NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Valeria V. Lakshina

THE FLUKE OF STOCHASTIC VOLATILITY VERSUS GARCH INEVITABILITY : WHICH MODEL CREATES BETTER FORECASTS?

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: Financial Economics
WP BRP 37/FE/2014

Valeria V. Lakshina[1]

THE FLUKE OF STOCHASTIC VOLATILITY VERSUS GARCH INEVITABILITY: WHICH MODEL CREATES BETTER FORECASTS?[2]

Abstract

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

JEL Classification: C01, C58, C51, G17.

Keywords: GARCH, stochastic volatility, markov switching multifractal, forecast performance.

---

[1] National Research University Higher School of Economics. Department of Mathematical Economics. Lecturer. E-mail: vlakshina@hse.ru. Nizhniy Novgorod, Russia.

[2] The author is grateful to Andrey M. Silaev (Higher School of Economics) for his careful guidance and Henry I. Penikas (Higher School of Economics) for his useful comments, which were received at the Second Russian Economic Congress, 2013.

1 Introduction

Predicting the 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 on which volatility measures predict the future volatility more efficiently in different scenarios. GARCH models are usually compared with implied and historical volatility in option pricing, but as yet no consensus has been reached. For example, (Lamoureux and Lastrapes, 1993) argue that implied volatility outperforms historical volatility in predicting the future volatility for a option's underlying assets. On the other hand (Ammann et al., 2009) came to the opposite conclusion after taking into account such factors as capitalization, beta, market-to-book ratio and price momentum. One possible explanation for these contradictory results is that the models considered do not capture some essential stylized facts of the financial time series, due to the absence of random term in the volatility process.

Another fundamentally different way of using volatility modeling is developed in stochastic volatility models (SV). The main difference between them and the 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, since the closed-form solution rarely exists. Examples of comparisons of GARCH with SV models can be found in (Danielsson, 1994) and (Shephard, 1996). See also (Kim et al., 1998) for the thorough simulation-based investigation of the SV model, including estimation, filtering, hypothesis testing and prediction.

In Chuang et al's (2013) more recent study, the aforementioned volatility measures (aside from SV) are compared with the Markov switching multifractal model (MSM), which was introduced in Calvet and Fisher, 2004) Unlike GARCH or implied volatility, MSM's multifractal structure 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 does outperform the implied volatility in the out-of-sample performance.

It is noteworthy that,as in the SV model, MSM also incorporates uncertainty into the volatility process but in a completely different way than SV (for details see Section 2). This meant that MSM, belonging to the class of stochastic volatility, has a closed-form likelihood function and can be estimated via the usual optimization procedure.

Consequently, this paper aims to compare two stochastic volatility models which both model the volatility via a random process but which substantially differ in terms of their computational efforts. However, GARCH models are used as traditional benchmarks in volatility estimations 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 discusses the results. Section 5 concludes the paper.

2 Model description

Volatility models can be divided into three main groups: autoregressive conditional heteroskedasticity (ARCH or GARCH which is generalized ARCH (Bollerslev, 1986)), stochastic volatility and realized volatility. The latter is usually used for analyzing intra-day data, therefore in our research we have focused on GARCH and stochastic volatility models.

From more than three hundred ARCH-type models (Hansen and Lunde, 2005) we chose four: the ordinary GARCH, exponential GARCH, Glosten-Jagannathan-Rünkle model (GJR) and threshold ARCH. For all these models, we estimated the simple specification with one ARCH and one GARCH term. The choice of the models is induced by the prevalence of this specification in the financial literature, particularly in the results' practical applicability. Examples can be found in (Ding and Meade, 2010), (Pederzoli, 2006).

We also chose two models from the second group: the original stochastic volatility model (or, as termed in the literature---the stochastic volatility) and the Markov switching multifractal model.

2.1 GARCH

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

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

where $E(x_t \mid \mathcal{F}_{t-1})$ is a conditional mean of daily returns x_t at time t, which is 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's price divided by yesterday's price: $x_t = log(\frac{P_t}{P_{t-1}})$. Conditional mean $E(x_t \mid \mathcal{F}_{t-1})$ is modelled by ARMA(p,q), see (2).

$$E(x_t \mid \mathcal{F}_{t-1}) = \omega + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j}, \tag{2}$$

where parameters α_i and β_j are the i th-order autoregressive (AR) and j th-order moving average (MA) terms.

Consequently, innovations y_t have zero mean and time-dependent variance σ_t^2, which is modeled by (3) and (4).

$$y_t = \sigma_t \eta_t, \;\; \eta_t \sim N(0,1) \tag{3}$$

$$\sigma_t^2 = c + \sum_{i=1}^{k} \kappa_i y_{t-i}^2 + \sum_{j=1}^{m} \mu_j \sigma_{t-j}^2, \tag{4}$$

where parameter κ_i represents i th-order ARCH term, μ_j ---the j th-order GARCH term, η_t are standardized innovations or standardized residuals, which are normally distributed with zero mean and unit variance. ARCH term in (4) allows us to capture the effects of volatility clustering and the GARCH term is responsible for the volatility autocorrelation estimated by μ_j.

The exponential GARCH (Nelson, 1991) also allows us to capture the leverage effect (i. e. the asymmetric volatility response to negative and positive returns) and ensure the simpler evaluation of shock persistence (5).

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

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 $I(\cdot)$ (6).

$$\sigma_t^2 = c + \sum_{i=1}^{k} \left(\kappa_i y_{t-i}^2 + \gamma_i I(y_{t-i}) y_{t-i}^2 \right) + \sum_{j=1}^{m} \mu_j \sigma_{t-j}^2, \tag{6}$$

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

Unlike the others, the TARCH model (Zakoian, 1994) is formulated for standard deviations σ_t (7).

$$\sigma_t = c + \sum_{i=1}^{k} \kappa_i \left(| y_{t-i} | - \gamma_i y_{t-i} \right) + \sum_{j=1}^{m} \mu_j \sigma_{t-j}. \tag{7}$$

This specific form allows to us observe the volatility's different reactions to different signs of the lagged innovations y_{t-i}.

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 natural to include another random term for volatility and state it as an autoregressive process. The next subsection describes this idea in detail.

2.2 Stochastic volatility

The set up for the basic stochastic volatility model (Taylor, 1994) (Tsyplakov 2007), is as follows: (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 the Markov Chain Monte Carlo simulations. For further details on MCMC see Brooks et al., (2011) and Jeliazkov and Lee (2010).

2.3 Markov switching multifractal model

In the 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. This means that, in each moment of time, the volatility component is equal to its previous value, or is drawn from a certain fixed distribution with a probability unique to each volatility component. The main difficulty that a researcher is likely to encounter is the estimation of the transition probability matrix for the Markov process, e.g. 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 the literature on multifractal models, which clarify that only five parameters need to be estimated. Moreover, the closed-form likelihood function exists and standard procedure of the maximum likelihood estimation may be carried out.

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 non-negative, statistically independent volatility components, $\bar{k}$ is the number of volatility components which is considered as the order of the 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).

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

where $k = 1,\ldots,\bar{k}$, M should be non-negative and have a unit of mathematical expectation. In the simplest case, the 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).

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

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: $\gamma_1 < \gamma_2 < \ldots < \gamma_{\bar{k}}$. Hence component $M_{1,t}$ has the lowest switching probability and $M_{\bar{k},t}$--- has the highest. Components with low switching probabilities are called low-frequency components and capture the most persistent variations of volatility, while high-frequency components capture the volatility's short-run dynamic. This feature distinguishes MSM from many other models where short-run and long-run variations of volatility are modeled separately. The inversed switching probability shows an average period of time, when the volatility component does not change its value and can be interpreted as a measure of the persistence of corresponding volatility shock.

Therefore, to identify a MSM process five parameters are needed:

- m_0 for distribution of volatility components $M_{k,t}$, $m_0 \in (1,2)$;
- scaling parameter σ, $\sigma \in (1,\infty)$;
- b for switching probabilities, $b \in (1,\infty)$;
- switching probability of the most high-frequency component $\gamma_{\bar{k}}$ (it is more convenient to estimate $\gamma_{\bar{k}}$ rather than γ_1 because $\gamma_{\bar{k}}$ has the same magnitude as the other parameters), $\gamma_{\bar{k}} \in (0,1)$;

• number of volatility components $\bar{k}$, $\bar{k} \in (1, \infty)$.

The details of the MSM model's estimation by the pseudo maximum likelihood method, the estimator's small sample properties and simulation results are discussed in (Calvet and Fisher, 2004) and (Calvet and Fisher, 2008).

3 Empirical results

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

3.1 Data description

The empirical analysis uses the daily prices of Aeroflot's company stocks (AFLT), Gazprom's company stocks (GAZP) and the rouble exchange rate against the US dollar (USD/RUB). The data covers the period from 01.01.06 to 16.05.12 and totals 5153 observations. Figure 1 shows the daily logarithmic returns of the three series. For each series, we calculated the mean, standard deviation, skewness and kurtosis coefficients and the minimum and maximum values, as shown in Table 1.

Table 1: Descriptive statistics for three log returns

Asset	Sample size	Mean	Standard deviation	Skewness	Kurtosis	Max	Min
AFLT	1581	-0.00010	0.0242	0.6067	12.7246	0.2104	-0.1662
GAZP	1572	-0.00027	0.0294	-0.0327	15.8053	0.2526	-0.2236
USD/RUB	2000	0.00004	0.0057	1.0285	22.2542	0.0639	-0.0567

Table 1 reveals that all three series are highly leptokurtic, as a financial time series ought to be. Moreover, in contrast to GAZP, which has a longer left tail, AFLT and USD/RUB are positively skewed.

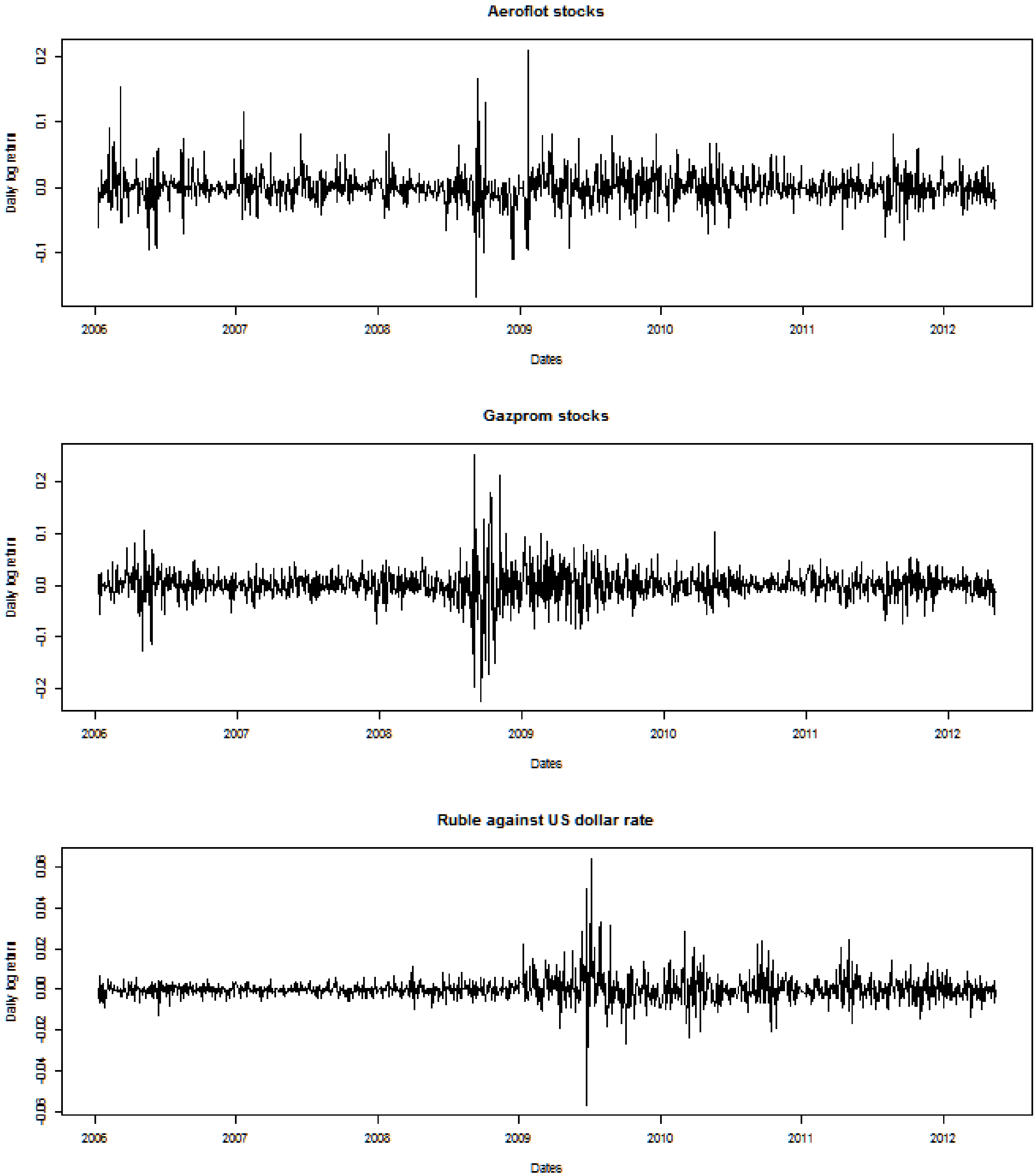

Fig. 1: Logarithmic returns of Aeroflot and Gazprom stock prices and USD/RUB exchange rate.

3.2 Estimation results

GARCH The parameter estimates of the four GARCH-type models are presented in Tables 2, 3 and 4. The log returns conditional mean, see Equation (2), 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 is available (see Equations **Ошибка! Источник ссылки не найден.**, (5), (6) and (7)).

Table 2: GARCH parameter estimates for Aeroflot

	GARCH	EGARCH	GJR	TARCH
ω	0.0004	0.0000	0.0000	-0.0001
	(0.0008)	(0.0002)	(0.0008)	(0.0006)
α_1	0.0926*	0.0990**	0.1012**	0.0806*
	(0.0480)	(0.0435)	(0.0454)	(0.0419)
c	0.0002**	-1.2932	0.0002**	0.0030
	(0.0001)	(0.7995)	(0.0001)	(0.0040)
κ_1	0.4648***	-0.0675	0.3306*	0.2744**
	(0.1368)	(0.0532)	(0.1871)	(0.1278)
μ_1	0.3582**	0.8222***	0.3869**	0.6864***
	(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
AIC	-5076.9072	-5100.7982	-5079.8464	-5092.4743
BIC	-5051.9836	-5070.8899	-5049.9381	-5062.5660
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$				

Table 3: GARCH parameter estimates for Gazprom

	GARCH	EGARCH	GJR	TARCH
ω	0.0005	-0.0003	0.0000	-0.0005
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
α_1	-0.0042	0.0007	0.0020	0.0005
	(0.0310)	(0.0301)	(0.0301)	(0.0298)
c	0.0000**	-0.1420	0.0000***	0.0007***
	(0.0000)	(0.0269)	(0.0000)	(0.0002)
κ_1	0.1066***	-0.0546	0.0586***	0.1129***
	(0.0261)	(0.0250)	(0.0185)	(0.0234)
μ_1	0.8784***	0.9794***	0.8719***	0.8902***
	(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
AIC	-4810.3982	-4803.0188	-4821.1548	-4805.7833
BIC	-4785.5164	-4773.1607	-4791.2967	-4775.9252

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

Table 4: GARCH parameter estimates for USD/RUB

	GARCH	EGARCH	GJR	TARCH
ω	-0.0003	-0.0002	-0.0002	-0.0002
	(0.0004)	(0.0001)	(0.0032)	(0.0000)
α_1	0.0408	0.0371	0.0441	0.0367
	(0.0640)	(0.0368)	(0.1839)	(0.0298)
c	0.0000	-0.0473	0.0000	0.0000
	(0.0000)	(0.0076)	(0.0001)	(0.0000)
κ_1	0.0760	0.0648***	0.0974	0.0616***
	(0.4479)	(0.0187)	(2.1944)	(0.0134)
μ_1	0.9229**	0.9944***	0.9367	0.9498***
	(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
AIC	-12184.8606	-12200.8207	-12207.3018	-12191.0736
BIC	-12158.2978	-12168.9454	-12175.4264	-12159.1983
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$				

We begin by examining the conditional mean equations. On the one hand, the lagged log returns coefficients are significant only for the Aeroflot data, meaning that Gazprom and the exchange rate exhibit a negligible autocorrelation in the conditional mean. On the other hand, the autocorrelation in volatility tends to be strongly significant (at 5% level or less) in all cases except the GJR model for USD/RUB. We also observed a substantial leverage effect in all three series, as estimated by the EGARCH model. According to the Schwarz information criterion reported in the last row of Tables 2--4, the GJR model provides the highest goodness of fit for Gazprom's stocks and USD/RUB exchange rate, and EGARCH provides the best fit for Aeroflot's stocks.

Stochastic volatility The parameter estimates for the stochastic volatility model in Equation (9) are presented in Table 5. We use the MCMC sampler, which is described in detail in Kastner and Friihwirth-Schnatter (2014). Following recommendations which were outlined in Friihwirth-Schnatter and Wagner (2010), and Kim et al. (1998), we define the prior distribution of δ as a Gaussian distribution, with mean equal to -10 and variance equal to 1. The prior for persistence coefficient ϕ is a beta distribution with parameters 20 and 1.1.

Table 5: 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)
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$			

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

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

Table 6 contains the parameter estimates of the MSM($\bar{k}$) model and the value of the log likelihood function, labeled as “LL”.

Table 6: Parameter estimates for the MSM model

$\bar{k}$	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

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. The switching probability's $\gamma_{\bar{k}}$ 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, $\gamma_{\bar{k}}$ tends to increase until the shortest volatility cycle declines to about a day and a half. The frequency parameter b increases with $\bar{k}$ but not monotonically, implying that the distance between switching probabilities becomes greater as the number of volatility components grows. The other assets generate parameters with similar behavior. m_0 tends to decrease with $\bar{k}$ in all cases, and 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 $\bar{k} = 4$. For the currency rate, the log likelihood function tends to increase with $\bar{k}$, which is compatible with the results found by Calvet and Fisher (2004).

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

The null hypothesis of Vuong's test is that two non-nested models, say F and G, fit the data equally well. The alternative implies that G outperforms F, so we concern ourselves with the left-sided criteria. The essence of Vuong's test is that, to compare the goodness-of-fit of two models, one should check that the difference between their log likelihood functions is significant. Based on the assumption of data normality, the following equation can be used:

$$V = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} ln \frac{f(x_t \mid x_{t-1}, \ldots, x_1)}{g(x_t \mid x_{t-1}, \ldots, x_1)} \sim N\left(0, \sigma_*^2\right), \quad (14)$$

where $f(\cdot)$ and $g(\cdot)$ are probability densities from models F and G respectively. The statistics are normally distributed, with zero mean and variance σ_*^2. The consistent estimator for σ_*^2 can be obtained as sample variance of terms in (14).

We take the MSM($\bar{k}$) with the highest log likelihood as the alternative (G model). Table 7 represents the results of the model selection procedure.

Table 7: Results of Vuong's 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	-

For AFLT's stocks, the null hypothesis is rejected at a 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. GAZP's results are slightly different. Only the MSM with $\bar{k} = 1$ and 2 reveal a poorer performance than the best performances from the GAZP MSM(4) model. In the case of the USD/RUB, the hypothesis of equal goodness of fit is rejected for all $\bar{k}$, which leads us to conclude that the more volatility components there are in the model specification for currency rate, the better this model can explain the data.

Hoping to decrease the computational costs, we choose the model with the minimum possible order for each asset, which does not perform any worse than the model with the maximum value of likelihood function. Therefore, the selected specifications are MSM(3) for Aeroflot and Gazprom and MSM(8) for the exchange rate.

4 Model comparison

4.1 In-sample analysis

In the previous section, we estimated four GARCH models, a stochastic volatility model and pick out MSM model specifications by Vuong's test. We cannot run the likelihood ratio test or similar tests to implement in-sample comparison, because these models are non-nested. It is possible however to use a probability integral transform (PIT), as used by Swanepoel and Van Graan (2002).

PIT is based on the simple idea that if U_0 is a uniformly distributed random variable (with values of [0,1]), then the random variable $X = F^{-1}(U)$ has the cumulative distribution function F. Conversely, 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 normally distributed, with zero mean and some time-dependent variance. According to PIT, applying the Gaussian cumulative distribution function with a zero mean and estimated variance to the log returns, we should obtain the uniformly distributed random variable. Using the Kolmogorov-Smirnov test, we checked how close the obtained random variable is to the uniform distribution. Table 8 presents the results.

Table 8: Kolmogorov-Smirnov test for probability integral transform

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

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

4.2 Out-of-sample analysis

We now investigate the competing models' out-of-sample performance over a 1-day forecasting horizon. For each asset, we estimate six models and leave 500 observations (or about one third of the sample) for an out-of-sample comparison. The comparison is conducted in two ways. The first one uses classical forecast performance measures of log returns and the second uses the Mincer-Zarnowitz regression, in order to directly evaluate the volatility forecast's accuracy.

4.2.1 Forecast performance measures

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

Table 9: Comparing forecast accuracy

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

According to Table 9, MSE is substantially lower for stochastic volatility than for the GARCH models. MAD shows similar results, aside from Gazprom's stock returns forecast, where MSM demonstrates higher MAD than in the GARCH models. For directional accuracy, the GARCH models match the actual sign of returns in about 30% of cases, which is essentially lower than approximately 65% and 87% for the original stochastic volatility and MSM respectively.

4.2.2 Mincer-Zarnowitz regression

The idea of Mincer-Zarnowitz regression is simple; using ordinary least squares we estimate the linear projection of squared log returns on the constant and one-day forecasts, shown in Equation (15).

$$y_t^2 = \alpha + \beta\sigma_t^2 \tag{15}$$

Here, squared log returns y_t^2 are proxies for volatility. Unbiased forecasts yield $\alpha = 0$ and $\beta = 1$. As suggested in West and McCracken, (1998) standard errors of α and β are corrected by the HAC variance estimator (Newey and West, 1987).

Table 10 reports the results. For all assets, α is either statistically insignificant or approaching zero, which provides evidence of unbiased forecasts. However, the slope coefficient β is significant in all cases and does not equal one. Interestingly, GARCH models tend to overestimate volatility due to $\beta < 1$ for all GARCH specifications. Conversely, stochastic volatility models give understated forecasts. Therefore, it might be more appropriate to use GARCH when estimating the upper boundary of “tomorrow’s” volatility and stochastic volatility.

Table 10: Mincer-Zarnowitz regression

	stand	exp	gjr	thresh	msm	stochvol
AFLT						
α	0.0002***	0.0001*	0.0002***	0.0001**	-0.0006***	-0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
β	0.3582***	0.4931***	0.3988***	0.4493***	1.4419***	1.9981***
	(0.0844)	(0.0944)	(0.0817)	(0.1011)	(0.0472)	(0.0742)
R^2	0.0349	0.0520	0.0456	0.0381	0.6517	0.5926
Adj. R^2	0.0330	0.0501	0.0437	0.0362	0.6510	0.5918
GAZP						
α	0.0001	0.0000	0.0000	0.0000	-0.0008***	-0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
β	0.5762***	0.7294***	0.7420***	0.7803***	1.5348***	1.7679***
	(0.1429)	(0.1325)	(0.1353)	(0.1391)	(0.0817)	(0.1607)
R^2	0.0316	0.0574	0.0570	0.0595	0.4149	0.1954
Adj. R^2	0.0297	0.0555	0.0551	0.0576	0.4137	0.1938
USD/RUB						
α	0.0000**	0.0000	0.0000**	0.0000	-0.0001***	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
β	0.6267***	0.6692***	0.6436***	0.7196***	4.3389***	1.9082***
	(0.1448)	(0.1449)	(0.1341)	(0.1569)	(0.1198)	(0.1311)
R^2	0.0362	0.0411	0.0442	0.0405	0.7248	0.2985
Adj. R^2	0.0343	0.0392	0.0423	0.0386	0.7242	0.2971
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$						

5 Conclusion

The article proposed the thorough investigation of the in-sample and out-of-sample performance of four GARCH and two stochastic volatility models. We applied the maximum likelihood method and Markov Chain Monte Carlo simulation to estimate the parameters and obtain one-day forecasts of ordinary GARCH, exponential GARCH, the Glosten-Jagannathan-Rünkle model, threshold ARCH, Markov switching multifractal and the stochastic volatility models. Using the probability integral transform, traditional forecast performance measures (MSE, MAD and DA) and a Mincer-Zarnowitz regression, we compared the aforementioned models and concluded that, in most cases, stochastic volatility models outperform GARCH, both in terms of explanation and prediction. One of the most important results was that the original stochastic volatility model is the only model where the log returns normality assumption was not rejected. We also demonstrated that GARCH models tend to overestimate the volatility forecasts, in contrast to stochastic volatility, where the forecasts are understated. The results of this research imply that it is more useful to involve multi-step forecasts in evaluating the out-of-sample performance of volatility models, and expanding the set of assets under consideration.

References

Ammann, M., Skovmand, D., and Verhofen, M. (2009). “Implied and realized volatility in the cross-section of equity options”. *International Journal of Theoretical and Applied Finance* 12.06, pp. 745-765.

Bollerslev, T. (1986). “Generalized autoregressive conditional heteroskedasticity”. *Journal of Econometrics* 31.3, pp. 307-327.

Brooks, S. et al. (2011). *Handbook of Markov Chain Monte Carlo.* CRC Press.

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, pp. 4983.

— (2008). *Multifractal volatility: theory, forecasting and pricing.* Academic Press.

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, pp. 168-187.

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

Ding, J. and Meade, N. (15, 2010). “Forecasting accuracy of stochastic volatility, GARCH and EWMA models under different volatility scenarios”. *Applied Financial Economics* 20.10, pp. 771783.

Friihwirth-Schnatter, S. and Wagner, H. (2010). “Stochastic model specification search for Gaussian and partial non-Gaussian state space models”. *Journal of Econometrics* 154.1, pp. 85100.

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, pp.1779-1801.

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, pp. 873-889.

Jeliazkov, I. and Lee, E. H. (2010). “MCMC perspectives on simulated likelihood estimation”. *Maximum simulated likelihood methods and applications.* Ed. by Greene, W and Hill, RC. Vol. 26. Advances in Econometrics. Howard House, Wagon Lane, Bingley, Yorkshire, England: Emerald Group Publishing Ltd, 3-39.

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

Kim, S., Shephard, N., and Chib, S. (1998). “Stochastic volatility: likelihood inference and comparison with ARCH models”. *The Review of Economic Studies* 65.3, pp. 361-393.

Lamoureux, C. G. and Lastrapes, W. D. (1993). “Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities”. *Review of Financial Studies* 6.2, pp. 293326.

Nelson, D. B. (1991). “Conditional Heteroskedasticity in Asset Returns: A New Approach”. *Econometrica* 59.2, p. 347.

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

Pederzoli, C. (1, 2006). “Stochastic Volatility and GARCH: a Comparison Based on UK Stock Data”. *The European Journal of Finance* 12.1, pp. 41-59.

Shephard, N. (1996). *Statistical aspects of ARCH and stochastic volatility.* Springer.

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, pp. 531-542.

Taylor, S. J. (1994). "Modeling stochastic volatility: A review and comparative study". *Mathematical Finance* 4.2, pp. 183-204.

Tsyplakov, A. (2007). "Revealing the arcane: an introduction to the art of stochastic volatility models". *Quantile* 8, pp. 69-122.

Vuong, Q. H. (1989). "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses". *Econometrica* 57.2, pp. 307-333.

West, K. and McCracken, M. (1998). "Regression-based tests of predictive ability". *International Economic Review* 39.4, 817-840.

Zakoian, J.-M. (1994). "Threshold heteroskedastic models". *Journal of Economic Dynamics and control* 18.5, pp. 931-955.

Valeria V. Lakshina

National Research University Higher School of Economics. Department of Mathematical Economics. Lecturer. Nizhniy Novgorod, Russia. E-mail: vlakshina@hse.ru, lakshinav@gmail.com