Stochastic volatility modelling of emission allowances futures prices in the EU ETS market

Fred Espen Benth, Marcus Eriksson and Sjur Westgaard

Abstract We conduct an empirical investigation of the logreturns on the futures prices of EU ETS emission certificates traded in the Nord Pool market from 2005-2013. Establishing stylized features such as heaviness, skewness and high kurtosis in the logreturns, we suggest to model the futures logprices with the Barndorff-Nielsen and Shephard (BNS) or the Heston stochastic volatility models. We carry out an empirical comparison between the performances of these models and investigate their stationary autocorrelation structure. In particular, as a consequence of allowing for skewness in the Heston model we find analytical expressions for the autocorrelation function of logreturns and their squares. Our analysis indicates the presence of short-range dependence in observed futures logprice returns. We conclude that the BNS model better describes the empirical features of observed futures prices than the Heston model. Our findings have relevance for the real option modelling of fossil-fueled power plants when considering emission costs.

1 Introduction

The European Union Emission Trading System (EU ETS) is a system to reduce the emissions of greenhouse gases in the European Union. EU ETS covers around 45% of EU's greenhouse gas emissions. The participants in the EU ETS can buy and sell emission allowances, called European Union Allowances (EUAs), which permit their holders to emit a certain amount of CO_2. There is a limit on the number of EUAs that can be traded each year. This limit is reduced each year in order to meet desirable emission levels in the future.

High-emitting industry players can buy allowances if needed or sell them if they judge that their current stock is more than needed. In this way companies can cut their emissions in a cost-effective way as the right to emit CO_2 becomes a tradable commodity. For example, coal- or gas-fired power plants in Europe have caps on the maximum amount of CO_2 that they can emit. All the production exceeding the limit must be covered by emission allowances. As production is stochastic, the managers of such power plants face a risk that they can hedge by actively trading in EUAs. Emissions are thus a cost when operating the fossil-fueled power plants, and their profitability depends both on the spread between the power and fuel prices and the emission costs, as recognized in the real options literature. Specifically in the real option modelling of these plants, the so-called clean spark (respectively dark) spread is of fundamental

Fred Espen Benth
Department of Mathematics, University of Oslo, P.O. Box 1053, Blindern, N–0316 Oslo, Norway, e-mail: fredb@math.uio.no

Marcus Eriksson
Department of Mathematics, University of Oslo, P.O. Box 1053, Blindern, N–0316 Oslo, Norway, e-mail: marcuskverikssson@gmail.com

Sjur Westgaard
Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, N-7491 Trondheim, Norway, e-mail: sjur.westgaard@iot.ntnu.no

importance, being the difference between the power price, the natural gas (respectively coal) price and the emission cost (see e.g. [11], [23], [6], [7], [13] and [15] for real option valuation and energy spark spreads). Secomandi and Seppi [20, Sect. 7.6] briefly discuss the operation of a power plant as a timing option where the plant produces when it is most profitable according to the spark (dark) spread. They view the usage of emission rights as an inventory disposal asset, whereby the inventory level is reduced when production takes place. Given a market for emission rights, the inventory of permits can vary over time due to consumption, purchasing or selling.

We perform an empirical investigation with the goal of specifying a continuous time model for the futures price evolution of EUAs on the EU ETS market. Our analysis relies on futures price data collected from Nord Pool between April 2005 until August 2013. From statistical analysis of the logreturns we observe skewness, heavy tails and leptokurtic behavior in the distribution, rejecting the normal distribution. To capture these stylized features we consider logprice models with stochastic volatility. We choose stochastic volatility models, rather than Lévy processes [4], since we observe volatility clustering in the time series for the logreturns, a feature that Lévy models do not capture. Two commonly used models in commodity markets with this property are the Barndorff-Nielsen and Shepard (BNS) model [1] and the Heston model [16].

The main objective of this chapter is to compare the performance of the BNS and Heston models in relation to observed data from the EU ETS market. That is, we focus on the calibration of these models to data. Our comparison will focus on how well these models capture key features of our data regarding distributional properties as well as autocorrelation functions.

We show that in the Heston model it is necessary to introduce an additional parameter in the drift for the conditional price process to account for skewness. As a consequence the stationary autocorrelation function for the logreturns is nonzero and shows a short-range dependence. We also find the stationary autocorrelation function for squared returns in the Heston model. This is based on analytical expressions for the autocorrelation function of logreturns and their squares, which are known for the BNS model and derived in this paper for the Heston model.

The Heston model, which is a mixture of geometric Brownian motion with volatility modelled by the square root of a Cox-Ingersoll-Ross (CIR) model [9], results in a stationary variance-gamma (VG) distribution for logreturns. This distribution is also obtained as a particular case in the BNS model. From the sample distribution of our data we use maximum-likelihood estimation (MLE) to obtain values for the parameters in the VG distribution and use them to estimate the parameters of the BNS and Heston models, respectively. As the BNS model is more flexible, e.g., allowing for generalized-hyperbolic (GH) distributions, we also investigate the fit of the normal inverse Gaussian (NIG) distribution with the sample distribution. Both the aforementioned distributions possess key features of our data such as skewness, peakedness and heavy tails, which the normal distribution cannot capture. However, a fundamental difference between the BNS model and the Heston model is the structure of the respective stationary autocorrelation functions. For the BNS model this function is zero, while it may not be zero for the Heston model. Since the sample stationary autocorrelation function is close to zero and due to a slightly better distributional fit with the NIG distribution compared with the VG distribution, we conclude that the BNS model is a better model of EU ETS futures prices than the Heston model.

The EU ETS market was introduced in January 2005, and had its pilot phase (phase 1) between 2005-2007. The second phase, the Kyoto commitment period, spans the years 2008-2012. Thereafter the post Kyoto period is planned until 2020. Most of the existing empirical studies on EU ETS prices only involves data from phase 1. Recent work investigating the CO_2 allowance price dynamics includes [5], [21],[19] and [10]. In [5] the authors propose Markov-switching and AR-GARCH models for stochastic modeling of the logprice returns. They find that these models capture key observed features of the data as skewness, excess kurtosis and non-constant volatility. In [19] the performance of different GARCH models is considered. These papers (and some of those that they cite) also give a description of the market mechanism and empirical investigation based on the spot prices in the first phase. The three main markets for EUAs within EU ETS, Powernext, Nord Pool and the European Climate Exchange (ECX), are studied in [10]. Banking of permits was prohibited during phase 1 and phase 2, which made emission allowances worthless at the end of phase 1, and thus not tradable during the whole life of the underlying contract. This issue is addressed in [10]. The authors of this paper also consider the cases of futures traded within a single phase (intra-period futures) and futures traded across phases (inter-period futures). They investigate an augmented

cost-of-carry relationship between futures and spot prices (see references therein for other research on EU ETS futures prices). Furthermore, the authors examine the ability of various diffusion and jump diffusion continuous-time processes in capturing the dynamics of EUA spot prices. In particular, they find that a geometric Brownian motion augmented with jumps significantly improves on the performance of its diffusion counterpart. The authors in the recent paper [14] use a nonlinear dynamics approach to analyze EUA price volatility, based on data from April 2005 until February 2009.

Compared to the exisiting literature we perform an extensive and up to date empirical analysis of Nord Pool futures prices, investigating the descriptive statistics of their logreturns. Moreover, we consider two continuous stochastic volatility models, specifically the BNS and Heston models for the futures price logreturns. This approach differs from [10], where a jump diffusion model is used to account for the overall volatility. Furthermore, we introduce an additional parameter in the Heston model to capture skewness. As a consequence we find a non-zero stationary autocorrelation function for the logreturns, having a short-range dependence in this modified model. This result is, to the best of our knowledge, new to the literature.

This chapter is organized as follows: In Sect. 2 we present the findings of our statistical analysis of our data set. In Sect. 3 we present the BNS and Heston models and their stationary autocorrelation functions. In Sect. 4 we propose and investigate the fit of the VG and NIG distributions, and estimating their parameters via MLE. We then compare the empirical performance of the BNS and Heston models. In Sect. 5 we conduct a statistical analysis for each individual phase of the EU ETS market. We conclude in Sect. 6. We give detailed proofs of Proposition 2 and 3 in Appendix 1 and 2, respectively. Appendix 3 includes the proofs of some Lemmas in Appendix 2.

2 Statistical data analysis of EU ETS futures prices

We analyze five series of futures prices for EUAs observed at the Nordic power exchange Nord Pool. The series represent successively the front contract, the second front and so on. The data is collected on a daily basis (5 days a week) from 22.04.2005-27.08.2013, giving 2148 observed futures prices. Due to a similarity in the data, our analysis and graphics will rely on the data from series 1. However, statistical characteristics for all series are collected in the tables below. In this section we present statistics, giving the foundation and motivation to investigate the BNS model and the Heston model for the futures logprice in upcoming sections. Thereafter, a deeper analysis regarding parameter estimations for the respective models will be made. We start to look at the time series for the futures price $X(t)$ in series 1, shown in Fig. 1.

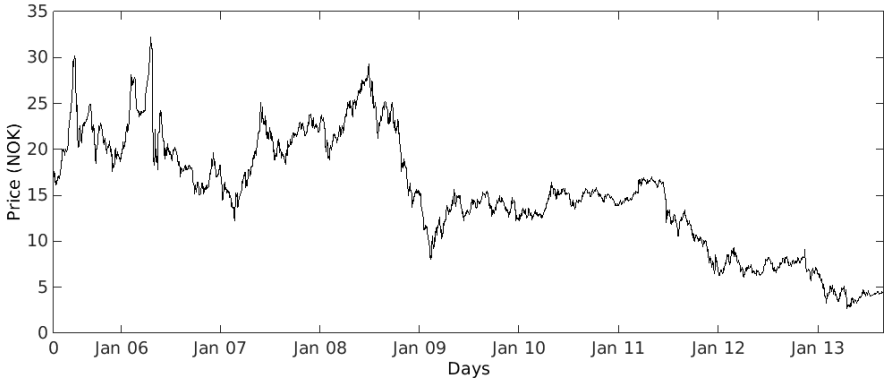


Fig. 1: Time series for the futures price $X(t)$.

We observe a clear downward trend in the futures price. The futures prices do not show any seasonal pattern, which is in line with the main finding in [21] for CO_2 spot prices. The first four sample moments of the futures price data series are collected in Table 1. The data are slightly skewed, and the significant

Table 1: Sample Moments for futures prices

	Mean	Variance	Skew ($\times 10^{-2}$)	Kurtosis
Serie 1	15.30	39.30	−2.28	2.31
Serie 2	15.55	41.83	−0.65	2.29
Serie 3	15.79	44.39	1.57	2.27
Serie 4	16.07	47.08	3.77	2.27
Serie 5	16.36	50.29	8.16	2.29

excessive kurtosis gives sign of a non-Gaussian distribution.

To proceed we investigate the logreturns $L(t)$, defined as follows: define the normalized[1] futures logprice S as

$$S(t) := \log \frac{X(t)}{X(0)}. \tag{1}$$

The logreturn at time t_i is defined as

$$L(t_i) = S(t_{i+1}) - S(t_i). \tag{2}$$

The partition will be such that $t_{i+1} - t_i = 1$, corresponding to one day in our data, i.e. we consider daily logreturns. We define daily logreturns in continuous time as

$$L(t) = S(t+1) - S(t). \tag{3}$$

A time series of the logreturns is showed in Fig. 2.

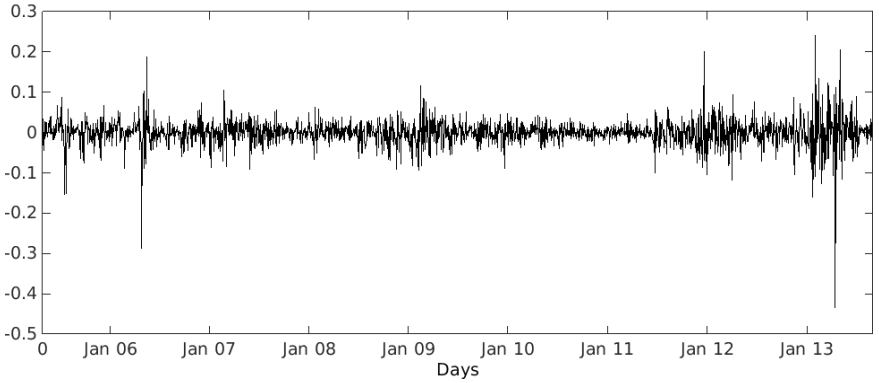
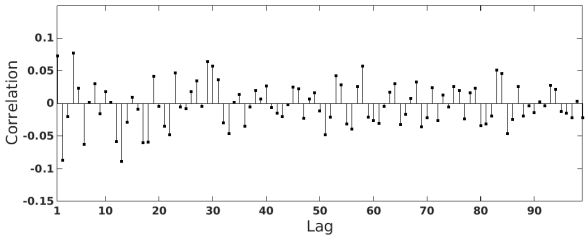


Fig. 2: Time series for the logreturns $L(t)$.

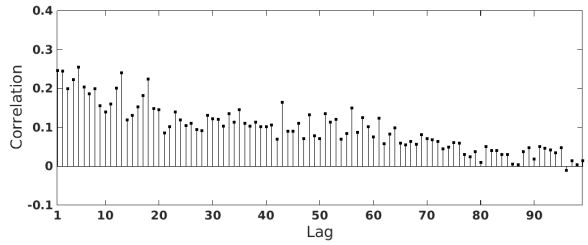
The observed volatility clustering in Fig. 2, which seems to have a non-deterministic appearance, indicates that a large (small) absolute change tend to be followed by a large (small) absolute change. In particular, we can observe several absolute changes greater than 10% [2]. This behavior is also consistent with the autocorrelation of the absolute logreturns in Fig. $3(b)$. The fact that the absolute logreturns are slowly decaying to zero, is considered as a typical indicator of volatility clustering, see [8]. The slow decay to zero can also be seen in the autocorrelation function for squared logreturns (see Fig. 6), but it is more apparent for the absolute logreturns. In Fig. $3(a)$ we plot the autocorrelation function of the logreturns, which, not unexpectedly, shows a variation around zero.

[1] Normalized in the sense that $S(0) = 0$.

[2] Most of the extreme changes in the market prices are due to weather conditions, e.g. dry summers, cold winters etc. and corresponding high or low prices of competing energy commodities such as oil and gas. However, in 2005 there was a surplus of granted EUAs in the market, causing a crash in April 2006, see [5].



(a) Autocorrelation function for logreturns.



(b) Autocorrelation function for absolute logreturns.

Fig. 3: Autocorrelation function for the logreturns (a) and the absolute logreturns (b).

In Fig. $3(a)$ we see that the correlation tends rapidly towards zero already for lag 1. From the sample moments, presented in Table 2 we see that the logreturns are skewed to the left. Furthermore, the high excessive kurtosis indicates a distribution of the returns that has a strong deviation away from normality. Hence, the increments of $S(t)$, i.e. the logreturns, are clearly not normally distributed. Thus, the absence of autocorrelation does not necessarily indicate independence, we may still have nonlinear dependence. However, this suggests a model having small or no-memory effects in the logreturns, i.e. a stationary autocorrelation function close to zero for all lags. This excludes any long range dependence, even though we may have a presence of short range dependence. That is, an autocorrelation function decreasing at a geometric rate, see [8].

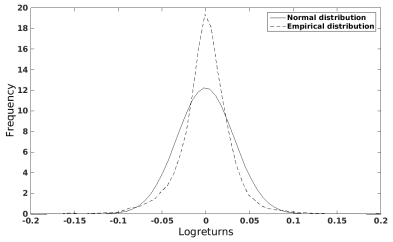
Table 2: Sample moments for the logreturns

	Mean ($\times 10^{-4}$)	Variance ($\times 10^{-4}$)	Skew	Kurtosis
Series 1	-6.2	11	-1.19	24.82
Series 2	-6.2	11	-1.16	24.59
Series 3	-6.2	10	-1.15	24.66
Series 4	-6.2	10	-1.14	24.48
Series 5	-6.2	10	-1.17	24.37

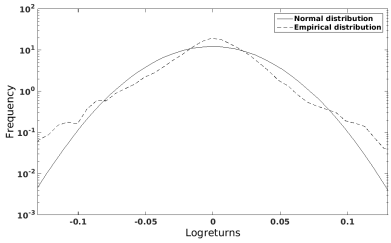
The non-normal behavior is very distinctive in the density plots in Fig. 4, visualizing the deficiency of the normal distribution to describe the empirical distribution. In Sect. 5 we shall also investigate the three phases separately to assess whether there have been significant differences among them, as well as a comparison with the full times series spanning over 2005-2013.

3 BNS and Heston model

In this section we investigate distributional and stationary properties of the BNS model and the Heston model. First we set up a stochastic modeling framework. Denote the square of the stochastic volatility process by Y, and define the filtration $\mathscr{F}_t^Y := \{Y_u, u \leq t\}$, i.e. the σ-algebra generated by Y, revealing all



(a) Distribution for the logreturns.



(b) Distribution for the logreturns. Logarithmic scale.

Fig. 4: The estimated distribution for logreturn, fitted with normal distribution $\mathcal{N}(\mu_s, \sigma_s)$. Here, μ_s, σ_s are the sample mean and variance, reported in Table 2.

information of the stochastic volatility up to time t. Similarly we define $\mathscr{F}_t^B$ to be the σ-algebra generated by the Brownian motion B. Let $(\Omega, \mathscr{F}, \mathbb{F}, \mathbf{P})$, where $\mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}$, be a complete filtered probability space satisfying the usual conditions. We define $\mathscr{F}_t$ to be the smallest σ-algebra containing all the information generated by the Brownian motion B and the stochastic volatility Y up to time t, i.e.

$$\mathscr{F}_t := \mathscr{F}_t^B \vee \mathscr{F}_t^Y.$$

Our data analysis in the previous section encountered skewness and clustering in the logreturn. We also saw that the empirical distribution had heavier tails than the normal distribution. As mentioned in [2] this type of data rejects the assumptions in the Black-Scholes (BS) model. The statistical features and the empirical autocorrelation indicate that we need to extend our model to incorporate stochastic volatility. We investigate two models with this property: the model proposed by Barndorff-Nielsen and Shephard known as the BNS model (see [2]), and the Heston model (see [16]). These models can explain stylized facts like heavy tails, skewness and excessive returns as observed in the time series above.

In this section we give some main properties of these models. For a more detailed analysis, see [2], [1], [3], [12] and [16].

3.1 The Barndorff-Nielsen and Shepard model

Instead of using the simple BS-model, we suggest to model the logarithmic futures price $S(t)$ with the BNS model, which incorporates stochastic volatility. The idea is to model the volatility $\sigma^2(t)$ as a superposition of independent stationary positive Ornstein-Uhlenbeck (OU) processes, driven by a Background Driving Lévy Process (BDLP), see [2]. To be more specific, we consider the logarithmic price

$$dS(t) = (\mu + \beta\sigma^2(t))dt + \sigma(t)dB(t). \tag{4}$$

We assume

$$\sigma^2(t) = Y(t), \tag{5}$$

where

$$dY(t) = -\lambda Y(t)dt + dU(\lambda t). \tag{6}$$

The process U is a subordinator, i.e. a Lévy process with positive drift and jumps, but no Brownian part. This ensures that Y is a positive process, and the volatility $\sigma(t) := \sqrt{Y(t)}$ is well defined. The unusual scaling by λ in $U, U(\lambda t)$, ensures that the stationary distribution of U does not depend on λ. Thus, the speed of mean reversion λ accounts only for the modeling of the autocorrelation function of squared returns. The process $U(t)$ is the BDLP of Y. In general, the process $\sigma^2(t)$ has the form

$$\sigma^2(t) = \sum_{i=1}^{n} w_i Y_i(t), \quad \sum_{i=1}^{n} w_i = 1 \tag{7}$$

when using an n-factor model for the stochastic volatility. The empirical analysis of the futures price data motivates the choice $n = 1$ to be sufficient for our purpose (see Sect. 4.1.1). In [2] it is shown that for a self-decomposable distribution D, there exists a stationary OU-process such that $Y \sim D$, satisfying (6) where U is a Lévy process with positive increments and the cumulant generating function, defined as the logarithm of the moment generating function, is

$$\psi(u) = -\int_{0+}^{\infty} (1 - \mathrm{e}^{-ux}) \ell(dx).$$

Here, ℓ is the Lévy measure of $U(1)$, i.e. a subordinator. The only requirement of a self-decomposable distribution D allows for a great flexibility in the choice of distributions for Y. In [2] and [1] the self-decomposable class of generalized inverse Gaussian (GIG) distributions is treated. This is a three parameter class, labeled $GIG(\nu, \delta, \gamma)$, with density

$$pdf_{GIG(\nu,\delta,\gamma)}(u) = \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} u^{\nu-1} \exp\left(-\frac{1}{2}(\delta^2 u^{-1} + \gamma^2 u)\right), \quad u > 0,$$

where K_ν is the Bessel function of third kind. For $Y \sim GIG(\nu, \delta, \gamma)$ independent of the Brownian motion B, the logreturns, $L(t) := S(t+1) - S(t)$, will be generalized hyperbolic (GH) distributed. The GH distribution is a five parameter family allowing us to control features such as skewness and kurtosis, which are desirable properties according to our previous data analysis. Define $\alpha := \sqrt{\beta^2 + \gamma^2}$, then the $GH(\nu, \alpha, \beta, \delta, \mu)$ density is given by

$$\begin{aligned} pdf_{GH}(u) := \frac{(\gamma/\delta)^\nu}{\sqrt{2\pi}\alpha^{\nu-1/2} K_\nu(\delta\gamma)} & \left(\delta^2 + (u-\mu)^2\right)^{(\nu-1/2)/2} \\ & \times K_{\nu-1/2}\left[\alpha\sqrt{\delta^2 + (u-\mu)^2}\right] \exp\left(\beta(u-\mu)\right). \end{aligned} \tag{8}$$

For our purpose we will consider two special cases: $\nu = -\frac{1}{2}$ leading to $Y \sim IG(\delta, \gamma)$ (inverse-Gaussian distribution) and $\delta = 0, \nu > 0$ leading to $Y \sim \Gamma(\nu, \gamma)$ (Gamma distribution), with densities

$$pdf_{IG(\delta,\gamma)}(u) = \frac{\delta}{\sqrt{(2\pi)}} \exp(\delta\gamma) u^{-3/2} \exp\left(-\frac{1}{2}(\delta^2 u^{-1} + \gamma^2 u)\right), \quad u > 0,$$

where $\delta > 0$ and $\gamma \geq 0$ and

$$pdf_{\Gamma(\nu,\gamma)}(u) = \frac{(\gamma^2/2)^\nu}{\Gamma(\nu)} u^{\nu-1} \exp(-u\gamma^2/2), \quad u > 0.$$

Using $Y \sim IG(\delta, \gamma)$ or $Y \sim \Gamma(\nu, \gamma)$ results in a NIG or Γ distributed logreturns respectively, which are both special cases of the GH distribution. In Sect. 4 we will estimate and analyze these distributions to our data series. From our discussions, we see that the BNS model is rather flexible in capturing distributional properties needed for empirical fitting. We remark that in [2] it is proposed to start with a model for the stationary distribution of Y, and shown how one can construct a BDLP U which gives this desired distribution. Such a BDLP exists whenever Y has a self-decomposable distribution. We end this first discussion of the BNS model with a known fact of the autocorrelation function, a result which will become useful in the parameter estimations.

Proposition 1. *The stationary autocorrelation of the logreturns and for squared logreturns, defined as in* (2) *with logprice given by* (4)*, are*

$$AutoCorr(L(t), L(t+k)) = 0, \tag{9}$$

$$AutoCorr(L^2(t), L^2(t+k)) = \mathrm{e}^{-\lambda k}, \tag{10}$$

for $k = 0, 1, 2 \ldots$

Proof. See [2].

As we saw in Fig. $3(a)$, the autocorrelation function for the logreturns is close to zero for all lags. Since the driving Brownian motion in (4) and U are independent, this is the case for the BNS model, as is seen in (9).

3.2 The Heston model

Introduce a stochastic volatility given by the process $Y(t)$, where $Y(t)$ is

$$dY(t) = -\gamma(Y(t) - \lambda)dt + \kappa\sqrt{Y(t)}dW(t). \tag{11}$$

The long-time mean of Y is here given by $\lambda > 0$, $\gamma > 0$ is the rate of relaxation to this mean, $\kappa > 0$ is the variance noise (the volatility of the volatility) and $W(t)$ is a standard Brownian motion, assumed to be independent of $B(t)$ (the independence is not a necessary requirement in general, however our investigation will rely on this assumption). In what follows we assume that the futures price $X(t)$ of the allowances is modeled by

$$dX(t) = (\mu + (\frac{1}{2} - \beta)\sigma^2(t))X(t)\,dt + \sigma(t)X(t)\,dB(t)\,, \tag{12}$$

with *stochastic volatility* $\sigma(t)$ defined by $\sigma^2(t) = Y(t)$. Here μ and β are constants. The bivariate process (X,Y) is known as the Heston model (see [16] and [12]), therefore we will also refer to the process Y in (11) as Heston type volatility model. Also, the introduction of the constant β in the dynamics (12) is necessary for obtaining the distributional characteristics, see Remark 1.

By Itô's Formula we get

$$dS(t) = (\mu - \beta\sigma^2(t))dt + \sigma(t)dB(t), \tag{13}$$

for the logprice defined in (1). The daily logreturns $L(t)$ is then, expressed in terms of Y, given by

$$L(t) := S(t+1) - S(t) = \int_t^{t+1} (\mu - \beta Y(s))ds + \int_t^{t+1} \sqrt{Y(s)}dB(s), \tag{14}$$

or approximately as

$$L(t) \approx (\mu - \beta Y(t)) + \sqrt{Y(t)}\Delta B(t).$$

Let us investigate the characteristics of $L(t)$ a bit closer. By conditioning on $Y(t) = y$, $L(t)$ is normally distributed with mean $\mu - \beta y$ and variance y. It is shown in [12] that the probability density function (pdf) for Y in stationarity is given by

$$\Pi(y) = \frac{\alpha}{\Gamma(\alpha)} \frac{y^{(\alpha-1)}}{\lambda^\alpha} \exp(\frac{-\alpha y}{\lambda}), \quad \alpha = \frac{2\gamma\lambda}{\kappa^2}.$$

The characteristic function $\phi_L(u)$ for L can now be calculated as

$$\begin{aligned}\phi_L(u) =& \mathbb{E}\left[\mathrm{e}^{iuL(t)}\right] \\ =& \mathbb{E}\left[\mathbb{E}\left[\mathrm{e}^{iuL(t)}|Y(t)\right]\right] \\ =& \mathbb{E}\left[\mathrm{e}^{iu(\mu-\beta Y(t))-\frac{1}{2}u^2Y(t)}\right] \\ =& \mathbb{E}\left[\mathrm{e}^{iu\mu+i(-\beta u+i\frac{1}{2}u^2)Y(t)}\right] \\ =& \mathrm{e}^{iu\mu}\mathbb{E}\left[\mathrm{e}^{i\xi Y(t)}\right] = \mathrm{e}^{iu\mu}\phi_Y(\xi),\end{aligned}$$

where $\xi = -\beta u + i\frac{1}{2}u^2$. The characteristic function $\phi_Y(\xi)$ for Y is obtained via the density $\Pi(y)$. The following calculation also shows that there is no problem of extending to complex-valued parameters ξ in the Lévy-Khintchine formula:

$$\begin{aligned}\phi_Y(\xi) =& \int_0^\infty \mathrm{e}^{i\xi y}\frac{\alpha}{\Gamma(\alpha)}\frac{y^{(\alpha-1)}}{\lambda^\alpha}\exp(\frac{-\alpha y}{\lambda})dy \\ =& \frac{\alpha^\alpha}{\Gamma(\alpha)\lambda^\alpha}\int_0^\infty y^{\alpha-1}\mathrm{e}^{-(\frac{\alpha}{\lambda}-i\xi)y}dy \\ =& (1-i\frac{\lambda}{\alpha}\xi)^{-\alpha},\end{aligned}$$

where we in the last step used the integral identity

$$\int_0^\infty x^m \mathrm{e}^{-qx^n}dx = \frac{\Gamma(p)}{nq^p}, \quad p = \frac{m+1}{n}$$

for $Re(n) > 0$, $Re(m) > -1$ and $Re(q) > 0$. By identification we have $Re(n) = 1$, $Re(m) = \alpha - 1$ and $Re(q) = \frac{\alpha}{\lambda} + \frac{1}{2}u^2$. Hence, the conditions for the integral identity are clearly satisfied. The characteristic function $\phi_L(u)$ then becomes, after inserting the expressions for ξ and α,

$$\phi_L(u) = \mathrm{e}^{iu\mu}\left[1 + i\frac{\beta\kappa^2}{2\gamma}u + \frac{\kappa^2}{4\gamma}u^2\right]^{-\frac{2\gamma\lambda}{\kappa^2}} \tag{15}$$

The conditional distribution of the CIR-model is non-centered chi-squared. Since $\gamma, \lambda > 0$ the CIR-model approaches a gamma distribution in stationarity. In this way the Heston model will have VG distributed logreturns. In the next section we shall obtain the parameters in the Heston model via MLE. However, rather than making a MLE for the parameters $(\mu, \beta, \gamma, \lambda, \kappa)$ in the Heston model directly, we make the parameter estimation for a process possessing the same key features observed in the statistical data analysis. Furthermore, since we know the characteristic function for the logreturns of the Heston model in stationary, we consider a process having a VG distribution, with known characteristic function. A natural process to choose with these properties is the variance-gamma process.

The *variance-gamma process* was introduced in [18] as a model for stock market logreturns, then as a two-parameter process. To account for skewness the model was extended in [17] to a three parameter model, and in [22] it was extended to a four parameter model for the logreturns of an underlying. The process is obtained by evaluating a Brownian motion with drift at a Γ-distributed random time, leading to a variance mixture model via subordination. By conditioning the underlying process on the random time we get a normally distributed process with two parameters that account for drift and (constant) volatility. The pdf of the VG-distribution is given by

$$pdf_{VG}(u) := \frac{2\exp\left(\frac{\theta(u-c)}{\sigma_{vg}^2}\right)}{\sigma_{vg}\sqrt{2\pi}\nu_{vg}^{\frac{1}{\nu_{vg}}}\Gamma(\frac{1}{\nu_{vg}})}\left(\frac{|u-c|}{\sqrt{\frac{2\sigma_{vg}^2}{\nu_{vg}}+\theta^2}}\right)^{\frac{1}{\nu_{vg}}-\frac{1}{2}} K_{\frac{1}{\nu_{vg}}-\frac{1}{2}}\left(\frac{|u-c|\sqrt{\frac{2\sigma_{vg}^2}{\nu_{vg}}+\theta^2}}{\sigma_{vg}^2}\right) \tag{16}$$

where K_η is the modified Bessel function of third kind with index η (see [22]). Here, c describes the long-time mean of the process, σ_{vg} is the volatility of the Brownian motion, ν_{vg} the variance rate of the gamma time change and θ the drift in the Brownian motion with drift, responsible for introducing skewness. In the next section we estimate the VG distribution to the logreturns of the futures price series we have at hand, and relate the VG parameters to the Heston model. To do that we need the following results for the autocorrelation of the logreturns and squared logreturns. This is needed in order to make the characteristic function (15) consistent with the pdf (16).

Proposition 2. *The stationary autocorrelation function of lag $k > 0$ for the logreturns, defined as in* (14)*, is of the form*

$$AutoCorr(L(t+k), L(t)) = Ce^{-\gamma k}. \tag{17}$$

The constant C is given by

$$\frac{\beta^2\kappa^2}{(\beta^2\kappa^2+2\gamma)\gamma^2}(e^{\gamma}+e^{-\gamma}-2). \tag{18}$$

Proof. See Appendix 1. Note that from the proof, the calculation of the non-stationary autocorrelation function is straightforward as well.

We see that the autocorrelation does not depend on the mean of the stochastic volatility, however it depends on the rate of relaxation γ to this mean and the volatility of the stochastic volatility κ. Furthermore, since $\gamma > 0$ the autocorrelation decays at a geometric rate, indicating short-range dependence (see [8]).

Proposition 3. *The stationary autocorrelation function of lag $k > 0$ for squared logreturns, defined as in* (14)*, is of the form*

$$AutoCorr(L^2(t+k), L^2(t)) = A_0 + A_1 e^{-\gamma k} + A_2 e^{-2\gamma k}. \tag{19}$$

Here the constants A_i depend on the parameters of the distribution of L.

Proof. See Appendix 2.

The characteristic function ϕ_L and the autocorrelation in Proposition 3 give us a bridge to connect the parameters in the Heston model with the four parameters $(c, \sigma_{vg}, \theta, \nu_{vg})$ in the VG process, which are the ones obtained via the estimation in R.

4 Estimation of the stochastic volatility models for EU ETS futures prices

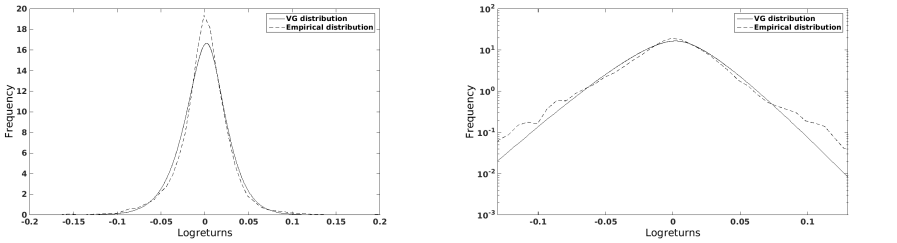
The deficiency of the normal distribution was already established in the Sect. 2, as it is unable to capture established key features in the EU ETS logprice. In the previous section we saw that a class of distributions possessing these properties is the GH-distributions. We consider two special cases, the VG distribution and the NIG distribution. To proceed, we analyze the BNS model and Heston model. The aim is to establish the best distributional fit of the Heston model and the BNS model as well as determine the parameters in the respective models. As stated, the Heston model gives us VG distributed returns. For the BNS model, the possible distributional properties are more flexible. We focus on two cases for the BNS model. First, when Y is IG distributed, resulting in NIG distributed returns. As a *"fair"* comparison with the Heston model we also derive the model parameters in the BNS model when Y is Γ-distributed, resulting in VG distributed returns. Both the Heston and the BNS model require the estimation of five parameters. In order to specify all parameters in the two models we also need, in addition to the sample distributional fit[3], to consider the autocorrelation of squared returns.

[3] The estimation of parameters in the distributions below has been made via maximum-likelihood estimation in R, using appropiate packages ("VarianceGamma" and "fBasics"). All figures and numerical calculations are generated in Matlab, and for symbolic calculation Maple has been used.

Before we proceed, we remark that the density functions that are used for the distributional fit in the maximum-likelihood estimation are given by (16) for the VG distribution and by (27) for the NIG distribution. We need to relate the parameters in the Heston model and the VG distributed BNS model to the parameters in (16). We relate the parameters in the NIG distributed BNS model to (27).

4.1 Variance-gamma distribution

Let us start to investigate the distributional fit of the VG distribution with the sample distribution. Density plots are shown in Fig. 5 [4]. We see that the VG distribution with pdf given in (16) capture the key features well, however there are minor shortcomings in both the tail and center. The estimated parameters in the VG distribution are displayed in Table 3.



(a) Distribution for the logreturns.

(b) Distribution for the logreturns. Logarithmic scale.

Fig. 5: The estimated distribution for logreturns fitted with the VG distribution.

Table 3: VG parameters for logreturns

	$\hat{c}$	$\hat{\sigma}_{vg}$	$\hat{\theta}$	$\hat{\nu}_{vg}$
Series 1	0.00379	0.0282	−0.00441	0.387
Series 2	0.0144	0.0786	−0.0180	0.118
Series 3	0.00732	0.0307	−0.00758	0.118
Series 4	0.0111	0.0296	−0.0116	0.118
Series 5	0.0690	0.0723	−0.0727	0.118

We obtain the VG distribution of the logreturns in stationarity by either using the Heston model or the BNS model.

4.1.1 Variance-gamma via the BNS model

To obtain the VG distribution via the BNS model we take Y to be $\Gamma(\nu,\gamma)$-distributed with $\gamma=\sqrt{\alpha^2-\beta^2}$. The logreturns will then be $GH(\nu,\alpha,\beta,0,\mu)$ distributed, i.e. have the VG-distribution. The corresponding pdf is obtained as the limit $\delta\to 0$ from (8),

$$pdf_{BNS_{VG}}(u)=\frac{\gamma^{2\nu}|u-\mu|^{\nu-1/2}K_{\nu-1/2}(\alpha|u-\mu|)}{\sqrt{\pi}\Gamma(\nu)(2\alpha)^{\nu-1/2}}\exp(\beta(u-\mu)). \tag{20}$$

[4] Figures with logarithmic scale are truncated due to a few number of data outliers, making the kernel density estimator in Matlab inaccurate.

The construction of such a process Y can be found in [2] using series representation. The parameters we need to specify in the BNS model is now $(\mu,\beta,\lambda,\nu,\gamma:=\sqrt{\alpha^2-\beta^2})$. To estimate the parameters, let us first consider the autocorrelation function for squared returns. The fitted exponential function (solid curve)

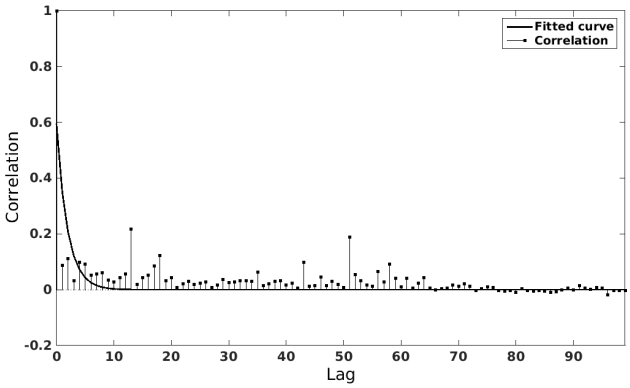


Fig. 6: Autocorrelation function for squared logreturns. The solid curve is fitted with the function $\exp(-\hat{\lambda}k)$, where $\hat{\lambda}=0.526$.

is shown in Fig. 6 where the estimated mean-reversion coefficient is $\hat{\lambda}=0.5260$. Except for some peaks around 15 and 50 lags we see that there is a nice decay tending to zero. The fit is far from perfect, but captures at least the main feature of a fast decaying autocorrelation function. Instead of using a multi-factor stochastic volatility model, that would require a much more sophisticated estimation procedure, we are confident with the first-order approximative model using a one-factor BNS model. The mean-reverting coefficient for the other series were the same up to three decimals of accuracy. By (10) in Proposition 1, we obtain the λ parameter as the decay coefficient in the autocorrelation for the squared returns. The other BNS parameters in (20) can be related to the estimated VG parameters in Table 3 via the pdf (16). By identification, to get equality between (16) and (20), we find the BNS parameters to be

$$(\mu,\beta,\lambda,\nu,\gamma)=\left(c,\frac{\theta}{\sigma_{vg}^2},\lambda,\frac{1}{\nu_{vg}},\sqrt{\frac{2}{\sigma_{vg}^2\nu_{vg}}}\right), \tag{21}$$

where $\gamma:=\sqrt{\alpha^2-\beta^2}$ and $\alpha=\frac{1}{\sigma_{vg}^2}\sqrt{\frac{2\sigma_{vg}^2}{\nu_{vg}}+\theta^2}$. This completely specifies the BNS model $(4)-(6)$ with $Y\sim\Gamma(\nu,\gamma)$. Numerically, using the estimated parameters $(\hat{c},\hat{\sigma}_{vg},\hat{\theta},\hat{\nu}_{vg})$ in Table 3 and $\lambda=\hat{\lambda}$ in (21) we find the VG distributed BNS model parameters to be

$$(\hat{\mu},\hat{\beta},\hat{\lambda},\hat{\nu},\hat{\gamma})=(0.00379,-5.56,0.526,2.59,80.8). \tag{22}$$

4.1.2 Variance-gamma via the Heston model

To obtain the parameters in the Heston model we make use of the characteristic function (15). It is shown in [17] (see also [22]) that the characteristic function $\phi_{L_{vg}}(u)$ for the logreturns is given by

$$\phi_{L_{vg}}(u)=\mathrm{e}^{iuc}\left[1-i\theta\nu_{vg}u+\frac{\sigma_{vg}^2\nu_{vg}}{2}u^2\right]^{-\frac{1}{\nu_{vg}}} \tag{23}$$

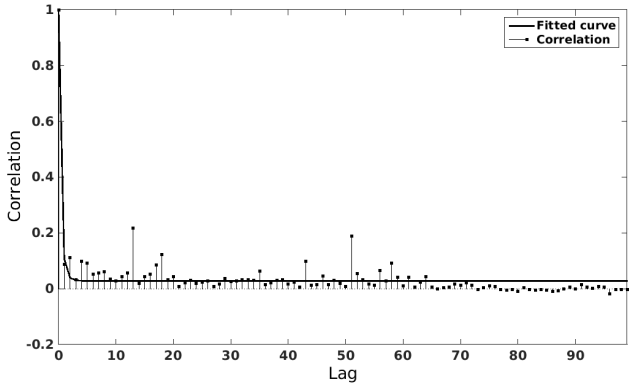


Fig. 7: Autocorrelation function for squared logreturn, fitted with the curve in Proposition 3 and using $[\hat{\gamma},\hat{A}_0,\hat{A}_1,\hat{A}_2] = [1.62, 0.0281, 1.31, 18.4]$.

By comparing the characteristic functions (15) and (23) we can identify the parameters $(\mu,\beta,\gamma,\lambda,\kappa)$ of our model (12) (with Heston type volatility (11)) with the parameters $(c,\sigma_{vg},\theta,\nu_{vg})$ of the VG distribution. We find the system of equations

$$\mu = c, \quad -\frac{\beta\kappa^2}{2\gamma} = \theta\nu_{vg}. \tag{24}$$

$$\frac{\kappa^2}{4\gamma} = \frac{\sigma_{vg}^2\nu_{vg}}{2}, \quad \frac{2\gamma\lambda}{\kappa^2} = \frac{1}{\nu_{vg}}. \tag{25}$$

Note that we have more unknowns than equations, making the system under-determined. Let $\tilde{\alpha} := \frac{\kappa^2}{\gamma}$, the above system $(24)-(25)$ can then be solved for $\tilde{\alpha}$, β and λ with the solution

$$(\tilde{\alpha},\beta,\lambda) = (2\sigma_{vg}^2\nu_{vg}, -\frac{\theta}{\sigma_{vg}^2}, \sigma_{vg}^2).$$

To solve this under-determined system, we can obtain the parameter γ from the autocorrelation function. The red curve in Fig. 7 is fitted with the function $A_0 + A_1 e^{-\gamma k} + A_2 e^{-2\gamma k}$. The estimation yields $[\hat{\gamma},\hat{A}_0,\hat{A}_1,\hat{A}_2] = [1.62, 0.0281, 1.31, 18.4]$. We obtain the corresponding parameters for the Heston model to be

$$(\mu,\beta,\gamma,\lambda,\kappa) = (c, -\frac{\theta}{\sigma_{vg}^2}, \hat{\gamma}, \sigma_{vg}^2, \sqrt{2\sigma_{vg}^2\nu_{vg}\hat{\gamma}}). \tag{26}$$

This completely specifies the Heston model (12) with Heston type volatility (11). Numerically, again using the estimated parameters $(\hat{c},\hat{\sigma}_{vg},\hat{\theta},\hat{\nu}_{vg})$ in Table 3 and $\gamma = \hat{\gamma}$ in (26) we find the Heston model parameters to be

$$(\hat{\mu},\hat{\beta},\hat{\gamma},\hat{\lambda},\hat{\kappa}) = (0.00379, 5.56, 1.62, 0.0008, 0.0316).$$

We see that the first two parameters in (21) and (26) are the same (in absolute value for β) for the VG-distributed BNS model and the Heston model. This is clear since, conditionally on the volatility in (4) and (13), they have the same dynamics. Note that the different signs on the β parameter is just a consequence of the different choices of sign in (4) and (13) for the BNS and Heston model respectively.

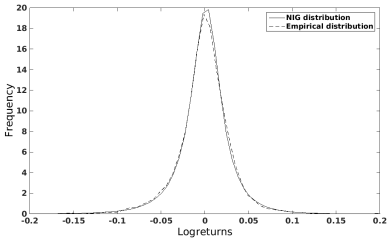
Remark 1. It is evident that $\beta = 0$ if and only if $\theta = 0$. Note that θ is the drift parameter in the VG process responsible for skewness. Hence, the introduction of the parameter β is necessary to account for skewness.

4.2 Normal inverse Gausssian distribution

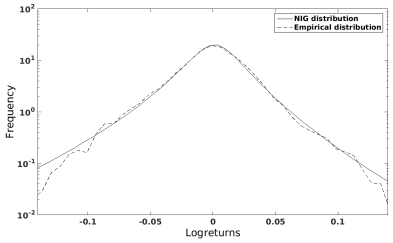
We now turn to the NIG distribution. A density plot of the NIG distribution is shown in Fig. 8. The pdf for the $NIG(\alpha,\beta,\mu,\delta)$ distribution is given by (see [1]),

$$pdf_{NIG}(u) := \frac{\delta\alpha}{\pi\sqrt{\delta^2+(u-\mu)^2}}\exp\left(\delta\sqrt{\alpha^2-\beta^2}+\beta(u-\mu)\right)K_1(\alpha\sqrt{\delta^2+(u-\mu)^2}) \tag{27}$$

where K_1 is the modified Bessel function of third order and index 1. Here $\alpha > 0$ controls the behavior of the tails. The steepness of the NIG distribution increases with increasing α. Hence, large values of α implies light tails and vice versa. The parameter β, $0 < |\beta| < \alpha$, controls the skewness of the distribution. μ is the location parameter and δ is the scale parameter. We see that stylized features such as tail and center behavior is well captured by the NIG distribution. The estimated parameters are displayed in Table 4.



(a) Distribution for the logreturns.



(b) Distribution for the logreturns. Logarithmic scale.

Fig. 8: The estimated distribution for logreturns fitted with NIG distribution, $NIG(\alpha,\beta,\mu,\delta)$. The parameters are given in Series 1 in Table 4. The heavy tails are seen to be well captured by the NIG distribution.

Table 4: NIG parameters for logreturns

	$\hat{\alpha}_{NIG}$	$\hat{\beta}_{NIG}$	$\hat{\mu}_{NIG}(\times 10^{-3})$	$\hat{\delta}_{NIG}(\times 10^{-2})$
Series 1	21.58	-2.59	1.90	2.10
Series 2	21.24	-2.44	1.75	2.05
Series 3	20.94	-2.54	1.83	2.02
Series 4	20.88	-2.56	1.85	2.01
Series 5	21.20	-2.63	1.92	2.04

By the flexibility of the BNS model, a $NIG(\alpha,\beta,\mu,\delta)$ distribution is obtained by taking Y to be $IG(\delta,\sqrt{\alpha^2-\beta^2})$ distributed, i.e. $\nu = -\frac{1}{2}$ in (8). One method of how to construct and simulate such an IG-process Y is described in [24].

Note that the NIG distribution (27) is defined as the distribution of the BNS model when Y is IG distributed with parameters $(\delta,\sqrt{\alpha^2-\beta^2})$ and the parameters in the NIG distribution exactly correspond to the parameters in that BNS model. Hence, the parameters in the BNS model are obtained by fitting the NIG density (27) to the sample density. As before, λ is estimated via the autocorrelation function (10) for squared logreturns. This completely specifies all the parameters in the BNS model (4) – (6) with $Y \sim IG(\delta,\gamma)$. We directly obtain the BNS parameters to be

$$(\mu,\beta,\lambda,\delta,\gamma) = (\hat{\mu}_{NIG},\hat{\beta}_{NIG},\hat{\lambda},\hat{\delta}_{NIG},\sqrt{\hat{\alpha}^2_{NIG}-\hat{\beta}^2_{NIG}}). \tag{28}$$

Numerically, the NIG-distributed BNS model has parameters

$$(\hat{\mu}, \hat{\beta}, \hat{\lambda}, \hat{\delta}, \hat{\gamma}) = (0.0019, -2.59, 0.526, 0.0210, 21.4).$$

Having investigated the distributional fit of the VG and NIG distribution, it is evident from Fig. 9 that the NIG distribution has a better fit with the sample distribution regarding the tails and center. Furthermore, we have tested the equality of the distribution for the logreturns with the VG and NIG distributions respectively, using the Kolmogorov-Smirnov test (KS-test). The p-values of the KS-test, presented in Table 5, indicates that the null hypothesis[5] for the VG distribution[6] can be rejected, whereas the KS-test fails to reject the null hypothesis for the NIG distribution.

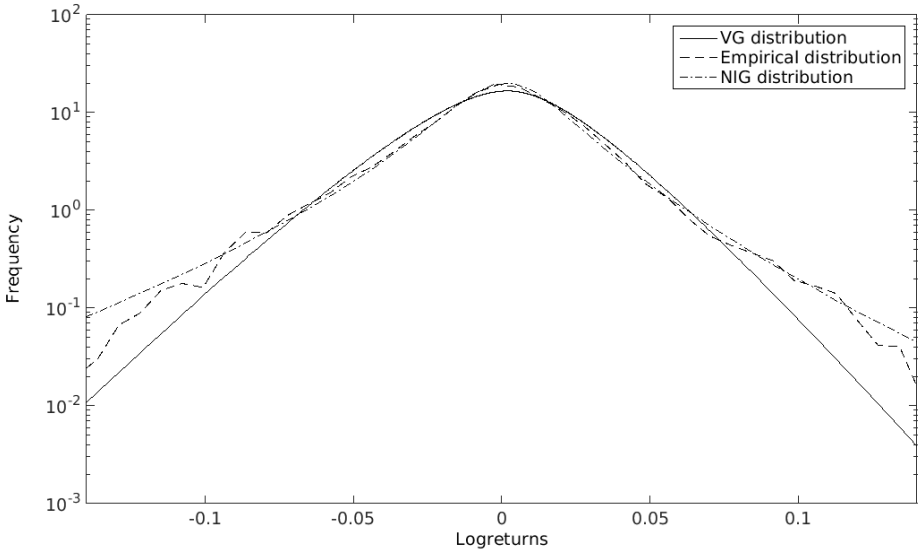


Fig. 9: Logarithmic sample distribution for logreturns. A comparison between the VG and the NIG distribution.

Table 5: Kolmogorov-Smirnov test: p-values.

	VG	NIG
Series 1	0.00155	0.318
Series 2	< 0.001	0.402
Series 3	*NA*	0.337
Series 4	< 0.001	0.231
Series 5	< 0.001	0.267

5 Empirical investigation for the individual phases

The three phases have been subjected to different regulations. Due to that we are motivated to analyze the prices for each phase separately. In this section we assess whether there have been significant differences

[5] The KS-test was conducted in R. Here, the null hypothesis states that the logreturns sample distribution comes from a VG distribution or a NIG distribution respectively. A low p-value indicates presumption against the null hypothesis.

[6] R was unable to calculate the p-value for the VG distribution in Series 3.

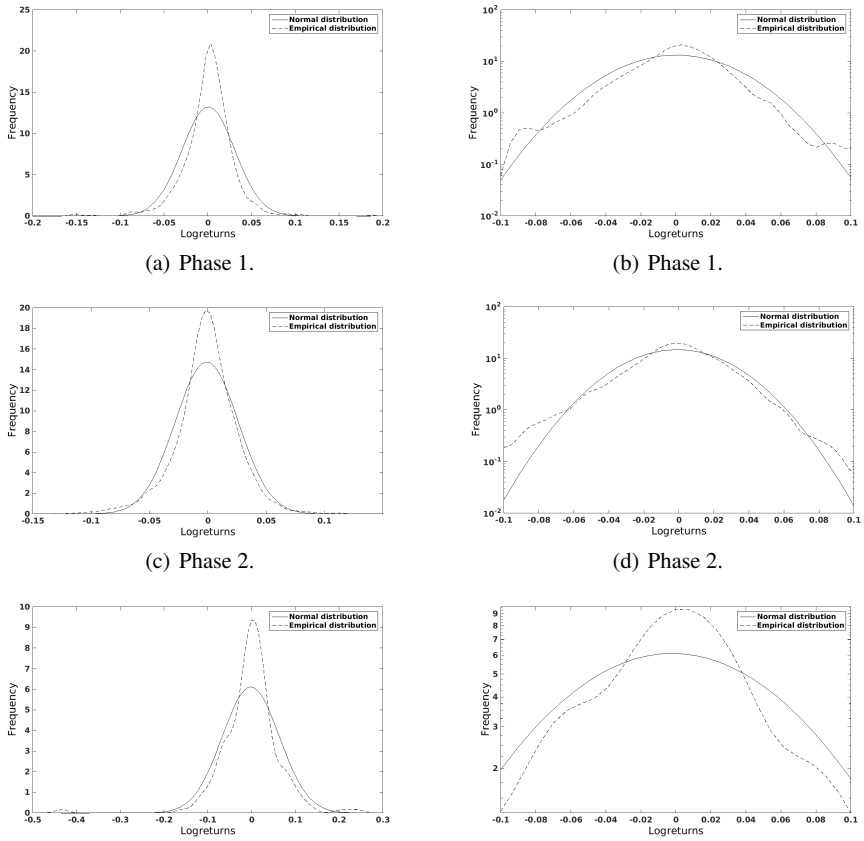
between the phases as well as between each phase and the total time span. In our analysis, we consider the time series from futures 1 only.

Phase 1 took place between 2005-2007, phase 2 between 2008-2012, and phase 3 from 2013 on. This corresponds to 690, 1289 and 169 futures logprices respectively in each phase. The sample moments for the logreturns are collected in Table 6. For each phase we observe skewness (most apparent in phase 1 and 3)

Table 6: Partial sample moments for the logreturns: Series 1.

	Mean $(\times 10^{-4})$	Variance $(\times 10^{-4})$	Skew	Kurtosis
Full series	-6.2	11	-1.19	24.82
Phase 1	4.0	9.1	-1.47	19.43
Phase 2	-10	7.3	0.0733	6.97
Phase 3	-20	43	-1.30	14.62

and high excessive kurtosis. This is in line with the time series in Fig. 2, and the deviation from normality is apparent in Fig. 10.



(a) Phase 1. (b) Phase 1.

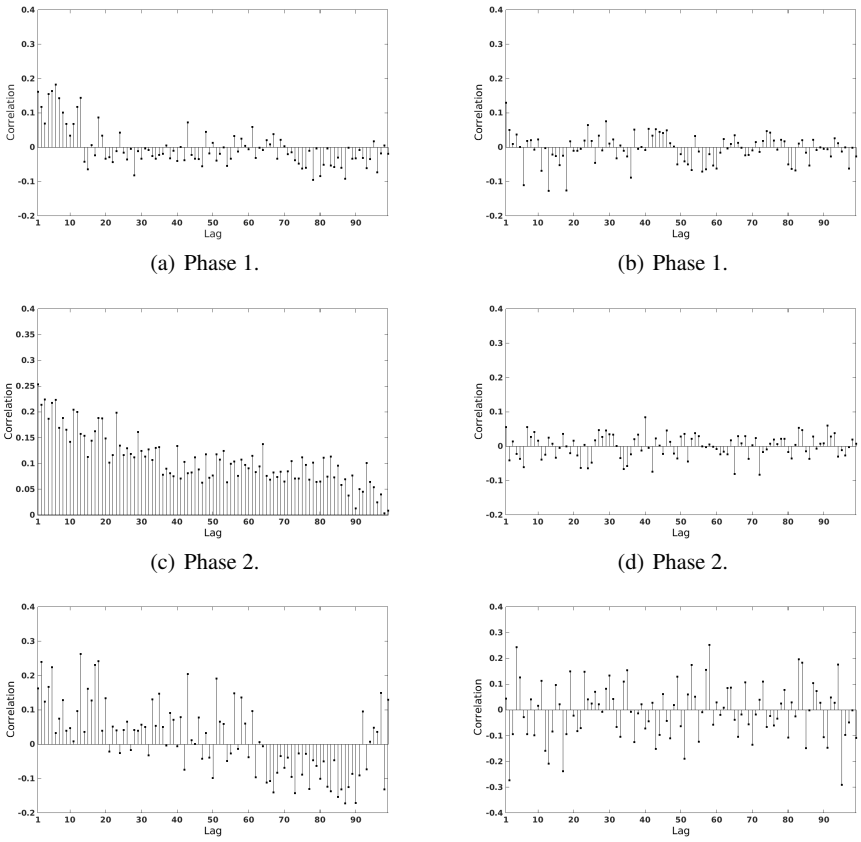
(c) Phase 2. (d) Phase 2.

(e) Phase 3. (f) Phase 3.

Fig. 10: The estimated distribution for logreturns fitted with the normal distribution. Logarithmic scale to the right.

The absolute autocorrelation function in Fig. 11 also indicate volatility clustering, as observed in the times series. For phase 1 and 2 the autocorrelation for logreturns are close to zero, however, for phase 3 we observe a modest autocorrelation (see Fig. $11(f)$). We also see that the excessive kurtosis is significantly

higher in phase 1 and 3, however all phases have lower excessive kurtosis than the full series. It is also notable that the variance of phase 3 is greater than the other phases, as well as the full time series. As we shall see, these observations are consistent with the parameter estimations below.



(a) Phase 1. (b) Phase 1.

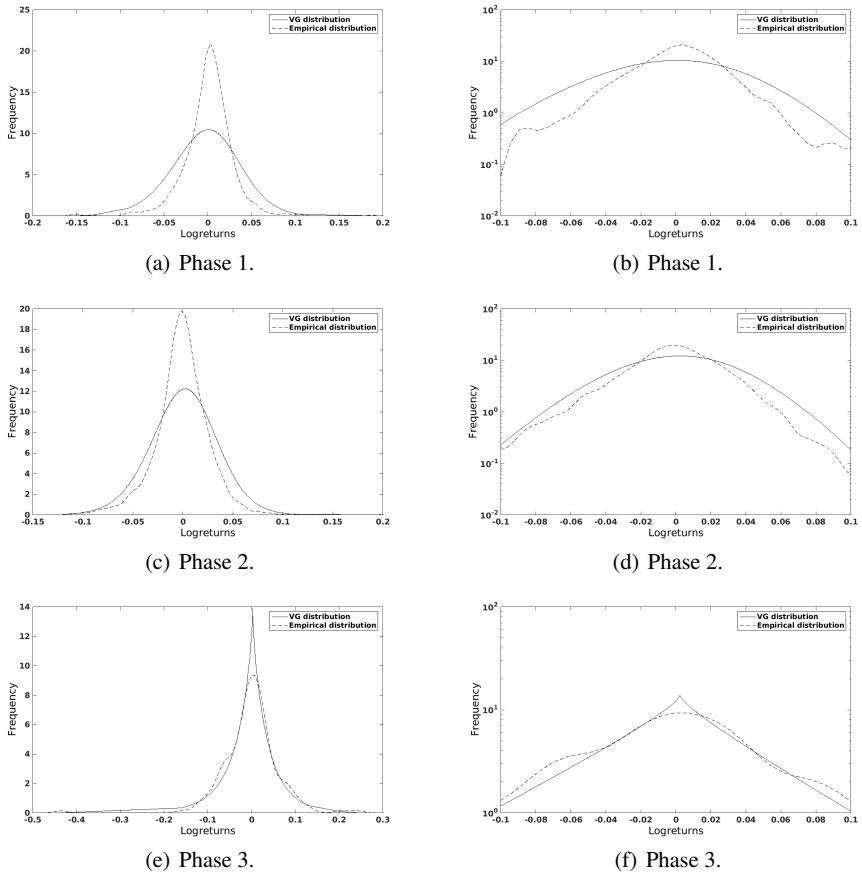
(c) Phase 2. (d) Phase 2.

(e) Phase 3. (f) Phase 3.

Fig. 11: Autocorrelation function for absolute logreturns (left) and logreturns (right).

5.1 Variance-gamma distribution

We now investigate the distributional fit of the VG distribution with the sample distribution of the three phases. Density plots are shown in Fig. 12. It is clear that the VG-distribution fails to make a good fit for the induvidual phases. In particular, it is deficient in the center for all phases. Also, in Fig. 12, it is seen that the the tails are too heavy in phase 1 and 2, but the tail behavior in phase 3 is better. This is in line with the lower excessive kurtosis for the individual phases compared to the full time series. Based on these observed distributional properties, the Heston model seems more appropriate to model the full time series rather than as a model for the individual phases. The estimated VG parameters are displayed in Table 7.



(a) Phase 1. (b) Phase 1.

(c) Phase 2. (d) Phase 2.

(e) Phase 3. (f) Phase 3.

Fig. 12: The estimated distribution for logreturns fitted with the VG distribution. Logarithmic scale to the right.

Table 7: VG parameters for series 1 for logreturns. Phase 1-3.

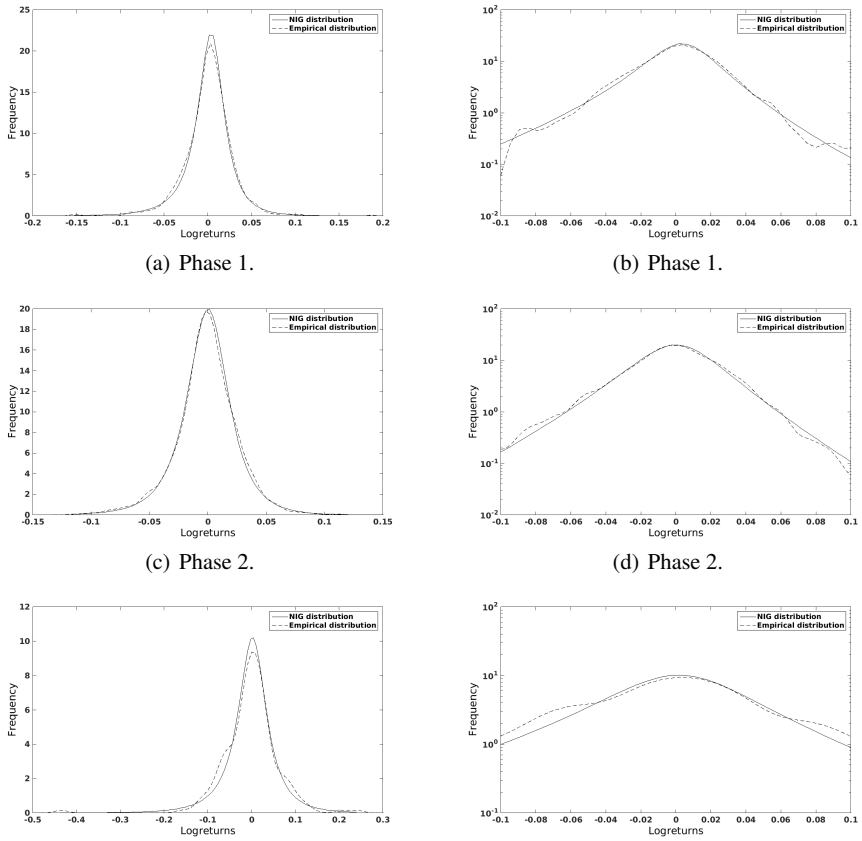
	$\hat{c}$	$\hat{\sigma}_{vg}$	$\hat{\theta}$	$\hat{\nu}_{vg}$
Full series	0.00379	0.0282	−0.00441	0.387
Phase 1	0.0183	0.0392	−0.0212	0.118
Phase 2	0.00983	0.0340	−0.00910	0.118
Phase 3	0.00232	0.0615	−0.00421	1.19

5.2 NIG distribution

We now turn to the NIG distribution. The estimated parameters in the NIG distribution are displayed in Table 8, and density plots are shown in Fig. 13.

Table 8: NIG parameters for series 1 for logreturns. Phase 1-3.

	$\hat{\alpha}_{NIG}$	$\hat{\beta}_{NIG}$	$\hat{\mu}_{NIG}$ $(\times 10^{-3})$	$\hat{\delta}_{NIG}$ $(\times 10^{-2})$
Full series	21.58	−2.59	1.90	2.10
Phase 1	23.75	−4.82	4.34	1.89
Phase 2	33.85	−2.56	0.83	2.46
Phase 3	10.14	−1.14	2.73	4.06



(a) Phase 1. (b) Phase 1.

(c) Phase 2. (d) Phase 2.

(e) Phase 3. (f) Phase 3.

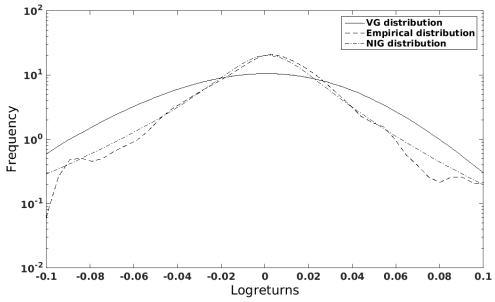
Fig. 13: Estimated distribution for logreturns fitted with the NIG distribution, $NIG(\alpha,\beta,\mu,\delta)$. The parameters are given in Table 8. Logarithmic scale to the right.

It is clear that the distributional fit with the NIG distribution is better than the VG distribution. In particular, the peak behavior is better captured for all phases. A comparison between the NIG and VG distribution is shown in Fig. 14, from which the NIG distribution seems most eligible for each phase as well as for the full time series (recall Fig. 9). Furthermore, we note that the main distributional difference between the phases is the heaviness of the tails. This is seen from the value of the α parameter in Table 8. The significantly higher (lower) value of α in phase 2 (phase 3) is consistent with the observed time series and the statistics in Table 6. Based on these distributional properties, the BNS model seems appropriate as a model for each individual phase as well as for the full time series.

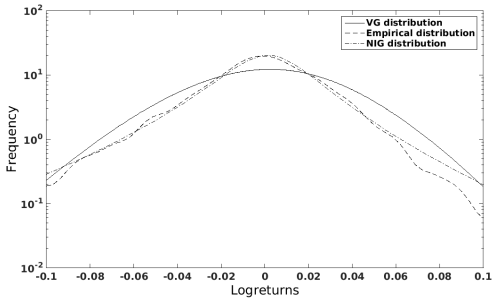
By using the autocorrelation functions the parameters for the BNS model (as well as the Heston model) can be estimated as in Sect. 4.

6 Discussion and Conclusion

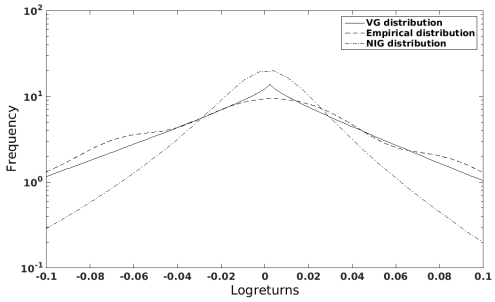
The BNS model and the Heston model can both explain the key empirical features of EU ETS futures prices observed at Nord Pool, such as skewness and heavy tails in the distribution. We noted that the introduction of β in the Heston model is necessary in order to obtain a model that could incorporate skewness. As a consequence the Heston model has a non-zero autocorrelation function. The zero autocorrelation function, obtained in the BNS model is more consistent with the empirical autocorrelation. Furthermore, the NIG distribution obtained from the BNS model by taking the stochastic volatility Y to be IG clearly fits the data



(a) Phase 1.



(b) Phase 2.



(c) Phase 3.

Fig. 14: A comparison between the VG and the NIG estimated distributions for logreturns. Logarithmic scale.

better in comparison with the VG distribution from the Heston model. Also, the BNS model is clearly more flexible as we can obtain different distributions for the returns by choosing a suitable self-decomposable distribution for the stochastic volatility process.

Furthermore, we noted in Sect. 5 that the major differences between the phases was the autocorrelation structure in phase 3 and the significant higher excessive kurtosis in phase 1 and 3. We also noted that the distributional properties of the Heston model had better alignment with the full time series rather than with the individual phases. In addition, the NIG distributed BNS model seemed to be eligible for modeling each phase as well.

Based on the autocorrelation structure and distributional properties, we conclude that the NIG distributed BNS model seems to offer the most adequate model for EU ETS allowance futures prices.

Acknowledgements: We are grateful to Montel for providing data. Financial support from Elcarbonrisk and Mawrem funded by the Norwegian Research Council within the ENERGIX program are gratefully acknowledged. We also thank an anonymous referee and the editor Nicola Secomandi for their suggestions and critics, leading to a significant improvement of the presentation of this chapter.

Appendix 1: Proof of Proposition 2

The covariance at lag k is defined by $cov(L(t+k),L(t)) := \mathbb{E}[L(t+k)L(t)] - \mathbb{E}[L(t+k)]\mathbb{E}[L(t)]$. We start to calculate $\mathbb{E}[L(t+k)L(t)]$ for $k>0$.

$$\begin{aligned}\mathbb{E}[L(t+k)L(t)] &= \mathbb{E}\left[\left(\int_{t+k}^{t+k+1}(\mu-\beta Y(s))ds+\int_{t+k}^{t+k+1}\sqrt{Y(s)}dB(s)\right)\right. \\ &\quad \left.\times\left(\int_{t}^{t+1}(\mu-\beta Y(s))ds+\int_{t}^{t+1}\sqrt{Y(s)}dB(s)\right)\right] \\ &= \mathbb{E}\left[\int_{t+k}^{t+k+1}(\mu-\beta Y(s))ds\int_{t}^{t+1}(\mu-\beta Y(u))du\right] \\ &\quad +\mathbb{E}\left[\int_{t+k}^{t+k+1}(\mu-\beta Y(s))ds\int_{t}^{t+1}\sqrt{Y(u)}dB(u)\right. \\ &\quad \left.+\int_{t}^{t+1}(\mu-\beta Y(u))du\int_{t+k}^{t+k+1}\sqrt{Y(s)}dB(s)\right] \\ &\quad +\mathbb{E}\left[\int_{t+k}^{t+k+1}\sqrt{Y(s)}dB(s)\int_{t}^{t+1}\sqrt{Y(u)}dB(u)\right].\end{aligned} \tag{29}$$

Recall the filtration $\mathscr{F}_t^Y := \{Y_u, u\le t\}$, i.e. the σ-algebra generated by Y, revealing all information of the stochastic volatility up to time t. By Fubini's Theorem and iterated expectations we obtain for the $ds\times dB$-integrals, for $k>0$,

$$\begin{aligned}&\mathbb{E}\left[\int_{t+k}^{t+k+1}(\mu-\beta Y(s))ds\int_{t}^{t+1}\sqrt{Y(u)}dB(u)\right] \\ &=\mathbb{E}\left[\int_{t+k}^{t+k+1}\left[(\mu-\beta Y(s))\int_{t}^{t+1}\sqrt{Y(u)}dB(u)\right]ds\right] \\ &=\int_{t+k}^{t+k+1}\mathbb{E}\left(\mathbb{E}\left[(\mu-\beta Y(s))\int_{t}^{t+1}\sqrt{Y(u)}dB(u)|\mathscr{F}_s^Y\right]\right)ds \\ &=\int_{t+k}^{t+k+1}\mathbb{E}\left((\mu-\beta Y(s))\mathbb{E}\left[\int_{t}^{t+1}\sqrt{Y(u)}dB(u)|\mathscr{F}_s^Y\right]\right)ds=0.\end{aligned}$$

The last line follows since $\mu-\beta Y(s)$ is $\mathscr{F}_s^Y$-measurable and $\mathbb{E}\left[\int_t^{t+1}\sqrt{Y(u)}dB(u)|\mathscr{F}_s^Y\right]=0$. Similarly we obtain

$$\begin{aligned}&\mathbb{E}\left[\int_{t}^{t+1}(\mu-\beta Y(s))ds\int_{t+k}^{t+k+1}\sqrt{Y(u)}dB(u)\right] \\ &=\int_{t}^{t+1}\mathbb{E}\left((\mu-\beta Y(s))\mathbb{E}\left[\int_{t+k}^{t+k+1}\sqrt{Y(u)}dB(u)|\mathscr{F}_s^Y\right]\right)ds=0.\end{aligned}$$

For the product of the Itô integrals we obtain by the Itô isometry

$$\begin{aligned}&\mathbb{E}\left[\int_{t+k}^{t+k+1}\sqrt{Y(s)}dB(s)\int_{t}^{t+1}\sqrt{Y(u)}dB(u)\right] \\ &=\mathbb{E}\left[\int_0^T\mathbf{1}(t+k\le s\le t+k+1)\sqrt{Y(s)}dB(s)\right. \\ &\quad\left.\times\int_0^T\mathbf{1}(t\le s\le t+1)\sqrt{Y(s)}dB(s)\right] \\ &=\mathbb{E}\left[\int_0^T\mathbf{1}(t+k\le s\le t+k+1)\mathbf{1}(t\le s\le t+1)Y(s)ds\right]=0.\end{aligned}$$

for T large enough and $k > 0$.

Let us now consider the remaining term in (29).

$$\begin{aligned}
&\mathbb{E}\left[\int_{t+k}^{t+k+1}(\mu-\beta Y(s))ds\int_{t}^{t+1}(\mu-\beta Y(u))du\right]\\
&=\mathbb{E}\left[\int_{t+k}^{t+k+1}\int_{t}^{t+1}(\mu-\beta Y(s))(\mu-\beta Y(u))duds\right]\\
&=\int_{t+k}^{t+k+1}\int_{t}^{t+1}\mathbb{E}\left[(\mu-\beta Y(s))(\mu-\beta Y(u))\right]duds,
\end{aligned}$$

where we used Fubini's theorem in the last step. The expectation expands to

$$\mathbb{E}\left[(\mu-\beta Y(s))(\mu-\beta Y(u))\right]=\mu^2-\mu\beta\left(\mathbb{E}\left[Y(s)\right]+\mathbb{E}\left[Y(u)\right]\right)+\beta^2\mathbb{E}\left[Y(s)Y(u)\right].$$

Note that we may write Y in explicit form by applying Itô's Formula to the function $f(y,t)=y\mathrm{e}^{\gamma t}$;

$$\begin{aligned}
df(Y(t),t)&=\gamma Y(t)\mathrm{e}^{\gamma t}dt+\mathrm{e}^{\gamma t}dY(t)\\
&=\gamma Y(t)\mathrm{e}^{\gamma t}dt+\mathrm{e}^{\gamma t}(-\gamma(Y(t)-\lambda)dt+\kappa\sqrt{Y(t)}dW(t))\\
&=\gamma\lambda\mathrm{e}^{\gamma t}dt+\kappa\mathrm{e}^{\gamma t}\sqrt{Y(t)}dW(t).
\end{aligned}$$

Integrating from 0 to t with $Y(t)=y$ at $t=0$, we obtain

$$Y(t)=y\mathrm{e}^{-\gamma t}+\lambda(1-\mathrm{e}^{-\gamma t})+\kappa\int_0^t\mathrm{e}^{\gamma(s-t)}\sqrt{Y(s)}dW(s). \tag{30}$$

Hence,

$$\mathbb{E}\left[Y(t)\right]=y_0\mathrm{e}^{-\gamma t}+\lambda\left(1-\mathrm{e}^{-\gamma t}\right). \tag{31}$$

From (30) and (31) we get

$$\begin{aligned}
Y(u)Y(s)=&\left(\mathbb{E}\left[Y(u)\right]+\kappa\int_0^u\mathrm{e}^{\gamma(v-u)}\sqrt{Y(v)}dW(v)\right)\\
&\times\left(\mathbb{E}\left[Y(s)\right]+\kappa\int_0^s\mathrm{e}^{\gamma(v-s)}\sqrt{Y(v)}dW(v)\right).
\end{aligned} \tag{32}$$

Upon taking expectation after multiplying, the Brownian integrals vanish. The Itô isometry yields

$$\begin{aligned}
&\mathbb{E}\left[\int_0^u\mathrm{e}^{\gamma(v-u)}\sqrt{Y(v)}dW(v)\int_0^s\mathrm{e}^{\gamma(v-s)}\sqrt{Y(v)}dW(v)\right]\\
&=\mathrm{e}^{-\gamma(u+s)}\mathbb{E}\left[\int_0^{min(s,u)}\mathrm{e}^{2\gamma v}Y(v)dv\right]\\
&=\mathrm{e}^{-\gamma(u+s)}\int_0^{min(s,u)}\mathrm{e}^{2\gamma v}\mathbb{E}\left[Y(v)\right]dv,
\end{aligned}$$

where we used Fubini's theorem in the last step. We obtain

$$\mathbb{E}[Y(u)Y(s)]=\mathbb{E}[Y(u)]E[Y(s)]+\kappa^2\mathrm{e}^{-\gamma(u+s)}\int_0^{min(s,u)}\mathrm{e}^{2\gamma v}\mathbb{E}\left[Y(v)\right]dv$$

and thus,

$$\mathbb{E}\left[(\mu-\beta Y(s))(\mu-\beta Y(u))\right]=\mu^2-\mu\beta\left(\mathbb{E}\left[Y(s)\right]+\mathbb{E}\left[Y(u)\right]\right)$$
$$+\beta^2\mathbb{E}[Y(u)]E[Y(s)]+\beta^2\kappa^2\mathrm{e}^{-\gamma(u+s)}\int_0^{min(s,u)}\mathrm{e}^{2\gamma v}\mathbb{E}\left[Y(v)\right]dv.$$

It follows that, for $k>0$

$$\mathbb{E}[L(t)L(t+k)]=\int_{t+k}^{t+k+1}\int_t^{t+1}\mu^2-\mu\beta\left(\mathbb{E}\left[Y(s)\right]+\mathbb{E}\left[Y(u)\right]\right)+\beta^2\mathbb{E}[Y(u)]E[Y(s)]duds$$
$$+\beta^2\kappa^2\int_{t+k}^{t+k+1}\int_t^{t+1}(\mathrm{e}^{-\gamma(u+s)}\int_0^u\mathrm{e}^{2\gamma v}\mathbb{E}\left[Y(v)\right]dv)duds \tag{33}$$

since $u<s$. To calculate (33), we plug in the expression (31) for the expectation of Y and

$$\mathbb{E}[Y(u)]\mathbb{E}[Y(s)]=\mathrm{e}^{-\gamma(s+u)}(y_0-\lambda)^2+\lambda(y_0-\lambda)(\mathrm{e}^{-\gamma s}+\mathrm{e}^{-\gamma u})+\lambda^2.$$

Also,

$$\mathrm{e}^{-\gamma(u+s)}\int_0^u\mathrm{e}^{2\gamma v}\mathbb{E}\left[Y(v)\right]dv=\frac{1}{2\gamma}\left[\mathrm{e}^{-\gamma(s+u)}(\lambda-2y_0)+2\mathrm{e}^{-\gamma s}(y_0-\lambda)+\lambda\mathrm{e}^{-\gamma(s-u)}\right]$$

yielding,

$$\mathbb{E}[L(t)L(t+k)]=\int_{t+k}^{t+k+1}\int_t^{t+1}\mu^2-\mu\beta y_0(\mathrm{e}^{-\gamma s}+\mathrm{e}^{-\gamma u})-2\mu\beta\lambda+\mu\beta\lambda(\mathrm{e}^{-\gamma s}+\mathrm{e}^{-\gamma u})duds \tag{34}$$
$$+\beta^2\int_{t+k}^{t+k+1}\int_t^{t+1}\mathrm{e}^{-\gamma(s+u)}(y_0-\lambda)^2+\lambda(y_0-\lambda)(\mathrm{e}^{-\gamma s}+\mathrm{e}^{-\gamma u})+\lambda^2duds$$
$$+\frac{\beta^2\kappa^2}{2\gamma}\int_{t+k}^{t+k+1}\int_t^{t+1}\mathrm{e}^{-\gamma(s+u)}(\lambda-2y_0)+2\mathrm{e}^{-\gamma s}(y_0-\lambda)+\lambda\mathrm{e}^{-\gamma(s-u)}duds$$
$$=\mu^2-\mu\beta y_0(\int_{t+k}^{t+k+1}\mathrm{e}^{-\gamma s}ds+\int_t^{t+1}\mathrm{e}^{-\gamma u}du)-2\mu\beta\lambda+\mu\beta\lambda(\int_{t+k}^{t+k+1}\mathrm{e}^{-\gamma s}ds+\int_t^{t+1}\mathrm{e}^{-\gamma u}du)$$
$$+\beta^2(y_0-\lambda)^2\int_{t+k}^{t+k+1}\int_t^{t+1}\mathrm{e}^{-\gamma(s+u)}duds$$
$$+\lambda(y_0-\lambda)\beta^2(\int_{t+k}^{t+k+1}\mathrm{e}^{-\gamma s}ds+\int_t^{t+1}\mathrm{e}^{-\gamma u}du)+\lambda^2\beta^2$$
$$+\frac{\beta^2\kappa^2}{2\gamma}\left[(\lambda-2y_0)\int_{t+k}^{t+k+1}\int_t^{t+1}\mathrm{e}^{-\gamma(s+u)}duds\right.$$
$$\left.+2(y_0-\lambda)\int_{t+k}^{t+k+1}\mathrm{e}^{-\gamma s}ds+\lambda\int_{t+k}^{t+k+1}\int_t^{t+1}\mathrm{e}^{-\gamma(s-u)}duds\right].$$

for $min(s,u)=u$.

Calculation of the integrals in (34):
For any $s>0$ we have

$$\int_{t+k}^{t+k+1}\mathrm{e}^{-\gamma s}ds=\left(1-\mathrm{e}^{-\gamma}\right)\frac{1}{\gamma}\mathrm{e}^{-\gamma(t+k)}, \tag{35}$$

$$\int_{t+k}^{t+k+1}\int_t^{t+1}\mathrm{e}^{-\gamma(s+u)}duds=\frac{1-2\mathrm{e}^{-\gamma}+\mathrm{e}^{-2\gamma}}{\gamma^2}\mathrm{e}^{-\gamma(2t+k)}, \tag{36}$$

$$\int_{t+k}^{t+k+1}\int_t^{t+1}\mathrm{e}^{-\gamma(s-u)}duds=\frac{1}{\gamma^2}(\mathrm{e}^{\gamma}+\mathrm{e}^{-\gamma}-2)\mathrm{e}^{-\gamma k}. \tag{37}$$

In the limit we see that (37) is the only term contributing. Hence, in stationary we get

$$\mathbb{E}[L(t)L(t+k)]=(\mu-\lambda\beta)^2+\frac{\beta^2\kappa^2\lambda}{2\gamma^3}\mathrm{e}^{-\gamma k}(\mathrm{e}^{\gamma}+\mathrm{e}^{-\gamma}-2).$$

The stationary covariance becomes

$$cov(L(t+k),L(t)) = \frac{\beta^2\kappa^2\lambda}{2\gamma^3}\mathrm{e}^{-\gamma k}(\mathrm{e}^{\gamma}+\mathrm{e}^{-\gamma}-2)$$

since $\mathbb{E}[L(t)] = \mathbb{E}[L(t+k)] = \mu - \lambda\beta$ in stationary. Clearly, this tends to zero as k becomes large. The stationary autocorrelation *Autocorr* becomes

$$Autocorr(L(t+k),L(t)) := \frac{cov(L(t+k),L(t))}{\sqrt{var(L(t+k))}\sqrt{var(L(t))}} = \frac{\beta^2\kappa^2}{(\beta^2\kappa+2\gamma)\gamma^2}\mathrm{e}^{-\gamma k}(\mathrm{e}^{\gamma}+\mathrm{e}^{-\gamma}-2).$$

Appendix 2: Proof of Proposition 3

To start, we need some intermediate results and lemmas. The lemmas are of technical nature and their detailed proofs can be found in Appendix 3. Define the processes X and Z as

$$X(s) := \int_t^s \Phi_1(u)du + \int_t^s \Psi_1(u)dB(u), \tag{38}$$

$$Z(s) := \int_t^s \Phi_2(u)du + \int_t^s \Psi_2(u)dB(u), \tag{39}$$

where Φ and Ψ will be defined below. Apply Itô formula to $f(x,z) = x^2z^2$:

$$\begin{aligned} df(X(u),Z(u)) = &\left[2X(u)Z^2(u)\Phi_1(u) + 2X^2(u)Z(u)\Phi_2(u)\right. \\ &\left. + Z^2\Psi_1^2(u) + 2X(u)Z(u)\Psi_1(u)\Psi_2(u) + X^2(u)\Psi_2^2(u)\right]du \\ &+ \left[2X(u)Z^2(u)\Psi_1(u) + 2X^2(u)Z(u)\Psi_2(u)\right]dB(u). \end{aligned}$$

Taking expectation yields,

$$\begin{aligned} &\mathbb{E}[X^2(s)Z^2(s)] \\ &= \mathbb{E}\left[\int_t^s \left(2X(u)Z^2(u)\Phi_1(u) + 2X^2(u)Z(u)\Phi_2(u)\right.\right. \\ &\quad \left.\left. + Z^2\Psi_1^2(u) + 2X(u)Z(u)\Psi_1(u)\Psi_2(u) + X^2(u)\Psi_2^2(u)\right)du\right]. \end{aligned} \tag{40}$$

Let $s = t+k+1$ and define

$$\begin{aligned} \Phi_1(u) &:= \mathbf{1}(t \le u \le t+1)(\mu - \beta Y(u)) \\ \Psi_1(u) &:= \mathbf{1}(t \le u \le t+1)\sqrt{Y(u)} \\ \Phi_2(u) &:= \mathbf{1}(t+k \le u \le t+k+1)(\mu - \beta Y(u)) \\ \Psi_2(u) &:= \mathbf{1}(t+k \le u \le t+k+1)\sqrt{Y(u)}. \end{aligned} \tag{41}$$

The covariance is given by $cov(L^2(t+k),L^2(t)) = \mathbb{E}[L^2(t+k)L^2(t)] - \mathbb{E}[L^2(t+k)]\mathbb{E}[L^2(t)]$. To find this we will make frequent use of Fubini's Theorem, the tower property and the Markov property of the process Y. Furthermore, we will end up with expressions involving conditional expectations of powers of Y. We start to find formulas to calculate the conditional expectation that will also simplify the notation. To begin, by the Markov property we have for $t < s$

$$\mathbb{E}[Y^n(s)|\mathscr{F}_t^Y] = \mathbb{E}[(Y^{t,y})^n(s)]_{y=Y(t)}, \tag{42}$$

where $\mathbb{E}[(Y^{t,y})^n(s)]_{y=Y(t)}$ is the expectation of the process Y^n starting at $y = Y(t)$ at time t for an integer $n = 1,2..$ indicating the nth power of Y. The nth moment is given by

$$\mathbb{E}[Y^n(s)] = \mathbb{E}[(Y^{0,y_0})^n(s)]_{y_0=Y(0)}.$$

Let

$$\xi(t) := Y^n(t).$$

We have the following result.

Lemma 1. *The conditional expectation for the process Y defined in* (30) *is given by*

$$\mathbb{E}[(Y^{t,y})^n(s)]_{y=Y(t)} = Y^n(t)\mathrm{e}^{-n\gamma(s-t)} + n\left(\lambda\gamma + \frac{1}{2}(n-1)\kappa^2\right)\int_t^s \mathrm{e}^{-n\gamma(s-u)}\mathbb{E}[(Y^{t,y})^{n-1}(u)]_{y=Y(t)}du.$$

The explicit expressions for $\mathbb{E}[(Y^{t,y})^n(s)]_{y=Y(t)}$ for $n = 1,2,3,4$ are given by $(54)-(60)$ in Appendix 3. Let $x \in \mathbb{R}_+$ and define

$$C_k := k\left(\lambda\gamma + \frac{1}{2}(k-1)\kappa^2\right), \quad k = 1,2\ldots \tag{43}$$

$$F_\theta^n(x) := \sum_{m=0}^n f_m^n(\theta)x^m, \tag{44}$$

where

$$f_m^n(\theta) := \frac{(1-\mathrm{e}^{-\gamma\theta})^{(n-m)}\mathrm{e}^{-m\gamma\theta}}{(n-m)!\gamma^{n-m}}\prod_{k=0}^{n-m} C_k \tag{45}$$

and $C_0 := 1$. Since $\gamma \geq 0$ clearly $f_m^n(\theta)$ are finite for all m,n and $\theta \geq 0$, and hence $F_\theta^n(x)$ is finite.

Lemma 2. *Let $Y(0) = y_0$ and for $s > u > v > t \geq 0$. Then*

$$\mathbb{E}[Y^n(s)|\mathscr{F}_t^Y] = F_{s-t}^n(Y(t)). \tag{46}$$

In particular,

$$\mathbb{E}[Y^n(s)|\mathscr{F}_0^Y] := \mathbb{E}[Y^n(s)] = F_s^n(y_0). \tag{47}$$

Also,

$$\mathbb{E}\left[F_{s-t}^n(Y(t))\right] = f_0^n(s-t) + \sum_{m=1}^n f_m^n(s-t)\sum_{i=0}^m f_i^m(t)y_0^i, \tag{48}$$

$$\mathbb{E}\left[Y^k(v)F_{s-u}^n(Y(u))\right] = f_0^n(s-u)\sum_{c=0}^k f_c^k(v)y_0^c + \sum_{m=1}^n f_m^n(s-u)\sum_{i=0}^m f_i^m(u-v)\sum_{j=0}^{k+i} f_j^{k+i}(v)y_0^j, \tag{49}$$

$$\mathbb{E}\left[Y^l(u)Y^k(v)F_{s-u}^n(Y(u))\right] = \sum_{m=0}^n f_m^n(s-u)\sum_{i=0}^{m+l} f_i^{m+l}(u-v)\sum_{j=0}^{k+i} f_j^{k+i}(v)y_0^j \tag{50}$$

and

$$\mathbb{E}\left[Y^k(u)F_{s-u}^n(Y(u))\right] = \sum_{m=0}^n f_m^n(s-u)\sum_{i=0}^{m+k} f_i^{m+k}(u)y_0^i. \tag{51}$$

Furthermore, all expectations above are finite for all m,n,k,l and $s > u > v > t \geq 0$.

Lemma 3. *The expectation $\mathbb{E}[L^2(t)L^2(t+k)]$ is given by*

$$\mathbb{E}[L^2(t)L^2(t+k)] = D_0(t,k)\mathbb{E}[L^2(t)] + D_1(t,k)\mathbb{E}[L^2(t)Y(t+1)] + D_2(t,k)\mathbb{E}[L^2(t)Y^2(t+1)],$$

for some finite deterministic functions $D_n(t,k)$.

The functions $D_n(t,k)$, for $n = 0,1,2$, are given by $(65)-(67)$ in Appendix 3.

Lemma 4. *The expectation* $\mathbb{E}[L^2(t)Y^n(t+1)]$ *is given by*

$$\begin{aligned}
&\mathbb{E}[L^2(t)Y^n(t+1)] \\
&= 2\int_t^{t+1}\int_t^u \Big[\mu^2\mathbb{E}\left[F^n_{t+1-u}(Y(u))\right] - \mu\beta\mathbb{E}\left[(Y(v)+Y(u))F^n_{t+1-u}(Y(u))\right] \\
&\quad + \beta^2\mathbb{E}\left[Y(u)Y(v)F^n_{t+1-u}(Y(u))\right]\Big] dvdu + \int_t^{t+1}\mathbb{E}\left[Y(u)F^n_{t+1-u}(Y(u))\right]du,
\end{aligned}$$

for $n>0$. *Furthermore,* $\mathbb{E}[L^2(t)Y^n(t+1)]\leq\infty$.

Proof of Proposition 3

We now have the necessary tools to prove Proposition 3. The stationary autocorrelation of squared logreturns is given by

$$\lim_{t\to\infty} Autocorr(L^2(t),L^2(t+k)) = \lim_{t\to\infty}\frac{\mathbb{E}[L^2(t)L^2(t+k)]-\mathbb{E}[L^2(t)]\mathbb{E}[L^2(t+k)]}{\sqrt{var(L^2(t+k))}\sqrt{var(L^2(t))}}.$$

Since the the expectations and the functions $D_n(t,k)$ are finite, it follows that $\mathbb{E}[L^2(t)L^2(t+k)]$ is finite. Furthermore, since L has finite moments of all orders we can write

$$\begin{aligned}
&\lim_{t\to\infty} Autocorr(L^2(t),L^2(t+k)) \\
&= \frac{\lim_{t\to\infty}\mathbb{E}[L^2(t)L^2(t+k)]-\lim_{t\to\infty}\mathbb{E}[L^2(t)]\lim_{t\to\infty}\mathbb{E}[L^2(t+k)]}{\sqrt{\lim_{t\to\infty}\mathbb{E}[L^4(t+k)]-(\lim_{t\to\infty}\mathbb{E}[L^2(t+k)])^2}\sqrt{\lim_{t\to\infty}\mathbb{E}[L^4(t)]-(\lim_{t\to\infty}\mathbb{E}[L^2(t)])^2}}.
\end{aligned}$$

But, for all $n\geq 0$

$$\lim_{t\to\infty}\mathbb{E}[L^n(t+k)] = \lim_{t\to\infty}\mathbb{E}[L^n(t)] = (-i)^n\frac{\partial^n\phi}{\partial u^n}(0) := \mathbb{E}[L^n],$$

where $\phi(u)$ is the stationary characteristic function for L defined in (15). It follows that $\mathbb{E}[L^n]$ is a constant only depending on the parameters in the distribution of L. Thus,

$$\lim_{t\to\infty} Autocorr(L^2(t),L^2(t+k)) = \frac{\lim_{t\to\infty}\mathbb{E}[L^2(t)L^2(t+k)]-(\mathbb{E}[L^2])^2}{\mathbb{E}[L^4]-(\mathbb{E}[L^2])^2}.$$

Combining Lemma 3 and Lemma 4 we get

$$\begin{aligned}
&\mathbb{E}[L^2(t)L^2(t+k) = D_0(t,k)\mathbb{E}[L^2(t)] \\
&+2D_1(t,k)\Bigg(\int_t^{t+1}\int_t^u\Big[\mu^2\mathbb{E}\left[F^1_{t+1-u}(Y(u))\right]-\mu\beta\mathbb{E}\left[(Y(v)+Y(u))F^1_{t+1-u}(Y(u))\right] \\
&\quad+\beta^2\mathbb{E}\left[Y(u)Y(v)F^1_{t+1-u}(Y(u))\right]\Big]dvdu+\frac{1}{2}\int_t^{t+1}\mathbb{E}\left[Y(u)F^1_{t+1-u}(Y(u))\right]du\Bigg) \\
&+2D_2(t,k)\Bigg(\int_t^{t+1}\int_t^u\Big[\mu^2\mathbb{E}\left[F^2_{t+1-u}(Y(u))\right]-\mu\beta\mathbb{E}\left[(Y(v)+Y(u))F^2_{t+1-u}(Y(u))\right] \\
&\quad+\beta^2\mathbb{E}\left[Y(u)Y(v)F^2_{t+1-u}(Y(u))\right]\Big]dvdu+\frac{1}{2}\int_t^{t+1}\mathbb{E}\left[Y(u)F^2_{t+1-u}(Y(u))\right]du\Bigg).
\end{aligned} \tag{52}$$

Applying Lemma 2 and the expressions for $D_n(t,k)$ given by (65) – (67) in Appendix 3, the expectation (52) can be calculated by elementary calculus. Passing to the limit, yields

$$\lim_{t\to\infty}\mathbb{E}[L^2(t),L^2(t+k)] = \tilde{A}_0+\tilde{A}_1\mathrm{e}^{-\gamma k}+\tilde{A}_2\mathrm{e}^{-2\gamma k},$$

for constants $\tilde{A}$ depending on the parameters of the distribution of L. Hence

$$\lim_{t\to\infty} Autocorr(L^2(t),L^2(t+k)) = A_0+A_1\mathrm{e}^{-\gamma k}+A_2\mathrm{e}^{-2\gamma k}.$$

Note that the calculation of the squared autocorrelation function is straightforward via (52).

Appendix 3: Proof of Lemmas in Appendix 2

Proof of Lemma 1

By Itô's Formula we get

$$\begin{aligned} d\xi(t) &= nY^{n-1}(-\gamma(Y(t)-\lambda))dt+nY(t)^{n-1}\kappa\sqrt{Y(t)}dW(t)+\frac{1}{2}n(n-1)Y^{n-2}(t)\kappa^2Y(t)dt \\ &= \left[-n\gamma Y^n(t)+Y^{n-1}(t)\left(n\lambda\gamma+\frac{1}{2}n(n-1)\kappa^2\right)\right]dt+nY(t)^{n-1}\kappa\sqrt{Y(t)}dW(t). \end{aligned} \tag{53}$$

To find an explicit solution for ξ, we apply Itô's Formula to the function $f(\xi(t),t) := \xi(t)\mathrm{e}^{n\gamma t}$,

$$\begin{aligned} df(\xi(u),u) &= n\gamma\xi\mathrm{e}^{n\gamma u}du+\mathrm{e}^{n\gamma u}d\xi(u) \\ &= \left[nY^{n-1}(u)\mathrm{e}^{n\gamma u}\left(\lambda\gamma+\frac{1}{2}(n-1)\kappa^2\right)\right]du+nY(u)^{n-1}\kappa\sqrt{Y(u)}\mathrm{e}^{n\gamma u}dW(u). \end{aligned}$$

Integrating from t to s yields

$$\begin{aligned} Y^n(s) = Y^n(t)\mathrm{e}^{-n\gamma(s-t)} &+n\left(\lambda\gamma+\frac{1}{2}(n-1)\kappa^2\right)\int_t^s\mathrm{e}^{-n\gamma(s-u)}Y^{n-1}(u)du \\ &+n\kappa\int_t^s\mathrm{e}^{-n\gamma(s-u)}Y^{n-1}(u)\sqrt{Y(u)}dW(u). \end{aligned}$$

Taking expectation and using Fubini's Theorem we obtain

$$\mathbb{E}[Y^{n,t,y}(s)]_{y=Y(t)} = Y^n(t)\mathrm{e}^{-n\gamma(s-t)}+n\left(\lambda\gamma+\frac{1}{2}(n-1)\kappa^2\right)\int_t^s\mathrm{e}^{-n\gamma(s-u)}\mathbb{E}[Y^{n-1,t,y}(u)]_{y=Y(t)}du.$$

This completes the proof of Lemma 1.

The explicit expression for $\mathbb{E}[Y^{n,t,y}(s)]_{y=Y(t)}$ for $n=1,2,3,4$ is calculated as;
For $n=1$:

$$\begin{aligned} \mathbb{E}[Y^{t,y}(s)]_{y=Y(t)} &= Y(t)\mathrm{e}^{-\gamma(s-t)}+C_1\int_t^s\mathrm{e}^{-\gamma(s-u)}du \\ &= Y(t)\mathrm{e}^{-\gamma(s-t)}+\frac{C_1}{\gamma}\left(1-\mathrm{e}^{-\gamma(s-t)}\right) \\ &:= Y(t)f_1^1(s-t)+f_0^1(s-t). \end{aligned} \tag{54}$$

In particular,

$$\begin{aligned}\mathbb{E}[Y(s)] =& y_0 \mathrm{e}^{-\gamma(s)} + \frac{C_1}{\gamma}\left(1-\mathrm{e}^{-\gamma(s)}\right)\\ :=& Y(s) f_1^1(s) + f_0^1(s).\end{aligned} \tag{55}$$

For $n=2$:

$$\begin{aligned}\mathbb{E}[Y^{2,t,y}(s)]_{y=Y(t)} &= Y^2(t)\mathrm{e}^{-2\gamma(s-t)} + \left(2\lambda\gamma+\kappa^2\right)\int_t^s \mathrm{e}^{-2\gamma(s-u)}\mathbb{E}[Y^{t,y}(u)]_{y=Y(t)}du\\ &= \mathrm{e}^{-2\gamma s}\left[Y^2(t)\mathrm{e}^{2\gamma t} + C_2\int_t^s \mathrm{e}^{2\gamma u}\left(\mathrm{e}^{-\gamma u}\left[Y(t)\mathrm{e}^{\gamma t} + \frac{C_1}{\gamma}(\mathrm{e}^{\gamma u}-\mathrm{e}^{\gamma t})\right]\right)du\right]\\ &= Y^2(t)\mathrm{e}^{-2\gamma(s-t)} + Y(t)\frac{C_2}{\gamma}(1-\mathrm{e}^{-\gamma(s-t)})\mathrm{e}^{-\gamma(s-t)} + \frac{C_1C_2}{2\gamma^2}(1-\mathrm{e}^{-\gamma(s-t)})^2\\ &:= Y^2(t) f_2^2(s-t) + Y(t) f_1^2(s-t) + f_0^2(s-t).\end{aligned} \tag{56}$$

In particular,

$$\begin{aligned}\mathbb{E}[Y^2(s)] &= y_0^2\mathrm{e}^{-2\gamma s} + y_0C_2(1-\mathrm{e}^{-\gamma s})\mathrm{e}^{-\gamma s} + \frac{C_1C_2}{2\gamma^2}(1-\mathrm{e}^{-\gamma s})^2\\ &:= Y^2(s) f_2^2(s) + Y(s) f_1^2(s) + f_0^2(s).\end{aligned} \tag{57}$$

For $n=3$:

$$\begin{aligned}\mathbb{E}[Y^{3,t,y}(s)]_{y=Y(t)} &= Y^3(t)\mathrm{e}^{-3\gamma(s-t)} + 3\left(\lambda\gamma+\kappa^2\right)\int_t^s \mathrm{e}^{-3\gamma(s-u)}\mathbb{E}[Y^{2,t,y}(u)]_{y=Y(t)}du\\ &= \mathrm{e}^{-3\gamma s}\left[Y^3(t)\mathrm{e}^{3\gamma t} + C_3\int_t^s \mathrm{e}^{3\gamma u}\left(\mathrm{e}^{-2\gamma u}\left[Y^2(t)\mathrm{e}^{2\gamma t} + Y(t)\mathrm{e}^{\gamma t}\frac{C_2}{\gamma}(\mathrm{e}^{\gamma u}-\mathrm{e}^{\gamma t})\right.\right.\right.\\ &\quad \left.\left.\left. + \frac{C_1C_2}{2\gamma^2}(\mathrm{e}^{2\gamma u} - 2\mathrm{e}^{\gamma(u+t)} + \mathrm{e}^{2\gamma t})\right]\right)du\right]\\ &= Y^3(t)\mathrm{e}^{-3\gamma(s-t)} + Y^2(t)\frac{C_3}{\gamma}(1-\mathrm{e}^{-\gamma(s-t)})\mathrm{e}^{-2\gamma(s-t)}\\ &\quad + Y(t)\frac{C_2C_3}{2\gamma^2}(1-\mathrm{e}^{-\gamma(s-t)})^2\mathrm{e}^{-\gamma(s-t)} + \frac{C_1C_2C_3}{6\gamma^3}(1-\mathrm{e}^{-\gamma(s-t)})^3\\ &:= Y^3(t) f_3^3(s-t) + Y^2(t) f_2^3(s-t) + Y(t) f_1^3(s-t) + f_0^3(s-t).\end{aligned} \tag{58}$$

In particular,

$$\begin{aligned}\mathbb{E}[Y^3(s)] &= y_0^3\mathrm{e}^{-3\gamma s} + y_0^2\frac{C_3}{\gamma}(1-\mathrm{e}^{-\gamma s})\mathrm{e}^{-2\gamma s}\\ &\quad + y_0\frac{C_2C_3}{2\gamma^2}(1-\mathrm{e}^{-\gamma s})^2\mathrm{e}^{-\gamma s} + \frac{C_1C_2C_3}{6\gamma^3}(1-\mathrm{e}^{-\gamma s})^3\\ &:= Y^3(s) f_3^3(s) + Y^2(s) f_2^3(s) + Y(s) f_1^3(s) + f_0^3(s).\end{aligned} \tag{59}$$

For $n=4$:

$$\begin{aligned}
\mathbb{E}[Y^{4,t,y}(s)]_{y=Y(t)} &= Y^4(t)\mathrm{e}^{-4\gamma(s-t)} + 2\left(2\lambda\gamma+3\kappa^2\right)\int_t^s \mathrm{e}^{-4\gamma(s-u)}\mathbb{E}[Y^{3,t,y}(u)]_{y=Y(t)}du \\
&= \mathrm{e}^{-4\gamma s}\Bigg[Y^4(t)\mathrm{e}^{4\gamma t)} + C_4\int_t^s \mathrm{e}^{4\gamma u}\mathrm{e}^{-3\gamma u}Y^3(t)\mathrm{e}^{3\gamma t}du \\
&\quad + C_4\int_t^s \mathrm{e}^{4\gamma u}\mathrm{e}^{-3\gamma u}Y^2(t)\mathrm{e}^{2\gamma t}\frac{C_3}{\gamma}(\mathrm{e}^{\gamma u}-\mathrm{e}^{\gamma t})du \\
&\quad + C_4\int_t^s \mathrm{e}^{4\gamma u}\mathrm{e}^{-3\gamma u}\frac{C_2C_3}{2\gamma^2}Y(t)\mathrm{e}^{\gamma t}(\mathrm{e}^{2\gamma u}-2\mathrm{e}^{\gamma(u+t)}+\mathrm{e}^{2\gamma t})du \\
&\quad + C_4\int_t^s \mathrm{e}^{4\gamma u}\mathrm{e}^{-3\gamma u}\frac{C_1C_2C_3}{6\gamma^3}\left(\mathrm{e}^{3\gamma u}-3\mathrm{e}^{\gamma(2u+t)}+3\mathrm{e}^{\gamma(u+2t)}+\mathrm{e}^{\gamma t}\right)du\Bigg] \\
&= Y^4(t)\mathrm{e}^{-4\gamma(s-t)} + Y^3(t)\frac{C_4}{\gamma}(1-\mathrm{e}^{-\gamma(s-t)})\mathrm{e}^{-3\gamma(s-t)} \\
&\quad + Y^2(t)\frac{C_3C_4}{2\gamma^2}(1-\mathrm{e}^{-\gamma(s-t)})^2\mathrm{e}^{-2\gamma(s-t)} \\
&\quad + Y(t)\frac{C_2C_3C_4}{6\gamma^3}(1-\mathrm{e}^{-\gamma(s-t)})^3\mathrm{e}^{-\gamma(s-t)} \\
&\quad + \frac{C_1C_2C_3C_4}{24\gamma^4}(1-\mathrm{e}^{-\gamma(s-t)})^4 \\
&:= Y^4(t)f_4^4(s-t) + Y^3(t)f_3^4(s-t) \\
&\quad + Y^2(t)f_2^4(s-t) + Y(t)f_1^4(s-t) + f_0^4(s-t).
\end{aligned} \tag{60}$$

In particular,

$$\begin{aligned}
\mathbb{E}[Y^4(s)] &= y_0^4\mathrm{e}^{-4\gamma s} + y_0^3\frac{C_4}{\gamma}(1-\mathrm{e}^{-\gamma s})\mathrm{e}^{-3\gamma s} \\
&\quad + y_0^2\frac{C_3C_4}{2\gamma^2}(1-\mathrm{e}^{-\gamma s})^2\mathrm{e}^{-2\gamma s} + y_0\frac{C_2C_3C_4}{6\gamma^3}(1-\mathrm{e}^{-\gamma s})^3\mathrm{e}^{-\gamma s} \\
&\quad + \frac{C_1C_2C_3C_4}{24\gamma^4}(1-\mathrm{e}^{-\gamma s})^4 \\
&:= Y^4(s)f_4^4(s) + Y^3(s)f_3^4(s) + Y^2(s)f_2^4(s) + Y(s)f_1^4(s) + f_0^4(s).
\end{aligned} \tag{61}$$

Proof of Lemma 2

First, since F_{s-t}^n is finite, the relation (46) is well defined. Clearly, the first part of the Lemma holds for $n = 1,2,3,4$ from the calculations above. The result for a general $n > 0$ follows directly by induction, proving (46). For the formulas we have, using (46),

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[Y^n(s)|\mathscr{F}_t^Y]\right] &= \sum_{m=0}^n f_m^n(s-t)\mathbb{E}[Y^m(t)] \\
&= f_0^n(s-t) + \sum_{m=1}^n f_m^n(s-t)\mathbb{E}[Y^m(t)].
\end{aligned}$$

Since $m > 0$ in the second term we can apply the first part of the Lemma again, i.e.

$$\mathbb{E}[Y^m(t)] = F_{t-0}^m(Y(0)) = \sum_{i=0}^m f_i^m(t)y_0^i.$$

Hence,

$$\mathbb{E}\left[\mathbb{E}[Y^n(s)|\mathscr{F}_t^Y]\right] = f_0^n(s-t) + \sum_{m=1}^{n} f_m^n(s-t) \sum_{i=0}^{m} f_i^m(t) y_0^i.$$

Proving (48). The remaining formulas are obtained as follows: Let $s > u > v > 0$ and $k > 0$:

$$\begin{aligned}
&\mathbb{E}\left[Y^k(v) F_{s-u}^n(Y(u))\right] \\
&= \mathbb{E}\left[Y^k(v) \sum_{m=0}^{n} f_m^n(s-u) Y^m(u)\right] = \sum_{m=0}^{n} f_m^n(s-u) \mathbb{E}\left[Y^k(v) Y^m(u)\right] \\
&= \sum_{m=0}^{n} f_m^n(s-u) \mathbb{E}\left[Y^k(v) \mathbb{E}[Y^m(u)|\mathscr{F}_v^Y]\right] = f_0^n(s-u)\mathbb{E}[Y^k(v)] \\
&\quad + \sum_{m=1}^{n} f_m^n(s-u) \mathbb{E}\left[Y^k(v) \mathbb{E}[Y^m(u)|\mathscr{F}_v^Y]\right] \\
&= f_0^n(s-u) F_v^k(Y(0)) + \sum_{m=1}^{n} f_m^n(s-u) \mathbb{E}\left[Y^k(v) F_{u-v}^m(Y(v))\right] \\
&= f_0^n(s-u) \sum_{c=0}^{k} f_c^k(v) y_0^c + \sum_{m=1}^{n} f_m^n(s-u) \mathbb{E}\left[Y^k(v) \sum_{i=0}^{m} f_i^m(u-v) Y^i(v)\right] \\
&= f_0^n(s-u) \sum_{c=0}^{k} f_c^k(v) y_0^c + \sum_{m=1}^{n} f_m^n(s-u) \sum_{i=0}^{m} f_i^m(u-v) \mathbb{E}\left[Y^{k+i}(v)\right] \\
&= f_0^n(s-u) \sum_{c=0}^{k} f_c^k(v) y_0^c + \sum_{m=1}^{n} f_m^n(s-u) \sum_{i=0}^{m} f_i^m(u-v) \sum_{j=0}^{k+i} f_j^{k+i}(v) y_0^j.
\end{aligned}$$

Similarly, it follows that, for $l > 0$,

$$\mathbb{E}\left[Y^l(u) Y^k(v) F_{s-u}^n(Y(u))\right] = \sum_{m=0}^{n} f_m^n(s-u) \sum_{i=0}^{m+l} f_i^{m+l}(u-v) \sum_{j=0}^{k+i} f_j^{k+i}(v) y_0^j$$

and

$$\mathbb{E}\left[Y^k(u) F_{s-u}^n(Y(u))\right] = \sum_{m=0}^{n} f_m^n(s-u) \sum_{i=0}^{m+k} f_i^{m+k}(u) y_0^i,$$

proving (50) and (51). The finiteness of these formulas follows from the fact that $f_m^n(\theta)$, for $\theta \geq 0$, is finite.

Proof of Lemma 3

Take $\theta \geq t+k+1$. Then, by definition of L, we get from (40) and (41)

$$\begin{aligned}
&\mathbb{E}[L^2(t)L^2(t+k)] = \mathbb{E}[X^2(\theta)Z^2(\theta)] \\
&= 2\int_t^{t+1} \mathbb{E}\left[X(u)Z^2(u)(\mu - \beta Y(u))\right] du + 2\int_{t+k}^{t+k+1} \mathbb{E}\left[X^2(u)Z(u)(\mu - \beta Y(u))\right] du \\
&\quad + \int_t^{t+1} \mathbb{E}\left[Y(u)Z^2(u)\right] du + 4\cdot 0 + \int_{t+k}^{t+k+1} \mathbb{E}\left[Y(u)X^2(u)\right] du \\
&= 2\int_{t+k}^{t+k+1} \mathbb{E}\left[X^2(u)Z(u)(\mu - \beta Y(u))\right] du + \int_{t+k}^{t+k+1} \mathbb{E}\left[Y(u)X^2(u)\right] du,
\end{aligned}$$

where the last equality follows since $Z(u) = 0$ for $u \in [t, t+1]$ by definition of Z in equation (39). We have, for $s \in [t+k, t+k+1]$:

$$\mathbb{E}[L^2(t)L^2(t+k)] = 2\int_{t+k}^{t+k+1} \mathbb{E}\left[X^2(s)Z(s)(\mu-\beta Y(s))\right] ds + \int_{t+k}^{t+k+1} \mathbb{E}\left[Y(s)X^2(s)\right] ds, \tag{62}$$

We start to calculate the last expectation. Note that by the definition of (41) it follows that $X(s)$, defined in (38) is $\mathscr{F}_{t+1}$-measurable. Whence,

$$\begin{aligned}\mathbb{E}\left[Y(s)X^2(s)\right] &= \mathbb{E}\left[X^2\mathbb{E}[Y(s)|\mathscr{F}_{t+1}]\right] \\ &= f_1^1(s-t-1)\mathbb{E}[X^2(s)Y(t+1)] + \lambda(1-f_1^1(s-t-1)\mathbb{E}[X^2(s)]).\end{aligned} \tag{63}$$

The last equality follows from (42) and (54). Now for the first term in (62) we have by the tower property

$$\mathbb{E}\left[X^2(s)Z(s)(\mu-\beta Y(s))\right] = \mathbb{E}\left[X^2(s)\mathbb{E}\left[Z(s)(\mu-\beta Y(s))|\mathscr{F}_{t+1}\right]\right],$$

since $X(s)$ is $\mathscr{F}_{t+1}$-measurable for $s \in [t+k, t+k+1]$. The conditional expectation

$$\mathbb{E}[Z(s)(\mu-\beta Y(s))|\mathscr{F}_{t+1}]$$

is calculated as,

$$\begin{aligned}&\mathbb{E}[Z(s)(\mu-\beta Y(s))|\mathscr{F}_{t+1}] \\ &= \mathbb{E}\left[(\mu-\beta Y(s))\left(\int_{t+k}^{s}(\mu-\beta Y(u))du + \int_{t+k}^{s}\sqrt{Y(u)}dB(u)\right)|\mathscr{F}_{t+1}\right] \\ &= \int_{t+k}^{s}\mathbb{E}[(\mu-\beta Y(u))\mathbb{E}[(\mu-\beta Y(s))|\mathscr{F}_u]\,|\mathscr{F}_{t+1}]\,du \\ &\quad + \mathbb{E}\left[(\mu-\beta Y(s))\mathbb{E}\left[\int_{t+k}^{s}\sqrt{Y(u)}dB(u)|\mathscr{F}_{t+1}\vee\mathscr{F}^Y\right]|\mathscr{F}_{t+1}\right].\end{aligned}$$

The last conditional expectation is zero. The remaining part yields,

$$\begin{aligned}&\mathbb{E}[Z(s)(\mu-\beta Y(s))|\mathscr{F}_{t+1}] \\ &= \int_{t+k}^{s}\mathbb{E}\left[(\mu-\beta Y(u))\left(\mu-\beta f_1^1(s-u)Y(u)-\beta\lambda(1-f_1^1(s-u))\right)|\mathscr{F}_{t+1}\right]du \\ &= \int_{t+k}^{s}\mathbb{E}\left[\mu^2-\mu\beta\lambda(1-f_1^1(s-u))+\beta^2 f_1^1(s-u)Y^2(u)\right. \\ &\quad \left.+Y(u)\left(\beta^2\lambda(1-f_1^1(s-u))-\mu\beta f_1^1(s-u)-\mu\beta\right)|\mathscr{F}_{t+1}\right]du \\ &= \mu^2(s-t-k)-\int_{t+k}^{s}\mu\beta\lambda(1-f_1^1(s-u))du \\ &\quad +\beta^2\int_{t+k}^{s}f_1^1(s-u)\left[Y^2(t+1)f_2^2(s-t-1)+Y(t+1)f_1^2(u-t-1)+f_0^2(u-t-1)\right]du \\ &\quad +\int_{t+k}^{s}\left[\left(\beta^2\lambda(1-f_1^1(s-u))-\mu\beta f_1^1(s-u)-\mu\beta\right)\right. \\ &\quad \left.\times\left(f_1^1(u-t-1)Y(t+1)+\lambda(1-f_1^1(u-t-1))\right)\right]du,\end{aligned}$$

by using (54) and (56). The expectation $\mathbb{E}\left[X^2(s)Z(s)(\mu-\beta Y(s))\right]$ becomes

$$
\begin{aligned}
&\mathbb{E}\left[X^2(s)Z(s)(\mu-\beta Y(s))\right]\\
&=\mathbb{E}\Bigg[X^2(s)\left(\mu^2(s-t-k)-\int_{t+k}^{s}\mu\beta\lambda(1-f_1^1(s-u))du\right)\\
&\quad+\beta^2\int_{t+k}^{s}X^2(s)f_1^1(s-u)\left[Y^2(t+1)f_2^2(u-t-1)+Y(t+1)f_1^2(u-t-1)+f_0^2(u-t-1)\right]du\\
&\quad+\int_{t+k}^{s}X^2(s)\left(\beta^2\lambda(1-f_1^1(s-u))-\mu\beta f_1^1(s-u)-\mu\beta\right)\\
&\quad\times\left[f_1^1(u-t-1)Y(t+1)+\lambda(1-f_1^1(u-t-1))\right]du\Bigg].
\end{aligned}
$$

Applying Fubini's Theorem yields,

$$
\begin{aligned}
&\mathbb{E}\left[X^2(s)Z(s)(\mu-\beta Y(s))\right] \qquad (64)\\
&=\mathbb{E}[X^2(s)]\left(\mu^2(s-t-k)-\int_{t+k}^{s}\mu\beta\lambda(1-f_1^1(s-u))du\right)\\
&\quad+\beta^2\int_{t+k}^{s}f_1^1(s-u)\mathbb{E}[X^2(s)Y^2(t+1)]f_2^2(u-t-1)du\\
&\quad+\beta^2\int_{t+k}^{s}f_1^1(s-u)\left[\mathbb{E}[X^2(s)Y(t+1)]f_1^2(u-t-1)+\mathbb{E}[X^2(s)]f_0^2(u-t-1)\right]du\\
&\quad+\int_{t+k}^{s}\left(\beta^2\lambda(1-f_1^1(s-u))-\mu\beta f_1^1(s-u)-\mu\beta\right)\\
&\quad\times\left[f_1^1(u-t-1)\mathbb{E}[X^2(s)Y(t+1)]+\mathbb{E}[X^2(s)]\lambda(1-f_1^1(u-t-1))\right]du.
\end{aligned}
$$

Since, for $s\in[t+k,t+k+1]$, $\mathbb{E}[X^2(s)Y^n(t+1)]$ is independent of s by definition of X. From the expressions (64) and (63) we get,

$$
\begin{aligned}
&\mathbb{E}[L^2(t)L^2(t+k)]\\
&=2\mathbb{E}[X^2(s)]\left(\int_{t+k}^{t+k+1}\left[\mu^2(s-t-k)+\frac{1}{2}f_0^1(s-t-1)\right.\right.\\
&\quad\left.\left.+\int_{t+k}^{s}\beta^2f_0^2(u-t-1)+(\beta^2f_0^1(s-u)-\mu\beta f_1^1(s-u)-\mu\beta)f_0^1(u-t-1)du\right]ds\right)\\
&\quad+2\mathbb{E}[X^2(s)Y(t+1)]\left(\int_{t+k}^{t+k+1}\left[\frac{1}{2}f_1^1(s-t-1)\right.\right.\\
&\quad\left.\left.+\int_{t+k}^{s}\beta^2f_1^2(u-t-1)+(\beta^2f_0^1(s-u)-\mu\beta f_1^1(s-u)-\mu\beta)f_1^1(u-t-1)du\right]ds\right)\\
&\quad+2\mathbb{E}[X^2(s)Y^2(t+1)]\left(\int_{t+k}^{t+k+1}\int_{t+k}^{s}\beta^2f_2^2(t+1,u)f_1^1(s-u)duds\right)\\
&:=D_0(t,k)\mathbb{E}[X^2(s)]+D_1(t,k)\mathbb{E}[X^2(s)Y(t+1)]+D_2(t,k)\mathbb{E}[X^2(s)Y^2(t+1)]\\
&:=D_0(t,k)\mathbb{E}[L^2(t)]+D_1(t,k)\mathbb{E}[L^2(t)Y(t+1)]+D_2(t,k)\mathbb{E}[L^2(t)Y^2(t+1)],
\end{aligned}
$$

where $D_n(t,k)$, $n=0,1,2$ are deterministic functions of t and the time lag k. The last equality follows from the definition of X. The expressions for $D_n(t,k)$ are given by

$$
\begin{aligned}
D_0(t,k):=&2\int_{t+k}^{t+k+1}\left[\mu^2(s-t-k)+\frac{1}{2}f_0^1(s-t-1)\right.\\
&\left.+\int_{t+k}^{s}\beta^2f_0^2(u-t-1)+(\beta^2f_0^1(s-u)-\mu\beta f_1^1(s-u)-\mu\beta)f_0^1(u-t-1)du\right]ds, \qquad (65)
\end{aligned}
$$

$$D_1(t,k) := 2\int_{t+k}^{t+k+1} \Big[\frac{1}{2} f_1^1(s-t-1) + \int_{t+k}^{s} \beta^2 f_1^2(u-t-1) + (\beta^2 f_0^1(s-u) - \mu\beta f_1^1(s-u) - \mu\beta) f_1^1(u-t-1)du\Big]ds, \tag{66}$$

$$D_2(t,k) := 2\int_{t+k}^{t+k+1}\int_{t+k}^{s} \beta^2 f_2^2(t+1,u) f_1^1(s-u)duds. \tag{67}$$

Since f_m^n are all finite, clearly $D_n(t,k)$ is well defined. The claim of the Lemma follows.

Proof of Lemma 4

For $n > 0$ we have

$$\begin{aligned}
&\mathbb{E}[L^2(t)Y^n(t+1)] = \mathbb{E}[X^2(s)Y^n(t+1)] \\
&= \mathbb{E}\left[Y^n(t+1)\left(\int_t^{t+1}(\mu-\beta Y(u))du + \int_t^{t+1}\sqrt{Y(u)}dB(u)\right)^2\right] \\
&= \mathbb{E}\left[Y^n(t+1)\left(\int_t^{t+1}(\mu-\beta Y(u))du\right)^2\right] \\
&\quad + 2\mathbb{E}\left[Y^n(t+1)\int_t^{t+1}(\mu-\beta Y(u))du\int_t^{t+1}\sqrt{Y(u)}dB(u)\right] \\
&\quad + \mathbb{E}\left[Y^n(t+1)\left(\int_t^{t+1}\sqrt{Y(u)}dB(u)\right)^2\right] \\
&= 2\int_t^{t+1}\int_t^{u}\mathbb{E}\left[(\mu-\beta Y(u))(\mu-\beta Y(v))\mathbb{E}\left[Y^n(t+1)|\,\mathscr{F}_u^Y\right]\right]dvdu \\
&\quad + 2\mathbb{E}\left[Y^n(t+1)\int_t^{t+1}(\mu-\beta Y(u))du\mathbb{E}\left[\int_t^{t+1}\sqrt{Y(u)}dB(u)|\mathscr{F}_{t+1}^Y\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[Y^n(t+1)\left(\int_t^{t+1}\sqrt{Y(u)}dB(u)\right)^2|\mathscr{F}_{t+1}^Y\right]\right] \\
&= 2\int_t^{t+1}\int_t^{u}\mathbb{E}\left[(\mu-\beta Y(u))(\mu-\beta Y(v))\mathbb{E}\left[Y^n(t+1)|\,\mathscr{F}_u^Y\right]\right]dvdu \\
&\quad + \int_t^{t+1}\mathbb{E}\left[Y(u)\mathbb{E}\left[Y^n(t+1)|\mathscr{F}_u^Y\right]\right]du.
\end{aligned}$$

Applying Lemma 2 we get

$$\begin{aligned}
\mathbb{E}[L^2(t)Y^n(t+1)] &= 2\int_t^{t+1}\int_t^{u}\mathbb{E}\left[(\mu^2-\mu\beta(Y(v)+Y(u))+\beta^2Y(u)Y(v))F_{t+1-u}^n(Y(u))\right]dvdu \\
&\quad + \int_t^{t+1}\mathbb{E}\left[Y(u)F_{t+1-u}^n(Y(u))\right]du \\
&= 2\int_t^{t+1}\int_t^{u}\Big[\mu^2\mathbb{E}\left[F_{t+1-u}^n(Y(u))\right] - \mu\beta\mathbb{E}\left[(Y(v)+Y(u))F_{t+1-u}^n(Y(u))\right] \\
&\quad + \beta^2\mathbb{E}\left[Y(u)Y(v)F_{t+1-u}^n(Y(u))\right]\Big]dvdu + \int_t^{t+1}\mathbb{E}\left[Y(u)F_{t+1-u}^n(Y(u))\right]du.
\end{aligned}$$

By Lemma 2, $\mathbb{E}[L^2(t)Y^n(t+1)]$ is finite.

References

1. Ole E. Barndorff-Nielsen. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13, 1997.
2. Ole E. Barndorff-Nielsen and Neil Shephard. Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(2):pp. 167–241, 2001.
3. Fred E. Benth. The stochastic volatility model of barndorff-nielsen and shephard in commodity markets. *Mathematical Finance*, 21(4):595–625, 2011.
4. Fred E. Benth, Jūratė Šaltytė Benth, and Steen Koekebakker. *Stochastic modelling of electricity and related markets*. World Scientific, Singapore, 2008.
5. Eva Benz and Stefan Trück. Modeling the price dynamics of {CO2} emission allowances. *Energy Economics*, 31(1):4–15, 2009.
6. Markus Burger, Bernhard Graeber, and Gero Schindlmayr. *Managing energy risk*. John Wiley & Sons, Ltd, Chichester, UK, 2014.
7. Les Clewlow and Chris Strickland. *Energy derivatives: pricing and risk management*. Lacima Publications, London UK, 2000.
8. Rama Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In Gilles Teyssire and AlanP. Kirman, editors, *Long memory in economics*, pages 289–309. Springer Berlin Heidelberg, 2007.
9. John C Cox, Jr Ingersoll, Jonathan E, and Stephen A Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, March 1985.
10. George Daskalakis, Dimitris Psychoyios, and Raphael N. Markellos. Modeling {CO2} emission allowance prices and derivatives: evidence from the european trading scheme. *Journal of Banking & Finance*, 33(7):1230 – 1241, 2009.
11. Avinash K. Dixit and Robert S. Pindyck. *Investment under uncertainty*. Princeton University Press, Princeton NJ, 1994.
12. Adrian A Drăgulescu and Victor M Yakovenko. Probability distribution of returns in the heston model with stochastic volatility. *Quantitative Finance*, 2(6):443–453, 2002.
13. Alexander Eydeland and Krzysztof Wolyniec. *Energy and power risk management: new developments in modeling, pricing, and hedging*. John Wiley & Sons, Inc., Hoboken NJ, 2003.
14. Zhen-Hua Feng, Le-Le Zou, and Yi-Ming Wei. Carbon price volatility: evidence from EU ETS. CEEP-BIT Working Papers 4, Center for Energy and Environmental Policy Research (CEEP), Beijing Institute of Technology, 2009.
15. Hélyette Geman. *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals and energy*. John Wiley & Sons, Inc., Chichester, UK, 2005.
16. Steven L Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–43, 1993.
17. Dilip B. Madan, Peter P. Carr, and Eric C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2(1):79–105, 1998.
18. Dilip B. Madan and Eugene Seneta. The variance gamma (v.g.) model for share market returns. *The Journal of Business*, 63(4):pp. 511–524, 1990.
19. Marc S. Paolella and Luca Taschini. An econometric analysis of emission allowance prices. *Journal of Banking & Finance*, 32(10):2022 – 2032, 2008.
20. Nicola Secomandi and Duane J. Seppi. Real options and merchant operations of energy and other commodities. *Foundations and Trends in Technology, Information and Operations Management*, 6(34):161–331, 2012.
21. Jan Seifert, Marliese Uhrig-Homburg, and Michael Wagner. Dynamic behavior of {CO2} spot prices. *Journal of Environmental Economics and Management*, 56(2):180 – 194, 2008.
22. Eugene Seneta. Fitting the variance-gamma model to financial data. *J. Appl. Probab.*, 41A:177–187, 02 2004.
23. Lenos Trigeorgis. *Real options: managerial flexibility and strategy in resource allocation*. The MIT Press, Cambridge, MA, 1996.
24. Shibin Zhang and Xinsheng Zhang. Exact simulation of ig-ou processes. *Methodology and Computing in Applied Probability*, 10(3):337–355, 2008.