A stochastic processes toolkit for risk management: Geometric Brownian motion, jumps, GARCH and variance gamma models

Received (in revised form): 8th January, 2009

Damiano Brigo*

is Managing Director of Fitch Solutions' Quantitative Innovation team and Visiting Professor at Imperial College London's Department of Mathematics. His work has been published in top journals for mathematical finance, systems theory, probability and statistics, and he has also written a field reference book on stochastic interest rate modelling. Damiano is Managing Editor of the *International Journal of Theoretical and Applied Finance*, a member of the Fitch Academic Advisory Board, and has been a charter member of Risk Who's Who since 2007. Damiano obtained a PhD in stochastic filtering with differential geometry in 1996 from the Free University of Amsterdam, following a BSc in mathematics with honours from the University of Padua.

Antonio Dalessandro

holds a PhD in mathematics from Imperial College London and is research associate at the University of Geneva's Department of Econometrics. From 2001 to 2004 he worked as a researcher at CERN and INFN in Geneva. He has a first-class honours MSc in electrical engineering from the Politecnico di Bari, and his research interests include probability, functional analysis, optimisation and high frequencies trading.

Matthias Neugebauer

is Senior Director of Fitch Ratings' structured credit group. His responsibilities include criteria development and credit market research, including design and development of the portfolio credit model. Prior to this, he worked with the company's European quantitative research team. He holds a master's degree in finance and economics from Durham University and an engineering degree from Technical University Cottbus. He also holds a master's degree in mathematical finance from Oxford University.

Fares Triki

obtained an engineering diploma from a French grande école, an MSc in finance from CNAM Paris, and a DEA diploma in economics from Paris Sorbonne University with summa cum laude distinction. After working for multinational consulting firms, he joined the Fitch Ratings Quantitative Financial Research team with the role of risk assessment and valuation of various financial products. Fares is now a macro finance expert at the financial stability directorate of Banque de France and is also a PhD student at the Paris School of Economics. His research areas include risk modelling and international economics and finance.

*Fitch Solutions, 101 Finsbury Pavement, London EC2A 1RS, UK
E-mail: damiano.brigo@fitchsolutions.com

Abstract In risk management it is desirable to grasp the essential statistical features of a time series representing a risk factor. This paper aims to introduce a number of different stochastic processes that can help in grasping the essential features of risk factors, describing different asset classes or behaviours. The paper does not aim to be exhaustive, but gives examples and a feeling for practically implementable models, allowing for stylised features in the data. These models can also be used as building blocks to build more complex models, although, for a number of risk management applications, the models developed here suffice for the first step in the quantitative analysis.

The broad qualitative features addressed here are fat tails. In the second part of this work to appear in a subsequent paper, mean reversion is addressed with and without fat tails. The paper gives some orientation on the initial choice of a suitable stochastic process and then explains how the process parameters can be estimated based on historical data. Once the process has been calibrated, typically through maximum likelihood estimation, one may simulate the risk factor and build future scenarios for the risky portfolio. On the terminal simulated distribution of the portfolio, one may then single out several risk measures, although the present paper focuses on the stochastic processes estimation preceding the simulation of the risk factors. Finally, this paper focuses on single time series. Correlation or more generally dependence across risk factors, leading to multivariate processes modelling, will be addressed in future work.

Keywords: *risk management, stochastic processes, risk factors, fat tails, mean reversion, jump diffusion processes, variance gamma processes*

INTRODUCTION

In risk management and in the rating practice it is desirable to grasp the essential statistical features of a time series representing a risk factor before beginning a detailed technical analysis of the product or the entity under scrutiny. The quantitative analysis is not the final tool, as it has to be integrated with judgmental analysis and possible stresses, but it represents a good starting point for the process leading to the final decision.

This paper aims to introduce a number of different stochastic processes that, according to the economic and financial nature of the specific risk factor, can help in grasping the essential features of the risk factor itself. For example, a family of stochastic processes that can be suited to capture foreign exchange behaviour might not be suited to model hedge funds or a commodity. Hence there is a need to have at one's disposal a sufficiently large set of practically implementable stochastic processes that may be used to address the quantitative analysis at hand to begin a risk management or rating decision process.

This paper does not aim to be exhaustive, as this would be a formidable task beyond the scope of a survey paper. Instead, the paper gives examples and a feeling for practically implementable models allowing for stylised features in the data. The reader may also use these models as building blocks to build more complex models, although for a number of risk management applications the models developed here suffice for the first step in the quantitative analysis.

The broad qualitative features addressed here are fat tails and mean reversion. This paper begins with the geometric Brownian motion (GBM) as a fundamental example of an important stochastic process featuring neither mean reversion nor fat tails; it then goes through generalisations featuring any of the two properties or both of them. According to the specific situation, the different processes are formulated either in (log-) returns space or in levels space, as is more convenient. Levels and returns can be easily converted into each other, so this is indeed a matter of mere convenience.

This paper will address the following processes:

- *Basic process:* arithmetic Brownian motion (ABM, returns) or GBM (levels).
- *Fat-tails processes:* GBM with lognormal jumps (levels), ABM with normal jumps (returns), GARCH (returns), variance gamma (returns).

In the second part of this paper the following are analysed:

- *Mean reverting processes:* Vasicek, CIR (levels of interest rates or spreads, or returns in general), exponential Vasicek (levels).
- *Mean reverting processes with fat tails:* Vasicek with jumps (levels of interest rates or spreads, or returns in general), exponential Vasicek with jumps (levels).

Different financial time series are better described by different processes. In general, when first presented with a historical time series of data for a risk factor, one should decide what kind of general properties the series features. The financial nature of the time series can give some hints at whether or not mean reversion and fat tails are expected to be present. Interest rate processes, for example, are often taken to be mean-reverting, whereas foreign exchange series are often supposed to be fat-tailed and credit spreads feature both characteristics. In general one should check for mean reversion or stationarity as well as for fat tails.

Checking for mean reversion or stationarity

Some of the tests mentioned in the report concern mean reversion. The presence of an autoregressive (AR) feature can be tested in the data, typically on the returns of a series or on the series itself if this is an interest rate or spread series. In linear processes with normal shocks, this amounts to checking stationarity. If this is present, this can be a symptom for mean reversion and a mean-reverting process can be adopted as a first assumption for the data. If the AR test is rejected, this means that there is no mean reversion, at least under normal shocks and linear models, although the situation can be more complex for nonlinear processes with possibly fat tails, where the more general notions of stationarity and ergodicity may apply. These are not addressed in this paper.

If the tests find AR features then the process is considered as mean-reverting and one can compute autocorrelation and partial autocorrelation functions to see the lag in the regression. Alternatively, one can go directly to the continuous time model and estimate it on the data through maximum likelihood. In this case, the main model to try is the Vasicek model.

If the tests do not find AR features then the simplest form of mean reversion for linear processes with Gaussian shocks is to be rejected. One must, however, be careful as the process could still be mean-reverting in a more general sense. In a way, there could be some sort of mean reversion even under non-Gaussian shocks; the jump-extended Vasicek or exponential Vasicek models represent an example of such a case, where mean reversion is mixed with fat tails, as shown below in the next point.

Checking for fat tails

If the AR test fails, either the series is not mean-reverting, or it is but with fatter tails than the Gaussian distribution and possibly nonlinear behaviour. To test for fat tails, a first graphical tool is the QQ-plot, comparing the tails in the data with the Gaussian distribution. The QQ-plot gives immediate graphical evidence on the possible presence of fat tails. Further evidence can be collected by considering the third and fourth sample moments, or better, the skewness and excess kurtosis, to see how the data differ from the Gaussian distribution. More rigorous tests on normality should also be run following this preliminary analysis (for example, the Jarque Bera, the Shapiro-Wilk and the Anderson-Darling tests, which are not addressed in this paper). If fat tails are not found and the AR test failed, this means that one is likely to be dealing with a process lacking mean reversion but with Gaussian shocks, that could be modelled as an arithmetic Brownian motion. If fat tails are found (and the AR test has been rejected), one may start calibration with models featuring fat tails and no mean reversion (GARCH, NGARCH, variance gamma, arithmetic Brownian motion with jumps) or fat tails and mean reversion (Vasicek with jumps, exponential Vasicek with jumps). Comparing the goodness of fit of the models may suggest which alternative is preferable. Goodness of fit is determined again graphically through QQ-plot of the model's implied distribution against the empirical distribution, although more rigorous approaches can be applied (examples are the Kolmogorov Smirnov test, likelihood ratio methods and the Akaike information criteria, as well as methods based on the Kullback Leibler information or relative entropy, the Hellinger distance and other divergences). The predictive power can also be tested, following the calibration and the goodness of fit tests (although not pursued here, Diebold Mariano statistics can be mentioned as an example for AR processes).

The above classification may be summarised in Table 1, where the referenced variable is typically the return process or the process itself in case of interest rates or spreads.

Once the process has been chosen and calibrated to historical data, typically through regression or maximum likelihood estimation, one may use the process to simulate the risk factor over a given time horizon and build future scenarios for the portfolio under examination. On the terminal simulated distribution of the portfolio, one may then single out several risk measures. This paper does not focus on the risk measures themselves but on the stochastic processes estimation preceding the simulation of the risk factors. In case more than one model is suited to the task, one can analyse risks with different models and compare the outputs as a way to reduce model risk.

Table 1: Model classification

	Normal tails	Fat tails
No mean reversion	ABM	ABM+Jumps, (N)GARCH, VG
Mean reversion	Vasicek	Exponential Vasicek CIR, Vasicek with Jumps

Finally, this paper focuses on single time series. Correlation or more generally dependence across risk factors, leading to multivariate processes modelling, will be addressed in future work.

Prerequisites

The following discussion assumes that the reader is familiar with basic probability theory, including probability distributions (density and cumulative density functions, moment generating functions), random variables and expectations. It is also assumed that the reader has some practical knowledge of stochastic calculus and of Itō's formula, and basic coding skills. For each section, the MATLAB® code is provided to implement the specific models. Examples illustrating the possible use of the models on actual financial time series are also presented.

MODELLING WITH BASIC STOCHASTIC PROCESSES: GEOMETRIC BROWNIAN MOTION (GBM)

This section provides an introduction to modelling through stochastic processes. All the concepts will be introduced using the fundamental process for financial modelling, the GBM with constant drift and constant volatility. The GBM is ubiquitous in finance, being the process underlying the Black and Scholes formula for pricing European options.

The geometric Brownian motion

The GBM describes the random behaviour of the asset price level $S(t)$ over time. The GBM is specified as follows:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) \tag{1}$$

Here, W is a standard Brownian motion, a special diffusion process (also referred to as a Wiener process) that is characterised by independent identically distributed (iid) increments that are normally (or Gaussian) distributed with zero mean and a standard deviation equal to the square root of the time step. Independence in the increments implies that the model is a Markov Process, which is a particular type of process for which only the current asset price is relevant for computing probabilities of events involving future asset prices. In other terms, to compute probabilities involving future asset prices, knowledge of the whole past does not add anything to knowledge of the present.

The d terms indicate that the equation is given in its continuous time version.

A continuous-time stochastic process is one where the value of the price can change at any point in time. The theory is very complex and actually this differential notation is just shorthand for an integral equation. A discrete-time stochastic process on the other hand is one where the price of the financial asset can only change at certain fixed times. In practice, discrete behaviour is usually observed, but continuous time processes prove to be useful to analyse the properties of the model, besides being paramount in valuation where the assumption of continuous trading is at the basis of the Black and Scholes theory and its extensions.

The property of independent identically distributed increments is very important and will be exploited in the calibration as well as in the simulation of the random behaviour of asset prices. Using some basic stochastic calculus the equation can be rewritten as follows:

$$d\log S(t) = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW(t) \tag{2}$$

where log denotes the standard natural logarithm. The process followed by the log is called an arithmetic Brownian motion. The increments in the logarithm of the asset value are normally distributed. This equation is straightforward to integrate between any two instants, t and u, leading to:

$$\begin{aligned}\log S(u) - \log S(t) &= \left(\mu - \frac{1}{2}\sigma^2\right)(u-t) \\ &\quad + \sigma(W(u) - W(t)) \\ &\quad \sim N\left(\left(\mu - \frac{1}{2}\sigma^2\right)(u-t), \sigma^2(u-t)\right).\end{aligned} \tag{3}$$

Through exponentiation and taking $u = T$ and $t = 0$, the solution for the GBM equation is obtained:

$$S(T) = S(0) \times \exp\left(\left[\mu - \frac{1}{2}\sigma^2\right]T + \sigma W(T)\right) \tag{4}$$

This equation shows that the asset price in the GBM follows a log-normal distribution, while the logarithmic returns $\log(S_{t+\Delta t}/S_t)$ are normally distributed.

The moments of $S(T)$ are easily found using the above solution due to the fact that $W(T)$ is Gaussian with mean zero and variance T, and finally the moment generating function of a standard normal random variable Z is given by:

$$\mathbb{E}[e^{aZ}] = e^{\frac{1}{2}a^2}. \tag{5}$$

Hence the first two moments (mean and variance) of $S(T)$ are:

$$\begin{aligned} \mathbb{E}[S(T)] &= S(0)e^{\mu T} \\ \mathrm{Var}[S(T)] &= e^{2\mu T}S(0)^2(e^{\sigma^2 T} - 1) \end{aligned} \tag{6}$$

To simulate this process, the continuous equation between discrete instants $t_0 < t_1 < \ldots < t_n$ needs to be solved as follows:

$$S(t_{i+1}) = S(t_i)\exp\left(\left[\mu - \frac{1}{2}\sigma^2\right](t_{i+1} - t_i) + \sigma\sqrt{t_{i+1} - t_i}Z_{i+1}\right) \tag{7}$$

where $Z_1, Z_2, \ldots, Z_n$ are independent random draws from the standard normal distribution. The simulation is exact and does not introduce any discretisation error, due to the fact that the equation can be solved exactly. Figure 1 plots a number of simulated sample paths using the above equation and the mean plus/minus one standard deviation and the first and 99th percentiles over time. The mean of the asset price grows exponentially with time.

The following MATLAB function simulates sample paths of the GBM using Equation (7), which was vectorised in MATLAB®:

Code 1: MATLAB® code to simulate GBM sample paths

```
function S=GBM_simulation(N_Sim,T,dt,mu,
  sigma,S0)
mean=(mu-0.5*sigma^2)*dt;
S=S0*ones(N_Sim,T+1);
BM=sigma*sqrt(dt)*normrnd(0,1,N_Sim,T);
S(:,2:end)=S0*exp(cumsum(mean+BM,2));
end
```

Parameter estimation

This section describes how the GBM can be used as an attempt to model the random behaviour of the FTSE 100 stock index, the log of which is plotted in the left chart of Figure 2.

The chart on the right of Figure 2 plots the quantiles of the log-return distribution against the quantiles of the standard normal distribution. This QQ-plot allows one to compare distributions and to check the assumption of normality. The quantiles of the historical distribution are plotted on the y-axis and the quantiles of the chosen modelling distribution on the x-axis. If the comparison distribution provides a good fit to the historical returns, then the QQ-plot approximates a straight line. In the case of the FTSE 100 log-returns, the QQ-plot shows that the historical quantiles in the tail of the distribution are significantly larger compared with the normal distribution. This is in line with the general

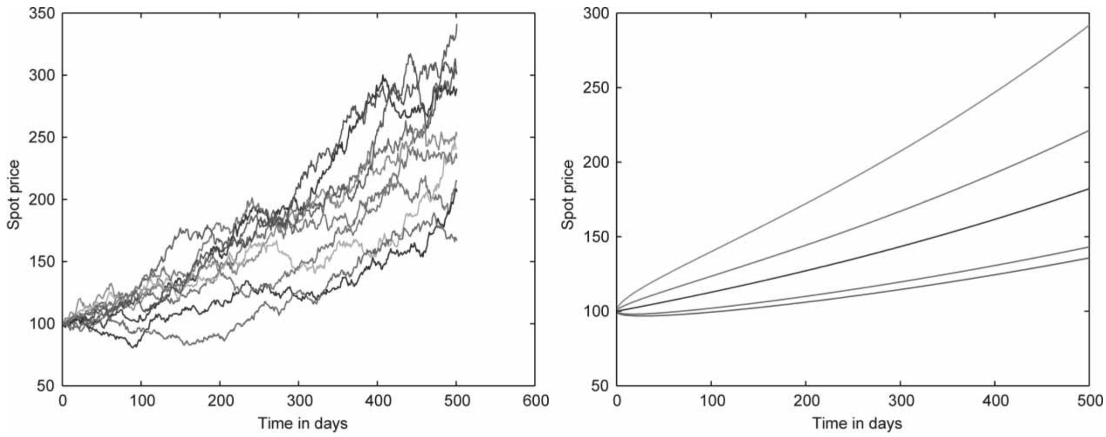


Figure 1: GBM sample paths and distribution statistics

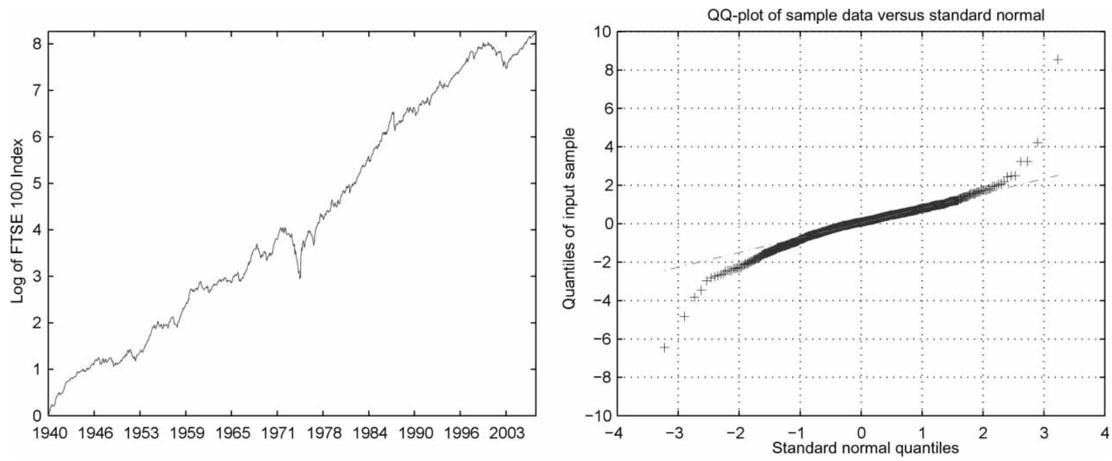


Figure 2: Historical FTSE 100 Index (daily since 1940) and QQ-plot FTSE 100

observation about the presence of fat tails in the return distribution of financial asset prices. Therefore, the GBM at best provides only a rough approximation for the FTSE 100. The next sections will outline some extensions of the GBM that provide a better fit to the historical distribution.

Another important property of the GBM is the independence of returns. Therefore, to ensure the GBM is appropriate for modelling the asset price in a financial time series, one has to check that the returns of the observed data are truly independent. The assumption of independence in statistical terms means that there is no autocorrelation present in historical returns. A common test for autocorrelation (typically on returns of a given asset) in a time series sample x_1, x_2, $\ldots$, x_n realised from a random process $X(t_i)$ is to plot the autocorrelation function of the lag k, defined as:

$$\mathrm{ACF}(k) = \frac{1}{(n-k)\hat{v}} \sum_{i=1}^{n-k} (x_i - \hat{m}) \times (x_{i+k} - \hat{m}), \quad k = 1, 2, \ldots \tag{8}$$

where $\hat{m}$ and $\hat{v}$ are the sample mean and variance of the series, respectively. Often, besides the auto-correlation function (ACF), one also considers the partial auto-correlation function (PACF). Without fully defining PACF, roughly speaking one may say that while ACF(k) gives an estimate of the correlation between $X(t_i)$ and $X(t_{i+k})$, PACF(k) informs on the correlation between $X(t_i)$ and $X(t_{i+k})$ which is not accounted for by the shorter lags from 1 to $k-1$, or more precisely, it is a sort of sample estimate of:

$$\mathrm{Corr}\Big(X(t_i) - \bar{X}_i^{i+1,\ldots,i+k-1},\; X(t_{i+k}) - \bar{X}_{i+k}^{i+1,\ldots,i+k-1}\Big)$$

where $\bar{X}_i^{i+1,\ldots,i+k-1}$ and $\bar{X}_{i+k}^{i+1,\ldots,i+k-1}$ are the best estimates (regressions in a linear context) of $X(t_i)$ and $X(t_{i+k})$ given $X(t_{i+1}), \ldots, X(t_{i+k-1})$. PACF also gives an estimate of the lag in an autoregressive process.

ACF and PACF for the FTSE 100 equity index returns are shown in Figure 3. For the FTSE 100, one does not observe any significant lags in the historical return time series, which means the independence assumption is acceptable in this example.

Maximum likelihood estimation

Having tested the properties of independence as well as normality for the underlying historical data, one can now proceed to calibrate the parameter set $\Theta = (\mu, \sigma)$ for the GBM based on the historical returns. To find Θ that yields the best fit to the historical dataset the method of maximum likelihood estimation (MLE) is used.

MLE can be used for both continuous and discrete random variables. The basic concept of MLE, as suggested by the name, is to find the parameter estimates Θ for the assumed probability density function f_Θ (continuous case) or probability mass function (discrete case) that will maximise the likelihood or probability of having observed the given data sample $x_1, x_2, x_3, \ldots, x_n$ for the random vector $X_1, \ldots, X_n$. In other terms, the observed sample $x_1, x_2, x_3, \ldots, x_n$ is used inside $f_{X1,\ X2,\ldots,Xn;\Theta}$, so that the only variable in f is Θ, and the resulting function is maximised in Θ. The likelihood (or probability) of

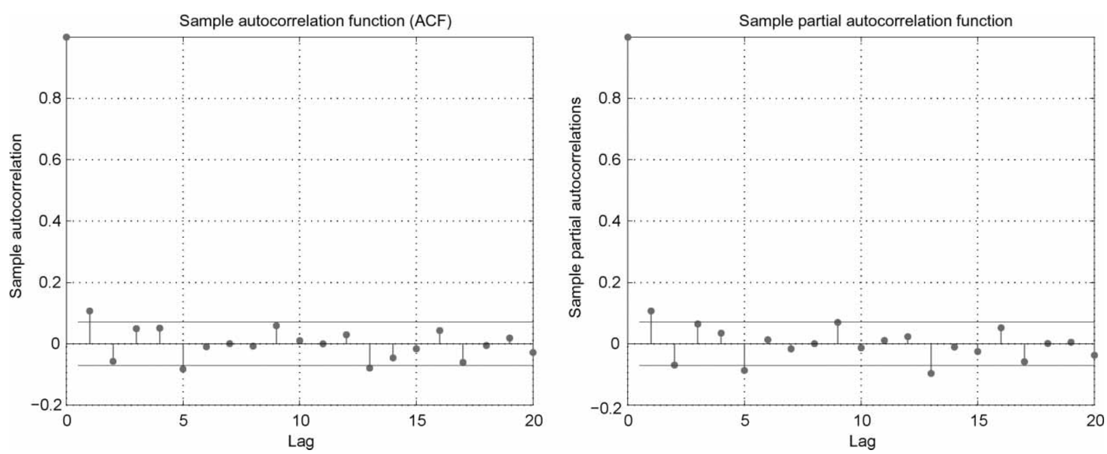


Figure 3: Autocorrelation and partial autocorrelation function for FTSE 100

observing a particular data sample, ie the likelihood function, will be denoted by $\mathcal{L}(\Theta)$.

In general, for stochastic processes, the Markov property is sufficient to be able to write the likelihood along a time series of observations as a product of transition likelihoods on single time steps between two adjacent instants. The key idea there is to notice that, for a Markov process x_t, one can write, again denoting f the probability density of the vector random sample:

$$\begin{aligned}\mathcal{L}(\Theta) &= f_{X(t_0),X(t_1),\ldots,X(t_n);\Theta} \\ &= f_{X(t_n)|X(t_{n-1});\Theta} \cdot f_{X(t_{n-1})|X(t_{n-2});\Theta} \\ &\quad \cdots f_{X(t_1)|X(t_0);\Theta} \cdot f_{X(t_0);\Theta}\end{aligned} \tag{9}$$

In the more extreme cases where the observations in the data series are iid, the problem is considerably simplified, as $fX(t_i)|X(t_{i-1});\Theta = fX(t_i);\Theta$ and the likelihood function becomes the product of the probability density of each data point in the sample without any conditioning. In the GBM case, this happens if the maximum likelihood estimation is done on log-returns rather than on the levels. By defining the X as

$$X(t_i) := \log S(t_i) - \log S(t_{i-1}), \tag{10}$$

one sees that they are already independent so that one does not need to use explicitly the above decomposition (9) through transitions. For general Markov processes different from GBM, or if one had worked at level space $S(t_i)$ rather than in return space $X(t_i)$, this can however be necessary: see, for example, the Vasicek example later on.

The likelihood function for iid variables is in general:

$$\mathcal{L}(\Theta) = f_\Theta(x_1, x_2, x_3, \ldots, x_n) = \prod_{i=1}^{n} f_\Theta(x_i) \tag{11}$$

The MLE estimate $\hat{\Theta}$ is found by maximising the likelihood function. As the product of density values could become very small, which would cause numerical problems with handling such numbers, the likelihood function is usually converted to the log likelihood $\mathcal{L}^* = \log \mathcal{L}$ (this is allowed, as maxima are not affected by monotone transformations). One must be careful not to confuse the log taken to move from levels S to returns X with the log used to move from the likelihood to the log-likelihood. The two are not directly related, in that one takes the log-likelihood in general even in situations when there are no log-returns but just general processes.

For the iid case, the log-likelihood reads:

$$\mathcal{L}^*(\Theta) = \sum_{i=1}^{n} \log f_\Theta(x_i) \tag{12}$$

The maximum of the log likelihood usually has to be found numerically using an optimisation algorithm (eg Excel Solver or fminsearch in MATLAB, although for potentially multimodal likelihoods, such as possibly the jump diffusion cases further on, one may have to preprocess the optimisation through a coarse grid global method such as genetic algorithms or simulated annealing, so as to avoid getting stuck in a local minima given by a lower mode). In the case of the GBM, the log-returns increments form normal iid random variables, each with a well known density $f_\Theta(x) = f(x; m, v)$, determined by mean and variance. Based on Equation (3) (with $u = t_{i+1}$ and $t = t_i$), the parameters are computed as follows:

$$m = \left[\hat{\mu} - \frac{1}{2}\widehat{\sigma^2}\right]\Delta t \quad v = \widehat{\sigma^2}\Delta t \quad (13)$$

can be corrected by multiplying the estimator by $n/(n-1)$; however, for large data samples the difference is very small.) Similarly, one can find the maximum likelihood estimates for both parameters simultaneously using the following numerical search algorithm.

Code 2: MATLAB® MLE function for iid normal distribution

```
function [mu sigma] = GBM_calibration(S,dt,params)
  Ret=price2ret(S);
  n=length(Ret);
  options = optimset('MaxFunEvals', 100000, 'MaxIter', 100000);
  fminsearch(@normalLL, params, options);
  function mll = normalLL(params)
    mu=params (1);sigma=abs(params(2));
    ll=n*log(1/sqrt(2*pi*dt)/sigma)+sum(-(Ret-mu*dt).^2/2/(dt*sigma^2));
    mll=-ll;
  end
end
```

The estimates for the GBM parameters are deduced from the estimates of m and v. In the case of the GBM, the MLE method provides closed form expressions for m and v through MLE of Gaussian iid samples.[1] This is done by differentiating the Gaussian density function with respect to each parameter and setting the resulting derivative equal to zero. This leads to the following well-known closed form expressions for the sample mean and variance of the log-returns samples x_i for $X_i = \log S(t_i) - \log S(t_{i-1})$

$$\hat{m} = \sum_{i=1}^{n} x_i/n \quad \hat{v} = \sum_{i=1}^{n}(x_i - \hat{m})^2/n. \quad (14)$$

(Note that the estimator for the variance is biased, as $\mathbb{E}(\hat{v}) = v(n-1)/n$. The bias

An important remark is in order for this important example: given that $x_i = \log s(t_i) - \log s(t_{i-1})$, where s is the observed sample time series for GBM S, $\hat{m}$ is expressed as:

$$\hat{m} = \sum_{i=1}^{n} x_i/n = \frac{\log(s(t_n)) - \log(s(t_0))}{n} \quad (15)$$

The only relevant information used on the whole S sample are its initial and final points. All of the remaining sample, possibly consisting of hundreds of observations, is useless. In the event that v is known, the drift estimation can be statistically consistent, converging in probability to the true drift value, only if the number of points times the time step tends to infinity.[2] This confirms that increasing the observations between two given fixed instants is useless; the only helpful limit would be letting the final time of the sample go to infinity, or letting the number of observations increase while keeping a fixed time step.

This is linked to the fact that drift estimation for random processes like GBM is extremely difficult. In a way, if one could really estimate the drift, one would know locally the direction of the market in the future, which is effectively one of the most difficult problems. The impossibility of estimating this quantity with precision is what keeps the market alive. This is also one of the reasons for the success of risk-neutral valuation: in risk-neutral pricing, one is allowed to replace this very difficult drift with the risk-free rate.

The drift estimation will improve considerably for mean-reverting processes, as shown in the second part of this paper.

Confidence levels for parameter estimates

For most applications it is important to know the uncertainty around the parameter estimates obtained from historical data, which is statistically expressed by confidence levels. Generally, the shorter the historical data sample, the larger the uncertainty as expressed by wider confidence levels.

For the GBM, one can derive closed form expressions for the confidence interval (CI) of the mean and standard deviation of returns, using the Gaussian law of independent return increments and the fact that the sum of independent Gaussians is still Gaussian, with the mean given by the sum of the mean values and the variance given by the sum of variances. Let $C_n = X_1 + X_2 + X_3 + \ldots + X_n$, then one knows that $C_n \sim N(nm, nv)$ (even without the assumption of Gaussian log-returns increments, but knowing only that log-returns increments are iid, one can reach a similar Gaussian limit for C_n, but only asymptotically in n, thanks to the central limit theorem). One then has:

$$P\left(\frac{C_n - nm}{\sqrt{v}\sqrt{n}} \leq z\right) = P\left(\frac{C_n/n - m}{\sqrt{v}/\sqrt{n}} \leq z\right) = \Phi(z) \qquad (16)$$

where $\Phi()$ denotes the cumulative standard normal distribution. C_n/n, however, is the MLE estimate for the sample mean m. Therefore the 95th confidence level for the sample mean is given by:

$$\frac{C_n}{n} - 1.96\frac{\sqrt{v}}{\sqrt{n}} \leq m \leq \frac{C_n}{n} + 1.96\frac{\sqrt{v}}{\sqrt{n}} \qquad (17)$$

This shows that as n increases the CI tightens, which means the uncertainty around the parameter estimates declines.

As a standard Gaussian squared is a chi squared, and the sum of n independent chi squared is a chi squared with n degrees of freedom, this gives:

$$P\left(\sum_{i=1}^{n}\left(\frac{(x_i - \hat{m})}{\sqrt{v}}\right)^2 \leq z\right) = P\left(\frac{n}{v}\hat{v} \leq z\right) = \chi_n^2(z), \qquad (18)$$

hence under the assumption that the returns are normal one finds the following CI for the sample variance:

$$\frac{n}{q_U^{\chi}}\hat{v} \leq v \leq \frac{n}{q_L^{\chi}}\hat{v} \qquad (19)$$

where q_L^χ, q_U^χ are the quantiles of the chi-squared distribution with n degrees of freedom corresponding to the desired confidence level.

Remark 1 (Confidence levels and simulated parameters distributions): *In the absence of these explicit relationships one can also derive the distribution of each of the parameters numerically, by simulating the sample parameters as follows. After the parameters have been estimated (either through MLE or with the best fit to the chosen model, call these the 'first parameters'), simulate a sample path of length equal to the historical sample with those parameters. For this sample path one then estimates the best-fit parameters as if the historical sample were the simulated one, to obtain a sample of the parameters. Then again, simulating from the first parameters, one goes through this procedure and obtains another sample of the parameters. By iterating, one builds a random distribution of each parameter, all starting from the first parameters. An implementation of the simulated parameter distributions is provided in the next MATLAB function.*

The charts in Figure 4 plot the simulated distribution for the sample mean and sample variance against the normal and chi-squared distribution given by the theory and the Gaussian distribution. In both cases, the simulated distributions are very close to their theoretical limits. The confidence levels and distributions of individual estimated parameters only tell part of the story. For example, in value-at-risk calculations, one would be interested in finding a particular percentile of the asset distribution.

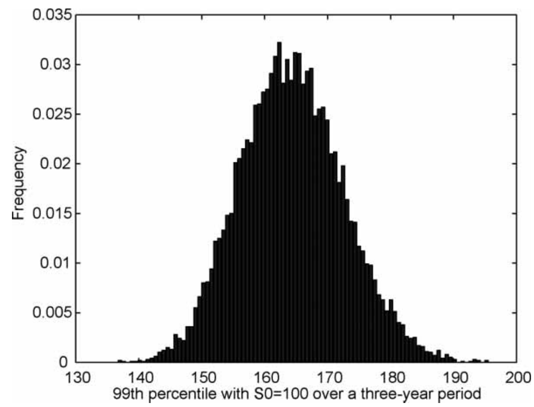


Figure 5: Distribution for 99th percentile of FTSE 100 over a three-year period

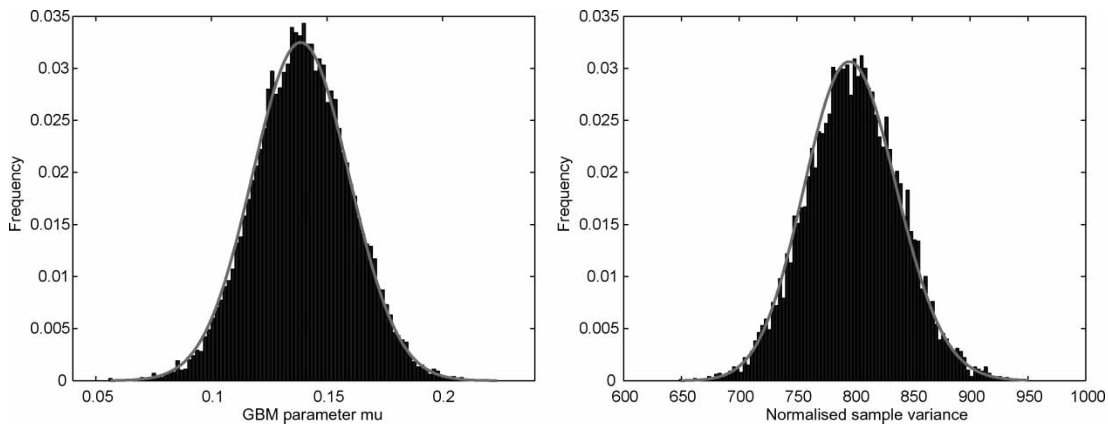


Figure 4: Distribution of sample mean and variance

Code 3: MATLAB® simulating distribution for sample mean and variance

```
%load 'C:\Research\QS CVO Technical Paper\Report\Chapter1\ftse100monthly1940.mat'
load 'ftse100monthly1940.mat'
S=ftse100monthly1940data;
ret=log(S(2:end,2)./S(1:end-1,2));
n=length(ret)

m=sum(ret)/n
v=sqrt(sum((ret-m).^2)/n)
SimR=normrnd(m,v,10000,n);

m_dist=sum(SimR,2)/n;
m_tmp=repmat(m_dist,1,799);
v_dist=sum((SimR-m_tmp).^2,2)/n;

dt=1/12;
sigma_dist=sqrt(v_dist/dt);
mu_dist=m_dist/dt+0.5*sigma_dist.^2;

S0=100;
pct_dist=S0*logninv(0.99,sigma_dist,mu_dist);
```

In expected shortfall calculations, one would be interested in computing the expectation of the asset conditional on exceeding such percentile. In the case of the GBM, one knows the marginal distribution of the asset price to be log normal and one can compute percentiles analytically for a given parameter set. The MATLAB® Code 3, which computes the parameter distributions, also simulates the distribution for the 99th percentile of the asset price after three years when re-sampling from simulation according to the first estimated parameters. The histogram is plotted in Figure 5.

FAT TAILS: GARCH MODELS

Generalised autoregressive conditional heteroscedasticity (GARCH) models were first introduced by Bollerslev.[3] The assumption underlying a GARCH model is that volatility changes with time and with the past information. In contrast, the GBM model introduced in the previous section assumes that volatility (σ) is constant. GBM could have been easily generalised to a time-dependent but fully deterministic volatility, not depending on the state of the process at any time. GARCH instead allows the volatility to depend on the evolution of the process. In particular, the GARCH(1,1) model, introduced by Bollerslev,[3] assumes that the conditional variance (ie conditional on information available up to time t_i) is given by a linear combination of the past variance $\sigma(t_{i-1})^2$ and squared values of past returns:

$$\frac{\Delta S(t_i)}{S(t_i)} = \mu\Delta t_i + \sigma(t_i)\Delta W(t_i)$$

$$\sigma(t_i)^2 = \omega\bar{\sigma}^2 + \alpha\sigma(t_{i-1})^2 + \beta\epsilon(t_{i-1})^2$$

$$\epsilon(t_i)^2 = (\sigma(t_i)\Delta W(t_i))^2 \tag{20}$$

where $\Delta S(t_i) = S(t_{i+1}) - S(t_i)$ and the first equation is just the discrete time counterpart of Equation (1).

If one had just written the discrete time version of GBM, $\sigma(t_i)$ would just be a known function of time and one would stop with the first equation above. But in this case, in the above GARCH model $\sigma(t)^2$ is assumed itself to be a stochastic process, although one depending only on past information. Indeed, for example,

$$\sigma(t_2)^2 = \omega\bar{\sigma}^2 + \beta\epsilon(t_1)^2 + \alpha\omega\bar{\sigma}^2 + \alpha^2\sigma(t_0)^2$$

so that $\sigma(t_2)$ depends on the random variable $\epsilon(t_1)^2$. Clearly conditional on the information at the previous time t_1 the volatility is no longer random, but unconditionally it is, contrary to the GBM volatility.

In GARCH, the variance is an autoregressive process around a long-term average variance rate $\bar{\sigma}^2$. The GARCH(1,1) model assumes one lag period in the variance. Interestingly, with $\omega = 0$ and $\beta = 1 - \alpha$, one recovers the exponentially weighted moving average model, which unlike the equally weighted MLE estimator in the previous section, puts more weight on the recent observations. The GARCH(1,1) model can be generalised to a GARCH(p, q) model with p lag terms in the variance and q terms in the squared returns. The GARCH parameters can be estimated by ordinary least squares (OLS) regression or by maximum likelihood estimation, subject to the constraint $\omega + \alpha + \beta = 1$, a constraint summarising the fact that the variance at each instant is a weighted sum where the weights add up to 1.

As a result of the time-varying state-dependent volatility, the unconditional distribution of returns is non-Gaussian and exhibits so-called 'fat' tails. Fat-tail distributions allow the realisations of the random variables having those distributions to assume values that are more extreme than in the normal/Gaussian case, and are therefore more suited to model risks of large movements than the Gaussian. Therefore, returns in GARCH will tend to be 'riskier' than the Gaussian returns of GBM.

GARCH is not the only approach to account for fat tails, and this paper will introduce other approaches in due course.

The particular structure of the GARCH model allows for volatility clustering (autocorrelation in the volatility), which means a period of high volatility will be followed by high volatility and conversely a period of low volatility will be followed by low volatility. This is an often observed characteristic of financial data. For example, Figure 8 shows the monthly volatility of the FTSE 100 index and the corresponding autocorrelation function, which confirms the presence of autocorrelation in the volatility.

A long line of research followed the seminal work of Bollerslev leading to customisations and generalisations including exponential GARCH (EGARCH) models, threshold GARCH (TGARCH) models and nonlinear GARCH (NGARCH) models, among others. In particular, the NGARCH model introduced by Engle and Ng[4] is an interesting extension of the GARCH(1,1) model, as it allows for asymmetric behaviour in the volatility such that 'good news' or positive returns yield subsequently lower volatility, while 'bad news' or negative returns yield a subsequent increase in volatility. The NGARCH is specified as follows:

$$\frac{\Delta S(t_i)}{S(t_i)} = \mu\Delta t_i + \sigma(t_i)\Delta W(t_i)$$

$$\sigma(t_i)^2 = \omega + \alpha\sigma(t_{i-1})^2 + \beta(\epsilon(t_{i-1}) - \gamma\sigma(t_{i-1}))^2$$
$$\epsilon(t_i)^2 = (\sigma(t_i)\Delta W(t_i))^2 \quad (21)$$

NGARCH parameters ω, α and β are positive and α, β and γ are subject to the stationarity constraint $\alpha + \beta(1 + \gamma^2) < 1$.

Compared with the GARCH(1,1), this model contains an additional parameter γ, which is an adjustment to the return innovations. Gamma equal to zero leads to the symmetric GARCH(1,1) model, in which positive and negative $\epsilon(t_{i-1})$ have the same effect on the conditional variance. In the NGARCH γ is positive, thus reducing the impact of good news ($\epsilon(t_{i-1}) > 0$) and increasing the impact of bad news ($\epsilon(t_{i-1}) < 0$) on the variance.

Code 4: MATLAB® code to simulate GBM with NGARCH vol

```
function S=NGARCH_simulation(N_Sim,T,params,S0)
  mu_=params(1);omega_=params(2);alpha_=params(3);beta_=params(4);
      gamma_=params(5);
  S=S0*ones(N_Sim,T);
  v0=omega_/(1-alpha_-beta_);
  v=v0*ones(N_Sim,1);
  BM=normrnd(0,1,N_Sim,T);
  for i=2:T
      sigma_=sqrt(v);
      mymean=(mu_-0.5*sigma_.^2);
      S(:,i)=S(:,i-1).*exp(mymean+sigma_.*BM(:,i));
      v=omega_+alpha_*v+beta_*(sqrt(v).*BM(:,i)-gamma_*sqrt(v)).^2;
  end
end
```

This routine extends Code 1 by simulating the GBM with the NGARCH volatility structure. Notice that there is only one loop in the simulation routine, which is the time loop. As the GARCH model is path-dependent it is no longer possible to vectorise the simulation over the time step. For illustration purposes, this model is calibrated to the historic asset returns of the FTSE 100 index. In order to estimate the parameters for the NGARCH GBM, the maximum-likelihood method introduced in the previous section is used.

Although the variance is changing over time and is state-dependent and hence stochastic conditional on information at the initial time, it is locally deterministic, which means that conditional on the information at t_{i-1} the variance is known and constant in the next step between time t_{i-1} and t_i. In other words, the variance in the GARCH models is locally constant. Therefore, conditional on the variance at t_{i-1} the density of the process at the next instant is still normal (note that the unconditional density of the NGARCH GBM is no longer normal, but an unknown heavy-tailed distribution). Moreover, the conditional returns are still independent. The log-likelihood function for the sample $x_1, \ldots, x_i, \ldots, x_n$ of:

$$X_i = \frac{\Delta S(t_i)}{S(t_i)}$$

is given by:

$$\mathcal{L}^*(\Theta) = \sum_{i=1}^{n} \log f_\Theta(x_i; x_{i-1}) \quad (22)$$

$$f_\Theta(x_i; x_{i-1}) = f_N(x_i; \mu, \sigma_i^2)$$
$$= \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma_i^2}\right) \quad (23)$$

$$\mathcal{L}^*(\Theta) = constant + \frac{1}{2}(-\log(\sigma_i^2) - (x_i-\mu)^2/\sigma_i^2) \quad (24)$$

where f_Θ is the density function of the normal distribution. Note the appearance of x_{i-1} in the function, which indicates that normal density is only locally applicable between successive returns, as noticed previously. The parameter set for the NGARCH GBM includes five parameters, $\Theta = (\mu, \omega, \alpha, \beta, \gamma)$. Due to the time-changing state-dependent variance, one no longer finds explicit solutions for the individual parameters and has to use a numerical scheme to maximise the likelihood function through a solver. To do so, one only needs to maximise the term in brackets in Equation (24), as the constant and the factor do not depend on the parameters.

Going back to the advantages of the NGARCH GBM over the GARCH GBM as regards the news effect, these are further illustrated in Figure 7. Figure 7 shows that, while the NGARCH GBM reproduces the asymmetric volatility reaction to the news usually observed in the data, the simpler GARCH GBM does not. Around the 100th day, both GARCH and NGARCH volatilities react significantly to the bad news on the markets. When the market faces a high level of very good news, which happens around the 500th day in the graph, the GARCH GBM model reacts with an important increase in the volatility. In contrast, the NGARCH GBM volatility shows a much more moderate reaction, as is more appropriate. Both log returns time series are compared with the plain GBM one, where volatility is not affected by good or bad market news.

Code 5: MATLAB® code to estimate parameters for GBM NGARCH

```
function [mu_ omega_ alpha_ beta_ gamma_]=NG_calibration(data,params)
  returns = price2ret(data);
  returnsLength=length(returns);
  options = optimset('MaxFunEvals', 100000, 'MaxIter', 100000);
  fminsearch(@NG_JGBM_LL, params, options);
  function mll = NG_JGBM_LL(params)
    mu_=params (1) ;omega_=abs(params (2)) ;alpha_=abs(params (3)) ;
    beta_=abs(params (4)) ;gamma_=params (5) ;
    denum = 1-alpha_-beta_*(1+gamma_^2);
    if denum < 0; mll = intmax; return; end
    var_t = omega_/denum;
    innov = returns(1) -mu_;
    LogLikelihood = log(exp(-innov^2/var_t/2)/sqrt(pi*2*var_t));
    for time=2:returnsLength
        var_t = omega_+alpha_*var_t+beta_*(innov-gamma_*sqrt(var_t))^2;
        if var_t<0; mll = intmax; return; end;
        innov = returns(time)-mu_;
        LogLikelihood = LogLikelihood +
        log(exp(-innov^2/var_t/2)/sqrt(pi*2*var_t));
        end
        mll = -LogLikelihood;
    end
  end
```

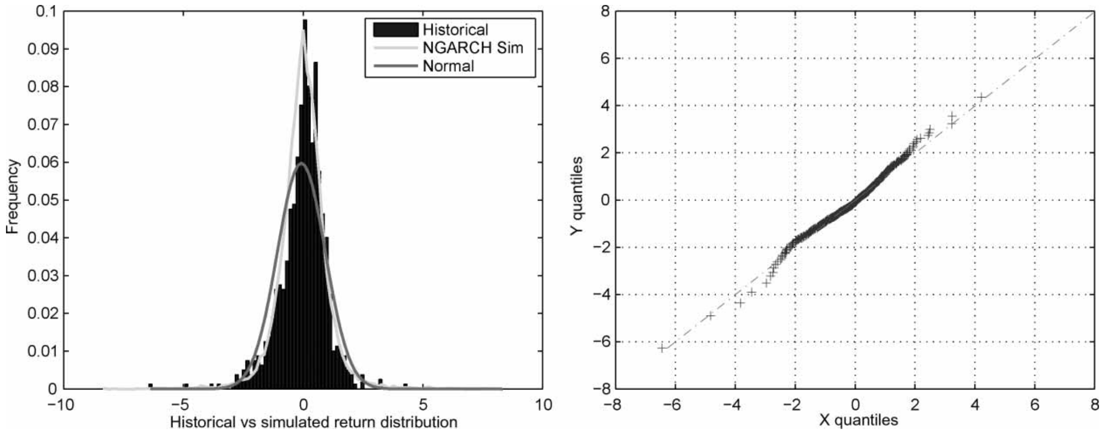


Figure 6: Historical FTSE 100 return distribution vs NGARCH return distribution

As a step further, having estimated the best-fit parameters, one can verify the fit of the NGARCH model for the FTSE 100 monthly returns using the QQ-plot. First, a large number of monthly returns using MATLAB Code 5 are simulated. The QQ-plot of the simulated returns against the historical returns shows that the GARCH model does a significantly better job when compared with the constant volatility model of the previous section. The quantiles of the simulated NGARCH distribution are much more aligned with the quantiles of the historical return distribution than was the case for the plain normal distribution (compare QQ-plots in Figures 2 and 6).

Remark 2 (Confidence levels for parameter estimates and distribution percentiles): *The GARCH model does not provide analytical solutions for either the parameter distribution or the distribution of specific percentiles of the underlying risk factor. These have to be derived using the bootstrapping technique as described in Remark 1. In general, the bootstrapping approach would be to simulate a sample path with the estimated best-fit parameters and re-estimate the parameters for each sample path.*

FAT TAILS: JUMP DIFFUSION MODELS

Compared with the normal distribution of the returns in the GBM model, the log-returns of a GBM model with jumps are often leptokurtotic (fat-tailed distribution). This section discusses the implementation of a jump-diffusion model with compound Poisson jumps. In this model, a standard time homogeneous Poisson process dictates the arrival of jumps, and the jump sizes are random variables with some preferred distribution (typically Gaussian, exponential or multinomial). Merton applied this model to option pricing for stocks that can be subject to idiosyncratic shocks, modelled through jumps.[5] The model stochastic differential equations (SDE) can be written as:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t) + S(t)dJ_t \tag{25}$$

where again W_t is a univariate Wiener process and J_t is the univariate jump

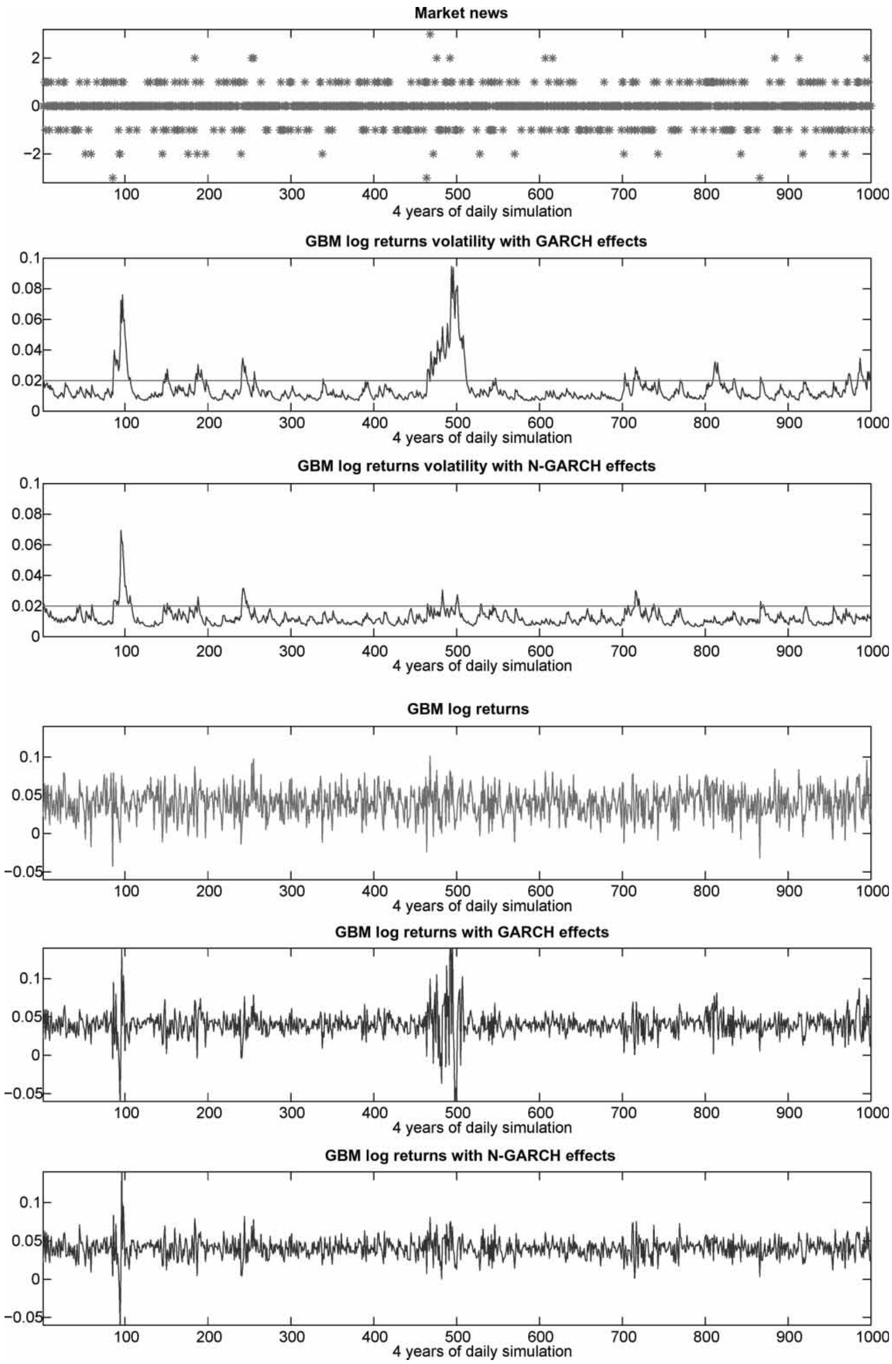


Figure 7: NGARCH news effect vs GARCH news effect

process defined by:

$$J_T = \sum_{j=1}^{N_T} (Y_j - 1), \text{ or } dJ(t) = (Y_{N(t)} - 1)dN(t), \qquad (26)$$

where $(N_T)_{T \geq 0}$ follows a homogeneous Poisson process with intensity λ, and is thus distributed like a Poisson distribution with parameter λT. The Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time. As the Poisson distribution is the limiting distribution of a binomial distribution where the number of repetitions of the Bernoulli experiment n tends to infinity, each experiment with probability of success λ/n, it is usually adopted to model the occurrence of rare events. For a Poisson distribution with parameter λT, the Poisson probability mass function is:

$$f_P(x, \lambda T) = \frac{\exp(-\lambda T)(\lambda T)^x}{x!},$$
$$x = 0, 1, 2, \ldots$$

(recall that by convention $0! = 1$). The Poisson process $(N_T)_{T \geq 0}$ is counting the number of arrivals in $[1, T]$, and Y_j is the size of the j-th jump. The Y are iid log-normal variables ($Y_j \sim \exp(\mathcal{N}(\mu_Y, \sigma_Y^2))$) that are also independent from the Brownian motion W and the basic Poisson process N.

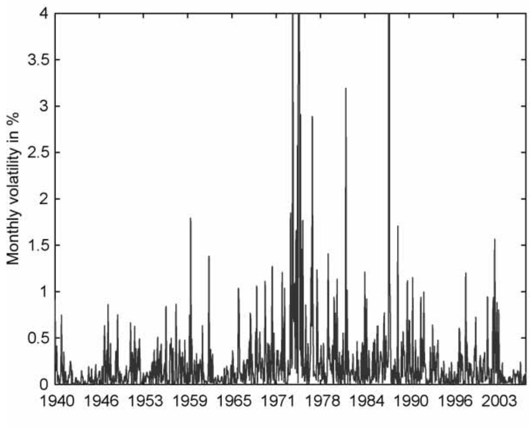
Also in the jump diffusion case, and for comparison with (2), it can be useful to write the equation for the logarithm of S, which in this case reads:

$$d\log S(t) = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW(t)$$
$$+ \log(Y_{N(t)})dN(t) \text{ or, equivalently}$$
$$d\log S(t) = \left(\mu + \lambda\mu_Y - \frac{1}{2}\sigma^2\right)dt$$
$$+ \sigma dW(t) + [\log(Y_{N(t)})dN(t)$$
$$- \mu_Y \lambda dt] \qquad (27)$$

where now both the jump (between square brackets) and diffusion shocks have zero mean, as is shown below. Here too it is convenient to do MLE estimation in log-returns space, so that here too one defines $X(t_i)$ values according to (10).

The solution of the SDE for the levels S is given easily by integrating the log equation:

$$S(T)$$
$$= S(0)\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W(T)\right)$$
$$\times \prod_{j=1}^{N(T)} Y_j \qquad (28)$$



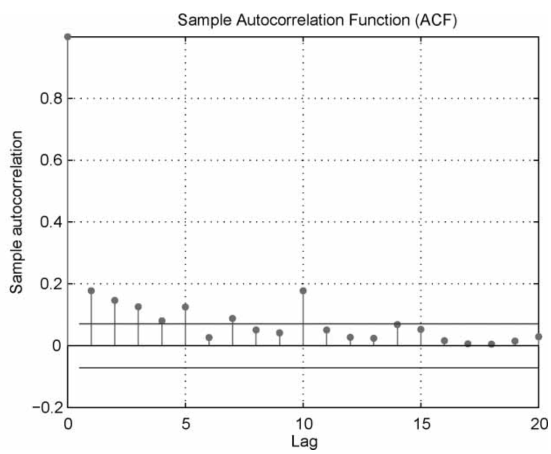


Figure 8: Monthly volatility of the FTSE 100 and autocorrelation function

Code 6: MATLAB® code to simulate GBM with jumps

```
function S = JGBM_simulation(N_Sim,T,dt,params,S0)
  mu_star=params (1) ;sigma_=params (2) ;lambda_=params (3) ;
  mu_y_=params (4) ;sigma_y_=params (5) ;
  M_simul = zeros(N_Sim,T);
  for t=1:T
    jumpnb = poissrnd(lambda_*dt,N_Sim,1);
    jump = normrnd(mu_y_*(jumpnb-lambda_*dt),sqrt(jumpnb)*sigma_y_);
    M_simul(:,t) = mu_star*dt + sigma_*sqrt(dt)*randn(N_Sim,1) + jump;
  end
  S=ret2price(M_simul',S0)';
end
```

The discretisation of Equation (28) for the given time step Δt is:

$$S(t) = S(t-\Delta t)\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}\varepsilon_t\right)\prod_{j=1}^{n_t} Y_j \tag{29}$$

where $\varepsilon \sim \mathcal{N}(0,1)$ and $n_t = N_t - N_{t-\Delta t}$ counts the jumps between time $t-\Delta t$ and t. When rewriting this equation for the log-returns $X(t) := \Delta\log(S(t)) = \log(S(t)) - \log(S(t-\Delta t))$ one obtains:

$$\begin{aligned} X(t) &= \Delta\log(S(t)) \\ &= \mu^*\Delta t + \sigma\sqrt{\Delta t}\,\varepsilon_t + \Delta J_t^* \end{aligned} \tag{30}$$

where the jumps ΔJ_t^* in the time interval Δt and the drift μ^* are defined as:

$$\begin{aligned} \Delta J_t^* &= \sum_{j=1}^{n_t}\log(Y_j) - \lambda\Delta t\mu_Y, \\ \mu^* &= \left(\mu + \lambda\mu_Y - \frac{1}{2}\sigma^2\right) \end{aligned} \tag{31}$$

so that the jumps ΔJ_t^* have zero mean. As the calibration and simulation are done in a discrete time framework, one uses the fact that, for the Poisson process with rate λ, the number of jump events occurring in the fixed small time interval Δt has the expectation of $\lambda\Delta t$. Compute:

$$\begin{aligned} \mathbb{E}(\Delta J_t^*) &= E\left(\sum_{j=1}^{n_t}\log Y_j\right) - \lambda\Delta t\mu_Y \\ &= E\left(E\left(\sum_{j=1}^{n_t}\log Y_j \middle| n_t\right)\right) - \lambda\Delta t\mu_Y \\ &= E(n_t\mu_Y) - \lambda\Delta t\mu_Y = 0 \end{aligned}$$

The simulation of this model is best done by using Equation (30).

Conditional on the jumps occurrence n_t, the jump disturbance ΔJ_t^* is normally distributed as:

$$\Delta J_t^*|_{n_t} \sim \mathcal{N}((n_t - \lambda\Delta t)\mu_Y, n_t\sigma_Y^2).$$

Thus the conditional distribution of $X(t) = \Delta\log(S(t))$ is also normal and the first two conditional moments are:

Code 7: MATLAB® code to calibrate GBM with jumps

```
function [mu_star sigma_ lambda_ mu_y_ sigma_y_] = JGBM_calibration(data,dt,params)
  returns = price2ret(data);
  dataLength=length(data);
  options = optimset('MaxFunEvals', 100000, 'MaxIter', 100000);
  params = fminsearch(@FT_JGBM_LL, params, options);
    mu_star=params (1) ;sigma_=abs(params (2)) ;lambda_=abs(params (3)) ;
    mu_y_=params (4) ;sigma_y_=abs(params (5)) ;

  function mll = FT_JGBM_LL(params)
  mu_=params (1) ;sigma_=abs(params (2)) ;lambda_=abs(params (3)) ;
  mu_y_=params (4) ;sigma_y_=abs(params (5)) ;
  Max_jumps = 5;
  factoriel = factorial(0:Max_jumps);
  LogLikelihood = 0;
  for time=1:dataLength
  ProbDensity = 0;
      for jumps=0:Max_jumps-1
      jumpProb = exp(-lambda_*dt)*(lambda_*dt)^jumps/factoriel(jumps+1);
      condVol = dt*sigma_^2+jumps*sigma_y_^2;
      condMean= mu_*dt+mu_y_*(jumps-lambda_*dt);
      condToJumps = exp(-(data(time)-condMean)^2/condVol/2)/sqrt(pi*2*condVol);
      ProbDensity = ProbDensity + jumpProb*condToJumps;
      end
      LogLikelihood = LogLikelihood + log(ProbDensity);
    end
    mll = -LogLikelihood;
  end
end
```

$$\mathbb{E}(\Delta\log(S(t))|n_t) = \mu^*\Delta t + (n_t - \lambda\Delta t)\mu_Y = (\mu - \sigma^2/2)\Delta t + n_t\mu_Y, \tag{32}$$

$$\mathrm{Var}\,(\Delta\log(S(t))|n_t) = \sigma^2\Delta t + n_t\sigma_Y^2 \tag{33}$$

The probability density function is the sum of the conditional probabilities density weighted by the probability of the conditioning variable, ie the number of jumps. The calibration of the model parameters can then be done using the MLE on log-returns X conditioning on the jumps:

$$\underset{\mu*,\mu_Y,\lambda>0,\sigma>0,\sigma_Y>0}{\arg\max} \mathcal{L}^* = \log(\mathcal{L}) \tag{34}$$

The log-likelihood function for the returns $X(t)$ at observed times $t = t_1, \ldots, t_n$ with values $x_1, \ldots, x_n$, with $\Delta t = t_i - t_{i-1}$, is then given by:

$$\mathcal{L}^*(\Theta) = \sum_{i=1}^{n} \log f\,(x_i; \mu, \mu_Y, \lambda, \sigma, \sigma_Y) \tag{35}$$

$$f\,(x_i; \mu, \mu_Y, \lambda, \sigma, \sigma_Y) = \sum_{j=0}^{+\infty} P(n_t = j) f_{\mathcal{N}}(x_i; (\mu - \sigma^2/2)\Delta t + j\mu_Y, \sigma^2\Delta t + j\sigma_Y^2) \tag{36}$$

This is an infinite mixture of Gaussian random variables, each weighted by a Poisson probability $P(n_t = j) = f_P(j;\lambda\Delta t)$. Note that, if Δt is small, typically the Poisson process jumps at most once. In that case, the above expression

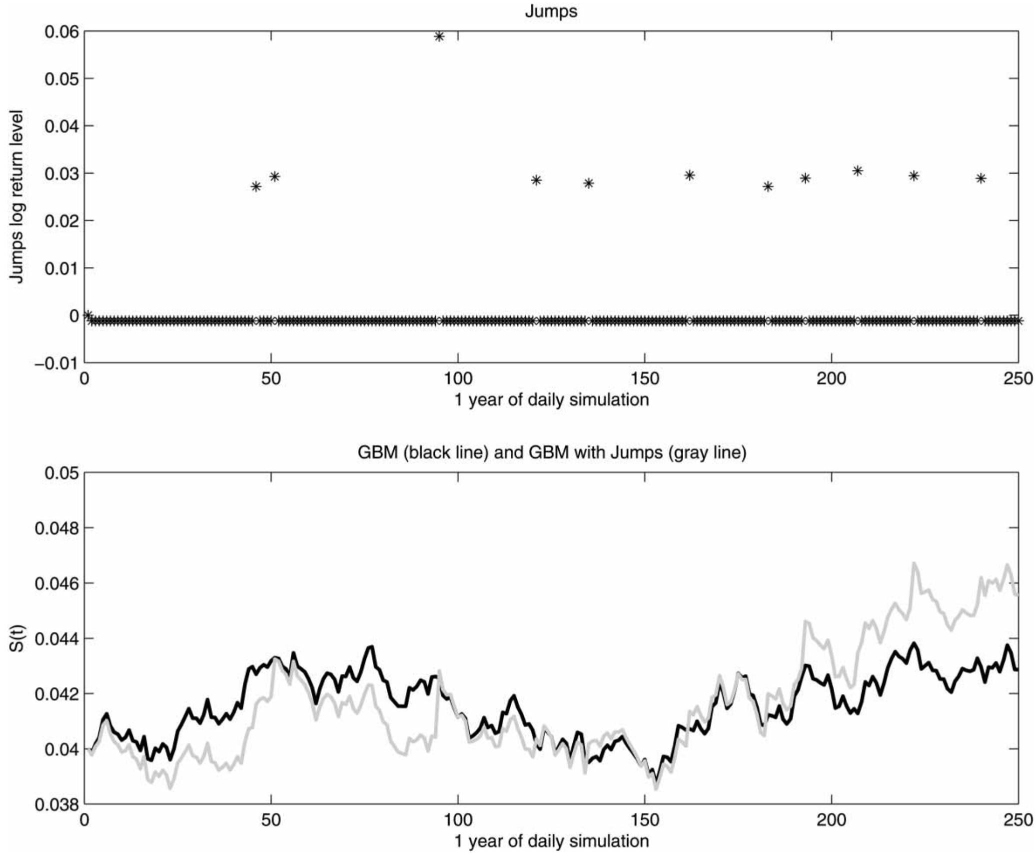


Figure 9: The effect of the jumps on the simulated GBM path and on the density probability of the log returns $\mu^* = 0$, $\sigma = 0.15$, $\lambda = 10$, $\mu_Y = 0.03$, $\sigma_Y = 0.001$

simplifies as:

$$
\begin{aligned}
f(x_i; \mu, \mu_Y, \lambda, \sigma, \sigma_Y) \\
&= (1 - \lambda\Delta t) f_{\mathcal{N}}(x_i; (\mu - \sigma^2/2)\Delta t, \sigma^2 \Delta t) \\
&\quad + \lambda \Delta t f_{\mathcal{N}}(x_i; (\mu - \sigma^2/2)\Delta t \\
&\quad + \mu_Y, \sigma^2 \Delta t + \sigma_Y^2)
\end{aligned}
$$

ie a mixture of two Gaussian random variables weighted by the probability of zero or one jump in Δt.

The model parameters are estimated by MLE for the FTSE 100 on monthly returns and different paths are then simulated. The goodness to fit is seen by showing the QQ-plot of the simulated monthly log-returns against the historical FTSE 100 monthly returns in Figure 10. This plot shows that the jumps model fits the data better than the simple GBM model.

Finally, one can add markedly negative jumps to the arithmetic Brownian motion return process (27), to then exponentiate it to obtain a new level process for S featuring a more complex jump behaviour. Indeed, if μ_Y is larger than one and σ_Y^2 is not large, jumps in the log-returns will tend to be markedly positive, which may not be realistic in some cases. Details on how to incorporate negative jumps and on the more complex mixture features this

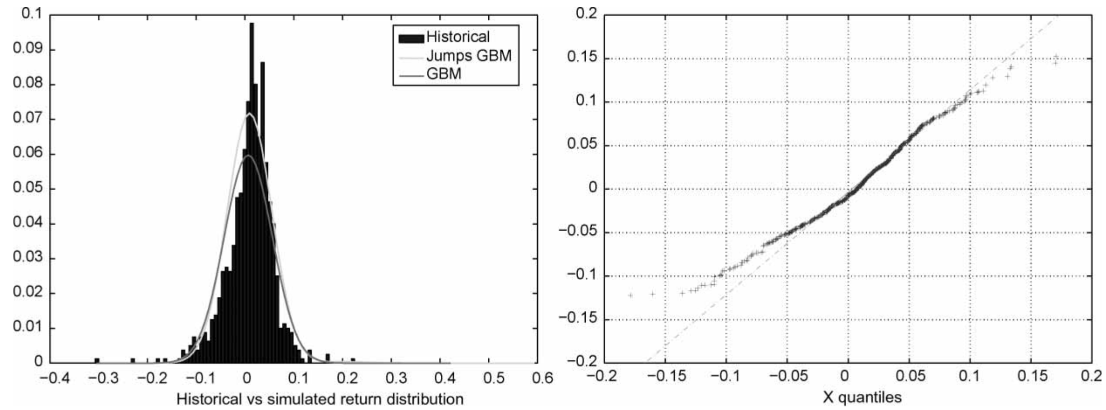


Figure 10: Historical FTSE 100 return distribution vs jump GBM return distribution

induces are given later in this paper, where the more general mean-reverting case is described.

Remark 3 (Confidence levels for parameter estimates and distribution percentiles): *A jump diffusion model such as the GARCH model does not provide analytical solutions for either the parameter distribution or the distribution of specific percentiles of the underlying risk factor. These have to be derived using the bootstrapping technique as described in Remark 1.*

FAT TAILS VARIANCE GAMMA (VG) PROCESS

The two previous sections have addressed the modelling of phenomena where numerically large changes in value ('fat tails') are more likely than in the case of normal distribution. This section presents yet another method to accomplish fat tails, and this is through a process obtained by *time changing* a Brownian motion (with drift $\bar{\mu}$ and volatility $\bar{\sigma}$) with an independent subordinator, that is an increasing process with independent and stationary increments. In a way, instead of making the volatility state-dependent like in GARCH or random like in stochastic volatility models, and instead of adding a new source of randomness like jumps, one makes the time of the process random through a new process, the subordinator. A possible related interpretative feature is that financial time, or market activity time, related to trading, is a random transform of calendar time.

Two models based on Brownian subordination are the variance gamma (VG) process and normal inverse Gaussian (NIG) process. They are a subclass of the generalised hyperbolic (GH) distributions. They are characterised by drift parameters, volatility parameter of the Brownian motion, and a variance parameter of the subordinator.

Both VG and NIG have exponential tails and the tail decay rate is the same for both VG and NIG distributions and it is smaller than for the normal distribution.

Therefore these models allow for more flexibility in capturing fat-tail features and high kurtosis of some historical financial distributions with respect to GBM.

Among fat-tail distributions obtained through subordination, this section will

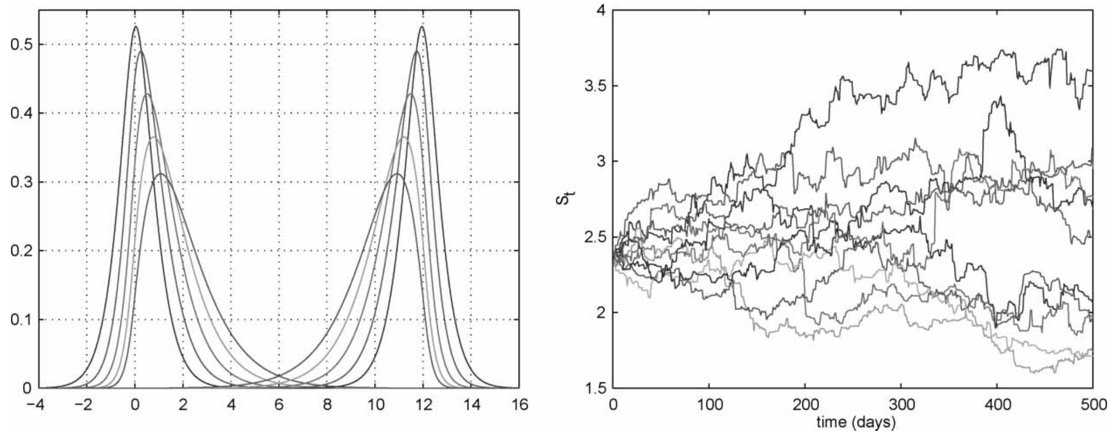


Figure 11: On the left side, VG densities for $\nu = 0.4$, $\bar{\sigma} = 0.9$, $\bar{\mu} = 0$ and various values of the parameter $\bar{\theta}$; on the right, simulated VG paths samples. To replicate the simulation, use Code 8 with the parameters $\bar{\mu} = \bar{\theta} = 0$, $\nu = 6$ and $\bar{\sigma} = 0.03$

focus on the VG distribution, which can also be thought of as a mixture of normal distributions, where the mixing weights density is given by the gamma distribution of the subordinator. The philosophy is not so different from the GBM with jumps seen before, where the return process is also a mixture. The VG distribution was introduced in the financial literature by Madan and Seneta.[6] A more advanced introduction and background on the above processes can be found elsewhere.[7–10]

The model considered for the log-returns of an asset price is of this type:

$$d \log S(t) = \bar{\mu} dt + \bar{\theta} dg(t) + \bar{\sigma} dW(g(t))$$
$$S(0) = S_0 \tag{37}$$

where $\bar{\mu}$, $\bar{\theta}$ and $\bar{\sigma}$ are real constants and $\bar{\sigma} \geq 0$ (one may also say that Equation (37) is a Lèvy process of finite variation but of relatively low activity with small jumps). This model differs from the 'usual' notation of Brownian motion mainly in the term g_t. In fact, one introduces g_t to characterise the market activity time. One can define the 'market time' as a positive increasing random process, $g(t)$, with stationary increments $g(u) - g(t)$ for $u \geq t \geq 0$. In probabilistic jargon, $g(t)$ is a subordinator. An important assumption is that $\mathbb{E}[g(u) - g(t)] = u - t$. This means that the market time has to reconcile with the calendar time between t and u on average. Conditional on the subordinator g, between two instants t and u:

$$\bar{\sigma}(W(g(u)) - W(g(t)))|_g \sim \bar{\sigma}\sqrt{(g(u) - g(t))}\boldsymbol{\epsilon}, \tag{38}$$

where $\boldsymbol{\epsilon}$ is a standard Gaussian, and hence:

$$(\log(S(u)/S(t)) - \bar{\mu}(u - t))|g \sim N(\bar{\theta}(g(u) - g(t)), \bar{\sigma}^2(g(u) - g(t))) \tag{39}$$

VG assumes $\{g(t)\} \sim \Gamma(\frac{t}{\nu}, \nu)$, a Gamma process with parameter ν that is independent from the standard Brownian motion $\{W_t\}_{t \geq 0}$. Notice also that the gamma process definition implies $\{g(u) - g(t)\} \sim \Gamma(u - \frac{t}{\nu}, \nu)$ for any u

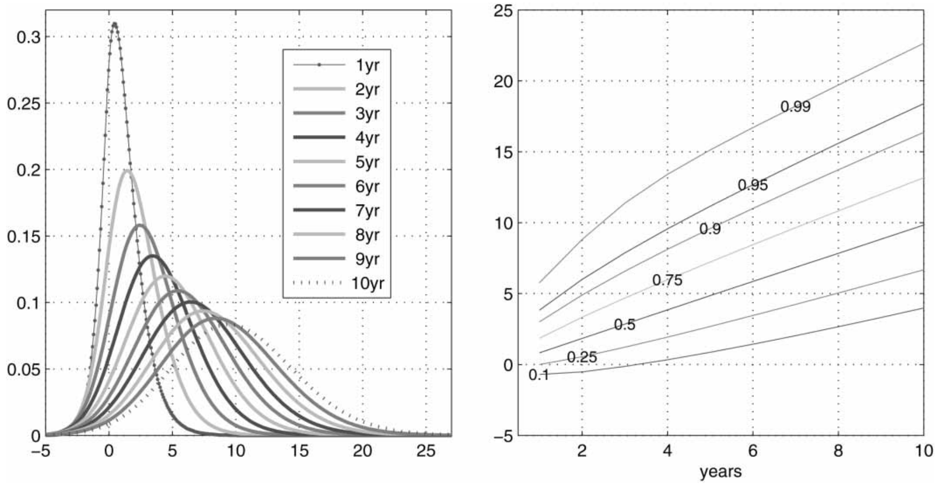


Figure 12: On the left side, VG densities for $\bar{\theta} = 1$, $\nu = 0.4$, $\bar{\sigma} = 1.4$, $\bar{\mu} = 0$ and time from 1 to 10 years; on the right side, on the y-axis, percentiles of these densities as time varies, x-axis

and t. Increments are independent and stationary gamma random variables. Consistently with the above condition on the expected subordinator, this gamma assumption implies $\mathbb{E}[g(u) - g(t)] = u - t$ and $\mathrm{Var}(g(u) - g(t)) = \nu(u - t)$.

Through iterated expectations, one has easily that:

$$\mathbb{E}[\log(S(u)) - \log(S(t))] = (\bar{\mu} + \bar{\theta})(u - t),$$

$$\mathrm{Var}(\log S(u) - \log S(t)) = (\bar{\sigma}^2 + \bar{\theta}^2 \nu) \times (u - t).$$

Mean and variance of these log-returns can be compared with the log-returns of GBM in Equation (3), and found to be the same if $(\bar{\mu} + \bar{\theta}) = \mu - \sigma^2/2$ and $\bar{\sigma}^2 + \bar{\theta}^2 \nu = \sigma^2$. In this case, the differences will be visible with higher order moments.

The process in Equation (37) is called a variance-gamma process.

Probability density function of the VG process

The unconditional density may then be obtained by integrating out g in Equation (39) through its gamma density. This corresponds to the following density function, where one sets $u - t =: \Delta t$ and $X_{\Delta t} := \log(S(t + \Delta t)/S(t))$:

$$f_{(X_{\Delta t})}(x) = \int_0^\infty f_N(x; \bar{\theta} g, \bar{\sigma}^2 g) f_\Gamma\left(g; \frac{\Delta t}{\nu}, \nu\right) dg \tag{40}$$

As mentioned above, this is again a (infinite) mixture of Gaussian densities, where the weights are given by the gamma densities rather than Poisson probabilities. The above integral

Code 8: MATLAB® code to simulate the asset model in Equation (37)

```
function S = VG_simulation(N_Sim,T,dt,params,S0)
theta = params (1) ; nu = params (2) ;sigma = params (3); mu = params (4) ;
g = gamrnd(dt/nu,nu,N_Sim,T);
S = S0*cumprod([ones(1,T); exp(mu*dt + theta*g + sigma*sqrt(g).*randn(N_Sim,T))]);
```

converges and the probability density function (PDF) of the VG process $X_{\Delta t}$ defined in Equation (37) is:

$$f_{(X_{\Delta t})}(x) = \frac{2e^{\frac{\bar{\theta}(x-\bar{\mu})}{\bar{\sigma}^2}}}{\bar{\sigma}\sqrt{2\pi}\nu^{\frac{\Delta t}{\nu}}\Gamma(1/\nu)} \times \left(\frac{|x-\bar{\mu}|}{\sqrt{\frac{2\bar{\sigma}^2}{\nu}+\bar{\theta}^2}}\right)^{\Delta t/\nu - 1/2} K_{\Delta t/\nu-1/2} \times \left(\frac{|s-\bar{\mu}|\sqrt{\frac{2\bar{\sigma}^2}{\nu}+\bar{\theta}^2}}{\bar{\sigma}^2}\right) \quad (41)$$

Here $K_\eta(\cdot)$ is a modified Bessel function of the third kind with index η, given for $\omega > 0$ by

$$K_\eta(x) = \frac{1}{2}\int_0^\infty y^{\eta-1}\exp\left\{-\frac{x}{2}(y^{-1}+y)\right\}dy$$

The four central moments of the return distribution over an interval of length Δt can be derived using the moment generating function:

$$\begin{aligned} E[X] &= \bar{X} = (\bar{\mu}+\bar{\theta})\Delta t \\ E[(X-\bar{X})^2] &= (\nu\bar{\theta}^2+\bar{\sigma}^2)\Delta t \\ E[(X-\bar{X})^3] &= (2\bar{\theta}^3\nu^2+3\bar{\sigma}^2\nu\bar{\theta})\Delta t \\ E[(X-\bar{X})^4] &= (3\nu\bar{\sigma}^4+12\bar{\theta}^2\bar{\sigma}^2\nu^2 \\ &\quad + 6\bar{\theta}^4\nu^3)\Delta t \\ &\quad + (3\bar{\sigma}^4+6\bar{\theta}^2\bar{\sigma}^2\nu \\ &\quad + 3\bar{\theta}^4\nu^2)(\Delta t)^2. \end{aligned}$$

The moment generating function of $X_{\Delta t} := \log(S(t+\Delta \mathrm{t})/S(t)), \mathbb{E}[e^{zX}]$ is given by:

Code 9: MATLAB® code of the variance gamma density function in Equation (41)

```
function fx = VGdensity(x, theta, nu, sigma, mu, T )
  v1 = 2*exp((theta*(x-mu)) / sigma^2) / ( (nu^(T/nu)) * sqrt(2*pi) * sigma * gamma(T/nu) );
  M2 = (2*sigma^2)/nu + theta^2;
  v3 = abs(x-mu)./sqrt(M2);
  v4 = v3.^(T/nu - 0.5) ;
  v6 = (abs(x-mu).*sqrt(M2))./sigma^2;
  K = besselk(T/nu - 0.5, v6);
  fx = v1.*v4.*K;
```

$$M_X(z) = e^{\bar{\mu}z}\left(1-\bar{\theta}\nu z-\frac{1}{2}\nu\bar{\sigma}^2z^2\right)^{-\frac{\Delta t}{\nu}} \quad (42)$$

Now considering the skewness and the kurtosis:

$$S = \frac{3\nu\bar{\theta}\Delta t}{((\nu\bar{\theta}^2+\bar{\sigma}^2)\Delta t)^{3/2}}$$

$$K = \frac{(3\nu\bar{\sigma}^4+12\bar{\theta}^2\bar{\sigma}^2\nu^2+6\bar{\theta}^4\nu^3)\Delta t + (3\bar{\sigma}^4+6\bar{\theta}^2\bar{\sigma}^2\nu+3\bar{\theta}^4\nu^2)(\Delta t)^2}{((\nu\bar{\theta}^2+\bar{\sigma}^2)\Delta t)^2}.$$

Calibration of the VG process

Given a time series of independent log-returns $\log(S(t_{i+1})/S(t_i))$ (with $t_{i+1} - t_i = \Delta t$), say $(x_1, x_2, \ldots, x_n)$, and denoting by $\pi = \{\bar{\mu}, \bar{\theta}, \bar{\sigma}, \nu\}$ the variance-gamma probability density function parameters that one would like to estimate, the maximum likelihood estimation (MLE) of π consists in finding $\bar{\pi}$ that maximises the logarithm of the likelihood function $\mathcal{L}(\pi) = \prod_{i=1}^{x} f(x_i; \pi)$.

and assuming $\bar{\theta}$ to be small, thus ignoring $\bar{\theta}^2$, $\bar{\theta}^3$, $\bar{\theta}^4$, obtain:

$$S = \frac{3\nu\bar{\theta}}{\bar{\sigma}\sqrt{\Delta t}}$$

$$K = 3\left(1 + \frac{\nu}{\Delta t}\right).$$

So the initial guess for the parameters π_0 will be:

$$\bar{\sigma} = \sqrt{\frac{V}{\Delta t}}$$

$$\nu = \left(\frac{K}{3} - 1\right)\Delta t$$

$$\bar{\theta} = \frac{S\bar{\sigma}\sqrt{\Delta t}}{3\nu}$$

$$\bar{\mu} = \frac{\bar{X}}{\Delta t} - \bar{\theta}$$

An example can be the calibration of the historical time series of the euro/US$ exchange rate, where the VG process is a more suitable choice than the GBM. The fit is performed on daily returns and the plot in Figure 13 illustrates the results. Note in particular that the difference between the arithmetic Brownian motion for the returns and the VG process is expressed by the time change parameter ν , all other parameters being roughly the same in the ABM and VG cases.

Remark 4 (Confidence levels for parameter estimates and distribution percentiles): *The VG model does not provide analytical solutions for the parameter distribution. These have to be derived using the bootstrapping technique as described in Remark 1. In general, the bootstrapping approach would be to simulate a sample path with the estimated best-fit parameters and re-estimate the parameters for each sample path. However, as the density*

Code 10: MATLAB® code of the MLE for variance gamma density function in Equation (41)

```
%function params = VG_calibration(data,dt)
% computation of the initial parameters
data = (EURUS$sv(:,1));
hist(data)
dt = 1;
data = price2ret(data);
M = mean(data);
V = std(data);
S = skewness(data);
K = kurtosis(data);
sigma = sqrt(V/dt);
nu = (K/3-1)*dt;
theta = (S*sigma*sqrt(dt))/(3*nu);
mu = (M/dt)-theta;
% VG MLE
pdf_VG = @(data,theta, nu, sigma, mu) VGdensity(data, theta, nu, sigma,
mu,dt);
start = [theta, nu, sigma, mu];
lb = [-intmax 0 0 -intmax];
ub = [intmax intmax intmax intmax];
options = statset('MaxIter',1000, 'MaxFunEvals',10000);
params = mle(data, 'pdf',pdf_VG, 'start',start,
'lower',lb, 'upper',ub, 'options',options);
```

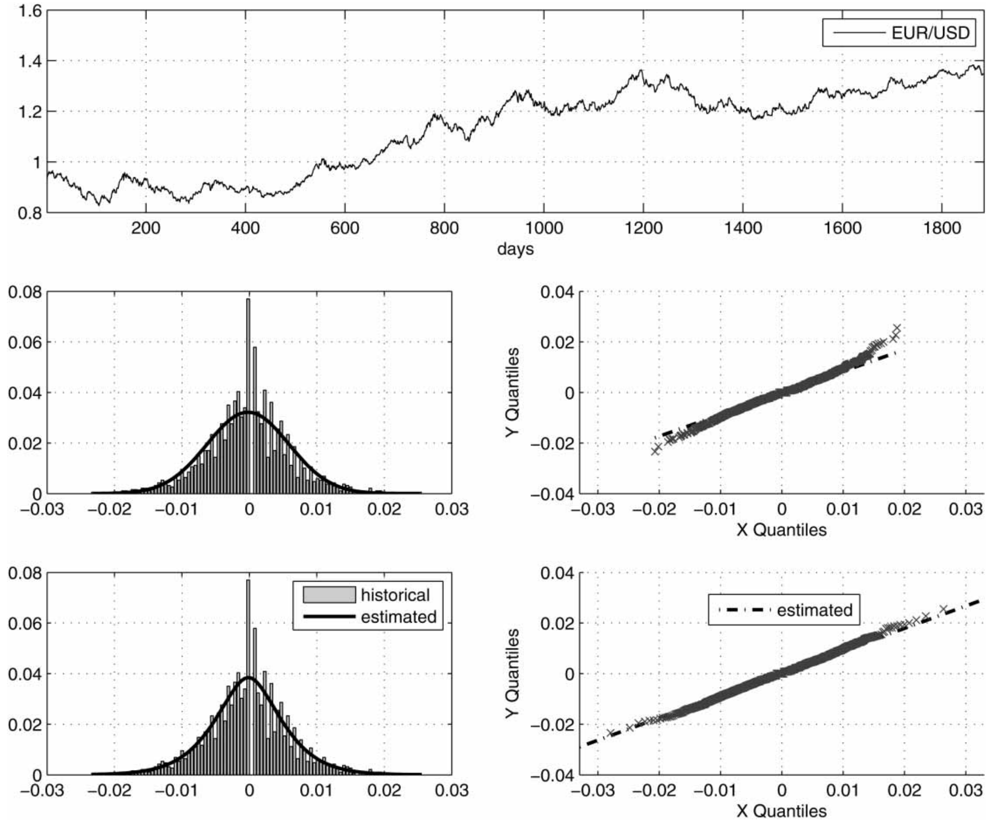


Figure 13: On the top chart, the historical trend of the euro/US$ is reported. In the middle chart the daily return distribution presents fat tails with respect to a normal distribution. The MLE parameters for the normal RV are $\mu = -0.0002$ and $\sigma = 0.0061$. The QQ-plot in the middle shows clearly the presence of tails more pronounced than in the normal case. The bottom chart is the plot of the daily return density against the estimated VG one, and the QQ-plot of the fitted daily historical FX return time series euro/US$ using a VG process with parameters $\overline{\theta} = -0.0001$, $\nu = 0.4$, $\overline{\sigma} = 0.0061$, $\overline{\mu} = -0.0001$

function for the VG process is known, it is relatively straightforward to derive percentile estimates for the underlying risk factor for each simulated set of parameters. This allows one to compute a distribution for the percentiles and derive corresponding confidence levels.

Code 11: MATLAB® code to calculate the percentiles of the VG distribution as in Figure 12.

```
clear all; close all

lb = 0; ub = 20;
options = optimset('Display','off', 'TolFun',1e-5,'MaxFunEvals',2500);
X = linspace(-15,30,500);

sigma = 1.4;
f = @(x,t) VGdensity(x, 1, .4, sigma, 0, t);

nT = 10;
t = linspace(1,10,nT);
pc = [.1 .25 .5 .75 .9 .95 .99];

for k = 1:nT
  Y(:,k) = VGdensity(X, 1, .4, sigma, 0, t(k) );
  f = @(x) VGdensity(x, 1, .4, sigma, 0, t(k));
  VG_PDF = quadl(f,-3,30);
```

```
for zz = 1:length(pc)
  percF = @(a) quadl(f,-3,a) - pc(zz);
  PercA(k,zz) = fzero(percF,1,options);
end
end

subplot(1,2,1)
plot(X,Y);
legend('1yr','2yr',' 3yr',' 4yr',' 5yr',' 6yr',' 7yr',' 8yr',' 9yr','10yr')
xlim([-5 27]);ylim([0 .32]); subplot(1,2,2); plot(t,PercA);
for zz = 1:length(pc)
  text(t(zz),PercA(zz,zz),[{'\fontsize{9}\color{black}',num2str(pc(zz))}],
  'HorizontalAlignment',...
  'center', 'BackgroundColor',[1 1 1]); hold on
end
xlim([.5 10]); xlabel('years'); grid on
```

References

1 Hogg, and Craig, 'Introduction to Mathematical Statistics' (4th edn), p. 205, example 4.

2 Jacod, 'Statistics of Diffusion Processes: Some Elements', CNR-IAMI 94.17, p. 2.

3 Bollerslev, T. (1986) 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, Vol. 31, pp. 307–327.

4 Engle, R. and Ng, V. K. (1993) 'Measuring and testing the impact of news on volatility', *Journal of Finance*, Vol. 48, pp. 1749–1778.

5 Merton, R. (1976) 'Option pricing when underlying stock returns are discontinuous', *Journal of Financial Economics*, Vol. 3, pp. 125–144.

6 Madan, D. B. and Seneta, E. (1990) 'The variance gamma (v.g.) model for share market returns', *Journal of Business*, Vol. 63, pp. 511–524.

7 Seneta, E. (2004) 'Fitting the variance-gamma model to financial data', *Journal of Applied Probability*, Vol. 41A, pp. 177–187.

8 Cont, R. and Tankov, P. (2004) 'Financial Modelling with Jump Processes', Chapman and Hall, Financial Mathematics Series.

9 Moosbrucker, T. (2006) 'Pricing CDO with correlated variance-gamma distributions', *Journal of Fixed Income*, Vol. 12, pp. 1–30.

10 Madan, D., Carr, P. and Chang, E. (1998) 'The variance gamma process and option pricing', *European Finance Review*, Vol. 2, pp. 79–105.