Financial Time Series & Volatility Modeling

T. V. Ramanathan

Department of Statistics
Savitribai Phule Pune University
Pune - 411007 (India).

ram@stats.unipune.ac.in

National Workshop on Financial Data Analytics

C R Rao AIMSCS, Hyderabad

December 27-30, 2014

Outline

1. Introduction to financial time series
2. Financial asset returns & some of their properties
3. Linear time series modeling of asset returns (Box-Jenkin's ARIMA modeling)
4. Volatility modeling: Conditional heteroskedastic models (ARCH & GARCH)
5. Asymmetric volatility models & generalizations of GARCH models
6. Gaussian assumption of asset returns
7. Need for non stationary modeling
8. Volatility modelling: Further extensions
9. Time varying ARCH/GARCH

1. Introduction to Financial Time Series

- Concerned with theory and practice of asset valuation over time.
- Financial theory & empirical time series consist of element of uncertainty - Statistical methods plays a crucial role.
- Stock prices, stock indices, exchange rates, interest rates, commodity prices, bond yields, options, futures, gold prices etc.
- Objectives:
  - To understand (to capture main characteristics) how prices behave.
  - To use the knowledge of price behaviour, in order to take better decisions.

Financial Asset Returns:

- One-period simple return:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}, \quad P_t \text{ is the price of an asset at time t}$$

- k-period simple return:

$$R_t[k] = \frac{P_t - P_{t-k}}{P_{t-k}}, \text{ k-period net return}$$

$$1 + R_t[k] = \prod_{i=0}^{k-1}(1 + R_{t-i}), \text{ k-period gross return}$$

- Continuously compounded return: Natural logarithm of simple gross return

$$r_t = ln\frac{P_t}{P_{t-1}} = p_t - p_{t-1} \qquad \text{where} \qquad p_t = ln(P_t)$$

Therefore, $r_t = ln(1 + R_t)$;

Financial Asset Returns:

- k-period continuously compound return:

$$r_t[k] = r_t + r_{t-1} + ... + r_{t-(k-1)}$$

- Portfolio return: $\sum_{i=1}^{N} \omega_{it} \ R_{it}$

- Dividend payment (periodically):

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1, \quad r_t = ln(P_t + D_t) - ln(P_{t-1})$$

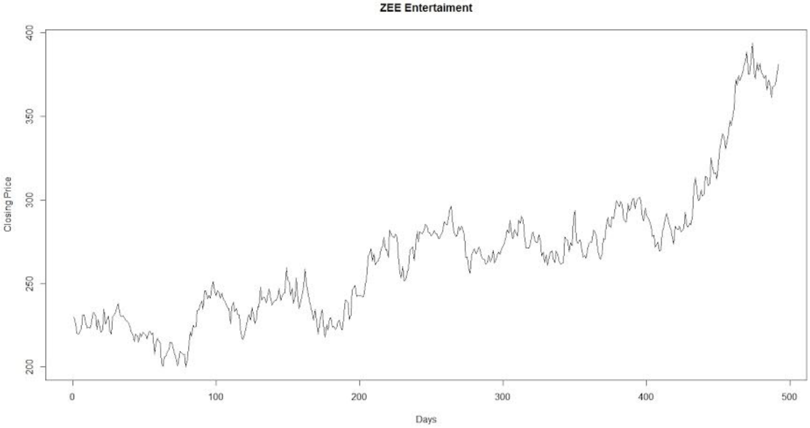
  D_t is the dividend payment between the dates $t-1$ and t.

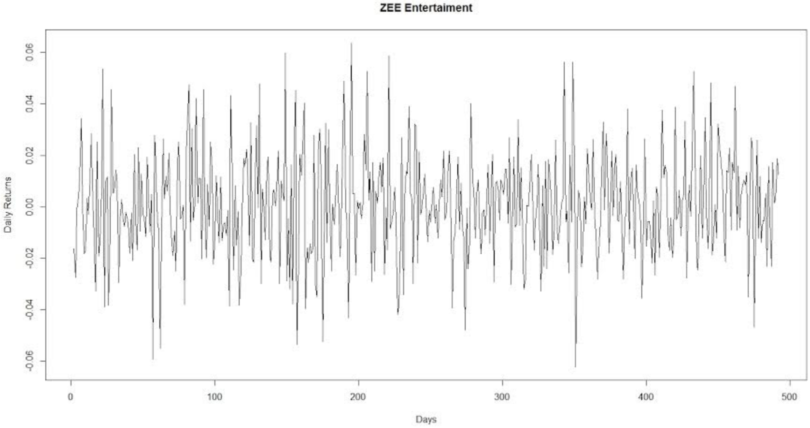
- Excess return: $Z_t = R_t - R_{0t}$, $\qquad z_t = r_t - r_{0t}$
  After subtracting some benchmark asset return (TB or GS)

Why returns instead of prices?

- For average investors, return of an asset is a complete and scale-free summary of the investment opportunity.
- Return series are easier to handle than price series because the former have more attractive statistical properties.

Following plots on daily closing price of Zee Entertainment Television stocks from 24 Dec 2012 to 19th Dec 2014 (492 time points) traded at NSE, clearly illustrate the difference between prices and returns.


ZEE Entertaiment
Closing Price
400
350
300
250
200
0
100
200
300
400
500
Days


ZEE Entertaiment
Daily Returns
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
0
100
200
300
400
500
Days

Long financial position: Owning the asset.

Short position: Selling an asset one does not own. This is accomplished by borrowing the asset from an investor who has purchased it.

Distribution of returns:

- Conditional/Marginal distributions
- Distributional assumptions: Normal, Log normal, Stable, Mixture of normal, **Generalized hyperbolic, NIG** etc.

- Exploratory analysis - Elimination of trend and Seasonality - (Trend fitting, filtering MA, exponential smoothing, Holt-Winter, differencing, seasonal differencing etc.)
- Stationarity
  1. Strict - $(r_{t_1}, r_{t_2}, ..., r_{t_n}) \stackrel{d}{=} (r_{t_1+h}, r_{t_2+h}, ..., r_{t_n+h})$ for any $(t_1, t_2, ..., t_n)$ and h
  2. Weak - $E(r_t)$ free from t, $Var(r_t) < \infty$, free from t and $Cov(r_t, r_s)$ is a function of $t - s$
- Autocorrelation function (ACF) - $\rho_r(s) = Corr(r_t, r_{t+s})$ lag-s autocorrelation - $\frac{\gamma_s}{\gamma_0}$
- Partial Autocorrelation function (PACF)-

$$\alpha_{kk} = Corr(r_k - \hat{r}_k, r_0 - \hat{r}_0)$$

- Correlation between r_k and r_0, after removing the linear effect of $(r_1, ..., r_{k-1})$ - lag-k PACF

- $AR(p)$ models -

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + ... + \phi_p r_{t-p} + \epsilon_t \quad \text{or} \quad \phi(B) r_t = \epsilon_t$$

- $MA(q)$ model -

$$r_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} \quad \text{or} \quad r_t = \theta(B) \epsilon_t$$

- $ARIMA(p, d, q)$ model -

$$\phi(B)(1 - B)^d r_t = \theta(B) \epsilon_t$$

- $ARIMA(p, d, q) \times (P, D, Q)_s$ seasonal ARIMA model -

$$\Phi_P(B^s)\phi(B)(1 - B^s)^D (1 - B)^d r_t = \Theta_Q(B^s)\theta(b)\epsilon_t$$

- Information Criteria - AIC, AICC, BIC, HQC, FIC, DIC etc.

$$AIC = -\frac{2}{T}\, ln(likelihood) + \frac{2}{T} \times (number\ of\ parameters),$$

- Residual analysis : Portmanteau tests, Plots of ACF, PACF, Normality tests, P-P, Q-Q plots etc.

3. Linear Time Series Modeling of Asset Returns

- Box-Jenkin's **stationary** ARIMA modeling and forecasting of conditional mean μ_t, where $\mu_t = E\left(r_t|\mathcal{F}_{t-1}\right)$, in the model $r_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2)$
- Order selection (ACF/PACF, Information criteria - XIC)
- Estimation: Likelihood (Under Gaussian White noise), conditional least squares,
- Forecasting (recursive) - The linear equation structure help to get easy forecasts
- Unit-root testing (ADF tests) and non-stationarity
- Brockwell & Davis (2003), Shumway & Stoffer (2010), Tsay (2010)

- Long memory modeling: ACF decreases to zero at a hyperbolic rate: ARFIMA models (fractionally differenced) Hosking (1981).

$$\phi(B)(1-B)^d r_t = \theta(B)\epsilon_t$$

- Using binomial theorem for non-integer powers
- Consider the case $\phi(B)=1,\ \theta(B)=1$. Then, for for large k, and $-1/2 < d < 1/2$, the $k-th$ order ACF

$$\rho_k = \frac{d(1+d)...(k-1+d)}{(1-d)(2-d)...(k-d)} \approx \frac{(-d)!}{(d-1)!}k^{2d-1}$$

when $\rho_1 = d/(1-d)$.

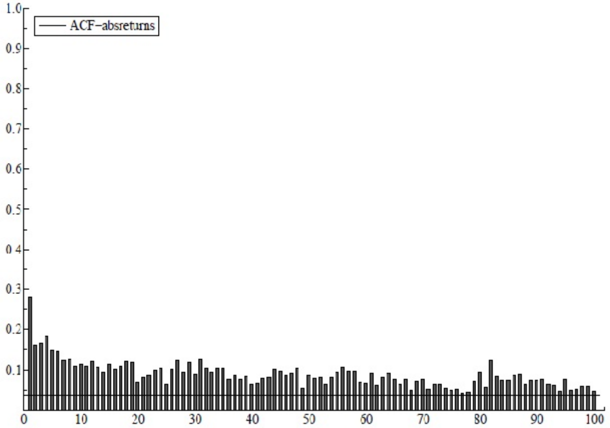


Figure 2: Autocorrelation function (ACF) of absolute returns

4. Volatility Modeling - ARCH/GARCH

- Option pricing, risk management (VaR, ES etc.), Portfolio management (mean-variance), In time series models for better forecasts
- Not directly observable, Implied volatility of a stock (obtained from observed prices of an option and then use B-S formula to get volatility)
- Characteristics: Volatility clustering, evolves over time in a continuous manner, does not diverge to infinity, often stationary, leverage effect
- $\sigma_t^2 = E\left((r_t - \mu_t)^2 \,|\mathcal{F}_{t-1}\right)$ - conditional variance
- ACF/PACF of absolute and square return shows time dependence - ARCH effect - need the modeling of σ_t^2
- Deterministic & stochastic

Model Building Strategy:

- Model the conditional mean
- Use the residuals of the mean equation to test for ARCH effects (graphically as well as analytically).
- Specify a volatility model if ARCH effects are statistically significant, and perform a joint estimation of the mean and volatility equations.
- Check the fitted model carefully and refine it if necessary.
- Volatility forecasting

ARCH models:

- Engle (1982): $\epsilon_t = \sigma_t Z_t, \ \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + ... + \alpha_m \epsilon_{t-m}^2$
- $Z_t \sim iidN(0,1)$, Normal, Student's t etc. are general choices
- Restrictions on parameters to ensure finiteness and non-negativity of the conditional variance.
- $E(\epsilon_t) = 0, \quad Var(\epsilon_t) = \alpha_0/(1 - \alpha_1 - ... - \alpha_m)$
- For $m = 1$, we can show that

$$Kurtosis = \frac{E(\epsilon_t^4)}{[Var(\epsilon_t)]^2} > 3,$$

  which implies that the tail of the distribution is heavier than normal
- Fitting (order determination, estimation, volatility-forecasting)

GARCH Models:

- Bollerslev (1986):

$$\epsilon_t = \sigma_t Z_t, \ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2$$

- Restrictions on parameters to ensure finiteness and non-negativity of the conditional variance.
- ARMA-type representation for ϵ_t^2

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{p} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + \eta_t - \sum_{j=1}^{r} \beta_j \eta_{t-j}$$

$$\eta_t = \epsilon_t^2 - \sigma_t^2, \quad p = max(m, r)$$

- $\{\eta_t\}$ is a martingale difference series (not iid like ARMA)
- Similar properties can be established
- Fitting (order determination, estimation, volatility-forecasting)

IGARCH Models:

- Unit root GARCH models
- IGARCH(1,1): $\epsilon_t = \sigma_t Z_t, \; \sigma_t^2 = \alpha_0 + \beta_1 \epsilon_{t-1}^2 + (1 - \beta_1)\sigma_{t-1}^2$
- Impact of past squared shocks on ϵ_t^2 is persistent.
- IGARCH phenomenon might be caused by occasional level shifts in volatility. The actual cause of persistence in volatility deserves a careful investigation.
- Nelson (1990) for more properties of IGARCH

Asymmetric volatility model: EGARCH

- Asymmetric behavior of market with respect to news Black (1976), French et. al. (1987, JFE), responsible for several asymmetric volatility models

- EGARCH (Nelson (1991)): Allow for asymmetric effects between positive and negative asset returns

$$\epsilon_t = \sigma_t z_t, \qquad (1 - \alpha_1 B)\ln(\sigma_t^2) = (1 - \alpha_1)\alpha_0 + g(\epsilon_{t-1})$$

$$g(\epsilon_t) = \theta\epsilon_t + \gamma\left[|\epsilon_t| - E(|\epsilon_t|)\right]$$

  θ and γ are real constants

- Use of $g(\epsilon_t)$ enables the model to respond asymmetrically to positive and negative lagged values of ϵ_t

- EGARCH uses logged conditional variance to relax the positiveness constraint of model coefficients.

5. Asymmetric volatility models & Generalizations of GARCH

- Glosten, Jagannathan and Runkle (1993, JF) or Zakoian (1994, JEDC) GJR/TGARCH (Threshold-GARCH):

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \delta_G I(\epsilon_{t-1} < 0)\epsilon_{t-1}^2$$

- APARCH/APGARCH Ding et. al. (1993, JEF)
- Engel and Ng (1993, JF) AGARCH, NAGARCH, VGARCH and Sentana (1995, RES) QGARCH.
- EGARCH and other asymmetric volatility models allow the good news and bad news to have different impact on volatility - GARCH does not - News Impact Curves (Engel and Ng, 1993, JF)
- Latest Reviews available on these models: Rohan (2010), Engel (2011 JFE), Rodriguez & Ruiz (2012, JFE, 637-668)

6. Gaussian Assumption of Asset Returns

- Classical portfolio theory - Mean-variance approach
- Value-at-Risk - Risk metrics approach
- Black-Scholes-Merton (BSM) option pricing formula - Geometric Brownian motion (GBM)
- All heavily depends on the assumption that asset return follows a Gaussian distribution
- Empirical evidence: $\rightarrow$ Excess kurtosis, Fat (thick) tails, Asymmetry (more extreme negative returns than positive returns), Crashes and Booms of financial markets (very frequent)

Why not Gaussian in BSM model?

- Market crashes appear more often than one would expect from a normal distribution.
- Brownian motion - continuous price paths : Jumps and discontinuities are features of real markets.
- Days with no change or little change happen more often than the normal.
- Return volatility varies stochastically over time. Also returns and volatility are serially correlated (negatively for equities).
- The drift parameter μ need not be constant over time.
- Real markets have mean reverting behaviour - inconsistent with independent increments assumption of BM
- Momentum effects are common - contradicting independent increment assumption

What is the way out?

- Mixture of normal modeling
  1. Continuous mixture (Student-t (inverted gamma), variance-gamma (gamma), hyperbolic (inverse Gaussian))
  2. Discrete mixture (Infinite Poisson mixture, arbitrary mixture)

- Infinite variance distributions - stable modeling - Nolan, Rachev etc. (Recent workshop in Pune - Ashis Sengupta - How circular distributions are useful in the estimation of parameters of infinite variance distributions?)

What is needed?

- Precise form of the tail of financial returns' distribution is difficult to determine (Cont (2001))
- In order for a parametric distributional model to reproduce the properties of the empirical distribution it must have at least four parameters: a location parameter, a scale parameter, a parameter describing the decay of the tails and an asymmetry parameter.
- Therefore, it is important to develop theoretical models based on other distribution classes, explaining asymmetry and heavy tail phenomena.

Class of Generalized Hyperbolic Distributions:

- Class includes standard hyperbolic distribution, normal-inverse Gaussian (NIG) distribution, scaled t-distribution, variance-gamma (VG) distribution.
- Generalized hyperbolic distribution

$$f(x;\lambda,\alpha,\beta,\delta,\mu)=\frac{(\sqrt{\alpha^2-\beta^2}/\delta)^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})}e^{\beta(x-\mu)} \times\frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\left(\sqrt{\delta^2+(x-\mu)^2}/\alpha\right)^{1/2-\lambda}}$$

$(\lambda,\alpha,\beta,\delta,\mu)\in\mathbf{R}^5$ with $\delta>0$ and $\alpha>|\beta|>0$,
K_{λ} is the modified Bessel function of the third kind with index λ,

Class of GH distributions:

- μ - location, δ - scale, α - shape, β - skewness, λ - kurtosis and charecterizes the classification.
- For $\lambda = 1$ - Standard hyperbolic distribution whose log-density is a hyperbola (parabola for Gaussian)
- For $\lambda = -1/2$ - NIG distribution; the only subclass of the GH distribution that is closed to convolutions (an appealing property for modeling financial returns) Bandorff-Nielsen (1998), Eberlein et al (1998)

$$X_1 \sim NIG(\alpha, \beta, \delta_1, \mu_1), X_2 \sim NIG(\alpha, \beta, \delta_2, \mu_2)$$

$$X_1 + X_2 \sim NIG(\alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2)$$

Class of GH distributions:

- GH distribution is normal variance-mean mixture, where mixing distributin is $GIG(\lambda, \delta, \gamma)$ having density

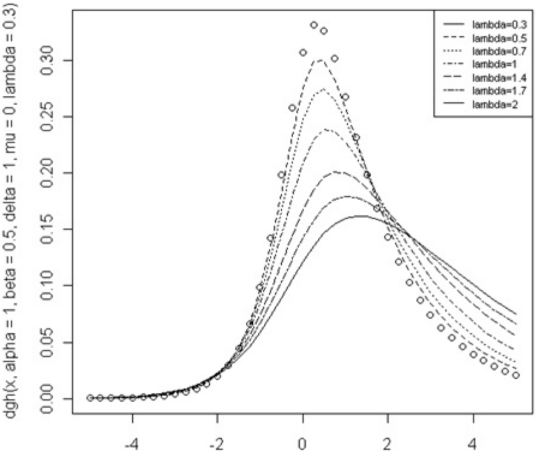
$$f(z; \lambda, \delta, \gamma) = \left(\frac{\gamma}{\delta}\right)^{\lambda} \frac{z^{\lambda-1}}{2K_{\lambda}(\gamma\delta)} exp\left\{-\frac{1}{2}(\delta^2 z^{-1} + \gamma^2 z)\right\}.$$

- $$X|\sigma^2 \ \sim \ N(\mu + \beta\sigma^2, \sigma^2), \ \sigma^2 \ \sim \ GIG(\lambda, \delta, \gamma)$$
$$X \ \sim \ GH\left(\lambda, \sqrt{\beta^2 + \gamma^2}, \beta, \delta, \mu\right)$$

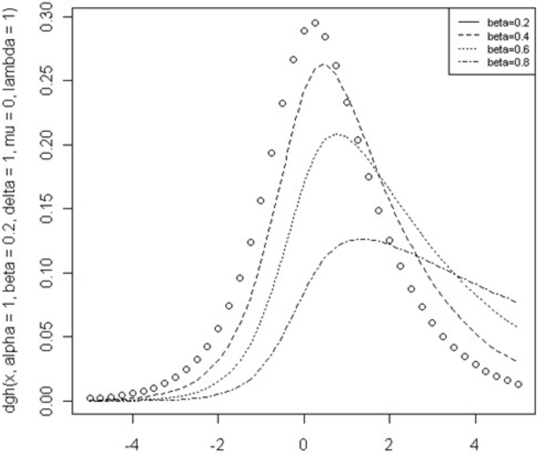
- The tail behaviour of the GH density is 'semi-heavy' (lighter than non-Gaussian stable, heavier than Gaussian)
- All moments exist.

Class of GH distributions:

Plot of density function of GHD for different values of λ (kurtosis)



Class of GH distributions: Plot of density function of GHD for different values of β (skewness)



Variance-Gamma Distribution:

- Introduced by Madan and Seneta (1998): Normal variance-mean mixture, where mixing density is gamma; sub class of GH distributions
- Tails decreases more slower than normal
- Two other non-Gaussian models that are used in financial modeling are
  1. Generalized error distribution:
  2. Generalized lambda distribution: 4 parameters; No closed form pdf, cdf; defined by the inverse distribution Ramberg and Schmeiser (1974)

GH skew Student's t-distribution:

- GH skew Student's t-distribution is a limiting case of GH distribution
- Consider $\lambda = -\nu/2,\ \alpha \to |\beta|$. The density can be obtained from that of GH density.
- Non-central (scaled) Student's t-distribution with ν degrees of freedom.
- Expression for mean, variance, skeweness, kurtosis etc. can be derived
- This distribution has got one heavy and one semi-heavy tail

Skew t-distribution:

- There are sevral definitions in the literature.
- Fernandez and Steel (1998)

$$f(x) = \frac{2\beta}{1+\beta^2}\left[t_\nu(\beta x)I(x<0) + t_\nu\left(\frac{x}{\beta}\right)I(x \geq 0)\right]$$

- Other two versions by Jones and Faddy (2003) and Azzalini and Capitanio (2003)
- All three versions have two tails behaving as polynomials
- May fit heavy tailed data well, but may not capture substantial skewness

Stable Distributions

- Possible reasons for the failure of the CLT in financial markets:
  1. infinite variance-distributions of the variables
  2. non-identical distributions of the variables
  3. dependences between variables
  4. any combination of the three
- Limiting distributions of the sum of such variables is stable, Nolan (2010).
- Samorodnitsky et al. (1994), Weron (1996, 2001), Rohit Deo (....), Rachev (....), Ashis SenGupta (....).

Stable Distributions:

- 4 parameters:
  1. $\alpha \in (0, 2]$ - index of stability - determine the rate at which tails of the distribution taper off.
  2. $\beta \in [-1, 1]$ - defines asymmetry
  3. μ - location, σ - scale
- No closed form for pdf, cdf; described by charecteristic functions, leptokutic, can accommodate fat tails and asymmetry
- $\alpha = 2, \beta = 0$ - Gaussian, $\alpha = 1, \beta = 0$ - Cauchy, $\alpha = 0.5, \beta = 1$- Levy

Goodness-of-fit exercise

- We investigate the fit of these distributions (including normal) for the Indian stock returns. This exercise was carried out on all the 30 stocks listed on BSE 30 index , BSE 30 INDEX and the NSE indices. The goodness-of-fit was carried out by Anderson-Darling test as well as some graphical procedures.
- Normal distribution did not fit in all the 30 cases for the data on daily returns during period from 01/01/2006 to 31/12/2010.

Goodness-of-fit exercise:

- Generalized hyperbolic distribution provided better fit than the normal and the other distributions considered in most of the cases.
- In some cases, the Normal inverse Gaussian distribution and hyperbolic distribution turn out to give a better fit. But , these are particular cases of Generalized Hyperbolic distribution. Hence, the Generalized Hyperbolic distribution can be used to model the data on returns.

Goodness-of-fit exercise: BSE 30 index

Distribution	AD	P − value
GHD	0.1817	0.9946
HD	0.2980	0.9396
NIG	0.3914	0.8572
Normal	16.9829	$4.84E-07$
Stable	1.01487	0.3496
GH t	0.8646	0.4367
$GLD(reg.4)$	4.1888	.0070
VG	10.9769	$1.22E-06$
Skew t	0.7983	0.4822
G error	21495.42	$4.84E-07$

Structural Breaks in volatility:

- Any significant impact of news can change the structure of the model
- All the above models fail to account this change
- Lamoureux and Lastrapes (1990, JBES) - Persistence in variance may be overstated if existence of structural breaks is unaccounted Asymmetry imposed into the data set could be a consequence of these structural breaks.
- Hwang and Pereira (2009) - persistence of conditional volatility in large samples could be exaggerated by the existence of structural breaks in the ARCH and GARCH parameters.
- Break points may also be responsible for long range dependence in the data set

Structural Breaks in volatility:

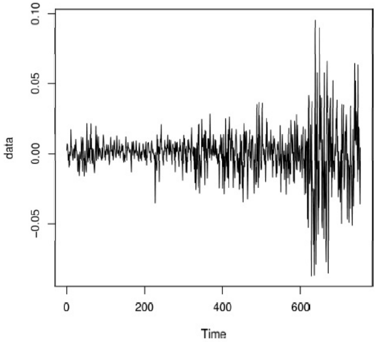


Fig. 5.1.1 S & P 500 data series

Structural Breaks in volatility:

- Cai (1994, JBES), Hamilton and Susmel (1994, JE), Kokoszka and Leipus (2000, Bernoulli), Lak (2008, WASJ) Markov-switching ARCH
- Gray (1996, JFE) , Klassen (2002, EE),Haas (1994, JF Econometrics), Gray (1996) propose a modified MS-GARCH model.
- Estimation of parameters is a challenging problem due to the non-Markovian structure of GARCH
- He and Maheu (2009, CS) propose particle filtering algorithm
- Bauwens, et. al. (2007) develop a Gibbs sampler algorithm for MS-GARCH models with a fixed number of regimes.
- Pastor and Stambaugh (2001), Kim, et. al. (2005, JBES) and Liu and Maheu (2008, JFE), Dueker (1997, JBES), Perignon and Smith (2007) discuss various methods to implement structural breaks

Structural Breaks in volatility:

- Various tests for testing the presence of structural breaks such as CUSUM, LM type tests and non parametric tests- Andreou and Ghysels (2002, JAE), Rapach and Strauss (2008, JAE), Smith (2008, JFE), Gao et. al. (2008, JE)
- None of the MS-GARCH models can account for shifts in the parameters of asymmetry.
- Structural breaks may shift the asymmetry and not accounting for this shift may lead to the spurious results
- There may be different asymmetric models appropriate for the different regimes.

Structural Breaks in volatility:

$$\epsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \begin{cases} \sigma_{1t}^2 & if \quad t \leq \tau_1 \\ \sigma_{2t}^2 & if \quad \tau_1 < t \leq \tau_2 \\ .. & .. \quad .. \\ .. & .. \quad .. \\ \sigma_{mt}^2 & if \quad \tau_{m-1} < t \leq \tau_m \\ \sigma_{(m+1)t}^2 & if \quad \tau_m < t \leq n \end{cases}$$

$$k \in \{1, 2, ..., m+1\}, \qquad \theta_k = (\omega_k, \alpha_k, \beta_k, b_k, c_k, \lambda_k, \nu_k)$$

$$\sigma_{kt}^2 = (\omega_k + \{\alpha_k \lambda_k [|\frac{\epsilon_{t-1}}{\sigma_{t-1}} - b_k| - c_k(\frac{\epsilon_{t-1}}{\sigma_{t-1}} - b_k)]^{\nu_k} + \beta_k\}\sigma_{t-1}^{\lambda_k})^{2/\lambda_k}$$

Break occurs at unknown time points $\Gamma_m = \{\tau_1, \tau_2, ..., \tau_m\}$
sufficient conditions for positivity are
$\omega_k > 0, \ \alpha_k \geq 0, \beta_k \geq 0, |c_k| \leq 1$

7. Need for Non Stationary Modeling

Structural Breaks in volatility:

- The above model can nest several symmetric and asymmetric volatility models
- It can take care about the shift in both the parameters of asymmetry, represented by b and c
- Capable of modeling different appropriate asymmetric volatility models in different regimes, which none of the MS-GARCH models can do.
- Hidden Markov model formulation Chib (1996, 1998, JE).
- Bayesian estimation - parameters and break points
- A new Bayesian method is suggested for assessing the suitability of the model
- Details may be found in Rohan & Ramanathan (2012)

Structural Breaks in volatility:

- We have analyzed S & P 500 and Dow Jones index series. The main finding of the analysis is the detection of the breaks around the same point in first and second weeks of September 2008.
- Financial crisis in the market started happening at the same time in the United states with bankruptcy of many major banks and insurance companies.
- We have also analyzed the S & P CNX Nifty series but no evidence for such a break found at all.

Stochastic Volatility Models:

- Introduce an innovation to the conditional variance equation - Taylor (1994), Harvey, Ruiz, and Shephard (1994), and Jacquier, Polson, and Rossi (1994)
- $\epsilon_t = \sigma_t Z_t, \quad (1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_m B^m) ln(\sigma_t^2) = \alpha_0 + \nu_t$
- $Z_t \sim iidN(0,1), \quad \nu_t \sim iidN(0, \sigma_\nu^2), \quad Z_t \perp \nu_t$
- Adding the innovation ν_t substantially increases the flexibility of the model in describing the evolution of σ_t^2, but it also increases the difficulty in parameter estimation.
- QML via Kalman filter, Bayesian (MCMC)
- Bovas Abraham & Balakrishna (....) for some of the latest developments on these models

Simple nonparametric volatility model:

- Fan & Yao (1998)
- y_t a strictly stationary stochastic process

$$y_t = \mu(y_{t-1}) + \sigma(y_{t-1})z_t, \quad E(z_t|\mathcal{F}_{t-1}) = 0, E(z_t^2|\mathcal{F}_{t-1}) = 1$$

- Obtain $\hat{\mu}(y) = \hat{\beta}_0(y)$ at $y = y_{t-1}; t = 1, 2, ..., n$ as the minimization intercept in

$$min_{\beta_0,\beta_1} n^{-1} \sum_{t=1}^{n} \{y_t - [\beta_0 + \beta_1(y_{t-1} - y)]\}^2 k_{h_1}(y_{t-1} - y)$$

- Obtain $\hat{\sigma}^2(y) = \hat{\gamma}_0(y)$ as the minimization intercept in

$$min_{\gamma_0,\gamma_1} n^{-1} \sum_{t=1}^{n} \{\hat{r}_t^2 - [\gamma_0 + \gamma_1(y_{t-1} - y)]\}^2 k_{h_2}(y_{t-1} - y)$$

$$r_t = y_t - \hat{\mu}(y_{t-1})$$

Simple nonparametric volatility model:

- It is shown that

$$\sqrt{nh_2}\left\{\hat{\sigma}^2(y) - \sigma^2(y) - B_n(y)\right\} \xrightarrow{d} N(0, V(y)),$$

where

$$B_n(y) = (1/2)\ h_2^2\ \mu_{2,1}\ \frac{\partial^2}{\partial y^2}\sigma^2(y) + o(h_1^2 + h_2^2),$$

$$V(y) = \mu_{0,2}\ \sigma^4(y)\ E\left\{\left(\epsilon_t^2 - 1\right)^2 | y_{t-1} = y\right\}/f(y)$$

$f(y)$ is the probability density function of y_t and

$$\mu_{s,t} = \int u^s k(u)^t du \quad for \quad s,t = 0,1,2.$$

Nonparametric volatility model - Extensions

- Hardle & Tsybakov (1997) - Joint estimation of conditional mean and variance functions
- Ziegelmann (2002) - Improvement - Addressed the problem of negative estimator of $\sigma^2(y)$ by considering an exponential tilting estimator
- Above methods are valid for ARCH(p) type information set.
- Extension to additive nonparametric volatility models where

$$\sigma^2(\mathbf{y}_{t-1}) = c + \sum_{i=1}^{d} \sigma_i^2(\mathbf{y}_{it-1})$$

- 'Curse of dimensionality'

Additive Models:

- Three approaches for the estimation of additive models:
  - Backfitting - Hastie & Tibshirani (1990); Estimates the additive functions through iterative calculations until certain convergence criterion is met.
  - Marginal integration - Newey (1994), Linton & Nielsen (1995); Based on averaging of multivariate kernels
  - Local instrumental variable estimation - Kim & Linton (2004);
- A review of above methods is available in Su et al. (2011)
- Most of the above methods do not ensure nonnegativity of the volatility estimators
- A generalization of additive model is available - Levine & Li (2007)

9. Time varying ARCH/GARCH Models

- Asian crisis - 1997, Russian crisis - 1998, Sept. 11 terrorist attack - 2001, Ongoing financial crisis 2008-; instability and hence the non-stationarity for the volatility process.
- Standard stationary volatility models such as ARCH, GARCH etc. may fail, non-stationary volatility models are more appropriate
- Dahlhaus & Rao (2006), Fryzlewicz et al. (2008) have proposed a time-varying ARCH (tvARCH) model for the volatility process
- ARCH models are of limited use as volatility process under this set up depend only on past few returns.
- Rohan & Ramanathan (2012) Generalized tvARCH to tv-GARCH. Such a model is expected to capture different dynamics in low and high volatility periods more efficiently

Dahlhaus and Suhasini Subba Rao (2006, 2007):

$$\epsilon_{t,T} = \sigma_{t,T} v_t,$$
$$\sigma^2_{t,T} = \omega(t/T) + \sum_{j=1}^{q} \alpha_j(t/T)\epsilon^2_{t-1,T}$$

tvARCH process can be locally approximated by a stationary ARCH process - Local stationarity.

Rescaling useful in asymptotics - Does not affect estimation procedure

Weighted quasi maximum likelihood estimation - Asymptotics properties including asymptotic normality (also extra bias due to nonstationarity of the process)

Fryzlewics, Sapatinas & Suhasini Subba Rao (2008):

Kernel-normalized least-squares estimation - outperforms kernel-QML - same rate of convergence - prediction based cross-validation for bandwidth selection - residual-based bootstrap - pointwise confidence intervals

A time varying GARCH model

tvGARCH:

$$\{\epsilon_t\}, \quad E(\epsilon_t|\mathcal{F}_{t-1}) = 0, \quad E(\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$$

$$\mathcal{F}_{t-1} = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots)$$

Suppose $\{v_t\}$ is $iid(0, \sigma^2)$ (independent of ϵ_t)

tvGARCH model with time varying parameters is defined as

$$\epsilon_t = \sigma_t v_t,$$
$$\sigma_t^2 = \omega(t) + \alpha(t)\epsilon_{t-1}^2 + \beta(t)\sigma_{t-1}^2$$

$\omega(\cdot)$, $\alpha(\cdot)$ and $\beta(\cdot)$ are certain non-negative functions of time.

Rescale the domain of the parameter functions to unit interval

$$\epsilon_t = \sigma_t v_t,$$
$$\sigma_t^2 = \omega\left(\tfrac{t}{n}\right) + \alpha\left(\tfrac{t}{n}\right)\epsilon_{t-1}^2 + \beta\left(\tfrac{t}{n}\right)\sigma_{t-1}^2, \;\; t = 1, 2, \ldots, n.$$

The sequence of stochastic processes $\{\epsilon_t, \;\; t = 1, 2, \ldots, n\}$ is said to follow a tvGARCH process.

$\omega(u), \alpha(u), \beta(u) \geq 0 \;\forall\; u \in (0, 1]$ ensure the non-negativity of σ_t^2.

Such a rescaling is a common technique in non-parametric regression and it does not affect the estimation procedure.

9. Time varying ARCH/GARCH Models

- We show that a non-stationary tvGARCH process can be locally approximated by stationary processes at specific time points.
- Therefore, the tvGARCH model is asymptotically locally stationary at every point of observation but it is globally non-stationary because of time-varying parameters.
- Such an approximation further helps us in deriving the asymptotic distribution of the estimators.

9. Time varying ARCH/GARCH Models

- An alternative approach: Cizek and Spokoiny (2009)
- Search for intervals of homogeneity over the entire period, such that the parameters of the process remain constant over each interval.
- Estimation using QML approach in each interval.
- Interval of homogeneity may not have enough observations to carry out the estimation.
- QML procedure is not very reliable when the sample size is small, Shephard (1996), Bose and Mukherjee (2003) and Fryzlewicz et. al. (2008)
- Non-Markovian structure of GARCH models as the volatility at time t in the GARCH model depends on the entire past and not only on the data inside the interval of homogeneity.
- The model and estimation procedure of Amado and Terasvirta (2008) also suffers from similar drawbacks.

Assumptions:

1. Parameter functions $\alpha(u)$ and $\beta(u)$ are such that

$$\sup_u E[log\left(\alpha(u)v_t^2 + \beta(u)\right)] < 0.$$

2. There exists $\delta > 0$ such that $0 < \alpha(u) + \beta(u) \le 1 - \delta, \; \forall \; 0 < u \le 1$ and a constant $0 < C < \infty$ such that $\sup_u \omega(u) \le C$.

3. There exist finite constants M_1, M_2 and M_3 such that $\forall \; u_1, u_2 \in (0, 1]$,

$$|\omega(u_1) - \omega(u_2)| \le M_1|u_1 - u_2|$$
$$|\alpha(u_1) - \alpha(u_2)| \le M_2|u_1 - u_2|$$
$$|\beta(u_1) - \beta(u_2)| \le M_3|u_1 - u_2|.$$

Proposition 1. *Let the Assumption 1 (i) hold. Then the variance process defined above has a well defined unique solution given by*

$$\bar{\sigma}_t^2 = \omega\left(\tfrac{t}{n}\right) + \sum_{i=1}^{\infty}\prod_{j=1}^{i}\left(\alpha\left(\tfrac{t-j+1}{n}\right)v_{t-j}^2 + \beta\left(\tfrac{t-j+1}{n}\right)\right)\omega\left(\tfrac{t-i}{n}\right),$$

such that $|\sigma_t^2 - \bar{\sigma}_t^2| \to 0$ *a.s., if* σ_0^2 *(starting point) is finite with probability one. Also,* $\inf_u \omega(u)/(1 - \inf_u \beta(u)) \leq \bar{\sigma}_t^2 < \infty\ \forall\ t$ *a.s., if for some* $0 < C < \infty$, $\sup_u \omega(u) \leq C$.

Proposition 2. *Suppose that the Assumptions 1 (i) and (ii) are satisfied for the tvGARCH process. Further assume that* $E|v_t|^4 < \infty$. *Then for a fixed* $k \geq 0$ *and* $0 < \delta < 1$,

$$Cov(\epsilon_t^2, \epsilon_{t+k}^2) = O\left((1-\delta)^k\right).$$

9. Time varying ARCH/GARCH Models

Define a stationary GARCH (1,1) process, which locally approximates the original process in the nbd. of a fixed point

$\widetilde{\epsilon}_t(u_0),\ u_0 \in (0,1],\ \ E(\widetilde{\epsilon}_t(u_0)|\mathcal{F}_{t-1}) = 0,\ \ E(\widetilde{\epsilon}_t^2(u_0)|\mathcal{F}_{t-1}) = \widetilde{\sigma}_t^2(u_0).$

$\{\widetilde{\epsilon}_t(u_0)\}$ a stationary GARCH process associated with original process at time point u_0

$$\widetilde{\epsilon}_t(u_0) = \widetilde{\sigma}_t(u_0)v_t,$$
$$\widetilde{\sigma}_t^2(u_0) = \omega(u_0) + \alpha(u_0)\widetilde{\epsilon}_{t-1}^2(u_0) + \beta(u_0)\widetilde{\sigma}_{t-1}^2(u_0).$$

Under Assump. 1(i), this is a stationary ergodic process. Assump. 1(ii) is sufficient for $\widetilde{\epsilon}_t(u_0)$ to be weakly stationary.

A unique stationary ergodic solution is

$$\bar{\sigma}_t^2(u_0) = \omega\left(u_0\right) + \sum_{i=1}^{\infty}\prod_{j=1}^{i}\Big(\alpha\left(u_0\right)v_{t-j}^2 + \beta\left(u_0\right)\Big)\,\omega\left(u_0\right).$$

Proposition 3. *Assump. 1 (i), (ii) and (iii) hold. The process $\{\epsilon_t^2\}$ can be approximated locally by a stationary ergodic process $\{\tilde{\epsilon}_t^2(u_0)\}$. That is, there exists a well defined stationary ergodic process V_t independent of u_0 and a constant $Q < \infty$ such that*

$$|\epsilon_t^2 - \tilde{\epsilon}_t^2(u_0)| \le Q \left(\left| \tfrac{t}{n} - u_0 \right| + \tfrac{1}{n} \right) V_t \quad a.s.$$

or equivalently $\epsilon_t^2 = \tilde{\epsilon}_t^2 + O_P \left(\left| \frac{t}{n} - u_0 \right| + \frac{1}{n} \right).$

We can also write the process by recursive substitution,

$$\sigma_t^2 = \alpha_0(\tfrac{t}{n}) + \sum_{k=1}^{t-1} \alpha_k(\tfrac{t}{n})\epsilon_{t-k}^2 + \sigma_0^2 \prod_{i=1}^{t} \beta\left(\tfrac{t-i+1}{n}\right),$$

where $\alpha_0(\frac{t}{n}) = \omega\left(\frac{t}{n}\right) + \sum_{k=1}^{t-1} \omega\left(\frac{t-k}{n}\right) \prod_{i=1}^{k} \beta\left(\frac{t-i+1}{n}\right),$

$$\alpha_k(\frac{t}{n}) = \alpha\left(\frac{t-k+1}{n}\right) \prod_{i=1}^{k-1} \beta\left(\frac{t-i+1}{n}\right), \quad k = 1, 2, \ldots t-1.$$

Local Polynomial Estimation: Two-stage estimation:

Step 1. First, we obtain a preliminary estimate of σ_t^2 using the following tvARCH (p) model;

$$\sigma_t^2 = \alpha_0(\frac{t}{n}) + \alpha_1(\frac{t}{n})\epsilon_{t-1}^2 \ldots + \alpha_p(\frac{t}{n})\epsilon_{t-p}^2$$

which can also be written as

$$\epsilon_t^2 = \alpha_0(\frac{t}{n}) + \alpha_1(\frac{t}{n})\epsilon_{t-1}^2 \ldots + \alpha_p(\frac{t}{n})\epsilon_{t-p}^2 + \sigma_t^2(v_t^2 - 1).$$

Here, p is such that $p = p_n \to \infty$ as $n \to \infty$. (i.e., p is $\log n$.)

Estimate the functions $\alpha_j(u)$, $j = 0, 1, \ldots p$, treating $\sigma_t^2(v_t^2 - 1)$ as error.

Local Polynomial Estimation:
Denote $(t/n) = u_t$.

$$\alpha_i(u_t) \approx \alpha_{i0} + \alpha_{i1}(u_t - u_0) + \ldots + \alpha_{id}(u_t - u_0)^d, \;\; i = 0, 1, \ldots, p$$

Given a Kernel function $K(\cdot)$, minimize

$$L = \sum_{i=p+1}^{n} \left(\epsilon_i^2 - \sum_{k=0}^{d} (\alpha_{0k} + \sum_{j=1}^{p} \alpha_{jk} \epsilon_{i-j}^2)(u_i - u_0)^k \right)^2 K_{h_1}(u_i - u_0)$$

where $K_{h_1}(\cdot) = (1/h_1)K(\cdot/h_1)$ and h_1 denotes the bandwidth.

Define

$$U_t = [1, (u_t - u_0), \ldots, (u_t - u_0)^d]_{1\times(d+1)} \;\; t = 1, 2, \ldots, n$$

$$X_1 = \begin{bmatrix} U_{p+1} & \epsilon_p^2 U_{p+1} & \dots & \epsilon_1^2 U_{p+1} \\ U_{p+2} & \epsilon_{p+1}^2 U_{p+2} & \dots & \epsilon_2^2 U_{p+2} \\ \vdots & \vdots & \ddots & \vdots \\ U_n & \epsilon_{n-1}^2 U_n & \dots & \epsilon_{n-p}^2 U_n \end{bmatrix},$$

$$W_1 = \operatorname{diag}(K_{h_1}(u_{p+1} - u_0), \dots, K_{h_1}(u_n - u_0)) \quad Y_1 = [\epsilon_{p+1}^2, \dots \epsilon_n^2]^\top.$$

$$\hat{\alpha}_i(u_0) = e_{i(d+1)+1,(p+1)(d+1)}^\top (X_1^\top W_1 X_1)^{-1} X_1^\top W_1 Y_1, \quad i = 0, 1, \dots, p.$$

An initial estimate of σ_t^2 is obtained by

$$\hat{\sigma}_t^2 = \hat{\alpha}_0(u_t) + \sum_{k=1}^{p} \hat{\alpha}_k(u_t)\epsilon_{t-k}^2.$$

We set $\epsilon_t^2 = 0, \ \forall \ t \leq 0$ for the practical implementation.

Step 2. Use the conditional variance initially estimated in Step 1 to get the estimates of the parameter functions

$$\omega(u_t) \approx \omega_{02} + \omega_{12}(u_t - u_0) + \ldots + \omega_{d2}(u_t - u_0)^d$$
$$\alpha(u_t) \approx a_{02} + a_{12}(u_t - u_0) + \ldots + a_{d2}(u_t - u_0)^d$$
$$\beta(u_t) \approx b_{02} + b_{12}(u_t - u_0) + \ldots + b_{d2}(u_t - u_0)^d$$

where ω_{i2}, a_{i2} and $b_{i2},\ i = 0, 1, \ldots, d$ are constants.

$$\epsilon_t^2 = \omega(\tfrac{t}{n}) + \alpha(\tfrac{t}{n})\epsilon_{t-1}^2 + \beta(\tfrac{t}{n})\hat{\sigma}_{t-1}^2 - \beta(\tfrac{t}{n})(\hat{\sigma}_{t-1}^2 - \sigma_{t-1}^2) + \sigma_t^2(v_t^2 - 1).$$

A particular choice of the Step 1 bandwidth $h_1 = o(h_2)$

$E(\hat{\sigma}_{t-1}^2 - \sigma_{t-1}^2)$ is asymptotically negligible.
The estimates are obtained by minimizing

$$L = \sum_{i=2}^{n}\left(\epsilon_i^2 - \sum_{k=0}^{d}(\omega_{k2} + a_{k2}\epsilon_{i-1}^2 + b_{k2}\hat{\sigma}_{i-1}^2)(u_i - u_0)^k\right)^2 K_{h_2}(u_i - u_0).$$

Define

$$X_2 = \begin{bmatrix} U_2 & \epsilon_1^2 U_2 & \hat{\sigma}_1^2 U_2 \\ U_3 & \epsilon_2^2 U_3 & \hat{\sigma}_2^2 U_3 \\ \vdots & \vdots & \vdots \\ U_n & \epsilon_{n-1}^2 U_n & \hat{\sigma}_{n-1}^2 U_n \end{bmatrix},$$

$$W_2 = \text{diag}(K_{h_2}(u_2 - u_0), \ldots, K_{h_2}(u_n - u_0)), \quad \text{and} \quad Y_2 = [\epsilon_2^2, \ldots, \epsilon_n^2]^\top.$$

Exact expressions for the estimators are given by

$$\begin{aligned} \hat{\omega}(u_0) &= e_{1,3(d+1)}^\top (X_2^\top W_2 X_2)^{-1} X_2^\top W_2 Y_2, \\ \hat{\alpha}(u_0) &= e_{d+2,3(d+1)}^\top (X_2^\top W_2 X_2)^{-1} X_2^\top W_2 Y_2 \quad \text{and} \\ \hat{\beta}(u_0) &= e_{2d+3,3(d+1)}^\top (X_2^\top W_2 X_2)^{-1} X_2^\top W_2 Y_2. \end{aligned}$$

Final estimates of σ_t^2 in tvGARCH can be obtained using these estimators.

Optimal rate of convergence if $d = 3$

Bandwidth selection: Two step estimator is not very sensitive to the choice h_1 as long as it is small enough (bias in first step is negligible)

Apply the standard univariate bandwidth selection procedures to select the smoothing parameter for Step 2.

The initial smoothing parameter can be chosen according to the second step bandwidth.

Practice: Select the optimal bandwidth (h_2) using the CV method based on the best linear predictor of ϵ_t^2 given the past (see Hart (1994))

$$CV(h_2) = \frac{1}{n-1} \sum_{t=2}^{n} \left(\epsilon_t^2 - \hat{\omega}^{-t}(u_t) - \hat{\alpha}^{-t}(u_t)\epsilon_{t-1}^2 - \hat{\beta}^{-t}(u_t)\sigma_{t-1}^2 \right)^2$$

$\hat{\omega}^{-t}(u_t), \hat{\alpha}^{-t}(u_t)$ and $\hat{\beta}^{-t}(u_t)$ - local polynomial estimators obtained by leaving the t^{th} observation.

Use pilot bandwidth initially to get the initial estimate of σ_{t-1}^2 using the full data.

Using the similar arguments as in Hart (1994) asymptotically it can be shown that such a bandwidth is a minimizer of the mean squared prediction error of ϵ_t^2.

Bandwidth selection procedure is computationally too cumbersome, specially when n is large. We provide a simplified version.

Modeling and forecasting volatility using tvGARCH:

1. Analyze the currency exchange rates between five major developing economies India (INR), China (CNY), Brazil (BRL), Russia (RUB) and South Africa (RND) (BRICS) and the developed economies viz. United States (USD) and Europe (EURO).
2. Daily percent log returns ranging from January 2000 (dates varying) to December 31, 2010 except NSE data, which start from January 2002.
3. Compare the in-sample performance of tvGARCH with several other well known existing models using aggregated mean squared error (AMSE):

$$\text{AMSE} = (1/n) \sum_{t=1}^{n} (\epsilon_t^2 - \hat{\sigma}_t^2)^2,$$

Aggregated mean squared errors of the 'in sample forecasts'.

Series	tvGARCH	tvARCH (1)	tvARCH (2)	GARCH	EGARCH	GJR	FIGARCH
INR/USD	35.59	36.37	**33.27**	40.23	38.03	40.26	38.68
INR/EURO	**2119.02**	2158.96	2137.97	2234.45	2524.825	2234.46	2249.93
CNY/USD	0.72	0.74	**0.71**	1.03	-	1.22	0.96
CNY/EURO	**76.64**	80.12	79.75	84.02	84.17	84.55	85.73
BRL/USD	1174.72	1276.72	**1117.56**	1249.60	1163.88	1312.59	1221.218
BRL/EURO	**3563.09**	4295.43	3844.53	4942.11	4402.06	5320.83	4861.54
RUB/EURO	**65.27**	68.77	68.34	73.98	72.81	74.04	69.34
RND/EURO	**935.43**	977.31	966.16	993.15	981.79	1016.55	989.07
S & P 500	**2154.41**	2979.67	2652.07	2614.41	2476.76	2679.29	2572.90
Dow Jones	**1715.59**	2330.29	2067.89	2075.91	1951.45	2125.98	2025.00
BSE	**5688.85**	6170.73	6026.22	6358.63	6095.25	6539.42	6381.01
NSE	**6205.13**	6764.44	6556.36	7134.17	6765.58	7398.32	7112.78

tvGARCH outperformed stationary as well as long memory models

In the Figure below for BRL/EURO data, the tvGARCH model has captured the ups and downs in the volatility more accurately.

The faint plot depicts the squared returns and the dark plot is the predicted volatility with the corresponding model

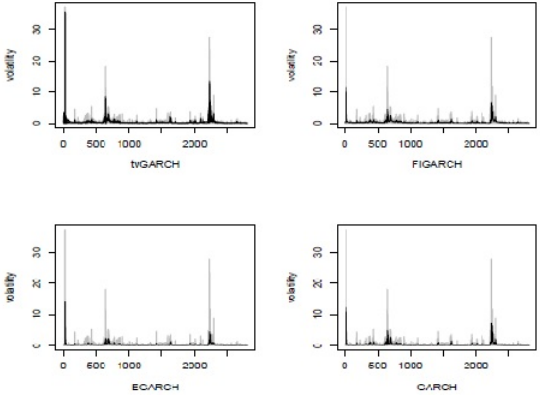


Figure 1 In sample volatility forecasts for the BRL/EURO data

Realized volatility forecasts: $\text{AMSE} = \sum_{t=1}^{[n/22]} (\bar{\epsilon}_t^2 - \bar{\sigma}_t^2)$

$\bar{\epsilon}_t^2 = \sum_{i=1}^{22} \epsilon^2_{22(t-1)+i}$ and $\bar{\sigma}_t^2 = \sum_{i=1}^{22} \hat{\sigma}^2_{22(t-1)+i}$; actual and predicted monthly realized volatilities (22 trading days).

Aggregated mean squared errors of the monthly realized volatility forecasts.

Series	tvGARCH	tvARCH (1)	tvARCH (2)	GARCH	EGARCH	GJR	FIGARCH
INR/USD	27.01	44.72	17.86	12.55	17.31	12.14	**11.85**
INR/EURO	**1788.12**	2450.51	2151.76	2430.32	9453.29	2374.42	2432.64
CNY/USD	0.32	0.76	0.23	0.08	-	0.17	**0.04**
CNY/EURO	**37.43**	102.92	102.19	85.73	99.35	85.21	92.23
BRL/USD	1867.04	3403.41	1617.07	**943.45**	2299.42	985.10	1177.51
BRL/EURO	**4057.67**	7853.92	6234.64	9161.12	8886.305	9894.01	9066.51
RUB/EURO	**45.32**	64.92	62.02	156.02	144.11	157.40	78.03
RND/EURO	**1058.51**	1978.03	1825.70	1928.56	2073.12	2040.18	2048.00
S & P 500	**679.53**	10972.46	5191.94	3253.30	3816.98	3727.57	2801.21
Dow Jones	**847.539**	9483.53	4448.66	2141.73	3112.366	2347.31	1611.60
BSE	5227.33	13691.05	11261.04	**2808.24**	5097.00	3109.73	3068.11
NSE	**2917.72**	11081.75	9009.67	9359.98	8601.01	9489.36	9351.92

Out of sample forecast: One week ahead daily forecasts and their AMSE

Performance of tvGARCH is much better than other models.

tvGARCH attains the lowest AMSE for 5 data sets, while FIGARCH is better in 3 cases.

There is very little difference in the AMSE of the models selected by FIGARCH and tvGARCH as compared to other models.

Aggregated mean squared errors of the one week out of sample volatility forecasts.

Series	tvGARCH	tvARCH (1)	tvARCH (2)	GARCH	EGARCH	GJR	FIGARCH
INR/USD	0.0211	0.0205	**0.0196**	0.0239	.0219	.0238	.0242
INR/EURO	4.7786	4.4323	5.0960	4.6835	5.7211	4.6890	**4.3101**
CNY/USD	0.0002	0.76	0.23	0.0056	-	0.0040	**0.0001**
CNY/EURO	0.0831	**0.0103**	0.0106	0.1043	0.0953	0.1043	0.0265
BRL/USD	0.0626	0.1760	0.1299	0.1501	0.1750	0.1290	**0.0366**
BRL/EURO	**0.0106**	0.1192	0.1167	0.0144	0.0111	0.0177	0.0120
RUB/EURO	**0.0538**	0.0576	0.0539	0.0575	0.0586	0.0572	0.0571
RND/EURO	**0.0431**	0.0439	0.0428	0.0559	0.0579	0.0535	0.0468
S & P 500	0.2971	0.2915	0.2843	0.0527	**0.0451**	0.0552	0.3140
Dow Jones	0.0080	0.0498	0.0678	0.0120	**0.0039**	0.0136	0.0047
BSE	**0.8749**	1.0130	1.0016	1.0521	1.0077	1.0576	1.0953
NSE	**0.8291**	0.8987	0.8738	0.9104	0.8631	0.9129	0.9572

Asymptotic Results: Assumption 2.

1. The functions $\omega(\cdot), \alpha(\cdot)$ and $\beta(\cdot)$ (and hence $\alpha_j(\cdot)$) have the bounded and continuous derivatives up to order $d+1$ $(d \geq 1)$, in the neighborhood of u_0, $u_0 \in (0,1]$.
2. $K(u)$ is a symmetric density function of bounded variation with a compact support.
3. The bandwidths h_1 and h_2 are such that $h_1 \to 0, h_2 \to 0$ and $nh_1 \to \infty, nh_2 \to \infty$ as $n \to \infty$.
4. $E|v_t|^4 < \infty$.

Notations.

$$
\begin{aligned}
\mu_i = & \int u^i K(u)du, \quad \nu_i = \int u^i K^2(u)du, \\
S = & S(u_0) = E\left([1, \tilde{\epsilon}^2_{t-1}(u_0), \ldots, \tilde{\epsilon}^2_{t-p}(u_0)]^\top [1, \tilde{\epsilon}^2_{t-1}(u_0), \ldots, \tilde{\epsilon}^2_{t-p}(u_0)]\right), \\
C_j = & C_j(u_0) = E(\tilde{\epsilon}^2_t(u_0)\ \tilde{\epsilon}^2_{t-j}(u_0)), \\
\Omega = & \Omega(u_0) = E\left(\tilde{\sigma}^4_t(u_0)[1, \tilde{\epsilon}^2_{t-1}(u_0), \ldots, \tilde{\epsilon}^2_{t-p}(u_0)]^\top [1, \tilde{\epsilon}^2_{t-1}(u_0), \ldots, \tilde{\epsilon}^2_{t-p}(u_0)]\right. \\
w_j = & E(\tilde{\epsilon}^j_t(u_0)), \quad \alpha_{tvARCH}(u_0) = [\alpha_0(u_0), \alpha_1(u_0), \ldots, \alpha_p(u_0)]^\top, \\
D_i = & [\mu_{d+1}, h_i\mu_{d+2}, \ldots, h_i^d \mu_{2d+1}]^\top, \quad i = 1, 2, \\
e_m = & \text{ a column vector of length } m \text{ with 1 everywhere,}
\end{aligned}
$$

$$A_i = \begin{bmatrix} 1 & h_i\mu_1 & \dots & h_i^d\mu_d \\ h_i\mu_1 & h_i^2\mu_2 & \dots & h_i^{d+1}\mu_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_i^d\mu_d & h_i^{d+1}\mu_{d+1} & \dots & h_i^{2d}\mu_{2d} \end{bmatrix},$$

$$B_i = \begin{bmatrix} \nu_0 & h_i\nu_1 & \dots & h_i^d\nu_d \\ h_i\nu_1 & h_i^2\nu_2 & \dots & h_i^{d+1}\nu_{d+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_i^d\nu_d & h_i^{d+1}\nu_{d+1} & \dots & h_i^{2d}\nu_{2d} \end{bmatrix}, \quad i = 1, 2.$$

Theorem 1. Let the Assumptions 1 and 2 be satisfied. Then the asymptotic bias of $\hat{\alpha}_j(u_0),\ j = 0, 1, \ldots, p$ is given by,

$$\text{Bias}(\hat{\alpha}_j(u_0)) = \frac{h_1^{d+1}}{(d+1)!} \left(\alpha_j^{(d+1)}(u_0)\right) e_{1,d+1}^{\top} A_1^{-1} D_1 + o_P(h_1^{d+1}).$$

If $E|v_t|^8 < \infty$, then

$$\begin{aligned} &Var(\hat{\alpha}_0(u_0), \ldots, \hat{\alpha}_p(u_0)) \\ &\qquad = \tfrac{1}{nh_1} e_{1,d+1}^{\top} A_1^{-1} B_1 A_1^{-1} e_{1,d+1} Var(v_t^2) S^{-1} \Omega S^{-1} (1 + o_P(1)) \end{aligned}$$

The bias for $\hat{\alpha}_j(u_0)$ depends on the $(d+1)^{th}$ derivative of $\alpha_j(u_0)$

MSE of the estimator $\hat{\alpha}_j(u_0)$ is $O_P(h_1^{2d+2} + (nh_1)^{-1})$.

This rate is far better even for $d = 1$, than the least-squares estimator of Fryzlewicz et. al. (2008), that is $O_P(h_1^{2\gamma} + (nh_1)^{-1}), \ 0 < \gamma < 1$.

The optimal bandwidth $h_1 = O(n^{-1/9})$ is used, then for $d = 3$, the local polynomial estimator achieves the optimal rate of convergence $O_P(n^{-8/9})$ (see Fan and Zhang (1999)) for estimating $\alpha_j(u_0)$.

Corollary 1. *Under the same assumptions as that of Theorem 1,*

$$\sqrt{nh_1}\left(\hat{\alpha}_{tvARCH}(u_0) - \alpha_{tvARCH}(u_0) - b(u_0)\right) \xrightarrow{D} N_{p+1}\left(0, e_{1,d+1}^{\top} A_1^{-1} B_1 A_1^{-1} e_{1,d+1} Var(v_t^2) S^{-1} \Omega S^{-1}\right)$$

where $b(u_0) = Bias(\hat{\alpha}_{tvARCH}(u_0))$

Corollary 2. *Let* $\hat{\sigma}_t^2 = \hat{\alpha}_{tvARCH}(u_t)^{\top}[1, \epsilon_{t-1}^2, \ldots, \epsilon_{t-p}^2]_{(p+1)\times 1}^{\top}$. *Then under the Assumptions 1 and 2,*

$$Bias(\hat{\sigma}_t^2) = E(\hat{\sigma}_t^2 - \sigma_t^2) = O_P(h_1^{d+1}) + O(\rho^{p_n})$$

where $0 < \rho < 1$ *and* $p_n \to \infty$ *as* $n \to \infty$.

Notations.

$$b_j = b_j(u_0) = \text{Bias}(\hat{\alpha}_j(u_0)),\ \delta_j = \delta_j(u_0) = \alpha_j(u_0) + b_j(u_0),\ j = 0, 1, \ldots, p,$$

$$\lambda_1 = \delta_0 + \sum_{j=1}^{p} \delta_j w_2,\ \lambda_2 = \delta_0 w_2 + \sum_{j=1}^{p} \delta_j C_j,$$

$$\lambda_3 = \delta_0^2 + 2\delta_0 w_2 \sum_{j=1}^{p} \delta_j + \sum_{j=1}^{p} \delta_j^2 w_4 + 2 \sum_{i,j=1(i<j)}^{p} \delta_i \delta_j C_{j-i},$$

$$\lambda_{1b} = b_0 + \sum_{j=1}^{p} b_j w_2,\ \lambda_{2b} = b_0 w_2 + \sum_{j=1}^{p} b_j C_j,$$

$$\lambda_{3b} = \delta_0 b_0 + (b_0 \sum_{j=1}^{p} \delta_j + \delta_0 \sum_{j=1}^{p} b_j) w_2 + \sum_{j=1}^{p} b_j \sum_{j=1}^{p} \delta_j w_4,$$

$$\Omega_2 =$$

$$E\left(\widetilde{\sigma}_t^4(u_0)[1, \widetilde{\epsilon}_{t-1}^2(u_0), \delta_0 + \sum_{j=1}^{p} \delta_j \widetilde{\epsilon}_{t-j}^2(u_0)]^\top [1, \widetilde{\epsilon}_{t-1}^2(u_0), \delta_0 + \sum_{j=1}^{p} \delta_j \widetilde{\epsilon}_{t-j}^2(u_0)]\right),$$

$$D^* = [1, h_2\mu_1, \ldots, h_2^d \mu_d]^\top, \quad S_2 = \begin{bmatrix} 1 & w_2 & \lambda_1 \\ w_2 & w_4 & \lambda_2 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{bmatrix}.$$

Theorem 2. Under the Assumptions 1 and 2, the asymptotic biases of the estimates of parameters in the two step procedure are given as

$$\begin{aligned}
\text{Bias}(\hat{\omega}(u_0)) = \; & \frac{h_2^{d+1}}{(d+1)!}\left(\omega^{(d+1)}(u_0)\right) e_{1,d+1}^{\top} A_2^{-1} D_2 \\
& - \frac{\beta(u_0)}{|S_2|}\left(\lambda_{1b}(\lambda_3 w_4 - \lambda_2^2) - \lambda_{2b}(\lambda_3 w_2 - \lambda_1 \lambda_2)\right. \\
& \left. + \lambda_{3b}(\lambda_2 w_2 - \lambda_1 w_4)\right) e_{1,d+1}^{\top} A_2^{-1} D^* + o_P(h_2^{d+1}), \\
\text{Bias}(\hat{\alpha}(u_0)) = \; & \frac{h_2^{d+1}}{(d+1)!}\left(\alpha^{(d+1)(u_0)}\right) e_{1,d+1}^{\top} A_2^{-1} D_2 \\
& - \frac{\beta(u_0)}{|S_2|}\left(-\lambda_{1b}(\lambda_3 w_2 - \lambda_1 \lambda_2) + \lambda_{2b}(\lambda_3 - \lambda_1^2)\right. \\
& \left. - \lambda_{3b}(\lambda_2 - \lambda_1 w_2)\right) e_{1,d+1}^{\top} A_2^{-1} D^* + o_P(h_2^{d+1}), \\
\text{Bias}(\hat{\beta}(u_0)) = \; & \frac{h_2^{d+1}}{(d+1)!}\left(\beta^{(d+1)}(u_0)\right) e_{1,d+1}^{\top} A_2^{-1} D_2 \\
& - \frac{\beta(u_0)}{|S_2|}\left(\lambda_{1b}(\lambda_2 w_2 - \lambda_1 w_4) - \lambda_{2b}(\lambda_2 - \lambda_1 w_2)\right. \\
& \left. + \lambda_{3b}(w_4 - w_2^2)\right) e_{1,d+1}^{\top} A_2^{-1} D^* + o_P(h_2^{d+1})
\end{aligned}$$

and under the additional assumption that, $E|v_t|^8 < \infty$, the asymptotic variance is

$$Var(\hat{\omega}(u_0), \hat{\alpha}(u_0), \hat{\beta}(u_0)) = \frac{1}{nh_2} e_{1,d+1}^{\top} A_2^{-1} B_2 A_2^{-1} e_{1,d+1} Var(v_t^2) S_2^{-1} \Omega_2 S_2^{-1} (1 + o_P(1)).$$

- the second part (containing $\lambda_{1b}, \lambda_{2b}$ and λ_{3b}) is due to the initial approximation of σ_t^2 as in Step 1 (see proof of Theorem 2).
- This term is $O_P(h_1^{d+1})$ as each $\lambda_{ib}, i = 1, 2, 3$ is $O_P(h_1^{d+1})$ (using Theorem 1).
- Therefore if we choose $h_1 = o(h_2)$ then the bias expression becomes free of the bias due to the first step asymptotically.

$$\text{Bias}(\hat{\omega}(u_0)) = \frac{h_2^{d+1}}{(d+1)!} \left(\omega^{(d+1)}(u_0)\right) e_{1,d+1}^\top A_2^{-1} D_2 + o_P(h_2^{d+1}),$$
$$\text{Bias}(\hat{\alpha}(u_0)) = \frac{h_2^{d+1}}{(d+1)!} \left(\alpha^{(d+1)(u_0)}\right) e_{1,d+1}^\top A_2^{-1} D_2 + o_P(h_2^{d+1}),$$
$$\text{Bias}(\hat{\beta}(u_0)) = \frac{h_2^{d+1}}{(d+1)!} \left(\beta^{(d+1)}(u_0)\right) e_{1,d+1}^\top A_2^{-1} D_2 + o_P(h_2^{d+1})$$

- Bias expressions are free of the derivatives of other parameter functions.
- If $h_1 = o(h_2)$, then $\delta_j = \alpha_j(u_0) + o_P(h_2^{d+1})$ and the variance of the estimator does not depend on the first step bandwidth.
- When the optimal bandwidth h_2 of order $n^{-1/9}$ is used, then the estimation remains unaffected for a large choice of initial step bandwidth.
- The estimation procedure relatively easy to implement.
- The MSE of the final estimator is $O_P(h_2^{2d+2} + (nh_2)^{-1})$, which is independent of the initial step bandwidth.
- This MSE achieves the optimal rate of convergence $(O_P(n^{-8/9}))$ for $d = 3$.

Corollary 3. *Under the same assumptions as that of Theorem 2,*

$$\sqrt{nh_2}\left(\hat{\beta}_{tvGARCH}(u_0) - \beta_{tvGARCH}(u_0) - b_{tvGARCH}(u_0)\right) \xrightarrow{D} N_3\left(0, e_{1,d+1}^{\top} A_2^{-1} B_2 A_2^{-1} e_{1,d+1} Var(v_t^2) S_2^{-1} \Omega_2 S_2^{-1}\right)$$

$$\beta_{tvGARCH}(u_0) = [\omega(u_0), \alpha(u_0), \beta(u_0)]^{\top}$$

$$b_{tvGARCH}(u_0) = [Bias(\hat{\omega}(u_0)), Bias(\hat{\alpha}(u_0)), \ Bias(\hat{\beta}(u_0))]^{\top}$$

. **Two Issues:**

1. Variance estimation; bootstrap, Fryzlewicz et. al. (2008) - residual based bootstrap, Bose and Mukherjee (2009)
2. Varying degree of smoothness for parametric functions - non optimality - Fan and Zhang (1999)

A feasible bandwidth selection criteria: A relation between the $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$ and $(\hat{\omega}^{-t}, \hat{\alpha}^{-t}, \hat{\beta}^{-t})$

Proposition A. *Let $\hat{\beta}_2(u_0)$ be the local polynomial estimator of $\beta_2(u_0)$ where $\beta_2 = (\omega_{02}, \omega_{12}, \ldots, \omega_{d2}, a_{02}, \ldots, a_{d0},\ b_{02}, \ldots, b_{d2})$. Suppose that $\hat{\beta}_2^{-t}(u_0)$ denotes the leave one out (obtained by eliminating the t^{th} observation) estimators of $\beta_2(u_0)$. Then,*

$$\begin{aligned} \hat{\beta}_2^{-i}(u_0) = \quad & \left(\hat{\beta}_2(u_0) - (X_2^\top W_2 X_2)^{-1} X_2^\top W_2 I_i^* Y_2\right) \\ & + Z_i \left(\hat{\beta}_2(u_0) - (X_2^\top W_2 X_2)^{-1} X_2^\top W_2 I_i^* Y_2\right) \end{aligned}$$

where
$Z_i = (X_2^\top W_2 X_2)^{-1} X_2^\top W_2 \left(I_{n-1} + I_i^* X_2 (X_2^\top W_2 X_2)^{-1} X_2^\top W_2\right)^{-1} I_i^* X_2$ *and* I_i^* *denotes a matrix of order $(n-1) \times (n-1)$ with $(i,i)^{th}$ element as one and rest of them as zero. Now*
$\hat{\omega}^{-i}(u_0) = e_{1,3(d+1)} \beta_2^{-i}(u_0),\ \hat{\alpha}^{-i}(u_0) = e_{d+1,3(d+1)} \beta_2^{-i}(u_0)$ *and*
$\hat{\beta}^{-i}(u_0) = e_{2d+3,3(d+1)} \beta_2^{-i}(u_0)$.

Some Remarks:

1. Comparative analysis of the bilateral exchange rates between five major emerging economies viz., Brazil, Russia, India, China and South Africa (BRICS)
2. These exchange rates have tremendously shifted themselves towards nonstationarity as tvGARCH model outperforms several stationary GARCH models.
3. A comparison with FIGARCH (Baillie(1996)) has evidenced that the bilateral exchange rates volatility has lost its widespread notion of long memory behavior and has shifted towards short memory and non-stationarity.
4. Two major Indian stock indices,which were presumed to be invulnerable to the financial crisis of 2008, have also shown the signs of instability.

1. The local linear estimator achieves the optimal rate of convergence asymptotically.
2. Our estimator of tvARCH achieves better rate of convergence than the least squares estimator of Fryzlewicz et al. (2008).
3. Even though this paper deals with tvGARCH (1,1) process only, the results presented here can be easily extended to a general tvGARCH (p, q) with slight modifications. See Rohan (2013)
4. Weighted bootstrap Bose Mukherjee (2009)

9. Time varying ARCH/GARCH Models

1. The global crisis vehemently turned the exchange rates volatility towards non-stationarity and short memory.
2. The frequent manipulation of the currencies may lead the currency rates to lose its widespread notion of the long memory behavior.
3. The out of sample forecasts are obtained using the last estimated coefficients at time n.
4. The model should be used cautiously for long term forecasting and it is advisable to re-estimate this model frequently to get more accurate forecasts due to its time varying nature of parameters.

1. The out of sample forecasts of tvGARCH are better than those of other models.
2. The tvGARCH attains the lowest AMSE for 5 data sets, while FIGARCH is better in 3 cases. However, even for those data in which FIGARCH has better forecasts, there is very little difference in the AMSE of the models selected by FIGARCH and tvGARCH as compared to other models.
3. For CNY/USD, tvGARCH has very low AMSE and the difference of AMSE of forecasts of FIGARCH and tvGARCH is just 0.0001.
4. EGARCH has shown good forecasts for S & P 500 and Dow Jones indices, while GARCH and GJR models are performing abysmally.

THANK YOU