopped when the improvement in the log-likelihood was negligible. Conrad and Loch (2015) also showed that as long as the selected K is large enough, the estimation results are robust with respect to the specific choice of the maximum number of lags included.

Next, an extended Markov switching models allowing for endogeneity in regime switching namely Endogenous RS GARCH and Endogenous RS GARCH MIDAS models are considered. These models are similar to RS GARCH and RS GARCH MIDAS with only difference in the transition probabilities. The transition probabilities for endogenous regime switching models are given in Equation (3.73) and if $\gamma = 0$, the transition probabilities become exogenous and our model reduces to the standard RS GARCH and RS GARCH MIDAS.

Thus, we fit 12 different models for the WTI crude oil returns: GARCH-N, GARCH-t, RS GARCH-N, RS GARCH-t, GARCH MIDAS-N, GARCH MIDAS-t, RS GARCH MIDAS-N, RS GARCH MIDAS-t, Endo RS GARCH-N, Endo RS GARCH-t, Endo RS GARCH MIDAS-N and Endo RS GARCH MIDAS-N. For now we model the low-frequency movements in conditional variance by realised volatility only. We consider macroeconomic variables, production and demand later in the chapter.

Parameter estimates along with their corresponding standard errors are given in Table 5.2 and Table 5.3. The degree of volatility persistence is given by $\rho = (\alpha + \beta)$. The log-likelihood value (LL) is also reported as a measure of the model's goodness of fit.

In general, all parameters in models with Student-t innovations, see Table 5.3, are statistically significant at 5% significance level and the estimates of individual parameters across different models are found to be relatively close to each other. On the

[3] We tried estimating GARCH MIDAS models using squared returns as natural explanatory variables of WTI oil returns, however, the estimation results were not something expected as it provided unreasonable long-term volatility.

Table 5.2: Parameter estimates with Gaussian innovations.

Parameters	GARCH(1,1)	RS GARCH(1,1)	GARCH MIDAS	RS GARCH MIDAS	Endo RS GARCH (1,1)	Endo RS GARCH MIDAS
ω_1		0.0075	-	0.0671	0.0593***	0.0539***
		(0.0269)		(0.0411)	(0.0085)	(0.0112)
ω_2	-	1.7910***	-	2.6475***	12.8481***	8.1577***
		(0.2955)		(0.4447)	(2.5696)	(1.4556)
θ	-	-	0.0438***	0.0362***	-	0.0140***
			(0.0009)	(0.0019)		(0.0038)
α	0.0605***	0.0542***	0.1361***	0.0227	0.0316***	0.0334***
	(0.0044)	(0.0082)	(0.0128)	(0.0124)	(0.0040)	(0.0047)
β	0.9385***	0.8977***	0.7355***	0.7242***	0.9317***	0.9068***
	(0.0045)	(0.0088)	(0.0316)	(0.0357)	(0.0055)	(0.0111)
p_{11}	-	0.9647***	-	0.9413***	-	-
		(0.0071)		(0.0081)		
p_{22}	-	0.7676***	-	0.4965***	-	-
		(0.0523)		(0.0588)		
κ	-	-	3.8431***	4.9575***	-	2.9773***
			(0.4703)	(0.5988)		(0.7686)
c_1	-	-	-	-	4.4286***	4.3637***
					(0.1691)	(0.1550)
c_2	-	-	-	-	74.7480	25.5582**
					(36.7130)	(11.0836)
γ_1	-	-	-	-	0.2502***	0.2648***
					(0.0436)	(0.0411)
γ_2	-	-	-	-	27.0086	16.2377**
					(13.4502)	(6.6193)
LL	-16130.07	-15981.30	-16084.94	-15866.90	-15863.23	-15858.71
ρ	0.9999	0.9519	0.8716	0.7469	0.9633	0.9402

Notes: This table reports the maximum likelihood estimates of different models for the daily WTI crude oil volatility. The distribution of error terms are assumed to follow Gaussian innovations. For MIDAS component RV is taken as the absolute value of returns. LL denotes the log-likelihood value and the persistence of the shocks are indicated by $\rho = \alpha + \beta$. Standard errors are given in brackets.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Table 5.3: Parameter estimates with Student-t innovations.

Parameters	GARCH(1,1)	RS GARCH(1,1)	GARCH MIDAS	RS GARCH MIDAS	Endo RS GARCH(1,1)	Endo RS GARCH MIDAS
ω_1	-	0.0841***		0.1785***	0.0400***	0.2254***
		(0.0219)		(0.0458)	(0.0092)	(0.0403)
ω_2	-	0.2716***	-	0.4044***	0.0667***	0.5018***
		(0.0685)		(0.1007)	(0.0138)	(0.0937)
θ	-	-	0.0411***	0.0262***	-	0.0264***
			(0.0013)	(0.0026)		(0.0023)
α	0.0519***	0.0694***	0.0788***	0.0944***	0.0614***	0.0834***
	(0.0044)	(0.0103)	(0.0124)	(0.0145)	(0.0059)	(0.0128)
β	0.9471***	0.8892***	0.8540***	0.7241***	0.9288***	0.6469***
	(0.0044)	(0.0182)	(0.0334)	(0.0579)	(0.0065)	(0.0460)
p_{11}	-	0.9963***	-	0.9961***	-	-
		(0.0014)		(0.0018)		
p_{22}	-	0.9984***	-	0.9986***	-	-
		(0.0002)		(0.0002)		
κ	-	-	3.4168***	4.4037***	-	5.8208***
			(0.7830)	(1.1940)		(1.0471)
v	6.9338***	6.8665***	6.5797***	6.6087***	6.1302***	6.7697***
	(0.4298)	(0.5295)	(0.4154)	(0.4710)	(0.4064)	(0.4991)
c_1	-	-	-	-	6.8644***	7.1089***
					(0.6891)	(1.3226)
c_2	-	-	-	-	8.1690***	6.6102***
					(1.2745)***	(0.5235)***
γ_1	-	-	-	-	0.1177***	1.6848***
					(0.0243)	(0.4432)
γ_2	-	-	-	-	0.4882***	0.4195***
					(0.1218)	(0.0656)
LL	-15864.48	-15816.76	-15847.7	-15808.86	-15822.12	-15808.54
ρ	0.9999	0.9586	0.9328	0.8185	0.9902	0.7303

Notes: This table reports the maximum likelihood estimates of different models for the daily WTI crude oil. The distribution of error terms are assumed to follow Student-t. For MIDAS component RV is taken as the absolute value of returns. LL denotes the log-likelihood value, the persistence of the shocks are indicated by $\rho = \alpha + \beta$ and v is the degrees of freedom. Standard errors are given in brackets.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

other hand, when the innovations are assumed to follow a normal distribution, Table 5.2, several parameters become statistically insignificant. The loss of statistical significance in certain parameters suggests that the normal distribution assumption may not adequately capture the inherent characteristics of the volatility process in financial time series data. This finding aligns with the results obtained from the simulation in Chapter 4.

Moreover, positive and significant α in all cases, except for RS GARCH MIDAS-N, confirms the presence of ARCH effects. An $\alpha < 0.10$ in all cases, except for GARCH MIDAS-N, indicates a low volatility period. Furthermore, GARCH persistence parameter β, is significant in all models, and where it is lower than 0.90 indicates volatility does not take long time to converge to average volatility. In cases when $\beta > 0.90$ indicates that volatility will persist for a long time following a market shock.

Looking at the volatility clustering in the data through the persistence parameter $\alpha + \beta$, we find that across all models in both innovations, the stationarity condition, $\alpha + \beta < 1$ is satisfied. The sum of these two parameters is lower in RS GARCH MIDAS models and lowest in Endo RS GARCH MIDAS-t compared to others where it is above 0.90. The higher persistence usually means that volatility of crude oil return is remarkably persistent over entire period.

The estimated degrees of freedom, v are greater than 4 in all models, suggesting that all the conditional moments up to the fourth order exist and a lower values of v indicate the inability of normal error to account for the fat tails.

Next, comparing the regime switching models with normal and t innovations, we observe that the regime persistence is more pronounced with Student-t innovations where the probabilities p_{11} and p_{22} are close to unity. This means that if current volatility belongs to the high volatility regime, the next day's volatility is more likely to stay in the same regime, i.e., the expectations of staying at both states are quite long. Whereas if the innovation is assumed to be Gaussian then ω_1 parameter in both regime switching GARCH and RS GARCH MIDAS models become insignificant and the transition between two state become more frequent due to lower p_{11} and p_{22} values.

Looking at the interpretation of the MIDAS component in GARCH MIDAS type models, both parameters θ and κ are significant. A positive θ in all GARCH MIDAS type models is an indicator of long-term oil volatility being positively related to realised volatility. In fact, bigger fluctuation in the RV would cause bigger long-term crude oil fluctuation. Whereas a lower value for κ means that all the lags have similar effect on RV thus the beta weights are declining slowly.

Next, in order to assess the performance of the models with respect to how well it describes the crude oil data, both log-likelihood values and model selection criteria are analysed. The most common approach involves selecting a model that minimises the Akaike's information criteria (AIC), Bayesian information criteria (BIC) and Hannan-Quinn information criteria (HQIC). These are calculated as follows:

$$\begin{aligned} AIC(P) &= -2(LL) + 2P, \\ BIC(P) &= -2(LL) + P\ln(N), \\ HQIC(P) &= -2(LL) + 2P\ln(\ln(\mathrm{N})), \end{aligned} \tag{5.2}$$

where P is the number of parameters estimated in the model, N is the number of observations and LL is log-likelihood value. Comparing the above equations BIC tends to select simpler models then those chosen by the AIC since it is stricter in penalising loss of degree of freedom than AIC. Using the log-likelihood values obtained through estimation we determine the best model is the one with lowest AIC, BIC and HQIC values.

The log-likelihood values and the model selection criteria for all models, divided into two panels for easier interpretation, are presented in Table 5.4. In panel A, the innovations are assumed to follow Gaussian whereas in panel B, the innovations are assumed to follow Student-t distribution. The highest LL and the lowest AIC, BIC, HQIC values amongst all models are given in red, whereas the highest LL and the lowest AIC, BIC, HQIC values among models with Gaussian innovations are given in blue.

In general, we observe that models with Student-t distributed errors generally have higher LL and lower selection criteria values than models with Gaussian distribution. Next, the regime switching models have the highest log-likelihood values and lowest

Table 5.4: Log-likelihood and model selection criteria.

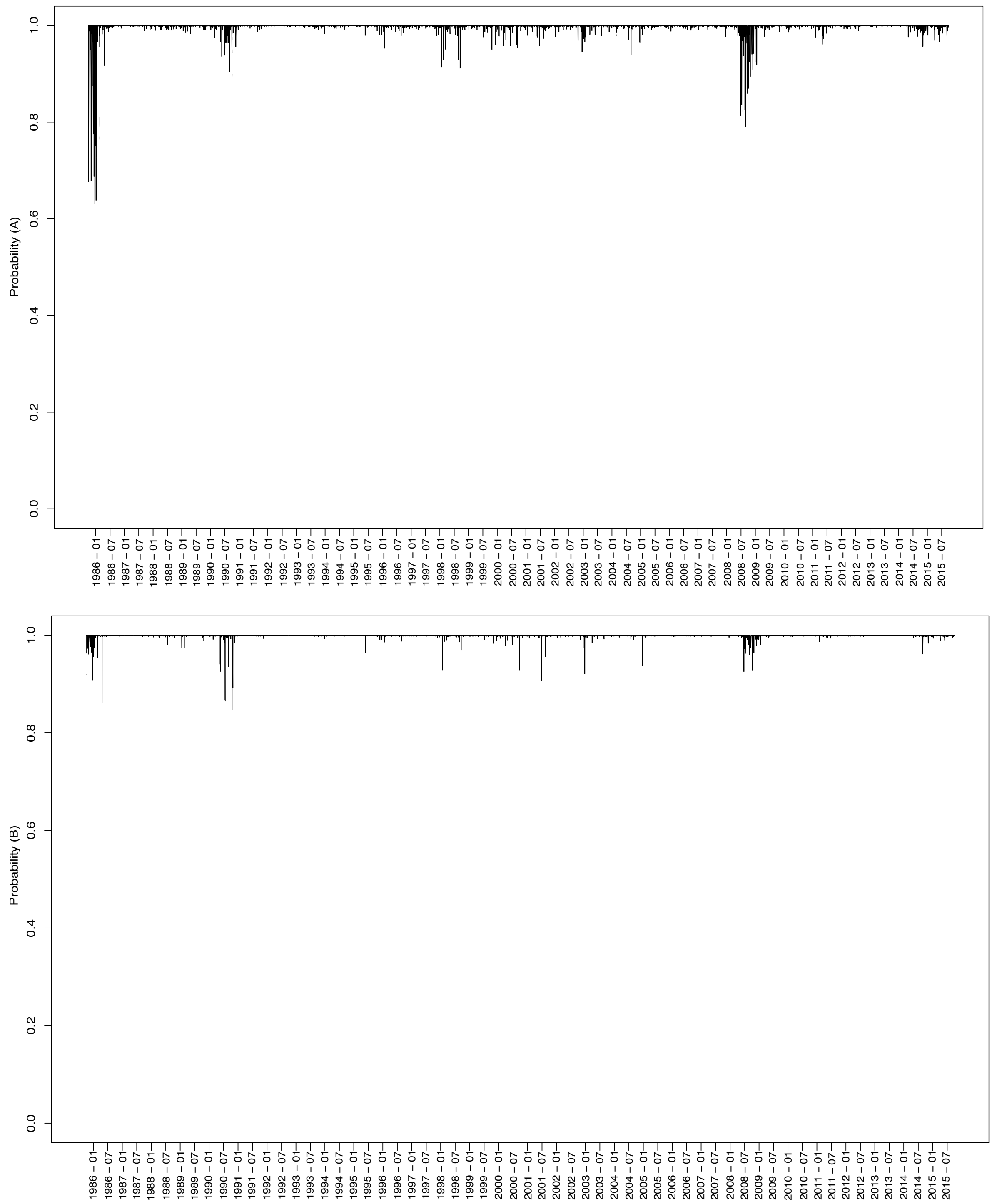
Model	*LL*	AIC	BIC	HQIC
Panel A: *Gaussian*				
GARCH	-16130.07	32264.14	32278.00	32268.90
RS GARCH	-15981.30	31974.60	32016.19	31988.88
GARCH MIDAS	-16084.94	32177.88	32205.61	32187.40
RS GARCH MIDAS	-15866.90	31749.80	31805.25	31768.83
Endo RS GARCH	-15863.23	31742.46	31797.91	31761.49
Endo RS GARCH MIDAS	-15858.71	31737.42	31806.74	31761.21
Panel B: *Student-t*				
GARCH	-15864.48	31732.96	31746.82	31755.24
RS GARCH	-15816.76	31647.52	31696.04	31664.17
GARCH MIDAS	-15847.70	31705.40	31740.06	31717.30
RS GARCH MIDAS	-15808.86	31635.72	31698.10	31657.13
Endo RS GARCH	-15822.12	31662.24	31724.62	31683.65
Endo RS GARCH MIDAS	-15808.54	31639.08	31715.33	31665.25

Notes: The Table reports the log-likelihood values, AIC, BIC and HQIC statistics for different model specifications. In panels A and B the innovations are normally or Student-t distributed. The computation of AIC, BIC and HQIC is given in Equation (5.2). The highest LL and lowest AIC, BIC and HQIC amongst all models is given in red, whereas for panel A they are given in blue.

AIC, BIC and HQIC statistics compared to their no regime switching counterparts. Specifically, among the models with Student-t errors (panel B), RS GARCH MIDAS-t stands out with the lowest AIC and HQIC values. RS GARCH-t, on the other hand, has the lowest BIC value. However, the difference of 2.04 in BIC between RS GARCH-t and RS GARCH MIDAS-t indicates there is a positive evidence in favour of the more complex model (Raftery, 1995; Fabozzi et al., 2014).

The comparison between Endo RS GARCH MIDAS-t and RS GARCH MIDAS-t models reveals some interesting findings. Although the estimated parameters of the Endogenous RS GARCH MIDAS-t model are statistically significant, the transition probabilities given in Figure 5.2 exhibit a behaviour similar to the constant transition probabilities produced by the exogenous RS GARCH MIDAS-t. Despite the inclusion of additional endogenous variables in the Endogenous RS GARCH MIDAS-t model, the time-varying nature of the transition probabilities remains relatively stable and does not deviate significantly from the constant transition probabilities of the exogenous model. This suggests that the exogenous model captures the primary drivers of regime

Figure 5.2: Transition probabilities of Endo RS GARCH MIDAS model with Student-t innovations.



Notes: This Figure shows the transition probabilities of Endogenous RS GARCH MIDAS-t. The Probability (A) and Probability (B) shows p_{11} and p_{22} respectively, calculated using parameters from Table 5.3 and Equation (3.73) .

shifts and adequately represents the volatility dynamics in the data.

In terms of log-likelihood, Endogenous RS GARCH MIDAS-t and RS GARCH MIDAS-t with log-likelihood values of -15808.54 and -15808.86, respectively, exhibit very similar performance. To determine if the difference in log-likelihood is statistically significant and to assess the goodness of fit, a likelihood ratio (LR) test can be applied. The likelihood ratio test compares the likelihoods of nested models, where one model is a restricted version of the other. In this case, RS GARCH MIDAS-t is nested within the more general model Endo RS GARCH MIDAS-t. Performing the LR test will help determine whether the additional parameters in the more complex model significantly improve the model fit. The test statistic is given as follows:

$$\text{LR} = 2[(LL_{ur}) - (LL_r)] \tag{5.3}$$

where LL_{ur} and LL_r are the log-likelihoods of unrestricted model and restricted model, respectively. This test gives us a p-value of 0.7261, thus failing to reject the null meaning that the restricted model, RS GARCH MIDAS-t, is preferred.

When considering the models with Gaussian errors, the AIC criterion favours the most complex model, Endo RS GARCH MIDAS-N. This preference for complexity in regime-switching models aligns with the findings of Kapetanios (2001). On the other hand, the BIC criterion tends to prefer endogenous RS GARCH-N, while HQIC is lowest in Endo RS GARCH MIDAS-N.

Overall, results obtained from the log-likelihood values and model selection criteria indicate that amongst the 12 model specifications the crude oil volatility is best captured by our proposed model, RS GARCH MIDAS-t.

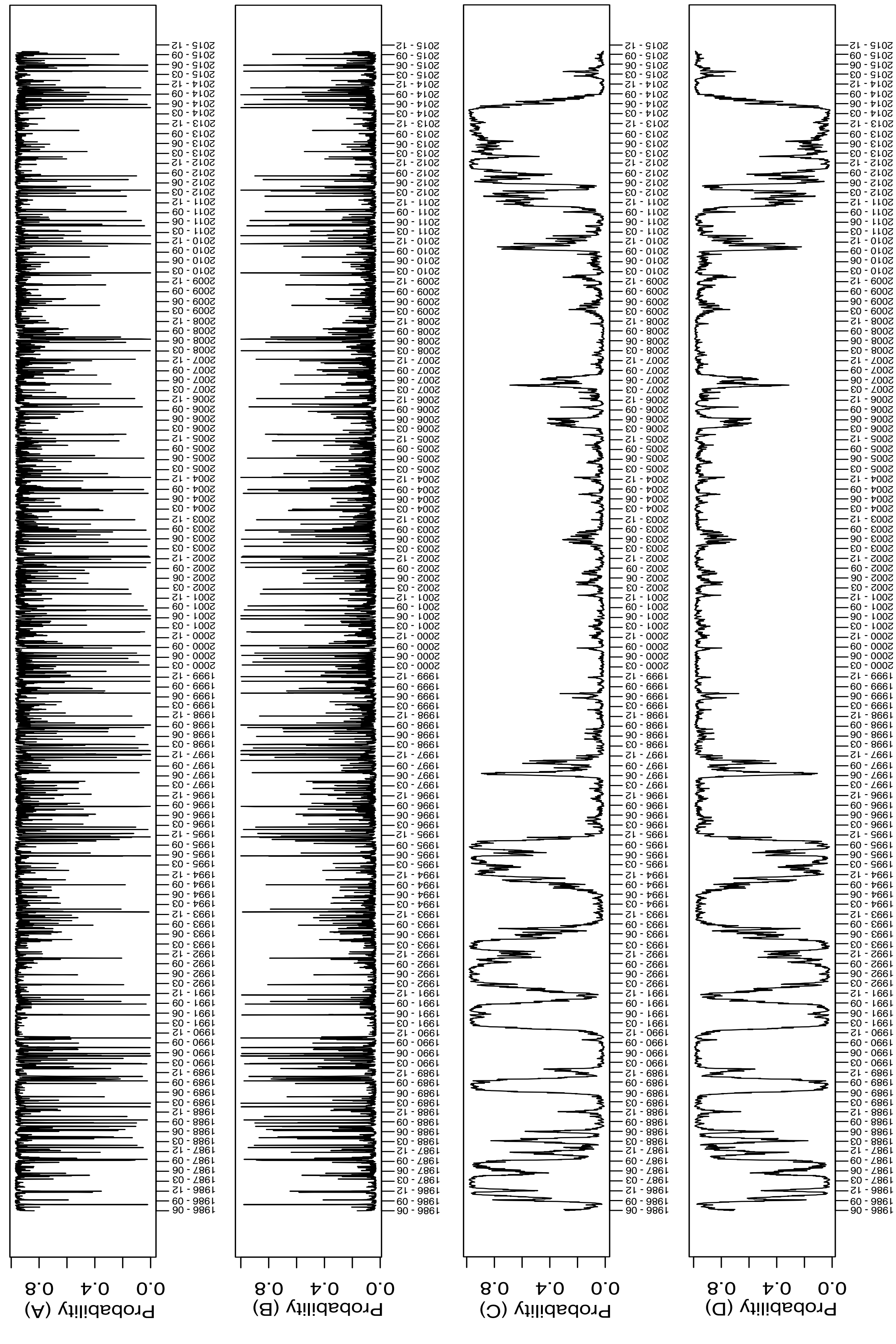
Comparing our results for RS GARCH MIDAS-N model with the parameter estimates obtained by Pan et al. (2017), we notice that the short-term volatility parameter estimates are quite similar whereas the transition probabilities and the long-term volatility parameters differ. A plausible explanations about the difference of reported values is that we use absolute returns as realised volatility measure and longer horizon (30 years) for in-sample estimation whereas Pan et al. (2017) considered the squared returns and their sample data covered the period from January 2, 1986 to December 31, 2015 (15 years). Furthermore, they applied Gray (1996) approach to tackle the path dependence issue whereas we follow an approach developed by Klaassen (2002) which is a modified version of Gray's approach. Lastly, we considered a fat tailed innovation while they assumed that errors are normally distributed.

Once we showed that WTI crude oil volatility is best captured by RS GARCH MIDAS model with the Student-t error distribution we turn our attention to in-depth analysis. For regime switching type models while considering the unconditional volatility we find that state 2 in RS GARCH MIDAS-t has two times higher volatility than state 1 whereas the volatility in regime 2 in RS GARCH MIDAS-N it is nearly 40 times higher than regime 1. The unconditional volatility in RS GARCH-t is 2.0314 and 6.5604 for regime 1 and regime 2, respectively, which again shows that the volatility in state 2 is 3 times higher than state 1. Hence, it is reasonable for us to assume that regime 1 is the low-volatility regime while regime 2 is the high-volatility regime.

To illustrate the difference between transition probabilities we plot the filter probabilities for RS GARCH MIDAS models in Figure 5.3. The Figure 5.3 (A) and (B) are the filter probabilities of being in low-volatility regime and high-volatility regime, respectively, for RS GARCH MIDAS-N model, whereas (C) and (D) are the probabilities of being in low-volatility regime and high-volatility regime according to RS GARCH MIDAS-t.

From Figure 5.3 we can see that RS GARCH MIDAS model with t innovations, plots (C) and (D), can easily identify the two volatility regimes, whereas the model with Gaussian innovations fails to do so. This is probably because RS GARCH MIDAS-N model tends to produce more frequent regime changes when there are volatile observations than the RS GARCH MIDAS-t. If the probability at time t is larger than 0.5, we can say that the volatility stays in Regime 1 (high-volatility regime) at that time, and the volatility regime switches to Regime 2 (low-volatility regime) if the probability is smaller than 0.5. The probability for RS GARCH MIDAS-N to stay in high-volatility is very low at 0.4965, meaning that it is more likely to stay in low-volatility for most of the time, which is not ideal to model volatility changes. Consequently, misspecified models with Gaussian errors tends to over-estimate the regime changes to compensate

Figure 5.3: Filter probabilities for RS GARCH MIDAS models.



Notes: This Figure shows filter probabilities for RS GARCH MIDAS models. The plots (A) and (B) are the probabilities of being in low-volatility regime 1 and high-volatility regime 2, respectively, according to RS GARCH MIDAS-N, whereas (C) and (D) are the filter probabilities of being in low-volatility regime and high-volatility regime according to RS GARCH MIDAS-t.

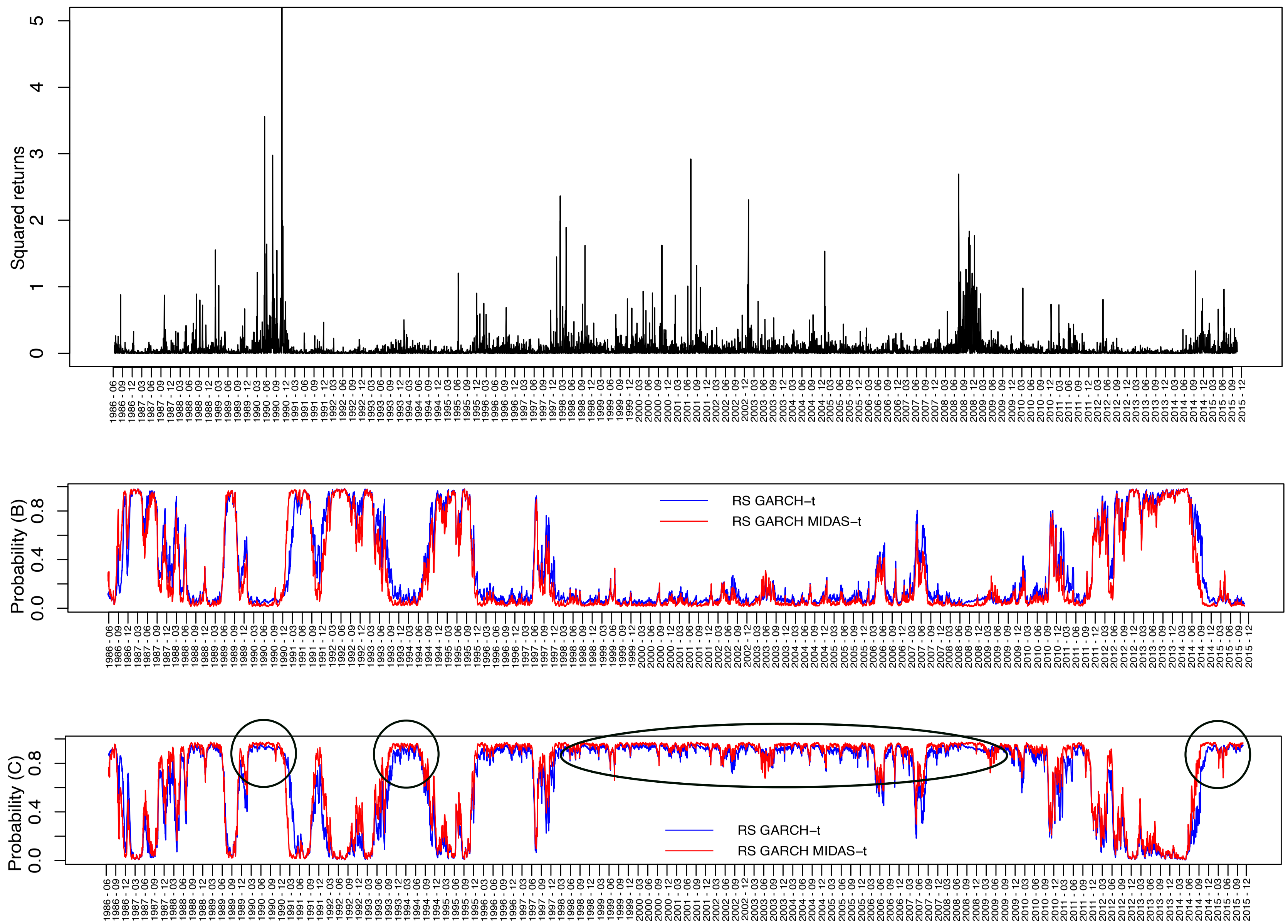
the additional variation and extreme observations. These results are in line with findings from our simulation in the previous chapter. Furthermore, the results in Figure 5.4 confirm the existence of regime switching due to the probabilities larger than 0.5.

Furthermore, plotting the filter probabilities of RS GARCH-t and RS GARCH MIDAS-t with the volatility proxy in Figure 5.4 suggests the volatility regimes identified by our model correlate well with major events affecting supply and demand for oil. For instance, the filter probabilities indicate a high-volatility regime around 1990-1991 which is when Iraq invaded Kuwait leading to the First Persian Gulf war causing the oil prices to increase due to lower production. Another high-volatility regime is observed around 1994 caused by excess OPEC supply, which can be seen clearer in Figure 5.5. Third high-volatility period is seen around 1996-1997 which coincided with backwardation in the oil market. The last two high-volatility periods that were identified are also discussed in Fong and See (2002). Since these periods are associated with supply/demand crises of oil disruptions, spikes of volatility are produced (Iglesias and Rivera-Alonso, 2022).

Crude oil prices show very strong volatility persistence during the period 1998 and 2010. During this period the economy experienced booming, recession and recovery periods hence all volatility is captured by high-volatility regime. For instance, the east Asian economic crisis affected the oil market and caused a collapse of oil prices in 1998 (Mabro, 2009). In 2001, the U.S. terrorist attack produced high levels and spikes of volatility (Zavadska et al., 2020).

The most remarkable surge in the price of oil occurred between mid-2003 and mid-2008 with the WTI crude oil price increasing from \$28 to \$134 per barrel. There is a general consensus that the surge in oil prices was not a result of oil supply disruptions but rather a cumulative effect of numerous small increases in the demand for crude oil over several years. Kilian (2009) and Hamilton (2009) among others argue that these demand shifts are associated with an unexpected expansion of the global economy. During the financial crisis of 2008 the demand for crude oil plummeted, causing a fall in the price of oil (Baumeister and Kilian, 2016). More specifically, these periods which were triggered by economic or financial crises are associated to higher volatility

Figure 5.4: Filter probabilities for regime switching models with Student-t errors.



Notes: This Figure shows the time series plot of filter probabilities in regime 1 (A) and in regime 2 (B) at time t according to RS GARCH-t and RS GARCH MIDAS-t. The plot (C) shows filter probabilities of the state 2 for RS GARCH-t and RS GARCH MIDAS-t for the period 1986 - 1995. The black line indicates the scaled realised volatility.

persistence, which is similar to findings of (Zavadska et al., 2020; Iglesias and Rivera-Alonso, 2022).

There have been a number of smaller demand and supply shocks in the oil market between 2010 and 2014. According to the filter probabilities the last period of high-volatility regime in our in-sample period is around 2014, the oil conflict of Saudi-Arabia with the US (Iglesias and Rivera-Alonso, 2022), when the oil prices began to decrease worldwide and continued to drop significantly until January 2015. The reason for this decline was an oversupply of oil compared to demand.

Additionally, in Figure 5.5 we observe that just after the high volatility in 1991, RS GARCH-t declines slower compared to RS GARCH MIDAS-t even though parameter estimates are very close to each other. A possible explanation for this is that accounting for long-term volatility in regime switching GARCH MIDAS-t reduces persistence in the short-term component.

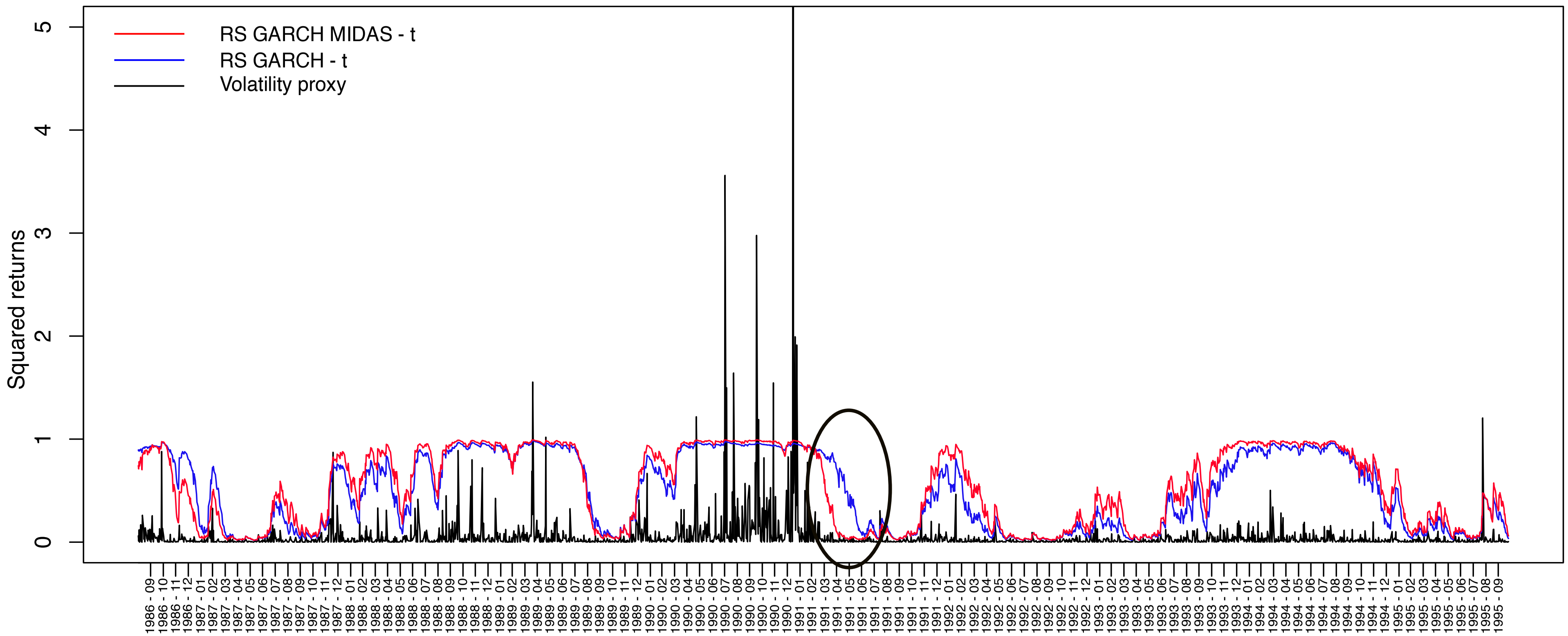
The estimated expected duration[4] provide insights into the dynamics of the RS GARCH MIDAS model and its ability to capture the volatility patterns in the data. In the case of the RS GARCH MIDAS model with Gaussian errors, the expected duration of the low-volatility regime is approximately 17 days, while for the high-volatility regime it is only 2 days. This can explain the frequent regime changes observed in Figure 5.2. A possible explanation for such pattern is due to the tendency of regimes in Gaussian to signal a regime shift whenever an unusual small or large observations occur within an otherwise calm regime (Haas and Paolella, 2012).

On the other hand, the RS GARCH MIDAS model with Student-t innovations exhibits longer expected durations for both the low-volatility and high-volatility regimes. Specifically, the expected duration for the low-volatility regime is approximately 1 year (257 days), while for the high-volatility regime it is slightly less than 2 years (715 days). This indicates that the model captures the persistence of volatility observed in crude oil returns, as higher volatility periods tend to persist for a longer duration.

The findings regarding the high regime persistence in the RS GARCH MIDAS-t model are consistent with the results of Herrera et al. (2018), who also found that the transition probabilities between regimes were close to one, indicating highly persis-

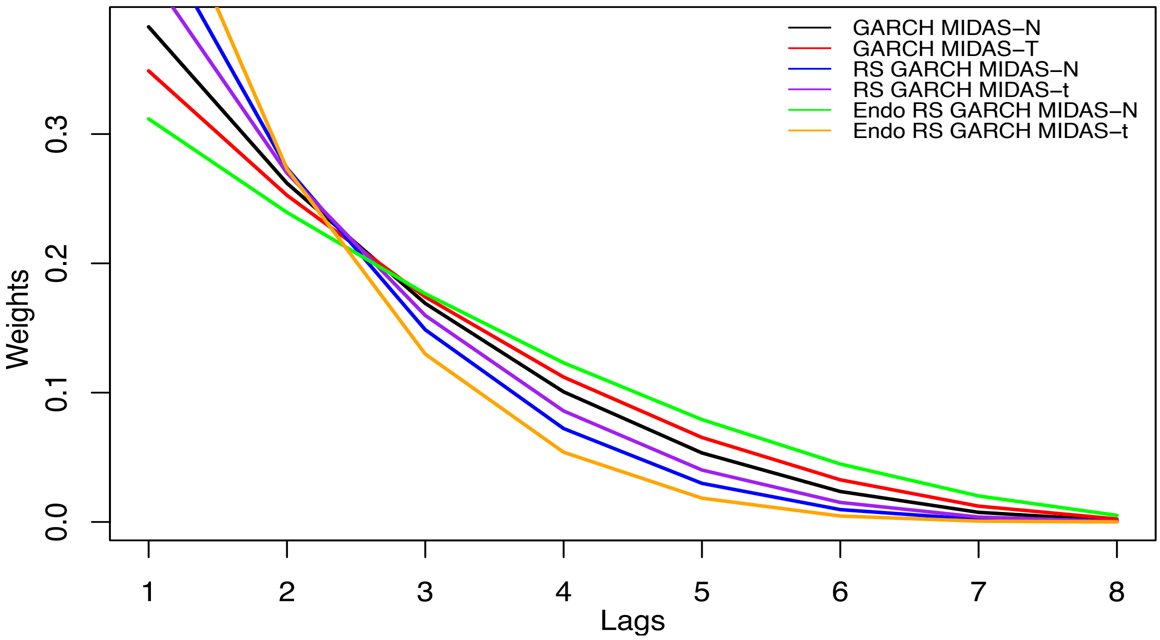
[4] Calculated as $(1 - p_{ii})^{-1}$ where i=1 or 2.

Figure 5.5: Filter probabilities for regime switching models with Student-t innovations.



Notes: This Figure shows the time series plot of filter probabilities for high-volatility regime for RS GARCH-t and RS GARCH MIDAS-t. The black line indicates the scaled squared returns.

Figure 5.6: Weighting function



Notes: This Figure shows the behaviour of weights as a function of the number of lags in GARCH MIDAS-N, GARCH MIDAS-t, RS GARCH MIDAS-N, RS GARCH MIDAS-t, endogenous RS GARCH MIDAS-N and endogenous RS GARCH MIDAS-t.

tent regimes. This suggests that a substantial majority of the observations belong to the high-volatility regime, further supporting the ability of the model to capture the persistence in volatility.

In terms of MIDAS approach, the parameters θ and κ along with realised volatility describes the long-term component in GARCH MIDAS models. A lower value for θ in RS GARCH MIDAS-t, implies lower persistence of the short-term volatility component and that monthly realised volatility lags contain information that help model the low-frequency component. Figure 5.6 illustrates the plot of weighting function, κ, for models with MIDAS component. As seen from the plot the weight function is monotonically decreasing for all models, this is similar to work of Engle et al. (2008) and Asgharian et al. (2013). A closer look at RS GARCH MIDAS-t shows the weight on lag 1 is 0.4251, followed by 0.2699 on lag 2 and last lag equals 0.0004. This shows that the lower weights are given to the most recent observations. Furthermore, RS GARCH MIDAS-t for the in-sample period gives an estimate of 0.0262 for θ and 4.4037 for κ. Hence, if a shock occurred in the current month, we would expect to see an increase of $e^{(0.0262\times0.4251)} - 1 = 1.1201\%$ in the long-term oil volatility the next month.

To illustrate the long-term component, τ_t for GARCH MIDAS-N, GARCH MIDAS-

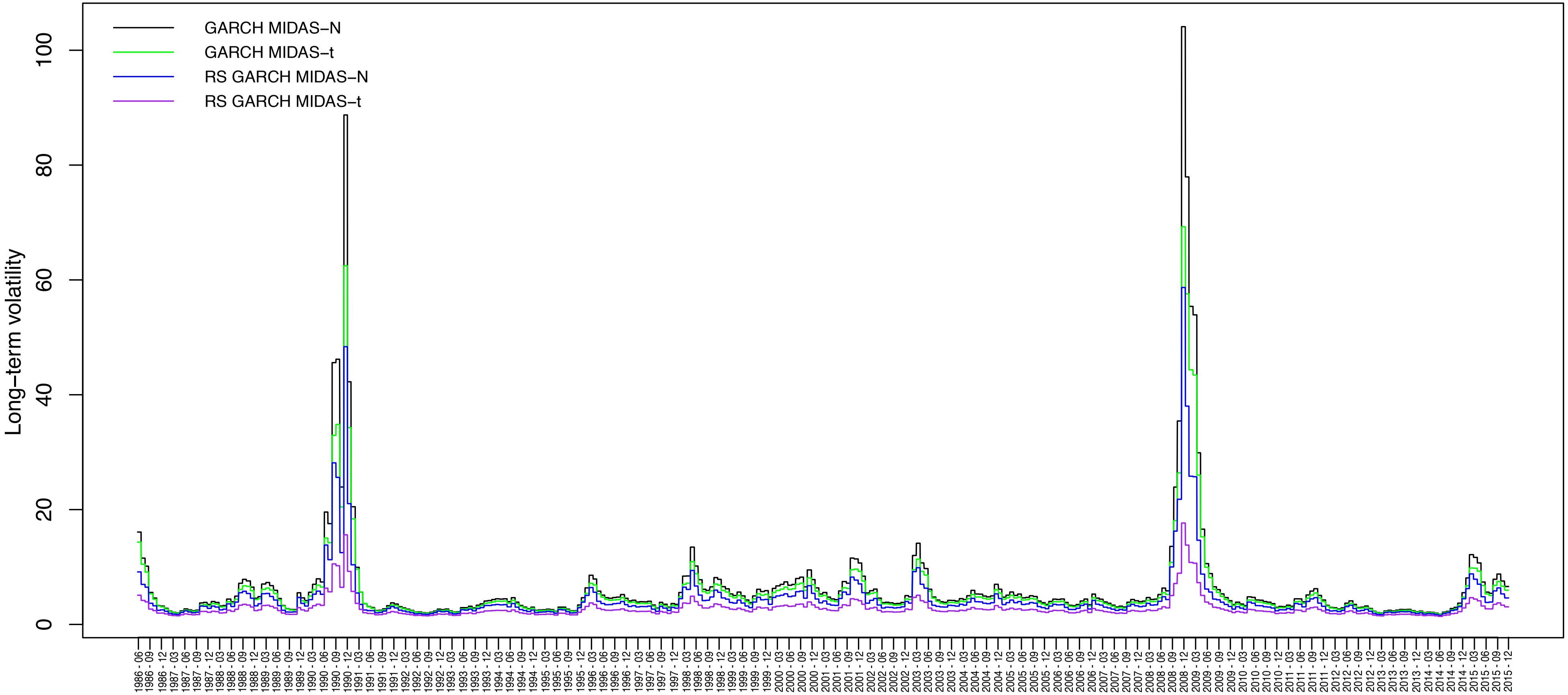
t, RS GARCH MIDAS-N and RS GARCH MIDAS-t models for the in-sample period we plot it in Figure 5.7. As can be seen from Figure 5.7, all four models give relatively similar pattern, most of the time. Moreover, the long-term volatility is high around 1990-1991 and 2009 which is the period when crude oil returns displayed high return volatility in Figure 5.1 (C). However, the long-term volatility is lowest in RS GARCH MIDAS-t and highest in GARCH MIDAS-N. This high value of τ_t is probably balanced by higher short-term component caused by regime switching and extreme volatility changes because of t innovations. Similar results were obtained in simulation.

It is worth mentioning that the models without the long-term volatility component, GARCH, RS GARCH and endogenous RS GARCH indicate a high degree of persistence in the conditional volatility of WTI crude oil returns while models with MIDAS approach have lower value of ρ. So whether this high persistence arises due to regime changes or any other reasons, we can conclude that accounting for long-term volatility reduces persistence in the short-term component.

Now we turn our attention to investigate whether or not the incorporation of macroeconomic variables in the long-term component can describe the WTI crude oil volatility better compared to RS GARCH MIDAS-t with RV only. When considering production and demand levels as macroeconomic variables, the models become: GARCH MIDAS-N (RV+Prod), GARCH MIDAS-t (RV+Prod), RS GARCH MIDAS-N (RV+Prod), RS GARCH MIDAS-t (RV+Prod), GARCH MIDAS-N (RV+Dem), GARCH MIDAS-t (RV+Dem), RS GARCH MIDAS-N (RV+Dem), RS GARCH-MIDAS-t (RV+Dem). The parameter estimates of GARCH MIDAS and RS GARCH MIDAS models with normal and Student-t errors where the long-term volatility is now described by a combination of realised volatility and macroeconomic variables are reported in Tables 5.5 and 5.6. To include these additional variables in the long-term volatility component the τ_t considered is given in Equation (3.48).

From the estimation results in Tables 5.5 and 5.6 we can see that the parameter estimates of GARCH MIDAS models are quite similar to the GARCH MIDAS model estimates where τ_t is described by RV only. On the contrary, in RS GARCH MIDAS models with macroeconomic variables some of the parameters estimates are different

Figure 5.7: In-sample estimation of long-term volatility.



Notes: This Figure shows the estimated long-term volatility, τ_t, of 4 different models: GARCH MIDAS-N, GARCH MIDAS-t, RS GARCH MIDAS-N and RS GARCH MIDAS-t. The estimation covers the period from September 1986 to December 2015.

Table 5.5: Parameter estimates with macroeconomic variables and Gaussian innovations.

Parameters	GARCH MIDAS (RV+Prod)	RS GARCH MIDAS (RV+Prod)	GARCH MIDAS (RV+Dem)	RS GARCH MIDAS (RV+Dem)
ω_1	-	0.0465	-	0.1103
		(0.1014)		(0.1021)
ω_2	-	2.4231***	-	2.7498***
		(0.4183)		(0.4607)
θ_1	0.0427***	0.0379***	0.0438***	0.0348***
	(0.0009)	(0.0020)	(0.0009)	(0.0020)
θ_2	0.5635***	0.3659***	-0.0015	-0.0005
	(0.1144)	(0.1062)	(0.0020)	(0.0021)
α	0.1431***	0.0285**	0.1365***	0.0162
	(0.0130)	(0.0140)	(0.0128)	(0.0124)
β	0.7165***	0.7289***	0.7350***	0.7022***
	(0.0312)	(0.0733)	(0.0314)	(0.0713)
p_{11}	-	0.9383***	-	0.9441***
		(0.0132)		(0.0131)
p_{22}	-	0.4606***	-	0.5492***
		(0.1200)		(0.0807)
κ_1	4.3554***	5.2883***	4.3770***	7.4000***
	(0.5234)	(0.8242)	(0.5334)	(1.3297)
κ_2	1.7725***	1.4952***	11.6071***	14.8860***
	(0.5007)	(0.4752)	(0.0030)	(0.0099)
LL	-16063.31	-15859.39	-16084.92	-15870.03
ρ	0.8596	0.7574	0.8715	0.7184

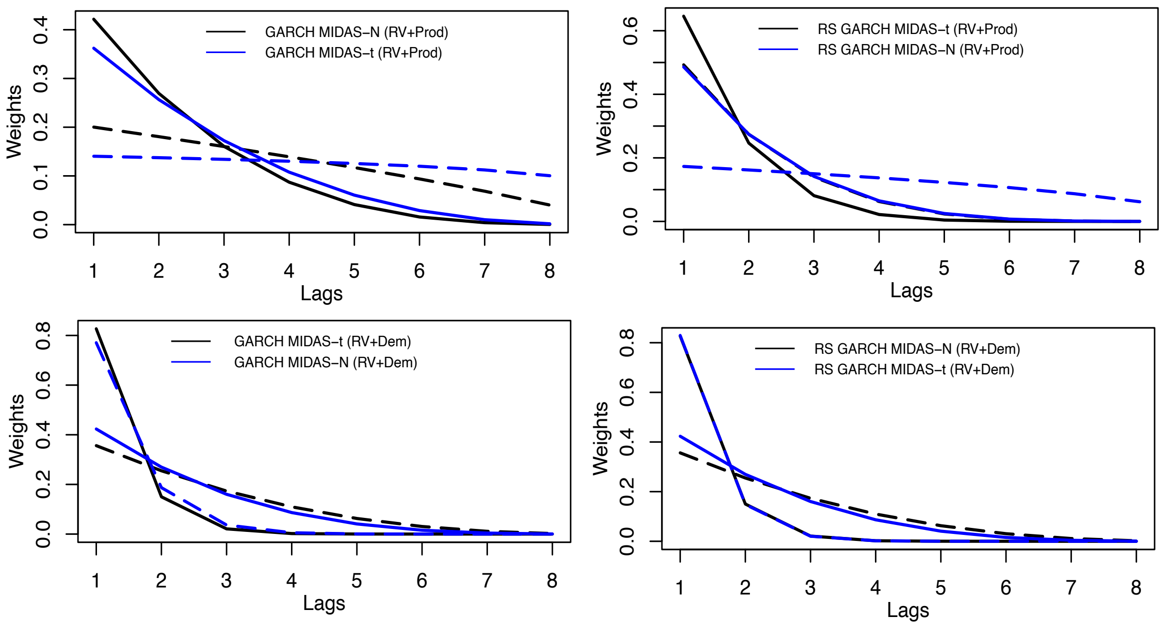
Notes: This table shows the estimation results of GARCH MIDAS and RS GARCH MIDAS for WTI Crude oil with Normal distribution. The first row corresponds to the MIDAS regressors, that are realised volatility and oil production level (RV+Prod), realised volatility and demand (RV+Dem). LL denotes the log-likelihood value. The persistence of the shocks are indicated by $\rho = \alpha + \beta$. The standard errors are reported in parenthesis. The parameters θ_2 and κ_2 tell us the quantitative effects of production/demand.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Table 5.6: Parameter estimates with macroeconomic variables and Student-t innovations.

Parameters	GARCH MIDAS RV+Prod	RS GARCH MIDAS RV+Prod	GARCH MIDAS RV+Dem	RS GARCH MIDAS RV+Dem
ω_1	-	0.2420***	-	0.1564
		(0.0492)		(0.0900)
ω_2	-	0.5504***	-	0.3660
		(0.1148)		(0.1893)
θ_1	0.0403***	0.0251***	0.0412***	0.0256***
	(0.0013)	(0.0026)	(0.0013)	(0.0050)
θ_2	0.5635***	0.1104***	-0.0004	-0.0014
	(0.1280)	(0.0463)	(0.0023)	(0.0024)
α	0.0808***	0.1039***	0.0808***	0.0940***
	(0.0139)	(0.0141)	(0.0131)	(0.0247)
β	0.8433***	0.6516***	0.8451***	0.7434***
	(0.0387)	(0.0517)	(0.0374)	(0.1353)
p_{11}	-	0.9957***	-	0.9966***
		(0.0004)		(0.0004)
p_{22}	-	0.9983***	-	0.9986***
		(0.0003)		(0.0001)
κ_1	3.5804***	8.2125***	3.5067***	3.7769**
	(0.8470)	(2.5230)	(0.8680)	(1.2956)
κ_2	1.1612***	5.3909***	13.7804***	13.8935***
	(0.2719)	(0.2534)	(0.0011)	(0.0375)
v	6.5184***	7.5240***	6.6293***	7.2694***
	(0.4192)	(0.6587)	(0.4252)	(0.6260)
LL	-15837.25	-15808.70	-15847.74	-15810.89
ρ	0.9241	0.7556	0.9259	0.8374

Notes: This table shows the estimation results of GARCH MIDAS and RS GARCH MIDAS for WTI Crude oil with Student-t innovations. The first row corresponds to the MIDAS regressors, that are realised volatility and oil production level (RV+Prod), realised volatility and demand (RV+Dem). LL denotes the log-likelihood value and v is the degrees of freedom and the persistence of the shocks are indicated by $\rho = \alpha + \beta$. The standard errors are reported in parenthesis. The parameters θ_2 and κ_2 tell us the quantitative effects of production/demand.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Figure 5.8: In-sample long-term volatility.



Notes: This Figure shows the behaviour of weights as a function of the number of lags in GARCH MIDAS and RS GARCH MIDAS models. Solid line indicates κ_1 value whereas dashed line indicates κ_2 value from the same model. The weights are calculated using Equation (3.50).

from our proposed model.

We also observe that production has a significantly positive impact on long-term crude oil volatility while demand has an insignificantly negative effect, meaning that demand has a negligible influence on the monthly component of crude oil daily volatility. These results align with the conclusions drawn by Wei et al. (2017) who also found that the demand factor is insignificant, meanwhile, the supply factor is positive and statistically significant in oil volatility modelling.

Furthermore, parameter θ_2 together with the weighting parameter κ_2 tells us the quantitative effects of oil fundamentals. For instance, in GARCH MIDAS-N model for production the estimate of θ_2 is 0.5635 while for demand θ_2 it is -0.0015. A significant κ_2 of 1.7725 means that all lags have similar effect on production thus, declining slowly, while the demand κ_2 of 11.6071 means that only the first couple of lags have significant effect on demand with last values being very close to zero (rapid decline), i.e. κ_2 value of 11.6071 puts a weight of 0.7708 on the first lag, 0.1870 on the second, 0.0364 on the third while the seventh and eighth input on the lags are 3.1682×10^{-7}, 2.0313×10^{-10} respectively. This can be seen in Figure 5.8.

Table 5.7: Log-likelihood and model selection criteria.

Model	*LL*	AIC	BIC	HQIC
Panel A: *(RV+Prod)*				
GARCH MIDAS-N	-16063.31	32138.62	32180.21	32152.90
GARCH MIDAS-t	-15837.25	31688.50	31737.02	31705.15
RS GARCH MIDAS-N	-15859.39	31738.78	31808.10	31762.57
RS GARCH MIDAS-t	-15808.70	31639.40	31715.65	31665.57
Panel B: *(RV+Dem)*				
GARCH MIDAS-N	-16084.92	32181.84	32223.43	32196.12
GARCH MIDAS-t	-15847.74	31709.48	31758.00	31726.13
RS GARCH MIDAS-N	-15870.03	31760.06	31829.38	31783.85
RS GARCH MIDAS-t	-15810.89	31643.78	31720.03	31669.95

Notes: This Table reports the log-likelihood values, AIC, BIC and HQIC statistics for different model specifications. In panels A and B the long-term component is driven by RV and explanatory variables, production and demand. The computation of AIC, BIC and HQIC is given in Equation (5.2). The highest log-likelihood and lowest AIC, BIC and HQIC are shown in red.

The overall results indicate that the monthly rates of production level have positive significant impact, whereas the demand level has a negative and insignificant impact on the WTI long-term volatility in both innovations. Moreover, as mentioned previously, the models with t innovations have higher log-likelihood values compared to models with Gaussian innovations.

Next, in order to assess the performance of the models with respect to how well it describes the crude oil data, both log-likelihood values and model selection criterion AIC, BIC and HQIC, which are presented in Table 5.7, are analysed.

We have divided the table into 2 panels for ease of interpretation. Panel A contains models with production as a macroeconomic variable whereas the models with demand are given in panel B. The red values indicate the lowest values amongst all models. Unsurprisingly, the regime switching models beat their single-regime counterparts regardless of any evaluation criteria. Similarly, models with Student-t innovations are superior compared to Gaussian. According to these results, the crude oil volatility is best described by RS GARCH MIDAS-t (RV + Prod). Since the log-likelihood values are similar, we also apply the LR test to see the relative performance of RS GARCH MIDAS-t and RS GARCH MIDAS-t (RV+Prod) model. The test statistic obtained is 0.32 and the corresponding p-value is 0.852, thus we fail to reject the null hypothesis.

Therefore, we can conclude that our proposed model RS GARCH MIDAS-t with RV only is still the best amongst all models. This result does not coincide with Pan et al. (2017) who found that both oil supply and demand have significant impacts on oil price volatility whereas our finding indicates that demand does not have a significant impact on the oil price volatility. This could again be because we used absolute returns as realised volatility.

An interesting finding is that ρ in RS GARCH MIDAS-t is higher than 0.85, whereas it much lower in RS GARCH MIDAS-t (RV+Prod) at 0.7556 and slightly lower in RS GARCH MIDAS-t (RV+Dem) at 0.8374. This implies that both production and demand are potential sources of volatility persistence. Similar result was shown by Pan et al. (2017).

5.3 Out-of-sample forecast performance

In the real world investors are more concerned about out-of-sample (OOS) forecasting performance compared with in-sample estimation since out-of-sample is more likely to show how the model behaves in the future. In our case we compare RS GARCH MIDAS-t with other models in terms of their out-of-sample forecasting performance.

To forecast the volatility out-of-sample, we predict it from the model fitted to all previous values. Hence all the predictions are one-step-ahead forecasts. Furthermore we use a rolling window approach to keep the sample size same. The forecasting horizon is from January 1, 2016 to July 27, 2020 where the model is refitted and parameters are re-estimated every month (22 days), as new information becomes available. It is common practice in the literature to use monthly frequency to re-estimate the model parameters. Furthermore, because of the long-term volatility component that changes monthly, it would be unlikely that forecasts will change much if we re-estimated parameters daily. Since the volatility is unobservable we compare the accuracy of our out-of-sample volatility forecasts from different models against the volatility proxy which is the daily squared returns.

The choice of loss functions and forecasting criteria is crucial for assessing the predictive accuracy of volatility models. In this study, we adopt the Mean Squared

Error (MSE) and QLIKE as our primary loss functions, following the findings of Patton (2006) who demonstrated their robustness to imperfect volatility proxies.

The MSE measures the average squared difference between the predicted volatility and the realized volatility. By comparing the MSE and QLIKE statistics across different models, we can assess their relative forecasting accuracy. Lower values of these statistics indicate better model performance. It is important to note that perfect forecasts would result in both the MSE and QLIKE being equal to zero. For a more detailed discussion on the rationale behind using MSE and QLIKE as evaluation criteria, please refer to Section 3.6.

In Table 5.8, we present the forecasting performances of different models using these two loss functions. Panel A considers models with Gaussian innovations, while panel B assumes Student-t distributed innovations and incorporates only the RV as the long-term volatility component. In panels C and D, we extend the models to include additional macroeconomic variables, namely, production or demand levels, to capture potential influences on volatility. These panels provide insights into the forecasting performance when considering the combined effects of RV and macroeconomic factors.

In the analysis of forecasting performance, two key factors are observed to influence the results: the choice of distribution and the loss function employed. It is evident that the RS GARCH model with Student-t innovations consistently outperforms all other models under both the MSE and QLIKE loss functions. It is important to note that all models with t innovations outperforms its Gaussian counterpart in both loss functions. This suggests that the inclusion of Student-t innovations improves the accuracy of volatility forecasting compared to models with Gaussian distribution. On the other hand, when comparing the RS GARCH MIDAS-t model with the GARCH model under MSE, the former does not exhibit superior performance. However, under the QLIKE loss function, RS GARCH MIDAS-t outperforms the GARCH model. This difference in performance may be attributed to the inclusion of the MIDAS component in the RS GARCH MIDAS-t model, which updates the long-term volatility every 22 days. This feature allows for a more dynamic and lagged effect of realized volatility, potentially enhancing forecasting accuracy.

Table 5.8: Forecasting performances of the models.

	MSE	QLIKE
Panel A		
GARCH-N	6669.9470	2.8064
RS GARCH-N	6661.7790	2.7644
GARCH MIDAS-N	43802.2400	2.8173
RS GARCH MIDAS-N	10282.6700	2.7898
Panel B		
GARCH-t	6736.0980	2.8186
RS GARCH-t	**6534.5100**	**2.7579**
GARCH MIDAS-t	13396.1900	2.8078
RS GARCH MIDAS-t	7744.8110	2.7836
Panel C: *(RV + Prod)*		
GARCH MIDAS-N	35157.8400	2.7901
RS GARCH MIDAS-N	18515.2700	2.8487
GARCH MIDAS-t	12712.7500	2.7865
RS GARCH MIDAS-t	10177.6400	2.7848
Panel D: *(RV + Dem)*		
GARCH MIDAS-N	33583.2000	2.8155
RS GARCH MIDAS-N	14791.0500	2.9969
GARCH MIDAS-t	12534.0600	2.8095
RS GARCH MIDAS-t	8822.3040	2.7852

Notes: This Table reports the evaluation results based on the loss functions of MSE and QLIKE given in Equation (3.74) for different model specifications. The forecasts are obtained via rolling window approach where the parameters are re-estimated every 22 steps. The daily squared returns are taken as volatility proxy. In panels A and B the innovations are Gaussian and Student-t distributed and τ_t is described by realised volatility only. In panels C and D the long-term component is driven by RV and explanatory variables, production and demand. Numbers in bold indicate the lowest loss function values.

To compare further the models with long-term volatility and test for the significance of the observed differences in the forecast accuracy, we employ the equal predictive ability test proposed by Diebold and Mariano (1995) on the equality of the MSE and QLIKE. The benchmark model chosen for comparison is RS GARCH MIDAS-t model, because of the following reasons. Firstly, within sample results indicate that this model outperforms all other models by having the highest LL, and smallest model selection criteria, see Table 5.4. Second reason is utilizing RS GARCH-MIDAS-t model, provides an appropriate setting to forecast high frequency oil market volatility using global predictors that are only available at low frequency.

The pair-wise comparison of the models, along with the corresponding p-values is presented in Table 5.9. Two forecasts have equal predictive accuracy if and only if the loss differential has zero expectation for all t. For the specific results and p-values of the Diebold-Mariano tests comparing forecasts within the same innovation and across different innovations, please refer to Appendix C.2, C.3 and C.4.

A preliminary examination of the DM test results reveals that most of the calculated p-values are statistically significant. This indicates that the forecasts generated by different models exhibit varying levels of forecasting accuracy when compared to the forecasts produced by the benchmark model.

When comparing models with a long-term volatility component, such as GARCH MIDAS, RS GARCH MIDAS models with only realized volatility, and GARCH MIDAS, RS GARCH MIDAS models with both RV and macroeconomic variables, the results in panels C and D of Table 5.8 indicate that models with Student-t innovations consistently outperform models with normal distribution in terms of both lower MSE and QLIKE values. Additionally, comparing the MSE values, models incorporating regime switching demonstrate superior performance compared to their single-regime counterparts. However, none of the models with macroeconomic variables are able to outperform the RS GARCH MIDAS-t model with RV only. This finding is consistent with the result of Conrad and Loch (2015) also demonstrated that models incorporating macroeconomic variables tended to perform less effectively compared to the benchmark GARCH MIDAS model with RV only, when forecasting long-term

Table 5.9: Diebold-Mariano test p-values.

	MSE	QLIKE
GARCH MIDAS-N	0.0136**	0.0437**
GARCH MIDAS-t	0.0006***	0.0047***
RS GARCH MIDAS-N	0.1180	0.0039***
GARCH MIDAS-N (RV + Prod)	0.0161**	0.0050***
GARCH MIDAS-t (RV + Prod)	0.0008***	0.0007***
RS GARCH MIDAS-N (RV + Prod)	0.0058***	0.0013***
RS GARCH MIDAS-t (RV + Prod)	0.0195**	0.9735
GARCH MIDAS-N (RV + Dem)	0.0077***	0.2290
GARCH MIDAS-t (RV + Dem)	0.0009***	0.0066***
RS GARCH MIDAS-N (RV + Dem)	0.0399**	0.0000***
RS GARCH MIDAS-t (RV + Dem)	0.0581*	0.7229

Notes: This table presents the p-values from Diebold-Mariano test of equal predictive accuracy using RS GARCH MIDAS-t with RV only forecasts as a benchmark.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

stock market volatility over a quarter horizon. Additionally, Fang et al. (2020), also presented compelling evidence suggesting that the significance of macroeconomic indicators in explaining stock market volatility may have been overstated. Their empirical analysis revealed that among three variables, namely, housing starts, default spread and realised volatility, RV emerged as the most powerful predictor of the long-term stock market volatility. On the contrary, these findings contradict the results of Pan et al. (2017). This discrepancy could be attributed to the fact that for out-of-sample forecasting, we utilized absolute returns and fat-tailed innovations, which differ from the methodology employed by Pan et al. (2017). Furthermore, the results might change if other macroeconomic variables are considered. It is also important to note that the economy was deeply affected by COVID and our dataset only contains information of the first couple of month, so in order to better analyse the crude oil volatility and the effect of macroeconomic variables during the COVID period more data would be needed.

In Figure 5.9, we present the volatility forecasts generated by the RS GARCH and RS GARCH MIDAS models, along with the daily volatility proxy of WTI oil returns.

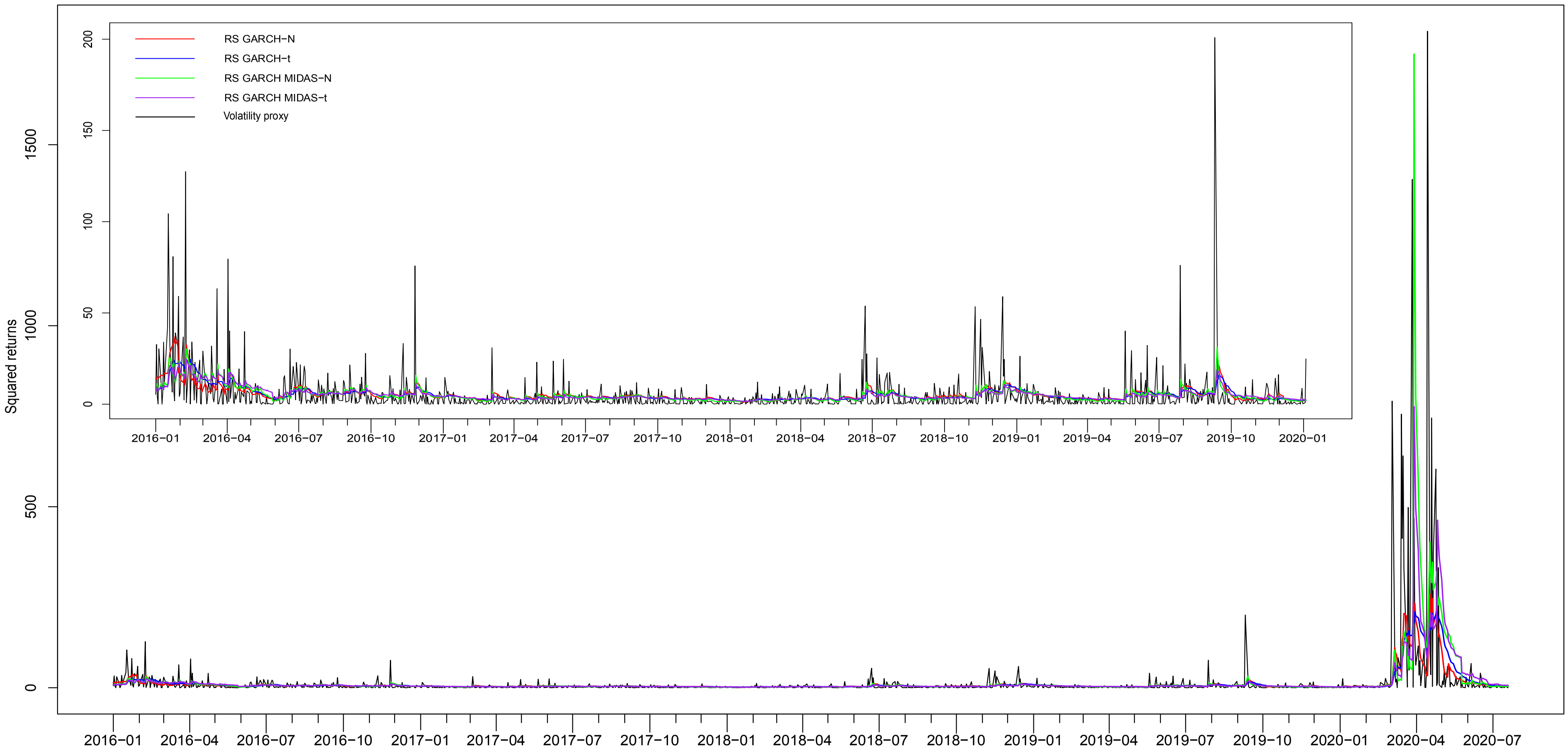
Please refer to Appendix C.5 for the plot comparing GARCH and RS GARCH forecasts. From the plot, it is evident that the highest volatility occurs between January 2020 and July 2020, which coincides with the period of the COVID-19 pandemic causing a significant drop in WTI crude oil prices. In contrast, the period from January 2016 to January 2020 exhibits relatively lower volatility. During the period of stable volatility, all forecasts exhibit a similar pattern, hence, to assess the differences between the forecasts generated by the RS GARCH and RS GARCH MIDAS models our focus is primarily on the high-volatility period in 2020.

We plot forecasts generated by RS GARCH, GARCH MIDAS and RS GARCH MIDAS-N, during high volatility periods in Figures 5.10 and 5.11 where RS GARCH MIDAS-t acts as the benchmark model[5]. Comparing two-regime GARCH models with the benchmark model we can see that RS GARCH models tend to under-predict the volatility during the high-volatility period. It also shows GARCH MIDAS-N model heavily over-predicts the volatility, while RS GARCH MIDAS can generate the forecasts closer to the true values. This is because RS GARCH MIDAS is more flexible in term of regime changes in volatility. Furthermore, changing the innovations to Student-t can also reduce the over-prediction of these models significantly. It also shows that GARCH MIDAS models are likely to increase the lagged effect of RV compared to RS GARCH MIDAS models.

The above results highlights the importance of correctly specifying the error distribution and that misspecified models can lead to over-prediction of volatility.

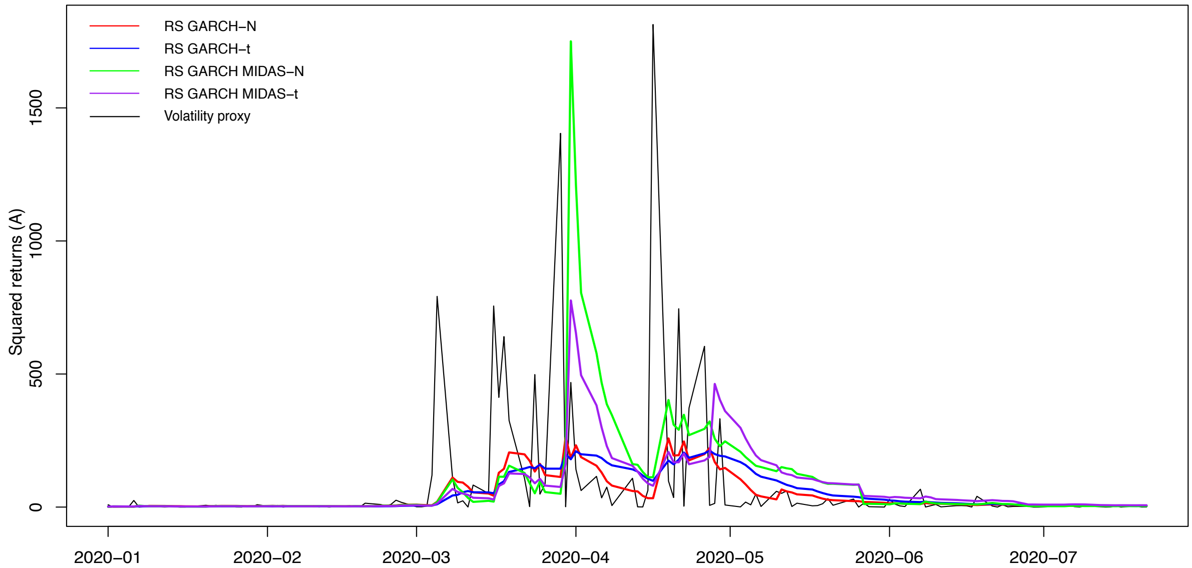
[5]Other forecast comparisons for the high volatility period can be found in Appendix C.6 and C.7.

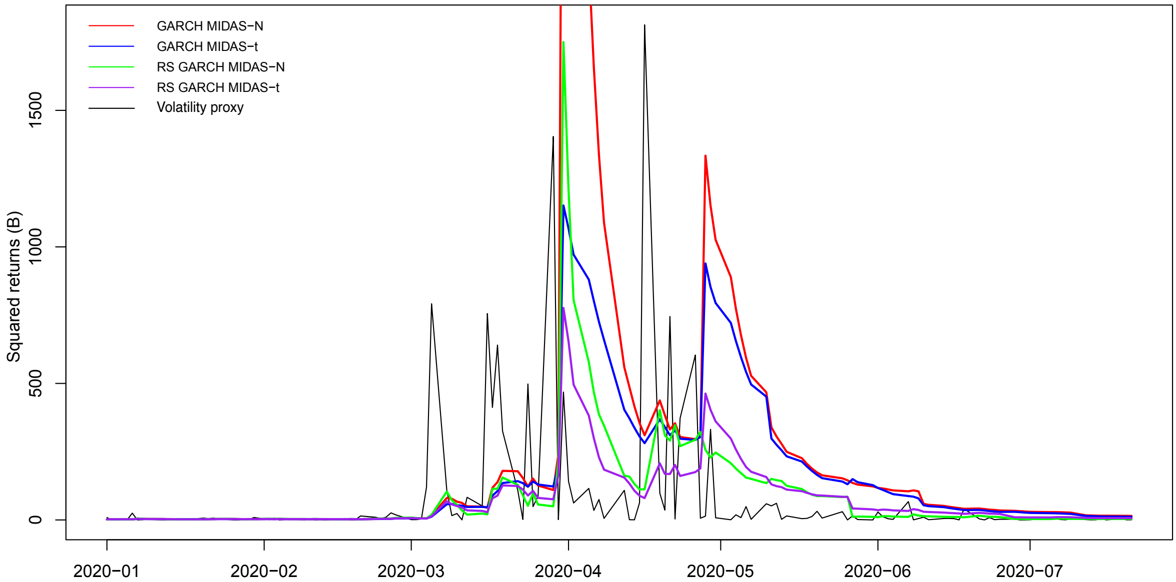
Figure 5.9: Out-of-sample forecasting comparison of RS GARCH and RS GARCH MIDAS models for 2016 - 2020.



Notes: This Figure shows the one-step-ahead out-of-sample volatility forecasts of WTI crude oil. The volatility forecasts are generated by RS GARCH-N, RS GARCH-t, RS GARCH MIDAS-N and RS GARCH MIDAS-t with RV as a macroeconomic variable for the later two models. The inner Figure is the closer plot for the period January 2016 - January 2020.

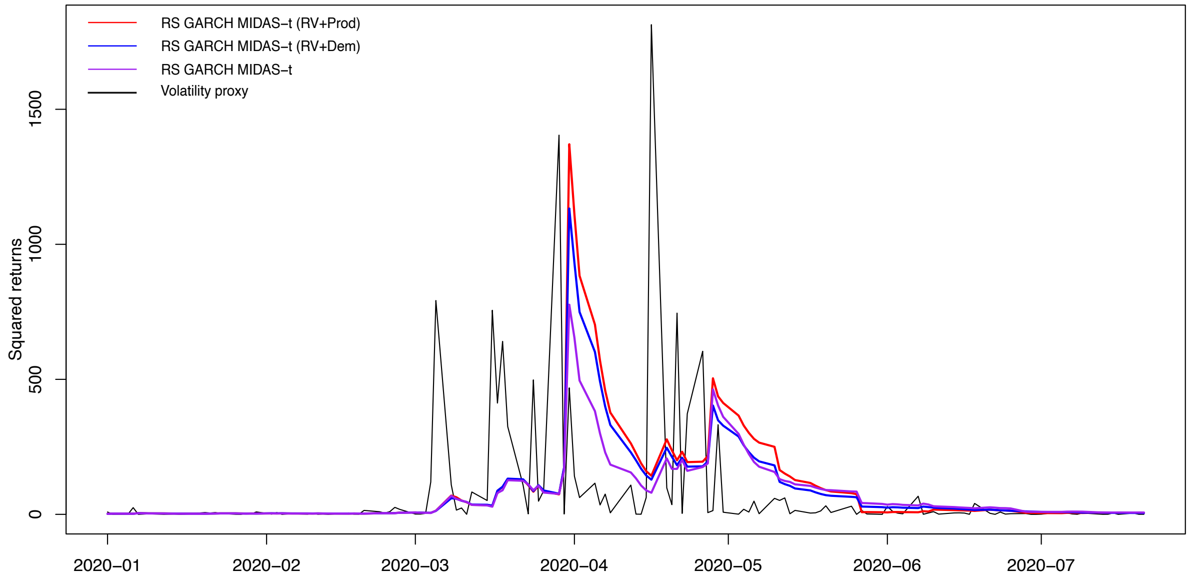
Figure 5.10: Out-of-sample forecasting comparison of models with RV only.

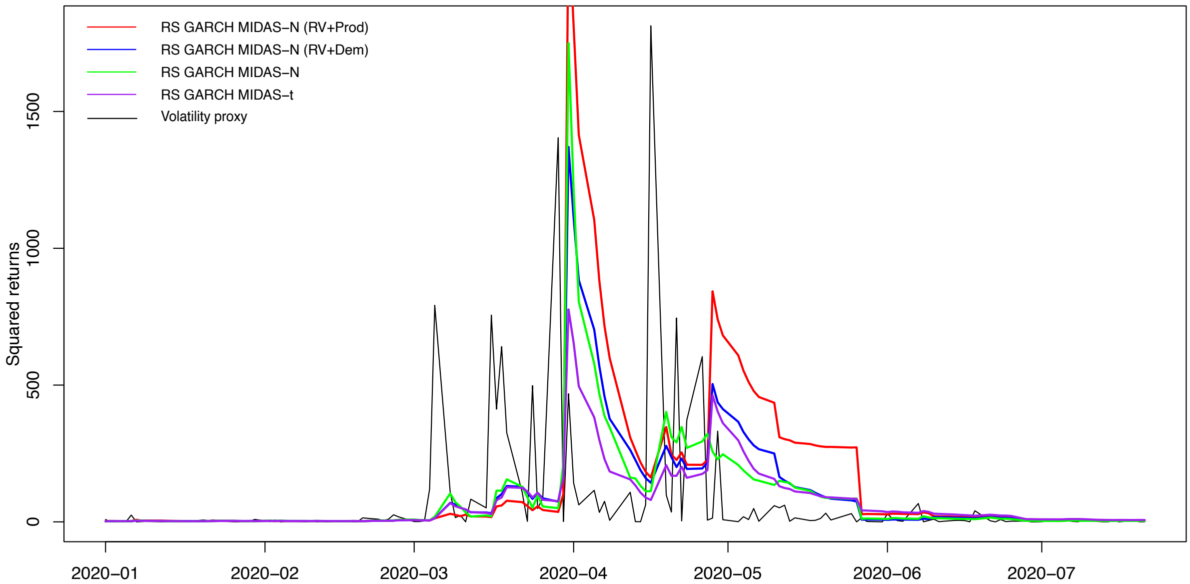




Notes: This Figure shows the volatility forecasts of WTI crude oil. The volatility forecasts are generated by two-regime GARCH models, single-regime GARCH MIDAS and RS GARCH MIDAS with normal and Student-t innovations where τ_t depends on RV only. The models are compared with the benchmark model RS GARCH MIDAS-t.

Figure 5.11: Out-of-sample forecasting comparison of models with macroeconomic variables.





Notes: This Figure shows the volatility forecasts of WTI crude oil. The volatility forecasts are generated by single-regime and two-regime GARCH MIDAS models where the macroeconomic information is realised volatility and production level (RV+Prod). The models are compared with the benchmark model RS GARCH MIDAS-t with RV only.

Chapter 6

Conclusion and Further research

In this thesis, the focus was on analysing the effectiveness of a regime-switching GARCH MIDAS model with non-Gaussian innovations in capturing volatility patterns observed in financial time series. A key emphasis of the research was to identify and understand how different components of the model capture changes in volatility over time and to demonstrate the importance and consequences of potential misspecification. We further assessed how well this model describes the fluctuations in volatility, compared to GARCH, RS GARCH, GARCH MIDAS, RS GARCH MIDAS-N, Endogeneous RS GARCH and Endogeneous RS GARCH MIDAS models with Gaussian and Student-t innovations. By investigating the consequences of misspecification and evaluating the model's performance against alternative approaches, the thesis aimed to shed light on the importance of choosing an appropriate volatility modelling framework.

The majority of prior research in the literature has been built on the assumption of a normal distribution, which does not adequately capture the characteristics of thick tails and spikes commonly found in financial data. Consequently, these findings may not fully reflect real market conditions. In contrast, a heavy-tailed distribution may be better suited to accurately describe the heavy tail behaviour observed in the data, making it a more realistic representation of financial data distribution. Therefore, the concept of employing an appropriate distribution to accommodate excess kurtosis has become crucial along with a correct model specification.

This thesis contributes to the current literature in several ways. First, we start with demonstrating the importance and consequences of potential misspecification through

Monte Carlo simulation. We considered the misspecification in terms of not considering regime switching, misspecifying the error term, omitting the long-term volatility component, or all three combined. By evaluating the finite-sample performance of QMLE in a Monte Carlo simulation we showed that QMLE is unbiased and the asymptotic standard errors are valid for general application.

Our findings of the Monte Carlo simulation can be summarised as follows. Firstly, we observed that the correctly specified model, RS GARCH MIDAS-t, exhibited significantly lower bias compared to the misspecified models. Next, the RS GARCH and RS GARCH MIDAS models with misspecified innovations had problems with accurately identifying volatility regimes. Gaussian specification in regime switching models could not correctly identify high- and low-volatility regimes, which resulted in larger parameter estimates for the ω_2 parameter and more frequent regime changes due to a lower p_{22} parameter. Interestingly, both GARCH and GARCH MIDAS models with misspecified innovations showed similar parameter estimate results, but the omission of regime switching had a notable impact on the GARCH MIDAS models. It led to an overestimation in the long-term volatility component. Ultimately, our simulation highlighted the crucial importance of correctly specifying the model, which in our case involved incorporating long- and short-term volatility components, regime switching components, and a fat-tailed error distribution.

In an empirical application to West Texas Intermediate crude oil returns, we compared the in-sample estimation and forecast performance of RS GARCH MIDAS-t model with a wide range of competitor models, such as GARCH, RS GARCH, GARCH MIDAS, Endogeneous RS GARCH and Endogeneous RS GARCH MIDAS. This analysis emphasizes the significance of correctly specifying the error distribution within the volatility models. In addition, following Ding et al. (1993) and Taylor (2008), we explored the advantages of calculating realised volatility as absolute value of returns.

The results obtained from the empirical analysis indicate that models incorporating Student-t innovations outperform those with Gaussian counterparts in terms of model selection criteria and regime identification. This suggests that the use of Student-t innovations is more effective in capturing the characteristics of crude oil market volatility. Moreover, the regime switching models with t innovations were able to capture major events in the crude oil history, such as the Gulf War in 1990, the 1996 backwardation in the oil market, the Asian financial crisis in 1997, the 2001 terrorist attack, and the U.S. invasion of Iraq in 2003, the financial crisis in 2008. Additionally, we found that models without long-term volatility component indicate a high degree of volatility persistence, hence we can conclude that accounting for long-term component could reduce the persistence in the short-term component. Furthermore, the analysis revealed that the RV and production had a significant positive effect on long-term volatility, while demand had an insignificant negative effect. Most importantly, among all models, the crude oil volatility was best captured by RS GARCH MIDAS-t.

For out-of-sample forecasting evaluation two loss functions, namely MSE and QLIKE were utilized along with DM test. The findings reveal that, overall, models with t innovations have lower MSE and QLIKE values compared to their Gaussian counterparts. Interestingly RS GARCH-t model performs best under both loss functions. In addition, while considering models with long-term volatility component only, we find that RS GARCH MIDAS-t with RV achieves the lowest MSE and QLIKE compared to models with macroeconomic variables, thus indicating that macroeconomic variables do not provide useful information regarding future oil volatility. This finding coincides with work of Conrad and Loch (2015), but contradicts to Pan et al. (2017).

We would like to conclude this thesis by outlining some limitations and ideas for future work. In empirical application we only considered Student-t distribution as a representative of fat tailed innovations. To generalise this, one might consider using other non-Gaussian innovations, such as skewed Student-t or GED. We also assumed that the long-term volatility component is fixed for all i in period t, however one can specify it to change daily or switch regimes similar to recent studies by Ma et al. (2021) and Wang et al. (2022) who allowed short-term, long-term or both terms to switch regimes in RS GARCH MIDAS models. A proper specification test could be developed to identify the necessity of regime switching in different components of RS GARCH MIDAS models. Furthermore, we considered endogenous regime switching model proposed by Choi (2009) where we assumed that current transition probabilities depend on the previous returns. However, Kim et al. (2008) and recently Chang et al. (2017) introduced several other ways to specify endogeneity in regime switching process. Hence, another possible extension could be to consider using a different specification of endogeneity in regime switching process. Additionally, in this thesis, we focused on the inclusion of two macroeconomic variables. However, there is a wide range of other factor that could be considered, such as financial uncertainty, default yield spread, GEPU, EPU, etc.

In recent years, there has been an increasing interest in applying machine learning techniques, specifically artificial neural networks, to financial volatility forecasting. For example, Bildirici and Ersin (2014) enhanced RS GARCH type models by incorporating artificial neural networks, to improve forecasting accuracy. Given that the current empirical literature primarily relies on econometric models for analysis, future studies could benefit from incorporating machine learning techniques to achieve even higher forecasting accuracy. Utilizing machine learning techniques can assist in identifying which macroeconomic variables are most relevant and important among the various options available. By expanding the scope of variables and leveraging machine learning methods, we can gain deeper insights into the factors that drive volatility and improve the accuracy of volatility forecasting in financial markets.

Bibliography

Abbara, O., Zevallos, M., 2023. Estimation and forecasting of long memory stochastic volatility models. Studies in Nonlinear Dynamics and Econometrics 27, 1–24.

Aloui, R., Gupta, R., Miller, S.M., 2016. Uncertainty and crude oil returns. Energy Economics 55, 92–100.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andreou, E., Ghysels, E., 2002. Detecting multiple breaks in financial market volatility dynamics. Journal of Applied Econometrics 17, 579–600.

Ardia, D., 2009. Bayesian estimation of a markov-switching threshold asymmetric garch model with student-t innovations. The Econometrics Journal 12, 105–126.

Asgharian, H., Hou, A.J., Javed, F., 2013. The importance of the macroeconomic variables in forecasting stock return variance: A garch-midas approach. Journal of Forecasting 32, 600–612.

Awartani, B.M., Corradi, V., 2005. Predicting the volatility of the s&p-500 stock index via garch models: the role of asymmetries. International Journal of Forecasting 21, 167–183.

Barsky, R.B., Kilian, L., 2004. Oil and the macroeconomy since the 1970s. Journal of Economic Perspectives 18, 115–134.

Baumeister, C., Kilian, L., 2015. Understanding the decline in the price of oil since June 2014. CFS Working Paper Series 501. Center for Financial Studies.

Baumeister, C., Kilian, L., 2016. Forty years of oil price fluctuations: Why the price of oil may still surprise us. Journal of Economic Perspectives 30, 139–60.

Bauwens, L., Preminger, A., Rombouts, J.V., 2010. Theory and inference for a markov switching garch model. The Econometrics Journal 13, 218–244.

Bildirici, M., Ersin, Ö., 2014. Modeling markov switching arma-garch neural networks models and an application to forecasting stock returns. The Scientific World Journal 2014.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics 31, 307–327.

Bollerslev, T., 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics 69, 542–547.

Cai, J., 1994. A markov model of unconditional variance in arch. Journal of Business and Economic Statistics 12, 309–316.

Chang, Y., Choi, Y., Park, J.Y., 2017. A new approach to model regime switching. Journal of Econometrics 196, 127–143.

Choi, S., 2009. Regime-switching univariate diffusion models of the short-term interest rate. Studies in Nonlinear Dynamics and Econometrics 13.

Chou, R.Y., 1988. Volatility persistence and stock valuations: Some empirical evidence using garch. Journal of Applied Econometrics 3, 279–294.

Chuang, O.C., Yang, C., 2022. Identifying the determinants of crude oil market volatility by the multivariate garch-midas model. Energies 15.

Conrad, C., Custovic, A., Ghysels, E., 2018. Long- and short-term cryptocurrency volatility components: A garch-midas analysis. Journal of Risk and Financial Management 11, 1–12.

Conrad, C., Kleen, O., 2020. Two are better than one: Volatility forecasting using multiplicative component garch-midas models. Journal of Applied Econometrics 35, 19–45.

Conrad, C., Loch, K., 2015. Anticipating long-term stock market volatility. Journal of Applied Econometrics 30, 1090–1114.

Conrad, C., Loch, K., Rittler, D., 2014. On the macroeconomic determinants of long-term volatilities and correlations in u.s. stock and crude oil markets. Journal of Empirical Finance 29, 26–40.

Dees, S., Karadeloglou, P., Kaufmann, R.K., Sanchez, M., 2007. Modelling the world oil market: Assessment of a quarterly econometric model. Energy policy 35, 178–191.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Ding, Z., Granger, C.W., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of empirical finance 1, 83–106.

Dueker, M.J., 1997. Markov switching in garch processes and mean-reverting stock-market volatility. Journal of Business and Economic Statistics 15, 26–34.

Engle, R., Rangel, J., 2008. The spline-garch model for low-frequency volatility and its global macroeconomic causes. Review of Financial Studies 21, 1187–1222.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987–1007.

Engle, R.F., Ghysels, E., Sohn, B., 2008. On the economic sources of stock market volatility. International Finance .

Engle, R.F., Ghysels, E., Sohn, B., 2013. Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics 95, 776–797.

Engle, R.F., Mustafa, C., 1992. Implied arch models from options prices. Journal of Econometrics 52, 289–311.

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. The journal of finance 48, 1749–1778.

Engle, R.F., White, H., et al., 1999. Cointegration, causality, and forecasting: a Festschrift in Honour of Clive WJ Granger. Oxford University Press on Demand.

Fabozzi, F.J., Focardi, S.M., Rachev, S.T., Arshanapalli, B.G., 2014. The basics of financial econometrics: Tools, concepts, and asset management applications. John Wiley and Sons.

Fan, Y., Zhang, Y.J., Tsai, H.T., Wei, Y.M., 2008. Estimating 'value at risk' of crude oil price and its spillover effect using the ged-garch approach. Energy Economics 30, 3156–3171.

Fang, T., Lee, T.H., Su, Z., 2020. Predicting the long-term stock market volatility: A garch-midas model with variable selection. Journal of Empirical Finance 58, 36–49.

Fernández, C., Steel, M.F., 1998. On bayesian modeling of fat tails and skewness. Journal of the american statistical association 93, 359–371.

Fong, W.M., See, K.H., 2002. A markov switching model of the conditional volatility of crude oil futures prices. Energy Economics 24, 71–95.

Forsberg, L., Ghysels, E., 2006. Why do absolute returns predict volatility so well? Journal of Financial Econometrics 5, 31–67.

Francq, C., Zakoı, J.M., et al., 2005. The l2-structures of standard and switching-regime garch models. Stochastic processes and their applications 115, 1557–1582.

Francq, C., Zakoian, J.M., 2019. GARCH models: structure, statistical inference and financial applications. John Wiley and Sons.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. Journal of Econometrics 131, 59–95.

Ghysels, E., Sinko, A., Valkanov, R., 2007. Midas regressions: Further results and new directions. Econometric reviews 26, 53–90.

Giles, D.E., 2008. Some properties of absolute returns as a proxy for volatility. Applied Financial Economics Letters 4, 347–350.

Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance 48, 1779–1801.

Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27–62.

Gunay, S., 2015. Markov regime switching generalized autoregressive conditional heteroskedastic model and volatility modeling for oil returns. International Journal of Energy Economics and Policy 5, 979–985.

Guérin, P., Marcellino, M., 2013. Markov-switching midas models. Journal of Business and Economic Statistics 31, 45–56.

Haas, M., Liu, J.C., 2018. A multivariate regime-switching garch model with an application to global stock market and real estate equity returns. Studies in Nonlinear Dynamics and Econometrics 22, 2016–0019.

Haas, M., Mittnik, S., Paolella, M.S., 2004. A new approach to markov-switching garch models. Journal of financial Econometrics 2, 493–530.

Haas, M., Paolella, M.S., 2012. Mixture and Regime-Switching GARCH Models. John Wiley and Sons, Ltd.

Hafner, C.M., Preminger, A., 2010. Deciding between garch and stochastic volatility via strong decision rules. Journal of Statistical Planning and Inference 140, 791–805.

Hamilton, J., 1990. Analysis of time series subject to changes in regime. Journal of Econometrics 45, 39–70.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.

Hamilton, J.D., 2009. Causes and Consequences of the Oil Shock of 2007-08. Brookings Papers on Economic Activity 40, 215–283.

Hamilton, J.D., 2020. Time series analysis. Princeton university press.

Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64, 307–333.

Hansen, B.E., 1994. Autoregressive conditional density estimation. International Economic Review , 705–730.

Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a garch (1, 1)? Journal of applied econometrics 20, 873–889.

Herrera, A.M., Hu, L., Pastor, D., 2018. Forecasting crude oil price volatility. International Journal of Forecasting 34, 622–635.

Hillebrand, E., 2005. Neglecting parameter changes in garch models. Journal of Econometrics 129, 121–138.

Ho, T.k., et al., 2004. How useful are regime-switching models in banking crises identification? Graduate Institute of International Economics, National Chung Cheng University .

Horv, L., Kokoszka, P., et al., 2003. Garch processes: structure and estimation. Bernoulli 9, 201–227.

Hung, N.T., Thach, N.N., Anh, L.H., 2018. Garch models in forecasting the volatility of the world's oil prices, in: International econometric conference of Vietnam, pp. 673–683.

Iglesias, E.M., Rivera-Alonso, D., 2022. Brent and wti oil prices volatility during major crises and covid-19. Journal of Petroleum Science and Engineering 211, 110182.

Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics letters 6, 255–259.

Kalimipalli, M., Susmel, R., 2004. Regime-switching stochastic volatility and short-term interest rates. Journal of Empirical Finance 11, 309–329.

Kapetanios, G., 2001. Model selection in threshold models. Journal of Time Series Analysis 22, 733–754.

Kilian, L., 2009. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. American Economic Review 99, 1053–69.

Kim, C.J., 1994. Dynamic linear models with markov-switching. Journal of Econometrics 60, 1–22.

Kim, C.J., 2004. Markov-switching models with endogenous explanatory variables. Journal of Econometrics 122, 127–136.

Kim, C.J., 2009. Markov-switching models with endogenous explanatory variables ii: A two-step mle procedure. Journal of Econometrics 148, 46–55.

Kim, C.J., Nelson, C.R., 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. Review of Economics and Statistics 80, 188–201.

Kim, C.J., Piger, J., Startz, R., 2008. Estimation of markov regime-switching regression models with endogenous switching. Journal of Econometrics 143, 263–273.

Kim, D., Kon, S.J., 1999. Structural change and time dependence in models of stock returns. Journal of Empirical Finance 6, 283–308.

Klaassen, F., 2002. Improving garch volatility forecasts with regime-switching garch. Empirical Economics 27, 363–394.

Lamoureux, C.G., Lastrapes, W.D., 1990. Persistence in variance, structural change, and the garch model. Journal of Business and Economic Statistics 8, 225–234.

Le, T.H., Boubaker, S., Bui, M.T., Park, D., 2023. On the volatility of wti crude oil prices: A time-varying approach with stochastic volatility. Energy Economics 117, 106474.

Lee, S.W., Hansen, B.E., 1994. Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. Econometric theory 10, 29–52.

Liu, W., Morley, B., 2009. Volatility forecasting in the hang seng index using the garch approach. Asia-Pacific Financial Markets 16, 51–63.

Lumsdaine, R.L., 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in igarch (1, 1) and covariance stationary garch (1, 1) models. Econometrica: Journal of the Econometric Society , 575–596.

Ma, F., Lu, X., Wang, L., Chevallier, J., 2021. Global economic policy uncertainty and gold futures market volatility: Evidence from markov regime-switching garch-midas models. Journal of Forecasting 40, 1070–1085.

Ma, R., Zhou, C., Cai, H., Deng, C., 2019. The forecasting power of epu for crude oil return volatility. Energy Reports 5, 866–873.

Mabro, R., 2009. The oil price crises of 1998-9 and 2008-9, in: Oxford Energy Forum, Oxford Institute for Energy Studies.

Mandelbrot, B., 1963. The variation of certain speculative prices. The Journal of Business 36, 394–419.

Marcucci, J., 2005. Forecasting stock market volatility with regime-switching garch models. Studies in Nonlinear Dynamics and Econometrics 9.

Mattera, R., Giacalone, M., et al., 2018. Alternative distribution based garch models for bitcoin volatility estimation. The Empirical Economics Letters 17, 1283–1288.

Mazzeu, J.H.G., Veiga, H., Mariti, M.B., 2019. Modeling and forecasting the oil volatility index. Journal of Forecasting 38, 773–787.

Melino, A., Turnbull, S.M., 1990. Pricing foreign currency options with stochastic volatility. Journal of Econometrics 45, 239–265.

Mikosch, T., Stărică, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the igarch effects. The Review of Economics and Statistics 86, 378–390.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347–70.

Nguyen, D.K., Walther, T., 2020. Modeling and forecasting commodity market volatility with long-term economic and financial variables. Journal of Forecasting 39, 126–142.

Pan, Z., Wang, Y., Wu, C., Yin, L., 2017. Oil price volatility and macroeconomic fundamentals: A regime switching garch-midas model. Journal of Empirical Finance 43, 130–142.

Patton, A., 2006. Volatility Forecast Comparison using Imperfect Volatility Proxies. Research Paper Series 175. Quantitative Finance Research Centre, University of Technology, Sydney.

Poon, S.H., Granger, C.W., 2003. Forecasting volatility in financial markets: A review. Journal of Economic Literature 41, 478–539.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociological methodology , 111–163.

Schwert, G., 1988. Why Does Stock Market Volatility Change Over Time? National Bureau of Economic Research Working Papers 2798. National Bureau of Economic Research, Inc.

Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. European Journal of Operational Research 235, 643–659.

Takahashi, M., Watanabe, T., Omori, Y., 2021. Forecasting daily volatility of stock price index using daily returns and realized volatility. Econometrics and Statistics .

Taylor, S.J., 1994. Modeling stochastic volatility: A review and comparative study. Mathematical Finance 4, 183–204.

Taylor, S.J., 2008. Modelling financial time series. world scientific.

Theodossiou, P., 1998. Financial data and the skewed generalized t distribution. Management Science 44, 1650–1661.

Triacca, U., 2007. On the variance of the error associated to the squared return as proxy of volatility. Applied Financial Economics Letters 3, 255–257.

Tsay, R.S., 2005. Analysis of financial time series. John wiley and sons.

Ural, M., 2016. The impact of the global financial crisis on crude oil price volatility. Yönetim ve Ekonomi Araştırmaları Dergisi 14, 64.

Vo, M.T., 2009. Regime-switching stochastic volatility: Evidence from the crude oil market. Energy Economics 31, 779–788.

Wang, F., Ghysels, E., 2015. Econometric analysis of volatility component models. Econometric Theory 31, 362–393.

Wang, L., Wu, J., Cao, Y., Hong, Y., 2022. Forecasting renewable energy stock volatility using short and long-term markov switching garch-midas models: Either, neither or both? Energy Economics 111, 106056.

Wei, Y., Liu, J., Lai, X., Hu, Y., 2017. Which determinant is the most informative in forecasting crude oil market volatility: Fundamental, speculation, or uncertainty? Energy Economics 68, 141–150.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica: Journal of the Econometric Society , 1067–1084.

Wilhelmsson, A., 2006. Garch forecasting performance under different distribution assumptions. Journal of Forecasting 25, 561–578.

Xie, Y., 2009. Consistency of maximum likelihood estimators for the regime-switching garch model. Statistics 43, 153–165.

Yu, X., Huang, Y., 2021. The impact of economic policy uncertainty on stock volatility: Evidence from garch–midas approach. Physica A: Statistical Mechanics and its Applications 570, 125794.

Zakoian, J.M., 1994. Threshold heteroskedastic models. Journal of Economic Dynamics and Control 18, 931–955.

Zavadska, M., Morales, L., Coughlan, J., 2020. Brent crude oil prices volatility during major crises. Finance Research Letters 32, 101078.

Zhang, Y.J., Yao, T., He, L.Y., Ripple, R., 2019. Volatility forecasting of crude oil market: Can the regime switching garch model beat the single-regime garch models? International Review of Economics and Finance 59, 302–317.

Zhao, J., 2022. Exploring the influence of the main factors on the crude oil price volatility: An analysis based on garch-midas model with lasso approach. Resources Policy 79, 103031.

Appendix A

Definitions

In this appendix, some basic concepts such as stationarity, and some definitions that are needed in the main chapters are introduced.

A time series, x_t, is a set of random variables indexed by time t.

Definition A.0.1 A time series $\{x_t\}$ is said to be strictly stationary if for any time points $t_1, t_2, \ldots, t_n$ and any $s \in Z$, the joint distribution of $\{x_{t_1}, x_{t_2}, \ldots, x_{t_n}\}$ is identical to that of $\{x_{t_1+s}, x_{t_2+s}, \ldots, x_{t_n+s}\}$. In other words, strict stationarity requires that the joint distribution of $\{x_{t_1}, x_{t_2}, \ldots, x_{t_n}\}$ is invariant under time shift.

Definition A.0.2 A time series $\{x_t\}$ is said to be weakly stationary if
(1) $E(x_t) = \mu$, a constant independent of t, and
(2) $Var(x_t) = \sigma^2$, a constant independent of t, and
(3) $Cov(x_t, x_s)$ is a function of $s - t$ only, for any t and s.

Definition A.0.3 Suppose there are only two states and the transition probability matrix is

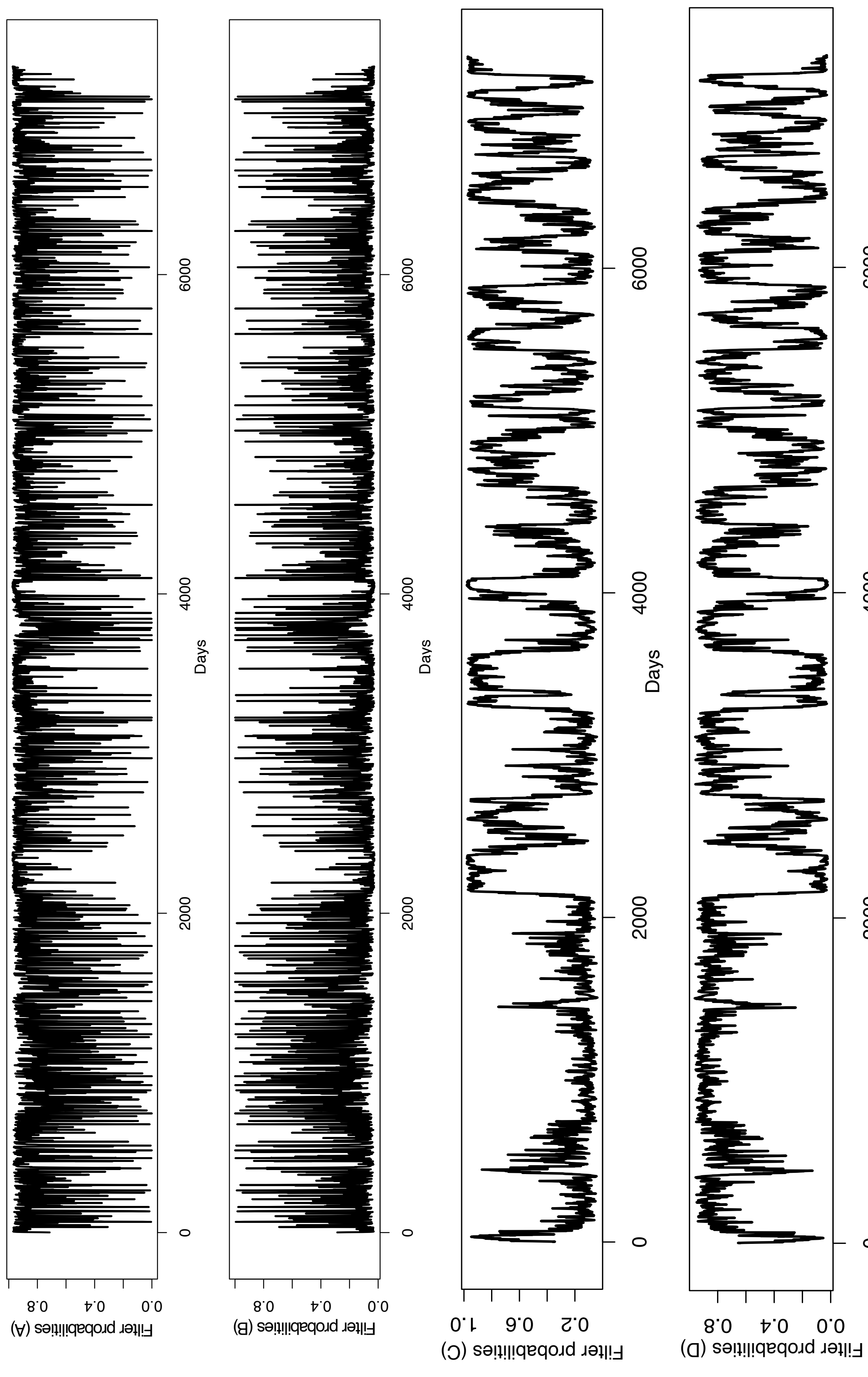
$$P = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22}. \end{bmatrix} \tag{A.1}$$

If $p_{11} = 1$, then $1 - p_{11} = 0$, meaning that once the process moves to state 1, there is no possibility of ever returning to state 2. In this case, state 1 would be considered as an absorbing state and the Markov chain is reducible. An irreducible Markov chain is the Markov chain that is not reducible. Therefore, a two-state Markov chain is irreducible if and only if $p_{11} < 1$ and $p_{22} < 1$.

Appendix B

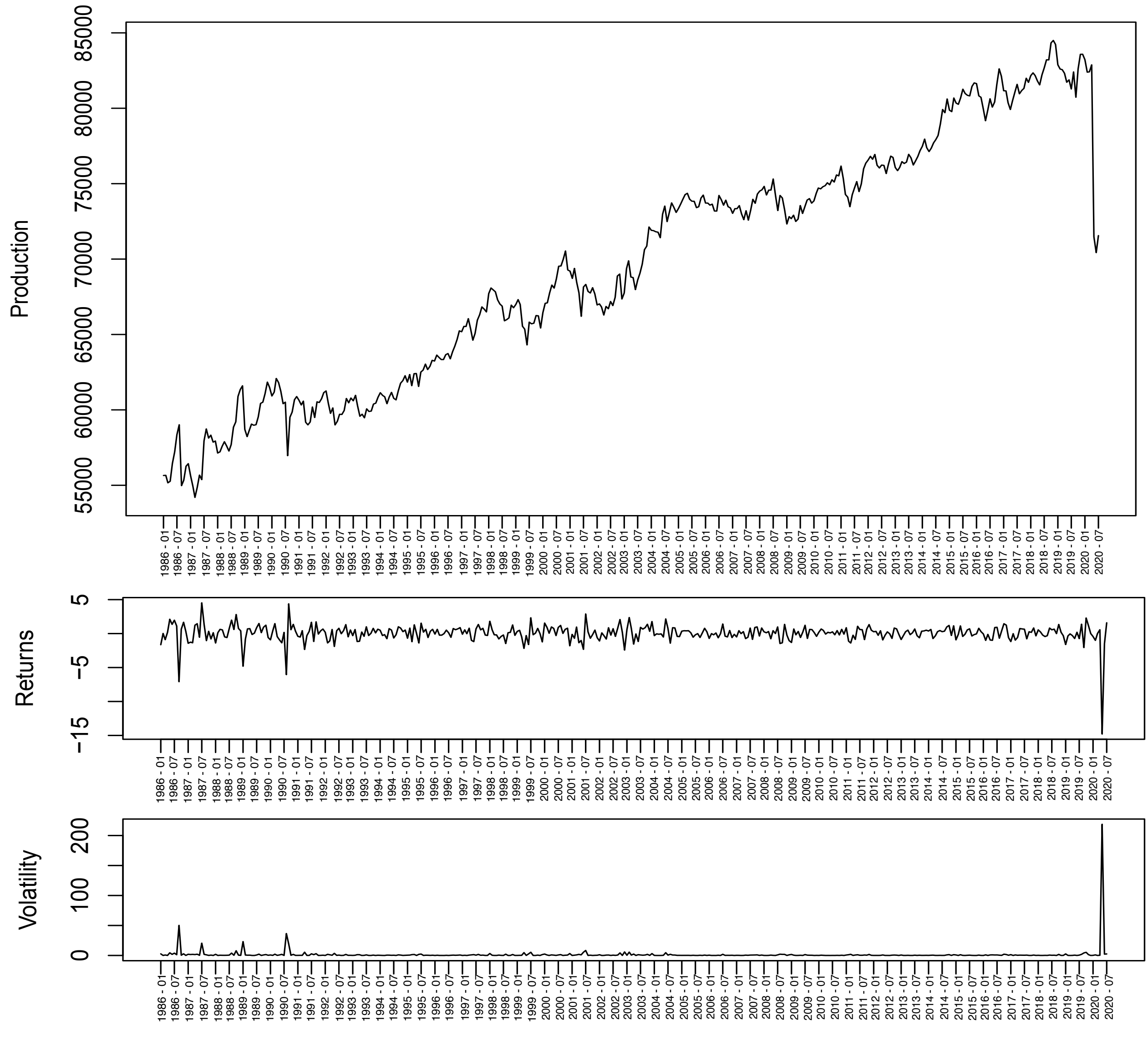
Supplementary materials for Chapters 4 and 5

Figure B.1: Filter probabilities for RS GARCH models.



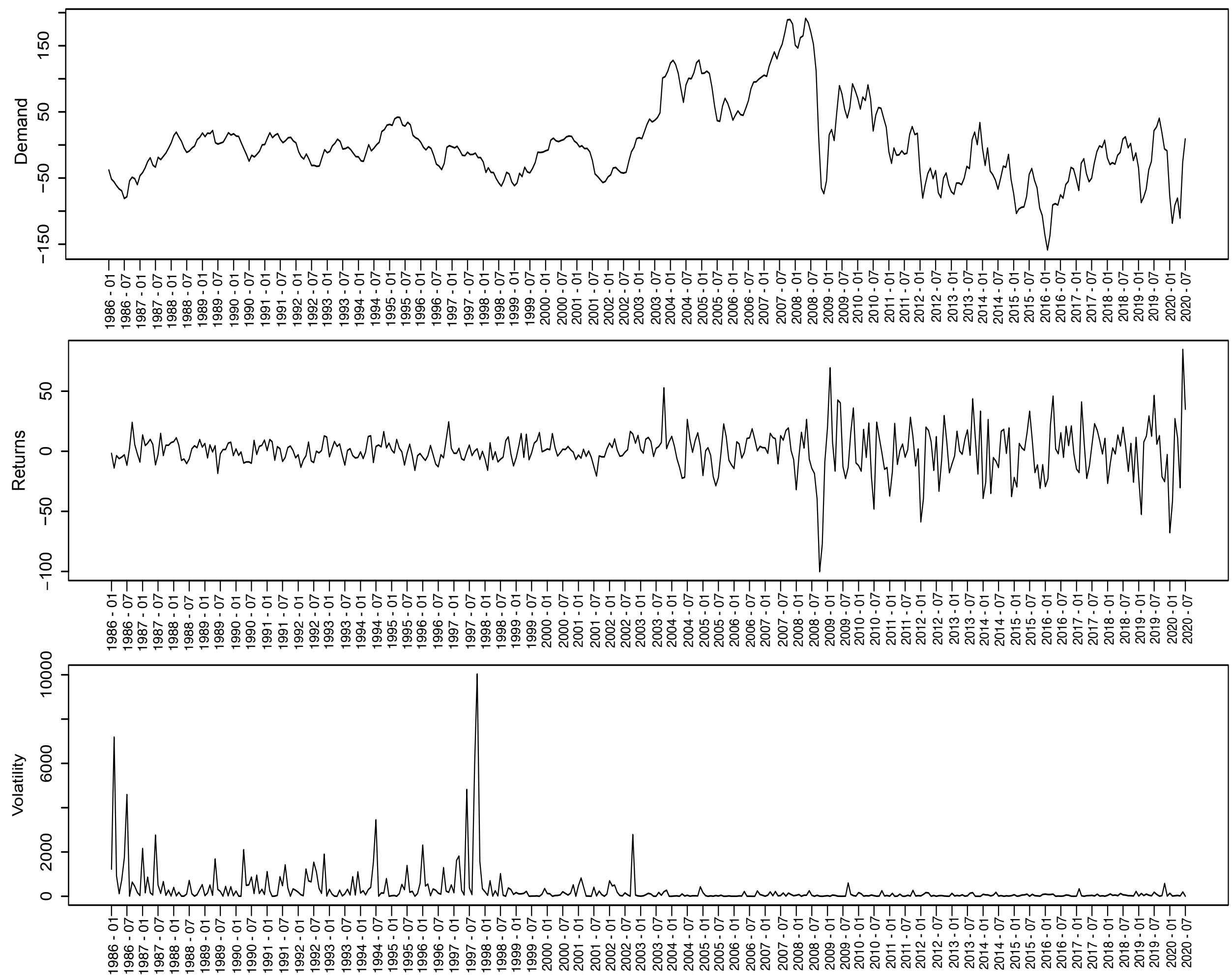
Notes: This Figure shows filter probabilities for RS GARCH model with Gaussian innovations. The (A) and (B) are the filter probabilities of being in low-volatility regime and high-volatility regime, respectively.

Figure B.2: WTI crude oil production levels.



Notes: This Figure shows global oil production obtained from EIA is used as a proxy of world oil supply. Plot (A) shows the production of crude oil of barrels per month, plot (B) shows the returns, and volatilities are displayed in plot (C) .

Figure B.3: WTI crude oil demand levels.



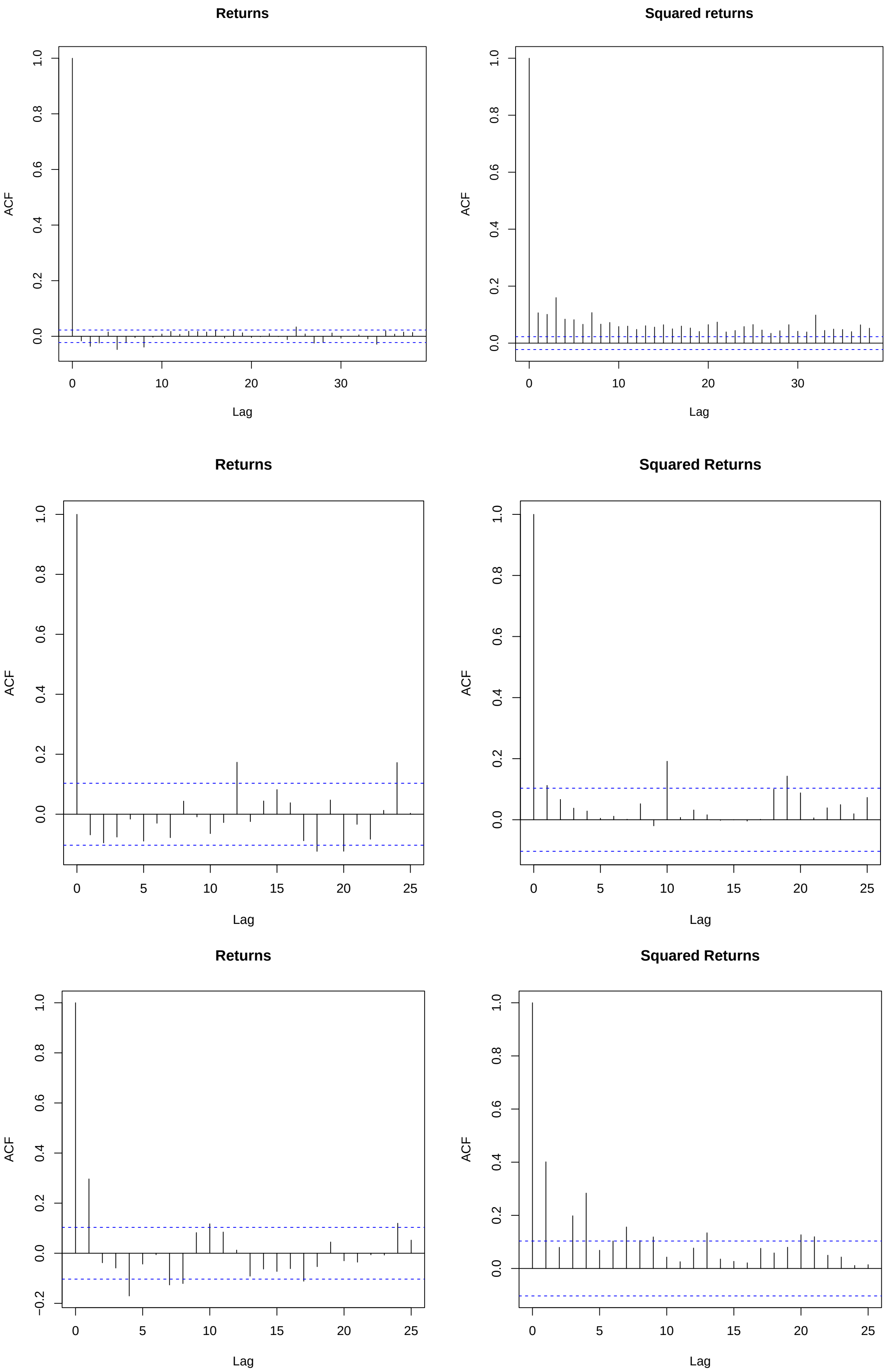
Notes: This Figure shows the crude oil demand level where index of Kilian (2009) is used as the signal for oil demand. Plot (A) shows the Kilian index per month, plot (B) shows the returns, and volatilities are displayed in plot (C).

Table B.1: Descriptive statistics of WTI crude oil for full sample.

	WTI	Production	Demand
Mean	0.0138	0.0566	0.1088
Var	7.4439	1.6269	298.9061
Min	-40.6396	-14.7902	-100.1826
Max	42.5832	4.5266	84.7775
Skewness	0.0151	-4.3704	-0.4667
Kurtosis	29.9217	49.7328	8.9368
JB ($\times 10^4$)	26.3138***	3.9484***	0.0634***
	(0.0000)	(0.0000)	(0.0000)
ADF(5)	-39.8320***	-8.8242***	-9.8954***
	(0.0100)	(0.0100)	(0.0100)
Q(5)	58.8780***	5.8412	85.3490***
	(0.0000)	(0.3220)	(0.0000)
$Q^2(5)$	1626.5000***	0.0817	120.9400***
	0.0000	(0.9999)	(0.0000)
ARCH(5)	1074.2000***	0.1960	70.7300***
	(0.0000)	(0.9992)	(0.0000)

Notes: This table presents the descriptive statistics for the full dataset from January 1986 to July 2020. Jarque-Bera (JB) Stat, augmented Dickey-Fuller (ADF) stat, Ljung-Box statistics, Q(5), and ARCH are the statistics testing for normal distribution, stationarity, serial correlation and heteroskedastic effects respectively. The corresponding *p*-values are given in brackets.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Figure B.4: ACF plots for log and squared returns of WTI crude oil, production and demand levels.



Notes: This Figure shows the ACF plots for WTI crude oil price (top), production (middle) and demand (bottom) returns and squared returns.

Table B.2: Diebold-Mariano test p-values comparison of innovations.

	GARCH-N	RS GARCH-N	GARCH MIDAS-N	RS GARCH MIDAS-N
GARCH-t	0.1731	0.8375	0.0214**	0.0481**
	0.2825	0.0230*	0.0031***	0.4493
RS GARCH-t	0.4496	0.5748	0.0120**	0.0369**
	0.0000**	0.4907	0.6593	0.1619
GARCH MIDAS-t	0.0002***	0.0006***	0.0253**	0.0767*
	0.0003**	0.1158	0.0111**	0.8325
RS GARCH MIDAS-t	0.0111*	0.0063***	0.0136*	0.1180
	0.0000**	0.6126	0.2612	0.0039***

Notes: This table presents the p-values from Diebold-Mariano test of equal predictive accuracy for one-day-ahead rolling window forecast over the period January 2016 to July 2020. The forecasts from the models with Gaussian innovation are compared with the forecasts from the models with Student-t innovations. The top value is p-value for MSE loss function while the bottom p-value is for QLIKE loss function. A p-value greater that 0.05 denotes the rejection of the null of equal predictive accuracy.

* denote the rejection of null hypothesis at 10% significance level.

** denote the rejection of null hypothesis at 5% significance level.

*** denote the rejection of null hypothesis at 1% significance level.

Table B.3: Diebold-Mariano test p-values for models with Gaussian innovations.

Gaussian	GARCH	RS GARCH	GM	RS GM	GM $(RV+Prod)$	RS GM $(RV+Prod)$	GM $(RV+Dem)$	RS GM $(RV+Dem)$
RS GARCH	0.9828	-	-	-	-	-	-	-
	0.0143*							
GM	0.0123*	0.0124*	-	-	-	-	-	-
	0.0005**	0.3428						
RS GM	0.0449*	0.0448*	0.0106**	-	-	-	-	-
	0.3643	0.0375*	0.1593					
GM	0.0140**	0.0142**	0.0122**	0.0117**	-	-	-	-
$(RV+Prod)$	0.0381**	0.1014	0.0107	0.7089				
RS GM	0.0042***	0.0045***	0.0185**	0.0010***	0.0272**	-	-	-
$(RV+Prod)$	0.6353	0.0616*	0.0909*	0.5286	0.3970			
GM	0.0066***	0.0068***	0.0417**	0.0048***	0.4903	0.0096***	-	-
$(RV+Dem)$	0.0007***	0.3255	0.1637	0.1868	0.0173**	0.1078		
RS GM								
$(RV+Dem)$								

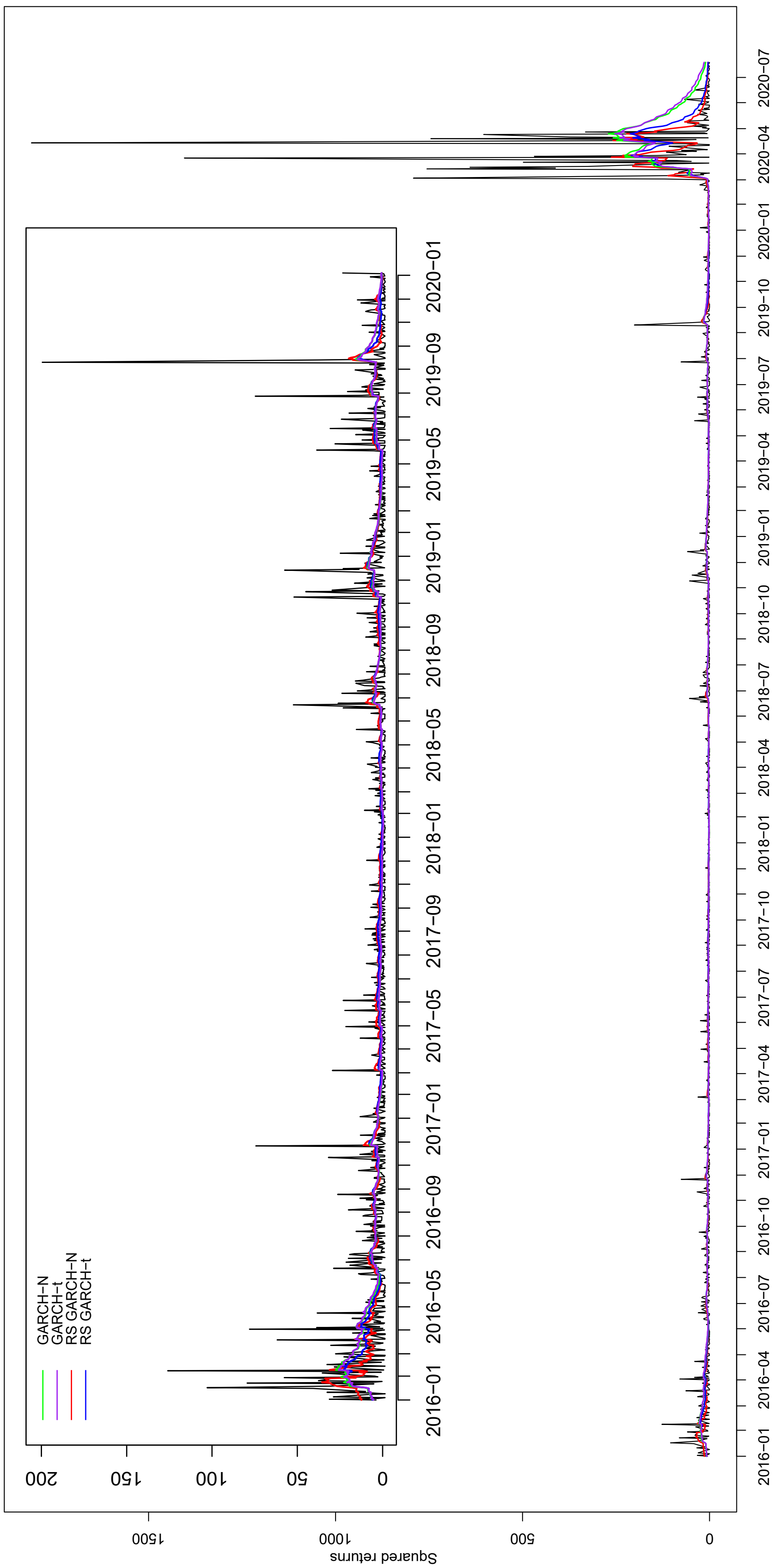
Notes: This table presents the p-values from Diebold-Mariano test of equal predictive accuracy. GM stands for GARCH MIDAS. The top value is the p-value for MSE loss function while the bottom p-value is for QLIKE loss function.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Table B.4: Diebold-Mariano test p-values for models with Student-t innovations.

Student-t	GARCH	RS GARCH	GM	GM ($RV + Prod$)	RS GM ($RV + Prod$)	GM ($RV + Dem$)	RS GM ($RV + Dem$)
RS GARCH	0.2075	-	-	-	-	-	-
	0.0000***						
GM	0.0214**	0.0003***	-	-	-	-	-
	0.0031***	0.0069***					
RS GM	0.0132**	0.0013***	0.0006***	-	-	-	-
	0.0000***	0.5434	0.0047***				
GM	0.0003***	0.0003***	0.0163**	-	-	-	-
($RV + Prod$)	0.1105	0.0000***	0.1324				
RS GM	0.0073***	0.0049***	0.0057***	0.0319**	-	-	-
($RV + Prod$)	0.0000***	0.5414	0.0010***	0.0000***			
GM	0.0004***	0.0004***	0.0249**	0.7333	0.0066***	-	-
($RV + Dem$)	0.0010***	0.0119**	0.0048***	0.0628*	0.0018***		
RS GM	0.0107**	0.0056***	0.0006***	0.0019***	0.0062***	0.0004***	-
($RV + Prod$)	0.0000***	0.3442	0.0000***	0.0000***	0.6668	0.0000***	

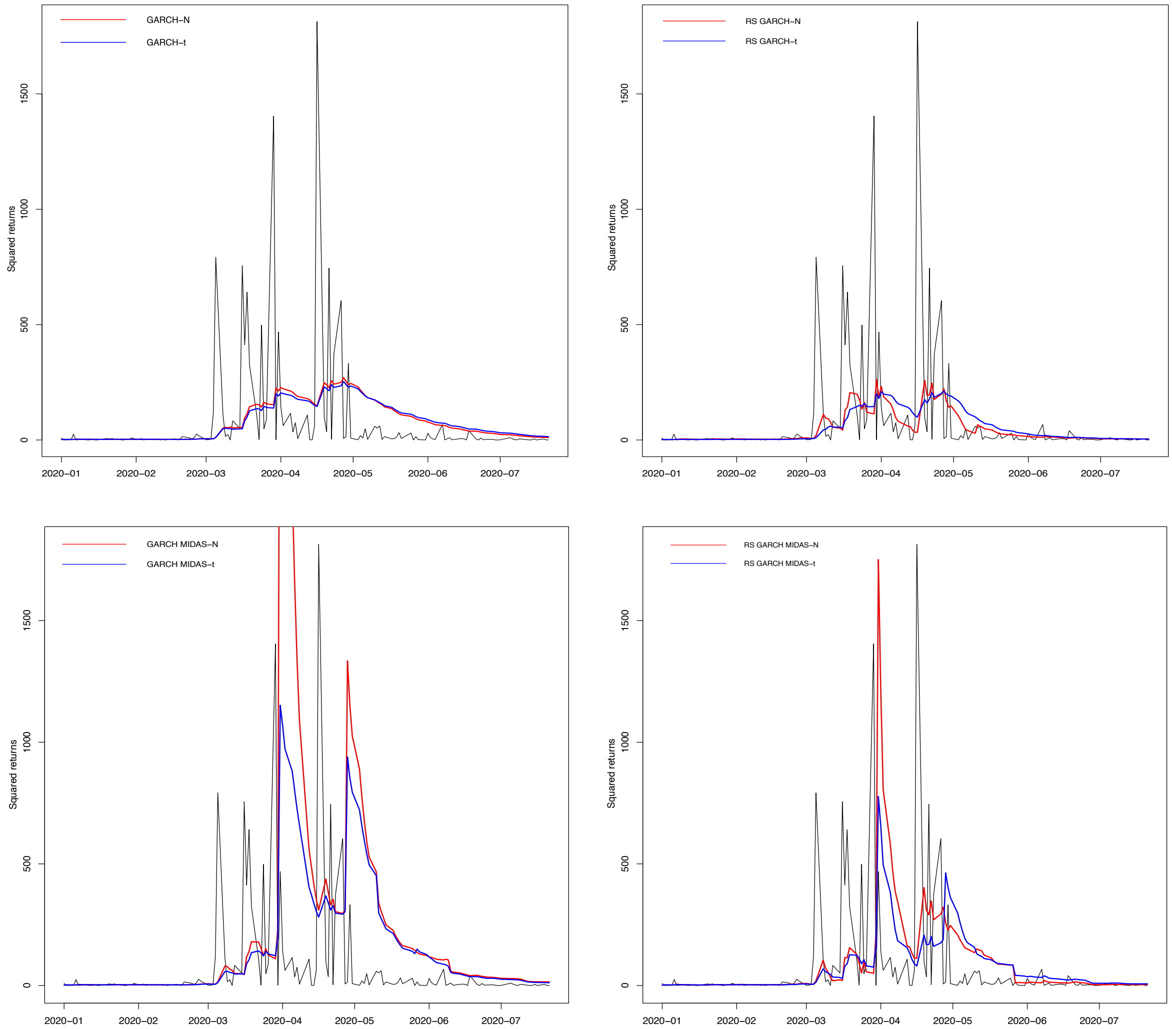
Notes: This table presents the p-values from Diebold-Mariano test of equal predictive accuracy. GM denotes GARCH MIDAS. The top value is p-value for MSE loss function while the bottom p-value is for QLIKE loss function. A p-value greater that 0.05 denotes the rejection of the null of equal predictive accuracy.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Figure B.5: Out-of-sample forecasting comparison of GARCH and RS GARCH models.



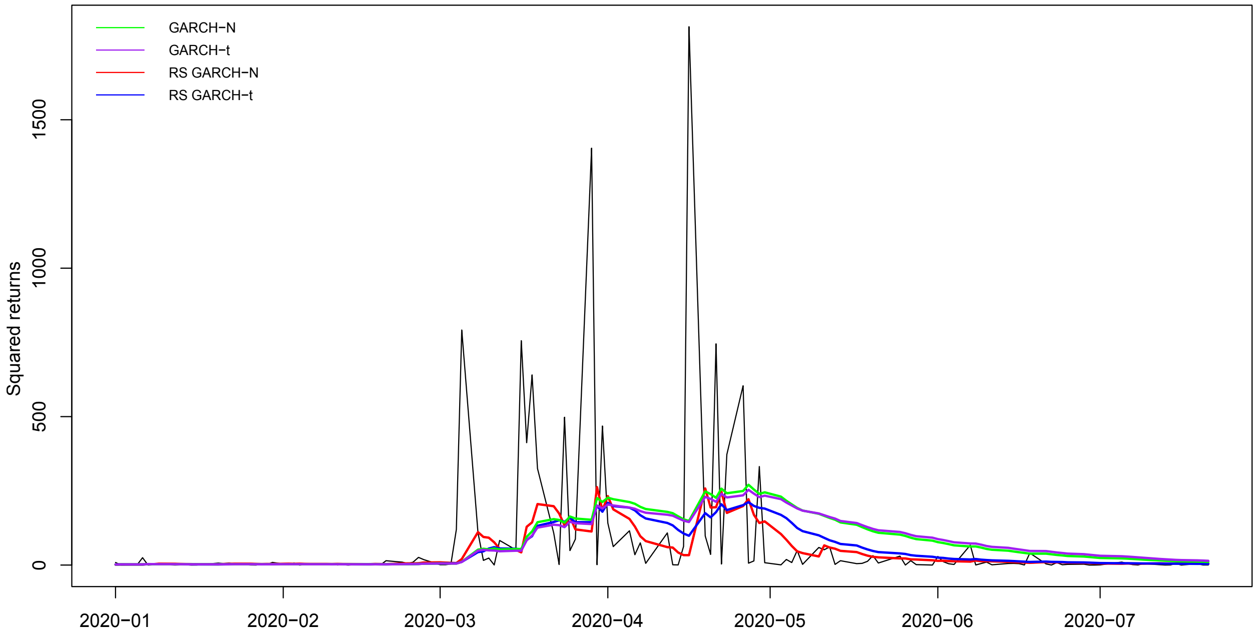
Notes: This Figure shows the one-step-ahead out-of-sample forecasts for GARCH-N, GARCH-t, RS GARCH-N and RS GARCH-t. The inner Figure is the zoomed version.

Figure B.6: Out-of-sample forecasting comparison between innovations.



Notes: This Figure shows the one-step-ahead out-of-sample comparison of GARCH-N against GARCH-t (left), and RS GARCH-N against RS GARCH-t (right) for the high volatility in 2020.

Figure B.7: Out-of-sample forecasting comparison of GARCH and RS GARCH models for 2020.



Notes: This Figure shows the one-step-ahead out-of-sample comparison of GARCH and RS GARCH model with Normal and Student-t distribution for 2020, the high volatile period.

Appendix C

R code

C.0.1 R code for data generating process

The below code is used for generating 1,000 Monte Carlo Simulations, used in Chapter 4, based on the true model RS GARCH MIDAS with Student-t innovations, described in (3.60).

```
ms.garch.midas.sim=function(months, days, r0, g0, K, w, omega0,
    omega1, theta, alpha, beta, p11, p22, df1){

    prob.matrix=matrix(c(p11, (1-p11), (1-p22), p22), nrow=2,
       ncol=2, byrow=TRUE)
    n=months*days
    e1=rt.scaled(n, df=df1, mean=0, sd=sqrt((df1-2)/df1))  #
       standardised student-t
    e=matrix(e1, nrow=days, ncol=months)
    st=matrix(0, nrow=days, ncol=months)
    x=numeric(n)
    phi=matrix(0, nrow=K, ncol=1)
```

```
tau=numeric(months)
g=matrix(0, nrow=days, ncol=months)
RV=matrix(0, nrow=1, ncol=months)
RV[1,1:8]=c(56.16643, 89.97468, 115.45625, 100.93382,
   68.64632, 49.45606, 72.00444, 53.52409) # values for
   absolute RV of crude oil 1-8 months

r=matrix(0, nrow=days, ncol=months)
parameter = NULL

for(o in 1:n){
  x[o]=runif(1, 0, 1)
}

#choosing states (0 or 1)
st[1]=1
for(d in 2:n){
  if(st[d-1]==0){
    if(x[d] < prob.matrix[1,1]){
      st[d]=0
    } else{
      st[d]=1
    }
  } else if(st[d-1]==1){
    if(x[d]<prob.matrix[2,1]){
      st[d]=0
    } else {
      st[d]=1
    }
  }
}
#the beta weighting scheme
```

```
b=1:K
phi[b]=((1-b/(K+1))^(w-1))/sum((1-b/(K+1))^(w-1))
rt=matrix(RV, nrow = months, ncol=1)

for (t in (K+1):months){

  #long-term component calculation
  tau[t]=exp(theta*(phi[1]*RV[t-1]+phi[2]*RV[t-2]+phi[3]*RV
     [t-3]+phi[4]*RV[t-4]+phi[5]*RV[t-5]+phi[6]*RV[t-6]+phi
     [7]*RV[t-7]+phi[8]*RV[t-8]))

  if(t==(K+1)){
    if(st[1,t]==0){
     #short-term component calculation
      g[1,t] = omega0 + (alpha * r0^2/tau[t]) + beta * g0
    }else if (st[1,t]==1){
      g[1,t] = omega1 + (alpha * r0^2/tau[t]) + beta * g0
    }
  }else{
    if(st[1,t]==0){
      g[1,t] = omega0 + (alpha * r[i,t-1]^2)/tau[t] + beta
         * g[i,t-1]
    }else if (st[1,t]==1){
      g[1,t] = omega1 + (alpha * r[i,t-1]^2)/tau[t] + beta
         * g[i,t-1]
    }
  }

  r[1,t]=sqrt(tau[t]*g[1,t])*e[1,t]

  for(i in 2:days){
```

```
        #short-term component calculation
        if(st[i,t]==0){
          g[i,t] = omega0 + (alpha * r[i-1,t]^2)/tau[t] + beta
             * g[i-1,t]
        }else if (st[i,t]==1){
          g[i,t] = omega1 + (alpha * r[i-1,t]^2)/tau[t] + beta
             * g[i-1,t]
        }
        r[i,t]=sqrt(tau[t]*g[i,t])*e[i,t]
      }
      RV[t]=sum(abs(r[,t]))
    }

    parameter$x=x
    parameter$state=st #simulated state
    parameter$return=r #simulated return
    parameter$g=g #short-term component
    parameter$tau=tau #long-term component
    parameter$RV=RV #realised volatility

    return(parameter)
  }

#run the simulation
true.sim=ms.garch.midas.sim(months = 340, days=22, r0=2.5266,
   g0=0.005, K=8, w=4.4037, df1=6.6087, omega0 = 0.1785, omega1
   = 0.4044, theta=0.0262, alpha=0.0944, beta=0.7241, p11
   =0.9961, p22=0.9986)

d.return=true.sim$return
```

C.0.2 R code for within-sample estimation

The below code is used to estimate parameters of RS GARCH MIDAS-t model, in Chapter 4, by obtaining Hamilton's filter probabilities and conditional density for QMLE. A detailed flow-chart is given in Figure 3.2.

```
ms.garch.midas.simulation.estimation=function(par, a){

  months=ncol(a)
  days=nrow(a)
  #parameters to estimate
  omega=par[1:2]
  theta=par[3]
  alpha=par[4]
  beta=par[5]
  p_ij=matrix(c(par[6], 1-par[6], 1-par[7], par[7]), nrow=2,
     ncol=2, byrow = TRUE)
  w1=par[8] #kappa weight
  df1=par[9] #degrees of freedom

  g_bar=array(rep(0, days), c(days, 2, months))
  g=array(rep(0, days), c(days, 2, months))
  tau=numeric(months)
  K=8
  phi=matrix(0, nrow=K, ncol=1)

  RV=matrix(0, nrow=1, ncol=months)
  RV[1,1:8]=c(56.16643, 89.97468, 115.45625, 100.93382,
     68.64632, 49.45606, 72.00444, 53.52409) # values for
     absolute RV of crude oil 1-8 months

  A=array(rep(0, days), c(2, 2, days, months))
```

```
B=array(rep(0, days), c(days, 2, months)) #first column state
   =1, second column state=2
C=array(rep(0, days), c(days, 2, months))
D=array(rep(0, days), c(days, 1, months))
E=array(rep(0, days), c(days, 2, months))
s.pi=matrix(c(1/3, 2/3), nrow=1, ncol=2) #stationary
   probability
parameter=NULL

p_hat_ij=array(rep(0, days), c(2, 2, days, months))
B.C=array(rep(0, days), c(days, 2, months)) #first column
   state=1, second column state=2

kl=1:K
phi[kl]=((1-kl/(K+1))^(w1-1))/sum((1-kl/(K+1))^(w1-1))

for(t in (K+1):(months)){
  RV[t]=sum(abs(a[,t]))
  tau[t]=exp(theta*(phi[1]*RV[t-1]+phi[2]*RV[t-2]+phi[3]*RV[t
     -3]+phi[4]*RV[t-4]+phi[5]*RV[t-5]+phi[6]*RV[t-6]+phi[7]*
     RV[t-7]+phi[8]*RV[t-8]))

  if(t==(K+1)){
    for(l in 1:2){
      for(j in 1:2){
        A[l,j,1,t]=p_ij[l,j]*s.pi[l] #pi = stationary
           probability
      }
    }
    for(j in 1:2){
      B[1,j,t]=sum(A[,j,1,t])
    }
```

```
    for(j in 1:2){
      B.C[1,j,t]=B[1,1,t]*p_ij[1,j]+(B[1,2,t]*p_ij[2,j]) #p(S
          _t+1=j|a_t-1)
    }
    for(i in 1:2){
      for(j in 1:2){
        p_hat_ij[j,i,1,t]= (p_ij[j,i]*B[1,j,t])/(B.C[1,i,t])
      }
    }
    g[1,,t]=omega+(alpha*2.5267^2)/tau[t]+beta*(2.5267)
    C[1,,t]=dt.scaled(x=a[1,t], df=df1, mean=0, sd=sqrt(tau[t
       ]*g[1,,t]*(df1-2)/df1))*B[1,,t]
    D[1,,t]=sum(C[1,,t])
    E[1,,t]=C[1,,t]/D[1,,t]

  }else{
    for(l in 1:2){
      for(j in 1:2){
        A[l,j,1,t]=p_ij[l,j]*E[i,l,t-1] #pi = stationary
            probability
      }
    }
    for(j in 1:2){
      B[1,j,t]=sum(A[,j,1,t])
    }
    for(j in 1:2){
      g_bar[1,j,t]=p_hat_ij[1,j,i,t-1]*g[i,1,t-1]+p_hat_ij[2,
          j,(i),t-1]*g[i,2,t-1]
    }
    for(j in 1:2){
      B.C[1,j,t]=B[1,1,t]*p_ij[1,j]+(B[1,2,t]*p_ij[2,j]) #p(S
          _t+1=j|a_t-1)
```

```
    }
    g[1,,t]=omega+(alpha*a[i,t-1]^2)/tau[t]+beta*g_bar[1,,t]

    for(z in 1:2){
      for(j in 1:2){
        p_hat_ij[j,z,1,t]= (p_ij[j,z]*B[1,j,t])/(B.C[1,z,t])
      }
    }
    C[1,,t]=dt.scaled(x=a[1,t], df=df1, mean=0, sd=sqrt(tau[t
       ]*g[1,,t]*(df1-2)/df1))*B[1,,t]
    D[1,,t]=sum(C[1,,t])
    E[1,,t]=C[1,,t]/D[1,,t]
  }

  for(i in 2:days){

    for(m in 1:2){
      for(j in 1:2){
        A[m,j,i,t]=p_ij[m,j]*E[(i-1),m,t]
      }
    }
    for(j in 1:2){
      B[i,j,t]=sum(A[,j,i,t])
    }
    for(j in 1:2){
      g_bar[i,j,t]=p_hat_ij[1,j,(i-1),t]*g[i-1,1,t]+p_hat_ij
          [2,j,(i-1),t]*g[i-1,2,t]
    }
    for(j in 1:2){
      B.C[i,j,t]=B[i,1,t]*p_ij[1,j]+(B[i,2,t]*p_ij[2,j]) #p(S
          _t+1=j|a_t-1)
    }
```

```
      g[i,,t]=omega+(alpha*a[i-1,t]^2)/tau[t]+beta*g_bar[i,,t]
      for(m in 1:2){
        for(j in 1:2){
          p_hat_ij[j,m,i,t]= (p_ij[j,m]*B[i,j,t])/(B.C[i,m,t])
        }
      }
      C[i,,t]=dt.scaled(x=a[i,t], df=df1, mean=0, sd=sqrt(tau[t
         ]*g[i,,t]*(df1-2)/df1))*B[i,,t]
      D[i,,t]=sum(C[i,,t])
      E[i,,t]=C[i,,t]/D[i,,t]
    }

  }

  log.likelihood=sum(log(D[,,(K+1):months]))
  return(log.likelihood) }

#initial parameters
ms.gm.par.t=c(0.1785, 0.4044, 0.0262, 0.0944, 0.7241, 0.9961,
   0.9986, 4.4037, 6.6087)
ms.midas.garch.result.t=optim(ms.gm.par.t, ms.garch.midas.
   simulation.estimation, a=d.return, method = "BFGS",control=
   list(fnscale=-1, trace=5, maxit=10000), hessian = F)
```

C.0.3 R code for out-of-sample forecasting

The below code is used in Chapter 5, for re-estimating parameters every 22 steps and obtaining one-step-ahead out-of-sample volatility forecasts for WTI crude oil returns.

```
times=seq(from = 7567, to = 8708, by = 22)
```

```
modified.ms.garch.midas.student=function(psi, lag.value, a){
  #parameters to be estimated
  months=length(RV.1)
  omega=psi[1:2]
  theta=psi[3]
  alpha=psi[4]
  beta=psi[5]
  p_ij=matrix(c(psi[6], 1-psi[6], 1-psi[7], psi[7]), nrow=2,
     ncol=2, byrow = T)
  w2=psi[8] #weight parameter
  df1=psi[9]

  g_bar=array(rep(0, length(a)), c(length(a), 2))
  g=array(rep(0, length(a)), c(length(a), 2))
  A=array(rep(0, length(a)), c(2, 2, length(a)))

  B=array(rep(0, length(a)), c(length(a), 2)) #first column
     state=1, second column state=2
  C=array(rep(0, length(a)), c(length(a), 2))
  D=array(rep(0, length(a)), c(length(a), 1))
  E=array(rep(0, length(a)), c(length(a), 2))
  s.pi=matrix(c(1/3, 2/3), nrow=1, ncol=2) #stationary
     probability
  parameter=NULL
  p_hat_ij=array(rep(0, length(a)), c(2, 2, length(a)))
  B.C=array(rep(0, length(a)), c(length(a), 2))  #first column
     state=1, second column state=2

  tau=numeric(months)
  K=lag.value
  phi=matrix(0, nrow=K, ncol=1)
  kl=1:K
```

```
phi[kl]=((1-kl/(K+1))^(w2-1))/sum((1-kl/(K+1))^(w2-1)) #Beta
   weighing scheme
rt=matrix(RV.1, nrow = months, ncol=1)

for(t in (K+1):months){
  tau[t]=exp(theta*(phi[1]*as.numeric(RV.1[t-1])+phi[2]*as.
     numeric(RV.1[t-2])+phi[3]*as.numeric(RV.1[t-3])+phi[4]*as
     .numeric(RV.1[t-4])+phi[5]*as.numeric(RV.1[t-5])+phi[6]*
     as.numeric(RV.1[t-6])+phi[7]*as.numeric(RV.1[t-7])+phi[8]
     *as.numeric(RV.1[t-8])))
}
srt = a[head(endpoints(a, "months") + 1, -1)]
tau.data=as.xts(tau,order.by=as.Date(index(srt)))
daily.tau=na.locf(merge(tau.data, foo=zoo(NA, order.by=index(
   a)))[, 1])
d=min(which(daily.tau!= 0))
start=d

#first day
for(i in 1:2){
  for(j in 1:2){
    A[i,j,start]=p_ij[i,j]*s.pi[i]
  }
}
for(j in 1:2){
  B[start,j]=sum(A[,j,start])
}
for(j in 1:2){
  B.C[start,j]=B[start,1]*p_ij[1,j]+(B[start,2]*p_ij[2,j])
     #p(S_t+1=j|a_t-1)
}
for(i in 1:2){
```

```
  for(j in 1:2){
    p_hat_ij[j,i,start]= (p_ij[j,i]*B[start,j])/(B.C[start,i
       ])
  }
}
g[start,]=omega+alpha*as.numeric(a[start-1])^2/as.numeric(
   daily.tau[start])+beta*(omega/(1-alpha-beta))
C[start,]=dt.scaled(x=as.numeric(a[start]), mean=0, df=df1,
   sd=sqrt(as.numeric(daily.tau[start])*g[start,]*(df1-2)/df1)
   )*B[start,]
D[start]=sum(C[start,])
E[start,]=C[start,]/D[start,]

#2nd step
for(k in (start+1):length(a)){

  for(i in 1:2){
    for(j in 1:2){
      A[i,j,k]=p_ij[i,j]*E[(k-1),i]
    }
  }
  for(j in 1:2){
    B[k,j]=sum(A[,j,k])
  }
  for(v in 1:2){
    g_bar[k,v]=p_hat_ij[1,v,(k-1)]*g[k-1,1]+p_hat_ij[2,v,(k
       -1)]*g[k-1,2]
  }
  for(j in 1:2){
    B.C[k,j]=B[k,1]*p_ij[1,j]+(B[k,2]*p_ij[2,j])    #p(S_t+1=j
       |a_t-1)
  }
```

```
    g[k,]=omega+alpha*as.numeric(a[k-1])^2/as.numeric(daily.tau
       [k])+beta*g_bar[k,]
    for(i in 1:2){
      for(j in 1:2){
        p_hat_ij[j,i,k]= (p_ij[j,i]*B[k,j])/(B.C[k,i])
      }
    }
    C[k,]=dt.scaled(x=as.numeric(a[k]), mean=0, df=df1, sd=sqrt
       (as.numeric(daily.tau[k])*g[k,]*(df1-2)/df1))*B[k,]
    D[k]=sum(C[k,])
    E[k,]=C[k,]/D[k]
}
  log.likelihood=sum(log(D[start:length(a)]))
  return(log.likelihood)
}


#Code for parameter re-estimation every 21 days.
for(i in seq(from = 7567, to = 8708, by = 22)){
  #initial parameters
  ms.gm.parameters.T=c(0.1785, 0.4044, 0.0262, 0.0944,  0.7241,
      0.9961, 0.9986, 4.4037, 6.6087)
  pos=which(times==i)
  RV.1<- apply.monthly(data2[(i-7566):(i)],abs.rv)
  ms.gm.result.student=optim(ms.gm.parameters.T, modified.ms.
     garch.midas.student, lag.value=8, a=data2[(i-7566):i],
     method = "Nelder-Mead", control=list( fnscale=-1,trace=5,
     maxit=5000), hessian=F)#, lower = c(0.000001, 0.000001,
     0.000001, 0.000001, 2.01, 0.000001, 0.000001), upper = c(
     Inf,Inf,0.9999,0.9999,Inf, 0.99999,0.99999))
  ms.gm.par.est.student[pos,]=ms.gm.result.student$par
  print(ms.gm.par.est.student[1:pos,])
}
```

```
#Code for one-step-ahead out of sample forecast
for(i in seq(from = 7567, to = 8708, by = 22)){
  pos=which(times==i)
  for(j in times[pos]:(times[pos+1]-1)){
    RV.1<- apply.monthly(data2[(j-7566):(j)],abs.rv)
    filter.ms.garch.midas=function(par, a){
      months=length(RV.1)
      #parameters to be estimated
      omega=par[1:2]
      theta=par[3]
      alpha=par[4]
      beta=par[5]
      p_ij=matrix(c(par[6], 1-par[6], 1-par[7], par[7]), nrow
         =2, ncol=2, byrow = TRUE)
      w1=par[8]
      df1=par[9]

      g_bar=array(rep(0, length(a)), c(length(a), 2))
      g=array(rep(0, length(a)), c(length(a), 2))

      #A=array(rep(0, length(a)), c(2, 2, days, months))
      A=array(rep(0, length(a)), c(2, 2, length(a)))

      B=array(rep(0, length(a)), c(length(a), 2)) #first column
          state=1, second column state=2
      C=array(rep(0, length(a)), c(length(a), 2))
      D=array(rep(0, length(a)), c(length(a), 1))
      E=array(rep(0, length(a)), c(length(a), 2))
      s.pi=matrix(c(1/3, 2/3), nrow=1, ncol=2) #stationary
         probability
      parameter=NULL
```

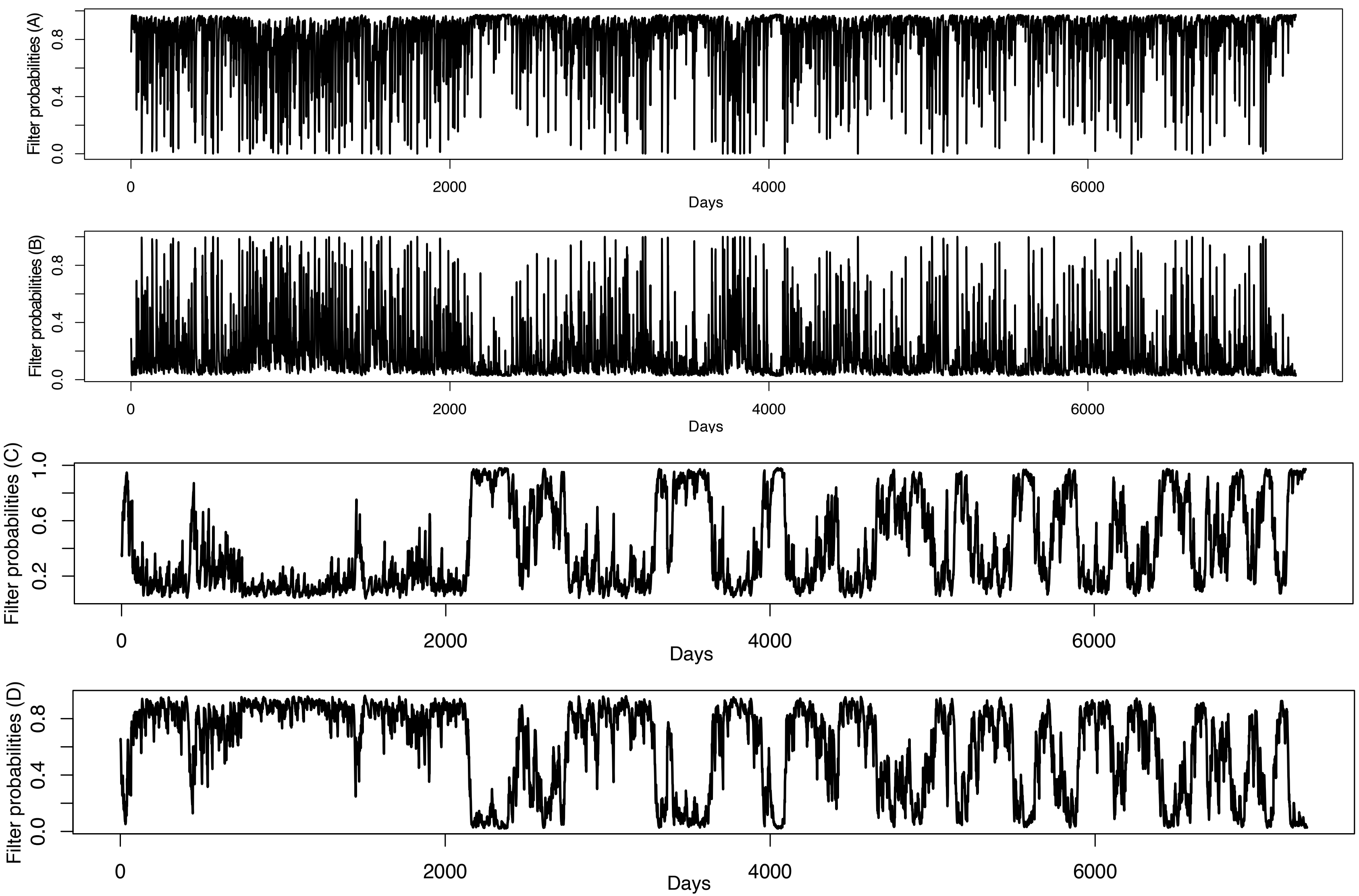
```
p_hat_ij=array(rep(0, length(a)), c(2, 2, length(a)))
B.C=array(rep(0, length(a)), c(length(a), 2))  #first
   column state=1, second column state=2
tau=numeric(months)
K=8
phi=matrix(0, nrow=K, ncol=1)

kl=1:K
phi[kl]=((1-kl/(K+1))^(w1-1))/sum((1-kl/(K+1))^(w1-1))
for(t in (K+1):months){
  tau[t]=exp(theta*(phi[1]*as.numeric(RV.1[t-1])+phi[2]*
     as.numeric(RV.1[t-2])+phi[3]*as.numeric(RV.1[t-3])+
     phi[4]*as.numeric(RV.1[t-4])+phi[5]*as.numeric(RV.1[t
     -5])+phi[6]*as.numeric(RV.1[t-6])+phi[7]*as.numeric(
     RV.1[t-7])+phi[8]*as.numeric(RV.1[t-8])))
}

#converting monthly tau to daily
srt = a[head(endpoints(a, "months") + 1, -1)]
tau.data=as.xts(tau,order.by=as.Date(index(srt)))
daily.tau=na.locf(merge(tau.data, foo=zoo(NA, order.by=
   index(a)))[, 1])
d=min(which(daily.tau!= 0))
start=d
for(l in 1:2){
  for(j in 1:2){
    A[l,j,start]=p_ij[l,j]*s.pi[l] #pi = stationary
       probability
  }
}
for(j in 1:2){
  B[start,j]=sum(A[,j,start])
```

```
}
for(j in 1:2){
  B.C[start,j]=B[start,1]*p_ij[1,j]+(B[start,2]*p_ij[2,j
     ])    #p(S_t+1=j|a_t-1)
}
for(i in 1:2){
  for(j in 1:2){
    p_hat_ij[j,i,start]= (p_ij[j,i]*B[start,j])/(B.C[
       start,i])
  }
}
g[start,]=omega+alpha*as.numeric(a[start-1])^2/as.numeric
   (daily.tau[start])+beta*as.numeric(a[start]^2)
C[start,]=dt.scaled(x=as.numeric(a[start]), mean=0, df=
   df1, sd=sqrt(as.numeric(daily.tau[start])*g[start,]*(
   df1-2)/df1))*B[start,]
D[start]=sum(C[start,])
E[start,]=C[start,]/D[start,]
for(i in (start+1):length(a)){
  for(m in 1:2){
    for(j in 1:2){
      A[m,j,i]=p_ij[m,j]*E[(i-1),m]
    }
  }
  for(j in 1:2){
    B[i,j]=sum(A[,j,i])
  }
  for(j in 1:2){
    g_bar[i,j]=p_hat_ij[1,j,(i-1)]*g[(i-1),1]+p_hat_ij[2,
       j,(i-1)]*g[(i-1),2]
  }
  for(j in 1:2){
```

```
      B.C[i,j]=B[i,1]*p_ij[1,j]+(B[i,2]*p_ij[2,j])    #p(S_t
         +1=j|a_t-1)
    }
    g[i,]=omega+(alpha*as.numeric(a[i-1]^2))/as.numeric(
       daily.tau[i])+beta*g_bar[i,]
    for(m in 1:2){
      for(j in 1:2){
        p_hat_ij[j,m,i]= (p_ij[j,m]*B[i,j])/(B.C[i,m])
      }
    }
    C[i,]=dt.scaled(x=as.numeric(a[i]), mean=0, df=df1, sd=
       sqrt(as.numeric(daily.tau[i])*g[i,]*(df1-2)/df1))*B[i
       ,]
    D[i]=sum(C[i,])
    E[i,]=C[i,]/D[i,]

  }

  log.likelihood=sum(log(D[start:length(a)]))
  parameter$filter.probabilities=E
  parameter$log.likelihood.value=log.likelihood
  parameter$g=g
  parameter$tau=tau
  parameter$daily.tau=daily.tau
  return(parameter)
}
f.p2=filter.ms.garch.midas(par=ms.gm.reest.par[pos,], a=
   data2[(j-7566):(j)])

prob1=(f.p2$filter.probabilities[(7567),1]*ms.gm.par.est.
   student[pos,6])+(f.p2$filter.probabilities[(7567),2]*(1-
   ms.gm.par.est.student[pos,7]))
```

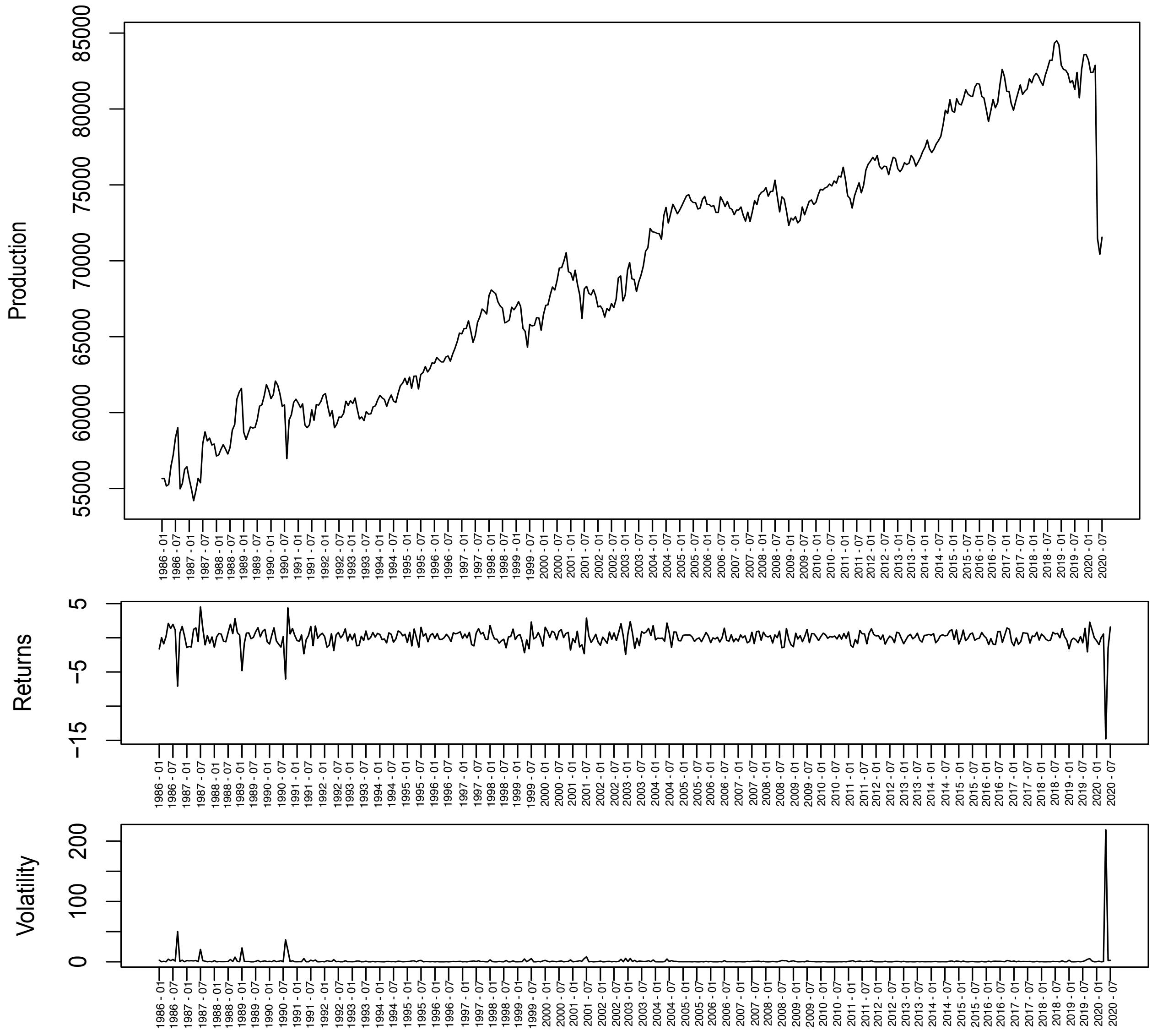
```
    prob2=(f.p2$filter.probabilities[(7567),1]*(1-ms.gm.par.est
       .student[pos,6]))+(f.p2$filter.probabilities[(7567),2]*(
       ms.gm.par.est.student[pos,7]))

    h0=(ms.gm.par.est.student[pos,1]+(ms.gm.par.est.student[pos
       , 4]*as.numeric(data2[j]^2))/as.numeric(f.p2$daily.tau
       [7567])+ms.gm.par.est.student[pos,5]*f.p2$g[(7567),1])*
       prob1
    h1=(ms.gm.par.est.student[pos,2]+(ms.gm.par.est.student[pos
       , 4]*as.numeric(data2[j]^2))/as.numeric(f.p2$daily.tau
       [7567])+ms.gm.par.est.student[pos,5]*f.p2$g[(7567),2])*
       prob2

    out.ms.gm.reest[j+1]=as.numeric(f.p2$daily.tau[7567])*(h0+
       h1)

  }
}
```

Figure C.1: Filter probabilities for RS GARCH models.



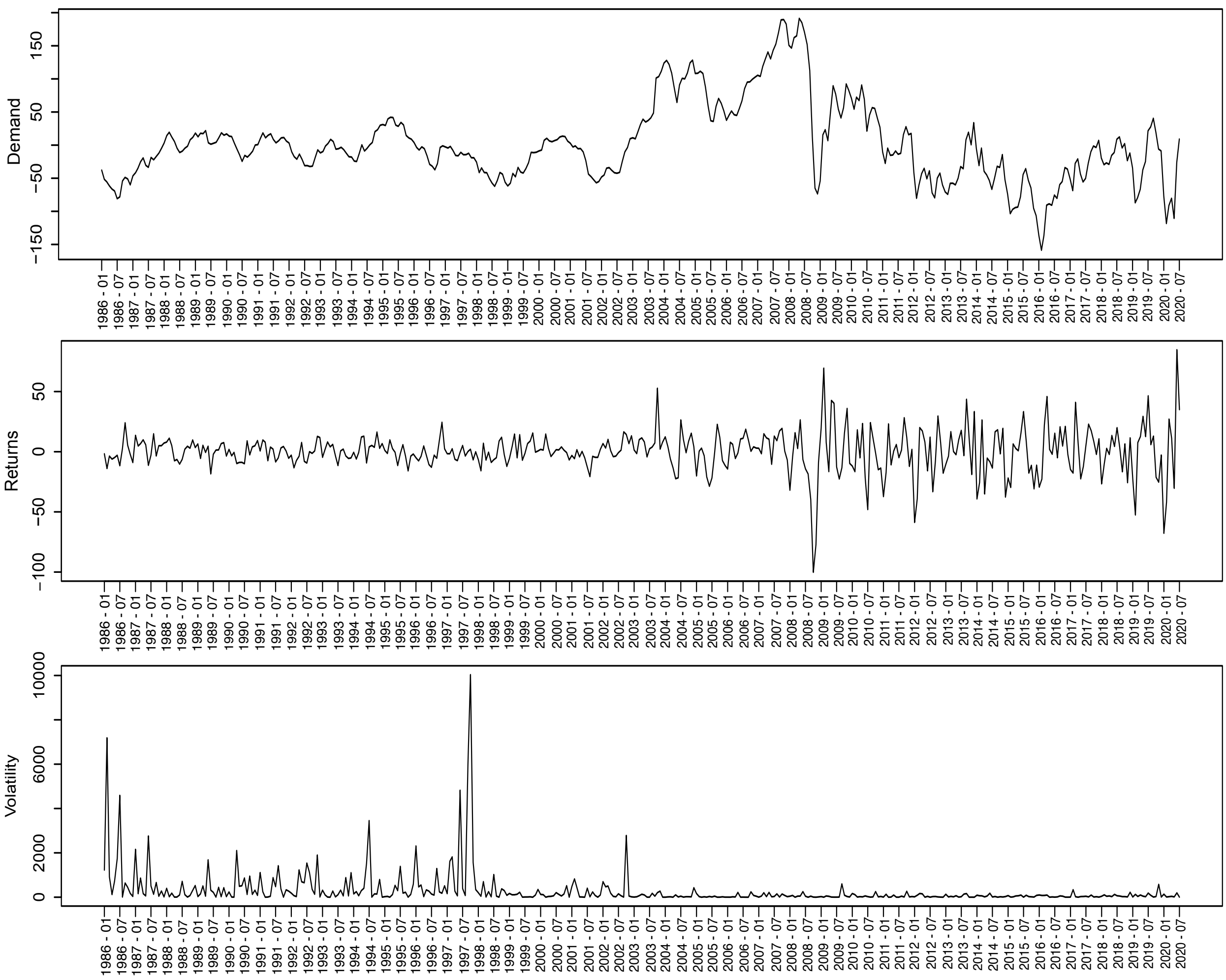
Notes: This Figure shows filter probabilities for RS GARCH model with Gaussian innovations. The (A) and (B) are the filter probabilities of being in low-volatility regime and high-volatility regime, respectively.

Figure C.2: WTI crude oil production levels.



Notes: This Figure shows global oil production obtained from EIA is used as a proxy of world oil supply. Plot (A) shows the production of crude oil of barrels per month, plot (B) shows the returns, and volatilities are displayed in plot (C) .

Figure C.3: WTI crude oil demand levels.



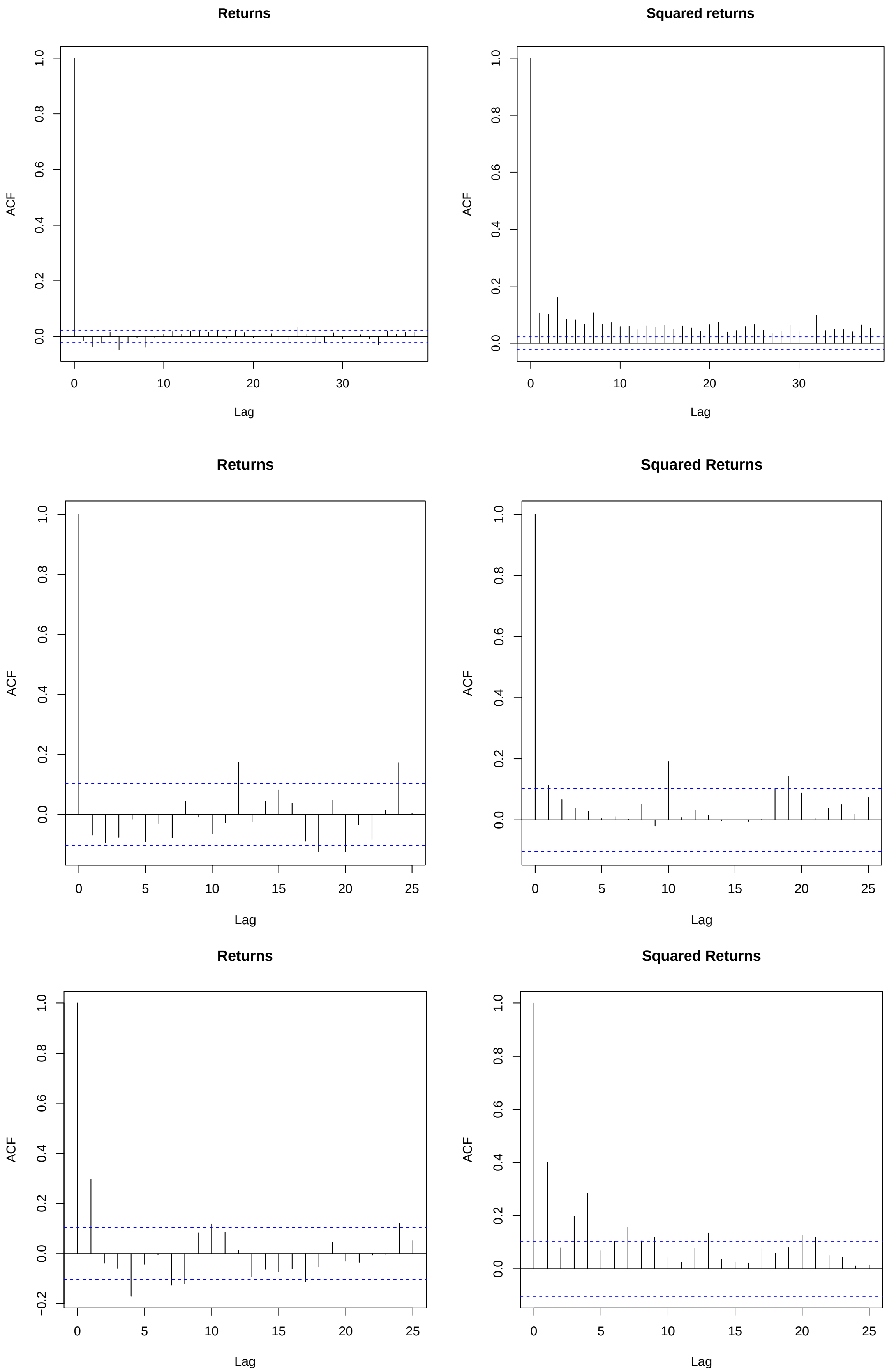
Notes: This Figure shows the crude oil demand level where index of Kilian (2009) is used as the signal for oil demand. Plot (A) shows the Kilian index per month, plot (B) shows the returns, and volatilities are displayed in plot (C).

Table C.1: Descriptive statistics of WTI crude oil for full sample.

	WTI	Production	Demand
Mean	0.0138	0.0566	0.1088
Var	7.4439	1.6269	298.9061
Min	-40.6396	-14.7902	-100.1826
Max	42.5832	4.5266	84.7775
Skewness	0.0151	-4.3704	-0.4667
Kurtosis	29.9217	49.7328	8.9368
JB ($\times 10^4$)	26.3138***	3.9484***	0.0634***
	(0.0000)	(0.0000)	(0.0000)
ADF(5)	-39.8320***	-8.8242***	-9.8954***
	(0.0100)	(0.0100)	(0.0100)
Q(5)	58.8780***	5.8412	85.3490***
	(0.0000)	(0.3220)	(0.0000)
$Q^2(5)$	1626.5000***	0.0817	120.9400***
	0.0000	(0.9999)	(0.0000)
ARCH(5)	1074.2000***	0.1960	70.7300***
	(0.0000)	(0.9992)	(0.0000)

Notes: This table presents the descriptive statistics for the full dataset from January 1986 to July 2020. Jarque-Bera (JB) Stat, augmented Dickey-Fuller (ADF) stat, Ljung-Box statistics, Q(5), and ARCH are the statistics testing for normal distribution, stationarity, serial correlation and heteroskedastic effects respectively. The corresponding p-values are given in brackets.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Figure C.4: ACF plots for log and squared returns of WTI crude oil, production and demand levels.



Notes: This Figure shows the ACF plots for WTI crude oil price (top), production (middle) and demand (bottom) returns and squared returns.

Table C.2: Diebold-Mariano test p-values comparison of innovations.

	GARCH-N	RS GARCH-N	GARCH MIDAS-N	RS GARCH MIDAS-N
GARCH-t	0.1731	0.8375	0.0214**	0.0481**
	0.2825	0.0230*	0.0031***	0.4493
RS GARCH-t	0.4496	0.5748	0.0120**	0.0369**
	0.0000**	0.4907	0.6593	0.1619
GARCH MIDAS-t	0.0002***	0.0006***	0.0253**	0.0767*
	0.0003**	0.1158	0.0111**	0.8325
RS GARCH MIDAS-t	0.0111*	0.0063***	0.0136*	0.1180
	0.0000**	0.6126	0.2612	0.0039***

Notes: This table presents the p-values from Diebold-Mariano test of equal predictive accuracy for one-day-ahead rolling window forecast over the period January 2016 to July 2020. The forecasts from the models with Gaussian innovation are compared with the forecasts from the models with Student-t innovations. The top value is p-value for MSE loss function while the bottom p-value is for QLIKE loss function. A p-value greater that 0.05 denotes the rejection of the null of equal predictive accuracy.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Table C.3: Diebold-Mariano test p-values for models with Gaussian innovations.

Gaussian	GARCH	RS GARCH	GM	RS GM	GM ($RV + Prod$)	RS GM ($RV + Prod$)	GM ($RV + Dem$)	RS GM ($RV + Dem$)
RS GARCH	0.9828	-	-	-	-	-	-	-
	0.0143*							
GM	0.0123*	0.0124*	-	-	-	-	-	-
	0.0005**	0.3428						
RS GM	0.0449*	0.0448*	0.0106**	-	-	-	-	-
	0.3643	0.0375*	0.1593					
GM	0.0140**	0.0142**	0.0122**	0.0117**	-	-	-	-
($RV + Prod$)	0.0381**	0.1014	0.0107	0.7089				
RS GM	0.0042***	0.0045***	0.0185**	0.0010***	0.0272**	-	-	-
($RV + Prod$)	0.6353	0.0616*	0.0909*	0.5286	0.3970			
GM	0.0066***	0.0068***	0.0417**	0.0048***	0.4903	0.0096***	-	-
($RV + Dem$)	0.0007***	0.3255	0.1637	0.1868	0.0173**	0.1078		
RS GM								
($RV + Dem$)								

Notes: This table presents the p-values from Diebold-Mariano test of equal predictive accuracy. GM stands for GARCH MIDAS. The top value is the p-value for MSE loss function while the bottom p-value is for QLIKE loss function.
* denote the rejection of null hypothesis at 10% significance level.
** denote the rejection of null hypothesis at 5% significance level.
*** denote the rejection of null hypothesis at 1% significance level.

Table C.4: Diebold-Mariano test p-values for models with Student-t innovations.

Student-t	GARCH	RS GARCH	GM	GM $(RV + Prod)$	RS GM $(RV + Prod)$	GM $(RV + Dem)$	RS GM $(RV + Dem)$
RS GARCH	0.2075	-	-	-	-	-	-
	0.0000***						
GM	0.0214**	0.0003***	-	-	-	-	-
	0.0031***	0.0069***					
RS GM	0.0132**	0.0013***	0.0006***	-	-	-	-
	0.0000***	0.5434	0.0047***				
GM	0.0003***	0.0003***	0.0163**	-	-	-	-
$(RV + Prod)$	0.1105	0.0000***	0.1324				
RS GM	0.0073***	0.0049***	0.0057***	0.0319**	-	-	-
$(RV + Prod)$	0.0000***	0.5414	0.0010***	0.0000***			
GM	0.0004***	0.0004***	0.0249**	0.7333	0.0066***	-	-
$(RV + Dem)$	0.0010***	0.0119**	0.0048***	0.0628*	0.0018***		
RS GM	0.0107**	0.0056***	0.0006***	0.0019***	0.0062***	0.0004***	-
$(RV + Prod)$	0.0000***	0.3442	0.0000***	0.0000***	0.6668	0.0000***	

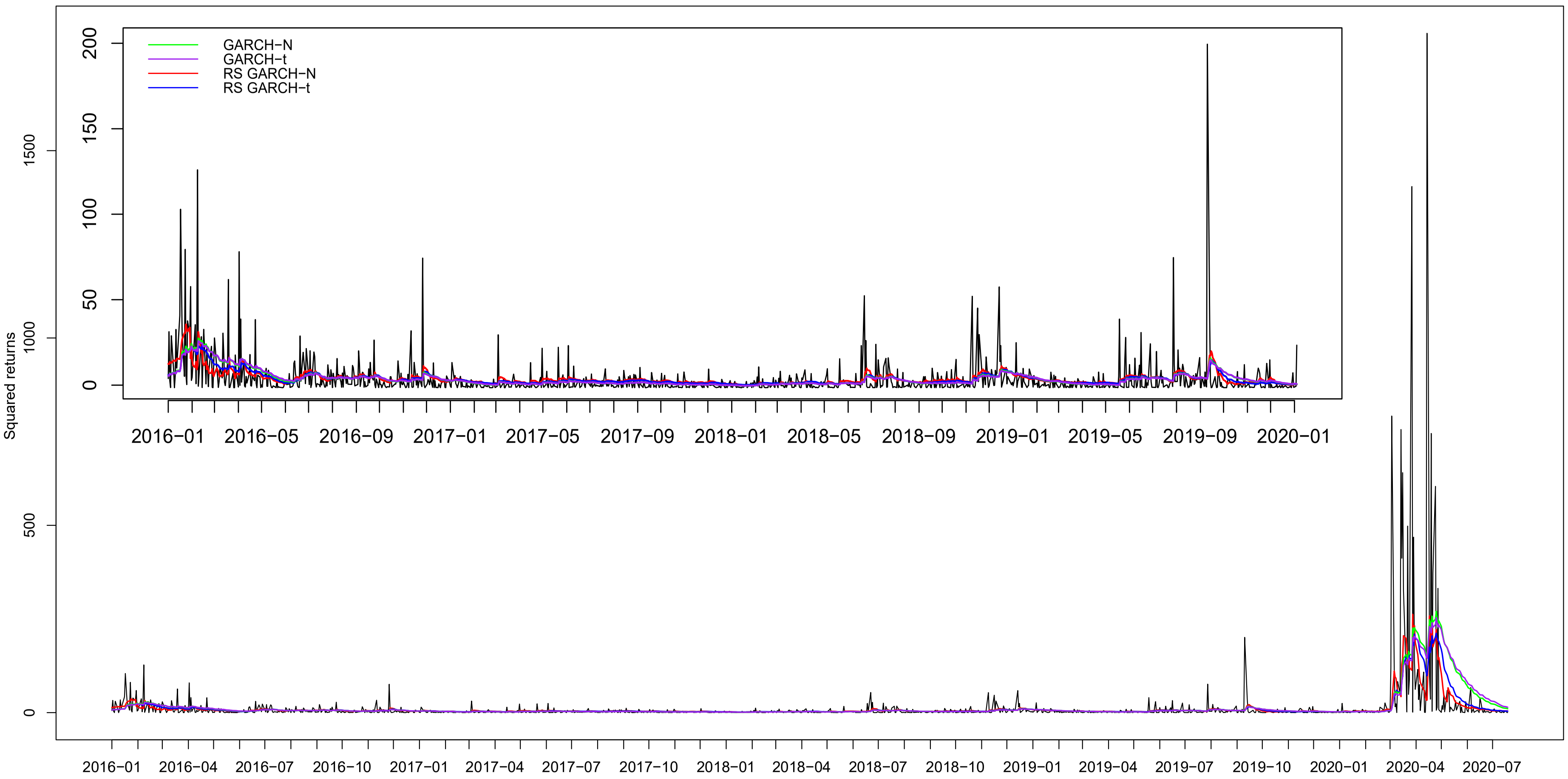
Notes: This table presents the p-values from Diebold-Mariano test of equal predictive accuracy. GM denotes GARCH MIDAS. The top value is p-value for MSE loss function while the bottom p-value is for QLIKE loss function. A p-value greater that 0.05 denotes the rejection of the null of equal predictive accuracy.

* denote the rejection of null hypothesis at 10% significance level.

** denote the rejection of null hypothesis at 5% significance level.

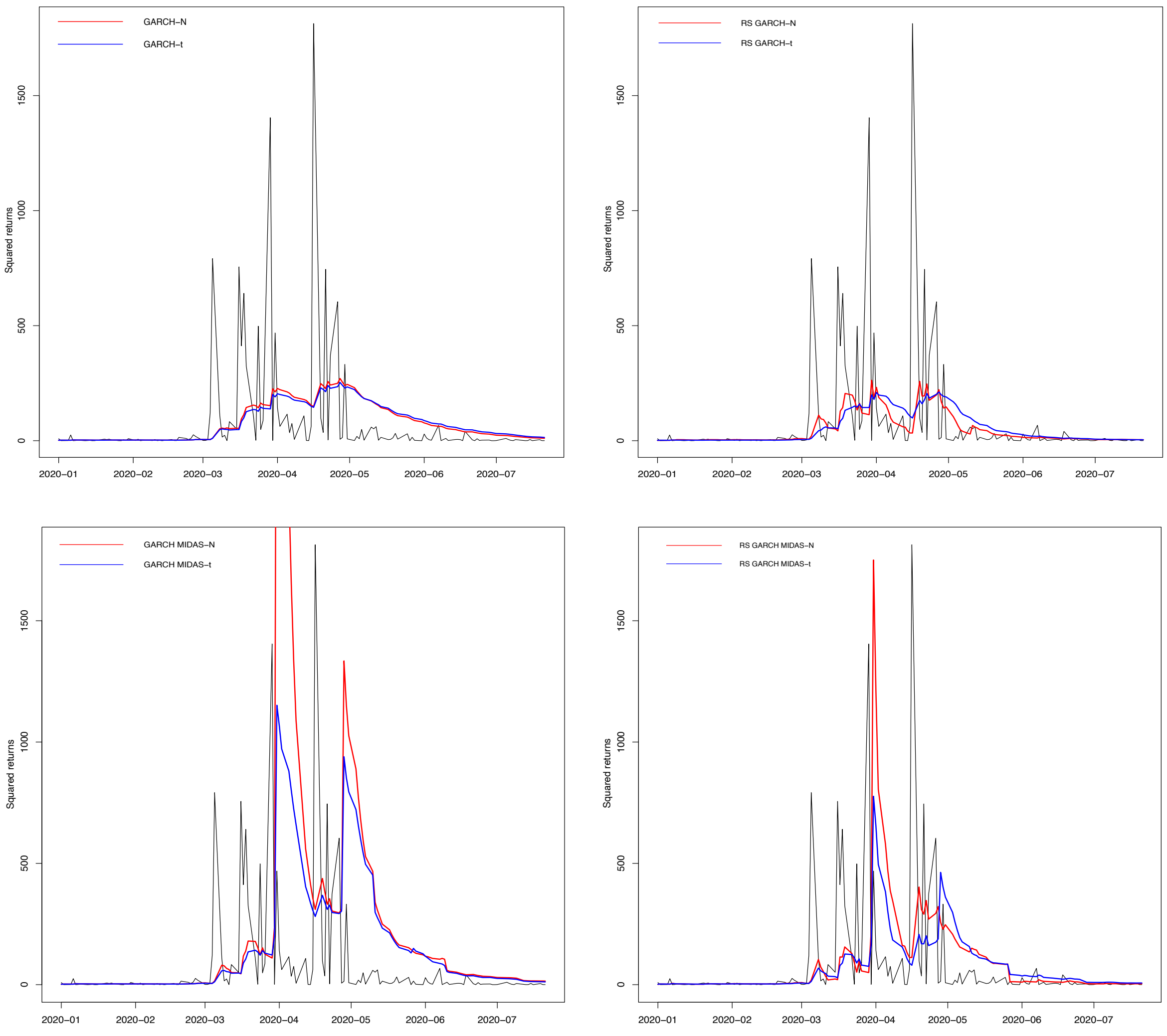
*** denote the rejection of null hypothesis at 1% significance level.

Figure C.5: Out-of-sample forecasting comparison of GARCH and RS GARCH models.



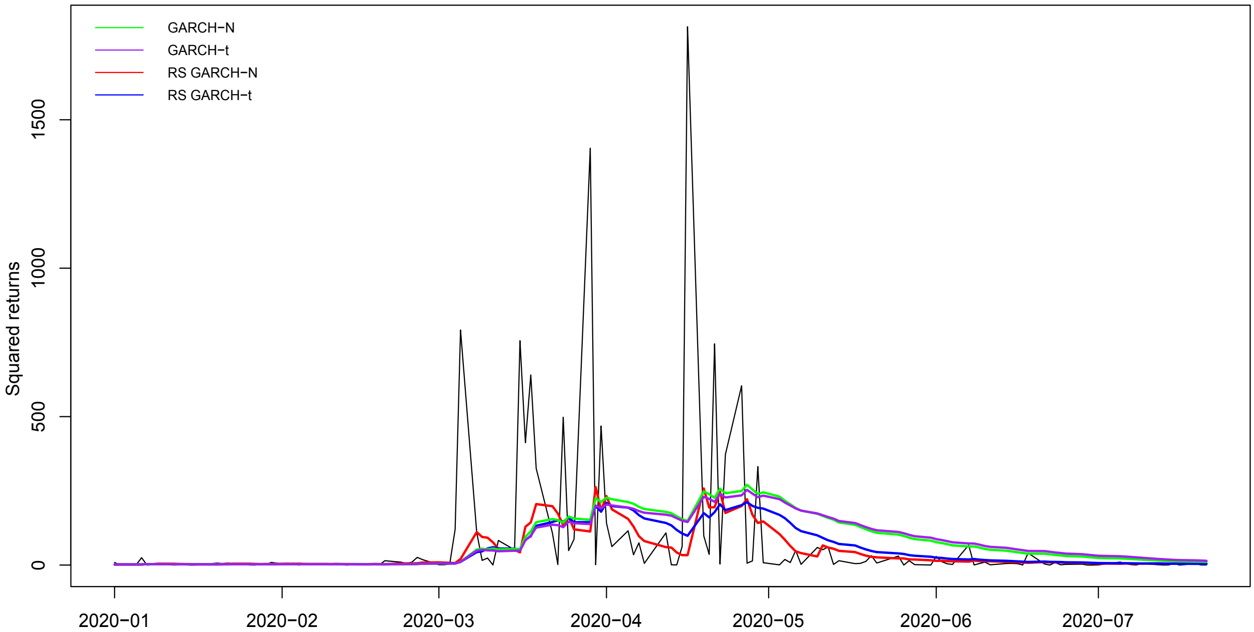
Notes: This Figure shows the one-step-ahead out-of-sample forecasts for GARCH-N, GARCH-t, RS GARCH-N and RS GARCH-t. The inner Figure is the zoomed version.

Figure C.6: Out-of-sample forecasting comparison between innovations.



Notes: This Figure shows the one-step-ahead out-of-sample comparison of GARCH-N against GARCH-t (left), and RS GARCH-N against RS GARCH-t (right) for the high volatility in 2020.

Figure C.7: Out-of-sample forecasting comparison of GARCH and RS GARCH models for 2020.



Notes: This Figure shows the one-step-ahead out-of-sample comparison of GARCH and RS GARCH model with Normal and Student-t distribution for 2020, the high volatile period.