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
January 16, 2025

Supervisor: Prof. Matei Demetrescu
Prof. Christoph Hanck

Contents

1 Introduction **1**

2 Data **3**

3 Methodology **7**

3.1 EGARCH 7

3.1.1 EGARCH (1,1) Model 7

3.1.2 Estimation 8

3.2 Heston model 10

3.2.1 Model 10

3.2.2 Estimation 12

3.2.3 MCMC 15

3.3 n-day ahead forecast 19

3.4 Evaluation Measurement 19

3.5 pseudo code 21

4 Empirical Analysis **21**

5 Conclusion **21**

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. It makes natural gas the third-largest source of energy after oil and coal. Moreover, it is published in The U.S. Energy Information Administration report that natural gas produces 50% less CO_2 than coal and 30% less CO_2 than oil when burned for energy. Therefore, natural gas has been becoming a reliable and relatively cleaner source of energy given the fact that it support the climate-neutral transformation scheme. Robust gas price forecast enables decision-making units to make informed decisions regarding energy consumption, investment strategies, resource allocation and risk management. While point forecast seems to be not sufficient 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, which can serve for multiple purposes, e.g. scenario analysis, complex derivative pricing, portfolio optimization,...

De Gooijer and Hyndman (2006), in their analysis of time series forecasting research from 1982 to 2005, which encompasses over 940 papers, conclude that the utilization of prediction intervals and densities, or probabilistic forecasting, has significantly increased 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 prevalent for several decades (Gigerenzer et al. 2005), 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; these challenges consist of meteorological and climatic forecasting (Collins & Knight 2007; Gneiting & Raftery 2005; Palmer 2002, 2012), projections regarding the accessibility of renewable energy resources (Pinson 2013, Zhu & Genton 2012), and economic and financial risk assessment (Groen et al. 2013, Timmermann 2000). As natural gas has been becoming an important financial commodity (Hong et al., 2020), there is a significant growth in the literature on natural gas price forecasting in the last decade where a wide range of different methodologies is employed: econometric model, statistical machine learning model and fundamental model. Ferrari et al. (2021) proposed dynamic factor models to forecast commodities' prices, including oil, natural gas an coal, when using a large dataset if macroeconomic variables. The penalized maximum likelihood

approach enables the the shrink to zero of parameters, which allows the model to get rid of the curve of dimension. They found that their model outperformed machine learning technique, such as LASSO and elastic net. Similarly, macroeconomics and technical trading rules are used to build commodities' price density forecast (Wang et al., 2020). It showed that technical indicators perform better than economic variables in predicting the density of commodity price.

In order to have robust probabilistic price forecast, accurate volatility prediction is required. Over the past few decades, there has been a significant evolution in the literature on energy market volatility. Introduced by Engle (1982), Autoregressive Conditional Heteroskedasticity (ARCH) is well-known time-varying volatility model, which was further developed in Bellerslev (1986) by including the ARMA structure, known as Generalized ARCH (GARCH) model. ARCH and GARCH have been extensively employed in econometric literature to model conditional variance characteristic of time series data due to the fact that they enable the representation of volatility cluster characteristic of financial return. Despite their popularity, GARCH model faces some limitations. First, asymmetric effects shocks on volatility is not covered in GARCH model, while it is a pronounced characteristic in energy market (Sadorsky, 1999) and Reboredo, 2011). Second, as volatility is required to be positive, the estimated parameters of GARCH model have to be non-negative, which make the GARCH model difficult to estimate. Exponential GARCH (EGARCH) model, introduced by Nelson (1991), not only capture the asymmetric effect of shocks on volatility but also simplify the estimation by eliminating the non-negativity condition for estimated parameters.

While EGARCH model has been widely employed for discrete-time series in econometric analysis, Heston model, which captures stochastic volatility as well as the correlation between price shock and volatility shock, is a benchmark model when it comes to model stochastic volatility from mathematical perspective. 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, 1999). Moreover, this model can capture the heavy-tailed nature of the return distributions, leverage effect, and volatility clustering, which make it one of the most famous stochastic volatility models for option pricing and risk management. Primarily, Heston model was used mainly for equity and currency (Adrian A Dragulescu & Victor M Yakovenko, 2002 and Mikhailov,

S., & Nögel, U., 2004). Recently, Oyuna, D., & Yaobin, L. (2021) applied Heston model to forecast crude oil price and their finding showed that the stochastic volatility model performs better than traditional GARCH model.

This paper 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.

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.

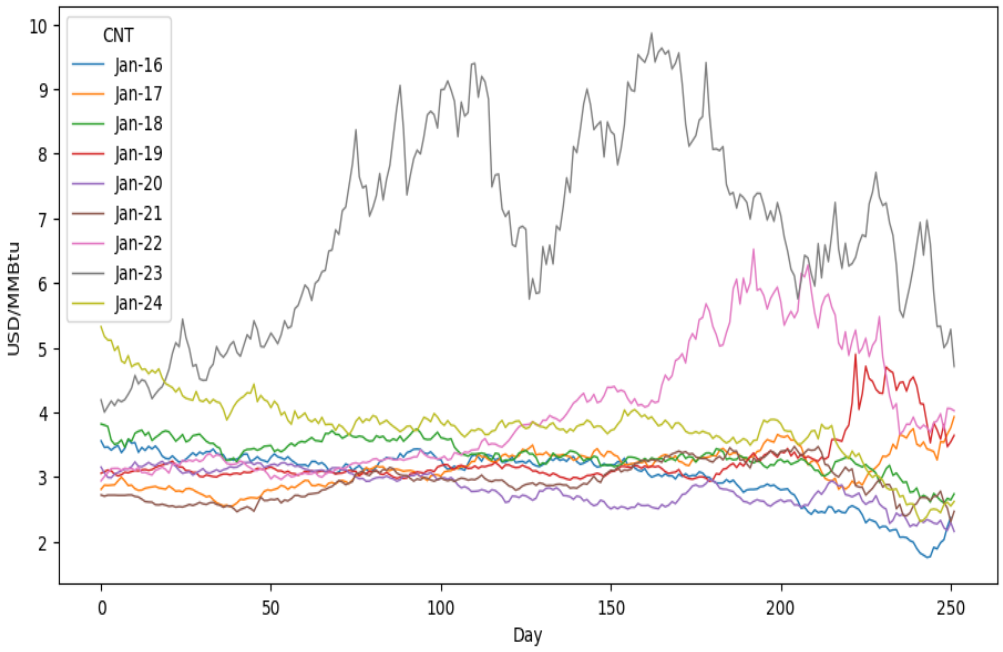

Figure 1: January Futures Contract daily prices within one year before expiry

Figure 3 and Figure 4 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.

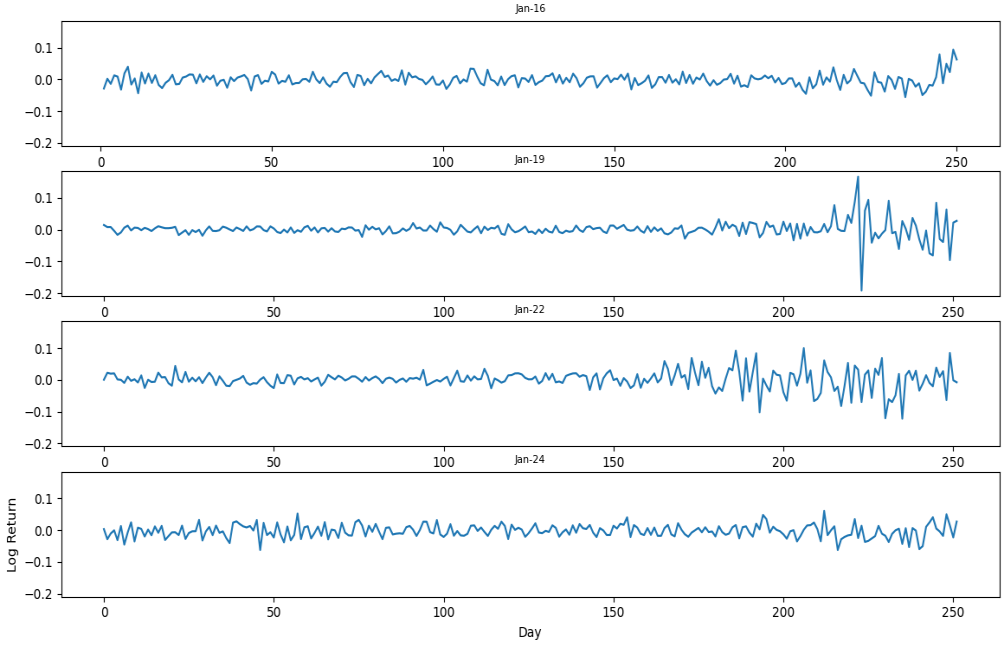

Figure 2: January Futures Contract daily log return within one year before expiry

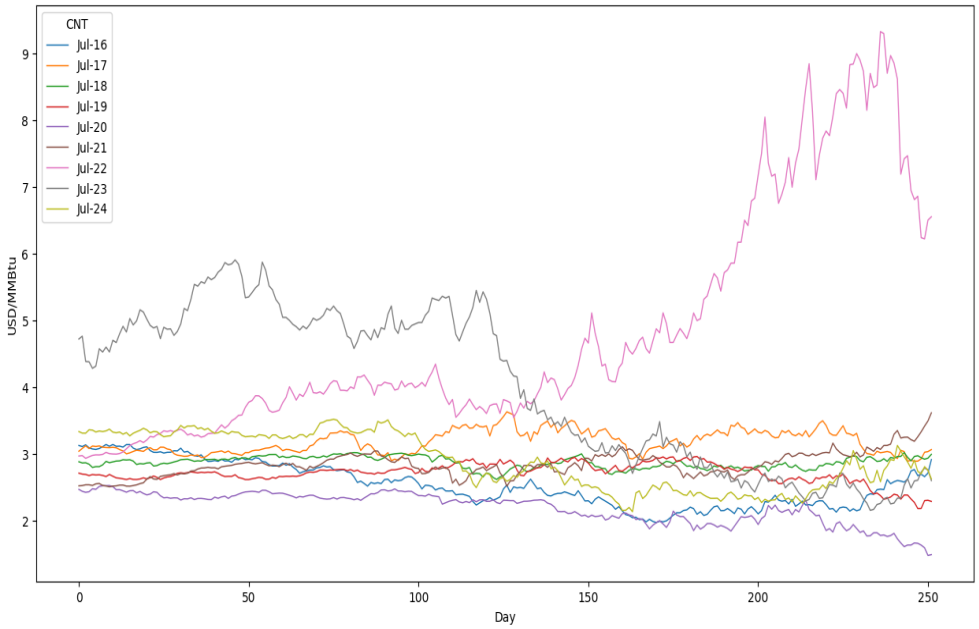

Figure 3: July Futures Contract daily prices within one year before expiry

Figure 5 shows the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of January 2016 logarithm return at first moment and second moment.

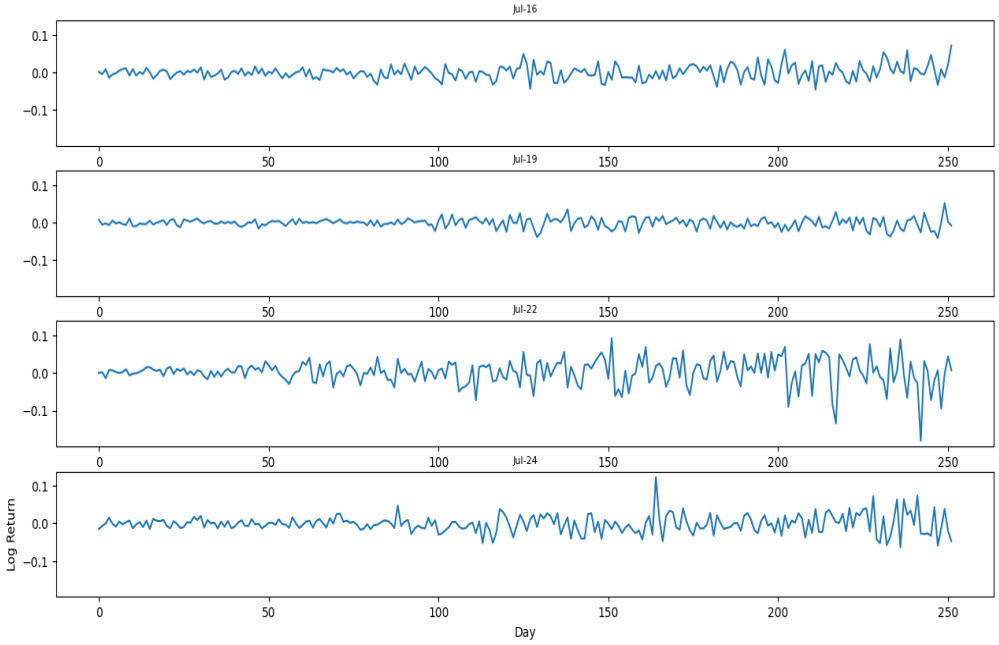

Figure 4: July Futures Contract daily log return within one year before expiry

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

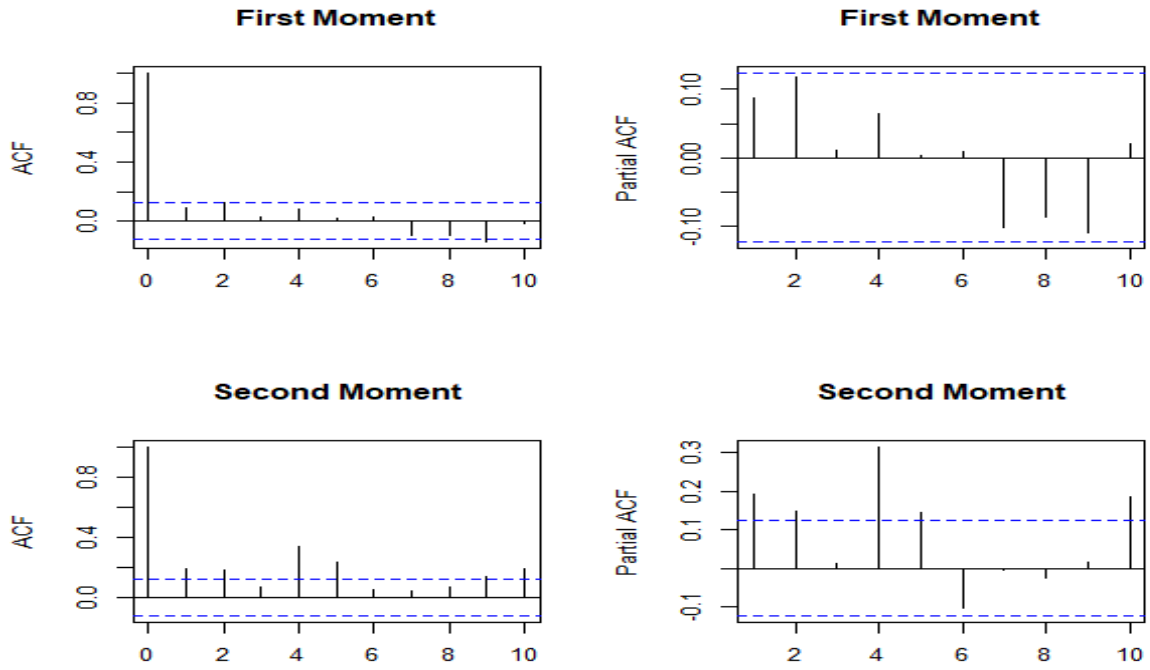

Figure 5: ACF and PACF of January 2016 log return and log return squared

3 Methodology

3.1 EGARCH

3.1.1 EGARCH (1,1) Model

The EGARCH(p,q) model, proposed by Nelson et al. (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. 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, $\epsilon_t \sim SST(0, \sigma_t)$. Berrisch, J., & Ziel, F. (2022) found that the Student-t distribution is useful in capturing the heavy tails 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.

3.1.2 Estimation

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

$$
\begin{aligned}
r_t &= \ln\left(\frac{S_t}{S_{t-1}}\right) = \mu + \epsilon_t, \\
\epsilon_t &= \sigma_t z_t, \\
z_t &\sim SST(0, 1, \nu, \delta), \\
\epsilon_t &\sim SST(0, \sigma_t, \nu, \delta), \\
\ln\left(\sigma_t^2\right) &= \omega + \alpha\left(|z_{t-1}| - \mathbb{E}[|z_{t-1}|]\right) + \gamma z_{t-1} + \beta \ln\left(\sigma_{t-1}^2\right)
\end{aligned}
$$

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

$$
\begin{aligned}
\tilde{\epsilon}_t &= r_t - \mu \\
\tilde{z}_t &= \frac{\tilde{\epsilon}_t}{\tilde{\sigma}_t} \\
\ln\left(\tilde{\sigma}_t^2\right) &= \tilde{\omega} + \tilde{\alpha}\left(|\tilde{z}_{t-1}| - \mathbb{E}[|\tilde{z}_{t-1}|]\right) + \tilde{\gamma}\tilde{z}_{t-1} + \tilde{\beta}\ln\left(\tilde{\sigma}_{t-1}^2\right)
\end{aligned}
$$

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

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

The degree of freedom ν is determined according to Akaike information criteria (AIC), which provide the information lost by a model and is usually used to decide the order of the model. ν that minimize the AIC value will be used throughout the modeling and forecasting in this study. AIC can be calculated by

$$\text{AIC } = 2m - 2\ln(L_n(\varphi))$$

with m is the number of estimated parameters in the model. For my comparative analysis, the order of the model is fixed, EGARCH(1,1), that make EGARCH setting similar to Heston model. Therefore, the number of estimated parameters is fixed and does not play any role in investigating the most suitable degree of freedom ν.

3.2 Heston model

3.2.1 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:

$$\begin{aligned} d\ln(S_t) &= \left(\mu - \frac{1}{2}V_t\right)dt + \sqrt{V_t}dW_t^S \\ dV_t &= \kappa(\theta - V_t)dt + \sqrt{V_t}\sigma_V dW_t^V \\ dW_t^1 dW_t^2 &= \rho \end{aligned}$$

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

$$\begin{aligned} \ln(S_t) - \ln(S_{t-\Delta t}) &= \left(\mu - \frac{1}{2}V_{t-\Delta t}\right)\Delta t + \sqrt{V_{t-\Delta t}}\sqrt{\Delta t}\epsilon_t^S \\ V_t - V_{t-\Delta t} &= \kappa(\theta - V_{t-\Delta t})\Delta t + \sqrt{V_{t-\Delta t}}\sqrt{\Delta t}\epsilon_t^V \\ \epsilon_t^S &\sim \mathcal{N}(0,1) \\ \epsilon_t^V &\sim \mathcal{N}(0,\sigma_V^2) \\ Corr(\epsilon_t^S, \epsilon_t^V) &= \rho \end{aligned}$$

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. 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

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

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

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. I briefly discuss the characteristic of Markov Chain and Monte Carlo in the Appendix. 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.

3.2.2 Estimation

I start with deriving likelihood function of (r_t, V_t). It can be seen from (3.12) and (3.2.2) 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
P(V_t \mid Y, V_{t+1}, V_{t-1}, \kappa, \mu, \theta, \psi, \Omega) &= P(Y, V_{t+1}, V_t, \mid V_{t-1}, \mu, \kappa, \theta, \psi, \Omega)\frac{P(V_{t-1} \mid \kappa, \theta, \psi, \Omega, \mu)}{P(r, V_{t+1}, V_{t-1} \mid \kappa, \theta, \psi, \Omega, \mu)} \\
&\propto P(Y, V_{t+1}, V_t \mid V_{t-1}, \kappa, \theta, \psi, \Omega, \mu) \\
&= \frac{1}{V_t \Delta t}\exp\left(-\frac{1}{2\Omega}\frac{(\Omega + \psi^2)(\frac{1}{2}V_t\Delta t + Y_{t+1} - \mu\Delta t)^2}{V_t\Delta t}\right. \\
&-\frac{1}{2\Omega}\frac{-2\psi(\frac{1}{2}V_t\Delta t + r_{t+1} - mu\Delta t)(-(1-\kappa\Delta t)V_t - \kappa\theta\Delta t + V_{t+1})}{V_t\Delta t} \\
&\left.-\frac{1}{2\Omega}\frac{(-(1-\kappa\Delta t)V_t - \kappa\theta\Delta t + V_{t+1})^2}{V_t\Delta t}\right) \\
&\cdot \exp\left(-\frac{1}{2\Omega}\frac{-2\psi(Y_t - \mu\Delta t + \frac{1}{2}V_{t-1}\Delta t)(V_t - \kappa\theta\Delta t - (1-\kappa\theta)V_{t-1})}{V_{t-1}\Delta t}\right. \\
&\left.-\frac{1}{2\Omega}\frac{(V_t - \kappa\theta\Delta t - (1-\kappa\Delta t)V_{t-1})^2}{V_{t-1}\Delta t}\right).
\end{aligned}
$$

3.2.3 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 space $V_1, V_2, ..., V_n$ are generated based on posterior distributions which are derived from prior distribution assumption and observed data. By construction, the sequence has Markov's properies includin convergence.

According to Michael Johannes and Nicholas Polson (2003), convergence of the sequence relied on the ergodic theory for Markov Chain. A g-step transition probability plays a key role to define a Markov Chain:

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

There are two important conditions for Markov Chain to converge, i.e. irreducible and aperiodic. A Markov chain is irreducible if it has positive probability of eventually entering any set which has $\pi-$positive probability regardless of its initial state. 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 the innital state. This study assume the convergence of estimated parameters after g-steps sampling. While the convergence condition is hard to verified, it is investigated through graphical illustration in this study by tracking the value of sampling each step.

After defining prior distribution and deriving posterior density of parameters set $\mu, \kappa, \theta, \rho$ and state space V_t, MCMC is employed to obtained draws from posterior distribution. MCMC approaches utilized for parameters and state space are different. In order to estimate parameters $\mu, \kappa, \theta, \rho$, Gibbs sampler is useful 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 in 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 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 parameters sampling that can be obtained easily with 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 approriate Joshua Cape et al. (2014). Similarly, the algorithm starts with initializing values for 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:

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

where $t \in 1, 2, ..., n$ and $V_t^{*(g)}$ is new proposal for $V_t^{(g)}$. σ_V^2 is volatility of volatility.

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:

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

and

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

$\mathcal{A}(V_t^{*(g)}, V_t^{(g-1)})$ 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 acceptane/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.

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

and

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

3.3 n-day ahead forecast

3.4 Evaluation Measurement

While mean square error (MSE) or mean absolute error (MAE) is popular when it comes to evaluation of forecast accuracy, probabilistic forecast, which provides forecasts for future and suitable measures of the uncertainty associated with them, requires different criteria to assess the performance of forecasters. Scoring rules play an essential role in both forecast quality measure (Gneiting, T., and Raftery, A. E., 2007). In some fields, researchers refer 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 for several reasons. First, the forecast density is illustrated in terms of samples by monte carlo simulation. Second, CRPS takes the distance of forecasts and actual value into account by giving credit for assigning high probabilities to values close to the real value.

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, 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. Negative orientation version is more popular in practice because it not only provide the computation in the same unit as observation that make 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 $X_1, ..., X_m$.

The intuition of CRPS is clearly illustrated in Figure 6. In real life, only one value x is observed, then the observed probability density function is

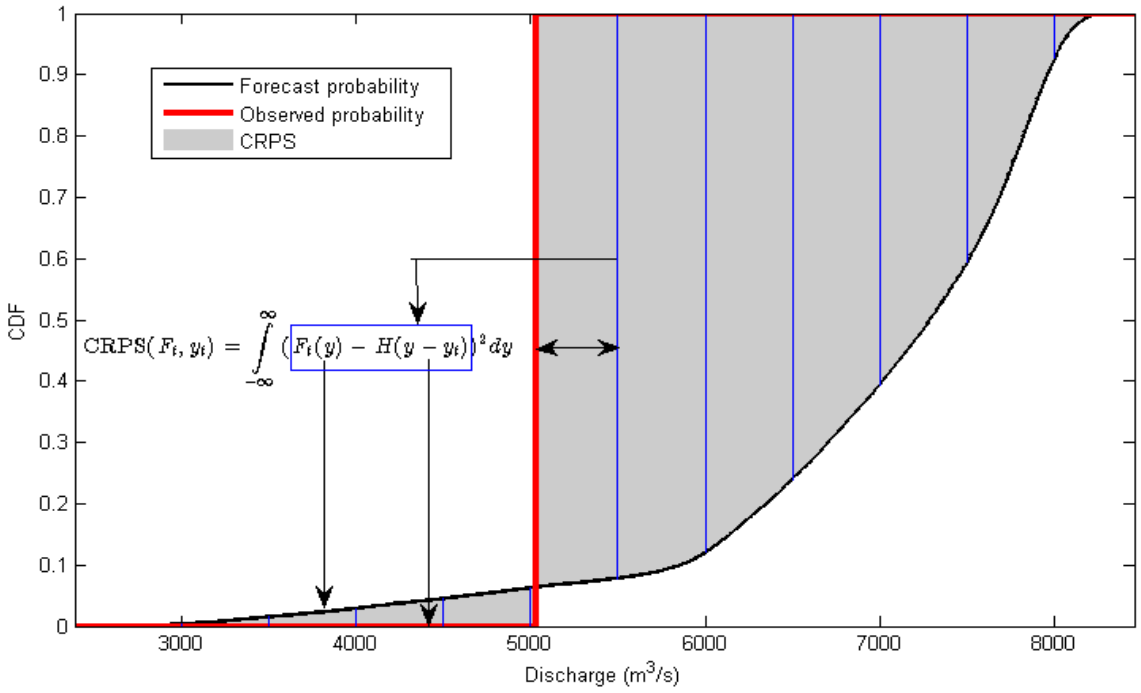

Figure 6: CPRS graphical illustration - Source: Durga Lal Shrestha (MATLAB)

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

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 6 that CRPS represents the gray area between the forecast CDF and observed CDF.The smaller the area, the more accurate the probabilistic forecast.

3.5 pseudo code

4 Empirical Analysis

5 Conclusion

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.

Appendix