Time Series Analysis

1. Introduction

This project is based on Chapter 7, Time Series Models for Financial Data, from Introduction to Time Series and Forecasting (Third Edition) by Brockwell and Davis. It explores time series models that are useful for analyzing financial data. For example, given P_t as the price of a stock or financial asset at time t, $t \in \mathbb{Z}$, the log returns, $\{Z_t \coloneqq log\ P_t - log\ P_{t-1}\}$, is often modelled as a stationary time series, which their statistical properties, such as mean and variance, remain constant over time. Traditional linear time series models, such as ARMA, assume constant conditional variance h_t for Z_t, which is independent of time and past data. However, this assumption is often violated in practice due to the changing variability of log returns Z_t.

Therefore, this project focuses on discrete-time series models designed to capture and forecast volatility. These include ARCH, GARCH, modified GARCH processes and stochastic volatility models, which can better reflect the stylized characteristics of financial time series, such as tail heaviness, asymmetry, volatility clustering and serial dependence without correlation, which the traditional linear time series models fail to capture.

2. ARCH (Autoregressive Conditional Heteroscedasticity) Models

The ARCH(p) model, introduced by Engle (1982), is designed to model changing volatility in time series by capturing the conditional variance, h_t, of a stationary process $\{Z_t\}$. The ARCH(p) process assumes $\{Z_t\}$ as a stationary solution of the equations:

$$Z_t = \sqrt{h_t} e_t,\ \{e_t\} \sim \text{IID N(0,1)}, \tag{2.1}$$

where h_t, the conditional variance, is a (positive) function of past squared values of Z_t:

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2, \text{ with } \alpha_0 > 0 \text{ and } \alpha_i \geq 0, i = 1, \dots, p. \tag{2.2}$$

ARCH(1) Model

The simplest case, ARCH(1), is obtained when $p = 1$. The recursion of (2.1) and (2.2) gives:

$$\begin{aligned} Z_t^2 &= \alpha_0 e_t^2 + \alpha_1 Z_{t-1}^2 e_t^2 \\ &= \alpha_0 e_t^2 + \alpha_1 \alpha_0 e_t^2 e_{t-1}^2 + {\alpha_1}^2 Z_{t-2}^2 e_t^2 e_{t-1}^2 \\ &= \cdots \\ &= \alpha_0 \sum_{j=0}^{n} \alpha_1^j e_t^2 e_{t-1}^2 \cdots e_{t-j}^2 + \alpha_1^{n+1} Z_{t-n-1}^2 e_t^2 e_{t-1}^2 \cdots e_{t-n}^2. \end{aligned}$$

If $\alpha_1 < 1$, and $\{Z_t\}$ is stationary and causal (i.e., Z_t is a function of $\{e_t, s \leq t\}$), then the last term which has expectation of $\alpha^{n+1} E Z_t^2$ converges to 0 as $n \to \infty$, and the first term which is non-decreasing in n converges as $n \to \infty$, and the expected value is bounded above by $\frac{\alpha_0}{1-\alpha_1}$.

Thus, for $\alpha_1 < 1$, ARCH(1) process has the unique causal stationary solution

$$Z_t = e_t \sqrt{\alpha_1 \left(1 + \sum_{j=1}^{\infty} \alpha_1^j e_{t-1}^2 \cdots e_{t-j}^2\right)}$$

and has the below properties:

- $E(Z_t) = E(E(Z_t | e_s, s < t) = 0,\ Var(Z_t) = \frac{\alpha_0}{1-\alpha_1}$

- $E(Z_{t+h}Z_t) = E(E(Z_{t+h}Z_t|e_s, s < t+h)) = 0$ for $h > 0$.

The ARCH(1) process with $\alpha_1 < 1$ is strictly stationary white noise, but is not an iid sequence because $E(Z_t^2|Z_{t-1}) = (\alpha_0 + \alpha_1 Z_{t-1}^2)E(e_t^2|Z_{t-1}) = \alpha_0 + \alpha_1 Z_{t-1}^2$, indicating dependence on past values. This shows that $\{Z_t\}$ is not Gaussian, since strictly stationary Gaussian white noise is necessarily iid. The distribution of Z_t is symmetric, i.e., Z_t and $-Z_t$ have the same distribution. The marginal distribution of Z_t is heavy-tailed, since $E({Z_t}^4)$ is finite if and only if $3\alpha_1^2 < 1$, which in general indicate that for every $\alpha_1$in the interval (0,1), $E(Z^{2k}) = \infty$, for some $k \in \mathbb{Z}^+$. If $E{Z_t}^4 < \infty$, the squared process Y_t= ${Z_t}^2$ has the same autocorrelation function (ACF) as the AR(1) process, defined by W_t=$\alpha_1 W_{t-1} + e_t$.

Conditional Gaussianity

The ARCH(p) process is conditionally Gaussian, i.e., given past observations $\{Z_s, s = t-1, t-2, .., t-p\}$, the current value Z_t follows a Gaussian distribution with known mean and variance. This allows estimation of parameters using conditional maximum likelihood by the likelihood of $Z_{p+1}, \dots, Z_n$ conditional on $\{Z_1, \dots, Z_p\}$, with numerical optimization methods. For example, the conditional likelihood of observations $\{z_2, \dots, z_n\}$ of ARCH(1) process given $Z_1 = z_1$, with zero mean and the conditional variance h_t, is

$$L = \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi(\alpha_0+\alpha_1 z_{t-1}^2)}} exp\left\{-\frac{z_t^2}{2(\alpha_0+\alpha_1 z_{t-1}^2)}\right\}$$

3. GARCH (Generalized ARCH) Models

GARCH(p,q) process, introduced by Bollerslev (1986), builds on the ARCH(p) model by incorporating past conditional variances. The GARCH(p,q) process is a stationary process $\{Z_t\}$ with the generalized form of (2.1):

$$Z_t = \sqrt{h_t} e_t,\ \{e_t\} \sim \text{IID } (0,1). \tag{3.1}$$

with the conditional variance equation in (2.2) replaced by

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{q} \beta_j\, h_{t-j}, \text{ with } \alpha_0 > 0,\ \alpha_i, \beta_j \geq 0. \tag{3.2}$$

For modelling purpose, distribution of e_t is usually assumed to follow either

- Standard normal distribution: $e_t{\sim}N(0,1)$, or (3.3)
- Scaled Student's t-distribution, with v degrees of freedom: $\sqrt{\frac{v}{v-2}}\, e_t{\sim}t_v,\ v > 2$, (3.4)

The scaling factor in (3.4) ensures that $Var(e_t) = 1$. Other distributions for e_t can also be used.

Parameter Estimation

Let $Y_t = a + Z_t$, where $\{Z_t\}$ is the GARCH(p, q) process. Similar to ARCH process, GARCH parameters are estimated by numerically maximizing the likelihood of $\tilde{Z}_{p+1}, \dots, \tilde{Z}_n$ conditional on the known values $\tilde{Z}_1, \dots \tilde{Z}_p$, where $\{\tilde{Z}_t\}$ denotes the mean-corrected observations. The sample mean can be used as an estimate of a in the model, otherwise a will be assumed 0. Initial values for $\tilde{Z}_t$ and h_t are assumed with $\tilde{Z}_t = 0$ and $h_t = \hat{\sigma}^2$ for $t \leq 0$, where $\hat{\sigma}^2$ is the sample variance of $\{\tilde{Z}_1, \dots \tilde{Z}_n\}$.

The conditional likelihood function for model Y_t with $\{Z_t\}$ defined by (3.1)-(3.3) is

$$L(\alpha_0, \dots, \alpha_p, \beta_1, \dots, B_q) = \prod_{t=p+1}^{n} \frac{1}{\sigma_t} \phi\left(\frac{\tilde{Z}_t}{\sigma_t}\right). \tag{3.5}$$

Parameters are estimated by maximizing (2.5) with respect to the coefficients $\alpha_0, \dots, \alpha_p$ and $\beta_1, \dots, B_q$, where ϕ denotes the standard normal density, and standard deviations $\sigma_t = \sqrt{h_t}, t \geq 1$ are computed recursively from (2.2) with Z_t replaced by $\tilde{Z}_t$ and with initial assumed values for $\tilde{Z}_t$ and h_t for $t \leq 0$.

The conditional likelihood function for model Y_t with $\{Z_t\}$ defined by (3.1), (3.2) and (3.4) is

$$L(\alpha_0, \dots, \alpha_p, \beta_1, \dots, B_q) = \prod_{t=p+1}^{n} \frac{\sqrt{v}}{\sigma_t \sqrt{v-2}} t_v \left(\frac{\tilde{Z}_t \sqrt{v}}{\sigma_t \sqrt{v-2}}\right). \tag{3.6}$$

The parameter estimation is similar to the model with Gaussian e_t, where maximization is carried out with respect to the coefficients, but with the conditional likelihood function in (3.6) and an additional parameter for the degrees of freedom v of the t-density t_v. It is often useful to initialize the coefficients of the model by first fitting a GARCH model with Gaussian e_t and then carrying out the optimization using *t*-distributed e_t to improve convergence.

The optimization (regardless of e_t distribution) is constrained with $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$ and $\hat{\alpha}_1 + \cdots + \hat{\alpha}_p + \hat{B}_1 + \cdots + \hat{B}_q$<1. This condition is necessary and sufficient for the GARCH model to have a causal weakly stationary solution, so that the conditional variances h_t do not grow indefinitely.

Model Diagnostics and Comparison

The standardized residuals should be IID, and approximately follow the respective e_t distribution. The sample ACF of absolute values and squares of the residuals can be used to verify independence.

Models with different orders p and q can be compared with the AICC (corrected Akaike Information Criterion), which is defined in terms of the conditional likelihood L:

$$AICC := -2\frac{n}{n-p} lnL + \frac{2(p+q+2)n}{n-p-q-3}, \tag{3.7}$$

Model with Gaussian e_t uses the conditional likelihood L in (3.5), while the model with t-distributed e_t uses the conditional likelihood L in (3.6) and replaces q by $q+1$, as an additional coefficient for the degrees of freedom v is considered. The factor $n/(n-p)$ in first term is adjusted for the reduced number of factors in (3.5) and (3.6).

Application

Both ARCH and GARCH models can capture volatility clustering, a phenomenon where large (small) fluctuations in the data tend to be followed by fluctuations of comparable magnitude. However, incorporating correlation in the sequence $\{h_t\}$ of conditional variances in the GARCH model makes it more flexible and capable of capturing more complex volatility clustering. GARCH models are suitable for time series which has sustained periods of high and low volatility, where the sample autocorrelations of the absolute values and squares of the data are significantly different from zero, which indicates dependence even if sample autocorrelation of raw data is small. Model with conditional t-distribution can accommodate heavy-tailed data, which may further improve model performance when the distribution of Z_t given $\{Z_s, s < t\}$ has a heavier tailed distribution.

4. Modified GARCH Processes

The GARCH model is able to account for marginal distributions with heavy tails, persistent volatility and aggregational Gaussianity. Specifically, aggregational Gaussianity states that the sum of the daily returns, $S_n = \sum_{t=1}^{n} Z_t$, where $Z_t = lnP_t - lnP_{t-1}$, is approximately normally distributed for large n, following martingale central limit theorem. EGARCH and FIGARCH models are introduced to account for asymmetry between negative and positive disturbances, and long-range dependence in volatility,

4.1 EGARCH (Exponential GARCH) Models

Nelson (1991) introduced EGARCH model to capture asymmetry by allowing negative and positive values of $\mathrm{e_t}$ in the GARCH process to affect future volatilities, h_s, $(s > t)$ differently.

EGARCH(1,1)

Consider the process $\{Z_t\}$ defined as (3.1) where $\{l_t \coloneqq ln\, h_t\}$ is the weakly and strictly stationary solution of

$$l_t = c + \alpha_1 g(e_{t-1}) + \gamma_1 l_{t-1}, \text{ where } c \in \mathbb{R},\ \alpha_1 \in \mathbb{R}, |\gamma_1| < 1. \quad (4.1.1)$$

Function $\mathrm{g(e_t)}$ introduces asymmetry, with $\mathrm{e_t}$ having a distribution symmetric about zero:

$$\mathrm{g(e_t)} = \mathrm{e_t} + \lambda(|\mathrm{e_t}| - \mathrm{E}|\mathrm{e_t}|), \quad (4.1.2)$$

Process is defined in terms of l_t to ensure that $\mathrm{h_t}(= \mathrm{e}^{\mathrm{l}t}) > 0$. Alternatively, (4.1.2) can be rewritten as

$$\mathrm{g(e_t)} = \begin{cases} (1+\lambda)\mathrm{e_t} - \lambda \mathrm{E}|\mathrm{e_t}|, & \text{if } \mathrm{e_t} \geq 0 \\ (1-\lambda)\mathrm{e_t} - \lambda \mathrm{E}|\mathrm{e_t}|, & \text{if } \mathrm{e_t} < 0 \end{cases}.$$

The slopes differ for positive and negative $\mathrm{e_t}$, which allows l_t to respond differently to positive and negative shocks $\mathrm{e_{t-1}}$ of the same magnitude. Typically, $\lambda < 0$, implying that large negative shocks have a greater impact on volatility than positive shocks of the same magnitude. There is no asymmetry when $\lambda = 0$.

$g(e_t)\}$ is iid with $Eg(e_t) = 0$ and $Var\big(g(e_t)\big) = 1 + \lambda^2 Var(|e_t|)$. The symmetry of e_t implies that e_t and $|e_t| - E|e_t|$ are uncorrelated.

General EGARCH(p,q)

The EGARCH(p,q) process generalizes $l_t \coloneqq ln\, h_t$ by replacing (4.1.1) by

$$l_t = c + \alpha(B)g(e_t) + \gamma(B)l_t, \text{ where } \alpha(B) = \sum_{i=1}^{p} \alpha_i B^i, \gamma(B) = \sum_{i=1}^{q} \gamma_i B^i. \quad (4.1.3)$$

Strict stationarity and causality of $\{l_t\}, \{h_t\}$ and $\{Z_t\}$ are ensured if 1- $\gamma(z) \neq 0$ for all complex z such that $|z| \leq 1$.

Generalized Error Distribution (GED)

Nelson proposed the use of generalized error distribution (GED) for e_t with density

$$f(x) = \frac{\nu\, exp[(-1/2)|x/\xi|^{\nu}]}{\xi \cdot 2^{1+1/\nu}\Gamma(1,\nu)}, \text{ where } \xi = \left\{\frac{2^{(-2/\nu)}\Gamma(1,\nu)}{\Gamma(3,\nu)}\right\}^{1/2} \text{ and } \nu > 0.$$

The value of ξ ensures that $Var(e_t) = 1$. Parameter ν determines the tail heaviness where distribution becomes normal when $\nu = 2$, while smaller ν values imply heavier tails. Density f is symmetric,

$\frac{1}{2}|e_t/\xi|^v$ has gamma distribution with parameters $\frac{1}{v}$ and 1, and $E|e_t|^k = \frac{\Gamma((k+1)/v)}{\Gamma(1/v)}\left[\frac{\Gamma(1/v)}{\Gamma(3/v)}\right]^{k/2}$.

Inference via conditional maximum likelihood

The conditional likelihood is computed as $L = \prod_{t=1}^{n} \frac{1}{\sqrt{h_t}} f\left(\frac{Z_t}{\sqrt{h_t}}\right)$, and objective is to minimize

$$-2\,ln\,L = \sum_{t=1}^{n} \ln h_t + \sum_{t=1}^{n} \left|\frac{Z_t}{\xi\sqrt{h_t}}\right|^v + 2n\ln\left(\frac{2\xi}{v} \cdot 2^{1/v}\Gamma(1,v)\right)$$

with respect to $c, \lambda, v, \alpha_1, \dots, \alpha_p, \gamma_1, \dots \gamma_q$. Assumed that $h_t = \hat{\sigma}^2$ and $e_t = 0, t \le 0$, then h_t and e_t are computed recursively from the observations $Z_1, Z_2, \dots$, in (3.1) and (4.1.3), with $e_t = Z_t/\sqrt{h_t}$. Since h_t is automatically positive, the only constraints in this optimization are $v > 0$ and $1 - \gamma(z) \neq 0$ for all complex z such that $|z| \le 1$.

4.2 FIGARCH (Fractionally Integrated GARCH)

To incorporate long range dependence in volatility into the GARCH models, Baillie et al. (1996) introduced FIGARCH. Before introducing FIGARCH, a brief overview of fractionally integrated ARMA (ARFIMA) processes and "long memory" is provided.

ARFIMA(Fractionally Integrated ARMA)

The ARFIMA(p, d, q) is a stationary time series with large lags decay slowly in the autocorrelation function. It is defined as the zero-mean stationary solution $\{X_t\}$ of the difference equation:

$$(1-B)^d \phi(B) X_t = \theta(B) Z_t,$$

where $d \in (0,0.5)$, $\{Z_t\} \sim WN(0, \sigma^2)$, B is the backward shift operator, and $\phi(z)$ and $\theta(z)$ are polynomials of degrees p and q respectively. These polynomials, with no common zeroes, satisfy $\phi(z) \neq 0$ and $\theta(z) \neq 0$ for all complex z such that $|z| \le 1$.

Fractional differencing, $(1-B)^r$, *is* defined via the power series expansion:

$$(1-z)^r \coloneqq 1 + \sum_{j=1}^{\infty} \frac{r(r-1)\dots(r-j+1)}{j!}(-z)^j, |z| < 1, r\epsilon\mathbb{R}.$$

The ARFIMA process $\{X_t\}$ has a mean-square convergent MA(∞) representation:

$$X_t = \sum_{j=0}^{\infty} \Psi_j Z_{t-j}, \text{ and } \Psi(z) = (1-z)^d \theta(z)/\ \phi(z), |z| < 1.$$

where Ψ_j is the coefficient of z^j in the power series expansion. Both Ψ_j and $\rho(j)$, autocorrelations of $\{X_t\}$ at lag j, converge to 0 at hyperbolic rates as $j \to \infty$, which is a much slower rate than those in ARMA process, leading to "long memory".

Motivation of FIGARCH(p,d,q)

Consider the GARCH model defined in (3.1), (3.2), with backward shift operator, it follows that

$$(1 - \alpha(B) - \beta(B)) Z_t^2 = \alpha_0 + (1 - \beta(B)) W_t,$$

where $\{W_t = Z_t^2 - h_t\}$ is white noise, $\alpha(B) = \sum_{i=1}^{p} \alpha_i B^i$ and $\beta(B) = \sum_{i=1}^{p} \beta_i B^i$. A causal weakly stationary solution exists for $\{Z_t\}$ if and only if the zeroes of $1 - \alpha(z) - \beta(z)$ have absolute value greater than 1 and there is then exactly one such solution. (Bollerslev 1986)

IGARCH (Integrated GARCH)

Engle and Bollerslev (1986) define the IGARCH(p,q) model satisfying (3.1), (3.2) and

$$\phi(B)(1-B)Z_t^2 = a_0 + (1-\beta(B))W_t,\ W_t = Z_t^2 - h_t,$$

with $1-\alpha(z)-\beta(z) = (1-z)\,\phi(z)$, where $\phi(z)$ is the polynomial with all of its zeroes outside the unit circle, i.e., the polynomial $1-\alpha(z)-\beta(z)$ has a simple zero at $z=1$, and that the other zeroes all fall outside the closed unit disc.

Bougerol and Picard (1992) showed that if the distribution of e_t has unbounded support and no atom at zero, then there is a unique strictly stationary causal solution. However, such solution has infinite variance, i.e., $EZ_t^2 = \infty$. Despite this, GARCH models often found to have $\alpha(1)+\beta(1) \approx 1$ in practice, supporting the practical relevance of IGARCH.

FIGARCH(p,d,q)

Baillie et al. (1996) define the FIGARCH(p,d,q) process as a causal strictly stationary solution of the equations in (3.1) and

$$\phi(B)(1-B)^d Z_t^2 = a_0 + \left(1-\beta(B)\right)W_t,\ 0 < d < 1, \tag{4.2.1}$$

where $W_t = Z_t^2 - h_t$, and the polynomials $\phi(z)$ and $1-\beta(z)$ are non-zero for all complex z such that$|z| \leq 1$. When substituting $W_t = Z_t^2 - h_t$ in (4.2.1), (4.2.1) is equivalent to the equation

$$h_t = \frac{\alpha_0}{1-\beta(1)} + \left[1 - \left(1-\beta(B)\right)^{-1}\phi(B)(1-B)^d\right]Z_t^2.$$

Thus, FIGARCH(p,q) process can be regarder as a special case of the IARCH(∞) process defined by (3.1) and

$$h_t = a_0 + \sum_{j=1}^{\infty} a_j Z_{t-j}^2 \text{ with } a_0 > 0 \text{ } and \sum_{j=1}^{\infty} a_j = 1.$$

The existence and uniqueness of strictly stationary solutions for IARCH(∞) (including FIGARCH) remain unresolved. Any strictly stationary solution must have infinite variance since

$$\sigma^2 := EZ_t^2 = Eh_t < \infty, \text{ but } \sum_{j=1}^{\infty} a_j = 1 \Rightarrow \sigma^2 = a_0 + \sigma^2,$$

which is a contradiction. However, Douc et al. (2008) have provided the sufficient conditions for the existence of causal strictly stationary solution of the IARCH(∞), particularly, FIGARCH.

Fractionally integrated EGARCH models have also been introduced (Bollerslev and Mikkelsen 1996) to account for both long memory and asymmetric effects of positive and negative shocks, e_t.

5. Stochastic Volatility Models

The general stochastic volatility (SV) model for $\{Z_t\}$ is defined similarly as (3.1), with $\{h_t\}$, volatility at time t, being a strictly stationary sequence of non-negative random variables. Unlike GARCH models where h_t depends on $e_s, s < t$ and which the processes $\{Z_t\}$ and $\{h_t\}$ are inextricably linked, $\{e_t\}$ and $\{h_t\}$ are independent in SV models, allowing more flexible modelling of the volatility process with any non-negative strictly stationary sequence. Inference for GARCH models is relatively easy using conditional likelihood based on observations of $Z_1, \ldots, Z_n$, but inference for an SV model based on observations $\{Z_t\}$ is more difficult since the process is driven by two independent random sequences rather than one and only $\{Z_t\}$ is observed, with $\{h_t\}$ being latent.

Lognormal SV

Lognormal SV is a widely used special case of SV model, due to Taylor (1982, 1986), defined similarly to (2.1) where $h_t = e^{l_t}$, $\{l_t\}$ is a (strictly and weakly) stationary solution of the equations

$$l_t = \gamma_0 + \gamma_1 l_{t-1} + \eta_t,\ \{\eta_t\} \sim \text{IID N}(0, \sigma^2),\ |\gamma_1| < 1 \tag{5.1}$$

and the sequences $\{\eta_t\}$ and $\{e_t\}$ are independent. The sequence $\{l_t\}$ is a Gaussian AR(1) process with mean $u_l = El_t = \frac{\gamma_0}{1-\gamma_1}$ and variance $v_l = Var(\,l_t) = \frac{\sigma^2}{1-\gamma_1^2}$.

Properties of {Zt}:

(i) $\{Z_t\}$ is strictly stationary

(ii) Moments: $EZ_t^r = E(e_t^r)E\,exp\left(\frac{rl_t}{2}\right)$

$$= \begin{cases} 0, & if\ r\ is\ odd \\ \left[\prod_{i=1}^{m}(2i-1)\right]\exp\left(\dfrac{m\gamma_0}{1-\gamma_1} \quad \dfrac{m^2\sigma^2}{+2(1-\gamma_1^2)}\right), & if\ r = 2m \end{cases}$$

(iii) Kurtosis: Marginal distribution of Z_t exhibit heavier tails than a normal distribution, since $\frac{Ez_t^4}{(EZ_t^2)^2} = 3\exp\left(\frac{\sigma^2}{1-\gamma_1^2}\right) \geq 3.$

(iv) Autocovariance function of $\{Z_t^2\}$: For t > s,

$$E(Z_t^2 Z_s^2 | e_u, n_u, u < t) = h_s h_t e_s^2 E(e_t^2 | e_u, n_u, u < t) = h_s h_t e_s^2\ ,$$

$$\Rightarrow\quad E(Z_t^2 Z_s^2) = E\exp(l_t +\ l_s)$$

$$\Rightarrow\quad Cov(Z_{t+h}^2, Z_t^2) = E\exp(l_{t+h} +\ l_t) - E\exp(l_{t+h})\,E\exp(\,l_t)$$

$$= \exp[2\mu_l + v_l(1+\gamma_1^2)] - \exp[2\mu_l + v_l] \text{ for } h > 0$$

Above is derived using the fact that $\{h_t\}$ and $\{e_t\}$ are independent, $h_s, h_{t,}$ and e_s^2 are each functions of $\{e_u, n_u, u < t\}$ which is independent of e_t^2, the relation $h_t = e^{l_t}$, the fact that l_{t+h} is normally distributed and $E\exp(\,X) = \exp[u + v/2]$ where $X \sim N(\mu, v)$.

$Var(Z_t^2) = 3\exp[2\mu_l + 2v_l] - \exp[2\mu_l + v_l]$. Hence, the autocorrelation function of $\{Z_t^2\}$ is:

$$\rho_{Z_t^2}(h) = \frac{Cov(Z_{t+h}^2, Z_t^2)}{Var(Z_t^2)} = \frac{\exp(v_l\gamma_1^h)-1}{3\exp(v_l)-1} \sim \frac{v_l}{3\exp(v_l)-1}\gamma_1^h \text{ as } \gamma_1 \to 0\ , \text{for } h > 0$$

which the approximation is similar to ARMA(1,1) process, akin to the squared GARCH(1,1) process.

(v) The process $\{\ln Z_t^2\}$: $lnZ_t^2 = l_l + lne_t^2$ has the autocovariance function of an ARMA(1,1) process with autocorrelation function

$$\rho_{lnZ_t^2}(h) = \frac{v_l\gamma_1^{|h|}}{v_l + 4.93}, \text{h} \neq 0$$

Since, if $e_t \sim N(0,1)$ then $E\ lne_t^2 = -1.27$ and $Var(lne_t^2) = 4.93$, then $Var(lnZ_t^2) = v_l + 4.93$ and $Cov(Z_{t+h}^2, Z_t^2) = v_l\gamma_1^{|h|}$ for h $\neq$ 0.

Parameter Estimation and Forecasting

Parameters σ^2, γ_0 and γ_1 in the defining equations (2.1) and (5.1) can be estimated by maximizing the Gaussian likelihood, applied to $\{Y_t \coloneqq lnZ_t^2 - E\ Z_t^2\}$ which satisfies the ARMA(1,1) equations,

$$Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}, \{Z_t\} \sim WN(0, \sigma_Z^2),$$

for some coefficients ϕ and θ in the interval (-1,1) and white-noise variance σ_Z^2. It is necessary and sufficient to constraint that $\phi + \theta > 0$, to ensure model validity. Maximizing Gaussian likelihood enables estimation of $\hat{\phi}$ and $\hat{\theta}$, which can then be used to derive for γ_1, v_l, σ^2 and γ_0 utilizing the properties of $\{Z_t\}$ and ARMA(1,1) structure. If the estimators $\hat{\phi}$ and $\hat{\theta}$ satisfy $\hat{\phi} + \hat{\theta} \leq 0$ then it suggests that the lognormal SV model is not appropriate in this case.

The minimum mean-squared error (MSE) predictor of l_{t+h} conditional on $\{l_s, s \leq t\}$ is found from (5.1) to be

$$P_t l_{t+h} = \gamma_1^h l_t + \gamma_0 \frac{1-\gamma_1^h}{1-\gamma_1}, \text{ with MSE } E(l_{t+h} + P_t l_{t+h})^2 = \sigma^2 \frac{1-\gamma_1^{2h}}{1-\gamma_1{}^2}.$$

Since l_t is not observed, Kalman recursions can be used to forecast l_{t+h} using past observations.

6. Implementation

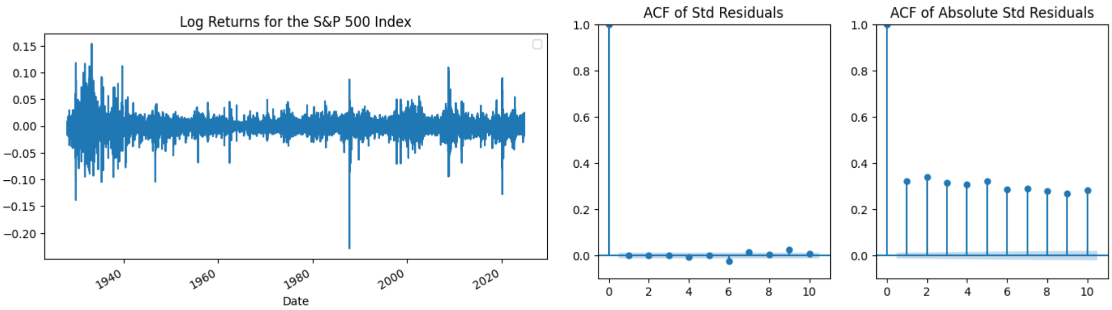

Fig.1

We will analyze the time series data of the S&P 500 Index, focusing on its adjusted closing prices from December 30th 1927 to November 15th 2024, used to compute the daily log returns. Fig. 1 (left) shows the log returns, which has sustained periods of high and low volatility. Notably, high volatility is observed before 1940, around 1995, 2010, and 2020. An ARMA(2,2) model, identified as the best ARIMA model, is fitted to the data and the residuals are evaluated.

Fig.1 (right) shows the autocorrelation of the standardized residuals and the absolute standardized residuals. While the autocorrelation of standardized residuals is small, the autocorrelation of absolute standardized residuals is significantly different from zero, suggesting the presence of volatility clustering in the data and potential suitability of GARCH model.

GARCH(1,1) models are fitted on the daily log returns, with both normal residuals and t-distributed residuals. The AICC values for the previously fitted ARMA(2,2) model, the GARCH(1,1) model with normal residuals, and the GARCH(1,1) model with t-distributed residuals are -146382, -159288 and -161049 respectively. Both GARCH models outperform the ARMA model, with the GARCH(1,1) model with t-distributed residuals yielded the lowest AICC, indicating the best fit.

Fig. 2 (left) shows the conditional volatility from GARCH(1,1) model with t-distributed residuals, reflecting the changing volatility patterns seen in Fig. 1 (left). Fig. 2 (right) shows that both the autocorrelation of the standardized residuals and the absolute standardized residuals from the GARCH model are low, indicating that the GARCH model has effectively captured the underlying data's structure.

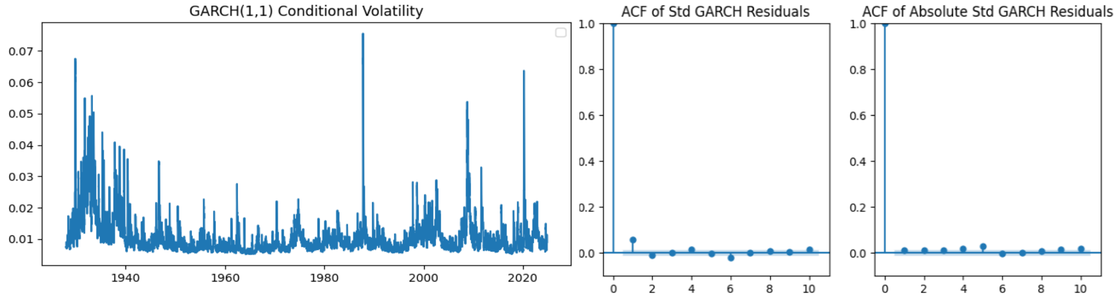

Fig. 2

The GARCH(1,1) model with t-distributed residuals reveals a mean close to zero, with a highly significant constant mean. The omega parameter is small but significant. The alpha (0.1) and beta (0.88) parameters suggest that 10% of current volatility is explained by past squared residuals, and 88% is explained by past volatility.

In summary, the GARCH(1,1) model with t-distributed residuals offers the best fit for the S&P 500 daily log returns, outperforming both the ARMA model and the GARCH model with normal residuals. The use of t-distributed residuals accounts for the heavy-tailed nature of the data (with kurtosis of 21.75), and the model successfully captures volatility clustering making it a robust model for modelling financial time series volatility.

7. Conclusion

This project has explored various models to capture volatility, focusing on the foundational concepts of ARCH and GARCH, as well as their extensions and the Stochastic Volatility (SV) model. The GARCH model was applied to the S&P 500 data, demonstrating its effectiveness in modelling volatility and highlighting the GARCH model's ability to provide a more accurate representation of market dynamics.

References

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 74, 3–30.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307–327.

Bollerslev, T.,&Mikkelsen, H.O. (1996).Modeling and pricing long memory in stock market volatility. Journal of Econometrics, 73, 151–184.

Bougerol,P., & Picard, N. (1992), Strict Stationarity of Generalized Autoregressive Processes. The Annals of Probability, 20(4), 1714 - 1730

Brockwell, P. J., & Davis, R. A. (2016). Introduction to Time Series and Forecasting. Springer Cham, 3, 195-212

Douc, R., Roueff, F., & Soulier, P. (2008). On the existence of some ARCH(∞)processes. Stochastic Processes and Their Applications, 118, 755–761.

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

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. Economic Review, 5, 1–50.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59, 347–370.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961–75. Time Series Analysis: Theory and Practice, 1, 203–226.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961–75. Time Series Analysis: Theory and Practice, 1, 203–226.

Taylor, S. J (1986). Modelling financial time series. New York: Wiley.