

Bayesian Techniques for Discrete Stochastic Volatility Models

Vladislav Soldatov

A Thesis Submitted for the Degree of MPhil at
Brunel University, the School of Information Systems, Computing and
Mathematics

August 21, 2013

Contents

Abstract	4
Acknowledgements	5
1 Introduction	6
2 Mathematical preliminaries	10
2.1 Basic concepts in mathematical statistics	10
2.2 Hypothesis testing	13
2.3 Bayesian statistics	14
2.3.1 Likelihood function and Bayesian theory	14
2.3.1.1 Prior distributions	16
2.3.2 Posterior inference	17
3 Dynamic Volatility Models	19
3.1 Properties of Financial Time Series	19
3.1.1 Distribution of returns	20
3.1.2 Normality tests	21
3.1.2.1 Likelihood function of returns	22
3.1.3 Elements of linear time series analysis	23
3.1.3.1 Stationarity	23
3.1.3.2 Correlation and Autocorrelation	23
3.1.3.3 Linear Time Series and White Noise	25
3.1.3.4 Stylized facts in financial time series	26
3.2 Modelling volatility	28
3.2.1 Types of volatility models	28
3.2.2 GARCH-type models	29
3.2.3 Research on ARCH/GARCH forecasting performance	32
3.2.4 Stochastic volatility models	34
3.3 Evaluation of volatility models	36

4	Parameter Estimation in Volatility Models	39
4.1	Numerical computation in Bayesian framework	39
4.2	Algorithms for posterior simulation	40
4.2.1	Markov chain Monte Carlo (MCMC) methods	40
4.3	MCMC convergence diagnostics	44
4.3.1	Geweke diagnostics.	45
4.3.2	Gelman and Rubin convergence diagnostics	46
4.3.3	Heidelberger and Welch diagnostics.	47
5	Empirical investigation of discrete SV models using MCMC	49
5.1	Description of the data	49
5.2	Bayesian analysis of Classical SV model	51
5.2.1	Theoretical model	51
5.2.1.1	Practical convergence assessment example	52
5.3	Bayesian techniques for stochastic volatility models with t-innovations	56
5.4	Bayesian techniques for stochastic volatility models with trading volume	57
6	Summary and future work	63
6.1	Contributions	63
6.2	Future work	64
	Appendix	66
A	Code Listings	66
A.1	Data preprocessing	66
A.1.1	Trading hours extraction	66
A.1.2	Specified period extraction	68
A.1.3	Aggregation	69
A.1.4	Realized volatility calculation	71
A.1.5	End of day data extraction for SV model calibration	73
A.1.6	Data analysis and hypothesis testing	74
A.2	JAGS models' description	76
A.2.1	Classic SV model	76
A.2.2	SV model with innovations having Student distribution	77
A.2.3	SV model with Gaussian innovations augmented with lag one trading volume	78
A.2.4	SV model with t-innovations augmented with lag one trading volume	79

A.2.5	SV model with Gaussian innovations augmented with lag two trading volume	80
A.2.6	SV model with t-innovations augmented with lag two trading volume	81
A.3	Data postprocessing	82
A.3.1	Parameter estimation for classic SV model	82
A.3.2	Forecasting performance analysis for classic SV model .	85

Bibliography		89
---------------------	--	-----------

Abstract

Reliable volatility forecasts are needed in many areas of finance, be it option pricing, risk management or portfolio allocation. Mathematical models that capture temporal dependencies between the returns of financial assets and their volatility and could be used for volatility forecasting generally fall into one of the following categories: historical volatility models, GARCH – type models and Stochastic Volatility (SV) models. This thesis will focus on the predictive ability of the discrete version of SV models. Six variants of discrete SV models will be estimated:

- classic SV model,
- SV model with innovations having Student distribution,
- SV model with Gaussian innovations augmented with lag one trading volume,
- SV model with t-innovations augmented with lag one trading volume,
- SV model with Gaussian innovations augmented with lag two trading volume,
- SV model with t-innovations augmented with lag two trading volume.

These models will be compared on the basis of their ability to predict volatility. Our study will show that SV model specification with Student t distribution with 3 degrees of freedom leads to a significant improvement in volatility forecasts, thus demonstrating good agreement with the empirical fact that financial returns have fat-tailed distribution. It will be shown that the influence of the trading volume is very small compared with the impact of different distributional assumptions on innovations.

Acknowledgements

I would like to thank my supervisor Dr Paresh Date for all his support and assistance during my research and writing of the thesis.

I am grateful to the Saratov State University and its pro-vice chancellor Tatyana Zakharova for the financial and organizational support of my studies through the programme “Innovative University”.

Chapter 1

Introduction

In finance, the term volatility is used to denote a measure of the variation of a particular asset. Mathematically it is often calculated as standard deviation of asset return [119]. Volatility plays an important part in option pricing, since it is the only quantity in the Black-Scholes option pricing formula that is not observed directly and needs to be estimated. Volatility modelling is useful in risk management for the value-at-risk calculations of a financial position. Asset allocation procedures based on mean-variance portfolio theory use volatility as a risk measure. Financial market volatility index is itself a financial instrument – VIX volatility index is compiled by the Chicago Board of Option Exchange (CBOE) and traded from March 2004. A number of mathematical models of volatility have been developed in order to provide the finance professionals with reliable volatility forecasts, e.g. historical volatility models, GARCH - type models and stochastic volatility models [121].

Since historical volatility models are built directly on the realized volatility from the previous periods, they are the easiest to estimate and produce forecasts. Riskmetrics EWMA (exponentially weighted moving average) from JP Morgan [97] is one of the widely used models from this class.

GARCH - type models were first designed by Engle [50] to capture the persistence in volatility and later proved to be a good fit for different kinds of financial time series [24]. Stochastic volatility models in continuous time are used mainly for the option pricing applications, one of the widely used model being the Heston model [77]. Discrete versions of stochastic volatility models, proposed by Taylor [146], are considered to be superior in capturing some basic features of financial time series (also called stylized facts). How-

ever, since parameter estimation for such models required computationally intensive simulation procedures (e.g., Markov chain Monte Carlo), they were not widely used for forecasting purposes. Recent developments in the area of Markov chain Monte Carlo (MCMC) methods for stochastic volatility (SV) models [139] and the availability of more or less standardized MCMC software tools (such as OpenBUGS software [99]) have made the application of SV models in volatility forecasting feasible.

There are a number of general advantages of Bayesian inference in applications to financial time series [81]. Firstly, for financial time series prior knowledge is often available: for example, dealers and traders usually have some information that they use to predict market patterns. This information could be easily incorporated into model priors. Secondly, Bayesian predictive densities automatically account for the parameter uncertainty in financial time series models: whereas a non-bayesian estimation of volatility models would require setting some fixed values for initial volatility, in Bayesian framework this problem is solved naturally by choosing a prior distribution for it. Finally, a strong motivation for the use of Bayesian methods stems from the ability of the MCMC methods to handle complicated models where either analytical expression for the likelihood function is not available or is nonlinear in the parameters (which is almost always the case for financial models), making optimization procedure very unreliable due to the existence of local optimal points [123].

A common approach for testing the forecasting capability of any model amounts to the comparison of the predicted values and subsequent realizations. But this approach is not applicable for volatility forecast evaluation due to the fact that volatility is a hidden variable and could not be observed directly. Since for a correctly specified model the mathematical expectation of squared returns is equal to their variance, squared returns could be used as a proxy for the unobserved ex-post volatility. It was shown by Lopez ([98]) that this produces a very noisy volatility estimate. Studies completed by Cumby [43], Figlewski [56], Jorion [86] were all based on this squared returns estimates and came to the conclusion that GARCH models are not able to satisfactorily explain the variability in ex-post returns.

Another approach to volatility forecast evaluation is based on the continuous-time volatility framework developed by Nelson[112] and Drost and Werker [46], that justified the usage of high-frequency trading data to construct the ex-post volatility estimate, which is called *Realized Volatility (RV)* and is calculated from the squared *intraday* returns. With the increase of the observation frequency RV converges to the unobserved actual volatility. Andersen

and Bollerslev [7] found that realized volatility, based on high-frequency returns, is a less noisy estimate of actual volatility (and has better stability) than the one that is estimated from squared daily returns. Their evaluation of volatility forecast based on the comparison with RV shows that GARCH models are able to explain at least 50% in the returns' variability.

There are a number of studies aimed at explanation of the contemporaneous correlation between trading volume and volatility. The first one was proposed by Clark [40] and is referred to as a mixture of distribution hypothesis (MDH). It postulates that both the variance of the daily price changes and trading volume are dependent on the same latent variable that represents the amount of information relevant to the price changes that arrives to the market. Then arrival of "good news" leads to a price increase, while "bad news" lead to a price decrease. Both the types of news lead to an increased market activity due to the market movement to a new price equilibrium. According to this concept a correlation between volatility and trading volume is due to their dependence on the latent information flow process.

Although ARCH and GARCH models are good at predicting the actual volatility values, they, as already noted above, do not explain the driving forces behind the volatility dynamics. Further attempt to explain them was made by Lamoureux and Lastrapes [92]. Their findings show that contemporaneous trading volume used as a proxy for the daily information flow in the volatility specification of the GARCH model has significantly stronger explanatory power compared to the previous price changes.

In our study we are going not only to extend the approach of Lamoureux and Lastrapes [92] to the discrete stochastic volatility models, but also attempt to construct a model suitable for volatility forecasting. To this end we will be trying to incorporate not the contemporaneous trading volume, but the one from the previous trading day. Using realized volatility estimates from intraday high-frequency trading data we will then compare the forecasting performance of different discrete stochastic volatility models with and without trading volume.

The rest of the thesis is organized as follows.

Chapter 2 collects together some facts from mathematical statistics that are used later in the thesis and establishes the notation used in the subsequent chapters.

Chapter 3 introduces general concepts of financial time series modelling and explains a number of volatility models that are used in empirical research.

Chapter 4 presents the Bayesian statistical framework with the emphasis on the numerical procedures that could be applied to the calibration of discrete SV models.

Chapter 5 contains the empirical investigation of six variants of discrete Stochastic Volatility model. Some results of augmentation of SV models with trading volume and “fat-tailed” innovations are reported and discussed.

Chapter 6 provides a summary of the main findings and outlines some directions for future research.

Appendix provides a collection of code listings which implement the numerical procedures used in the thesis.

Chapter 2

Mathematical preliminaries

2.1 Basic concepts in mathematical statistics

We first collect together some facts regarding mathematical statistics and will also establish notation used in the subsequent chapters.

Statistical distributions and their moments Let R^k be the k -dimensional Euclidean space and $x \in R^k$ denotes a point in it. Let $X = (X_1, X_2, \dots, X_k)$ and $Y = (Y_1, Y_2, \dots, Y_q)$ be continuous random vectors such that X is in the subspace $A \subset R^k$ and Y is in the subspace $B \subset R^q$.

Joint distribution The function

$$F_{X,Y}(x, y; \theta) = P(X \leq x, Y \leq y; \theta), \quad (2.1.1)$$

where $x \in R^p$, $y \in R^q$ (inequality \leq acts componentwise), is called a joint distribution function of X and Y with parameter θ . If the joint probability density function $f_{x,y}(x, y, ; \theta)$ of continuous random vectors X and Y exists, then

$$F_{X,Y}(x, y; \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{x,y}(w, z; \theta) dw dz. \quad (2.1.2)$$

Marginal distribution The marginal distribution of X is given by

$$F_X(x, \theta) = F_{X,Y}(x, \infty, \dots, \infty; \theta).$$

In scalar case (when $k = 1$) the distribution function becomes

$$F(x, \theta) = P(X \leq x; \theta),$$

where $x \in \mathbb{R}$, and is called a cumulative distribution function (CDF) of X . For a continuous random variable X

$$f(x) = \frac{d(F(x))}{dx}$$

defines the probability density function (PDF), assuming the CDF is differentiable. For a given probability p , the smallest q_p such that $p \leq F(q_p)$ is called the 100pth quantile of the random variable X :

$$q_p = \inf_x \{q | p \leq F(q)\}.$$

Conditional distribution The conditional distribution of X given $Y \leq y$ is given by

$$F_{x|Y \leq y}(x; \theta) = \frac{P(X \leq x, Y \leq y; \theta)}{P(Y \leq y; \theta)}.$$

If the probability density function for the random vectors X and Y exists, then the conditional density of X given $Y = y$ is

$$f_{x|y}(x; \theta) = \frac{f_{x,y}(x, y; \theta)}{f_y(y; \theta)}, \quad (2.1.3)$$

where the marginal density function $f_y(y; \theta)$ is obtained as

$$f_y(y; \theta) = \int_{-\infty}^{\infty} f_{x,y}(x, y; \theta) dx.$$

From the Eq. (2.1.3) we can get the following relation among joint, marginal and conditional PDFs:

$$f_{x,y}(x, y; \theta) = f_{x|y}(x; \theta) \times f_y(y; \theta). \quad (2.1.4)$$

This relation is widely used in maximum likelihood estimation and is also used to define the independence property, namely: random vectors X and Y are independent if and only if

$$f_{x|y}(x; \theta) = f_x(x; \theta).$$

For independent random vectors X and Y the relation (2.1.4) takes the form

$$f_{x,y}(x, y; \theta) = f_x(x; \theta) f_y(y; \theta). \quad (2.1.5)$$

Moments of a random variable

$$m'_\ell = E(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x) dx$$

is called the ℓ -th moment of a continuous variable X . The first moment $\mu_x = m'_1 = E(x)$ is called the *mean* or *expectation* of X and is used as a measure of the central location of the distribution.

The ℓ -th central moment of X is defined as

$$m_\ell = E[(X - m'_1)^\ell] = \int_{-\infty}^{\infty} (x - m'_1)^\ell f(x) dx.$$

The second central moment $\sigma^2 = m_2$ is called the *variance* of X and is used as a measure of variability of X . Its positive square root σ is a *standard deviation* of X .

The third central moment

$$S(x) = E \left[\frac{(X - \mu)^3}{\sigma^3} \right] \quad (2.1.6)$$

is called the *skewness* of X and is used as a measure of asymmetry of X 's distribution. The fourth central moment

$$K(x) = E \left[\frac{(X - \mu)^4}{\sigma^4} \right] \quad (2.1.7)$$

is called *kurtosis* and is used as a measure of X 's distribution tail thickness. There is a special term *excess kurtosis* for the expression $K(x) - 3$ due to the fact that $K(x) = 3$ for the normal distribution (excess kurtosis for the normal random variable is equal to 0). Positive excess kurtosis indicates that a distribution has heavier tails, that is more of its mass is located on the tails of its support (compared to the normal distribution). In financial analysis this is considered as an indicator that a random sample from such distribution could contain more extreme values. Distribution with positive excess kurtosis is called *leptokurtic*. There is also a term *platykurtic* which denotes a distribution with negative excess kurtosis and thinner tails.

For the random sample $\{x_1, \dots, x_T\}$ of X with T observations the sample variants of these moments are computed using the following formulas:

sample mean

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T x_T,$$

sample variance

$$\hat{\sigma} = \frac{1}{T-1} \sum_{t=1}^T (x_T - \hat{\mu})^2,$$

sample skewness

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}^3} \sum_{t=1}^T (x_T - \hat{\mu})^3,$$

sample kurtosis

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}^4} \sum_{t=1}^T (x_T - \hat{\mu})^4.$$

2.2 Hypothesis testing

When working with statistical models, hypotheses take the form of statements about model parameters. Using the information about a parameter that is contained in the data sample, hypothesis testing procedure is used to draw a conclusion about the hypothesis [79]. The *null hypothesis* specifies a value for the parameter being estimated in the form $H_0: \hat{\theta} = c$. This null hypothesis represents the belief that is considered to be true unless the sample evidence convinces us that it is not true. In the latter case we *reject* the null hypothesis. Every null hypothesis is accompanied by the *alternative hypothesis* that will be accepted if the null hypothesis is rejected. For the null hypothesis $H_0: \hat{\theta} = c$ there are three alternative hypothesis possible:

- $H_1: \hat{\theta} > c$,
- $H_1: \hat{\theta} < c$,
- $H_1: \hat{\theta} \neq c$.

The *test statistic* is a quantitative measure of the information about the null hypothesis contained in the sample. The value of a test statistic is used to decide whether to reject the null hypothesis. Test statistic is specially constructed in such a way that its probability distribution has one (theoretically known in advance) form if the null hypothesis is true and another form if the null hypothesis is not true. The *rejection region* is a range of values of the test statistic that leads to the rejection of the null hypothesis. When the null hypothesis is true, all values from the rejection region have low probability of occurring under the theoretical distribution of test statistic. Whether the

probability is “high” or “low” is determined by choosing the *significance level* α . Results of the hypothesis’ test is given by reporting the *p-value* (probability value) of the test. The decision on whether to reject or not the null hypothesis is made by comparing the p-value with the chosen level of significance α . If $p \leq \alpha$, then H_0 is rejected. We will use a variety of hypothesis tests in our empirical investigation later in this work.

2.3 Bayesian statistics

While there are a number of books on Bayesian statistics in general, the most appropriate one for financial application, in our opinion, is the one by Rachev et al [123], on which most of the discussion in this section is based.

2.3.1 Likelihood function and Bayesian theory

Statistical models are usually constructed to investigate possible relationships between the response variable Y and a number of explanatory variables $X = \{x_1, \dots, x_m\}$. Suppose we are interested in analyzing the evolution of stock prices (or returns) given as a time series object. A common statistical approach to this problem would be to try to guess what process generated the data. In the framework of Bayesian statistics such guess would be represented in the form of a probability density $f(y|\theta)$, where y is a realization of the random variable Y and θ is a vector of parameters that define the distribution f . Having the data set y , we are able to draw a conclusion about the value of θ . A likelihood function is a statistical construct introduced to measure our ability to extract information about parameter θ from the available data $y = \{y_1, y_2, \dots, y_n\}$. The joint probability density function of Y for a given value of θ would be

$$f(y_1, y_2, \dots, y_n; \theta) = \prod_{i=1}^n f(y_i|\theta). \quad (2.3.1)$$

We can treat this function as a function of the unknown parameters θ given a set of observed values $y = \{y_1, y_2, \dots, y_n\}$. Such function of the unknown parameters of a statistical distribution is called a *likelihood function* and is denoted as

$$L(\theta|y_1, y_2, \dots, y_n) = f(y_1, y_2, \dots, y_n; \theta). \quad (2.3.2)$$

The main philosophical principle of Bayesian statistics is that probability is a measure of the degree of belief one has about an uncertain event. Bayesian

inference is a process of updating these beliefs when some new information in the form of observed data arrives. The following theorem is a mathematical representation of this updating principle.

Bayes' theorem. Suppose that prior to observing our data y we have some evidence E about the values of the parameter θ , and the degree of the researcher's belief in it can be expressed as probability $p(E)$. According to the Bayes' theorem, after we have observed the data D our belief in E should be adjusted according to the following formula:

$$p(E|D) = \frac{p(D|E) \times p(E)}{p(D)}, \quad (2.3.3)$$

where:

- $p(D|E)$ is a conditional probability of the data given that the prior evidence E is true,
- $p(D)$ is the unconditional (marginal) probability of the data (the probability of D irrespective of E).

The probability of E before seeing the data, $p(E)$, is called the prior probability, whereas the updated probability, $p(E|D)$, is called the posterior probability. It could easily be seen that the probability $p(D|E)$, when treated as a function of E , is actually a likelihood function (2.3.2).

For financial applications that we are going to investigate in this work we will mostly need the continuous version of the Bayes theorem:

$$p(\theta|x) = \frac{L(\theta|x)\pi(\theta)}{f(x)}, \quad (2.3.4)$$

where $f(x)$ is the unconditional (marginal) distribution of the random variable X , given by

$$f(x) = \int L(\theta|x)\pi(\theta)d\theta. \quad (2.3.5)$$

Since $f(x)$ doesn't depend on θ , (2.3.5) usually written as

$$p(\theta|x) \propto L(\theta|x)\pi(\theta), \quad (2.3.6)$$

where

- θ = unknown parameter whose inference we are interested in,
- y = a set of recorded observations,

- $\pi(\theta)$ = prior distribution of θ depending on one or more parameters, called hyperparameters,
- $L(\theta|y)$ = likelihood function for θ ,
- $p(\theta|y)$ = posterior (updated) distribution of θ .

Schematically, Bayesian model analysis includes the following steps [94]:

1. Choose a specification for the data generating process of the observed data X . This is usually expressed in the form of probability density conditioned on the unknown parameter θ :

$$X \sim f(X|\theta).$$

2. Choose a specification for the prior distribution $\pi(\theta)$ for θ :

$$\theta \sim \pi(\theta).$$

3. Make inference on θ by computing integrals with respect to the posterior distribution:

$$\pi(\theta|X) = \frac{\pi(\theta)L(\theta|X)}{\int \pi(\theta)L(\theta|X)d\theta},$$

where $L(\theta|X) \propto f(X|\theta)$ is a likelihood of θ given X .

2.3.1.1 Prior distributions

Informative priors. To apply Bayesian computational methods, the prior information about parameter θ should be expressed in analytical form, that is a distribution of a certain class should be selected and the values for its parameters (which in Bayesian statistics are called hyperparameters) assigned. A value that we consider as the most typical for θ could be taken as a mode of this distribution (for symmetrical distributions it would be, in fact, its mean value). Distributions selected in this way are called *informative prior distributions*. A more difficult question is how to express our beliefs about the spread of the distribution. One approach is to make some guesses about the lower and upper quartiles of the prior distribution, which in turn could be used to get the spread estimate. If selected distribution has finite second-order moment, its value could be used to express our degree of belief in the value of θ .

Noninformative prior distributions. If one does not have any particular beliefs in the value of θ , a *noninformative prior* could be used. There are two main approaches to construction of noninformative priors. First option – employ a uniform distribution which support is set to the range of all possible values of θ . For example, diffuse prior for the mean of normal distribution could be taken in the form

$$\pi(\mu) \propto 1,$$

which is a uniform distribution with support $(-\infty, +\infty)$. Such prior is an example of *improper priors*, because

$$\int_{-\infty}^{+\infty} \pi(\mu) d\mu = \int_{-\infty}^{+\infty} 1 d\mu = \infty.$$

In theoretical Bayesian statistics this usually does not constitute a real problem since theoretically derived posterior distributions are proper. Second option to get a noninformative prior – use a proper prior distribution, but select specific hyperparameters' values to make it noninformative. Such noninformative distribution are called *diffuse priors* and are usually constructed selecting such hyperparameters that produce large second-order moments. For example, we can get a *normal diffuse prior* by setting a variance of a normal distribution to a significantly large value: $N(0, 10^6)$.

2.3.2 Posterior inference

As the posterior distribution $p(\theta|X)$ is obtained by applying the Bayes' theorem, it represents a combination of the data and the prior and encompasses all information about the parameter θ . Inference in Bayesian framework is performed by computing certain numerical characteristics of the posterior distribution (this gives results compared to frequentist approach). Most used for this purpose are posterior mean and posterior standard deviation. When posterior distribution is available in the closed form (as it is in the case of using a conjugate prior), these calculations could be done analytically. When we do not have posterior distribution in closed analytical form and compute it numerically (using simulation techniques such as MCMC), posterior mean is also computed numerically.

Bayesian Credible Intervals. Credible intervals in Bayesian statistical inference are used for the same purpose as confidence intervals in frequentist statistics. Procedure used to construct credible interval is (like many

definitions in Bayesian statistics) subjective. We will stick to a widely used technique when a credible interval is based on the highest posterior density (h.p.d). Then $100(1 - \alpha)\%$ credible interval is a set constructed as

$$R_{1-\alpha}^\pi = \{\theta : \pi(\theta|X) \geq \pi(\theta_{1-\alpha}|X)\}$$

for some $\theta_{1-\alpha}$ satisfying

$$\Pr(\theta \in R_{1-\alpha}^\pi) = 1 - \alpha.$$

There is an alternative credible set definition which is obtained by replacing the posterior density $\pi(\theta|X)$ with the likelihood $L(\theta|X)$

$$R_{1-\alpha}^L = \{\theta : \pi(L|X) \geq L(\theta_{1-\alpha}|X)\}$$

for some $\theta_{1-\alpha}$ satisfying

$$\Pr(\theta \in R_{1-\alpha}^L) = 1 - \alpha.$$

This second definition is useful when no prior information is available for θ .

Chapter 3

Dynamic Volatility Models

3.1 Properties of Financial Time Series

According to Campbell et al [33], there are several reasons to use returns instead of prices in financial time series models:

1. It is natural for the investors to think in terms of returns rather than in terms of prices because this is the characteristic of their investment that they are actually interested in;
2. Statistical properties of returns series make them more suitable as inputs to time series models.

There are several definitions of asset returns. If P_t denotes the price of an asset at time t , then two most used definitions have the form:

Simple return (SR)

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}, \quad (3.1.1)$$

Continuously compounded return (CCR)

$$r_t = \ln \frac{P_t}{P_{t-1}} = p_t - p_{t-1}, \quad (3.1.2)$$

where $p_t = \ln(P_t)$.

3.1.1 Distribution of returns

In general, a mathematical model for the log returns $\{r_t; t = 1, \dots, T\}$ is represented as their joint distribution function:

$$F_r(r_1, \dots, r_n; Y, \theta), \quad (3.1.3)$$

where vector Y can represent some exogenous factors and θ is a vector of parameters that define the function $F_r(\cdot)$. In time series analysis this joint distribution (3.1.3) is often represented in the following form

$$\begin{aligned} F(r_1, \dots, r_T; \theta) &= \\ &= F(r_1)F(r_2|r_1) \cdot \dots \cdot F(r_T|r_{T-1}, \dots, r_1) \\ &= F(r_1) \prod_{t=2}^T F(r_t|r_{t-1}, \dots, r_1), \end{aligned} \quad (3.1.4)$$

which is used to emphasize the dependencies of r_t on time. Specification of time evolution for the conditional distribution $F(r_t|r_{t-1})$ leads to different theories. Under the random walk hypothesis [55] conditional distribution $F(r_t|r_{t-1}, \dots, r_1)$ is equal to the marginal distribution $F(r_t)$, which means that returns are independent from their own previous values, and, consequently, unpredictable.

Although asset returns are measured at discrete time moments, for the theoretical purposes they are treated as continuous random variables. Assuming that a probability density function exists, (3.1.4) could be written as:

$$\begin{aligned} f(r_1, \dots, r_T; \theta) &= \\ &= f(r_1)f(r_2|r_1) \cdot \dots \cdot f(r_T|r_{T-1}, \dots, r_1) \\ &= f(r_1) \prod_{t=2}^T f(r_t|r_{t-1}, \dots, r_1). \end{aligned} \quad (3.1.5)$$

Several classes of statistical distributions can be used to model the asset returns. We will briefly discuss these choices and their consequences for the asset returns models.

Normal distribution According to this approach it is assumed that the simple returns $\{R_t|t = 1, \dots, T\}$ are independently and identically distributed normal random variables with fixed mean and variance. The normality assumption leads to tractable properties of returns and simplifies estimation of

parameters (e.g., maximum likelihood methods). But asset returns' normality also leads to several difficulties. By definition, simple return can not fall below -1, while normal random variable is unbounded. Statistical investigation of empirical asset returns (that is real returns from the stock exchange) does not support the normal distribution hypothesis either – we will give details on this later in the discussion devoted to the financial time series stylized facts.

Lognormal distribution According to this approach it is assumed that the log returns $\{r_t|t = 1, \dots, T\}$ are independently and identically distributed (iid) normal random variables with fixed mean μ and variance σ^2 . In this case the simple returns R_t are iid lognormal random variables with the following mean and variance:

$$E(R_t) = \exp\left(\mu + \frac{\sigma^2}{2}\right) - 1, \quad \text{Var}(R_t) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]. \quad (3.1.6)$$

If m_1 and m_2 are the mean μ and variance σ^2 of the lognormally distributed simple return R_t , then

$$E(r_t) = \ln \left[\frac{m_1 + 1}{\sqrt{1 + \frac{m_2}{(1+m_1)^2}}} \right], \quad \text{Var}(r_t) = \ln \left[1 + \frac{m_2}{(1+m_1)^2} \right]. \quad (3.1.7)$$

With this distribution assumption simple returns, which could be calculated from CCR using $R_t = \exp(r_t) - 1$, have the necessary lower bound.

3.1.2 Normality tests

Since the normality of returns is used in all the models investigated in this thesis, we need statistical tests to check such distribution assumption – normality tests. We will employ two of them.

Shapiro – Wilk test It is one of the standard tests for normality implemented in many statistical packages [137]. According to Shapiro and Wilk [138], the test is conducted as follows. Let x_1, \dots, x_n be a sample of independent and identically distributed values and $x_{(1)}, \dots, x_{(n)}$ a set of their ordered values. Set $s^2 = (n - 1)^{-1} \sum (x_i - \bar{x})^2$, where $\bar{x} = n^{-1} \sum x_i$. Then the Shapiro-Wilk statistic is computed as

$$W = \sum \frac{a_i x_{(i)}}{s},$$

where a_1, \dots, a_n depend on the expected values of the order statistics from a standard normal distribution and are tabulated in [138]. The null hypothesis H_0 is that the random sample is from a normal distribution against the alternative H_1 that the random sample is not from a normal distribution. The null hypothesis H_0 is rejected at the significance level α if $\text{p-value}(W) < \alpha$. To calculate $\text{p-value}(W)$ statistical packages use numerical approximations proposed by Royston [130, 131, 132].

Jarque and Bera normality test Sample excess and kurtosis can be used to test the normality of asset returns. Jarque and Bera [84] showed that under normality assumption the test statistics

$$JB = \frac{\widehat{S}^2}{\frac{6}{T}} + \frac{[\widehat{K} - 3]^2}{\frac{24}{T}}.$$

is asymptotically distributed as $\chi^2(2)$. The null hypothesis H_0 (r is normally distributed) is rejected at the significance level α if $\text{p-value}(JB) < \alpha$, where p-value stands for the probability of the JB (calculated from data) to occur under the assumption that H_0 is true.

3.1.2.1 Likelihood function of returns

If the conditional distribution $f(r_t|r_{t-1}, \dots, r_1, \theta)$ is normal with mean μ_t and variance σ_t^2 , then $\theta = (\mu_t, \sigma_t^2)$ and, using the partitioning formulas (3.1.4) or (3.1.5), we can obtain the likelihood function for the observed log returns r_t :

$$f(r_1, \dots, r_T; \theta) = f(r_1; \theta) \prod_{t=2}^T \frac{1}{\sqrt{2\pi}\sigma_t} \exp \left[-\frac{(r_t - \mu_t)^2}{2\sigma_t^2} \right], \quad (3.1.8)$$

Maximum likelihood estimate (MLE) of θ is obtained as the value of θ that maximizes the likelihood function. In some calculations, it is more convenient to use the log-likelihood function

$$\ln f(r_1, \dots, r_T; \theta) = \ln f(r_1; \theta) - \frac{1}{2} \sum_{t=2}^T \left[\ln(2\pi) + \ln(\sigma_t^2) + \frac{(r_t - \mu_t)^2}{\sigma_t^2} \right]. \quad (3.1.9)$$

Since logarithm is a concave function, the argument θ which maximizes 3.1.9 also maximizes 3.1.8.

3.1.3 Elements of linear time series analysis

Our analysis in subsequent chapters will be based mainly on the concepts from time series theory which could be found in the books by Box et al [28] and Brockwell and Davis [31]. Some of the fundamental concepts are briefly reviewed below.

3.1.3.1 Stationarity

A time series $\{r_t\}$ is called strictly stationary if the joint distributions of $(r_{t_1}, \dots, r_{t_k})$ and $(r_{t_1+t}, \dots, r_{t_k+t})$ are identical for all t , where k is an arbitrary positive integer and (t_1, \dots, t_k) is a collection of k positive integers. Since it is not always possible to verify this strong condition empirically, a less restrictive concept of *weak stationarity* is often used. A time series $\{r_t\}$ is called *weakly stationary* if the mean of r_t and the covariance between r_t and $r_{t-\ell}$ are time invariant for an arbitrary integer ℓ , where *covariance* for the random variables x and y with finite second moments is defined as $\text{Cov}(x, y) = E\{(x - \mu_x)(y - \mu_y)\}$. If $x = y$, then $\gamma_\ell = \text{Cov}(r_t, r_{t-\ell})$ is called the lag- ℓ autocovariance of the time series r_t .

$\{r_t\}$ is weakly stationary if:

- $E(r_t) = \mu$ (constant in time),
- $\text{Cov}(r_t, r_{t-\ell}) = \gamma_\ell$ (depends on ℓ).

On a time plot of the weakly stationary time series its values fluctuate with constant variation around a fixed level.

If r_t is strictly stationary and its first two moments are finite, then r_t is also weakly stationary. The reverse in general does not hold, but for the normally distributed time series r_t weak stationarity is equivalent to strict stationarity. Weak stationarity is one of the common assumptions in the study of financial time series.

3.1.3.2 Correlation and Autocorrelation

The correlation coefficient between two random variables X and Y is defined as

$$\rho_{x,y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E(X - \mu_x)^2 E(Y - \mu_y)^2}}. \quad (3.1.10)$$

It can be shown that $-1 \leq \rho_{x,y} \leq 1$ and $\rho_{x,y} = \rho_{y,x}$. The value of $\rho_{x,y}$ measures the strength of linear dependence between X and Y , where $\rho_{x,y} = 0$ indicates that X and Y are uncorrelated. If X and Y are normal random variables, then $\rho_{x,y} = 0$ if and only if X and Y are independent. Estimates of $\hat{\rho}_{x,y}$ for the sample $\{(x_t, y_t)\}_{t=1}^T$ are obtained using the formula

$$\hat{\rho}_{x,y} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2 \sum_{t=1}^T (y_t - \bar{y})^2}}, \quad (3.1.11)$$

where \bar{x} and \bar{y} stand for the sample means of X and Y .

Autocorrelation function (ACF) Let r_t be a weakly stationary time series of returns. The correlation coefficient between r_t and $r_{t-\ell}$ is called the lag- ℓ autocorrelation and is denoted by ρ_ℓ . In case of a weak stationarity it is a function of ℓ :

$$\rho_\ell = \frac{\text{Cov}(r_t, r_{t-\ell})}{\sqrt{\text{Var}(r_t) \text{Var}(r_{t-\ell})}} = \frac{\text{Cov}(r_t, r_{t-\ell})}{\text{Var}(r_t)} = \frac{\gamma_\ell}{\gamma_0}, \quad (3.1.12)$$

since for a weakly stationary time series $\text{Var}(r_t) = \text{Var}(r_{t-\ell})$. From this definition follows that $\rho_0 = 1$, $\rho_\ell = \rho_{-\ell}$, $-1 \leq \rho_\ell \leq 1$ and weakly stationary time series is not serially correlated if and only if $\rho_\ell = 0$ for all $\ell > 0$.

Estimate of the autocorrelation for the sample $\{r_t\}_{t=1}^T$ is calculated using the formula

$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad 0 \leq \ell < T - 1. \quad (3.1.13)$$

Hypothesis testing for ACF To test $H_0 : \rho_\ell = 0$ versus $H_1 : \rho_\ell \neq 0$, one can use the statistic

$$t = \frac{\rho_\ell}{\sqrt{\frac{1 + 2 \sum_{i=1}^{\ell-1} \rho_i^2}{T}}}, \quad (3.1.14)$$

which for a stationary Gaussian time series $\{r_t\}$ under assumption $\rho_j = 0$ for $j > \ell$ is asymptotically distributed as a standard normal random variable [31, 28, 58], and H_0 is rejected at the significance level α if $|t| > q_{\alpha/2}$.

Portmanteau test It is a common task in financial applications to test that several autocorrelations of r_t are zero – statistical tests of this kind are called "portmanteau tests". Box and Pierce [29] proposed the test statistic

$$Q^*(m) = T \sum_{\ell=1}^m \widehat{\rho}_\ell^2 \quad (3.1.15)$$

for the null hypothesis $H_0 : \rho_1 = \dots = \rho_m = 0$ against the alternative hypothesis $H_1 : \rho_i \neq 0$ for some $i \in \{1, \dots, m\}$. $Q^*(m)$ is asymptotically distributed as a chi-squared random variable with m degrees of freedom. Ljung and Box [96] proposed a statistic with increased power for the finite samples:

$$Q(m) = T(T+2) \sum_{\ell}^m \frac{\widehat{\rho}_\ell}{T-\ell}. \quad (3.1.16)$$

H_0 is rejected if $Q(m) > \chi_\alpha^2(m)$.

3.1.3.3 Linear Time Series and White Noise

A time series r_t is called a white noise if $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance [150]. If r_t is normally distributed with zero mean and variance σ^2 , the series is called a Gaussian white noise. ACF of a white noise series is zero at all lags – this fact is often used as a criteria of white noise property for empirical returns. For the serially correlated time series it is necessary to build a model for this serial dependance before applying other analysis techniques.

Linear Time Series A linear time series r_t can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \quad (3.1.17)$$

where μ is the mean of r_t , $\psi_0 = 1$, $\{a_t\}$ is a white noise and the coefficients $\{\psi_i\}$ are called *weights* of r_t . Since a_t represents the new information available at time t it is often referred to as *innovation* or *shock* at time t . If r_t is weakly stationary, its mean and variance could be obtained from (3.1.17) using the independence of $\{a_t\}$:

$$E(r_t) = \mu, \quad \text{Var}(r_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2, \quad (3.1.18)$$

where $\sigma_a^2 = \text{Var}(a_t)$. From the finite variance requirement $\text{Var}(r_t) < \infty$ follows that $\{\psi_i^2\}$ must be a convergent sequence and $\psi_i^2 \rightarrow 0$ as $i \rightarrow \infty$. Consequently, past shocks a_{t-i} have less impact on r_t as i increases.

The lag- ℓ autocovariance of r_t is

$$\begin{aligned} \gamma_\ell &= \text{Cov}(r_t, r_{t-\ell}) = E \left[\left(\sum_{i=0}^{\infty} \psi_i a_{t-i} \right) \left(\sum_{j=0}^{\infty} \psi_j a_{t-\ell-j} \right) \right] \\ &= E \left(\sum_{i,j=0}^{\infty} \psi_i \psi_j a_{t-i} a_{t-\ell-j} \right) \\ &= \sum_{j=0}^{\infty} \psi_{j+\ell} \psi_j E(a_{t-\ell-j}^2) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+\ell}. \end{aligned} \quad (3.1.19)$$

Then autocorrelations of r_t could be calculated from weights $\{\psi_i\}$:

$$\rho_\ell = \frac{\gamma_\ell}{\gamma_0} = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+\ell}}{1 + \sum_{i=1}^{\infty} \psi_i^2}, \quad \ell \geq 0, \quad (3.1.20)$$

where $\psi_0 = 1$. As noted above, for a weakly stationary time series $\psi_i \rightarrow 0$ as $i \rightarrow \infty$, and, hence, $\rho_\ell \rightarrow 0$ as $\ell \rightarrow \infty$. In terms of asset returns this means that linear dependence of the current return r_t on the past return $r_{t-\ell}$ diminishes with the increase of ℓ .

Based on [150], one may enumerate the steps in building a time series model of asset returns as follows:

1. Test for serial dependence in the returns data using Box-Pierce or Ljung-Box test. If time series has linear dependence, then specify a model for the time evolution of the mean (e.g., ARMA model).
2. Use residuals of the model built in step 1 to estimate volatility model.
3. If the series is not serially dependent, simply subtract the sample mean of the series from its values.

3.1.3.4 Stylized facts in financial time series

Empirical studies on financial time series from a *statistical* point of view indicate that prices of different classes of financial assets share some (sometimes non-trivial) common statistical properties. Such properties that could be observed for time series of different instruments, markets and time periods are called *stylized facts*.

Following Cont [41], financial assets share the following common properties:

1. **Absence of autocorrelations.** Linear autocorrelations of asset returns are approximately zero for all lags, with the exception of intraday data with time step less than 20 minutes where non-zero autocorrelations could be present due to microstructure effects.
2. **Heavy tails in unconditional distribution of returns.** Yet Mandelbrot [102] has pointed out that normal distribution is not sufficient to model the distribution of asset returns. Excess kurtosis of the returns' distribution (2.1.7) is often used as a measure of deviation from normal distribution. Positive values of K indicate a "fat tail" – a slow asymptotic decay of the PDF. Different distribution could be used to model this behaviour, one of them being t-distribution [89, 21].
3. **Assymetry in upward and downward price movements.** For most of the stock prices and indices one can observe significant downward price movements without the upward movements of an equal size.
4. **Aggregation distribution dependence.** Distribution of returns depends upon the time scale Δt over which returns are calculated. With the increase of Δt the distribution of returns converges to normal.
5. **Volatility clustering.** Realized volatility displays a positive autocorrelation over several days and events with high volatility tending to cluster in time.
6. **Heavy tails in conditional distribution of returns.** Residual returns time series obtained after correcting returns for volatility clustering (e.g., using GARCH-type models) still exhibit heavy tails, although they are less heavy than in unconditional distribution.
7. **Slow decay in autocorrelations of absolute returns.** Autocorrelation of absolute returns as a function of time lag decays slowly, approximately following a power law with exponent $\beta \in [0.2, 0.4]$.
8. **Leverage Effect.** Volatility of an asset is negatively correlated with the returns on that asset.
9. **Volume/volatility correlation.** Trading volume is correlated with realized volatility.

3.2 Modelling volatility

Poon ([119]) defines volatility as a measure of “the spread of all likely outcomes of an uncertain variable.” When working with time series of asset returns in financial applications, sample standard deviation is often used as a measure of volatility:

$$\hat{\sigma} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - \hat{\mu})^2},$$

where r_t is daily asset return, μ — sample average return.

We use standard deviation and not variance here because variance is less stable while performing numerical computations, and standard deviation is measured in the same units as as the mean value — the latter is more convenient while working with financial time series. From the pure theoretical point of view these measures are absolutely equal while comparing spreads of financial time series.

The term “volatility” could be used in two different meanings :

- Realized volatility – calculated from past observations using formula (3.3),
- Model (theoretical volatility) – volatility estimate produced by theoretical models like GARCH or stochastic volatility, where volatility itself is a hidden variable.

3.2.1 Types of volatility models

Volatility models could be classified into the following major types [119]:

1. historical volatility,
2. ARCH/GARCH models,
3. stochastic volatility models (discrete and continuous time versions).

Historical volatility models (HISVOL). Models of this type use lagged values of realized volatility and are quite easy to construct and manipulate [120]. While the main input into GARCH and SV estimation procedures are

the past returns, HISVOL models do not need them and use estimates of realized volatility from the previous periods:

$$\widehat{\sigma}_t = \psi_1\sigma_{t-1} + \psi_2\sigma_{t-2} + \dots + \sigma_\tau\sigma_{t-\tau},$$

where

- $\widehat{\sigma}_t$ – expected standard deviation at time t
- ψ_i – weight parameter
- $\sigma_{t-\tau}$ – historical values of standard deviation for preceding periods.

If sufficient data are available for every time period, historical volatilities $\sigma_{t-1}, \sigma_{t-2}, \dots, \sigma_{t-\tau}$ are calculated from historical returns, and then this model could be estimated using regression techniques. In case of insufficient daily data *volatility proxies* could be used.

Widely used Riskmetrics EWMA (exponentially weighted moving average) model developed by JP Morgan [15] is a HISVOL model and has the form

$$\widehat{\sigma}_t^2 = \alpha\sigma_{t-1}^2 + (1 - \alpha)r_{t-1}^2. \quad (3.2.1)$$

A number of studies reveals that HISVOL models have good forecasting performance in comparison to other classes of models. Taylor ([148], [147]), Figlewski ([56]), Figlewski and Green ([70]), Andersen, Bollerslev, Diebold and Labys([8]), Taylor, J. ([145]) claim that HISVOL models produce better volatility forecasts than GARCH or SV.

3.2.2 GARCH-type models

Autoregressive conditionally heteroscedastic (ARCH) model of Engle [50] and generalized ARCH (GARCH) of Bollerslev [24] specify a functional dependence of the conditional variance of price changes on the price changes in previous time periods. This variance specification enables these models to capture the volatility clustering which is often observed in empirical returns time series.

ARCH models. The ARCH (AutoRegressive Conditional Heteroscedasticity) was proposed by Engle ([50]) and was historically a first systematic framework for volatility modelling. Extensive investigations on this model were performed by Bera and Higgins ([18]), Bollerslev, Chou and Kroner ([25]), Bollerslev, Engle and Nelson ([26]) Diebold and Lopez ([45]).

ARCH models differ from HISVOL models in the fact that they do not use past values of realized volatilities, but treat volatility as a hidden variable and estimate its values using the method of maximum likelihood. The model itself is usually formulated as follows:

$$\begin{aligned} r_t &= \mu + \varepsilon_t, \\ \varepsilon_t &= \sqrt{h_t} z_t, \end{aligned} \tag{3.2.2}$$

$h_t = \sigma_t^2$ is a conditional variance (common denotation in ARCH literature), $z_t \sim N(0, 1)$ is a white noise. This model is interpreted in the following way: white noise process z_t is scaled by the conditional variance h_t which is a function of past squared returns. Original conditional variance specification proposed by Engle ([50]) has the form

$$h_t = \omega + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2, \tag{3.2.3}$$

where restrictions $\omega > 0$ and $\alpha_j \geq 0$ are imposed to insure that h_t is strictly positive.

It could be seen from 3.2.3 that past squared volatility shocks ε_{t-j}^2 lead to large conditional variance h_t in subsequent periods, which, in turn, leads to high innovation values ε_t . In other words, big shocks tend to be followed by big shocks [150], which allows a good reproduction of volatility clustering observed in empirical asset returns. Several general properties of ARCH model give a good insight into its suitability for the volatility modelling. It could be shown [150], that excess kurtosis of ε_t is positive and the tail distribution of ε_t is heavier than that of a normal distribution. As a consequence, any shock ε_t of a Gaussian ARCH model is more likely to produce “outliers” than a Gaussian white noise series, which is in good agreement with the stylized facts on financial time series.

ARCH model also has some disadvantages in relation to volatility modelling [150]:

1. Since volatility in ARCH depends on the previous squared shocks, the effects of the positive and negative shocks on the volatility in ARCH framework are the same, while empirical time series of returns clearly show that this is not the case.
2. ARCH model imposes certain restrictions on its coefficients, which, especially for the high-order ARCH models with Gaussian innovations, limits its ability to capture excess kurtosis.

3. In general, volatility forecasts produced by ARCH models tend to be higher than the actual volatility.

Generalized ARCH Further extension to ARCH was proposed by Bollerslev ([24]), where volatility h_t also depends on its own lagged values:

$$h_t = \omega + \sum_{i=1}^p \beta_i h_{t-i} + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2,$$

$\omega > 0$. In order to ensure that h_t is strictly positive a number of constraints should be imposed on α_i and β_i . For example, in the simplest case of GARCH(1,1) it is required that $\alpha_i \geq 0$ and $\beta_i \geq 0$, discussion of more complex cases could be found in [113].

Integrated GARCH [51] When $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j = 1$, the unconditional variance of a GARCH(p,q) process $\sigma^2 \rightarrow \infty$ and is no longer definite and the time series r_t is not covariance stationary. In this case conditional variance is described as an integrated GARCH (IGARCH) process with infinite fourth moment.

Exponential GARCH This modification was proposed by Nelson. One of the motivations was to specify conditional variance in a form that would render obsolete the parameter constraints (which were used in GARCH to avoid negative variance):

$$\ln h_t = \omega + \sum_{j=1}^q \beta_j \ln h_{t-j} + \sum_{k=1}^p \left[\theta_k \varepsilon_{t-k} + \gamma_k \left(|\varepsilon_{t-k}| - \sqrt{\frac{2}{\pi}} \right) \right],$$

$$\varepsilon_t = \frac{\varepsilon_t}{\sqrt{h_t}}.$$

In EGARCH hidden variance h_t depends not only on the magnitude of ε_t , but also on its sign. This allows to reproduce one of the stylized facts – large negative returns lead to higher conditional variance in the following period than large positive returns.

GJR-GARCH Was proposed by Glosten, Jagannathan and Runkle [68].

$$h_t = \omega + \sum_{i=1}^p \beta_i h_{t-i} + \sum_{j=1}^q (\alpha_j \varepsilon_{t-j}^2 + \delta_j D_{j,t-1} \varepsilon_{t-j}^2),$$

$$D_{t-1} = \begin{cases} 1, & \text{if } \varepsilon_{t-1} < 0 \\ 0, & \text{if } \varepsilon_{t-1} \geq 0 \end{cases}$$

This model allows to account for the leverage effect typically observed in financial returns since it is formulated in such a way that negative residuals generate higher volatility in subsequent periods than positive residuals of the same magnitude do.

TGARCH Threshold GARCH (TGARCH) was proposed by Zakoïan [158] and is quite similar to GJR-GARCH. The main difference is that it uses absolute values of the returns to specify conditional variance:

$$h_t = \omega + \sum_{i=1}^p [\alpha_i |\varepsilon_{t-i}| + \gamma_i D_{i,t-i} |\varepsilon_{t-i}|] + \sum_{j=1}^q \beta_j \sigma_{t-j}.$$

NGARCH Nonlinear Asymmetric GARCH was proposed by Engle and Ng [52] and its equation is given by:

$$h_t = \omega + \alpha(\varepsilon_{t-1} - \theta\sqrt{h_{t-1}})^2 + \beta h_{t-1},$$

where $\alpha, \beta \geq 0; \omega > 0$. Parameter θ is a measure of the leverage effect.

3.2.3 Research on ARCH/GARCH forecasting performance

From the theoretical point of view one can list several deficiencies of ARCH models:

1. Since the model is built on squared returns, positive and negative shocks have the same effects on volatility. In practice, however, prices of financial assets respond differently to positive and negative shocks.
2. ARCH model does not explain theoretically the evolution of variance in financial time series. It simply provides an algorithm to calculate the estimate of conditional variance using past returns.

Good forecasting performance of GARCH-type models is supported by a number of studies. Akgiray [3] discovered that daily time series of stock market returns demonstrate higher statistical dependence (than had been

reported in earlier studies), which could be explained if one assumes that the prices themselves are generated by the nonlinear stochastic process. In general, the paper gives clear indications that GARCH models provide better volatility forecasts than EWMA and HISVOL.

Pagan and Schwert [115] reported that EGARCH forecasting performance is better than that of GARCH and non-parametric forecasting methods. This research compared several statistical models for the monthly stock return volatility using U.S. data from the years 1835-1925. GARCH and EGARCH models were employed along with some non-parametric methods (which are beyond the scope of this thesis). For the in-sample forecasting evaluation the authors concluded that in most cases the non-parametric procedures provided a better explanation of the squared returns than any of the parametric models, although predictive power of EGARCH was almost equivalent to that of the non-parametric models. Different results were obtained in the out-of-sample prediction experiments, where parametric models (and EGARCH in particular) performed significantly better than non-parametric ones.

Cumby, Figlewski and Hasbrouk [43] found EGARCH better than HISVOL. Figlewski [56] finds that GARCH is the best choice for the stock market and short-term forecasts.

It could be noticed that models with asymmetrical specification of volatility dependence on past returns demonstrate better forecasting performance - behaviour closely connected with the well known for the financial time series stylized fact.

Using monthly volatility series of stock returns, Cao and Tsay [34] compared performance of threshold autoregressive (TAR), linear ARMA, GARCH and EGARCH. Results of their comparison could be summarized into the following points:

1. the TAR models outperform the linear ARMA models in multi-step ahead forecasts for large stocks,
2. the TAR models provide better forecasts than the GARCH and EGARCH models for the volatilities of large stocks,
3. EGARCH gives the best long-term volatility forecast for small stock returns.

Heynen and Kat [78], Lee [93], indicate that EGARCH is superior for the volatility prediction of the stock indices and foreign exchange rates, but, according to Blailsford and Faff [30] and Taylor [145], for the stock indices GJR-GARCH can outperform GARCH. Bali [14] successfully applied

several GARCH-type models (GARCH, NGARCH, VGARCH, AGARCH, QGARCH, GJR-GARCH, TGARCH, TS-GARCH) to forecast volatility of US T-bills, finding that NGARCH performed the best.

There are also studies where it was not possible to identify the best model: Lee [93], West and Cho [153], Brailsford and Faff [153], Brooks [32] (which also discusses trading volume), McMillan, Speight, Gwilym [104].

As parameter estimates in ARCH/GARCH models are usually obtained using maximum likelihood methods, and ML, in turn, uses numerical optimisation procedures, lack of convergence which arise when data period is not long enough or there are significant changes in the volatility level is a common problem in practical applications of these models – parameter estimates could become very unstable. ARCH/GARCH convergence problems were studied in [152], [148], [151], [90].

3.2.4 Stochastic volatility models

Stochastic Volatility (SV) model for financial time series was introduced by Tauchen and Pitts [144] and Taylor [146]. For a number of reasons SV was generally considered more as a purely theoretical concept than a technique that could be applied in practical setting to get reliable volatility forecasts. This was mostly due to the fact that in general case estimation of SV required time-consuming simulation methods – for example, Markov chain Monte Carlo method, which we will use later in this thesis. Stochastic volatility model is specified as

$$r_t = \mu + \varepsilon_t,$$

with innovations in the form

$$\varepsilon_t = z_t e^{\frac{h_t}{2}}.$$

with the following volatility process specification

$$h_t = \mu + \varphi(h_{t-1} - \mu) + u_t,$$

where $z_t \sim N(0, 1)$ and $u_t \sim N(0, \tau^2)$.

Compared to ARCH/GARCH volatility specification, here hidden volatility is itself a stochastic process (like returns process), which leads to more flexibility and can lead to a better fit (than with GARCH-type models). As volatility in this model is itself a stochastic variable, this allows the model to

better reproduce the fat tails in the return distribution. Volatility process specification in form of the autoregressive process (AR(1) in our case) allows for the volatility persistence.

As a result of introduction of the stochastic term in the volatility specification of this model (contrary to ARCH/GARCH), it is not possible to get the likelihood function in the closed form, and, consequently, this model could not be estimated by maximum likelihood method.

Ruiz [133] used quasi-maximum likelihood estimation (QMLE) and generalized method of moments (GMM). Andersen and Sorensen ([10]) provided some comments on the efficiency of these methods applied to the estimation of SV models.

A number of other methods was applied to the estimation of SV models: generalized method of moments (using simulation) by Duffie and Singleton ([47]), possibilities of obtaining an analytical solution were explored by Singleton ([141]), likelihood approach with numerical integration by Fridman and Harris ([57]), Monte Carlo integration with importance sampling by Danielsson ([44]), Pitt and Shepard ([140]), Durbin and Koopman ([49]), Monte Carlo integration with Markov Chain techniques by Jacquier, Polson and Rossi ([82]), and Kim, Shephard and Chib ([88]).

Forecasting performance of Stochastic Volatility models There are not many studies where SV forecasting performance is investigated.

Heynen and Kat ([78]) use SV to forecast volatility of stock indices and exchange rates. SV provided the best forecast for indices but not so satisfactory results for exchange rates.

Yu([156]) obtained good results in forecasting volatility for the New Zealand's stock market.

Lopez ([98]) found no difference between SV and other conventional volatility forecasting methods.

Bluhm and Yu ([22]), Dunis, Laws and Chauvin ([48]), Hol and Koopman ([80]) compared SV forecast with implied volatility forecasts. Bluhm and Yu ([22]), Hol and Koopman ([80]) concluded that implied volatility forecasts for the stock index volatility are better than SV forecasts, in [22] SV was found equally precise as GARCH.

3.3 Evaluation of volatility models

As we have already mentioned before, volatility is a latent variable. Volatility estimation procedure depends upon the type and amount of information available for each interval Δt and the whole period T of the time series used for estimation.

When $\exists k \in \mathbb{N}$ such that $k \cdot \Delta t \leq T$, definition (3.2) could be used:

$$\hat{\sigma} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - \mu)^2}.$$

This case applies, for example, when daily data is available and monthly volatility should be calculated.

As many economic and financial time series are available only on daily basis, it is quite common that $\Delta t = T$, which presents a more difficult case. In financial applications we usually have end of day stock market data and daily volatility estimation is required. Several approaches were developed in order to tackle this case. In general they propose some function of daily returns to be used as a *proxy* for daily volatility. Several studies have shown that such proxies give only a rough estimate of daily volatility. According to Figlewski ([56]), absolute value of returns as a proxy it is a very poor estimate of daily volatility. Squared returns as volatility proxy provide unbiased but very imprecise estimations of volatility. (see [98]). More precise volatility estimate could be obtained with the volatility estimators based on several prices from the trading day. If H_t and L_t are the highest and lowest prices on day t , then according to derivations in ([23]) we have volatility estimator in the form

$$\hat{\sigma}^2 = \frac{(\ln H_t - \ln L_t)^2}{4 \ln 2}.$$

This volatility estimator was studied by Parkinson ([116]), Garman and Klass ([62]), Beckers ([17]), Rogers and Satchell ([128]), Wiggins ([154]), Rogers, Satchell and Yoon ([129]), Alizadeh, Brandt and Diebold ([4]). Another estimator is proposed in [62] as an extension to [116] and uses high (H_t), low (L_t) and close (C_t) prices:

$$\hat{\sigma}^2 = 0.5 \left(\ln \frac{H_t}{L_t} \right)^2 - 0.39 \left(\ln \frac{C_t}{C_{t-1}} \right)^2.$$

Original form of stock market data is a time series of tick-by-tick prices: each tick is produced by such events as a quote or a transaction. Their time

spacing is irregular due to the fact that the actions that produce them are irregularly spaced in time. Another reason why intraday data are so important is that participants of financial markets (traders) base their decisions on it (cf. live market data feeds from Reuters, Bloomberg, dtm). Generally, intraday data comes in the form of tick data, that is the data itself (price of a financial asset, or a trading volume, etc) accompanied by the time stamp. As in recent years tick data from stock markets become available, *realized volatility* calculated using intraday (5min or 15min) squared returns was proposed as a measure of daily volatility. It could be shown ([7], [59], [87]), that for time series with zero mean and no jumps this realized volatility is a good approximation for the theoretical continuous time volatility. Realised volatility $RV(t_i)$ at time t_i is calculated from historical data and is also sometimes called *historical volatility* [65]:

$$RV(t_i) = RV(\Delta t, n, p; t_i) = \left[\frac{1}{n} \sum_{j=1}^n |r(\Delta t; t_{i-n+j})|^p \right]^{\frac{1}{p}},$$

where r are regularly spaced returns, n is a number of observations, Δt – time interval between returns, $n\Delta t$ – time span of the entire data set used for calculations. It is a common practice to set $p = 2$. To compare volatility calculated using different types of datasets and procedures, it is convenient to scale it to a certain time period. Assuming that returns are normally distributed (which is not exactly the case, but for scaling purposes this is a usual practice), it can be shown that $RV^2 \propto \Delta t$, and volatility scaling rule takes the form

$$RV_{scaled} = \sqrt{\frac{\Delta t_{scale}}{\Delta t}} vol$$

Then, for example, volatility scaled to 1 year (using daily returns) is calculated as

$$RV_{scaled} = \sqrt{\frac{252}{1}} RV, \quad (3.3.1)$$

assuming 252 working days in a year.

Let us denote \widehat{RV}_t the value of the hidden volatility predicted by the model under study and RV_t the value of realized volatility calculated using intraday data and formula (3.3). Then the following error statistics are used to compare volatility forecasts:

Mean Error(ME)

$$ME = \frac{1}{N} \sum_{t=1}^N \left(\widehat{RV}_t - RV_t \right),$$

Mean Square Error (MSE)

$$MSE = \frac{1}{N} \sum_{t=1}^N \left(\widehat{RV}_t - RV_t \right)^2,$$

Root Mean Square Error (RMSE)

$$MSE = \sqrt{\frac{1}{N} \sum_{t=1}^N \left(\widehat{RV}_t - RV_t \right)^2},$$

Mean Absolute Error (MAE)

$$MSE = \frac{1}{N} \sum_{t=1}^N \left| \widehat{RV}_t - RV_t \right|,$$

Mean Absolute Percent Error (MAPE)

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{\left| \widehat{RV}_t - RV_t \right|}{RV_t},$$

In the next chapter, we will look at how to calibrate some of the volatility models mentioned earlier using empirical data (that is, using end-of-the-day stock returns). The above measures of error will be used in empirical investigation of models' performance in chapter 5 where realized volatility will be calculated from intra-day stock returns using 3.3.1 and compared with volatility predicted by different models.

Chapter 4

Parameter Estimation in Volatility Models

4.1 Numerical computation in Bayesian framework

Numerical techniques in Bayesian framework are applied when posterior distribution does not have a closed analytical form. As any inference in Bayesian statistics is made using the posterior density of the investigated parameter θ , this usually amounts to the computation of the posterior mean for some function of θ , namely, $E(g(\theta))$ (in the simplest and most common case $g \equiv x$):

$$E(g(\theta)|y) = \int g(\theta)p(\theta|y)d\theta,$$

where $p(\theta|y)$ is θ 's posterior distribution. In general case this integral could not be evaluated analytically. The *Central Limit Theorem* constitutes a theoretical basis for the numerical computation of the approximation to this integral. If we are able to obtain a sample $(\theta_1, \theta_2, \dots, \theta_M)$ from posterior distribution $p(\theta|y)$ and compute

$$\widehat{g_M(\theta)} = \frac{1}{M} \sum_{m=1}^M g(\theta_m),$$

then according to the central limit theorem it converges to $Eg(\theta|y)$ as M goes to infinity. This procedure is called *Monte Carlo integration*. From the statistical point of view the computed quantity $\widehat{g_M(\theta)}$ is nothing more than

a sample average. Theoretically the asymptotic variance of $\widehat{g_M(\theta)}$ is equal to $\frac{\sigma^2}{M}$, and an estimate to σ^2 could also be obtained from the simulation results:

$$s_M^2 = \sqrt{\frac{1}{M} \sum_{m=1}^M \left(g(\theta^{(m)}) - \widehat{g_M(\theta)} \right)^2}.$$

Precision of the Monte Carlo integration is often estimated using the value of *Monte Carlo Standard Error* (MCSE)

$$\text{MCSE} = \sqrt{\frac{s_M^2}{M}}. \quad (4.1.1)$$

It should be noted that direct simulation from the posterior distribution is not always possible. In such cases one should use special simulation algorithms.

4.2 Algorithms for posterior simulation

For practical reasons we can divide algorithms for posterior simulations into two groups:

1. Algorithms producing independent and identically distributed samples from the posterior distribution,
2. Algorithms, which in the convergence phase produce nearly identically distributed, but not independent samples from posterior distribution.

In fact, algorithms from the first group are used as building blocks for the second, and in practice a combination of algorithms from these two groups is used.

4.2.1 Markov chain Monte Carlo (MCMC) methods

It is not always possible to sample independent identically distributed values from a posterior density. Markov chain Monte Carlo is a collective terms used to denote a group of algorithms used to produce draws from a posterior distribution. These algorithms work iteratively and provide approximate samples that are not independent.

Algorithms of this group use iterative methods to *approximately* sample from complex posterior densities. Sample values of θ which are used to compute

sample averages are obtained as realisation of Markov chain. When using MCMC algorithms, chains of sufficient length should be produced to guarantee convergence to the posterior distribution.

We will discuss two widely used MCMC algorithms — Metropolis-Hastings [107, 75] and Gibbs sampler [35].

Metropolis – Hastings algorithm (M–H) As above, $p(\theta|y)$ is the posterior density, direct sampling from which is not possible, $\theta = (\theta_1, \dots, \theta_K)$ – K dimensional parameter vector. Denote by $q(\theta|\theta^{(t-1)})$ the approximating density (also called proposal density, candidate-generating density). The function of the proposal density is to recursively generate a sample from θ conditional on its value in the previous iteration. Each step of the algorithm consists of generating a draw from the proposal density and an accept/reject decision. Schematically, Metropolis – Hasting algorithm works in the following way:

1. Choose a starting value $\theta^{(0)}$.
2. Iteration number t . Draw a realization θ^* from the proposal density $q(\theta|\theta^{(t-1)})$, where $\theta^{(t-1)}$ is the value of parameter θ at the previous step of M–H algorithm.
3. Compute the acceptance probability:

$$a(\theta^*, \theta^{(t-1)}) = \min \left\{ 1, \frac{p(\theta^*)/q(\theta^*|\theta^{(t-1)})}{p(\theta^{(t-1)})/q(\theta^{(t-1)}|\theta^*)} \right\}, \quad (4.2.1)$$

4. Draw u from the uniform distribution $U(0, 1)$. Then,
 - If $u \leq a(\theta^*, \theta^{(t-1)})$, set $\theta^{(t)} = \theta^*$.
 - Otherwise, set $\theta^{(t)} = \theta^{(t-1)}$.

5. Go to step 2.

Note that in practical simulations one should use only the samples produced after the chain converged to its stationary state.

Three main versions of the M–H algorithm exist according to the method that is used to select a proposal density: *Random Walk M–H Algorithm* and *Independence Chain M–H Algorithm*.

Random Walk Metropolis – Hastings algorithm The general idea of Random Walk M–H is that if there is no idea on the good approximation to the posterior density, the solution is to construct a Markov Chain that will explore all areas of parameter space no matter high or low their posterior probability is. The states of such chain are obtained as

$$\theta^{(t+1)} = \theta^{(t)} + \epsilon^{(t+1)}, \quad (4.2.2)$$

where ϵ^{t+1} is a K -dimensional random variable with a chosen distribution q and zero mean. In most cases ϵ is chosen to have a normal distribution, in which case proposal distribution has the form

$$q(\theta^*|\theta^{(t-1)}) = N(\theta^{(t-1)}, \Sigma). \quad (4.2.3)$$

If the chosen proposal distribution q is symmetric, the formula for the acceptance probability has the simpler form:

$$a(\theta^*, \theta^{(t-1)}) = \min \left\{ 1, \frac{p(\theta^*)}{p(\theta^{(t-1)})} \right\}. \quad (4.2.4)$$

It could be easily seen from the definition above that the idea of this algorithm could be expressed in the following way: the proposed draw is always accepted when it has a higher posterior probability than the current draw (then $a = 1$), and accepted with probability a when its posterior probability is lower than that of the current draw. Realization simplicity of the Random Walk M–H is an obvious advantage of this algorithm. The drawback of this algorithm is that there is nothing in it that could prevent the chain from making large steps and getting into the regions of parameter space with low posterior probability, which, in turn, would lead to low acceptance probability and rejection of the significant portion of the proposed draws. On the other hand, when the steps get too small, the chain would stay in the same region of parameter space, although the acceptance probability would be high and most of the draws would be accepted. In practical implementation the jump size could be tuned by choosing a special value for the covariance matrix of the proposal distribution Σ .

Independence Chain Metropolis – Hastings algorithm While Random Walk M–H obtains the next candidate draw starting from the previous one, in Independence Chain M–H the next candidate draw doesn't depend on the current state of the chain (thus the name "Independence Chain"). Proposal distribution is usually set to multivariate normal or multivariate Student [63], [127], [66]. Selected proposal density is then centered and scaled

for it to fit the posterior. The following algorithm describes this "fitting" procedure mathematically.

1. Find the mode $\hat{\theta}$ of the posterior distribution. In most cases this is done by means of numerical optimization algorithms.
2. Compute the Hessian $H = H(\log p(\theta|y))|_{\theta=\hat{\theta}}$
3. In the case of the normal proposal density set it to

$$q(\theta|\theta^{(t-1)}) = q(\theta) = N(\hat{\theta}, -H^{-1}). \quad (4.2.5)$$

In the case of the Student proposal density set it to

$$q(\theta|\theta^{(t-1)}) = q(\theta) = t \left(\nu, \hat{\theta}, -H^{-1} \frac{(\nu - 2)}{\nu} \right). \quad (4.2.6)$$

Degree of freedom parameter ν allows to regulate the "heavy-tailedness" of the t-distribution – smaller ν give heavier tails.

Block Structure M–H Algorithm Sometimes it is not feasible to set the multivariate proposal distribution $q(\theta)$ for the whole parameter vector θ , but such distributions for certain blocks of θ make sense. Consider the simple case of partitioning θ in two blocks θ_i , $i = 1, 2$,

$$\theta = (\theta_1, \theta_2)$$

with conditional posterior densities

$$p_1(\theta_1|\theta_2, y) \quad \text{and} \quad p_2(\theta_2|\theta_1, y)$$

and proposal densities

$$q_1(\theta_1|\theta_1^{(t-1)}, \theta_2) \quad \text{and} \quad q_2(\theta_2|\theta_2^{(t-1)}, \theta_1).$$

In this settings the following modification of M-H algorithm is used.

1. Draw θ_1^* from the conditional proposal density $q_1(\theta_1|\theta_1^{(t-1)}, \theta_2^{(t-1)})$.
2. Accept or reject θ_1^* as in simple M-H.
3. Draw θ_2^* from the conditional proposal density $q_2(\theta_2|\theta_1^{(t)}, \theta_2^{(t-1)})$, where $\theta_1^{(t)}$ is the value of θ_1 from step 1 and $\theta_2^{(t-1)}$ is the value of θ_2 from the previous iteration of this algorithm.
4. Accept or reject θ_2^* as in simple M-H.

The block partitioning structure is often governed by the structure of the model itself (e.g., model parameters constitute one block and hyperparameters - another one).

The Gibbs sampler. This sampler constitutes a further development of the block-structure M–H and is used when one is able to sample directly from the full conditional posterior distributions of the blocks into which θ is partitioned. To employ Gibbs sampler the parameter vector θ is partitioned into s components:

$$\theta = (\theta_1, \theta_2, \dots, \theta_s).$$

Full conditional posterior distribution for the blocks θ_i , $i = 1, \dots, s$ is represented as

$$p(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_s, y) \equiv p(\theta_i | \theta_{-i}, y). \quad (4.2.7)$$

The steps of the Gibbs sampler are then the following ones:

1. Select initial values for all blocks: $\theta_i^{(0)}$, $i = 1, \dots, s$.
2. Iteration t :
 - Draw $\theta_1^{(t)}$ from $p(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_s^{(t-1)}, y)$.
 - Draw $\theta_2^{(t)}$ from $p(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_s^{(t-1)}, y)$.
 - Draw $\theta_3, \dots, \theta_s$ in the same way.
3. Repeat step 2 until the chain converges.

4.3 MCMC convergence diagnostics

Reliability of results obtained from posterior simulation procedures depends on whether the Markov chain has converged, which means that the simulated samples are indeed from the desired posterior distribution. Another desirable property of a simulated Markov Chain is "good chain mixing", which means that Markov chain explores the entire parameter space equally well. When the autocorrelations between successive samples in Markov chain are high and decay slowly, simulated chain values stay in the certain region of a parameter space for a large number of steps. Though, of course, undesirable, high correlation does not exclude convergence, but the latter takes more steps to reach.

Markov property of chains produced by MCMC algorithms means that after a number of steps influence of the starting values on the produced draws eventually vanishes. To minimize the effect of the chain's initial state, a certain number of simulation from the start up to a step t_{bip} , called a *burn in period*, is often discarded, and posterior inference is based on the the samples $\{\theta_t : t > t_{bip}\}$ only.

It is necessary to point out that the definition of convergence itself applied to the simulation using Markov chain Monte Carlo methods is fundamentally different from the usual concept of convergence for the iterative procedures in numerical analysis (e.g., convergence for Newton-type root finding algorithms for nonlinear equations). Indeed, the object that is produced by MCMC in its convergence phase is not a single number (or vector in multidimensional case) and not even a distribution, but a *sample* from distribution.

Research efforts in the area of MCMC convergence fall mainly into two groups [42]. Methods within the first group analyze the Markov transition kernel of the chain and try to predict a number of iterations that will guarantee convergence to the stationary distribution within a specified tolerance. For example, Roberts and Polson [126] investigated conditions under which Gibbs sampler converges at a geometric rate. Although such techniques have great theoretical importance, they require sophisticated mathematical calculation for each and every particular model under consideration. In addition to this, in most cases these theoretically obtained convergence bounds suggest numbers of iterations that could differ by several orders of magnitude.

Due to these evident drawbacks of purely theoretical approaches applied research that uses MCMC methods relies mainly on the second approach to detect convergence, which is based on the application of certain diagnostic procedures applied to the output of MCMC simulation algorithms.

4.3.1 Geweke diagnostics.

When the draws produced by Gibbs sampler are used to estimate the mean of some function g of the parameter θ that is being simulated, Geweke [67] proposed to use methods from spectral analysis. Computing values $g(\theta^{(j)})$ after each iteration of the Gibbs sampler, we obtain a sequence which could be regarded as time series. Important assumption of the Geweke's method is that MCMC simulation process and function g imply that the above constructed time series has a spectral density $S_g(\omega)$ that is continuous at $\omega = 0$. Under this assumption it is possible to construct an estimator for $E[g(\theta)]$ using n iterations of the Gibbs sampler:

$$\bar{g}_n = \frac{\sum_{i=1}^n g(\theta^{(i)})}{n}, \quad (4.3.1)$$

which variance is $S_g(0)/n$. The standard error of the mean (or “numeric standard error”, NSE, in terms of Geweke) could be obtained as a square root of this variance.

After n iterations of the Gibbs sampler Geweke's convergence diagnostic is calculated by taking the difference between $\bar{g}(\theta)_n^A$ (mean based on the first n_A iterations) and $\bar{g}(\theta)_n^B$ (mean based on the last n_B iterations) and dividing it by the asymptotic standard error of the difference (computed as above from spectral density estimates). For fixed ratios n_A/n and n_B/n such that $n_A + n_B < n$, applying the central limit theorem, one can get that distribution of Geweke's diagnostic approaches normal when $n \rightarrow \infty$. As to the ratios n_A/n and n_B/n , Geweke suggested using $n_A/n = 0.1$ and $n_B/n = 0.5$. According to Geweke, the test is used to decide how many initial iterations to discard. After discarding, MCMC simulation is run for the number of iterations sufficient to get the desired precision (given by NSE).

4.3.2 Gelman and Rubin convergence diagnostics

Gelman and Rubin ([64]) proposed a convergence test that uses normal theory approximations to exact Bayesian posterior inference. Full Gelman and Rubin diagnostic procedure employs sampling from several Markov chains and consists of two parts. At part 1, that is before any sampling is made, one should obtain overdispersed starting values for the desired number of chains. After running chain simulations for $2n$ steps, part 2 uses the last n iterations to approximate the target distribution of the scalar parameter (which distribution we are trying to approximate with the Markov Chain) with a Student's t -distribution with the scale parameter based on between-chain and within-chain variance. Convergence criterium is based on the estimation of the scale parameter shrinking factor which is given by the statistic

$$\sqrt{\widehat{R}} = \sqrt{\left(\frac{n-1}{n} + \frac{m+1}{mn} \frac{B}{W}\right) \frac{df}{df-2}}, \quad (4.3.2)$$

where B is the variance between the mean values of the m parallel chains, W is the average of the m within-chain variances, df is the number of degrees of freedom of the approximating t distribution. According to Gelman and Rubin step 2 should be repeated until "shrink factors" for all scalar parameters of interest are close to 1. Although Gelman and Rubin initially proposed this technique as a convergence estimator for the Gibbs sampler, it could be successfully applied to the chains produced by any MCMC algorithm. Gelman and Rubin approach is heavily criticised for its reliance on the user's ability to find overdispersed starting distributions, which, in turn, requires some knowledge about the target distribution. Secondly, the reliance on the normal approximation to the target distribution is also a weak point since

MCMC simulation is used mostly in cases when target distribution is not normal.

4.3.3 Heidelberg and Welch diagnostics.

Heidelberg and Welch [76] proposed a procedure to construct a confidence interval for the mean when the Markov chain (from which it was estimated) has an “initial transient period”, that is, when at the beginning of the simulation process the chain has a distribution other than the desired stationary one. First part of Heidelberg and Welch diagnostics is based on the results of Schruben [136] and Schruben et al. [135]. The null hypothesis is that a sequence of iterates is from a stationary ψ mixing process. Applied to markov chains, this condition is equivalent to uniform ergodicity. According to Meyn and Tweedie [109], this approach is also applicable to geometrically ergodic chains, and, consequently, for chains produced by Gibbs and Metropolis-Hastings algorithms. If $\theta^{(j)}$ is a j th iterate in the chain, $S(0)$ is a spectral density of the chain evaluated at 0, n is the total number of iterations, and

$$T_0 = 0, \quad (4.3.3)$$

$$T_k = \sum_{j=1}^k \theta^{(j)}, \quad k \geq 1, \quad (4.3.4)$$

$$\bar{\theta} = \frac{\sum_{j=1}^n \theta^{(j)}}{n}, \quad (4.3.5)$$

and

$$B_n(t) = \frac{T_{[nt]} - [nt]\bar{\theta}}{\sqrt{nS(0)}}, \quad 0 \leq t \leq 1, \quad (4.3.6)$$

where $[\cdot]$ denotes the largest integer not greater than \cdot , then for large n under the null hypothesis $B_n = \{B_n(t), 0 \leq t \leq 1\}$ is approximately distributed as a Brownian bridge and the Cramer-von Mises statistic

$$\int_0^1 B_n(t)^2 dt \quad (4.3.7)$$

could be used to test the hypothesis. The unknown quantity $S(0)$ is estimated from the data and plugged into the formula for $B_n(t)$.

Heidelberg and Welch method consists of the following steps. Initially, two parameters are specified: the maximum number of iterations j_{max} and desired half-width for confidence intervals ε . $j_1 = 0.1j_{max}$ iterates of the chain are

produced and the above described Schruben's stationarity test is conducted ($S(0)$ is estimated from second half of this chain in order to diminish the influence of the initial transient period). If the null hypothesis is rejected, first 10% of the iterates are discarded and stationarity test is conducted for the remaining part of the chain. This process is repeated until either a chain of length not less than $0.5j_1$ is found that passes this stationarity test or the remaining chain contains less than 50% of original iterates (but the test still fails). If the test is passed, then the standard error of the mean is estimated as $\sqrt{\hat{S}(0)}/n_p$, where n_p is a remained chain's length. Process stops when the half width of the confidence interval obtained in this way is less than ε times the sample mean of the remained chain.

If the stationarity test was failed or the confidence interval is too wide, all the discarded iterates are restored in the chain and simulations continue to obtain a chain with the length $j_2 = 1.5j_1$. New longer chain is again tested using stationarity test and confidence interval width. This process could be repeated with the increasing chain lengths $j_{k+1} = \min(1.5j_k, j_{\max})$ until either a confidence interval of the necessary width is obtained or the chain length reaches j_{\max} .

Chapter 5

Empirical investigation of discrete SV models using MCMC

5.1 Description of the data

For the purposes of empirical investigation we randomly picked 5 companies from the list of the most actively traded stocks on the NASDAQ stock exchange:

1. Cisco Systems, Inc. (ticker symbol CSCO),
2. Microsoft Corporation (ticker symbol MSFT),
3. Intel Corporation (ticker symbol INTC)
4. NVIDIA Corporation (ticker symbol NVDA)
5. Dell Inc. (ticker symbol DELL)

Exchange trading data was obtained from the DTNiQ data feed (<http://www.dtniq.com/>) as a set of records with the following fields for each 1 minute period:

1. Time stamp,
2. Opening price (“Open”),
3. Highest price (“High”),
4. Lowest price (“Low”),

5. Closing price (“Close”),
6. Trading volume for the current period (“IncVol”),
7. Cumulative trading volume (“Volume”, counting from the start of the trading day).

Data covers the period from 01/03/2011 to 31/05/2011. This period selection was arbitrary, since the purpose was to gain insight into the MCMC methodology and to see if adding transaction volume as a descriptor variable makes any difference. The length was chosen so that, on the one hand it contained enough data to perform Bayesian estimation, and, on the other hand, to keep the computation time reasonable. Daily prices were converted to the continuously compounded returns (3.1.2). Sample skewness and kurtosis are given in table 5.1. Based on the kurtosis’ values greater than 3 we can suppose that only Microsoft and Dell returns could have normal distribution. Results of Shapiro – Wilk and Jarque – Bera normality tests (table 5.1) also indicate that only Microsoft and Dell returns have normal distribution.

	CSCO	MSFT	INTC	NVDA	DELL
Skewness	0.18726	0.04493	1.4928	-0.8019	0.4388
Kurtosis	3.73830	-0.0900	6.5467	2.432	0.3612

Table 5.1: Data characteristics: Second order moments

	CSCO	MSFT	INTC	NVDA	DELL
Shapiro - Wilk	0.0015	0.7801	0.0000	0.0112	0.3479
Jarque - Bera	0.0000	0.9889	0.0000	0.0000	0.259

Table 5.2: Data characteristics: Normality tests, p-values

We also applied Box–Pierce [29] and Ljung–Box [96] portmanteau tests which showed that autocorrelations for lags from 1 to 10 for all five time series are statistically not different from 0. Since autocorrelations are 0, there is no need to construct a trend model, and we applied a normalization transformation only, that is our time series were transformed to

$$r_t^{norm} = \frac{r_t - \hat{\mu}}{\hat{\sigma}}, \quad (5.1.1)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the sample mean and sample standard deviation respectively.

The following discrete stochastic volatility models will be tested:

1. Classical SV model with normal innovations,
2. SV model with t-innovations,
3. SV model with normal innovations and volatility process augmented with volume,
4. SV model with t-innovations and volatility process augmented with volume.

Model evaluation will be done in sample by comparing one step ahead volatility forecast produced by each of the mentioned above models 5.1 and realized volatility calculated for that day using 5-min intraday data according to formula (3.3) with $p = 2$. The forecasts will be evaluated using the error metrics discussed in section 3.3 on page 37.

5.2 Bayesian analysis of Classical SV model

5.2.1 Theoretical model

In this section we will give a general scheme for the implementation of the Bayesian analysis for “classic” stochastic volatility model

$$h_t = \mu + \phi(h_{t-1} - \mu) + u_t, \quad u_t \sim N(0, \tau^2), \quad (5.2.1)$$

where μ , ϕ , τ are model parameters and with innovations in the form

$$\varepsilon_t = e^{\frac{h_t}{2}} z_t, \quad z_t \sim N(0, 1). \quad (5.2.2)$$

To calibrate this model we will use the normalized returns from (5.1.1), that is $\varepsilon_t \equiv r_t^{norm}$.

Application of the classical frequentist parameter estimation methods to the SV models presents a difficult task since its likelihood function doesn't have a tractable form. Some of the classical methods have been applied to the estimation of SV models: generalized method of moments was used by Melino and Turnbull [106] and Sorensen [142]; quasi-maximum likelihood by Harvey, Ruiz and Shephard [73]; efficient method of moments by Gallant, Hsieh and Tauchen [60]; simulated maximum likelihood by Danielsson [44], Sandmann and Koopman [134]; approximate maximum likelihood by Fridman and Harris [57]. Parameter estimation for SV models using MCMC was proposed by Jacquier, Polson and Rossi [82], Shephard and Pitt [140], Kim, Shephard

and Chib [88]. Andersen, Chung and Sorensen [9] compared MCMC with other estimation tools in Monte Carlo framework.

As a framework for implementing MCMC computation we will use JAGS (Just Another Gibbs Sampler) [117] – one of the implementations of BUGS (Bayesian analysis Using Gibbs Software) modelling software [99]. As an example of SV estimation using this approach one can refer to the paper by Meyer and Yu [108].

There exists a standard tool that can be used in combination with BUGS software that implements some standard tests for MCMC convergence diagnostics - a CODA package [118], that could be called from R statistical software suite [122].

We can regard volatility specification written in the form (5.2.2) – (5.2.1) as a nonlinear state-space model [74]. In our case it is a state-space model for the market returns r_t (observations) with hidden volatility process h_t (states), the hidden state h_t determines the value of the volatility on day t , and the parameter ϕ could be interpreted as *volatility persistence* (see financial time series stylized facts outlined in section 3.1.3.4).

As above, we denote the vector of model parameters by θ :

$$\theta = (\mu, \phi, \tau^2).$$

The likelihood function for θ could be written as

$$L(\theta|r) = \prod_{t=1}^T \int f(r_t|\sigma_1^2, \dots, \sigma_T^2) f(\sigma_1^2, \dots, \sigma_T^2|\theta) d\sigma_1^2 \dots d\sigma_T^2 \quad (5.2.3)$$

Bayesian computation in this case amounts to the simulations (sampling) of hidden volatilities from their conditional distribution with model parameters.

5.2.1.1 Practical convergence assessment example

Before making any inference on model parameters, it is necessary to carry out MCMC convergence test and ensure that chain has reached a stationary distribution. In our usage of diagnostic routines implemented in CODA we will follow the extensive description [20]. Originally CODA (convergence diagnostics and output analysis) was a set of functions implemented in S-Plus to serve as an output processor for BUGS-type software. We used CODA implementation in R programming language (available as a package for R system). It is important to stress that MCMC convergence tests implemented

in CODA do not guarantee the convergence of the chain under study – it is the absence (or lack) of convergence that they can reliably indicate. We will give commented examples for the Cisco time series. For all other data sets given below, it is assumed that convergence phase has been reached.

Visual inspection of plots There are some visual tools that could help to see how well the chain is *mixing*. The term *mixing* refers to the movements of the chain around the parameter space. If due to the nature of the sampling process the chain stays in the same area of parameter space and, consequently, requires longer time to explore the whole parameter space, then such chain is called poorly mixing. The first aid in exploring the mixing property is a *traceplot* – a plot of the parameter draw’s value against the number of the iteration. Function `plot.mcmc` produces a graphical summary of the iterates for each monitored variable and chain in the JAGS output. Figure 5.1 shows plots of the MCMC chains for SV model parameters (μ , ϕ) with trends superimposed on them.

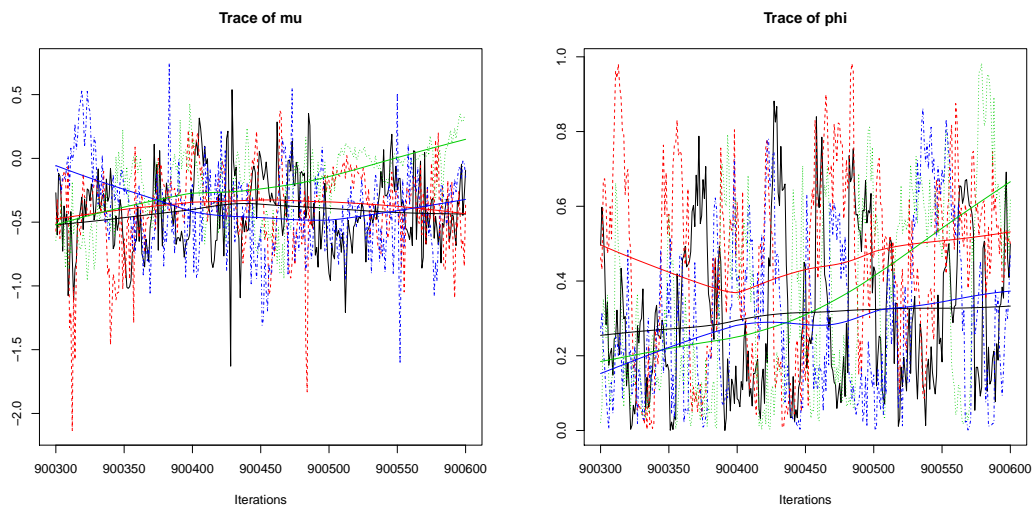


Figure 5.1: MCMC chains for the μ and ϕ parameters of the classic SV model

Chain statistics Function `summary.mcmc` produces summary statistics for each monitored variable and chain in the BUGS output. We will use the sample mean as an estimate for each parameter of interest in our model.

Geweke diagnostics Function `geweke.plot` generates graphical representation of Geweke’s diagnostics (Figure 5.2). in the following way. The entire chain for each variable is split into a number of segments. For a chain of length N , the first segment contains all N samples in the chain; the second

contains the last $N-n$ samples (i.e. iterations $n+1-N$); the third contains samples from iterations $2n+1-N$ and so on. The final segment is chosen so that it contains at least the last 50 samples. The default bin size is $n=10$ for chains of 500 iterations or less. For longer chains, the bin size is determined by splitting the chain equally into 50 bins. However these sizes may be changed by the user. Geweke's diagnostic is computed for each segment, and the resulting Z-scores are plotted against the number of the first iteration in the segment. Horizontal dotted lines at $Z = \pm 1.96$ are added to the plot to indicate the 95% confidence interval for a $N(0, 1)$ distribution. A large number of Z-scores falling outside this interval suggests possible convergence failure. These plots may take some time to produce due to the large number of computations needed to produce them, but the process may be speeded up by reducing the total number of bins, or increasing the bin size.

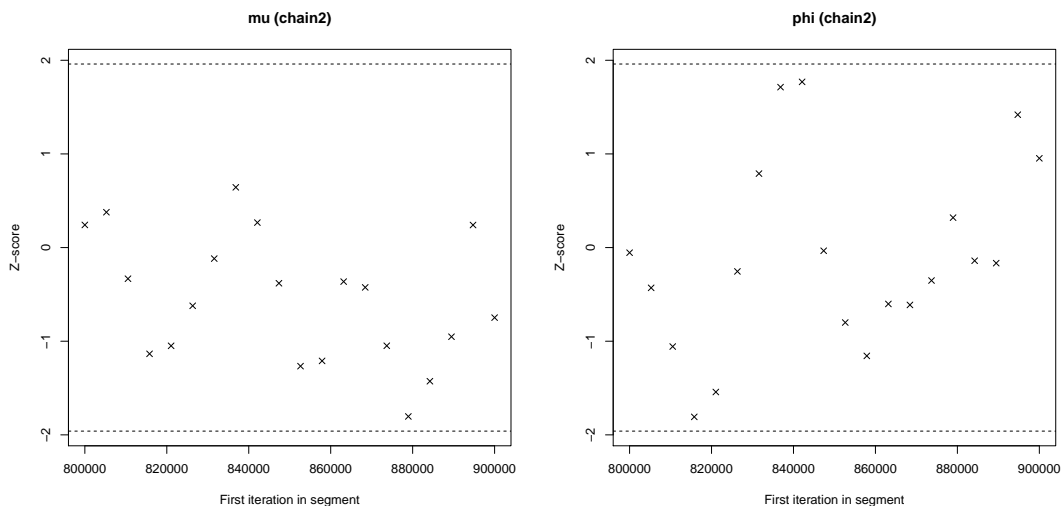


Figure 5.2: Plot of Geweke's diagnostics for the μ and ϕ parameters of the classic SV model

Gelman and Rubin convergence diagnostics Function `gelman.plot` produces plots of Gelman and Rubin diagnostics for each variable. (Figure 5.3). These plots are produced by splitting the chain for each variable into a number of segments as follows: the first contains samples 1:50; the second contains samples 1:(50 + n); the third contains samples 1:(50 + 2n) and so on. The default bin size is $n=10$ for chains of 500 iterations or less. For longer chains, the bin size is determined by splitting the chain equally into 50 bins. Gelman and Rubin's diagnostic is computed for each segment, and the median and 97.5% quantile of the sampling distribution for the result-

ing shrink factor are plotted against the maximum iteration number for the segment.

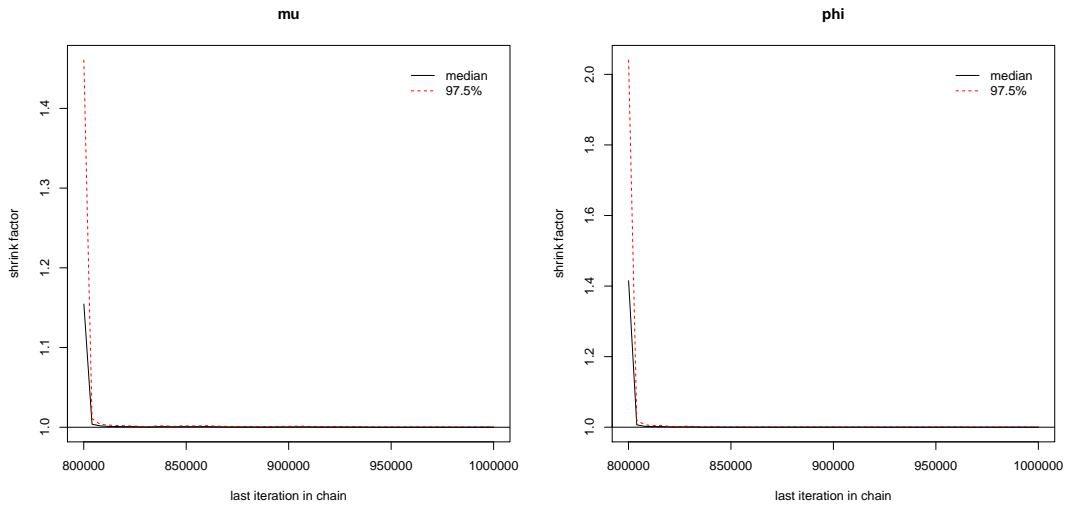


Figure 5.3: Plot of Gelman and Rubin diagnostics for the μ and ϕ parameters of the classic SV model

Heidelberger and Welch diagnostics Function `heidel.diag` implements a method devised by Heidelberger and Welch ([76]) for detecting an initial transient in simulated sequences of discrete events. If the stationarity test is passed, CODA reports the number of iterations to keep (i.e. which are diagnosed to arise from a stationary process), the number of initial iterations to discard and the Cramer-von-Mises statistic. Below is a typical test output indicating convergence.

```
[[1]]
```

	Stationarity test	start iteration	p-value
mu	passed	1	0.0701
beta	passed	1	0.0609
phi	passed	1	0.2699
tau2	passed	1	0.1090
itau2	passed	1	0.0552

	Halfwidth Mean test	Halfwidth

mu	passed	-0.444	0.00774
beta	passed	0.810	0.00323
phi	passed	0.337	0.00396
tau2	passed	0.760	0.01578
itau2	failed	4.731	2.49332

These convergence tests were carried out for all the models that we discuss further in the thesis and demonstrated suitable convergence. The parameter values and the error metrics obtained for this model, along with parameter values and error metrics for other volatility models are listed and compared in section 5.4.

5.3 Bayesian techniques for stochastic volatility models with t-innovations

It is a stylized fact of the financial time series behaviour that the returns distributions are non-normal and have heavy tails. For two out of five data sets that we investigate in this work hypothesis of normal distribution was rejected by normality tests. One of the techniques that could be used to account for heavy tails is to use Student distribution for the innovations in stochastic volatility model, that is to specify it as

$$r_t = \mu + \varepsilon_t, \quad t = 1 \dots T, \quad (5.3.1)$$

with innovations in the form

$$\varepsilon_t = e^{\frac{h_t}{2}} z_t, \quad z_t \sim t(k), \quad (5.3.2)$$

where k is a number of degrees of freedom, and volatility process specification is the same as for the classical model:

$$h_t = \mu + \varphi(h_{t-1} - \mu) + u_t, \quad u_t \sim N(0, \tau^2). \quad (5.3.3)$$

Modelling returns with t-distribution introduces a new parameter – number of degrees of freedom k – that needs to be estimated. Theoretically for the Student distribution with k degrees of freedom its excess kurtosis could be computed analytically and equals

$$K(k) = \frac{6}{k-4}. \quad (5.3.4)$$

On the one hand, this could give a convenient way of estimating the degree of freedom parameter of a t-distribution from empirical data. But on the other hand, sample kurtosis could not be viewed as a reliable estimate of population kurtosis (Mosteller and Tukey [110], [13]). There also exists a MCMC modification that allows to use k as a parameter in the estimation procedure, but it is found to make the procedure prohibitively expensive in terms of computation. Hence we decided to use a simulation approach and estimated the model with different degrees of freedom assumptions, namely, we tested the simulation performance of the model for $k = \{3, 4, 5, 6, 7, 8\}$.

	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$
μ	-0.8987	-0.7682	-0.6883	-0.6279	-0.5902	-0.5651
$q_{2.5\%}$	-1.4330	-1.3050	-1.2148	-1.1786	-1.1314	-1.1225
$q_{97.5\%}$	-0.3633	-0.2453	-0.1738	-0.1085	-0.0801	-0.0464
ϕ	0.4600	0.4531	0.4455	0.4365	0.4291	0.4229
$q_{2.5\%}$	0.0220	0.0216	0.0206	0.0199	0.0196	0.0192
$q_{97.5\%}$	0.9571	0.9549	0.9537	0.9470	0.9418	0.9410
MSE	0.0973	0.1105	0.1193	0.1262	0.1306	0.1335
RMSE	0.3119	0.3324	0.3453	0.3552	0.3614	0.3654
MAE	0.1621	0.1725	0.1791	0.1840	0.1870	0.1890
MAPE	0.2272	0.2266	0.2260	0.2253	0.2247	0.2242

Table 5.3: Parameters of SV model with t innovations

It could clearly be seen that the all error measures increase with the increase of the degrees of freedom parameter k . Based on these results, in all other models with t-innovations degrees of freedom parameter is set to $k = 3$.

5.4 Bayesian techniques for stochastic volatility models with trading volume

Investigation of the relationship between stock return volatility and trading volume could find many applications in financial economics, the most important being dynamic portfolio allocation and option pricing. Clark ([40]) proposed Mixture of Distributions Hypothesis (MDH): stock returns and trading volume depend on the same underlying latent variable – the flow of information relevant to the trading process itself. Research findings by Epps and Epps ([40]), Tauchen and Pitts ([40]), Harris ([72, 71]) agree with MDH. On the other hand, findings by Lamoreux and Lastrapes ([92]), Richard-

son and Smith([125]), Liesenfeld([95]) draw attention to some weak points of this theory. Richardson and Smith([125]) pointed out that using *normal distribution* to relate returns and trading volume with information flow may not be the right specification. According to Lamoreux and Lastrapes ([92]) mixing variable (judging by its time series behaviour) does not fully explain volatility persistence in financial time series.

Substantial number of papers is devoted to the relationship between return volatility and trading volume in the field of market microstructure research. In this area of research it is usually assumed that arrival of new information causes price movements. Market microstructure theory mostly explores intra-day relationship between volatility and volume. Andersen ([6]) tries to combine MDH and market microstructure theory and create empirical model of the daily return using market microstructure models proposed in papers by Glosten and Milgrom ([69]), Kyle ([91]), Admati and Pfleiderer ([2, 1]). MDH and stochastic volatility are combined with main features of microstructure models – asymmetric information structure, liquidity, noise traders. The resulting model is called Modified Mixture Model (MMM) and is specified as AR(1) stochastic volatility process for the latent information arrival by Andersen ([5]).

Mahieu [101] took a Bayesian approach to perform simulations for modified mixture model (MMM) in the framework proposed by Andersen [6].

All the studies mentioned above proposed models with trading volume contemporaneous to the returns, which makes these models not suitable for prediction purposes. We are trying to establish a model that uses lagged volume and could be useful in volatility prediction. In particular, we are constructing a stochastic volatility model which incorporate the trading volume from the previous trading day as exogenous factor.

We propose the following modifications of classical stochastic volatility model:

1. SV with Gaussian innovations

$$\varepsilon_t = e^{\frac{h_t}{2}} z_t, \quad z_t \sim N(0, 1), \quad (5.4.1)$$

and volatility process with trading volume lagged for one day

$$h_t = \mu + \varphi(h_{t-1} - \mu) + \psi \log v_{t-1} + u_t, \quad (5.4.2)$$

where $\log v_{t-1}$ is a natural logarithm of the lagged trading volume.

2. SV with t-innovations

$$\varepsilon_t = e^{\frac{h_t}{2}} z_t, \quad z_t \sim t(k), \quad (5.4.3)$$

and volatility process with trading volume lagged for one day

$$h_t = \mu + \varphi(h_{t-1} - \mu) + \psi \log v_{t-1} + u_t. \quad (5.4.4)$$

3. SV with Gaussian innovations

$$\varepsilon_t = e^{\frac{h_t}{2}} z_t, \quad z_t \sim N(0, 1), \quad (5.4.5)$$

and volatility process with trading volume lagged for two days

$$h_t = \mu + \varphi(h_{t-1} - \mu) + \psi \log v_{t-2} + u_t. \quad (5.4.6)$$

4. SV with t-innovations

$$\varepsilon_t = e^{\frac{h_t}{2}} z_t, \quad z_t \sim t(k), \quad (5.4.7)$$

and volatility process with trading volume lagged for two days

$$h_t = \mu + \varphi(h_{t-1} - \mu) + \psi \log v_{t-2} + u_t. \quad (5.4.8)$$

Table 5.4 – 5.4 give the results of model estimation for different data sets. We use the following abbreviations to denote the models:

1. *SV* – classic SV model,
2. *SV-t* – SV model with t-innovations,
3. *SV_VOL* – SV model with Gaussian innovations augmented with lag one trading volume,
4. *SV_VOL-t* – SV model with t-innovations augmented with lag one trading volume,
5. *SV_VOL-2* – SV model with Gaussian innovations augmented with lag two trading volume,
6. *SV_VOL-t-2* – SV model with t-innovations augmented with lag two trading volume.

As a measure of parameter's estimate significance in Bayesian framework we use the position of the interval $(q_{2.5\%}, q_{97.5\%})$ – if this interval contains 0, then the parameter is not significant. From the tables 5.4 – 5.4 we can see that trading volume does not add any extra useful information to the models, since its coefficient ψ in all models is not significant. This conclusion holds

	SV	SV-t	SV_VOL	SV_VOL-t	SV_VOL-2	SV_VOL-t-2
μ	-0.4410	-0.8987	-0.4552	-0.8775	-0.4229	-0.8570
signif	0	1	0	1	0	1
$q_{2.5\%}$	-1.0547	-1.4330	-1.2600	-1.422	-1.469	-1.4018
$q_{97.5\%}$	0.1490	-0.3633	0.1518	-0.3347	0.2633	-0.2982
ϕ	0.3364	0.4600	0.3164	0.3877	0.4723	0.4378
signif	1	1	1	1	1	1
$q_{2.5\%}$	0.0139	0.0220	0.0107	0.0157	0.0164	0.0173
$q_{97.5\%}$	0.8097	0.9571	0.9600	0.9360	0.9884	0.9611
ψ	N/A	N/A	0.1167	0.0713	-0.0773	-0.0014
signif			0	0	0	0
$q_{2.5\%}$			-0.3870	-0.3394	-0.4470	-0.3604
$q_{97.5\%}$			0.6079	0.9360	0.4272	0.4421
MSE	0.1466	0.0973	0.1462	0.0979	0.1569	0.1030
RMSE	0.3829	0.3119	0.3823	0.3129	0.3961	0.3210
MAE	0.1983	0.1621	0.1949	0.1606	0.2070	0.1705
MAPE	0.2209	0.2272	0.2219	0.2245	0.2271	0.2362

Table 5.4: Parameter estimates and error statistics for Cisco

	SV	SV-t	SV_VOL	SV_VOL-t	SV_VOL-2	SV_VOL-t-2
μ	-0.0268	-0.5078	-0.0265	-0.5122	-0.0018	-0.48080
signif	0	0	0	0	0	0
$q_{2.5\%}$	-0.4444	-1.018	-0.4537	-1.0059	-0.4084	-0.9781
$q_{97.5\%}$	0.3966	0.018	0.3942	-0.002	0.4091	0.02947
ϕ	0.4561	0.4571	0.4142	0.3939	0.3726	0.3788
signif	1	1	1		1	
$q_{2.5\%}$	0.0204	0.0211	0.0167	0.0163	0.0119	0.0124
$q_{97.5\%}$	0.9552	0.9574	0.9470	0.9357	0.9251	0.9401
ψ	N/A	N/A	-0.0316	-0.0270	-0.0995	-0.1189
signif			0	0	0	
$q_{2.5\%}$			-0.2971	-0.3938	-0.4849	-0.5664
$q_{97.5\%}$			0.2538	0.3500	0.2783	0.2891
MSE	0.1334	0.0825	0.1285	0.0827	0.1385	0.0886
RMSE	0.3652	0.2872	0.3585	0.2876	0.3722	0.2979
MAE	0.2705	0.2127	0.2649	0.2164	0.2800	0.2255
MAPE	0.2715	0.2715	0.2666	0.2794	0.2802	0.2878

Table 5.5: Parameter estimates and error statistics for Microsoft

for both lags two and one. Another fact that could be clearly seen for all models is that models with t-innovations provide more accurate forecasts.

We also performed an out of sample forecast evaluation using the trading data of the Cisco Systems, Inc for the period 01/06/2011 – 31/08/2011. Results of this evaluation, which could be found in Table 5.4 and also involve comparison with the GARCH(1,1) model, show that SV model with t-distributed innovations gives the most accurate forecast.

	SV	SV-t	SV_VOL	SV_VOL-t	SV_VOL-2	SV_VOL-t-2
μ	-0.3319	-0.7422	-0.3482	-0.7685	-0.2938	-0.7409
signif	0	1	0	1	0	1
$q_{2.5\%}$	-0.9287	-1.2469	-0.9170	-1.2492	-0.9092	-1.2319
$q_{97.5\%}$	0.2318	-0.2242	0.1947	-0.2782	0.2807	-0.2425
ϕ	0.4320	0.4759	0.3346	0.3377	0.4215	0.3732
signif	1	1	1		1	
$q_{2.5\%}$	0.0200	0.0231	0.0111	0.0140	0.0170	0.0174
$q_{97.5\%}$	0.9222	0.9674	0.9243	0.8554	0.9553	0.8884
ψ	N/A	N/A	0.1427	0.2471	-0.0082	0.1535
signif			0	0		0
$q_{2.5\%}$			-0.2453	-0.1453	-0.3127	-0.1777
$q_{97.5\%}$			0.6361	0.7501	0.3854	0.5364
MSE	0.2099	0.1418	0.2132	0.1522	0.2183	0.1461
RMSE	0.4582	0.3765	0.4618	0.3901	0.4672	0.3823
MAE	0.2203	0.1796	0.2269	0.1980	0.2285	0.1908
MAPE	0.2311	0.2310	0.2420	0.2633	0.2370	0.2476

Table 5.6: Parameter estimates and error statistics for Intel

	SV	SV-t	SV_VOL	SV_VOL-t	SV_VOL-2	SV_VOL-t-2
μ	-0.2028	-0.7069	2.7756	0.6701	0.006	0.1650
signif	0					
$q_{2.5\%}$	-0.7589	-1.2056	-11.93	-19.94	-15.21	-15.86
$q_{97.5\%}$	0.3104	-0.2032	20.17	13.83	18.82	20.26
ϕ	0.3473	0.4431	0.1862	0.2185	0.2169	0.2208
signif						
$q_{2.5\%}$	0.0143	0.0202	0.0064	0.006	0.007	0.0063
$q_{97.5\%}$	0.8765	0.9535	0.5727	0.6705	0.6173	0.6364
ψ	N/A	N/A	-0.1400	-0.0498	-0.0378	-0.0742
signif						
$q_{2.5\%}$			-1.074	-0.753	-1.046	-1.172
$q_{97.5\%}$			0.6598	1.0907	0.7525	0.7506
MSE	0.1004	0.0618	9.7898	0.48113	0.1859	0.3991
RMSE	0.3169	0.2487	3.1288	0.6936	0.4311	0.6317
MAE	0.2177	0.1686	2.7752	0.4935	0.2837	0.5015
MAPE	0.2397	0.2387	0.6657	0.2979	0.2396	0.4154

Table 5.7: Parameter estimates and error statistics for NVIDIA

	SV	SV-t	SV_VOL	SV_VOL-t	SV_VOL-2	SV_VOL-t-2
μ	-0.0703	-0.5461	-21.46	-11.36	-1.888	-10.43
signif						
$q_{2.5\%}$	0.5548	-1.0423	-34.27	-28.68	-15.55	-34.07
$q_{97.5\%}$	0.3686	-0.0237	-8.408	1.400	11.40	10.86
ϕ	0.4778	0.4618	0.1766	0.2552	0.2138	0.1394
signif						
$q_{2.5\%}$	0.0243	0.0218	0.0194	0.0195	0.006	0.0031
$q_{97.5\%}$	0.9621	0.9579	0.3470	0.6255	0.6772	0.5946
ψ	N/A	N/A	1.0564	0.5064	0.1009	0.5598
signif						
$q_{2.5\%}$			0.3701	-0.0831	-0.5502	-0.5694
$q_{97.5\%}$			1.779	1.418	0.8632	1.9247
ME	-0.0101	-0.0079	0.1664	0.2414	0.5179	0.6146
MSE	0.1475	0.0911	0.0292	0.0598	0.2927	0.3848
RMSE	0.3841	0.3019	0.1711	0.2445	0.5410	0.6203
MAE	0.2346	0.1838	0.1664	0.2414	0.5261	0.6146
MAPE	0.2343	0.2330	8314	77.69	1.5649	124.79

Table 5.8: Parameter estimates and error statistics for Dell

	SV	SV-t	SV_VOL	SV_VOL-t	SV_VOL-2	SV_VOL-t-2	GARCH(1,1)
MSE	0.3318	0.2015	0.3152	0.2053	0.3358	0.2158	0.7366
RMSE	0.5760	0.4489	0.5615	0.4531	0.5795	0.4645	0.8925
MAE	0.2648	0.2001	0.2605	0.2061	0.2608	0.2131	0.5623
MAPE	0.2919	0.2679	0.2907	0.2785	0.2770	0.2872	0.6371

Table 5.9: Parameter estimates and error statistics for Cisco (out of sample)

Chapter 6

Summary and future work

6.1 Contributions

Parameters of different modifications of SV models were estimated using Markov chain Monte Carlo methods. As calibration data we used stock prices of five companies listed on NASDAQ stock exchange: Cisco, Microsoft, Intel, NVIDIA and Dell. Data was acquired through the DTNiQ data feed (<http://www.dtniq.com/>) and used for calibration and realized volatility calculation. Upon examination of the sample skewness and kurtosis of the data Shapiro-Wilk and Jarque-Bera normality test were performed which indicated that distribution of the returns is not normal and has heavy tails. Box-Pierce and Ljung-Box portmanteau tests did not reject the hypothesis of zero autocorrelation in returns' time series. Calibration of SV models augmented with lagged trading volume and evaluation of their forecasting power by comparison with realized volatility constitutes the main contribution of the thesis, since most of the widely cited studies in this field dealt either with contemporaneous volume or employed GARCH models [32, 92, 6, 101].

Six variants of the SV model were calibrated using the above described data:

- classic SV model,
- SV model with innovations having Student distribution,
- SV model with Gaussian innovations augmented with lag one trading volume,
- SV model with t-innovations augmented with lag one trading volume,

- SV model with Gaussian innovations augmented with lag two trading volume,
- SV model with t-innovations augmented with lag two trading volume.

Calibration results showed that Markov chain Monte Carlo methods allow to get reliable parameter estimates for discrete SV models, although the estimation process is quite demanding computationally. On average, it takes about 1.5 hours on Intel Core 2 Duo Processor T7300 (2.0 GHz, 4MB L2 cache, 800MHz FSB), which could pose a problem for applications where frequent updates of parameters are needed.

Classic SV model was used to explore the influence of the distributional assumptions on the innovations on the forecasting performance and it was demonstrated that the most reliable forecasts are obtained with the innovations having t distribution with 3 degrees of freedom.

Forecasting power of volatility models was assessed by comparing the predicted volatility with realized volatility. For this purpose realized volatility for each day using intraday 5 minute trading data was computed and then compared for every day with one step ahead forecast obtained from each of the SV models. As an integral measure of forecast quality different functions of forecasting error were used.

Our experiments indicated that trading volume may not be useful as an exogenous factor providing extra information for discrete stochastic volatility models. This confirms the common belief that all information necessary for trading purposes is already contained in prices. Further, the experiments indicate that a stochastic volatility model with t-innovations appears to perform better than the one with Gaussian innovations. These results are in agreement with the findings for the GARCH models augmented with volume obtained by Brooks [32], who reported that augmenting of GARCH models with lagged trading volume leads only to very modest improvements in their forecasting performance.

6.2 Future work

Possible modifications to the classical SV model usually include heavy tails in the distribution of returns, leverage effect and jump components in the returns equation; see e.g. [39], [83],[157], [114], [19], [111]. One of the popular methods to account for heavy tails is to use Student distribution to model

return – an approach explored in the thesis. However, employing only different specification for the returns’ distribution is often considered insufficient and jumps in price process were introduced to explain the fat-tail behaviour [54]. Several studies compare general performance of SV models with jumps, among them [53], [38], [124], [37], [100]. Hence, investigation of SV models with leverage effects and jumps in returns’ process constitutes one the possible directions for future work.

As a second direction we could propose testing our main conclusion about volume in SV models using trading data from different markets, since empirical price-volume relationships could be different among stock markets. It is possible that, in informationally less efficient emerging markets, lagged volume does have predictive power. However, the intra-day data for emerging markets is harder to access.

As a third direction, there could be different approaches to calculate realized volatility [65], which could affect the assessment of predictive power of the volatility models [103]. Another promising line of research within this field could be aimed at using the realised volatility for the estimation of dynamic asset pricing models [11]. For example, Barndorf-Nielsen and Shepard [16] decomposed realized volatility into actual volatility and realized volatility error. Considering this representation as a state-space model, they applied Kalman to estimate volatility models. Building on the findings of Meddahi [105], Bollerslev and Zhou [27] and Garcia et al. [61] estimated stochastic volatility diffusion models using conditional moments of realized volatility. Technique was developed further for the models with jumps by Todorov [149].

There are also some technical problems that need to be solved. MCMC algorithms for many stochastic volatility models update volatility in a single-state procedure [85]. Development of efficient block routines, which could be computationally less intensive, would allow to analyze sequentially a wider range of models.

Appendix A

Code Listings

99

Appendix provides R code used in the numerical experiments performed in this thesis as well as model description in BUGS [143] notation.

A.1 Data preprocessing

A.1.1 Trading hours extraction

Script in R statistical programming language to extract the data corresponding to trading hours.

```
#process nasdaq quotes downloaded through dtmq  
rm(list = ls())  
source("J:\\brunel\\R\\r.funcs\\r.hft.funcs.R")
```

```

instr<- "MSFTTS.sorted"
fname.ext<- "csv"
instr.name<-paste(instr, ".", fname.ext, sep="")
instr.data.dir<- "J:\\ dtniq\\ dtniq.prog\\ data.r.proc.msft\\"
fname<-paste(instr.data.dir, instr.name, sep="")
#####
instr.trading.session<- "MSFTTS.sorted.trading.session"
instr.name.trading.session<-paste(instr.trading.session,
                                   ".", fname.ext, sep="")
fname.trading.session<-paste(instr.data.dir, instr.name.trading.session,
                              sep="")
#####
dframe<-read.table(fname, header=TRUE, sep="," , stringsAsFactors=FALSE)
aux.tStamps<-paste(aux.date.chr, dframe$Time, sep=" ")
aux.tStamps.POSIXct<-as.POSIXct(aux.tStamps)
dframe$aux.tStamps.POSIXct<-aux.tStamps.POSIXct

dframe$trading.time.start<-as.POSIXct(paste(aux.date.chr, "09:30:00"))
dframe$trading.time.end<-as.POSIXct(paste(aux.date.chr, "16:00:00"))
dframe.trading.session<-subset(dframe,
                               subset=(dframe$aux.tStamps.POSIXct>=dframe$trading.time.start) &
                                       (dframe$aux.tStamps.POSIXct<=dframe$trading.time.end))
#delete auxiliary columns
dframe.trading.session$aux.tStamps.POSIXct<-NULL
dframe.trading.session$trading.time.start<-NULL
dframe.trading.session$trading.time.end<-NULL
#write trading hours data
write.table(dframe.trading.session, file=fname.trading.session, quote=FALSE,
            row.names=FALSE, col.names=TRUE, sep="," )

```

A.1.2 Specified period extraction

Script in R statistical programming language to extract the data corresponding to chosen days.

```
88 #process nasdaq quotes downloaded through dtmq
#additional functions
rm(list = ls())
source("J:\\brunel\\R\\r.funcs\\r.hft.funcs.R")
instr<-"MSFTTS.sorted.trading.session"
fname.ext<-"csv"
instr.name<-paste(instr,".",fname.ext,sep="")
instr.data.dir<-"J:\\dtmq\\dtmq.prog\\data.r.proc.msft\\"
fname<-paste(instr.data.dir,instr.name,sep="")
#####
instr.mod<-"MSFTTS.mod.1.1_min"
instr.name.mod<-paste(instr.mod,".",fname.ext,sep="")
fname.mod<-paste(instr.data.dir,instr.name.mod,
                 sep="")
#####
dframe<-read.table(fname,header=TRUE,sep=" ",stringsAsFactors=FALSE)

names(dframe)<-c("Date","Time","Open","High","Low","Close",
               "IncVol","Volume")
aux.tStamps<-paste(dframe$Date,dframe$Time,sep=" ")
aux.tStamps.POSIXct<-as.POSIXct(aux.tStamps)
dframe$aux.tStamps.POSIXct<-aux.tStamps.POSIXct

dframe$time.start<-as.POSIXct("2011-03-01_09:30:00")
dframe$time.end<-as.POSIXct("2011-05-31_16:00:00")
dframe.mod<-subset(dframe,subset=
                  (dframe$aux.tStamps.POSIXct>=dframe$time.start) &
                  (dframe$aux.tStamps.POSIXct<=dframe$time.end))
```

```

#delete auxiliary columns
dframe.mod$aux.tStamps.POSIXct<-NULL
dframe.mod$time.start<-NULL
dframe.mod$time.end<-NULL
#write data
write.table(dframe.mod, file=fname.mod, quote=FALSE,
            row.names=FALSE, col.names=TRUE, sep=" ")

```

A.1.3 Aggregation

Script in R statistical programming language to aggregate 1 min intraday trading data to 5 min data using *timeSeries* package [155].

69

```

#process nasdaq quotes downloaded through dtmq
library("timeSeries")
rm(list = ls())
source("J:\\brunel\\R\\r.funcs\\r.hft.funcs.R")
instr<-"MSFTTS.mod.1.1_min"
fname.ext<-"csv"
instr.name<-paste(instr, ".", fname.ext, sep="")
instr.data.dir<-"J:\\dtmq\\dtmq.prog\\data.r.proc.msft\\"
fname<-paste(instr.data.dir, instr.name, sep="")
#####
## output
instr.mod.name<-"MSFTTS.mod.1.5_min"
instr.mod.name.part<-paste(instr.mod.name, ".", fname.ext, sep="")
fname.mod.name.full<-paste(instr.data.dir, instr.mod.name.part,
                           sep="")
#####
dframe<-read.table(fname, header=TRUE, sep=" ", stringsAsFactors=FALSE)

```

```

names(dframe)<-c("Date", "Time", "Open", "High", "Low", "Close",
               "IncVol", "Volume")
aux.tStamps<-paste(dframe$Date, dframe$Time, sep="_")
tSeries<-timeSeries(data=dframe, charvec=aux.tStamps)
#aggregate calculating daily volatility
#aggregation parameters
aggr.start<-time(tSeries)[1]
aggr.len<-length(time(tSeries))
aggr.end<-time(tSeries)[aggr.len] + 24*3600
aggr.by<-timeSequence(from=aggr.start, to=aggr.end, by = 5*60)
tSeries.aggr<-aggregate(tSeries,
                        from=aggr.start,
                        to=aggr.end,
                        by=aggr.by,
                        x.start)
tSeries.aggr.as.dframe<-as.data.frame(tSeries.aggr)
DateTime<-time(tSeries.aggr)
DateTime.chr<-as.character(DateTime)
DateTime.chr.split<-strsplit(DateTime.chr, "_")
DateTime.chr.split.unlisted<-unlist(DateTime.chr.split)
DateTime.chr.split.unlisted.matrix<-matrix(DateTime.chr.split.unlisted,
                                           ncol=2, byrow=T)
Dates<-DateTime.chr.split.unlisted.matrix[,1]
Times<-DateTime.chr.split.unlisted.matrix[,2]
tSeries.dframe<-data.frame(Date=Dates, Time=Times,
                           Open=tSeries.aggr.as.dframe$Open,
                           High=tSeries.aggr.as.dframe$High,
                           Low=tSeries.aggr.as.dframe$Low,
                           Close=tSeries.aggr.as.dframe$Close)

#write data
write.table(tSeries.dframe, file=fname.mod.name.full, quote=FALSE,

```

```
row.names=FALSE, col.names=TRUE, sep=" ,")
```

A.1.4 Realized volatility calculation

Script in R statistical programming language to calculate realized volatility from 5 min intraday data using *timeSeries* package [155].

71

```
#process nasdaq quotes downloaded through dtmq
#additional functions
library("timeSeries")
rm(list = ls())
source("J:\\brunel\\R\\r.funcs\\r.hft.funcs.R")
instr<-"MSFTTS.mod.1.5_min"
fname.ext<-"csv"
instr.name<-paste(instr, ".", fname.ext, sep="")
instr.data.dir<-"J:\\dtmq\\dtmq.prog\\data.r.proc.msft\\"
fname<-paste(instr.data.dir, instr.name, sep="")
#####
## output
instr.mod.name<-"MSFTTS.mod.1.5_min.intravol"
instr.mod.name.part<-paste(instr.mod.name, ".", fname.ext, sep="")
fname.mod.name.full<-paste(instr.data.dir, instr.mod.name.part,
                           sep="")
#####
dframe<-read.table(fname, header=TRUE, sep=" ,", stringsAsFactors=FALSE)

names(dframe)<-c("Date", "Time", "Open", "High", "Low", "Close")
aux.tStamps<-paste(dframe$Date, dframe$Time, sep=" ")

tSeries<-timeSeries(data=dframe, charvec=aux.tStamps)
```



```

data.prices<-cbind(Open=dframe$Open,
                    High=dframe$High,Low=dframe$Low,Close=dframe$Close)
tSeries.prices<-timeSeries(data.prices, charvec=aux.tStamps)
tSeries.prices.returns<-returns(tSeries.prices,method="continuous")
#aggregate calculating daily volatility
#aggregation parameters
aggr.start<-time(tSeries.prices.returns)[1] + 6*3600 + 30*60
aggr.len<-length(time(tSeries.prices.returns))
aggr.end<-time(tSeries.prices.returns)[aggr.len] + 24*3600
aggr.by<-timeSequence(from=aggr.start, to=aggr.end, by = "day")
tSeries.prices.returns.day.volatility<-aggregate(tSeries.prices.returns,
                                                from=aggr.start,
                                                to=aggr.end,
                                                by=aggr.by,
                                                sq.ret)

tSeries.prices.as.dframe<-as.data.frame(tSeries.prices)
tSeries.prices.returns.as.dframe<-as.data.frame(tSeries.prices.returns)
tSeries.prices.returns.day.volatility.as.dframe<-as.data.frame(tSeries.prices.returns.day.volatility)
names(tSeries.prices.returns.day.volatility.as.dframe)<-c("SdOpen","SdHigh","SdLow","SdClose")
DateTime<-time(tSeries.prices.returns.day.volatility)
DateTime.chr<-as.character(DateTime)
DateTime.chr.split<-strsplit(DateTime.chr,"_")
DateTime.chr.split.unlisted<-unlist(DateTime.chr.split)
DateTime.chr.split.unlisted.matrix<-matrix(DateTime.chr.split.unlisted, ncol=2,byrow=T)
Dates<-DateTime.chr.split.unlisted.matrix[,1]
Times<-DateTime.chr.split.unlisted.matrix[,2]
delta_t<-5 # minutes
delta_t_scale<-24*60 #minutes in a day
#scaling factor
scfactor<-sqrt((24*60)/(6*60+30))
#scfactor<-1
tSeries.prices.returns.day.volatility.as.dframe<-scfactor*tSeries.prices.returns.day.volatility.as.dframe

```

```

volatility.dframe<-data.frame(Date=Dates,Time=Times,
                             SdOpen=tSeries.prices.returns.day.volatility.as.dframe$SdOpen,
                             SdHigh=tSeries.prices.returns.day.volatility.as.dframe$SdHigh,
                             SdLow=tSeries.prices.returns.day.volatility.as.dframe$SdLow,
                             SdClose=tSeries.prices.returns.day.volatility.as.dframe$SdClose)
write.table(volatility.dframe, file=fname.mod.name.full, quote=FALSE,
            row.names=FALSE, col.names=TRUE, sep=" ")

```

A.1.5 End of day data extraction for SV model calibration

Script in R statistical programming language to extract end of day trading data from 1 min data.

73

```

#process nasdaq quotes downloaded through dtmq
#additional functions
rm(list = ls())
source("J:\\brunel\\R\\r.funcs\\r.hft.funcs.R")
instr<-"MSFTTS.mod.1.1_min"
fname.ext<-"csv"
instr.name<-paste(instr, ".", fname.ext, sep="")
instr.data.dir<-"J:\\dtmq\\dtmq.prog\\data.r.proc.msft\\"
fname<-paste(instr.data.dir, instr.name, sep="")
#####
instr.mod.end.of.day<-"MSFTTS.mod.1.end"
instr.name.mod.end.of.day<-paste(instr.mod.end.of.day, ".", fname.ext, sep="")
fname.mod.end.of.day<-paste(instr.data.dir, instr.name.mod.end.of.day,
                             sep="")
#####
dframe<-read.table(fname, header=TRUE, sep=" ", stringsAsFactors=FALSE)

names(dframe)<-c("Date", "Time", "Open", "High", "Low", "Close",

```

```

      "IncVol", "Volume")
aux.tStamps<-paste(aux.date.chr, dframe$Time, sep=" ")
aux.tStamps.POSIXct<-as.POSIXct(aux.tStamps)
dframe$aux.tStamps.POSIXct<-aux.tStamps.POSIXct
dframe$time.end<-as.POSIXct(paste(aux.date.chr, " 16:00:00"))
dframe.mod.end.of.day<-subset(dframe, subset=(dframe$aux.tStamps.POSIXct==dframe$time.end))
#delete auxiliary columns
dframe.mod.end.of.day$aux.tStamps.POSIXct<-NULL
dframe.mod.end.of.day$time.end<-NULL
write.table(dframe.mod.end.of.day, file=fname.mod.end.of.day, quote=FALSE,
            row.names=FALSE, col.names=TRUE, sep=",")

```

A.1.6 Data analysis and hypothesis testing

74

Script in R statistical programming language to perform the following tasks:

1. Convert time series of prices to time series of returns and write it to file in format suitable for JAGS,
2. Calculate basic statistics including skewness and kurtosis,
3. Perform Shapiro–Wilk and Jarque–Bera normality tests,
4. Conduct Box–Pierce and Ljung–Box portmanteau test for serial correlation,

```

#prepare jags files
rm(list = ls(all = TRUE))
source("E:\\brunel\\R\\r.funcs\\r.hft.funcs.R")
source("E:\\brunel\\R\\r.funcs\\model.funcs.R")
source("E:\\brunel\\R\\r.funcs\\Box.test.wrap.R")
library("fBasics")
data.in.dir<-"E:\\brunel\\brunel.science\\sv.with.rv\\round.2\\mod.1.2\\data.in\\"

```

```

data.stat.dir<-paste("E:\\brunel\\brunel.science\\sv.with.rv\\",
                    "round.2\\mod.1.2\\data.stat\\",sep="")

data.src.file<-"MSFTTS.mod.1.end.csv"
data.file<-paste(data.in.dir,data.src.file,sep="")
price.data.frame<-read.csv(file=data.file,header=TRUE)
price.data.frame<-adjust.volume(price.data.frame)

returns.list<-get.log.return.list(price.data.frame)

log.returns<-returns.list$log.return.frame$LogReturn
log.volume<-returns.list$log.return.frame$LogVolume
log.incvol<-returns.list$log.return.frame$LogIncVol
log.volume.0<-returns.list$LogVolume0
log.incvol.0<-returns.list$LogIncVol0
volume<-returns.list$log.return.frame$Volume
volume0<-returns.list$Volume0
incvol<-returns.list$log.return.frame$IncVol
incvol0<-returns.list$IncVol0

#write data for jags
data.stat.file<-paste(data.stat.dir,data.src.file,".","data.stat.1",".","dat",sep="")
parse.opts<-c("quoteExpressions")
#data for jags
n<-length(log.returns)
logret<-log.returns
logret<-(logret-mean(logret))/sd(logret)
vol<-volume
incv<-incvol
vol_0<-volume0
incv_0<-incvol0
logvol<-log.volume

```

```

logincv<-log.incvol
logvol_0<-log.volume.0
logincv_0<-log.incvol.0
fl<-file(data.stat.file,"w")
sink(fl)
basicStats(logret)
shapiroTest(logret)
jarqueberaTest(logret)
box.test.dframe<-Box.test.wrap(x=logret,lag=10,type="Box-Pierce")
print(box.test.dframe,digits=5)
box.test.dframe<-Box.test.wrap(x=logret,lag=10,type="Ljung-Box")
print(box.test.dframe,digits=5)
sink(file=NULL)
close(fl)

```

76

A.2 JAGS models' description

A.2.1 Classic SV model

```

model {
#likelihood: joint distribution of ys
for (t in 1:n)
{
yisigma2[t]<-1/exp(theta[t])
logret[t] ~ dnorm(0,yisigma2[t])
}

# prior distributions

```

```

mu ~ dnorm(0,1.0E-4)
itau2 ~ dgamma(0.001,0.001)
beta <- exp(mu/2)
phi ~ dbeta(1,1)
tau <- sqrt(1/itau2)
tau2 <- 1/itau2
theta0 ~ dnorm(mu, itau2)
thmean[1] <- mu + phi*(theta0-mu)
theta[1] ~ dnorm(thmean[1], itau2)
for (t in 2:n)
{
  thmean[t] <- mu + phi*(theta[t-1]-mu)
  theta[t] ~ dnorm(thmean[t], itau2)
}
}

```

77

A.2.2 SV model with innovations having Student distribution

```

model {
#likelihood: joint distribution of ys
for (t in 1:n)
{
  yisigma2[t] <- 1/exp(theta[t])
  logret[t] ~ dt(0,yisigma2[t],3)
}

# prior distributions

mu ~ dnorm(0,1.0E-4)
itau2 ~ dgamma(0.001,0.001)
beta <- exp(mu/2)
phi ~ dbeta(1,1)

```

```

tau <- sqrt(1/itau2)
tau2 <- 1/itau2
theta0 ~ dnorm(mu, itau2)
thmean[1] <- mu + phi*(theta0-mu)
theta[1] ~ dnorm(thmean[1], itau2)
for (t in 2:n)
  {
    thmean[t] <- mu + phi*(theta[t-1]-mu)
    theta[t] ~ dnorm(thmean[t], itau2)
  }
}

```

A.2.3 SV model with Gaussian innovations augmented with lag one trading volume

87

```

model {
  #likelihood: joint distribution of ys
  for (t in 1:n)
    {
      yisigma2[t] <- 1/exp(theta[t])
      logret[t] ~ dnorm(0, yisigma2[t])
    }

  # prior distributions

  mu ~ dnorm(0, 1.0E-4)
  itau2 ~ dgamma(0.001, 0.001)
  beta <- exp(mu/2)
  phi ~ dbeta(1, 1)
  psi ~ dnorm(0, 1.0E-4)
  tau <- sqrt(1/itau2)
  tau2 <- 1/itau2
  theta0 ~ dnorm(mu, itau2)
}

```

```

thmean[1] <- mu + phi*(theta0-mu) + psi*logvol_0
theta[1] ~ dnorm(thmean[1], itau2)
for (t in 2:n)
  {
    thmean[t] <- mu + phi*(theta[t-1]-mu) + psi*logvol[t-1]
    theta[t] ~ dnorm(thmean[t], itau2)
  }
}

```

A.2.4 SV model with t-innovations augmented with lag one trading volume

69

```

model {
#likelihood: joint distribution of ys
for (t in 1:n)
  {
    yisigma2[t] <- 1/exp(theta[t])
    logret[t] ~ dt(0, yisigma2[t], 3)
  }

# prior distributions

mu ~ dnorm(0, 1.0E-4)
itau2 ~ dgamma(0.001, 0.001)
beta <- exp(mu/2)
phi ~ dbeta(1, 1)
psi ~ dnorm(0, 1.0E-4)
tau <- sqrt(1/itau2)
tau2 <- 1/itau2
theta0 ~ dnorm(mu, itau2)
thmean[1] <- mu + phi*(theta0-mu) + psi*logvol_0
theta[1] ~ dnorm(thmean[1], itau2)

```



```

for (t in 2:n)
{
  thmean[t] <- mu + phi*(theta[t-1]-mu) + psi*logvol[t-1]
  theta[t] ~ dnorm(thmean[t], itau2)
}
}

```

A.2.5 SV model with Gaussian innovations augmented with lag two trading volume

80

```

model {
#likelihood: joint distribution of ys
for (t in 2:n)
{
  yisigma2[t] <- 1/exp(theta[t])
  logret[t] ~ dnorm(0, yisigma2[t])
}

# prior distributions

mu ~ dnorm(0, 1.0E-4)
itau2 ~ dgamma(0.001, 0.001)
beta <- exp(mu/2)
phi ~ dbeta(1, 1)
psi ~ dnorm(0, 1.0E-4)
tau <- sqrt(1/itau2)
tau2 <- 1/itau2
theta0 ~ dnorm(mu, itau2)
thmean[2] <- mu + phi*(theta0-mu) + psi*logvol_0
theta[2] ~ dnorm(thmean[2], itau2)
for (t in 3:n)
{

```

```

    thmean[t] <- mu + phi*(theta[t-1]-mu) + psi*logvol[t-2]
    theta[t] ~ dnorm(thmean[t], itau2)
  }
}

```

A.2.6 SV model with t-innovations augmented with lag two trading volume

```

model {
#likelihood: joint distribution of ys
for (t in 2:n)
  {
    yisigma2[t] <- 1/exp(theta[t])
    logret[t] ~ dt(0, yisigma2[t], 3)
  }

# prior distributions

mu ~ dnorm(0, 1.0E-4)
itau2 ~ dgamma(0.001, 0.001)
beta <- exp(mu/2)
phi ~ dbeta(1, 1)
psi ~ dnorm(0, 1.0E-4)
tau <- sqrt(1/itau2)
tau2 <- 1/itau2
theta0 ~ dnorm(mu, itau2)
thmean[2] <- mu + phi*(theta0-mu) + psi*logvol_0
theta[2] ~ dnorm(thmean[2], itau2)
for (t in 3:n)
  {
    thmean[t] <- mu + phi*(theta[t-1]-mu) + psi*logvol[t-2]
    theta[t] ~ dnorm(thmean[t], itau2)
  }
}

```

```
}  
}
```

A.3 Data postprocessing

A.3.1 Parameter estimation for classic SV model

Script in R statistical programming language to perform the following tasks:

1. Load Markov chain data simulated in JAGS,
2. Obtain parameter estimates of SV model from chain data,
3. Calculate highest posterior density intervals,
4. Plot chains' trace plots and autocorrelation function,
5. Perform convergence diagnostics using Geweke, Gelman and Rubin, Heidelberger and Welch tests.

Script uses function from R packages CODA [118] and TSA [36].

```
library("coda")  
library("TSA")  
#####  
# read data from chain text files  
#####  
#paths  
#####  
chain.data.dir<-paste("J:\\brunel\\brunel.science\\sv.with.rv\\round.2\\",  
                      "mod.1.2\\jags.linux.batches\\batch.1\\coda",sep="")
```

```

coda.out.dir<-paste("J:\\brunel\\brunel.science\\sv.with.rv\\round.2\\",
                    "mod.1.2\\jags.linux.batches\\batch.1\\conv.analysis.1.params.out\\",sep="")
#####
#read jags output into mcmc object
#####
chain.start<-1
chain.end<-1000000
chain.thin<-1
setwd(chain.data.dir)
mcmc.obj.all<-read.openbugs(start=chain.start,end=chain.end,thin=chain.thin)
window.chain.start<-800000
window.chain.end<-1000000
window.chain.thin<-1
mcmc.obj<-window(mcmc.obj.all,start=window.chain.start,end=window.chain.end,
                 thin=window.chain.thin)
#####
#TSA autocorrs
#####
autocorr.file<-paste(coda.out.dir,"autocorr.pdf",sep="")
pdf(autocorr.file)
autocorr.plot(mcmc.obj,auto.layout=FALSE)
dev.off()
#####
#Raftery and Lewis diagnostics
#####
raftery.lewis.diag.file<-paste(coda.out.dir,"raftery.lewis.diag.dat",sep="")
sink(file=raftery.lewis.diag.file)
raftery.diag(mcmc.obj,q=0.5)
sink()
#####
# coda summary
#####

```

```

summary.file<-paste(coda.out.dir,"coda.summary.dat",sep="")
sink(file=summary.file)
summary.obj<-summary(mcmc.obj)
print(summary.obj)
sink()
summary.obj.file<-paste(coda.out.dir,"coda.summary.obj.dat",sep="")
dump("summary.obj",file=summary.obj.file)
#####
#####
# Highest Posterior Density intervals
#####
hpd.file<-paste(coda.out.dir,"hpd.dat",sep="")
sink(file=hpd.file)
HPDinterval(mcmc.obj)
sink()
#####
#trace plots
#####
plot.mcmc.file<-paste(coda.out.dir,"plot.mcmc.bmp",sep="")
bmp(plot.mcmc.file)
plot(mcmc.obj,density=FALSE,auto.layout=FALSE)
dev.off()
#####
#geweke diagnostics
#####
geweke.diag.file<-paste(coda.out.dir,"geweke.diag.dat",sep="")
sink(file=geweke.diag.file)
geweke.diag(mcmc.obj)
sink()
geweke.plot.diag.file<-paste(coda.out.dir,"geweke.plot.diag.pdf",sep="")
pdf(geweke.plot.diag.file)
geweke.plot(mcmc.obj,auto.layout=FALSE)

```

```

dev.off()
#####
#####
#gelman diagnostics
#####
gelman.diag.file<-paste(coda.out.dir,"gelman.diag.dat",sep="")
sink(file=gelman.diag.file)
gelman.diag(mcmc.obj)
sink()
gelman.plot.diag.file<-paste(coda.out.dir,"gelman.plot.diag.pdf",sep="")
pdf(gelman.plot.diag.file)
gelman.plot(mcmc.obj,auto.layout=FALSE)
dev.off()
#####
#####
#Heidelberger and Welch's convergence diagnostic
#####
heidel.diag.file<-paste(coda.out.dir,"heidel.diag.dat",sep="")
sink(file=heidel.diag.file)
heidel.diag(mcmc.obj)
sink()
#####

```

A.3.2 Forecasting performance analysis for classic SV model

```

rm(list = ls())
source("J:\\brunel\\R\\r.funcs\\r.hft.funcs.R")
source("J:\\brunel\\R\\r.funcs\\model.funcs.R")
data.in.dir<-"J:\\brunel\\brunel.science\\sv.with.rv\\round.2\\mod.1.2\\data.in\\"
volatility.src.file<-"MSFTTS.mod.1.5_min.intravol.csv"

```

```

chain.data.dir<-paste("J:\\brunel\\brunel.science\\sv.with.rv\\round.2\\",
                      "mod.1.2\\jags.linux.batches\\batch.1\\coda",sep="")
coda.out.dir<-paste("J:\\brunel\\brunel.science\\sv.with.rv\\round.2\\",
                    "mod.1.2\\jags.linux.batches\\batch.1\\conv.analysis.1\\",sep="")

#####
#forecast evaluation file
#####
forecast.eval.file<-paste(coda.out.dir,"forecast.eval.dat",sep="")
forecast.measures.file<-paste(coda.out.dir,"forecast.eval.measures.dat",sep="")
forecast.errors.file<-paste(coda.out.dir,"forecast.eval.errors.csv",sep="")
#####
#read realized volatility file
volatility.file<-paste(data.in.dir,volatility.src.file,sep="")
summary.obj.file<-paste(coda.out.dir,"coda.summary.obj.dat",sep="")
source(file=summary.obj.file)
#extract model parameters
mu<-summary.obj$statistics["mu","Mean"]
beta<-summary.obj$statistics["beta","Mean"]
phi<-summary.obj$statistics["phi","Mean"]
tau2<-summary.obj$statistics["tau2","Mean"]
itau2<-summary.obj$statistics["itau2","Mean"]
#read file with daily annualized volatility calculated from
# 5-min intervals
intvol.dframe<-read.table(volatility.file,header=TRUE,sep="," ,stringsAsFactors=FALSE)
intvol<-intvol.dframe$SdClose
mean.intvol<-mean(intvol)
scaling<-exp(0.5*mu)/mean.intvol
intvol<-scaling*intvol
log.intvol2<-log(intvol*intvol)
intvol.len<-length(intvol)
intvol.rv.to.compare<-intvol[2:intvol.len]
intvol.from.mod.to.compare<-rep(0,(intvol.len-1))

```

```

log.sigma.2<-rep(0,(intvol.len-1))
n.repl<-1
intvol.from.mod.to.compare.mat<-matrix(rep(0,n.repl*(intvol.len-1)),ncol=n.repl)
for(rep in 1:n.repl)
{
log.vol2.stoch.cmpn<-rep(0,intvol.len-1)
for(i in 1:(intvol.len-1))
{
log.sigma.2[i]<-mu+phi*(log.intvol2[i]-mu)+log.vol2.stoch.cmpn[i]
}
}
intvol.from.mod.to.compare.mat[,rep]<-exp(0.5*log.sigma.2)
}
intvol.from.mod.to.compare<-apply(intvol.from.mod.to.compare.mat,1,mean)
Mod.RV<-intvol.from.mod.to.compare-intvol.rv.to.compare
ME<-mean(Mod.RV)
ME.series<-Mod.RV
MSE<-mean(Mod.RV^2)
MSE.series<-Mod.RV^2
RMSE<-sqrt(MSE)
RMSE.series<-abs(Mod.RV)
MAE<-mean(abs(Mod.RV))
MAE.series<-abs(Mod.RV)
MAPE<-mean(abs(Mod.RV)/intvol.rv.to.compare)
MAPE.series<-abs(Mod.RV)/intvol.rv.to.compare
forecast.measures.lst<-list(ME=ME,MSE=MSE,RMSE=RMSE,MAE=MAE,MAPE=MAPE)
dump("forecast.measures.lst",file=forecast.measures.file)

forecast.errors.dframe<-data.frame(ME.series=ME.series,MSE.series=MSE.series,
RMSE.series=RMSE.series,MAE.series=MAE.series,
MAPE.series=MAPE.series)
write.table(forecast.errors.dframe,file=forecast.errors.file,quote=FALSE,

```



```
row.names=FALSE, col.names=TRUE)
```

Bibliography

- [1] A.R. Admati and P. Pfleiderer. A theory of intraday patterns: Volume and price variability. *Review of Financial Studies*, 1(1):3–40, 1988. 58
- [2] A.R. Admati and P. Pfleiderer. Divide and conquer: A theory of intraday and day-of-the-week mean effects. *Review of Financial Studies*, 2(2):189–223, 1989. 58
- [3] V. Akgiray. Conditional heteroscedasticity in time series of stock returns: evidence and forecasts. *Journal of business*, pages 55–80, 1989. 32
- [4] S. Alizadeh, M.W. Brandt, and F.X. Diebold. Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57(3):1047–1091, 2002. 36
- [5] T.G. Andersen. Stochastic autoregressive volatility: a framework for volatility modeling. *Mathematical finance*, 4(2):75–102, 1994. 58
- [6] T.G. Andersen. Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance*, pages 169–204, 1996. 58, 63
- [7] T.G. Andersen and T. Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, pages 885–905, 1998. 8, 37
- [8] T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003. 29
- [9] T.G. Andersen, H.J. Chung, and B.E. Sørensen. Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of econometrics*, 91(1):61–87, 1999. 52

- [10] T.G. Andersen and B.E. Sørensen. GMM and QML asymptotic standard deviations in stochastic volatility models: Comments on Ruiz (1994). *Journal of Econometrics*, 76(1):397–403, 1997. 35
- [11] Torben G Andersen and Luca Benzoni. Stochastic volatility. In *Complex Systems in Finance and Econometrics*, pages 694–726. Springer, 2011. 65
- [12] T. Ané and L. Ureche-Rangau. Does trading volume really explain stock returns volatility? *Journal of International Financial Markets, Institutions and Money*, 18(3):216–235, 2008.
- [13] Kevin P Balanda and HL MacGillivray. Kurtosis: a critical review. *The American Statistician*, 42(2):111–119, 1988. 57
- [14] T.G. Bali. Testing the empirical performance of stochastic volatility models of the short-term interest rate. *Journal of Financial and Quantitative Analysis*, 35(2):191–216, 2000. 33
- [15] J.P.M. Bank. Riskmetrics Technical Document. *New York, JP Morgan Bank*, 1996. 29
- [16] Ole E Barndorff-Nielsen. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002. 65
- [17] S. Beckers. Variances of security price returns based on high, low, and closing prices. *Journal of Business*, pages 97–112, 1983. 36
- [18] A.K. Bera and M.L. Higgins. ARCH models: properties, estimation and testing. *Journal of economic surveys*, 7(4):305–366, 1993. 29
- [19] Andreas Berg, Renate Meyer, and Jun Yu. Deviance information criterion for comparing stochastic volatility models. *Journal of Business & Economic Statistics*, 22(1):107–120, 2004. 64
- [20] NG Best, M.K. Cowles, and K. Vines. CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, Version 0.30. *MRC Biostatistics Unit, Cambridge*, 1995. 52
- [21] Robert C Blattberg and Nicholas J Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, pages 244–280, 1974. 27
- [22] H. Bluhm and J. Yu. Forecasting volatility: evidence from the German stock market. 2001. 35

- [23] B. Bollen and B. Inder. Estimating daily volatility in financial markets utilizing intraday data. *Journal of Empirical Finance*, 9(5):551–562, 2002. 36
- [24] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986. 6, 29, 31
- [25] T. Bollerslev, R.Y. Chou, and K.F. Kroner. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1):5–59, 1992. 29
- [26] T. Bollerslev, R.F. Engle, and D.B. Nelson. ARCH models. 1994. 29
- [27] Tim Bollerslev and Hao Zhou. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109(1):33–65, 2002. 65
- [28] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time series analysis: forecasting and control*. Forecasting and Control Series. Prentice Hall, 1994. 23, 24
- [29] G.E.P. Box and D.A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526, 1970. 25, 50
- [30] T.J. Brailsford and R.W. Faff. An evaluation of volatility forecasting techniques. *Journal of Banking & Finance*, 20(3):419–438, 1996. 33
- [31] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer, 2009. 23, 24
- [32] C. Brooks. Predicting stock index volatility: can market volume help? *Journal of Forecasting*, 17(1):59–80, 1998. 34, 63, 64
- [33] J.Y.A. Campbell, A.A.W.C. Lo, and A.C.A. MacKinlay. *The Econometrics of Financial Markets*. Princeton Univers. Press, 1997. 19
- [34] C.Q. Cao and R.S. Tsay. Nonlinear time-series analysis of stock volatilities. *Journal of Applied Econometrics*, 7(S1):S165–S185, 1992. 33
- [35] G. Casella and E.I. George. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174, 1992. 41
- [36] Kung-Sik Chan and Brian Ripley. *TSA: Time Series Analysis*, 2012. R package version 1.01. 82

- [37] Wing H Chan and John M Maheu. Conditional jump dynamics in stock market returns. *Journal of Business & Economic Statistics*, 20(3):377–389, 2002. 65
- [38] Mikhail Chernov, A Ronald Gallant, Eric Ghysels, and George Tauchen. Alternative models for stock price dynamics. *Journal of Econometrics*, 116(1):225–257, 2003. 65
- [39] Siddhartha Chib, Federico Nardari, and Neil Shephard. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316, 2002. 64
- [40] P.K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, pages 135–155, 1973. 8, 57
- [41] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. 27
- [42] M.K. Cowles and B.P. Carlin. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, pages 883–904, 1996. 45
- [43] R. Cumby, S. Figlewski, and J. Hasbrouck. Forecasting volatilities and correlations with EGARCH models. *The Journal of Derivatives*, 1(2):51–63, 1993. 7, 33
- [44] J. Danielsson. Stochastic volatility in asset prices estimation with simulated maximum likelihood. *Journal of Econometrics*, 64(1):375–400, 1994. 35, 51
- [45] F.X. Diebold and J.A. Lopez. Modeling volatility dynamics. *Macroeconomics: developments, tensions and prospects*, 427(472):1995, 1995. 29
- [46] F.C. Drost and B.J.M. Werker. Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics*, 74(1):31–57, 1996. 7
- [47] D. Duffie and K.J. Singleton. Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica: Journal of the Econometric Society*, pages 929–952, 1993. 35
- [48] C.L. Dunis, J. Laws, and S. Chauvin. The use of market data and model combination to improve forecast accuracy. *Development in Forecasts Combination and Portfolio Choice (Wiley, Oxford)*, 2001. 35

- [49] J. Durbin and S.J. Koopman. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):3–56, 2000. 35
- [50] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982. 6, 29, 30
- [51] R.F. Engle and T. Bollerslev. Modelling the persistence of conditional variances. *Econometric reviews*, 5(1):1–50, 1986. 31
- [52] R.F. Engle and V.K. Ng. Measuring and testing the impact of news on volatility. Technical report, National Bureau of Economic Research, 1991. 32
- [53] Bjørn Eraker. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance*, 59(3):1367–1404, 2004. 65
- [54] Bjørn Eraker, Michael Johannes, and Nicholas Polson. The impact of jumps in volatility and returns. *The Journal of Finance*, 58(3):1269–1300, 2003. 65
- [55] E.F. Fama. Random walks in stock market prices. *Financial Analysts Journal*, pages 55–59, 1965. 20
- [56] S. Figlewski. Forecasting volatility. *Financial Markets, Institutions & Instruments*, 6(1):1–88, 1997. 7, 29, 33, 36
- [57] M. Fridman and L. Harris. A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business & Economic Statistics*, pages 284–291, 1998. 35, 51
- [58] W.A. Fuller. *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. Wiley, 1995. 24
- [59] W.K.H. Fung and D.A. Hsieh. Empirical analysis of implied volatility: Stocks, bonds and currencies. *Manuscript, Duke University*, 1991. 37
- [60] A.R. Gallant, D. Hsieh, and G. Tauchen. Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, 81(1):159–192, 1997. 51
- [61] René Garcia, Marc-André Lewis, and E Renault. Estimation of objective and risk-neutral distributions based on moments of integrated

- volatility. *Manuscript, Department of Economics, University of Montreal and CIRANO*, 2001. 65
- [62] M.B. Garman and M.J. Klass. On the estimation of security price volatilities from historical data. *Journal of business*, pages 67–78, 1980. 36
- [63] A. Gelman, G. Roberts, and W. Gilks. Efficient metropolis jumping hules. *Bayesian statistics*, 5:599–608, 1996. 42
- [64] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992. 46
- [65] R. Gençay, M. Dacorogna, U.A. Muller, O. Pictet, and R. Olsen. *An Introduction to High-Frequency Finance*. Elsevier Science, 2001. 37, 65
- [66] J. Geweke. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica: Journal of the Econometric Society*, pages 1317–1339, 1989. 42
- [67] John Geweke. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Federal Reserve Bank of Minneapolis, Research Department, 1991. 45
- [68] L.R. Glosten, R. Jagannathan, and D.E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of finance*, pages 1779–1801, 1993. 31
- [69] L.R. Glosten and P.R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1):71–100, 1985. 58
- [70] T.C. Green and S. Figlewski. Market risk and model risk for a financial institution writing options. *The Journal of Finance*, 54(4):1465–1499, 1999. 29
- [71] L. Harris. Cross-security tests of the mixture of distributions hypothesis. *Journal of financial and Quantitative Analysis*, 21(1):39–46, 1986. 57
- [72] L. Harris. Transaction data tests of the mixture of distributions hypothesis. *Journal of Financial and Quantitative Analysis*, 22(2):127–141, 1987. 57
- [73] A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264, 1994. 51

- [74] A.C. Harvey. Forecasting, Structural Time Series Models and the Kalman Filter. *Cambridge Books*, 1991. 52
- [75] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. 41
- [76] P. Heidelberger and P.D. Welch. Simulation run length control in the presence of an initial transient. *Operations Research*, pages 1109–1144, 1983. 47, 55
- [77] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993. 6
- [78] R.C. Heynen and H.M. Kat. Volatility Prediction. *The Journal of Derivatives*, 2(2):50–65, 1994. 33, 35
- [79] R.C. Hill, W.E. Griffiths, and G.C. Lim. *Principles of Econometrics*. Wiley, 2007. 13
- [80] E. Hol and S.J. Koopman. *Forecasting the Variability of Stock Index Returns with Stochastic Volatility Models and Implied Volatility*. Tinbergen Institute, 2000. 35
- [81] D. Insua, F. Ruggeri, and M. Wiper. *Bayesian Analysis of Stochastic Process Models*. Wiley Series in Probability and Statistics. Wiley, 2012. 7
- [82] E. Jacquier, N.G. Polson, and P.E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of business and Economic Statistics*, 20(1):69–87, 2002. 35, 51
- [83] E. Jacquier, N.G. Polson, and P.E. Rossi. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185–212, 2004. 64
- [84] C.M. Jarque and A.K. Bera. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, pages 163–172, 1987. 22
- [85] Michael Johannes and Nicholas Polson. MCMC methods for financial econometrics. *The Handbook of Financial Econometrics*, 2002. 65
- [86] P. Jorion. Predicting volatility in the foreign exchange market. *Journal of Finance*, pages 507–528, 1995. 7

- [87] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer, 1991. 37
- [88] S. Kim, N. Shephard, and S. Chib. Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393, 1998. 35, 52
- [89] Stanley J Kon. Models of stock returnsA comparison. *The Journal of Finance*, 39(1):147–165, 2012. 27
- [90] T.Y. Kuen and T.S. Hoong. Forecasting volatility in the Singapore stock market. *Asia Pacific Journal of Management*, 9(1):1–13, 1992. 34
- [91] A.S. Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985. 58
- [92] C.G. Lamoureux and W.D. Lastrapes. Heteroskedasticity in stock return data: volume versus GARCH effects. *Journal of Finance*, pages 221–229, 1990. 8, 57, 58, 63
- [93] K.Y. Lee. Are the GARCH models best in out-of-sample performance? *Economics Letters*, 37(3):305–308, 1991. 33, 34
- [94] F. Liang, C. Liu, and R. Carroll. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley Series in Computational Statistics. Wiley, 2010. 16
- [95] R. Liesenfeld. Dynamic bivariate mixture models: modeling the behavior of prices and trading volume. *Journal of Business & Economic Statistics*, pages 101–109, 1998. 58
- [96] G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978. 25, 50
- [97] J. Longerstaeay and L. More. *Introduction to RiskMetrics*. Morgan Guaranty Trust Company, 1995. 6
- [98] J.A. Lopez and Federal Reserve Bank of New York. Evaluating the predictive accuracy of volatility models. *Journal of Forecasting*, 20(2):87–109, 2001. 7, 35, 36
- [99] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067, 2009. 7, 52

- [100] John M Maheu and Thomas H McCurdy. News arrival, jump dynamics, and volatility components for individual stock returns. *The Journal of Finance*, 59(2):755–793, 2004. 65
- [101] R. Mahieu and R. Bauer. A Bayesian analysis of stock return volatility and trading volume. *Applied financial economics*, 8(6):671–687, 1998. 58, 63
- [102] Benoit Mandelbrot. The variation of certain speculative prices. *Journal of business*, pages 394–419, 1963. 27
- [103] Michael McAleer and Marcelo C Medeiros. Realized volatility: A review. *Econometric Reviews*, 27(1-3):10–45, 2008. 65
- [104] D. McMillan, A. Speight, and O.A.P. Gwilym. Forecasting UK stock market volatility. *Applied Financial Economics*, 10(4):435–448, 2000. 34
- [105] Nour Meddahi. Moments of continuous time stochastic volatility models. *Unpublished paper: University of Montreal*, 2002. 65
- [106] A. Melino and S.M. Turnbull. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45(1):239–265, 1990. 51
- [107] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21:1087, 1953. 41
- [108] R. Meyer and J. Yu. BUGS for a Bayesian analysis of stochastic volatility models. *The econometrics journal*, 3(2):198–215, 2002. 52
- [109] Sean P Meyn, Richard L Tweedie, and Peter W Glynn. *Markov chains and stochastic stability*, volume 2. Cambridge University Press Cambridge, 2009. 47
- [110] Frederick Mosteller and John W Tukey. Data analysis and regression. A second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods, Reading, Mass.: Addison-Wesley, 1977*, 1, 1977. 57
- [111] Jouchi Nakajima and Yasuhiro Omori. Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics & Data Analysis*, 53(6):2335–2353, 2009. 64
- [112] D.B. Nelson. ARCH models as diffusion approximations. *Journal of Econometrics*, 45(1):7–38, 1990. 7

- [113] D.B. Nelson and C.Q. Cao. Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics*, pages 229–235, 1992. 31
- [114] Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449, 2007. 64
- [115] A.R. Pagan and G.W. Schwert. Alternative models for conditional stock volatility. *Journal of Econometrics*, 45(1-2):267–290, 1990. 33
- [116] M. Parkinson. The extreme value method for estimating the variance of the rate of return. *Journal of Business*, pages 61–65, 1980. 36
- [117] M. Plummer. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. March, pages 20–22, 2003. 52
- [118] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11, 2006. 52, 82
- [119] S.H. Poon. *A Practical Guide to Forecasting Financial Market Volatility*. The Wiley Finance Series. Wiley, 2005. 6, 28
- [120] S.H. Poon and C. Granger. Practical issues in forecasting volatility. *Financial Analysts Journal*, pages 45–56, 2005. 28
- [121] S.H. Poon and C.W.J. Granger. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539, 2003. 6
- [122] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0. 52
- [123] S.T. Rachev, J.S.J. Hsu, B.S. Bagasheva, and CFA Frank J. Fabozzi. *Bayesian Methods in Finance*. Frank J. Fabozzi Series. Wiley, 2008. 7, 14
- [124] Davide Raggi and Silvano Bordignon. Comparing stochastic volatility models through Monte Carlo simulations. *Computational statistics & data analysis*, 50(7):1678–1699, 2006. 65

- [125] M. Richardson and T. Smith. A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information. *Journal of Financial and Quantitative Analysis*, 29(1), 1994. 58
- [126] G. O. Roberts and N. G. Polson. On the geometric convergence of the Gibbs sampler. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–384, 1994. 45
- [127] G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997. 42
- [128] L.C.G. Rogers and S.E. Satchell. Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, pages 504–512, 1991. 36
- [129] L.C.G. Rogers, S.E. Satchell, and Y. Yoon. Estimating the volatility of stock prices: a comparison of methods that use high and low prices. *Applied Financial Economics*, 4(3):241–247, 1994. 36
- [130] JP Royston. Algorithm AS 181: The W test for normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 31(2):176–180, 1982. 22
- [131] JP Royston. An extension of Shapiro and Wilk’s W test for normality to large samples. *Applied Statistics*, pages 115–124, 1982. 22
- [132] Patrick Royston. Remark AS R94: A remark on algorithm AS 181: The W-test for normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(4):547–551, 1995. 22
- [133] E. Ruiz. Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics*, 63(1):289–306, 1994. 35
- [134] G. Sandmann and S.J. Koopman. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, 87(2):271–301, 1998. 51
- [135] Lee Schruben, H Singh, and L Tierney. Optimal tests for initialization bias in simulation output. *Operations Research*, 31(6):1167–1178, 1983. 47
- [136] Lee W Schruben. Detecting initialization bias in simulation output. *Operations Research*, 30(3):569–590, 1982. 47
- [137] A. Sen and M.S. Srivastava. *Regression Analysis: Theory, Methods, and Applications*. Springer Texts in Statistics. Springer, 1990. 21

- [138] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. 21, 22
- [139] N. Shephard. *Stochastic volatility: selected readings*. Advanced texts in econometrics. Oxford University Press, 2005. 7
- [140] N. Shephard and M.K. Pitt. Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84(3):653–667, 1997. 35, 51
- [141] K.J. Singleton. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102(1):111–141, 2001. 35
- [142] M. Sørensen. Prediction-based estimating functions. *The Econometrics Journal*, 3(2):123–147, 2000. 51
- [143] D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn. WinBUGS, version 1.4.3, 2007. 66
- [144] G.E. Tauchen and M. Pitts. The price variability-volume relationship on speculative markets. *Econometrica: Journal of the Econometric Society*, pages 485–505, 1983. 34
- [145] J.W. Taylor. Volatility forecasting with smooth transition exponential smoothing. *International Journal of Forecasting*, 20(2):273–286, 2004. 29, 33
- [146] SJ TAYLOR. Financial returns modelled by the product of two stochastic processes-A study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice*, 1:203–226, 1982. 6, 34
- [147] S.J. Taylor. Forecasting the volatility of currency exchange rates. *International Journal of Forecasting*, 3(1):159–170, 1987. 29
- [148] S.J. Taylor. *Modelling Financial Time Series*. World Scientific, 2008. 29, 34
- [149] Viktor Todorov. Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics*, 148(2):131–148, 2009. 65
- [150] R.S. Tsay. *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2010. 25, 26, 30
- [151] Y.K. Tse. Stock returns volatility in the Tokyo Stock Exchange. *Japan and the World Economy*, 3(3):285–298, 1991. 34

- [152] D.M. Walsh and G.Y.G. Tsou. Forecasting index volatility: Sampling interval and non-trading effects. *Applied Financial Economics*, 8(5):477–485, 1998. 34
- [153] K.D. West and D. Cho. The predictive ability of several models of exchange rate volatility. *Journal of Econometrics*, 69(2):367–391, 1995. 34
- [154] J.B. Wiggins. Estimating the volatility of S&P 500 futures prices using the extreme-value method. *Journal of Futures Markets*, 12(3):265–273, 1992. 36
- [155] Diethelm Wuertz and Yohan Chalabi. *timeSeries: Rmetrics - Financial Time Series Objects*, 2012. R package version 2160.95. 69, 71
- [156] J. Yu. Forecasting volatility in the New Zealand stock market. *Applied Financial Economics*, 12(3):193–202, 2002. 35
- [157] Jun Yu. On leverage in a stochastic volatility model. *Journal of Econometrics*, 127(2):165–178, 2005. 64
- [158] J.M. Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5):931–955, 1994. 32