technische universität
dortmund

Master Thesis

Winter 2024-2025

A Comparative Analysis of Econometric and Mathematical Approaches to Gas Probabilistic Forecasting: EGARCH vs. Heston Models

Thanh Quyen Nguyen
June 22, 2025

Supervisor: Prof. Matei Demetrescu
Prof. Christoph Hanck

Contents

List of Figures

List of Tables

1 Introduction

Natural gas is essential to global energy systems, representing over 23% of worldwide energy use, as reported by the International Energy Agency in August 2024. Additionally, natural gas markets are noticeably liberalized, making natural gas an essential financial commodity (Hong et al., 2020). According to the U.S. Energy Information Administration report, natural gas produces 50% less CO_2 than coal and 30% less CO_2 than oil when burned for energy. Therefore, natural gas has become a reliable and relatively cleaner source of energy, given that it supports the climate-neutral transformation scheme. A robust gas price forecast enables decision-making units to make informed decisions regarding energy consumption, investment strategies, resource allocation and risk management. While point forecasts seem to be insufficient to achieve the above-mentioned goals, probabilistic prediction, providing a full predictive distribution, quantifying uncertainty, and offering probabilities for different outcomes, is a powerful tool that can serve multiple purposes, e.g., scenario analysis, complex derivative pricing, and portfolio optimization.

In their analysis of 940 time-series forecasting researches from 1982 to 2005, De Gooijer and Hyndman (2006) concluded that the utilization of prediction intervals and densities, or probabilistic forecasting, has increased significantly over time as practitioners have recognized the limitations of point forecasts. While probability forecasts for binary events (e.g., an 80% likelihood of rain today, a 10% likelihood of a financial crisis by year-end) have been popular for several decades (Gigerenzer et al., 2005), the focus has increasingly turned to probabilistic forecasts for a broader range of variables and events. Significant challenges in science and society have driven this research area, including the accessibility of renewable energy resources (Wan et al., 2013; Zhu & Genton, 2012) and economic and financial risk assessment (Timmermann, 2000).

To achieve a robust probabilistic price forecast, accurate volatility prediction is essential. Introduced by Engle (1982), Autoregressive Conditional Heteroskedasticity (ARCH) is a well-known time-varying volatility model that was further developed by Engle (1982) and Bollerslev (1986) by including the ARMA structure, known as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. ARCH and GARCH have been extensively employed in the econometric literature to model the conditional variance characteristics of time series data because they enable the representation of the volatility clustering feature of financial returns. Despite their popularity, the GARCH model faces

some limitations. First, asymmetric effects shocks on volatility are not covered in the GARCH model, while it is a pronounced characteristic in the energy market (Sadorsky, 1999); Reboredo, 2011). Second, as volatility is required to be positive, the estimated parameters of the GARCH model have to be non-negative, which makes the GARCH model difficult to estimate. The exponential GARCH (EGARCH) model, introduced by Nelson (1991), not only captures the asymmetric effect of shocks on volatility but also simplifies the estimation by eliminating the non-negativity condition for estimated parameters.

While the EGARCH model has been widely employed for discrete-time series in econometric analysis, the Heston model, which captures both stochastic volatility and the correlation between price shocks and volatility shocks, is a benchmark model from a mathematical perspective for modeling stochastic volatility. Heston model, proposed by Steven L. Heston in 1993, extends the Black and Scholes model (BSM) by considering stochastic volatility driven by a Cox–Ingersoll–Ross (CIR) process (Sircar, 1985). Moreover, this model can capture the heavy-tailed nature of the return distributions, leverage effect, and volatility clustering, which makes it one of the most famous stochastic volatility models for option pricing and risk management.

As natural gas has become a significant financial commodity (Hong et al., 2020), there has been substantial growth in the literature on natural gas price forecasting over the last decade, with a wide range of methodologies employed, including econometric models, statistical machine learning models, and fundamental models. Ferrari et al. (2021) proposed dynamic factor models to forecast the prices of commodities, including oil, natural gas, and coal, using a large dataset of macroeconomic variables. The penalized maximum likelihood approach enables the shrinkage of parameters to zero, which allows the model to eliminate the dimensionality of the curve. They found that their model outperformed machine learning techniques, such as LASSO and elastic net. Similarly, macroeconomics and technical trading rules are employed to construct commodity price density forecasts (Wang et al., 2020). The results showed that technical indicators outperform economic variables in predicting the density of commodity prices.

Jonathan Berrisch and Florian Ziel (2022) proposed models to forecast day-ahead and one-month-ahead natural gas prices. The time series models covered skewed and heavy-tailed distributions, which enabled them to capture the stylized facts of financial data. Additionally, autocorrelation, seasonality, risk premia, temperature, storage levels,

the price of a CO2 certificate, and other energy prices were included in the model. Jinpei Liu et al. (2025) introduced probabilistic connectivity network (PCnet) link models for predicting natural gas prices. The empirical analysis demonstrated the outstanding performance of the price prediction approach; however, the method consisted of four data-intensive steps, which maked it challenging to verify the generalized representation of the model. Moreover, Ambya A. et al. (2020) employed econometric models, including ARIMA and GARCH, to forecast future natural gas prices. According to the empirical result, AR(1)-GARCH(1,1) performs the best among models from the same class.

The Heston model, which enables the capture of stochastic volatility, has been primarily used to forecast and analyze the prices of stock-exchange markets (Bianca Reichert and Adriano Mendonça Souza, 2022). In 2011, Beth introduced a non-Gaussian stochastic volatility model by Barndorff-Nielsen and Shephard and compared the natural gas price forecast performance of this model with that of the Heston model. Afterwards, Heston becomes more well-known for its application in the commodity market (Dong et al., 2020; Hsu et al., 2017). Recently, Oyuna, D., & Yaobin, L. (2021) applied the Heston model to forecast crude oil prices, and their finding showed that the stochastic volatility model performs better than the traditional GARCH model.

This thesis contributes to the literature by undertaking a detailed comparative analysis of EGARCH assuming skewed-t-distribution and Heston model, evaluating their performance in modeling and forecasting natural gas probabilistic price. For performance evaluation, I compare not only Continuous Ranked Probability Score (CRPS) but also investigate the investment strategy built on probabilistic forecast. This approach provides insightful comparison for both academic and practitioner as it brings both statistical and commercial perspectives while analyzing the accuracy of price prediction. I found that there is no significant difference between the performance of EGARCH and Heston model based on the result of Wilcoxon test for probabilistic forecast errors and delta hedges errors. However, it can be shown graphically that both forecast errors and delta hedges errors of Heston model is noticeably smaller than errors of EGARCH model, implying the advantage of Heston usage during extreme events that can cause demand-supply shocks.

The remainder of the paper is as follows: Section 2 explains the data used, giving overview of the US gas price and many stylized facts. The proposed models, estimation method and evaluation criteria are presented and discussed in Section 3. Section 4 presents

the comparative analysis of EGARCH and Heston models in terms of their predictive performance. The conclusion is discussed in Section 5.

2 Data

This study focuses on analyzing daily prices of NYMEX Henry Hub Natural Gas Futures in North America which is one of the most actively traded energy futures contracts globally. The underlying asset of this futures contract is delivered at the Henry Hub in Erath, Louisiana, USA. Monthly contracts that trades for all 12 months of the calendar year (from January to December) for the current year plus next 12 years are available for trading. The lack of liquidity in futures contracts that are distant from maturity results in an inefficient market, which can adversely affect data quality. Therefore, in this thesis, the data analysis, empirical modeling and price forecast will make used of the daily close price of futures contracts within 1 year before expiry even though the more data is available for each contract. In total, 108 futures contracts from Jan-2016 to Dec-2024 are considered, which provides comprehensive analysis for natural gas price dynamic in different regimes from before-after Covid 19 and before-after Russia-Ukraine conflict.

Figure 1 and Figure 2 show the overview of January contracts which represents for winter months. January prices are extremely high with many significant price movements in the first two years after Russia-Ukraine conflict, which lead to a heavy-tailed return distribution. Similarly, January price in 2019 also experienced a big price changes within 2 months before its maturity. prices before Covid 19 and in 2024 after the tense stage of the recent political conflict are likely to move smoothly and continuously. It is consistent with what can be seen from Figure 2 that time-varying volatility and volatility clustering is observed for only January 19 and January 22. Volatility clustering is clearly seen when January 19 and January 22 came close to expiry, that turbulent (high volatility) sub-periods are followed by quiet (low-volatility) sub-periods. This pattern does not have periodical feature. In general, model with homoscedasticity, constant variance, is not effectively capture the dynamic of volatility and prices.

Figure 3 depicts the histogram of January contracts' log returns. The heavy-tailed and skewed returns distribution is caused by some large spikes which can be seen in Figure 2. As a result, it justifies the application of EGARCH with skewed-t distribution

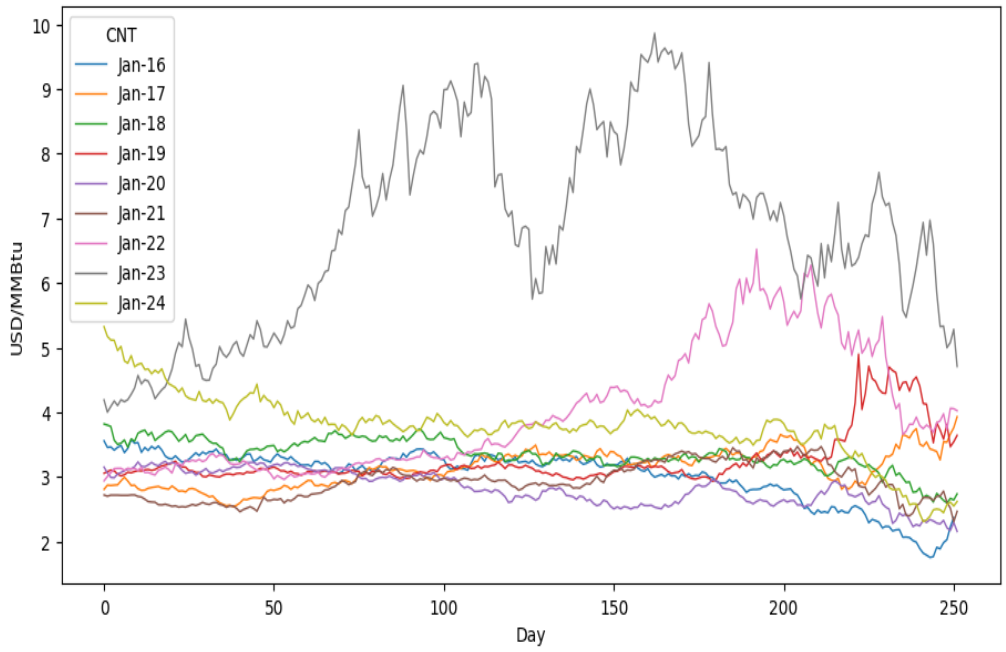

Figure 1: January Futures Contract daily prices within one year before expiry

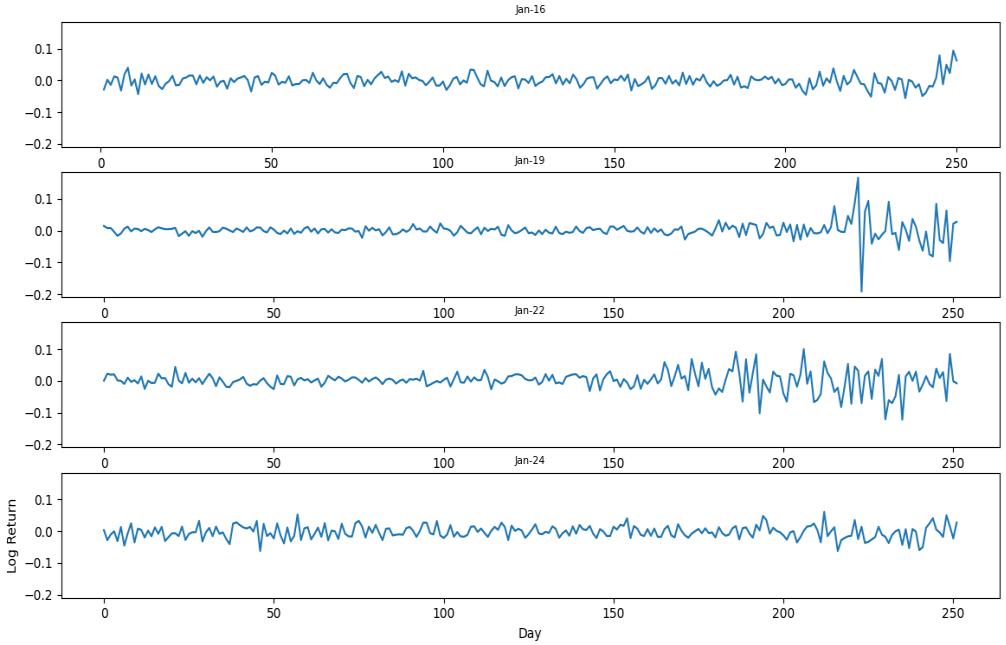

Figure 2: January Futures Contract daily log return within one year before expiry

and HESTON models in this study.

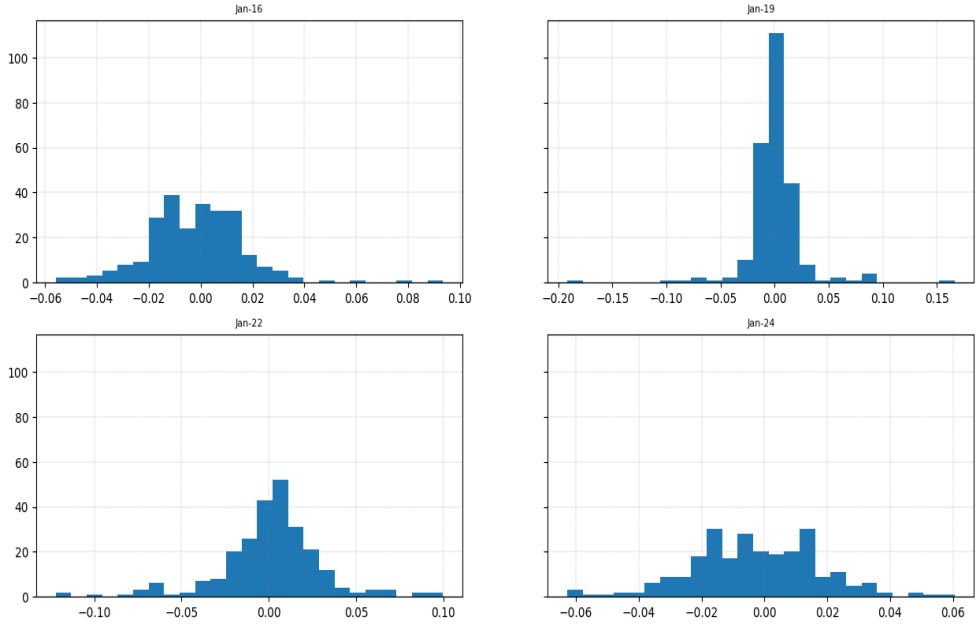

Figure 3: January Futures Contract daily log return histogram

Figure 4 and Figure 5 represent the price and log return time series of July contracts. It is noteworthy that Covid 19 seems to not have effect on prices of July 2019 while the political conflict had a significant influence in price of July 2022 and July 2023. On the other hand, volatility cluster is observed for all July 2016, July 2019, July 2022 and July 2024, which is the most important argument for using time-varying volatility models like EGARCH and HESTON.

Figure 6 shows the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of January 2016 logarithm return at first moment and second moment. The finding is in line with one of the literature of financial returns' s stylized facts that there is no clear presense of autocorralation of return while there are significant auto-correltion of squared returns. These ACF and PACF features quantitatively indicate the volatility clustering which justifies the used of EGARCH and HESTON models. More graphical illustration for ACF and PACF of other futures contracts can be found in Appendix.

In this study, I assume Efficient Market Hypothesis (EMH) developed by Eugene F. Fama in the 1960s holds true that prices fully reflect all available information at any given time. Based on EMH, futures prices are random and there is no need to use past

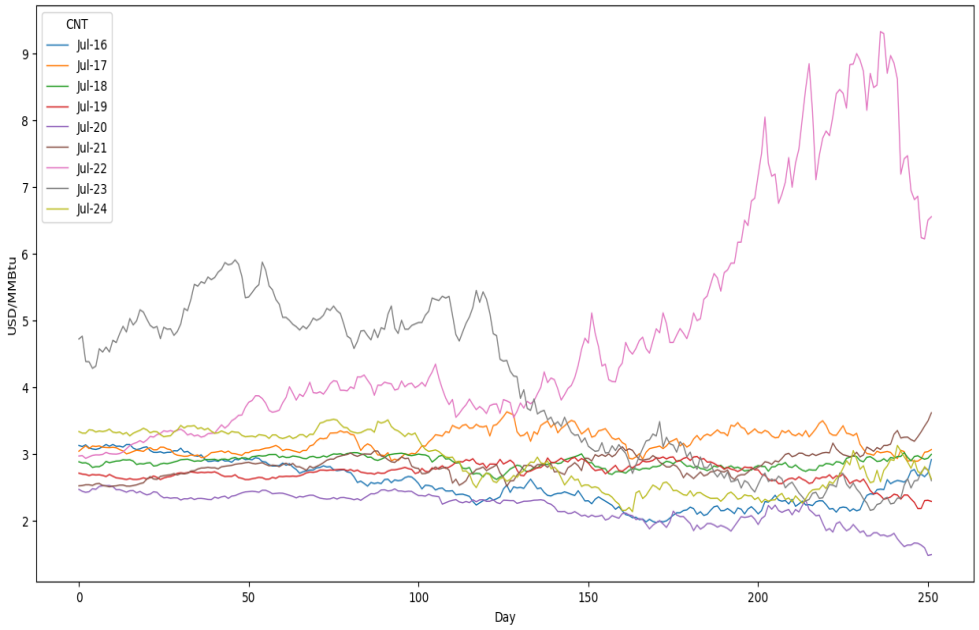

Figure 4: July Futures Contract daily prices within one year before expiry

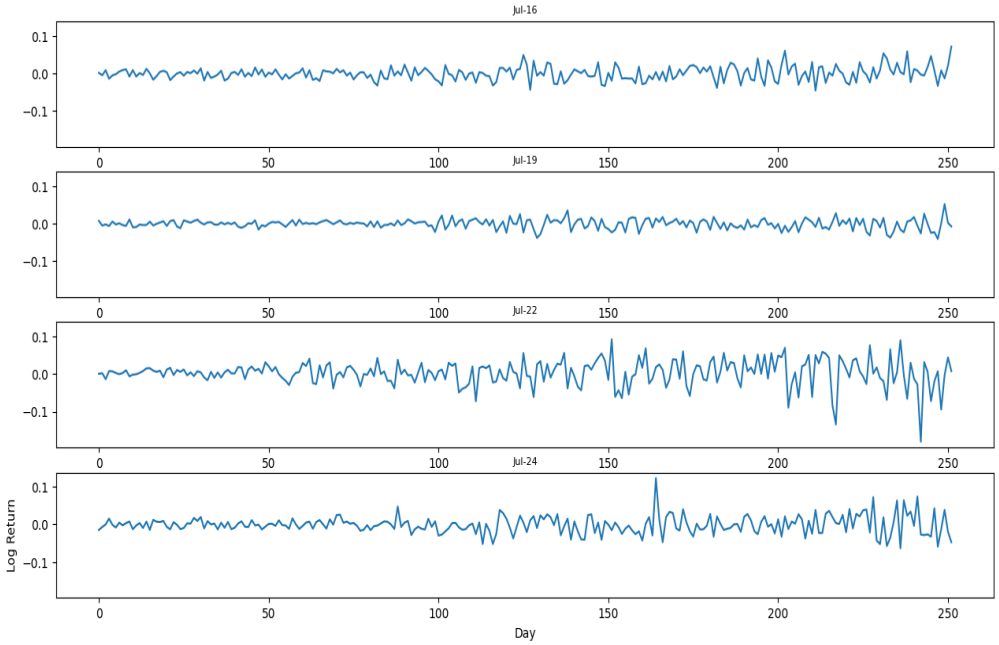

Figure 5: July Futures Contract daily log return within one year before expiry

information or exogenous variables to systematically predict prices. Therefore, this study utilizes only prices series, and focus on analyzing its dynamic employing EGARCH and

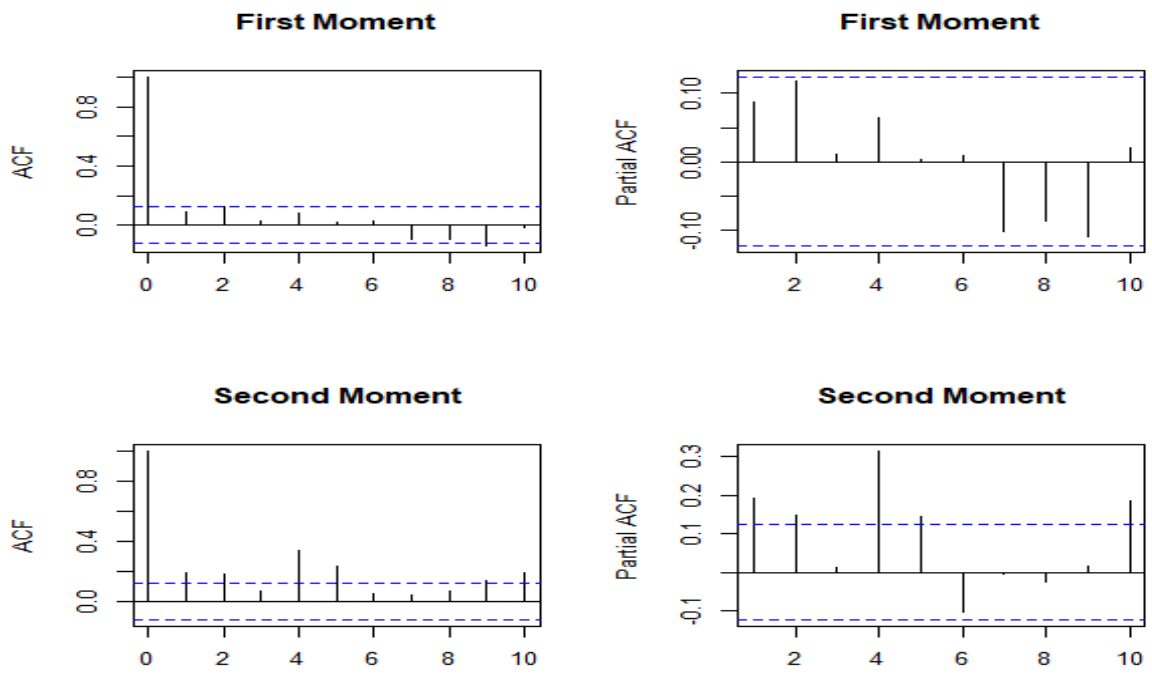

Figure 6: ACF and PACF of January 2016 log return and log return squared

HESTON model.

3 Methodology

3.1 EGARCH

3.1.1 Background and motivation

Traditional econometric models, which assume homoskedasticity, do not represent the well-documented stylized facts of financial time series, such as volatility clustering, leverage effects, and fat-tailed distributions. In 1982, autoregressive conditionally heteroscedastic (ARCH) models were first developed by Engle. The ARCH model is the first formal framework to capture volatility clustering. It describes volatility as a function of past squared residuals (shocks).

While the ARCH model is simple and successfully describes the important characteristic of financial returns, it usually requires a large number of lagged squared residuals to sufficiently represent the volatility process, resulting in estimation inefficiency. To address this issue, some alternative models were introduced. An extension of the ARCH model,

general autoregressive conditionally heteroscedastic (GARCH), proposed by Bollerslev (1986), is one of the most well-known approaches.

A key feature of (G)ARCH models is the conditional variance, which allows the variance at time t to depend on the variances and shocks at time u with $u < t$.

A strong GARCH(p,q) model is given by

$$\epsilon_t = \sigma_t \eta_t,$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where the α_i and β_j are non-negative constant and ω is positive constant, and $\eta_t \sim \mathbb{N}(0, 1)$.

If all $\beta_j = 0$, function of variance becomes

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2$$

and the process has a form of ARCH(p).

As the GARCH(p,q) model has ARCH(∞), it helps capture the variance process without adding too many lags of shocks and variances. However, there are some technical and practical issues that one must consider when employing GARCH(p,q) to model financial volatility clusters. First, α_i and β_j have to be positive to ensure that $\sigma^2 > 0$. The restriction makes it hard to have an efficient optimizer. Second, ARCH and GARCH both assume a symmetric effect of negative and positive shocks, although there is an asymmetric influence of price shocks in practice. The exponential general autoregressive conditionally heteroscedastic (EGARCH) model is one of the solutions to both issues while maintaining the capability to capture volatility clustering.

3.1.2 EGARCH (1,1) Model

EGARCH captures volatility clustering as the conditional variance (volatility) is represented by a function of past squared returns, absolute returns, and other variables, in a way that accounts for the persistence of volatility over time. The EGARCH(p,q) model, proposed by Nelson (1991), is defined as follows:

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) = \mu + \epsilon_t,$$

$$\epsilon_t = \sigma_t z_t,$$

$$\ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{p}\left\{\alpha_i\left(|z_{t-i}| - \mathbb{E}[|z_{t-i}|]\right) + \gamma_i z_{t-i}\right\} + \sum_{j=1}^{q}\beta_j \ln\left(\sigma_{t-j}^2\right)$$

where S_t is natural gas futures price at time t, r_t represents the logarithmic returns at time t, μ is the mean of return, $\sigma_t = Var(\epsilon_t|\epsilon_u, u < t)$ is the conditional variance, z_t is an i.i.d standardized residuals (innovations), ω is constant term, α_i measures the impact of magnitude of past shocks (symmetric effect), γ_i represent the sign effect of z_t (asymmetric effect), β_j is coefficient of past conditional variances, p and q are the orders of EGARCH model. The term $\sum_{j=1}^{q}\beta_j \ln\left(\sigma_{t-j}^2\right)$ captures the persistence of volatility. Specifically, high β implies that large volatility in previous period tends to result in large volatility in subsequent periods. Since $\ln(\sigma_t^2)$ can be negative, the positivity of the variance is automatically satisfied; therefore, there is no restrictions for parameters.

This thesis considers EGARCH(1,1) in order to have similar model setting with Heston model. The conditional variance of logarithmic returns under EGARCH(1,1) becomes:

$$\ln\left(\sigma_t^2\right) = \omega + \alpha\left(|z_{t-1}| - \mathbb{E}[|z_{t-1}|]\right) + \gamma z_{t-1} + \beta \ln\left(\sigma_{t-1}^2\right)$$

By construction, γz_t and $\alpha(|z_t| - \mathbb{E}[|z_t|])$ have mean zero. If the distribution of z_t is symmetric, the two elements are orthogonal. The conditional variance process responds asymmetrically to increase and decrease in price, which can be demonstrated as follows:

$$\begin{cases} \ln\left(\sigma_t^2\right) & = \omega + (\alpha+\gamma) z_{t-1} + \beta \ln\left(\sigma_{t-1}^2\right), 0 < z_t < \infty \\ \ln\left(\sigma_t^2\right) & = \omega + (-\alpha+\gamma) z_{t-1} + \beta \ln\left(\sigma_{t-1}^2\right), -\infty < z_t \leq 0 \end{cases} \tag{3.1}$$

Additionally, z_t is assumed to be skewed-student-t distributed (SST) with mean 0 and variance 1. As a result, $\epsilon_t \sim SST(0, \sigma_t)$. The Student-t distribution is useful in capturing the heavy tails and skewness feature of empirical log-return time series. Developed by Fernandez and Steel in 1998, the density of $SST(\sigma, \nu, \delta)$ is defined as follows:

$$f(z|\sigma,\nu,\delta) = \frac{2}{\delta + \frac{1}{\delta}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma} \left[1 + \frac{z^2}{\sigma^2\nu}\left(\frac{1}{\delta^2}\mathbb{1}_{[0,\infty)}(z) + \delta^2 \mathbb{1}_{(-\infty,0)}(z)\right)\right]^{-\frac{\nu+1}{2}} \tag{3.2}$$

with:

$$\nu > 2 \text{ is degree of freedom,} \tag{3.3}$$

$$\delta > 0 \text{ represents skewness parameter} \tag{3.4}$$

$$\delta^2 = \frac{P(z \geq 0|\sigma, \nu, \delta)}{P(z < 0|\sigma, \nu, \delta} \tag{3.5}$$

If $\delta = 1$, SST is reduced to Student-t distribution, which was developed in Bollerslev (1987). When $\delta \in (0, 1)$, SST is left-skewed, and when $\delta > 1$, SST is right-skewed.

3.1.3 Estimation

I consider EGARCH(1,1) model that cover all leverage effect, heavy tail and skew innovation. The proposed model is defined as follow:

$$r_t = \ln\left(\frac{S_t}{S_{t-1}}\right) = \mu + \epsilon_t,$$

$$\epsilon_t = \sigma_t z_t,$$

$$z_t \sim SST(0, 1, \nu, \delta),$$

$$\epsilon_t \sim SST(0, \sigma_t, \nu, \delta),$$

$$\ln\left(\sigma_t^2\right) = \omega + \alpha\left(|z_{t-1}| - \mathbb{E}[|z_{t-1}|]\right) + \gamma z_{t-1} + \beta \ln\left(\sigma_{t-1}^2\right)$$

The parameter set $\varphi = (\mu, \alpha, \beta, \gamma, \delta, \nu)$ is estimated using Quasi-Maximum Likelihood (QML) approach (Francq, C. and Zakoian, J. M., 2019). In order to adopt QML, besides the i.i.d skewed-Student-t distribution of innovation z_t, some initial values z_{t-1}, σ_{t-1} have to be chosen according to the estimated parameters as well as the observed r_t, which can be described as follows:

$$\sigma_0 = \exp\left(\frac{\tilde{\omega}}{1 - \tilde{\beta}}\right)^{0.5}, \tag{3.6}$$

$$z_0 = \frac{r_0 - \tilde{\mu}}{\sigma_0} \tag{3.7}$$

with $\tilde{\omega}$ and $\tilde{\beta}$ are candidate for estimated parameters. (3.6) is achieved by assuming $z_t = 0$ and $\sigma_t = \sigma_{t-1}$ in the long-run. Although it is not exact, it would be efficient to have a good initial values which is necessary for QML estimation.

The $\tilde{\epsilon}_t$, t = 1,..., n, and $\tilde{\sigma}_t^2$, t = 1,..., n are recursively determined by

$$\tilde{\epsilon}_t = r_t - \mu$$

$$\tilde{z}_t = \frac{\tilde{\epsilon}_t}{\tilde{\sigma}_t}$$

$$\ln\left(\tilde{\sigma}_t^2\right) = \tilde{\omega} + \tilde{\alpha}\left(|\tilde{z}_{t-1}| - \mathbb{E}[|\tilde{z}_{t-1}|]\right) + \tilde{\gamma}\tilde{z}_{t-1} + \tilde{\beta}\ln\left(\tilde{\sigma}_{t-1}^2\right)$$

The conditional skewed-Student-t quasi-likelihood is as follows

$$L_n(\varphi) = \prod_{t=1}^{n} f(\tilde{z}_t|1, \tilde{\nu}, \tilde{\delta})],$$

with

$$f(\tilde{z}_t|1, \tilde{\nu}, \tilde{\delta}) = \frac{2}{\tilde{\delta} + \frac{1}{\tilde{\delta}}} \frac{\Gamma\left(\frac{\tilde{\nu}+1}{2}\right)}{\Gamma\left(\frac{\tilde{\nu}}{2}\right)\sqrt{\pi\tilde{\nu}}} \left[1 + \frac{z^2}{\tilde{\nu}}\left(\frac{1}{\tilde{\delta}^2}\mathbb{1}_{[0,\infty)}(\tilde{z}) + \tilde{\delta}^2\mathbb{1}_{(-\infty,0)}(\tilde{z})\right)\right]^{-\frac{\tilde{\nu}+1}{2}}$$

The QML estimated parameters $\hat{\varphi}$ is achieved by maximize the likelihood $L_n(\varphi)$ or log-likelihood $\ln(L_n(\varphi))$. It is equivalent to minimize the loss function given below

$$\hat{\varphi} = \arg\min_{\varphi} \left[-\ln(L_n(\varphi))\right]$$

3.2 Heston model

3.2.1 Background and motivation

Black-Scholes model

Black-Scholes model, which assumes Geometric Brownian Motion (GBM) has been become the benchmark of models for financial derivatives pricing. The price of equity or commodity is given by GBM as follows

$$dS_t = \mu S_t dt + \sqrt{V} S_t dW_t^s$$

where

- S_t: asset price at time t

- μ: expected return
- $\sqrt{V}$: constant volatility
- W_t^s: Wiener process (Brownian motion)

Back-Scholes model is well-known thanks to its closed-form solution for European options pricing. However, it has several issues for practitioners when applying in real world. First, volatility $\sqrt{V}$ is assumed to be constant over time. It can be seen from empirical data that volatility is time-varying and has feature of clustering. Secondly, in real world, volatility tends to revert to long-term mean, but this feature is not covered by GBM either. To address these issues, stochastic volatility Heston models were developed in 1993 where underlying price follows GBM with stochastic volatility be represented by Ornstein-Uhlenbeck process.

Ornstein-Uhlenbeck

The mean reversion Ornstein-Uhlenbeck model is employed in Heston model for describing stochastic volatility process

$$dV_t = \kappa(\theta - V_t)dt + \sqrt{V_t}\sigma_V dW_t^V$$

where

- V_t: variance at time t
- κ: speed of mean reversion
- θ: long-term volatility
- σ_V: volatility of volatility
- W_t^v: Wiener process (Brownian motion)

Ornstein-Uhlenbeck process has property of mean reversion that volatility fluctuates but tends to revert to long-term level. Additionally, Brownian motion W_t^v make the process stochastic.

3.2.2 Model

The Heston model is an extension of the Black and Scholes model (BSM) that Heston model considers stochastic volatility (Heston 1993). Heston model is given by two stochastic differential equations:

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^S \tag{3.8}$$

$$dV_t = \kappa(\theta - V_t)dt + \sqrt{V_t}\sigma_V dW_t^V \tag{3.9}$$

$$dW_t^S dW_V^2 = \rho \tag{3.10}$$

$$2\kappa\theta \geq \sigma_V^2 \tag{3.11}$$

with the (3.8) represents asset price process at time t, and stochastic volatility of asset price is assumed to admit an Ornstein-Uhlenbeck process given in (3.9) where μ is growth rate or expected return of asset price, θ represent the long-term average from which the volatility diverges and to which it then returns κ is mean reversion speed coefficient (the larger the k, the longer it takes to return to θ), σ_V is volatility of the volatility, and it is generally responsible for the "scale" of randomness of the volatility process and dW^S, dW^V are two correlated standard Brownian motions with a $\rho \neq 0$ correlation. The Feller condition in (3.11) ensures the non-negative of variance.

By applying Ito's Lemma, which is discussed in Appendix, Heston model becomes:

$$d\ln(S_t) = \left(\mu - \frac{1}{2}V_t\right)dt + \sqrt{V_t}dW_t^S$$

$$dV_t = \kappa(\theta - V_t)dt + \sqrt{V_t}\sigma_V dW_t^V$$

$$dW_t^1 dW_t^2 = \rho$$

In order to estimate parameter and to forecast the price process, Heston model needs to be discretized. By Euler–Maruyama discretisation scheme (Kloeden and Platen, 1992), Heston model can be described as

$$\ln(S_t) - \ln(S_{t-\Delta t}) = \left(\mu - \frac{1}{2}V_{t-\Delta t}\right)\Delta t + \sqrt{V_{t-\Delta t}}\sqrt{\Delta t}\epsilon_t^S$$

$$V_t - V_{t-\Delta t} = \kappa(\theta - V_{t-\Delta t})\Delta t + \sqrt{V_{t-\Delta t}}\sqrt{\Delta t}\epsilon_t^V$$

$$\epsilon_t^S \sim \mathcal{N}(0,1)$$

$$\epsilon_t^V \sim \mathcal{N}(0,\sigma_V^2)$$

$$Corr(\epsilon_t^S, \epsilon_t^V) = \rho$$

Consider $\Delta t = \frac{1}{252}$ where 252 stands for the number of business days in a year, $\ln(S_t) - \ln(S_{t-\Delta t}) = r_t$ is daily return, and all parameters can be interpreted annually. According to Joshua Cape et al. (2014), in order to facilitate the estimation, let $\psi = \rho\sigma_V$ and $\Omega = \sigma_V^2(1-\rho^2)$, Heston model can be illustrated as

$$r_t = \left(\mu - \frac{1}{2}V_{t-\Delta t}\right)\Delta t + \sqrt{V_{t-\Delta t}}\sqrt{\Delta t}\epsilon_t^S \tag{3.12}$$

$$V_t = \kappa\theta\Delta t + V_{t-\Delta t}(1-\kappa\Delta t) + \sqrt{V_{t-\Delta t}}\sqrt{\Delta t}\epsilon_t^V \tag{3.13}$$

Equation (3.13) shows that volatility squared V_t is a function of $V_{t-\Delta_t}$ and other parameters. Therefore, HESTON can capture volatility clustering feature of returns. Furthermore, stochastic volatility setup allows random spike or drop in volatility. As a result, it will increase the probability of extreme movements in price, which then causes heavy-tailed distribution of returns because large price changes occurs more frequently in HESTON model than in standard Black-Scholes model.

After reorganization, it can be seen that

$$\epsilon_t^S = \frac{r_t - \mu\Delta t + \frac{1}{2}V_{t-\Delta t}}{\sqrt{V_{t-\Delta t}}\sqrt{\Delta t}},$$

$$\epsilon_t^V = \frac{V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-\Delta t}}{\sqrt{V_{t-\Delta t}}\sqrt{\Delta t}},$$

$$(\epsilon_t^S, \epsilon_t^V) \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & \rho\sigma_V \\ \rho\sigma_V & \sigma_V^2 \end{pmatrix}\right) = \mathcal{N}\left((0,0), \begin{pmatrix} 1 & \psi \\ \psi & \psi^2 + \Omega \end{pmatrix}\right)$$

Over the last decades, parameter estimation for stochastic differential equation has been investigated thoroughly. There are several approaches that have been studied and employed in financial mathematics. According to Gruszka, Jarosław, and Janusz Szwabinski (2023), Bayesian approach performs well in estimating parameters from the model class. In this study, Heston model is estimated by using Markov Chain Monte Carlo (MCMC) method, which bases on Bayesian inference. In general, the prior distribution for each parameter is assumed, then together with data observed, the posterior distribution is formalized. Under the assumption of Markov Chain and the repeated random sampling by Monte Carlo, the estimated parameters are achieved. The main content solely focus on the prior density assumption and posterior density derivation, which is developed in Joshua Cape et al. (2014).

MCMC distinguishes between parameter set and state space, where parameter set includes $\mu, \kappa, \theta, \sigma_V, \rho$, where state space $V_0, V_1, ...V_T$. These two components are estimated jointly using different random sampling methods which are discussed later in this chapter.

3.2.3 Estimation

The estimation for Heston model using MCMC starts with deriving likelihood function of (r_t, V_t). It can be seen from Equation (3.12) and (3.13) that (r_t, V_t) is a linear transformation of $(\epsilon_t^S, \epsilon_t^V)$ respectively. As $(\epsilon_t^S, \epsilon_t^V)$ is jointly normal distributed, (r_t, V_t) follows bivariate normal distribution with joint probability density function given below

$$P(r, V|\mu, \kappa, \theta, \psi, \Omega) = \Omega^{-n/2} \left(\prod_{t=1}^{n} \frac{1}{V_{t-1}} \right) \exp \left(-\frac{1}{2\Omega} \sum_{t=1}^{n} [(\Omega + \psi^2)(\epsilon_t^S)^2 - 2\psi \epsilon_t^S \epsilon_t^V + (\epsilon_t^V)^2] \right)$$

The joint likelihood of r_t, V_t is then used with the assumptions of prior distribution for parameters to obtain posterior distributions of parameters.

A. Posterior distribution of μ

Assuming that $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$, according to Bayes rule, the posterior of μ is given by

$$
\begin{aligned}
P(\mu \mid r, V, \kappa, \theta, \psi, \Omega) &\propto P(r, V \mid \mu, \kappa, \theta, \psi, \Omega) \cdot P(\mu) \\
&\propto \exp\left(-\frac{1}{2\Omega}\sum_{t=1}^{n}\left[(\Omega+\psi^2)(\epsilon_t^S)^2 - 2\psi\epsilon_t^V\epsilon_t^S\right]\right) \cdot \exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\sum_{t=1}^{n}\left[\frac{\Omega+\psi^2}{\Omega}\left(\frac{r_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t}{\sqrt{V_{t-1}\Delta t}}\right)^2\right.\right. \\
&\left.\left. -\frac{2\psi}{\Omega}\left(\frac{V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1}}{\sqrt{V_{t-1}\Delta t}}\right)\left(\frac{-\mu\Delta t}{\sqrt{V_{t-1}\Delta t}}\right)\right]\right) \\
&\cdot \exp\left(-\frac{\mu^2 - 2\mu_0\mu}{2\sigma_0^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\left(\sum_{t=1}^{n}\frac{\Omega+\psi^2}{\Omega V_{t-1}}\right)\mu^2\Delta t\right.\right. \\
&\left.\left. -2\sum_{t=1}^{T}\left(\frac{(\Omega+\psi^2)(r_t + \frac{1}{2}V_{t-1}\Delta t)}{\Omega V_{t-1}} - \frac{\psi\left(V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1}\right)}{\Omega V_{t-1}}\right)\mu\right]\right) \\
&\cdot \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma_0^2}\mu^2 - 2\frac{\mu_0}{\sigma_0^2}\mu\right]\right).
\end{aligned}
$$

After simplification, the $\mu \sim \mathcal{N}(\mu^*, \sigma^{*2})$ where

$$
\mu^* = \frac{\sum_{t=1}^{n}((\Omega+\psi^2)(r_t + \frac{1}{2}V_{t-1}\Delta t)/\Omega V_{t-1}) - \sum_{t=1}^{n}(\psi(V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1})/\Omega V_{t-1}) + \mu_0/\sigma_0^2}{\Delta t\sum_{t=1}^{n}((\Omega+\psi^2)/\Omega V_{t-1}) + 1/\sigma_0^2}
$$

$$
\sigma^{*2} = \frac{1}{\Delta t\sum_{t=1}^{n}((\Omega+\psi^2)/\Omega V_{t-1}) + 1/\sigma_0^2}
$$

B. Posterior distribution of ψ and Ω

Let the prior distribution assumptions be $\Omega \sim \mathcal{IG}(\tilde{\alpha}, \tilde{\beta})$ and $\psi_{|\Omega} \sim \mathcal{N}(\psi_0, \Omega/p_0)$. In Bayesian statistics, the inverse gamma distribution is commonly used as a prior for the variance of a normal distribution because it ensures the conjunction property.

First, the joint posterior density of (ψ, Ω) is as follows

$$
\begin{aligned}
P(\psi, \Omega \mid Y, V, \kappa, \theta, \mu) &\propto P(Y, V \mid \psi, \Omega, \kappa, \theta, \mu) \cdot P(\psi \mid \Omega) \cdot P(\Omega) \\
&\propto \Omega^{-n/2} \exp\left(-\frac{1}{2\Omega} \sum_{t=1}^{n} \left[(\Omega + \psi^2)(\epsilon_t^S)^2 + (\epsilon_t^V)^2 - 2\psi \epsilon_t^S \epsilon_t^V \right] \right) \\
&\cdot \sqrt{\frac{p_0}{\Omega}} \exp\left(-\frac{(\psi - \psi_0)^2}{2\Omega/p_0} \right) \cdot \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \Omega^{-\tilde{\alpha}-1} \exp\left(-\frac{\tilde{\beta}}{\Omega} \right) \\
&\propto \Omega^{-n/2-\tilde{\alpha}-1} \cdot \frac{1}{\Omega^{1/2}} \cdot \exp\left(-\frac{1}{2\Omega} \sum_{t=1}^{n} \left[(\Omega + \psi^2)(\epsilon_t^S)^2 + (\epsilon_t^V)^2 - 2\psi \epsilon_t^S \epsilon_t^V \right] \right. \\
&\left. -\frac{1}{2} \frac{(\psi - \psi_0)^2}{\Omega/p_0} - \frac{\tilde{\beta}}{\Omega} \right) \\
&\propto \Omega^{-n/2-\tilde{\alpha}-1} \cdot \exp\left(-\frac{1}{\Omega} \left[\tilde{\beta} + \frac{1}{2} \sum_{t=1}^{n} (\epsilon_t^V)^2 \right] \right) \cdot \exp\left(\frac{-1}{2\Omega} p_0 \psi_0^2 \right) \frac{1}{\Omega^{1/2}} \\
&\cdot \exp\left(-\frac{1}{2\Omega} \left[\left(p_0 + \sum_{t=1}^{n} (\epsilon_t^S)^2 \right) \psi^2 - 2 \left(p_0 \psi + \sum_{t=1}^{n} \epsilon_t^S \epsilon_{t=1}^V \right) \psi \right. \right. \\
&\left. \left. + \frac{(p_0 \psi_0 + \sum_{t=1}^{n} \epsilon_t^S \epsilon_t^V)^2}{p_0 + \sum_{t=1}^{n} (\epsilon_t^S)^2} - \frac{(p_0 \psi_0 + \sum_{t=1}^{n} \epsilon_t^S \epsilon_t^V)^2}{p_0 + \sum_{t=1}^{n} (\epsilon_t^S)^2} \right] \right)
\end{aligned}
$$

Consider the terms only relating to Ω, $\Omega \sim \mathcal{IG}(\alpha_*, \beta_*)$ with

$$\alpha_* = \frac{n}{2} + \tilde{\alpha}$$

and

$$\beta_* = \tilde{\beta} + \frac{1}{2} \sum_{t=1}^{n} (\epsilon_t^V)^2 + \frac{1}{2} p_0 \psi_0^2 - \frac{1}{2} \frac{(p_0 \psi_0 + \sum_{t=1}^{n} \epsilon_t^S \epsilon_t^V)^2}{p_0 + \sum_{t=1}^{n} (\epsilon_t^S)^2}$$

Additionally, $\psi_{|\Omega} \sim \mathcal{N}(\psi_*, \sigma_\psi^{*2})$ where

$$\psi_* = \frac{p_0 \psi_0 + \sum_{t=1}^{n} \epsilon_t^S \epsilon_t^V}{p_0 + \sum_{t=1}^{n} (\epsilon_t^S)^2}$$

and

$$\sigma \psi^{*2} = \frac{\Omega}{p_0 + \sum_{t=1}^{n} (\epsilon_t^S)^2}$$

C. Posterior distribution of θ

Assume that the prior of $\theta \sim \mathcal{N}(\theta_0, \sigma_\theta^2)$, then the posterior of θ is given by

$$
\begin{aligned}
P(\theta \mid r, V, \mu, \kappa, \psi, \Omega) &\propto P(r, V \mid \mu, \kappa, \theta, \psi, \Omega) \cdot P(\theta) \\
&\propto \exp\left(-\frac{1}{2\Omega}\sum_{t=1}^{n}\left[(\epsilon_t^V)^2 - 2\psi\epsilon_t^V\epsilon_t^S\right]\right) \cdot \exp\left(-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\sum_{t=1}^{n}\left[\frac{1}{\Omega}\left(\frac{V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1}}{\sqrt{V_{t-1}\Delta t}}\right)^2\right.\right. \\
&\quad \left.\left. -\frac{2\psi}{\Omega}\left(\frac{r_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t}{\sqrt{V_{t-1}\Delta t}}\right)\left(\frac{-\kappa\theta\Delta t}{\sqrt{V_{t-1}\Delta t}}\right)\right]\right) \\
&\quad \cdot \exp\left(-\frac{\theta^2 - 2\theta_0\theta}{2\sigma_\theta^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\left(\sum_{t=1}^{n}\frac{\kappa^2}{\Omega V_{t-1}}\right)\theta^2\Delta t\right.\right. \\
&\quad \left.\left. -2\sum_{t=1}^{n}\left(\frac{\kappa(V_t - (1-\kappa\Delta t)V_{t-1})}{\Omega V_{t-1}} - \frac{\psi(r_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t)\kappa}{\Omega V_{t-1}}\right)\theta\right]\right) \\
&\quad \cdot \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma_\theta^2}\theta^2 - 2\frac{\theta_0}{\sigma_\theta^2}\theta\right]\right).
\end{aligned}
$$

It requires some more reorganization to see that $\theta \sim \mathcal{N}(\theta^*, \sigma_\theta^2)$ with

$$
\theta^* = \frac{\sum_{t=1}^{n}(\kappa(V_t - (1-\kappa\Delta t)V_{t-1}))/\Omega V_{t-1}) - \sum_{t=1}^{n}(\psi(r_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t)\kappa/\Omega V_{t-1}) + \theta_0/\sigma_\theta^2}{\Delta t\sum_{t=1}^{n}(\kappa^2/\Omega V_{t-1}) + 1/\sigma_\theta^2}
$$

and

$$
\sigma^{*2} = \frac{1}{\Delta t\sum_{t=1}^{n}(\kappa^2/\Omega V_{t-1}) + 1/\sigma_\theta^2}
$$

Posterior distribution of κ

Let $\kappa \sim \mathcal{N}(\kappa_0, \sigma_\kappa^2)$, similarly the posterior density is as follows

$$
\begin{aligned}
P(\kappa \mid r, V, \mu, \theta, \psi, \Omega) &\propto P(r, V \mid \mu, \kappa, \theta, \psi, \Omega) \cdot P(\kappa) \\
&\propto \exp\left(-\frac{1}{2\Omega}\Sigma_{i=1}^{n}[(\epsilon_i^V)^2 - 2\psi\epsilon_i^V\epsilon_i^S]\right) \cdot \exp\left(-\frac{(\kappa-\kappa_0)^2}{2\sigma_k^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\Sigma_{i=1}^{T}\left[\frac{1}{\Omega}\left(\frac{V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1}}{\sqrt{V_{t-1}\Delta t}}\right)^2\right.\right. \\
&\left.\left. -\frac{2\psi}{\Omega}\left(\frac{r_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t}{\sqrt{V_{t-1}\Delta t}}\right)\left(-\frac{\kappa(\theta - V_{t-1})\Delta t}{\sqrt{V_{t-1}\Delta t}}\right)\right]\right) \\
&\cdot \exp\left(-\frac{(\kappa^2 - 2\kappa_0\kappa)}{2\sigma_k^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\left(\sum_{t=1}^{n}\frac{(V_{t-1}-\theta)^2}{\Omega V_{t-1}}\right)\kappa^2\Delta t\right.\right. \\
&\left.\left. -2\sum_{t=1}^{T}\left(\frac{(\theta - V_{t-1})(V_t - V_{t-1})}{\Omega V_{t-1}} - \frac{\psi(r_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t)(\theta - V_{t-1})}{\Omega V_{t-1}}\right)\kappa\right]\right) \\
&\cdot \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma_\kappa^2}\kappa^2 - 2\frac{\kappa_0}{\sigma_\kappa^2}\kappa\right]\right).
\end{aligned}
$$

After some transformations, posterior distribution of $\kappa \sim \mathcal{N}(\kappa_*, \sigma_\kappa^{*2})$ with

$$
\kappa^* = \frac{\sum_{t=1}^{n}((\theta - V_{t-1})(V_t - V_{t-1})/\Omega V_{t-1}) - \sum_{t=1}^{n}(\psi(r_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t)(\theta - V_{t-1})/\Omega V_{t-1}) + \kappa_0/\sigma_\kappa^2}{\Delta t\sum_{t=1}^{n}((V_{t-1}-\theta)^2/\Omega V_{t-1}) + 1/\sigma_\kappa^2}
$$

and

$$
\sigma_\kappa^{*2} = \frac{1}{\Delta t\sum_{t=1}^{n}((V_{t-1}-\theta)^2/\Omega V_{t-1}) + 1/\sigma_\kappa^2}
$$

Posterior distribution of V_t

$$P(V_t \mid r, V_{t+1}, V_{t-1}, \kappa, \mu, \theta, \psi, \Omega) = P(r, V_{t+1}, V_t, \mid V_{t-1}, \mu, \kappa, \theta, \psi, \Omega) \frac{P(V_{t-1} \mid \kappa, \theta, \psi, \Omega, \mu)}{P(r, V_{t+1}, V_{t-1} \mid \kappa, \theta, \psi, \Omega, \mu)}$$

$$\propto P(Y, V_{t+1}, V_t \mid V_{t-1}, \kappa, \theta, \psi, \Omega, \mu)$$

$$\propto \frac{1}{V_t \Delta t} \exp\left(-\frac{1}{2\Omega} \left[(\Omega + \psi^2)(\epsilon_{t+1}^S)^2 - 2\psi \epsilon_{t+1}^S \epsilon_{t+1}^V + (\epsilon_{t+1}^V)^2 \right] \right.$$

$$\left. -\frac{1}{2\Omega} \left[(\Omega + \psi^2)(\epsilon_i^S)^2 - 2\psi \epsilon_i^S \epsilon_i^V + (\epsilon_i^V)^2 \right] \right)$$

$$= \frac{1}{V_t \Delta t} \exp\left(-\frac{1}{2\Omega} \frac{(\Omega + \psi^2)(\frac{1}{2} V_t \Delta t + r_{t+1} - \mu \Delta t)^2}{V_t \Delta t} \right.$$

$$-\frac{1}{2\Omega} \frac{-2\psi(\frac{1}{2} V_t \Delta t + r_{t+1} - \mu \Delta t)(-(1 - \kappa \Delta t) V_t - \kappa \theta \Delta t + V_{t+1})}{V_t \Delta t}$$

$$-\frac{1}{2\Omega} \frac{(-(1 - \kappa \Delta t) V_t - \kappa \theta \Delta t + V_{t+1})^2}{V_t \Delta t}$$

$$-\frac{1}{2\Omega} \frac{-2\psi(r_t - \mu \Delta t + \frac{1}{2} V_{t-1} \Delta t)(V_t - \kappa \theta \Delta t - (1 - \kappa \Delta t) V_{1-1})}{V_{t-1} \Delta t}$$

$$\left. -\frac{1}{2\Omega} \frac{(V_t - \kappa \theta \Delta t - (1 - \kappa \Delta t) V_{t-1})^2}{V_{t-1} \Delta t} \right)$$

$$= \frac{1}{V_t \Delta t} \exp\left(-\frac{1}{2\Omega} \frac{(\Omega + \psi^2) \left(\frac{1}{2} V_t \Delta t + Y_{t+1} - \mu \Delta t \right)^2}{V_t \Delta t} \right)$$

$$-\frac{1}{2\Omega} \frac{-2\psi(\frac{1}{2} V_t \Delta t + r_{t+1} - \ mu \Delta t)(-(1 - \kappa \Delta t) V_t - \kappa \theta \Delta t + V_{t+1})}{V_t \Delta t}$$

$$\left. -\frac{1}{2\Omega} \frac{(-(1 - \kappa \Delta t) V_t - \kappa \theta \Delta t + V_{t+1})^2}{V_t \Delta t} \right)$$

$$\cdot \exp\left(-\frac{1}{2\Omega} \frac{-2\psi(r_t - \mu \Delta t + \frac{1}{2} V_{t-1} \Delta t)(V_t - \kappa \theta \Delta t - (1 - \kappa \theta) V_{t-1})}{V_{t-1} \Delta t} \right.$$

$$\left. -\frac{1}{2\Omega} \frac{(V_t - \kappa \theta \Delta t - (1 - \kappa \Delta t) V_{t-1})^2}{V_{t-1} \Delta t} \right).$$

3.2.4 MCMC

In general, MCMC generates random samples from a given target distribution. In this study, a sequence of sampling of target parameters $\mu, \kappa, \theta, \psi, \Omega$ and spaces $V_1, V_2, ..., V_n$ is generated based on posterior distributions which are derived from prior distribution assumption and observed data. By construction, the sequence has Markov's properties,

including convergence.

The convergence of the sequence relied on ergodic theory for Markov Chains. A g-step transition probability plays a key role in defining a Markov Chain:

$$P^{(g)}(x, A) = P[\theta^{(g)} \in A \mid \theta^0 = x]$$

There are two important conditions for the Markov Chain to converge, i.e., irreducible and aperiodic. First, a Markov chain is irreducible if it has a positive probability of eventually entering any set which has $\pi-$positive probability regardless of its initial state. Second, a chain is aperiodic if there are no portions of the state space that the chain visits at regularly spaced time intervals. When two conditions are met, the equilibrium distribution of the chain can be achieved:

$$\lim_{g\to\infty} P[\theta^{(g)} \in A \mid \theta^{(0)}] = \pi(A)$$

It can be seen from the equation that the chain will converge regardless of the initial state. This study assumes the convergence of estimated parameters after g-steps sampling. While the convergence condition is complex to verify, it is investigated through graphical illustration in this study by tracking the value of sampling at each step.

After defining the prior distribution and deriving the posterior density of parameters set $\mu, \kappa, \theta, \rho$ and state space V_t, MCMC is employed to obtain draws from the posterior distribution. MCMC approaches used for parameters and state space differ. In order to estimate parameters $\mu, \kappa, \theta, \rho$, the Gibbs sampler is helpful because the posterior distributions are normal distribution and inversed gamma distribution, which are well-known and straightforward to sample conditioned on other parameters. Gibbs sampler is the simplest MCMC algorithm that was introduced by Geman and Geman (1984).

The procedure of Gibbs sampler for parameter set starts with initializing a set of values $\Theta = \mu^{(0)}, \kappa^{(0)}, \theta^{(0)}, \psi^{(0)}, \Omega^{(0)}, V_0^{(0)}, V_1^{(0)}, ..., V_n^{(0)}$. First, the distribution of μ is identified by the density mentioned above and these initial values, then a sampling $\mu^{(1)}$ can be obtained from this posterior distribution of μ. Parameter set now is updated to $\Theta = \mu^{(1)}, \psi^{(0)}, \Omega^{(0)}, \kappa^{(0)}, \theta^{(0)}, V_0^{(0)}, V_1^{(0)}, ..., V_n^{(0)}$. The same process will be applied to obtain draws of $\psi^{(1)}$ and $\Omega^{(1)}$, and $\Theta = \mu^{(1)}, \psi^{(1)}, \Omega^{(1)}, \kappa^{(0)}, \theta^{(0)}, V_0^{(0)}, V_1^{(0)}, ..., V_n^{(0)}$. $\kappa^{(1)}$ and $\theta^{(1)}$ can be acquired by following steps mentioned above.

Unlike parameter sampling that can be obtained easily with the Gibbs sampler, the

distribution of state space $V = V_0, V_1, ..., V_n$ is not straightforward to simulate. In this case, a random walk Metropolis-Hastings approach is appropriate Joshua Cape et al. (2014). Similarly, the algorithm starts with initializing values for the 0th step:

$$V_0^{(0)}, V_1^{(0)}, ..., V_n^{(0)}$$

.

For $g \in 1, 2, ...G$, after drawing parameters using Gibbs sampler, the algorithm is run to obtain

$$V_0^{(g)}, V_1^{(g)}, ..., V_n^{(g)}$$

. Specifically, $V_t^{(g)}$ for $t \in 1, 2, ..., n$ is generating as follows:

$$V_t^{*(g)} = V_t^{(g-1)} + \epsilon_t, \text{ with } \epsilon_t \sim \mathcal{N}(0, \sigma_N^2)$$

where $t \in 1, 2, ..., n$ and $V_t^{*(g)}$ is new proposal for $V_t^{(g)}$. σ_N^2 is pre-defined rejection rate.

New proposal $V_t^{(g)}$ is rejected or accepted based on the comparison of likelihood of $V_t^{(g-1)}$ and new one. The computation of acceptance rate for new proposal is given by:

$$\mathcal{A}(V_t^{*(g)}, V_t^{(g-1)}) = min\left(\frac{\pi(V_t^{*(g)})}{\pi(V_t^{(g-1)})}, 1\right)$$

.

where $\pi(V_t^{*(g)})$ and $\pi(V_t^{(g-1)})$ are likelihood calculated as follows:

$$\begin{aligned}
\pi(V_t^{*(g)}) = \frac{1}{V_t^{*(g)}\Delta t} \exp\Bigg(& -\frac{1}{2\Omega}\frac{(\Omega+\psi^2)(\frac{1}{2}V_t^{*(g)}\Delta t + r_{t+1} - \mu\Delta t)^2)}{V_t^{*(g)}\Delta t} \\
& -\frac{1}{2\Omega}\frac{-2\psi(\frac{1}{2}V_t^{*(g)}\Delta t + r_{t+1} - \mu\Delta t)(-(1-\kappa\Delta t)V_t^{*(g)} - \kappa\theta\Delta t + V_{t+1}^{(g-1)})}{V_t^{*(g)}\Delta t} \\
& -\frac{1}{2\Omega}\frac{(-(1-\kappa\Delta t)V_t^{*(g)} - \kappa\theta\Delta t + V_{t+1}^{(g-1)})^2}{V_t^{*(g)}\Delta t}\Bigg) \\
& \cdot \exp\Bigg(-\frac{1}{2\Omega}\frac{-2\psi(r_t - \mu\Delta t + \frac{1}{2}V_{t-1}^{(g)}\Delta t)(V_t^{*(g)} - \kappa\theta\Delta t - (1-\kappa\theta)V_{t-1}^{(g)})}{V_{t-1}^{(g)}\Delta t} \\
& -\frac{1}{2\Omega}\frac{(V_t^{*(g)} - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1}^{(g)})^2}{V_{t-1}^{(g)}\Delta t}\Bigg).
\end{aligned}$$

and

$$\pi(V_t^{(g-1)}) = \frac{1}{V_t^{(g-1)}\Delta t}\exp\left(-\frac{1}{2\Omega}\frac{(\Omega+\psi^2)(\frac{1}{2}V_t^{(g-1)}\Delta t + Y_{t+1} - \mu\Delta t)^2)}{V_t^{(g-1)}\Delta t}\right.$$
$$-\frac{1}{2\Omega}\frac{-2\psi(\frac{1}{2}V_t^{(g-1)}\Delta t + r_{t+1} - \mu\Delta t)(-(1-\kappa\Delta t)V_t^{(g-1)} - \kappa\theta\Delta t + V_{t+1}^{(g-1)})}{V_t^{(g-1)}\Delta t}$$
$$\left.-\frac{1}{2\Omega}\frac{(-(1-\kappa\Delta t)V_t^{(g-1)} - \kappa\theta\Delta t + V_{t+1}^{(g-1)})^2}{V_t^{(g-1)}\Delta t}\right)$$
$$\cdot\exp\left(-\frac{1}{2\Omega}\frac{-2\psi(r_t - \mu\Delta t + \frac{1}{2}V_{t-1}^{(g)}\Delta t)(V_t^{(g-1)} - \kappa\theta\Delta t - (1-\kappa\theta)V_{t-1}^{(g)})}{V_{t-1}^{(g)}\Delta t}\right.$$
$$\left.-\frac{1}{2\Omega}\frac{(V_t^{(g-1)} - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1}^{(g)})^2}{V_{t-1}^{(g)}\Delta t}\right).$$

$\mathcal{A}(V_t^{*(g)}, V_t^{(g-1)})$ then is compared with random sampling U from $\mathcal{U}[0,1]$. Intuitively, the probability of acceptance equals $\mathcal{A}(V_t^{*(g)}, V_t^{(g-1)})$. The acceptance/rejection decision rule is described as

$$\text{If } U < \mathcal{A}(V_t^{*(g)}, V_t^{(g-1)}), \text{ then } V_t^{(g)} = V_t^{*(g)},$$
$$\text{Otherwise, } V_t^{(g)} = V_t^{(g-1)}$$

Metropolis-Hastings sampling is iterated for $t \in 1, 2, ..., n$ in order to determine state space $V_1^{(g)}, V_2^{(g)}, ...V_n^{()}$. For $V_0^{(g)}$, the $\pi(V_0^{*(g)})$ and $\pi(V_0^{(g-1)})$ is computed considering only the first exponential part.

$$\pi(V_0^{*(g)}) = \frac{1}{V_0^{*(g)}\Delta t}\exp\left(-\frac{1}{2\Omega}\frac{(\Omega+\psi^2)(\frac{1}{2}V_0^{*(g)}\Delta t + Y_1 - \mu\Delta t)^2)}{V_0{*}(g)\Delta t}\right.$$
$$-\frac{1}{2\Omega}\frac{-2\psi(\frac{1}{2}V_0^{*(g)}\Delta t + r_1 - \mu\Delta t)(-(1-\kappa\Delta t)V_0^{*(g)} - \kappa\theta\Delta t + V_1^{(g-1)})}{V_0^{*(g)}\Delta t}$$
$$\left.-\frac{1}{2\Omega}\frac{(-(1-\kappa\Delta t)V_0^{*(g)} - \kappa\theta\Delta t + V_1^{*(g-1)})^2}{V_0^{*(g)}\Delta t}\right)$$

and

$$\pi(V_0^{(g-1)}) = \frac{1}{V_0^{(g-1)}\Delta t}\exp\left(-\frac{1}{2\Omega}\frac{(\Omega+\psi^2)(\frac{1}{2}V_0^{(g-1)}\Delta t + r_1 - \mu\Delta t)^2)}{V_0^{(g-1)}\Delta t}\right.$$
$$-\frac{1}{2\Omega}\frac{-2\psi(\frac{1}{2}V_0^{(g-1)}\Delta t + r_1 - \mu\Delta t)(-(1-\kappa\Delta t)V_0^{(g-1)} - \kappa\theta\Delta t + V_1^{(g-1)})}{V_0^{(g-1)}\Delta t}$$
$$\left.-\frac{1}{2\Omega}\frac{(-(1-\kappa\Delta t)V_0^{(g-1)} - \kappa\theta\Delta t + V_1^{(g-1)})^2}{V_0^{(g-1)}\Delta t}\right)$$

Similarly, V_{T+1} can be obtained by considering the second exponential part which is given by

$$\pi(V_{T+1}^{*(g)}) = \frac{1}{V_{T+1}^{*(g)}\Delta t}\exp\left(-\frac{1}{2\Omega}\frac{-2\psi(r_T - \mu\Delta t + \frac{1}{2}V_T^{(g)}\Delta t)(V_T^{*(g)} - \kappa\theta\Delta t - (1-\kappa\theta)V_T^{(g)})}{V_T^{(g)}\Delta t}\right.$$
$$\left.-\frac{1}{2\Omega}\frac{(V_{T+1}^{*(g)} - \kappa\theta\Delta t - (1-\kappa\Delta t)V_T^{(g)})^2}{V_T^{(g)}\Delta t}\right).$$

and

$$\pi(V_{T+1}^{(g-1)}) = \frac{1}{V_{T+1}^{(g-1)}\Delta t}\exp\left(-\frac{1}{2\Omega}\frac{-2\psi(r_T - \mu\Delta t + \frac{1}{2}V_T^{(g)}\Delta t)(V_{T+1}^{(g-1)} - \kappa\theta\Delta t - (1-\kappa\theta)V_T^{(g)})}{V_T^{(g)}\Delta t}\right.$$
$$\left.-\frac{1}{2\Omega}\frac{(V_{T+1}^{(g-1)} - \kappa\theta\Delta t - (1-\kappa\Delta t)V_T^{(g)})^2}{V_T^{(g)}\Delta t}\right).$$

3.3 Evaluation Measurement

3.3.1 Continuous ranked probabilistic score

While mean square error (MSE) or mean absolute error (MAE) is popular when it comes to the evaluation of forecast accuracy, probabilistic forecast, which provides forecasts for the future and suitable measures of the uncertainty associated with them, requires different criteria to assess the performance of forecasters. Scoring rules play a crucial role in both forecast quality measures (Gneiting, T., and Raftery, A. E., 2007). In some fields, researchers refer to the evaluation task served by scoring rules as forecast verification. Specifically, scoring rules provide a single numerical score based on the forecast curve and the actual value. In this paper, continuous ranked probabilistic score (CRPS), which is

derived based on predictive cumulative distribution functions, is an appropriate measure for prediction performance.

Introduced in Gneiting and Raftery (2007), Let X be a random variable, F be the cumulative distribution function (CDF) of X, such as $F(y) = P[X \leq y]$, x be the observation and F be the CDF associated with an empirical probabilistic prediction, CPRS is defined as follows

$$CRPS(F, x) = -\int_{-\infty}^{\infty} (F(y) - \mathbb{I}_{(y \geq x)})^2 \, dy$$

$\mathbb{I}$ is the indicator function that attains:

$$\begin{cases} \mathbb{I}_{(y \geq x)} & = 1, if y \geq x \\ \mathbb{I}_{(y \geq x)} & = 0, if y < x \end{cases}$$

CRPS can be equivalently written as

$$CRPS(F, x) = \frac{1}{2} E_F[|X - X'|] - E_F[|X - x|],$$

where X and X' are independent samples of random variable with distribution function F and x is the observed value.

Moreover, CRPS can usually computed in a closed form. For example, the CPRS of predictive density which is Gaussian distributed $\mathcal{N}(\mu, \sigma^2)$

$$CRPS(\mathcal{N}(\mu, \sigma^2), x) = \sigma \left[\frac{1}{\pi} - 2\varphi \left(\frac{x - \mu}{\sigma} \right) - \frac{x - \mu}{\sigma} \left(2\Phi \left(\frac{x - \mu}{\sigma} \right) - 1 \right) \right]$$

with φ and Φ is the probability density function and cumulative distribution function of a standard Gaussian distribution respectively.

In addition, CRPS is usually utilized with negative orientation as follows

$$CRPS^*(F, x) = -CRPS(F, x) = E_F|X - x| - \frac{1}{2} E_F|X - X'|$$

The lower the $CRPS^*$ is, the better the forecast is. The negative orientation version is more popular in practice because it provides the computation in the same unit as the observation, which makes the $CRPS^*$ easier to interpret. $CRPS^*$ is reduced to MAE if F is a deterministic forecast (point forecast with probability 1):

$$CRPS^*(\hat{X}, x) = E[|\hat{X} - x|]$$

In case of evaluation samples instead of density, $CRPS^*$ can be simply estimated as follows

$$CRPS^*(F, x) = \frac{1}{m}\sum_{i=1}^{m}|X_i - x| - \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}|X_i - X_j|,$$

where x is observed value, F denotes forecast distribution given m discrete sample $X_1, ..., X_m$.

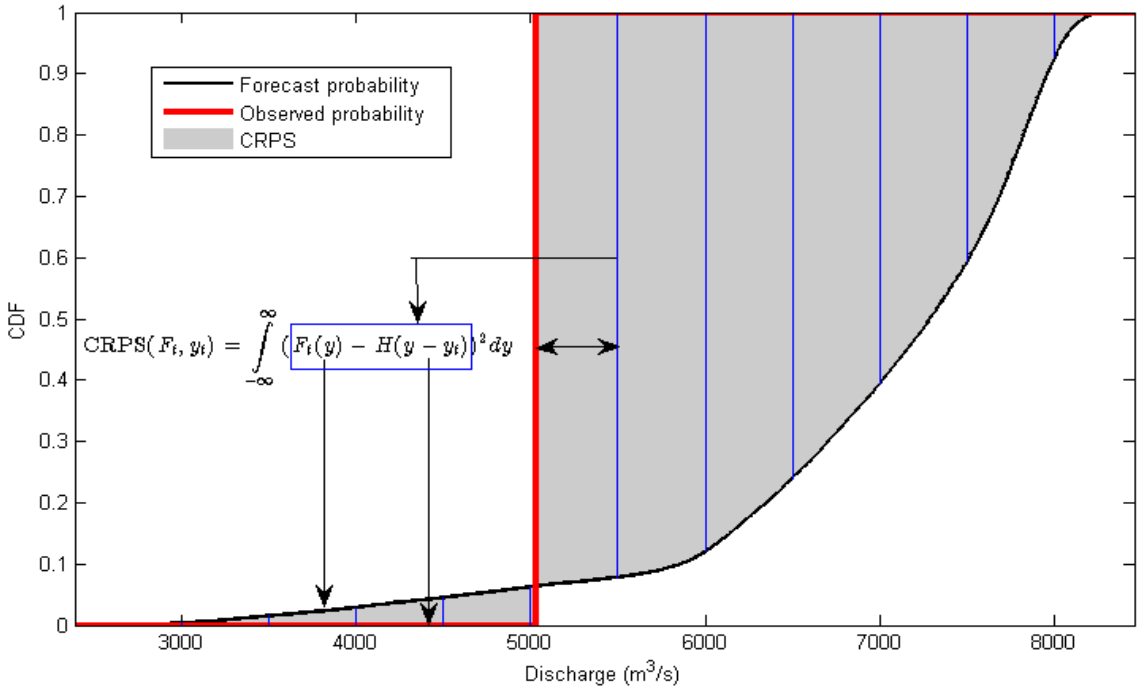

Figure 7: CPRS graphical illustration - Source: Durga Lal Shrestha (MATLAB)

The intuition of CRPS is clearly illustrated in Figure 7. In real life, only one value x is observed, then the observed probability density function is

$$P(y) = \begin{cases} 1, & y = y_t \\ 0 & otherwise \end{cases}$$

Consequently, observed CDF becomes

$$F(y) = \begin{cases} 0, & y < y_t \\ 1, & y \geq y_t \end{cases}$$

Put simply, it is shown in Figure 7 that CRPS represents the gray area between the forecast CDF and observed CDF. The smaller the area, the more accurate the probabilistic forecast.

In this thesis, the forecast accuracy of Heston and EGARCH models is investigated using CRPS. A comparative analysis is conducted not only descriptively but also via hypothesis testing. To statistically compare the performance of the two models, the Wilcoxon signed-rank test, which is an alternative to the t-test when the normal distribution assumption does not hold, is employed to analyze whether one of the two models performs significantly better than the other. The normally distributed assumption is replaced by a weaker assumption that the difference between the performances of 2 models is symmetric around its mean. Wilcoxon signed-rank test then investigates if the central value is significantly different from zero.

The Wilcoxon signed-rank test evaluates the median difference between the CRPS of the Heston and EGARCH models, with the hypothesis as follows.

- Null hypothesis (H0): The median difference between CRPSs of Heston and EGARCH is zero. Neither model performs better than the other.
- Alternative hypothesis (H1): The median difference is not zero. There is a significant difference in the models' performance.

3.3.2 Delta-neutral replication performance

A natural gas price forecast is an intermediate step required for various purposes, such as risk management, derivative pricing, and investment strategy; therefore, accurate option pricing and hedging are more important than forecasting prices alone. Moreover, the performance of delta hedges, which helps reduce the risk associated with price changes, is another way to represent the model's ability to capture price dynamics.

The general idea of delta hedges is to adjust the replicating portfolio periodically, aiming to maintain the value of the portfolio as closely as possible to the value of the derivative. This thesis investigates only the delta hedge accuracy used to hedge the risk associated with price volatility. In cases where multiple risk sources contribute to price and volatility, delta hedging is likely insufficient to assess the hedging performance. For the sake of simplicity and to maintain a fair ground for comparison, although it is

evident that delta hedges are not perfect, imperfect hedging is conducted to evaluate the performance of EGARCH and Heston model. Furthermore, to maintain relevance to real-world practice, daily hedge accuracy is taken into account in the comparative performance analysis of EGARCH and the Heston model.

To evaluate the futures price forecast, a portfolio consisting of a call option and underlying asset is considered, where a call option is a financial contract that gives the holder the right, but not the obligation, to buy a specific amount of an underlying asset (such as stocks, commodities, or indices) at a predetermined price (known as the strike price) within a specified period (until the expiration date). The value of the call option is determined by the expected payoff at expiration as follows

$$C(F, K) = e^{-rT}\mathbb{E}\left[(F_T - K)^+\right]$$

with F_T is the price of the underlying asset at expiration and K is strike price of the option. The option's price is computed numerically given by

$$C_t(F_t, K) = \frac{1}{n}\sum_{i=1}^{n}(\hat{F}_i - K)\mathbb{1}_{(K,\infty)}(\hat{F}_i)$$

with $\hat{F}_i$ is the price simulation by relevant model. In the context of this thesis, underlying asset is futures contract of natural gas and its $\hat{F}_T$ is the curve generated by relevant models. Considering a portfolio Π consisting of:

- Short position in call option C
- Long position in Δ amount of the underlying F
- A cash account B

By construction, the portfolio $\Pi = \Delta F - C + B$, with $\Delta = \frac{\partial C}{\partial F}$ measuring the sensitivity if an option's price to small changes in the price of the underlying asset. Similarly to call option's price, Δ is computed numerically as follows:

$$\Delta_t = \frac{C_t(F_t + \epsilon, K) - C_t(F_t - \epsilon, K)}{2\epsilon}$$

with small ϵ.

By continuously adjusting the position in the underlying asset by long $\Delta_{t+1} - \Delta_t$ amount of underlying, the portfolio is delta-neutral. Together with the assumption for

EGARCH model that price process has only one source of randomness from ϵ, the portfolio is risk-free.

The non-arbitrage condition of the market is ensured by minimizing the value of the portfolio, which results in a fair price of the option. Consequently, the absolute value of the portfolio at the expiration date, as determined by the relevant models, serves as a metric to assess the performance of models in terms of probabilistic forecasting.

$$|\Pi_T| = |\Delta_{T-1} F_T - C_T + B_{T-1}|$$

In simple words, the call option with specific strike price (i) is short at time t, while the replicating portfolio includes $\Delta_{i,t}$ units of the asset and cash position of $B_{i,t}$. The mean absolute hedging error at the maturity date of N options with different moneyness is measured as performance metric given as

$$PnL = \frac{1}{N} \sum_{i=1}^{N} |\Pi_{i,T}|$$

where moneyness is defined as

$$moneyness = \frac{K}{S_0} \text{ with K: Strike price, and } S_0\text{: asset price at valuation date}$$

The absolute error is investigated in order to evaluate the hedging accuracy rather than the expected PnL. Computing absolute error ensures that every deviation from zero of the PnL counts and positive and negative errors do not cancel out. The smaller the hedge error at the expiration date, the more effective the model is in generating probabilistic forecasts.

3.4 Forecasting design

The analysis, including parameterization and forecasting performance of the models, was conducted for each futures contract. Specifically, the analysis was performed twice for each contract: once 3 months before the expiry and once 1 month before the expiry, denoted as T. The estimation window is fixed at 3 months of data; then, CPRS was computed using the model forecasts and the price at expiry.

EGARCH is estimated by minimizing its negative log-likelihood, while Markov Chain Monte Carlo is used to estimate Heston. I implemented them in Python and utilized some built-in packages, including Pandas, NumPy, SciPy, and the SST distribution (Berrisch, 2021). The procedure of parameter estimation and probabilistic forecast is as follows:

3.4.1 EGARCH

For each futures contract:

- EGARCH(1,1) is fitted on 3 months log return using MLE approach by minimizing negative log likelihood in equation (3.2) to obtain $\hat{\varphi} = (\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \hat{\nu})$ only once.
- Initial volatility is computed as $\sigma_0 = (e^{\frac{\hat{\omega}}{1-\hat{\beta}}})^{0.5}$

Condition variance and price ($\hat{S}_T$) are computed recursively by

- Sampling 1000 i.i.d random variable z_t from skewed Student-t distribution SST(mean = 0, variance = 0, skew = $\hat{\delta}$, degree of freedom = $\hat{\nu}$).
- Computing $\hat{\sigma}_t$ where $\ln(\hat{\sigma}_t^2) = \hat{\omega} + \hat{\alpha}(|z_{t-1}| - \mathbb{E}[|z_{t-1}|]) + \hat{\gamma} z_{t-1} + \hat{\beta} \ln(\hat{\sigma}_{t-1}^2)$
- Computing $\hat{\epsilon}_t = \hat{\sigma}_t z_t$
- $\hat{S}_t = \hat{S}_{t-1} e^{\hat{\mu} + \hat{\epsilon}_t}$

Finally, $CRPS^*$ is computed for $\hat{S}_T$ and observed price at maturity S_T.

3.4.2 Heston

First, the prior distribution for parameters is partly taken from Cape et al. (2014) with some modification according to the observations in real data

$$\mu \sim N(0, 1)$$
$$\kappa \sim N(0, 1)$$
$$\theta \sim N(0, 1)$$
$$\Omega \sim IG(2, 0.01)$$
$$\Psi \sim N(0, \Omega/2)$$
$$V \sim N(0.09, 0.1)$$

MCMC algorithm is run to estimate the Heston's parameters as follows

- Randomly choose initial values of μ_0, κ_0, θ_0, Σ_0, $\Psi_0 \sim N(0, \Sigma_0/2)$
- Sampling $V_0, V_1, ..., V_T$ from chosen truncate normal distribution to ensure the non-negative value for variance
- Computing mean and variance for μ according to the posterior distribution of μ. Sampling new value of μ_0 from the defined distribution of μ
- Computing β parameter for IG according to posterior distribution of β. Replacing Ω_0 by random sample from $IG(2, \beta)$
- Repeating same procedure for Ψ, θ, κ.
- Propose new V = V + N(0, 0.001), then compute the likelihood of current V and proposed V
- Compute the acceptance rate and compare with random variable from Uniform distribution U(0,1).

In this study, MCMC is repeated 15000 times with 5000 sampling as burn-in period and only the last 10000 iterations is used to estimate parameters. After achieving the estimated parameters, forecast for returns and eventually price are computed following the procedure

- Sampling $Z_V \sim N(0,1)$ and $Z_{corr} \sim N(0,1)$ independently. Then $Z_S = \hat{\rho} Z_V + \sqrt{1-\hat{\rho}^2} Z_{corr}$.
- Compute simulated volatility according to equation (3.13).
- Compute simulated returns according to equation (3.12), then the forecast price is $\hat{S}_t = \hat{S}_{t-1} e^{\hat{r}_t}$.

When considering CRPS, the probabilistic price forecast is sufficient to get the final metric. On the other hand, additional computation is required to derive the performance of the EGARCH and the Heston model in delta hedging. To avoid computational cost, the model's parameters are estimated once and used throughout the valuation period. Secondly, probabilistic price forecasts are achieved based on the estimated parameters,

assuming different initial prices. Third, the delta is calculated according to the equation. Finally, the replicating portfolio is adjusted according to delta. The procedure is repeated until the expiry date for three moneyness levels: 0.95, 1, and 1.05. The mean of absolute delta hedge error represents the performance of models.

4 Empirical Analysis

4.1 Model diagnostics

4.1.1 Estimated parameters

Following the estimation method outlined in the methodology chapter, the parameters are estimated for both models for three-month and one-month-ahead forecasts. This thesis focuses on comparing the forecast performance of the Heston and EGARCH models; however, it is also essential to validate the parameters estimation and in-sample fitted values.

Figure 8 and Figure 9 show the time series of Heston parameters for three month-ahead and one month-ahead forecast, respectively. The range of parameters is similar for both cases, except for long-term volatility θ, which can be explained by the liquidity of the futures contract when the time to expiry becomes smaller. Specifically, when the contract expires in a shorter time frame, there is a higher demand for the commodity; therefore, volatility increases.

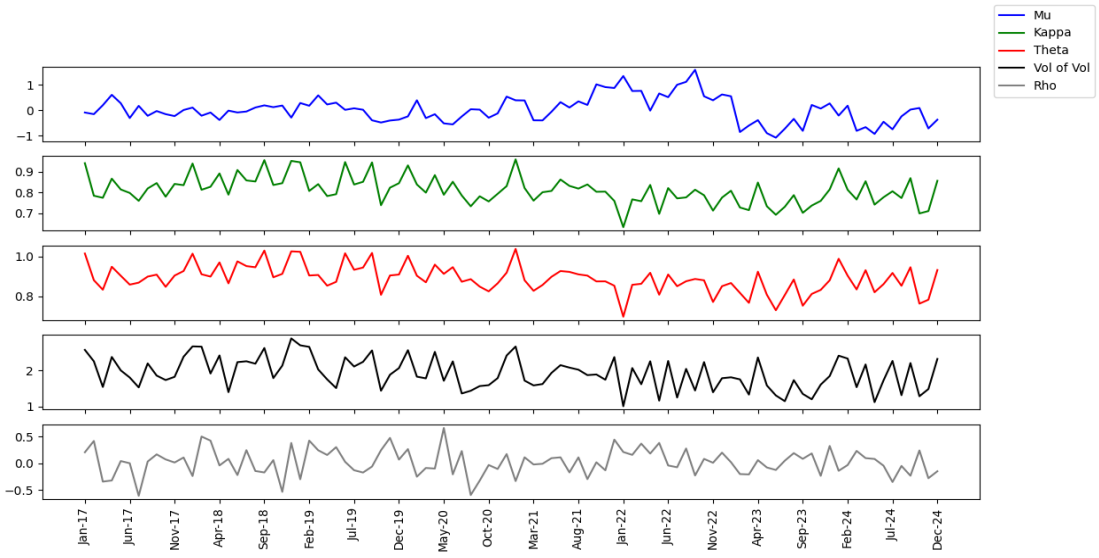

Figure 8: Heston parameter estimates from price calibration for 3 month-ahead

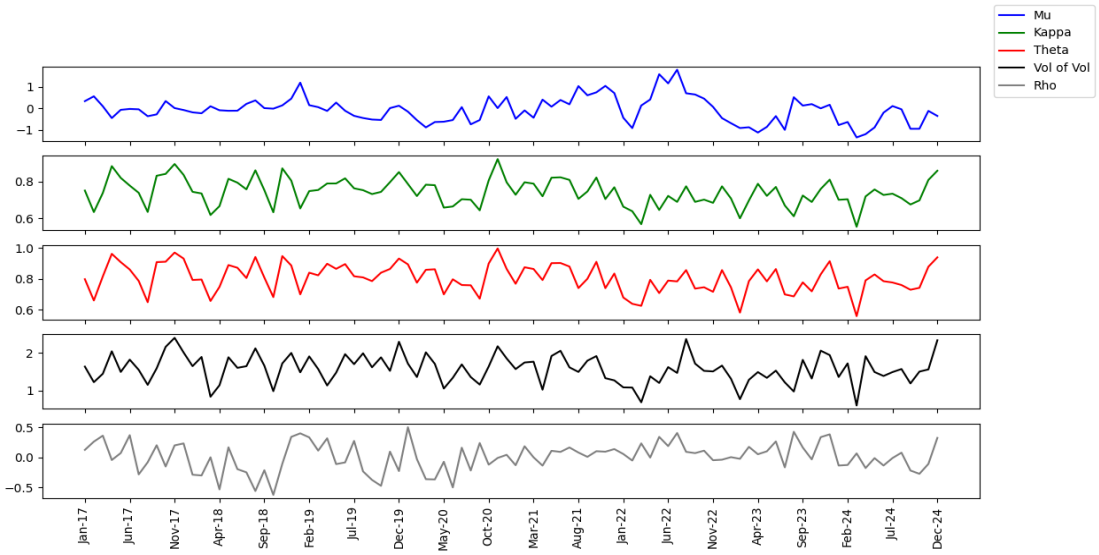

Figure 9: Heston parameter estimates from price calibration for 1 month-ahead

The estimated parameters of the EGARCH model are presented in Figure 10 for the three-month-ahead forecasts and in Figure 11 for the one-month-ahead forecasts. These graphs demonstrate a remarkable similarity in the parameter patterns between the two forecasting horizons, suggesting a consistent underlying volatility structure across different time frames. One of the most notable findings is the value of the skewness, δ, which is estimated to be 1 in both cases. This result holds significant implications for the distributional properties of returns over time. Specifically, a δ value of 1 indicates that the standardized innovations follow a symmetric Student's t-distribution, meaning that extreme positive and negative returns are equally likely. This symmetry in the conditional distribution implies that the model does not detect asymmetric tail behavior—where large adverse shocks might have a different impact on volatility than large positive shocks of the same magnitude. Furthermore, the consistency in parameter patterns across the one-month and three-month horizons suggests that the volatility dynamics are stable and persistent over different forecasting windows.

4.1.2 In-sample performance

In-sample CRPS of Heston and EGARCH are illustrated in Figure 12 and Figure 13, respectively. The results shown in this graph indicate that, in general, the EGARCH model provides a more accurate fitted distribution compared to the Heston model across most estimation periods. However, it is noticeable that the Heston model performs relatively

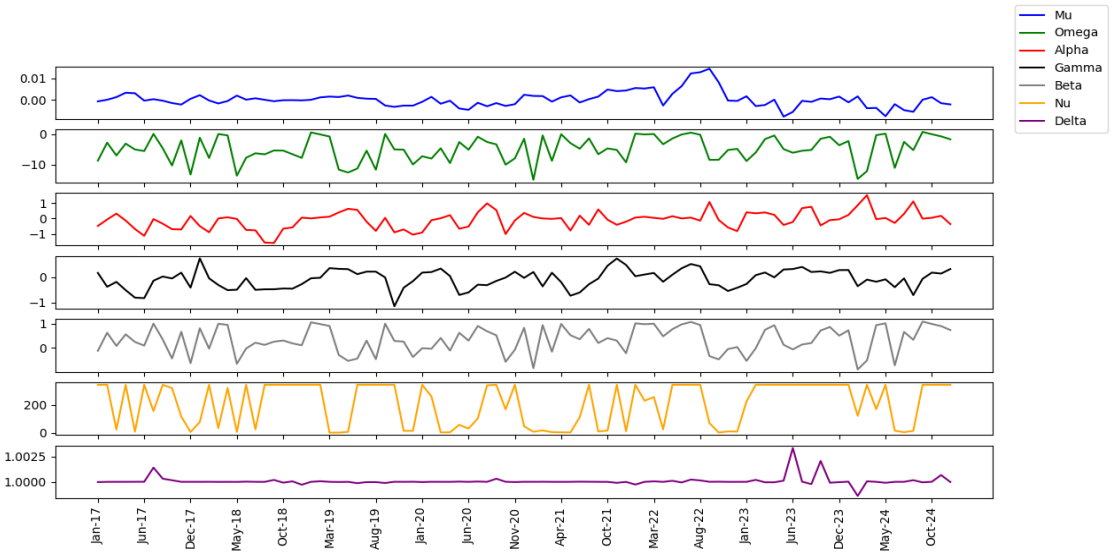

Figure 10: EGARCH parameter estimates from price calibration for 3 month-ahead

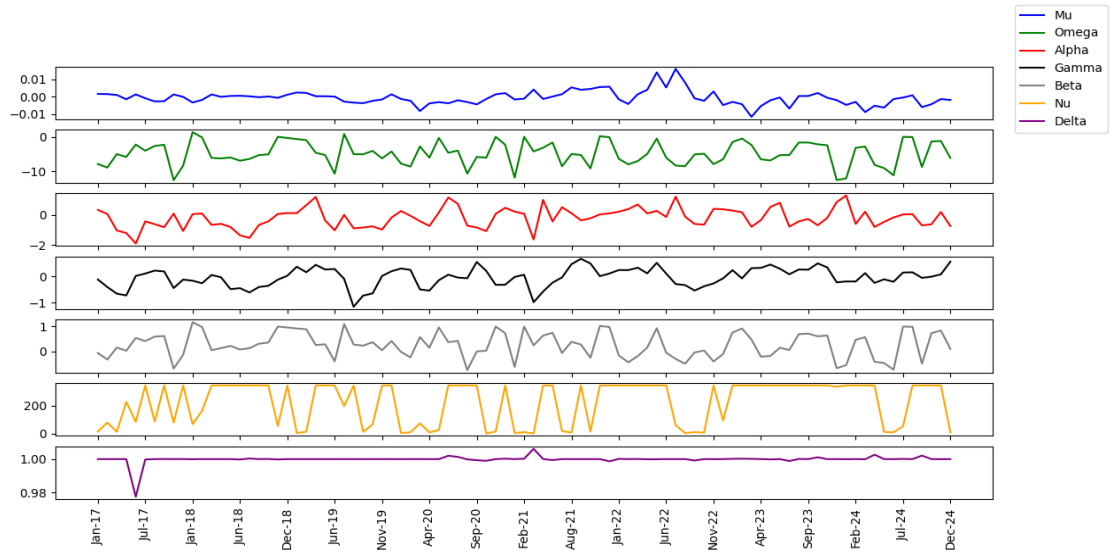

Figure 11: EGARCH parameter estimates from price calibration for 1 month-ahead

better during extreme market events. To be specific, during the Russian-Ukrainian war in 2022, while there was a dramatic geopolitical uncertainty that caused a shock in natural gas supply, the Heston model performed better in forecasting price distribution for three-month-ahead predictions. A similar pattern is observed for the one-month-ahead forecast in early 2020, when COVID-19 caused COVID-19 caused market stress, and in 2022.

In order to statistically assess the comparison of the performance of the in-sample forecast of the two models in terms of CRPS, Wilcoxon signed-rank tests were conducted for

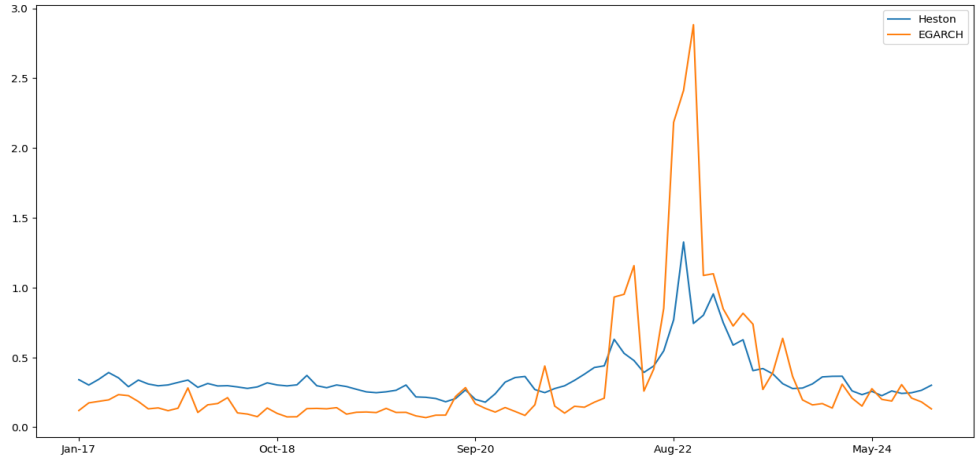

Figure 12: In-sample CPRS of 3 month-ahead models

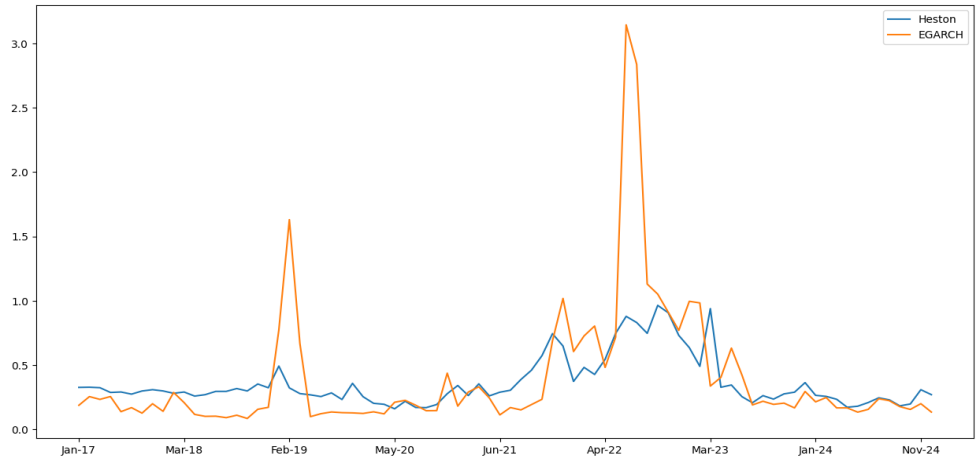

Figure 13: In-sample CPRS of 1 month-ahead models

both the one-month-ahead and three-month-ahead forecast horizons. The tests evaluated the following hypothesis:

H0: The performances of in-sample fitted distribution by Heston and EGARCH model are indifferent.

H1: EGARCH achieves lower (better) CRPS values than Heston.

It is shown in Table 1 that the Wilcoxon tests rejected H0 at $p = 0.01$ for both forecast horizons, confirming that EGARCH's in-sample CRPS is statistically superior to that of Heston. The result aligns with the graphical illustration from Figure 12 and Figure 13 that

N month-ahead	p-value	Conclusion
3 month-ahead	0.0002	In-sample fitted distribution of EGARCH is more accurate
1 month-ahead	0.025	In-sample fitted distribution of EGARCH is more accurate

Table 1: Wilcoxon signed-rank tests for in-sample CRPS

EGARCH's fitted distribution is close to the observed price in most periods. However, the consistent in-sample outperformance of EGARCH might be the result of overfitting, where EGARCH's parameters estimation captures noises rather than the actual process.

4.2 Predictive performance

4.2.1 3 month-ahead forecast

This study empirically evaluates the in-sample and out-of-sample forecasting performance of the EGARCH and Heston models for natural gas futures prices, using probabilistic metrics (CRPS) across 96 contracts from 2017 to 2024. The probabilistic forecast for three-month-ahead and one-month-ahead natural gas with the observed time series is shown in Figure 14 and Figure 15, respectively. It is obvious from the graphs that the density of price forecast changes over the study period. From January 2017 to July 2021, the forecast density spreads less, which is the result of lower conditional variance. On the other hand, the volatility is larger in the volatile period from late 2021 to the end of 2022, which leads to a dramatically increasing wide range of price forecasts. The median of the probabilistic forecast is also illustrated in the graphs. Overall, both models can identify the volatile phases; however, the Heston model provides a probabilistic forecast with a smaller range and is closer to the observed values. Next, CRPS is presented, and the Wilcoxon test on CRPS is conducted to thoroughly compare the performance of the EGARCH and the Heston model.

Figure 16 shows the forecasting performance of 3 month-ahead models. It can be seen from the Figure 16that in 2022, CRPS of both models are considerably higher than CRPS during the rest of the time. It is clear in **??** that there is no obvious difference between the CRPS of Heston and EGARCH except for the period of the beginning of Russian-Ukrainian conflict. Moreover, Heston clearly outperform EGARCH during the 2022. In order to have statistical evidence for comparative analysis, non-parametric Wilcoxon tests

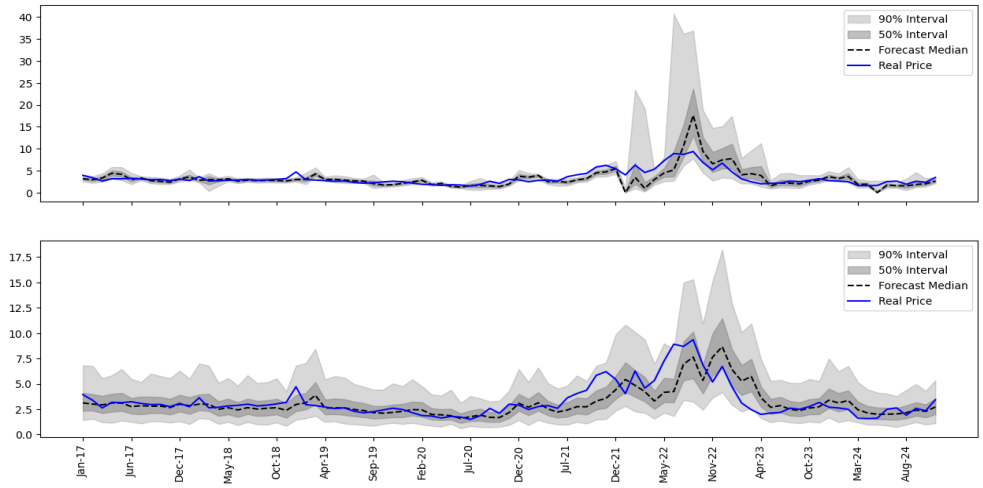

Figure 14: Three month-ahead price probabilistic price forecast by EGARCH and Heston

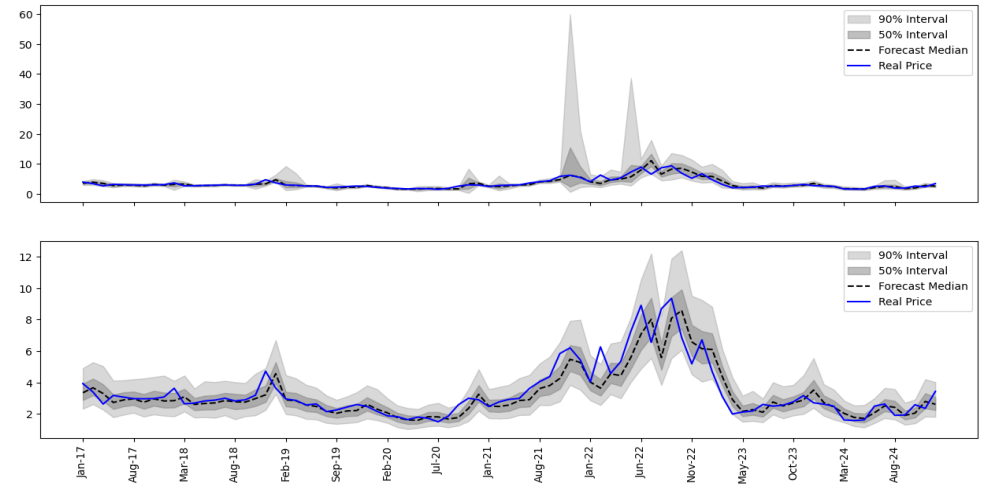

Figure 15: One month-ahead price probabilistic price forecast by EGARCH and Heston

are run for the whole period without 2022 and for the whole period from 2017 to 2024.

Wilcoxon tests are employed to test the following hypothesis:

H0: The performances of Heston and EGARCH model are indifferent.

H1: The performances of 2 models are different.

Table 1 show the result of Wilcoxon tests. In both cases, the null hypothesis is not rejected which means that there is no significant different between the forecasting performance of Heston and EGARCH model for 3 month-ahead forecast.

CRPS of EGARCH when extreme events happen are noticeably higher than CRPS of

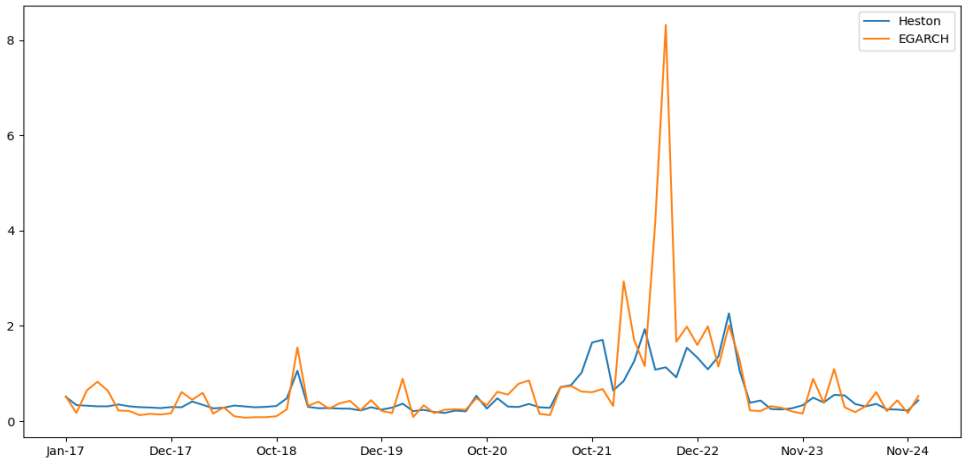

Figure 16: CPRS of 3 month-ahead models

Investigation period	p-value	Conclusion
2017-2024	0.36	Fail to reject null hypothesis
2017-2024 (without 2022)	0.88	Fail to reject null hypothesis

Heston model for both in-sample and out-sample for 3 month-ahead forecast. Therefore, even though Heston is computationally expensive and not statistically significant better than EGARCH, one should consider employing HESTON model when there is a concern that an essential event happens unexpectedly.

4.2.2 1 month-ahead forecast

The comparison of CRPS from EGARCH and Heston for 1 month-ahead forecast can be seen from Figure 17. The finding for 1 month-ahead prediction is similar to the one for 3 month-ahead analyzed above, that EGARCH's CRPS is noticeably higher than CRPS of Heston during Covid time in 2019 and Russian-Ukrainian war in 2022. On the other hand, the performance of both models is not dramatically different in terms of CRPS for the rest of the time.

Two-sided Wilcoxon tests are conducted to investigate the comparison of CRPS from two models, which shows the results as follows:

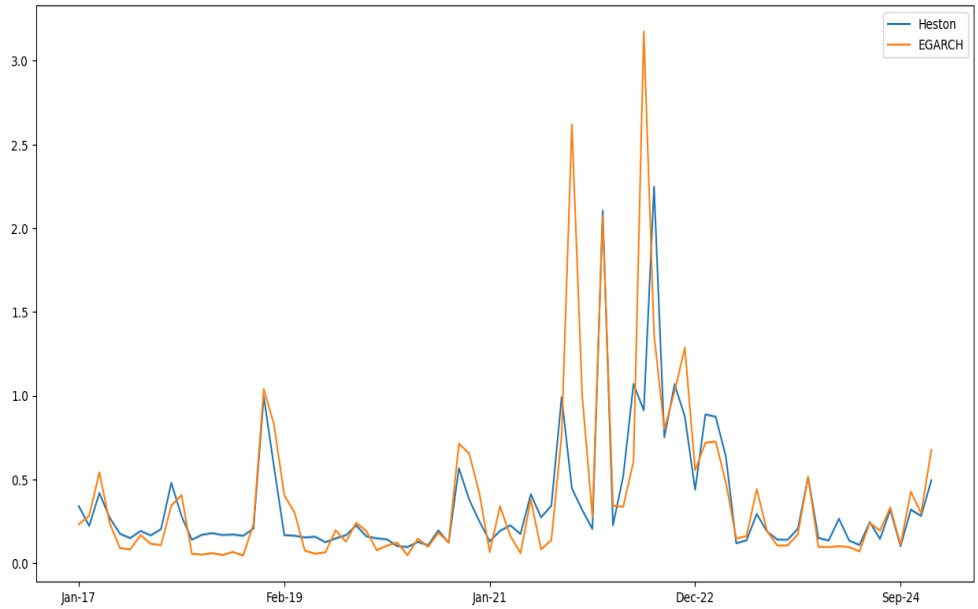

Figure 17: CPRS of 1 month-ahead models

Investigation period	p-value	Conclusion
2017-2024	0.311	Fail to reject null hypothesis
2017-2024 (without 2022)	0.186	Fail to reject null hypothesis

4.3 Delta Hedge Performance

Figure 18 shows the performance of EGARCH and Heston model when they are implement to delta hedge a call option with one-month to expiry. It is noticeable that profit and loss (PnL) of Heston model is more stable than the one based on EGARCH price forecast. Then fluctuation in PnL of EGARCH is clear during the 2022 when one of the main supply of natural gas was disrupted because of Russian-Ukrainian war. It is not obvious whether Heston outperforms EGARCH model in terms of delta hedge efficiency; therefore, I run Wilcoxon test to answer the question if there is a significant difference between these PnL from two models. The hypothesis is as follows

- H0 EGARCH and Heston's delta hedged PnL performance is not significantly different.
- H1 There is a significant difference in the delta hedged PnL performance of EGARCH

and Heston model.

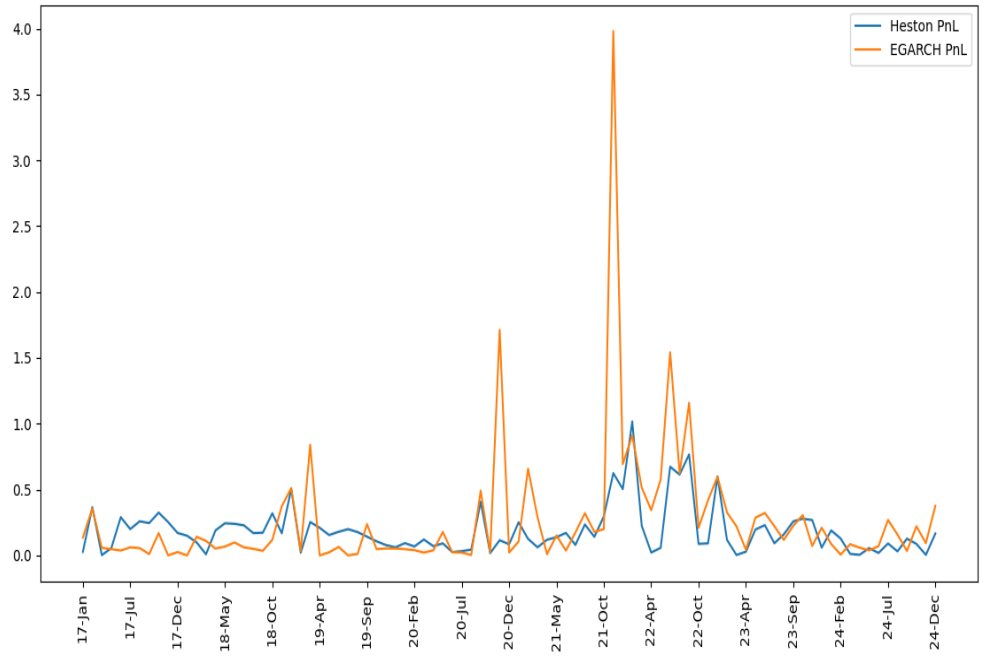

Figure 18: Absolute PnL Delta Hedge for Call option with one month to expiry

Investigation period	p-value	Conclusion
2017-2024	0.655	Fail to reject null hypothesis

Although PnL based on Heston is noticeably better than one for EGARCH during 2022, the result of Wilcoxon test shows that there is no significant difference in the performance of the two model in terms of delta hedges errors.

5 Conclusion

This thesis conducts a comparative analysis of the EGARCH and Heston models in terms of their performance in probabilistic forecasting for natural gas, considering futures prices from January 2017 to December 2024. In general, EGARCH and Heston models can be used to represent the volatility clustering of financial time series. EGARCH is primarily used as an econometric model to derive value at risk (VaR), a well-known metric in risk management. At the same time, Heston is renowned for its application in derivative pricing. Probabilistic price forecasts play a key role in achieving VAR or derivative prices.

To maintain a fair ground for comparison, I investigate the EGARCH model, assuming that error terms follow a skewed Student-t distribution. As a result, both models can capture the heavy tails and skewness characteristics of empirical log returns. EGARCH(1,1) and Heston are employed to forecast three-month-ahead and one-month-ahead probabilistic prices at maturity date for 96 futures contracts during the period from 2017 to 2024. The econometric model EGARCH is estimated by maximizing the log-likelihood, assuming a skewed student t distribution of residuals. Meanwhile, the Markov Chain Monte Carlo method is employed to estimate parameters for the Heston model. In-sample fitted values and out-of-sample forecasts are used to assess the models' representation of the natural gas price process. The focus of the study is to evaluate the forecast performance using various probabilistic metrics, including the CRPS and absolute errors of delta hedges.

I highlight four main findings. First, EGARCH's estimated skewness parameter δ is typically one for most contracts, which implies the symmetric of residuals. Additionally, the parameters of EGARCH and the Heston model fluctuate within the same range for both cases of one-month-ahead and three-month-ahead forecast, except for the long-term volatility of the Heston model, which can be explained by the fact that there are more trades for futures contracts when the time to expiry is getting close, leading to high volatility.

Secondly, the in-sample performance of the EGARCH model is better than that of Heston in terms of CRPS, as determined by the Wilcoxon test. The same relative result applies for both one month ahead and three months ahead. However, in the case of extreme events, such as the Russian-Ukrainian war, the in-sample fitted distribution is not accurate, as evidenced by the CRPS.

Third, I find limited evidence for the relative difference in probabilistic forecast performance of the EGARCH and Heston models with respect to CRPS. Even though it is clear from graphical illustrations that Heston outperforms EGARCH for one-month and three-month ahead forecasts in the period when important events leading to supply disruptions, there is no significant statistical testing result that supports this finding.

Fourth, the performance of delta hedges of EGARCH and Heston are investigated by applying a replicating portfolio approach. The mean of absolute delta hedge errors of multiple options with different strikes is utilized as a metric to measure the accuracy of probabilistic forecasts. The finding is similar to the result of comparing CRPS that the delta-hedge error of the Heston model is smaller during 2019 and 2022. On the other hand, it is not clear whether EGARCH or Heston performs better in the comparative analysis for the remainder of the study period.

Overall, it is more computationally expensive to estimate the parameters of the Heston model, as many sampling iterations have to be run in order to get the parameters to converge. Although there is no testing proof for the improvement in probabilistic forecasts evaluated by CRPS and delta hedge errors when using the Heston model, it is recommended to use the Heston model when it is uncertain what will happen in the future, affecting the supply of natural gas.

Finally, a simple prior distribution is utilized when estimating the Heston model, which does not consider much industry knowledge in order to create a fair ground for comparison. One can consider modifying the prior distribution to reflect the real world to achieve better results. Furthermore, it is noteworthy that delta hedges are not perfect as they consider only the uncertainty in the price process. For that reason, a more sophisticated hedging procedure, including vega hedge $\frac{\Delta C}{\Delta \sigma}$, can be investigated to measure the performance of models accurately.

References

[1] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31. 307-327.

[2] Cape et al. (2014). Estimating Heston's and Bates' models parameters using Markov chain Monte Carlo simulation. *Journal of Statistical Computation and Simulation*, 85. 2295-2314.

[3] International Energy Agency. (2024). Natural gas demand growth picks up in 2024 amid uncertainties over supply. International Energy Agency. https://www.iea.org/news/natural-gas-demand-growth-picks-up-in-2024-amid-uncertainties-over-supply

[4] Energy Information Administration. (2024). Carbon Dioxide Emissions Coefficients. U.S. Department of Energy. https://www.eia.gov/environment/emissions/CO2-vol-mass.php

[5] De Gooijer JG, Hyndman RJ (2006). 25 years of time series forecasting. *Int J Forecast*, 22(3). 443-327.

[6] Ferrari, D., Ravazzolo, F., & Vespignani, J. (2021). Forecasting energy commodity prices: A large global dataset sparse approach. *Energy Economics*, 98, 105268.

[7] Wang, Y., Liu, L., & Wu, C. (2020). Forecasting commodity prices out-of-sample: Can technical indicators help?*International Journal of Forecasting*, 36(2), 666–683

[8] Sadorsky, P., 1999. Oil price shocks and stock market activity. *Energy Economics*, 21(5), 449–469.

[9] Reboredo, J., 2011. How do crude oil prices co-move?: A copula approach. *Energy Economics*, 33(5), 948–955.

[10] Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society* , 987–1007.

[11] Nelson, D., 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, 347–370.

[12] Dragulescu, A. A., & Yakovenko, V. M. (2002). Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative finance*, 2(6), 443.

[13] Mikhailov, S., & Nögel, U. (2004). Heston's stochastic volatility model: Implementation, calibration and some extensions. *John Wiley and Sons*.

[14] Oyuna, D., & Yaobin, L. (2021). Forecasting the Crude Oil Prices Volatility With Stochastic Volatility Models. *Sage Open*, 11(3)

[15] Berrisch, J., & Ziel, F. (2022). Distributional modeling and forecasting of natural gas prices. *Journal of Forecasting*, 41(6), 1065–1086.

[16] Gruszka, Jarosław, & Janusz Szwabinski. 2023. Parameter Estimation of the Heston Volatility Model with Jumps in the Asset Prices. *Econometrics*, 11: 15.

[17] Cape, J., Dearden, W., Gamber, W., Liebner, J., Lu, Q., & Nguyen, M. L. (2014). Estimating Heston's and Bates' models parameters using Markov chain Monte Carlo simulation. *Journal of Statistical Computation and Simulation*, 85(11), 2295–2314.

[18] Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). "A 30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts?. *Risk Analysis: An International Journal*, 25(3), 623-629.

[19] Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, 7, 376–388

[20] Wan, C., Xu, Z., Pinson, P., Dong, Z. Y., & Wong, K. P. (2013). Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE transactions on power systems*, 29(3), 1033-1044.

[21] Zhu, X., & Genton, M. G. (2012). Short-term wind speed forecasting for power system operations. *International Statistical Review*, 80(1), 2-23.

[22] Pesaran, M. H., & Timmermann, A. (2000). A recursive modelling approach to predicting UK stock returns. *The Economic Journal*, 110(460), 159-191.

[23] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2), 327-343.

[24] Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2), 385-407.

[25] Liu, J., Qiu, B., Du, P., Zhao, X., & Zhu, J. (2025). A novel probabilistic connectivity network link prediction model for natural gas price based on an improved K-shell algorithm. *Physica A: Statistical Mechanics and its Applications*, 130672.

[26] Ambya, A., Gunarto, T., Hendrawaty, E., Kesumah, F. S. D., & Wisnu, F. K. (2020). Future natural gas price forecasting model and its policy implication. *International Journal of Energy Economics and Policy*, 10(5), 64-70.

[27] Reichert, B., & Souza, A. M. (2022). Can the Heston model forecast energy generation? A systematic literature review. *International Journal of Energy Economics and Policy*, 12(1), 289-295.

[28] Benth, F. E. (2011). The stochastic volatility model of Barndorff-Nielsen and Shephard in commodity markets. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 21(4), 595-625.

[29] Dong, Y., Wen, S.H., Hu, X.B. & Li, J.C. (2020), Stochastic resonance of drawdown risk in energy market prices. *Physica A: Statistical Mechanics and its Applications*, 540, 123098.

[30] Hsu, C.C., Chen, A.S., Lin, S.K. & Chen, T.F. (2017), The affine styled-facts price dynamics for the natural gas: Evidence from daily returns and option prices. *Review of Quantitative Finance and Accounting*, 48(3), 819-848.

[31] Fama, E. F. (1970). Efficient capital markets. *Journal of finance, 25(2)*, 383-417.

[32] Nelson, D. B., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59: 347-370.

[33] Fernández, C., & Steel, M. F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American statistical association*, 93(441), 359-371.

[34] Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542-547.

[35] Francq, C., & Zakoian, J. M. (2019). GARCH models: structure, statistical inference and financial applications. *John Wiley & Sons*.

[36] Kloeden, P. E., Platen, E., Kloeden, P. E., & Platen, E. (1992). Stochastic differential equations (pp. 103-160). *Springer Berlin Heidelberg*.

[37] Gruszka, J., & Szwabiński, J. (2023). Parameter estimation of the Heston volatility model with jumps in the asset prices. *Econometrics*, 11(2), 15.

[38] Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 721-741.

[39] Gneiting, T., Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477), 359-378.

[40] Berrisch, J. (2021). sstudentt 0.1.1: A python implementation of the skewed student-t distribution.

Appendix