Economics Bulletin

Volume 36, Issue 4

Fluke of stochastic volatility versus GARCH inevitability or which model creates better forecasts?

Valeriya V. Lakshina
National Research University Higher School of Economics

Andrey M. Silaev
National Research University Higher School of Economics

Abstract

The paper proposes the thorough investigation of in-sample and out-of-sample performance of five GARCH and two stochastic volatility models, estimated on the Russian financial data. The data includes prices of Aeroflot and Gazprom stocks and Ruble against US dollar exchange rates. In our analysis we use probability integral transform for in-sample comparison and Mincer-Zarnowitz regression along with classical forecast performance measures for out-of-sample comparison. Studying both the explanatory and the forecasting power of the considered models we came to the conclusion that stochastic volatility models perform equally or in some cases better than GARCH models.

Citation: Valeriya V. Lakshina and Andrey M. Silaev, (2016) "Fluke of stochastic volatility versus GARCH inevitability or which model creates better forecasts?", *Economics Bulletin*, Volume 36, Issue 4, pages 2368-2380
Contact: Valeriya V. Lakshina - vlakshina@hse.ru, Andrey M. Silaev - asilaev@hse.ru.
Submitted: September 19, 2016. **Published:** December 10, 2016.

1. Introduction

Predicting volatility of financial assets is an important task for the purposes of asset pricing, portfolio allocation and risk management. There is a long-standing discussion about what volatility measures predict the future volatility more efficient in various scenarios. Concerning option pricing GARCH models are usually compared with implied and historical volatility, but still no consensus is reached.

Another fundamentally different way of volatility modeling is developed in stochastic volatility models (SV). The main difference between them and GARCH-type models is that the former contains an additional innovation term for volatility dynamics, which may or may not be related to the returns' innovations. Moreover stochastic volatility models require more sophisticated estimation techniques based on simulations because the closed-form solution is rarely exists. The examples of comparison of GARCH with SV models can be found in (Danielsson, 1994) and (Shephard, 1996).

In more recent study of (Chuang et al., 2013) the above-mentioned volatility measures (except SV) are compared with Markov switching multifractal model (MSM) introduced in (Calvet and Fisher, 2004). Unlike GARCH or implied volatility the multifractal structure of MSM is able to capture not only the clustering feature of volatility process but also the outliers and long-memory behavior of volatility. As a result the authors recognize that MSM do outperforms the implied volatility in the out-of-sample performance.

Notably that like SV model MSM also incorporates uncertainty in the volatility process but completely in different way than in SV (for details see Section 2). This resulted in the fact that MSM belonging to the stochastic volatility class has a closed-form likelihood function and can be estimated via usual optimization procedure.

The paper is aimed at the comparison of two stochastic volatility models which both model the volatility via random process but substantially differ in terms of computational efforts. However, GARCH models are used as a traditional benchmarks in volatility estimation and forecasting.

The paper is organized as follows. Section 2 describes the set of models to be compared. Section 3 presents the data and parameter estimates for the chosen models. Section 4 considers the goodness-of-fit and forecast performance issues and discuss the results. Section 5 concludes.

2. Models' description

From more than three hundred ARCH-type models (Hansen and Lunde, 2005) we pick four: ordinary GARCH, exponential GARCH, Glosten-Jagannathan-Rünkle model (GJR) and threshold ARCH. For all these models we estimate simple specification with one ARCH and one GARCH terms. The choice of the models is induced by the prevalence of this specifications in the financial literature, especially when it comes to the applicability of the results in practice.

We also take original stochastic volatility (or as adopted in literature—the stochastic volatility) and Markov switching multifractal.

2.1 GARCH

All GARCH-type model have similar set up, distinguishing on the volatility equations. Firstly we have a time series x_t of T daily log returns:

$$x_t = E(x_t|\mathcal{F}_{t-1}) + y_t\,,\ \ t = 1, \ldots, T,$$

where $E(x_t|\mathcal{F}_{t-1})$ is a conditional mean of daily returns x_t at time t conditional on all available at $t-1$ information F_{t-1}, y_t are usually called innovations. Returns x_t are calculated as a logarithm of today price divided by the price yesterday: $x_t = log(\frac{P_t}{P_{t-1}})$. Conditional mean $E(x_t|\mathcal{F}_{t-1})$ is modelled by ARMA(p,q), see (1).

$$E(x_t|\mathcal{F}_{t-1}) = \omega + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j}\,, \tag{1}$$

where parameters α_i and β_j are the ith-order autoregressive (AR) and jth-order moving average (MA) terms. Consequently innovations y_t have zero mean and a time-dependent variance σ_t^2, which is modeled by (2) and (3).

$$y_t = \sigma_t \eta_t\,,\ \ \eta_t \sim N(0,1) \tag{2}$$

$$\sigma_t^2 = c + \sum_{i=1}^{k} \kappa_i y_{t-i}^2 + \sum_{j=1}^{m} \mu_j \sigma_{t-j}^2\,, \tag{3}$$

where parameter κ_i represents ith-order ARCH term, μ_j—the jth-order GARCH term, η_t are standardized innovations or standardized residuals, which are normally distributed with zero mean and unit variance. ARCH term in (3) allows to capture the effects of volatility clustering and GARCH term is responsible for volatility autocorrelation estimated by μ_j.

Exponential GARCH (Nelson, 1991) also allows to capture the leverage effect (i. e. the asymmetric volatility response to negative and positive returns) and ensure simpler evaluation of shock persistence (4).

$$\ln(\sigma_t^2) = c + \sum_{i=1}^{k} \left(\kappa_i \eta_{t-i} + \gamma_i \left(|\eta_{t-i}| - E(|\eta_{t-i}|)\right)\right) + \sum_{j=1}^{m} \mu_j \ln(\sigma_{t-j}^2), \tag{4}$$

where $\eta_t = {}^{y_t}/_{\sigma_t}$ are standardized innovations, γ_i estimates the leverage effect.

The GJR model (Glosten et al., 1993) solves the same problem of leverage effect modeling via the use of the indicator function $I(\cdot)$ (5).

$$\sigma_t^2 = c + \sum_{i=1}^{k} \left(\kappa_i y_{t-i}^2 + \gamma_i I(y_{t-i}) y_{t-i}^2\right) + \sum_{j=1}^{m} \mu_j \sigma_{t-j}^2, \tag{5}$$

where function I takes the value of 1 if $y_{t-i} \leq 0$ and 0 otherwise, γ_i again estimates the leverage effect.

The difference of TARCH model (Zakoian, 1994) is in the fact that it's formulated for standard deviations σ_t (6).

$$\sigma_t = c + \sum_{i=1}^{k} \kappa_i \left(|y_{t-i}| - \gamma_i y_{t-i}\right) + \sum_{j=1}^{m} \mu_j \sigma_{t-j}. \tag{6}$$

This specific form allows different reactions of the volatility to different signs of the lagged innovations y_{t-i}[1].

In order to cover the case where volatility demonstrates non-stationary behavior, we include integrated GARCH (IGARCH, (Engle and Bollerslev, 1986)) in our comparison list. This model assumes that the persistence of volatility implementing the following coefficient restriction to (3):

$$1 - \sum_{i=1}^{k} \kappa_i - \sum_{j=1}^{m} \mu_j = 0 \tag{7}$$

The estimation issues of IGARCH has been discussed in (Engle et al., 1987; Bollerslev et al., 1988).

In GARCH-type models there is only one source of uncertainty, η_t, which drives the dynamics of both returns and volatility. It seems more naturally to include another random term for volatility and state it as an autoregressive process. The subsection 2.2 describes this idea in details.

2.2 Stochastic volatility

The set up for the basic stochastic volatility model, see (Tsyplakov, 2010), is the following: (8) and (9).

$$y_t = \exp(\sigma_t/2)\eta_t, \tag{8}$$

$$\sigma_t = \delta + \phi\sigma_{t-1} + \sigma_\varepsilon \varepsilon_t, \tag{9}$$

where σ_t is the logarithm of variance, δ is its level, ϕ estimates the persistence, σ_ε is the variance of log-variance, y_t and η_t have the same meaning as before. The process σ_t is unobserved and usually interpreted as the latent time-varying volatility process. One of the main difficulties in estimating this model is the impossibility of obtaining the closed-form likelihood function. Parameters can be estimated by applying numerical methods such as Markov Chain Monte Carlo simulations.

2.3 Markov switching multifractal

In Markov Switching Multifractal model (further MSM), introduced in (Calvet et al., 1997), volatility also has its own source of uncertainty and consists of several volatility components which follow a first-order Markov process, i. e. in each moment the volatility component is equal to its previous value or is drawn from some fixed distribution with

[1]Smooth Transition GARCH, developed by (González-Rivera, 1998), can be considered as the generalization of the threshold GARCH in the sense that the former allows more than two states of volatility. We think that threshold GARCH is more appropriate to the Russian financial market because on integrated markets the prices dynamics is assumed to be smoother and on segmented markets the prices usually evolve more discontinuously due to lower level of liberalization (Bekaert and Harvey, 1997)

a probability which is unique for each volatility component. The main difficulty a researcher faces is the estimation of transition probability matrix for the Markov process (for example, if a volatility component can take only two values, then k volatility components generally needs to be parameterized by 2^{2k} variables). In MSM this problem is solved by introducing model restrictions taken from multifractal literature. Due to them the number of parameters to be estimated is only five. Moreover closed-form likelihood function and standard procedure of maximum likelihood estimation are available.

The dynamics of volatility is described in (10).

$$\sigma_t^2 = \sigma^2 \left(\sum_{k=1}^{\bar{k}} M_{k,t} \right), \tag{10}$$

where σ is a positive constant, $M_{k,t}$ are nonnegative, statistically independent volatility components, $\bar{k}$ is the number of volatility components which is considered as the order of MSM model. Due to their Markov chain nature each component can be in its previous state with probability $1 - \gamma_k$ or switch with probability γ_k, (11).

$$M_{k,t} = \begin{cases} M \text{ with probability } \gamma_k \\ M_{k,t-1} \text{ with probality } 1 - \gamma_k \end{cases}, \tag{11}$$

where $k = 1, \ldots, \bar{k}$, M should be nonnegative and have unit mathematical expectation. In the simplest case distribution of M is a sum of two Dirac delta functions $\delta(\cdot)$ (12).

$$f(M) = 0.5\delta\left(M - m_0\right) + 0.5\delta\left(M - m_1\right), m_1 = 2 - m_0, \tag{12}$$

Each component has its own switching probability γ_k defined by (13).

$$\gamma_k = 1 - \left(1 - \gamma_1\right)^{b^{k-1}}, \tag{13}$$

where $\gamma_1 \in (0,1)$, $b \in (1, \infty)$. This means that $\gamma_k < 1$ for all $k = 1, \ldots, \bar{k}$ and all γ_k are ordered as follows: $\gamma_1 < \gamma_2 < \cdots < \gamma_{\bar{k}}$. Hence component $M_{1,t}$ has the lowest switching probability and $M_{\bar{k},t}$–the highest. Components with low switching probabilities are called low-frequency components and capture the most persistence variations of volatility, while high-frequency components capture short-run dynamic of volatility. This feature distinguishes MSM from many other models where short-run and long-run variations of volatility are modeled separately.

The details of estimation of MSM model by pseudo maximum likelihood method, small sample properties of the estimator and simulation results are exhaustively discussed in (Calvet and Fisher, 2004).

3. Empirical results

We apply MLE to the GARCH and MSM models and MCMC estimator to the stochastic volatility model and obtain the preferred specifications for three financial time series—two stocks and an exchange rate.

3.1 Data description

The empirical analysis uses daily prices of Aeroflot company stocks (AFLT), Gazprom company stocks (GAZP) and exchange rate for ruble against US dollar (USD/RUB), taken from (*Yahoo! Finance* 2015). Data covers the period from 01.01.06 to 16.05.12 and includes 5153 observations.

The choice of the assets is determined by the specifics of Russian financial market, and Russian stock market in particular, which belongs to developing markets. As we wrote above, the character of asset price movement is connected with the degree, to which the financial market is integrated in the global market. We chose Gazprom and Aeroflot since they, in our opinion, can be considered as one of the assets, which are closest to the global financial market. And so they can successfully be used in GARCH and stochastic volatility modeling.

3.2 Estimation results

The parameter estimates of four GARCH-type models are presented in Tables I, II and III. The log returns conditional mean, see Equation (1), for all series is modeled as an autoregressive process of order 1. Each volatility equation includes one ARCH and one GARCH term and a term for measuring the leverage effect when it's available, see Equations (3), (4), (5), (6) and (7). μ_1 for IGARCH has no standard error since μ_1 is calculated from restriction (7).

Table I: GARCH parameter estimates for Aeroflot

	GARCH	EGARCH	GJR	TARCH	IGARCH
ω	0.0004	0.0000	0.0000	-0.0001	0.0004
	(0.0008)	(0.0002)	(0.0008)	(0.0006)	(0.0007)
α_1	0.0926^{*}	0.0990^{**}	0.1012^{**}	0.0806^{*}	0.0980^{***}
	(0.0480)	(0.0435)	(0.0454)	(0.0419)	(0.0369)
c	0.0002^{**}	-1.2932	0.0002^{**}	0.0030	0.0001^{***}
	(0.0001)	(0.7995)	(0.0001)	(0.0040)	(0.0000)
κ_1	0.4648^{***}	-0.0675	0.3306^{*}	0.2744^{**}	0.4232^{***}
	(0.1368)	(0.0532)	(0.1871)	(0.1278)	(0.0906)
μ_1	0.3582^{**}	0.8222^{***}	0.3869^{**}	0.6864^{***}	0.5768
	(0.1617)	(0.1070)	(0.1850)	(0.2515)	(.–)
γ_1		0.5189^{***}	0.2253	0.1212	
		(0.1178)	(0.1735)	(0.1375)	
LL	2543.4536	2556.3991	2545.9232	2552.2372	2539.5482
AIC	-5076.9072	-5100.7982	-5079.8464	-5092.4743	-5071.0964
BIC	-5051.9836	-5070.8899	-5049.9381	-5062.5660	-5051.1575

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

We begin by examining the conditional mean equations. The lagged log returns coefficients are significant only for Aeroflot data, meaning that Gazprom and the exchange rate exhibit negligible autocorrelation in conditional mean. On the other hand the autocorrelation in volatility tends to be strongly significant (at 5% level or less) in all cases

Table II: GARCH parameter estimates for Gazprom

	GARCH	EGARCH	GJR	TARCH	IGARCH
ω	0.0005	−0.0003	0.0000	−0.0005	0.0005
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
α_1	−0.0042	0.0007	0.0020	0.0005	−0.0048
	(0.0310)	(0.0301)	(0.0301)	(0.0298)	(0.0325)
c	0.0000**	−0.1420	0.0000***	0.0007***	0.0000***
	(0.0000)	(0.0269)	(0.0000)	(0.0002)	(0.0000)
κ_1	0.1066***	−0.0546	0.0586***	0.1129***	0.1179***
	(0.0261)	(0.0250)	(0.0185)	(0.0234)	(0.0185)
μ_1	0.8784***	0.9794***	0.8719***	0.8902***	0.8821
	(0.0189)	(0.0037)	(0.0264)	(0.0218)	(.−)
γ_1		0.2072***	0.0965**	0.2876**	
		(0.0377)	(0.0480)	(0.1276)	
LL	2410.1991	2407.5094	2416.5774	2408.8917	2408.6767
AIC	−4810.3982	−4803.0188	−4821.1548	−4805.7833	−4809.3534
BIC	−4785.5164	−4773.1607	−4791.2967	−4775.9252	−4789.4480

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table III: GARCH parameter estimates for USD/RUB

	GARCH	EGARCH	GJR	TARCH	IGARCH
ω	−0.0003	−0.0002	−0.0002	−0.0002	−0.0003***
	(0.0004)	(0.0001)	(0.0032)	(0.0000)	(0.0001)
α_1	0.0408	0.0371	0.0441	0.0367	0.0405
	(0.0640)	(0.0368)	(0.1839)	(0.0298)	(0.0288)
c	0.0000	−0.0473	0.0000	0.0000	0.0000***
	(0.0000)	(0.0076)	(0.0001)	(0.0000)	(0.0000)
κ_1	0.0760	0.0648***	0.0974	0.0616***	0.0772***
	(0.4479)	(0.0187)	(2.1944)	(0.0134)	(0.0090)
μ_1	0.9229**	0.9944***	0.9367	0.9498***	0.9228
	(0.4056)	(0.0006)	(1.8292)	(0.0133)	(.−)
γ_1		0.1364***	−0.0729	−0.5741	
		(0.0214)	(0.1725)	(0.1852)	
LL	6097.4303	6106.4104	6109.6509	6101.5368	6097.6928
AIC	−12184.8606	−12200.8207	−12207.3018	−12191.0736	−12187.3856
BIC	−12158.2978	−12168.9454	−12175.4264	−12159.1983	−12166.1354

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table IV: Stochastic volatility parameter estimates for three assets

	AFLT	GAZP	USD/RUB
δ	-8.2330	-7.9188	-11.0047
	(0.1082)	(0.4202)	(0.2241)
ϕ	0.8132^{***}	0.9818^{***}	0.9669^{***}
	(0.0304)	(0.0076)	(0.0124)
σ_ε	0.7278^{***}	0.1906^{***}	0.2981^{***}
	(0.0626)	(0.0276)	(0.0526)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

except GJR model for USD/RUB. We also observe substantial leverage effect, estimated by EGARCH model, in all three series. According to Schwarz information criterion reported in the last row of Tables I–III, GJR model provides the highest goodness of fit for Gazprom stocks and USD/RUB exchange rate, EGARCH—for Aeroflot stocks.

The parameter estimates for the stochastic volatility model in Equation (9) are presented in Table IV. We use MCMC sampler described in detail in (Kastner and Frühwirth-Schnatter, 2014). Following the recommendations in (Frühwirth-Schnatter and Wagner, 2010) we define the prior distribution of δ as a Gaussian distribution with mean equals to -10 and variance equals to 1. The prior for persistent coefficient ϕ is beta distribution with parameters 20 and 1.1.

The latent volatility processes for the three series appear to be highly persistent due to the ϕ coefficient is significant and close to one. The larger ϕ the lower σ_ε is, meaning that near unit root volatility process has lower unconditional variance in the estimated model.

The estimation is held for eight different specifications of MSM($\bar{k}$) model with $\bar{k}$ ranges from 1 to 8. Interestingly, if $\bar{k} = 1$ one obtains a usual Markov model, where volatility takes only two values: m_0 and $2 - m_0$. When $\bar{k} > 1$ number of volatility states grows as $2^{\bar{k}}$ and reaches 256 in our calculations.

Table V contains the parameter estimates of MSM($\bar{k}$) model and the value of log likelihood function, labeled as "LL".

We begin by examining the Aeroflot data. The multiplier parameter m_0 tends to decline with $\bar{k}$ (with some exceptions) because when the number of volatility components increases, they are able to capture the fluctuations in volatility without much variability in themselves. The estimates of σ vary across $\bar{k}$ with no particular pattern. As the switching probability γ_{bark} is concerned its reciprocal characterizes the average length of the shortest volatility cycle. When $\bar{k} = 1$ the only $M_{t,1}$ has a duration of approximately two months. As $\bar{k}$ increases γ_{bark} tends to grow until the shortest volatility cycle declines to about a day and a half. The frequency parameter b increases with $\bar{k}$ but not monotonically, implied that the spacing between switching probabilities becomes larger, when number of volatility components grows. The other assets generate parameters with similar behavior. m_0 tends to decrease with $\bar{k}$ in all cases, the magnitude of m_0 varies in approximately the same range for all assets.

For stock log returns the log likelihood function reaches its maximum if $\bar{k} = 4$. In case of currency rate the log likelihood function tends to grow with $\bar{k}$, what is compatible

Table V: MSM parameter estimates

$\bar{k}$	1	2	3	4	5	6	7	8
				AFLT				
b	2.4990	1.0100	4.2670	4.8864	9.7742	5.7669	5.7971	5.8032
$\gamma_{\bar{k}}$	0.0164	0.1330	0.5683	0.6457	0.3381	0.7070	0.7093	0.7098
m_0	1.9994	1.7811	1.6553	1.6173	1.7119	1.6205	1.6205	1.6205
σ	0.8760	0.0304	0.0250	0.0332	0.1546	0.0891	0.1448	0.2351
LL	3719.38	3965.91	3973.23	3978.79	3971.78	3976.35	3975.64	3974.94
				GAZP				
b	2.5046	2.6766	2.9058	4.3465	3.5778	4.6400	4.3450	3.9919
$\gamma_{\bar{k}}$	0.0143	0.0225	0.0458	0.1349	0.0479	0.0612	0.0579	0.1545
m_0	1.9993	1.6481	1.6124	1.4933	1.6142	1.6153	1.6155	1.4618
σ	0.8768	0.0469	0.0378	0.0409	0.0986	0.1622	0.1274	0.0195
LL	3484.92	3708.28	3725.70	3729.22	3723.39	3722.79	3723.66	3727.51
				USD/RUB				
b	2.5093	1.0100	8.4035	3.5970	4.0528	3.7041	2.5883	4.3180
$\gamma_{\bar{k}}$	0.0131	0.3030	0.7893	0.8763	0.6887	0.9900	0.9557	0.9900
m_0	2.0000	1.8039	1.7021	1.6379	1.7258	1.7009	1.4798	1.5031
σ	0.8763	0.0062	0.0064	0.0058	0.0140	0.0068	0.0054	0.0046
LL	7638.10	8019.33	8102.25	8097.64	8084.44	8092.40	8117.16	8125.73

with the results in (Calvet and Fisher, 2004).

After estimating MSM models of different orders we need to choose one of them for each asset. The usual way to compare models in-sample is comparing them by information criteria, but it's correct if considerated models are nested. MSM with different $\bar{k}$ are non-nested, therefore we implement model selection procedure presented in (Vuong, 1989). Vuong test for model selection resulted in final sets of parameters for all three financial assets.

The null hypothesis of Vuong test is in the fact that two non-nested models fit the data equally well.

We take MSM($\bar{k}$) with the highest log likelihood as the alternative. Table VI represents the results of model selection procedure.

For AFLT stocks the null hypothesis is rejected on 5% level for models with $\bar{k} = 1, 2, 6, 7$ and 8. It means that MSM(4) with the highest log likelihood outperforms the above mentioned specifications. The situation with GAZP is slightly different. Only MSM with $\bar{k} = 1$ and 2 reveal poorer performance than the best for GAZP MSM(4) model. As for USD/RUB the hypothesis of equal goodness of fit is rejected for all $\bar{k}$, what leads us to the fact that the more volatility components are in the model specification for currency rate, the better this model explains the data.

Willing to decrease the computational costs we choose for each asset the model with minimum possible order which performs at least not worse than the model with maximum value of likelihood function. To sum up, the selected specifications are MSM(3) for Aeroflot and Gazprom and MSM(8) for the exchange rate.

Table VI: Results of Vuong test

$\bar{k}$	1	2	3	4	5	6	7	8
				AFLT				
V	−6.5241	−0.3239	−0.1396	–	−0.1762	−0.0612	−0.0792	−0.0967
p-value	0.0000	0.0499	0.1656	–	0.0542	0.0318	0.0201	0.0117
				GAZP				
V	−6.1615	−0.5281	−0.0889	–	−0.1471	−0.1621	−0.1401	−0.0432
p-value	0.0000	0.0023	0.2380	–	0.1105	0.0711	0.1191	0.3126
				USD/RUB				
V	−10.9038	−2.3792	−0.5251	−0.6282	−0.9233	−0.7452	−0.1916	–
p-value	0.0000	0.0000	0.0045	0.0014	0.0000	0.0030	0.0371	–

4. Models comparison

In the previous section we estimate four GARCH models, a stochastic volatility model and pick out MSM model specifications by Vuong test. We can not run likelihood ratio or similar tests to implement in-sample comparison, because the considered models are non-nested. But it's possible to use probability integral transform (PIT) in this case, (Swanepoel and Van Graan, 2002).

4.1 In-sample analysis

PIT is based on a simple idea that If U_0 is a uniformly distributed random variable (with values in [0,1]), then the random variable $X = F^{-1}(U)$ has the cumulative distribution function F. Vice versa if X has the cumulative distribution function F, random variable $F(X)$ is uniformly distributed on the interval [0,1]: $F(X) \sim U(0,1)$.

A common assumption in all the models in our paper is that the log returns are distributed normally with zero mean and some time-dependent variance. According to PIT applying Gaussian cumulative distribution function with zero mean and estimated variance to the log returns we should obtain the uniformly distributed random variable. Using Kolmogorov-Smirnov test we check how close the obtained random variable is to the uniform distribution. Table VII presents the results.

The first column contains the name of the volatility model (the first four are GARCH models, the last two—stochastic volatility models). The second and the third columns present the Kolmogorov-Smirnov test statistics and its p-value respectively. Evidently, GARCH models fail to fit the assumption of normally distributed log returns. On the other hand, stochastic volatility models exhibit much better performance according to KS statistics and the null is not rejected for the log returns of stocks in case of basic stochastic volatility model.

4.2 Out-of-sample analysis

We now investigate the out-of-sample performance of the competing models over 1-day forecasting horizon. For each asset we estimate six models and leave 500 observations (or about one third of the sample) for out-of-sample comparison. The comparison is

Table VII: Kolmogorov-Smirnov test for probability integral transform

	stand	exp	gjr	thresh	integr	msm	stochvol
				AFLT			
KS	0.9910	0.9878	0.9904	0.9857	0.0930	0.1216	0.0320
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2177
				GAZP			
KS	0.9856	0.9795	0.9817	0.9790	0.0492	0.0970	0.0250
p-value	0.0000	0.0000	0.0000	0.0000	0.0111	0.0000	0.5154
				USD/RUB			
KS	0.9637	0.9631	0.9642	0.9877	0.0777	0.0898	0.0889
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

held in two ways. The first one uses such classical forecast performance measures of log returns and the second one employs Mincer-Zarnowitz regression in order to evaluate the volatility forecast accuracy directly.

Out-of-sample comparison is conducted using such measures as mean squared error (MSE), mean absolute error (MAE) and directional accuracy (DA) of log returns. The latter is calculated as percent of cases when the signs of real and predicted log returns match: $\mathrm{DA} = {}^1\!/_T \sum_{t=1}^{T} I(\mathrm{sgn}(y_t) = \mathrm{sgn}(\widehat{y}_t))$. Evidently, better models reveal lower MSE and MAD and higher DA.

Table VIII: Comparing forecast accuracy

	stand	exp	gjr	thresh	integr	msm	stochvol
				AFLT			
MSE	0.6631	0.6159	0.6515	0.6225	0.6864	0.3167	0.3311
MAD	0.5005	0.4882	0.4977	0.4838	0.5061	0.4009	0.2772
DAC	0.3106	0.3186	0.3267	0.3166	0.3327	0.9238	0.7555
				GAZP			
MSE	0.5950	0.5748	0.5752	0.5720	0.6045	0.5463	0.5039
MAD	0.4344	0.4303	0.4329	0.4304	0.4437	0.5787	0.3541
DAC	0.2966	0.2926	0.3287	0.2946	0.2966	0.8116	0.5912
				USD/RUB			
MSE	0.0062	0.0061	0.0061	0.0061	0.0062	0.0046	0.0049
MAD	0.0421	0.0432	0.0423	0.0419	0.0423	0.0262	0.0354
DAC	0.3046	0.3046	0.3146	0.3206	0.3046	0.9038	0.6192

According to Table VIII MSE is substantially lower for stochastic volatility than for GARCH models. MAD shows similar results except for Gazprom stock returns forecast where MSM demonstrates higher MAD than GARCH models. As for directional accuracy GARCH models match the actual sign of returns in about 30% of cases, what is essentially lower than approximately 65% and 87% for original stochastic volatility and MSM correspondingly.

The idea of Mincer-Zarnowitz regression is pretty simple: using ordinary least squares we estimate the linear projection of squared log returns on the constant and one-day forecasts, Equation (14).

$$y_t^2 = \alpha + \beta\sigma_t^2 \tag{14}$$

Here squared log returns y_t^2 are proxies for volatility. Unbiased forecasts yield $\alpha = 0$ and $\beta = 1$. Standard errors of α and β are corrected by HAC variance estimator (Newey and West, 1987).

Table IX: Mincer-Zarnowitz regression

	stand	exp	gjr	thresh	integr	msm	stochvol
				AFLT			
α	0.0002***	0.0001*	0.0002***	0.0001**	0.0002***	−0.0006***	−0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
β	0.3582***	0.4931***	0.3988***	0.4493***	0.3367***	1.4419***	1.9981***
	(0.0844)	(0.0944)	(0.0817)	(0.1011)	(0.0760)	(0.0472)	(0.0742)
R^2	0.0349	0.0520	0.0456	0.0381	0.0379	0.6517	0.5926
R^2_{adj}	0.0330	0.0501	0.0437	0.0362	0.0360	0.6510	0.5918
				GAZP			
α	0.0001	0.0000	0.0000	0.0000	0.0001**	−0.0008***	−0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
β	0.5762***	0.7294***	0.7420***	0.7803***	0.5102***	1.5348***	1.7679***
	(0.1429)	(0.1325)	(0.1353)	(0.1391)	(0.1265)	(0.0817)	(0.1607)
R^2	0.0316	0.0574	0.0570	0.0595	0.0316	0.4149	0.1954
R^2_{adj}	0.0297	0.0555	0.0551	0.0576	0.0297	0.4137	0.1938
				USD/RUB			
α	0.0000**	0.0000	0.0000**	0.0000	0.0000**	−0.0001***	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
β	0.6267***	0.6692***	0.6436***	0.7196***	0.6176***	4.3389***	1.9082***
	(0.1448)	(0.1449)	(0.1341)	(0.1569)	(0.1427)	(0.1198)	(0.1311)
R^2	0.0362	0.0411	0.0442	0.0405	0.0363	0.7248	0.2985
R^2_{adj}	0.0343	0.0392	0.0423	0.0386	0.0343	0.7242	0.2971

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table IX reports the results. For all assets α is either statistically insignificant or very close to zero, what provides evidence of unbiased forecasts. On the other hand, the slope coefficient β is significant in all cases and does not equal to one. Interestingly, that GARCH models tend to overestimate volatility due to $\beta < 1$ for all GARCH specifications.

Stochastic volatility models on the contrary give understated forecasts. Consequently it could be more appropriate to use GARCH when one wants to estimate the upper bound of tomorrow volatility and stochastic volatility otherwise.

5. Conclusion

The article proposes the thorough investigation of in-sample and out-of-sample performance of four GARCH and two stochastic volatility models. We apply maximum likelihood method and Markov Chain Monte Carlo simulation to estimate the parameters and obtain one-day forecasts of ordinary GARCH, exponential GARCH, Glosten-Jagannathan-Rünkle model, threshold ARCH, Markov switching multifractal and the stochastic volatility models. Using probability integral transform, traditional forecast performance measures (MSE, MAD and DAC) and Mincer-Zarnowitz regression we compare the above mentioned models and come to the conclusion that in most cases stochastic volatility models outperform GARCH both in explanation and prediction aspects. One of the most important results is that original stochastic volatility model is the only model where the log returns normality assumption is not rejected. We also demonstrate that GARCH models tend to overestimate the volatility forecasts in contrast to stochastic volatility, where the forecasts are understated. The research perspectives could imply involving multi-step forecasts in evaluating the out-of-sample performance of volatility models and expanding the set of assets under consideration.

References

Bekaert, G. and Harvey, C. R. (1997) “Emerging equity market volatility” *Journal of Financial economics* **43.1**, 29–77.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988) “A capital asset pricing model with time-varying covariances” *The Journal of Political Economy* **96.1**, 116–131.

Calvet, L. E. and Fisher, A. J. (2004) “How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes” *Journal of Financial Econometrics* **2.1**, 49–83.

Calvet, L. E., Fisher, A. J., and Mandelbrot, B. B. (1997) “Large deviations and the distribution of price changes” *Cowles Foundation Discussion Paper* **1165**.

Chuang, W.-I., Huang, T., and Lin, B.-H. (2013) “Predicting volatility using the Markov-switching multifractal model: Evidence from S&P 100 index and equity options” *The North American Journal of Economics and Finance* **25**, 168–187.

Danielsson, J. (1994) “Stochastic volatility in asset prices estimation with simulated maximum likelihood” *Journal of Econometrics* **64.1**, 375–400.

Engle, R. F. and Bollerslev, T. (1986) “Modelling the persistence of conditional variances” *Econometric reviews* **5.1**, 1–50.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987) “Estimating time varying risk premia in the term structure: The ARCH-M model” *Econometrica: Journal of the Econometric Society* **55.2**, 391–407.

Frühwirth-Schnatter, S. and Wagner, H. (2010) “Stochastic model specification search for Gaussian and partial non-Gaussian state space models” *Journal of Econometrics* **154.1**, 85–100.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993) “On the relation between the expected value and the volatility of the nominal excess return on stocks” *The journal of finance* **48**.**5**, 1779–1801.

González-Rivera, G. et al. (1998) “Smooth transition GARCH models” *Studies in nonlinear dynamics and econometrics* **3**.**2**, 61–78.

Hansen, P. R. and Lunde, A. (2005) “A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?” *Journal of Applied Econometrics* **20**.**7**, 873–889.

Kastner, G. and Frühwirth-Schnatter, S. (2014) “Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models” *Computational Statistics & Data Analysis* **76**, 408–423.

Nelson, D. B. (1991) “Conditional Heteroskedasticity in Asset Returns: A New Approach” *Econometrica* **59**.**2**, 347.

Newey, W. K. and West, K. D. (1987) “Hypothesis testing with efficient method of moments estimation” *International Economic Review* **28**.**3**, 777–787.

Shephard, N. (1996) “Statistical aspects of ARCH and stochastic volatility” *Time Series Models in Econometrics, Finance and Other Fields* ed. by D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen Springer-Science and Business Media, B.Y. chap. 1, 1–67.

Swanepoel, J. W. and Van Graan, F. C. (2002) “Goodness-of-fit tests based on estimated expectations of probability integral transformed order statistics” *Annals of the Institute of Statistical Mathematics* **54**.**3**, 531–542.

Tsyplakov, A. (2010) “Revealing the arcane: an introduction to the art of stochastic volatility models” *Quantile* **8**, 69–122.

Vuong, Q. H. (1989) “Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses” *Econometrica* **57**.**2**, 307–333.

Yahoo! Finance (2015) URL: `http://finance.yahoo.com` (visited on 01/22/2014).

Zakoian, J.-M. (1994) “Threshold heteroskedastic models” *Journal of Economic Dynamics and control* **18**.**5**, 931–955.