Parameter Estimation in a Stochastic Drift Energy Model with a Price Cap

Janko Hernández * David Saunders † Luis. A Seco ‡

July 13, 2007

Abstract

We consider a class of models which share with the Pilipović's model of electricity prices the property of having a hidden stochastic drift. We show that under certain assumptions, the model parameters can be estimated using the method of moments. The ergodic properties of the model allow us to introduce an important condition derived from actual market operations: price caps.

Keywords: energy prices, price cap, mean reversion, stochastic drift, Hidden Markov Models, ergodicity, mixing, method of moments

1 Introduction

Energy prices have proved to be very difficult to model. There are several reason for this. For example, observed price series not only have high variations within the same day, but also show high regional variations within relatively small geographical areas. Prices also exhibit extreme spikes, which are not consistent with the usual modeling via diffusion processes. Another of the problem is that energy prices exhibit significantly greater volatility than in other markets, such as those for stocks or bonds. It is even possible for prices to be zero or negative at times, although this is very rare.

One important characteristic of energy prices is their tendency to revert to a mean level within a time scale of days, or at most weeks. In 1998 Pilipović introduced a model that accounts for at least some of the economics of electricity pricing (see [Pil98]). It is written as follows

$$\left.\begin{aligned} dS_t =& \rho(L_t - S_t)\,dt + \sigma S_t\,dB_t \\ dL_t =& \mu L_t\,dt + \nu L_t\,dW_t \end{aligned}\right\}.$$

*Professor at the Department of Business Administration, ITAM, Mexico

†Assistant Professor at the Department of Statistics and Actuarial Science at the University of Waterloo, Canada

‡RiskLab, University of Toronto, Canada

Here, S_t refers to the spot price, whereas L_t is an unobserved variable which represents a long-term (stochastic) equilibrium price. The B_t, W_t terms denote independent Brownian motions. The equilibrium price L_t satisfies a geometric Brownian motion (exponential growth subject to noise), with μ its growth rate and ν its volatility. The process S_t *reverts* to the level L_t at the rate ρ, with volatility σ.

Other diffusion-based models for electricity prices include those by Lucia and Schwartz (2002) (see [LS02]) and Barlow (2002) (see [Bar02]). See also [LLSW] where the authors review one-factor, two-factors and three-factors models.

Another behaviour exhibited by some energy markets is the existence of *price caps*. These are limits on prices introduced by electricity market regulators, which modify the "real" price arising from the clearing process to a maximum allowed price, whenever the clearing price is higher than the allowed maximum.

Using Pilipović's model as an example, what this means is that instead of observing S_t, what we really observe is $\min(S_t, S_{\max})$. S_{max} is the cap imposed by regulatory authorities. The problem in this case is that the mean price grows exponentially fast. Therefore, as time evolves, the cap is in effect for longer periods of time. In other words, Pilipović's model is not consistent with the existence of a price cap. For such a model, the cap should be expressed in terms of quantiles of the price distribution, as opposed to being fixed.

It is not our main goal to propose a new model that improves the existing ones, but instead to study the ergodic properties of a model which is consistent with the existence of a price cap in the market. In in this paper we consider a class of models which share with Pilipović's model the property of having a hidden stochastic equilibrium state. Under certain assumptions, the model is stationary, and its parameters can be estimated using the method of moments. The ergodic properties of that model allow us to introduce a price cap.

More precisely, we consider the stochastic processes introduced in [Sau01] and defined as the solution to the following equations:

$$\left.\begin{aligned} dY_t =& \rho(V_t - Y_t)dt + \sigma\, dB_t \\ dV_t =& b(V_t;\theta)dt + a(V_t;\theta)\, dW_t \\ X_t =& \min(Y_t, M) \end{aligned}\right\},$$

where $\rho, \sigma > 0$, θ is a vector of parameters of the non-observable process, B and W are a two independent standard Brownian motions, and M is a known constant. We interpret Y_t and M as the logarithm of the spot price and the cap respectively (that is, $Y_t = \ln S_t$ and $M = \ln S_{\max}$). As in Pilipović's model, V_t represents a stochastic equilibrium (log)-price.

We want to point out that Pilipović's model and the model proposed here should be fairly similar over not too long periods of time. For long periods of time, the behavior of both models is very different. This is due to the fact mentioned before that Pilipović's model grows in time, while one important feature of our model is its stationarity. Pilipović's model is an inflationary model, better suited for modeling prices quoted in nominal terms. Our model is more appropriate for prices already corrected for inflation.

Another observation is that, in contrast to what occurs with the prices of natural gas, the existence of a cap should nor affect the model for the log-price Y. The reason for this is that electricity prices are set by an auction mechanism. The quoted price depends on the demand at any time, and the maximum amongst the costs given by all the suppliers: hydro plants, nuclear plants, natural gas plants, etc. Finally, through a rebate process, the effective price is a function of the mean generation cost, instead of the maximum. This mechanism makes the behavior of the "uncapped price" independent of the value of the cap.

The main goal of this paper is to use the Ergodic Theorem to justify the Method of Moments for the estimation of the parameters. The general idea is to assume stationary and ergodic properties for the hidden process V, and from those to induce the ergodic properties of Y first, and of the observable process X afterward. Once we have established the ergodicity of Y and X, the next practical problem for estimation purposes is to compute explicitly their moments.

Since V is a one-dimensional diffusion, its ergodic properties have been well studied. Moreover, for some particular choices of its drift and volatility, the moments can be found explicitly. It is easy to show that X is ergodic if Y is. Hence, the difficult problems are to obtain the ergodicity of Y from the ergodicity of V, and to compute the moments of Y once we know the moments of V. We were able to solve both problems by a discretization, followed by suitable passages to the limit.

In [GCJL00] the ergodic properties of models of the form:

$$\left.\begin{aligned} dY_t =& \mu(\sigma_t^2)\,dt + \sigma_t\,dB_t \\ d(\sigma_t^2) =& b(\sigma_t^2)\,dt + a(\sigma_t^2)\,dW_t \end{aligned}\right\}$$

are studied. Here μ, a and b are real functions satisfying some technical conditions. For example, the choice $\mu(v) = \left(m - \frac{v}{2}\right)$, $a(v) = \nu\sqrt{v}$, and $b(v) = \alpha(\beta - v)$, is the celebrated Heston model (see [Hes93]). Defining a convenient Hidden Markov Model (see Appendix A), the authors show that if σ^2 is ergodic, then the difference process $Z_n = Y_{(n+1)h} - Y_{nh}$ is ergodic for a discretization of any step size h.

In [Sau01] the same model we are considering here was introduced, except for the price cap. Using similar ideas as in [GCJL00], the author also showed that the discrete process

$$Z_n = Y_{(n+1)h} - e^{-\rho h} Y_{nh}$$

is ergodic for any step size h. Our idea was then to take the limit as $h \to 0$ to obtain the ergodicity of the continuous-time process Y. A very similar passage to the limit has been done before in [GIY04] for regime switching models, which are slightly simpler.

Finally, we realized that taking the limit $h \to \infty$ instead, the exact moments of Y can be obtained from the exact moments of Z, which in turn can be computed from the moments of V. We would like to point out that taking the limit $h \to \infty$ is a rather unusual direction to take, but it yields useful results.

In Section 2 we define the model without the cap, and describe all the technical conditions we are going to assume through the paper. In Section 3 we establish the ergodicity not only of Z, but also of any discretization $(Y_{nh})_{n\in\mathbb{N}}$ of Y. In that same section we also obtain some explicit bounds for the moments that will be needed later. In Section 4 we make $h \to 0$ and obtain the ergodicity of the continuous-time process Y. In Section 5 we show how making $h \to \infty$ we can obtain the moments and auto-covariance of Y if we know those for Z. In Section 6 we introduce some explicit forms for the drift and volatility of V and compute the moments of Y explicitly in those examples. In Section 7 we introduce the price cap and explain how to use the Method of Moments to construct estimators. In Section 8 we give some numerical examples through simulations. Finally, in Appendix A we recall the definitions and properties of mixing coefficients and Hidden Markov Models (HMM).

2 A Stochastic Drift Model

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}^+}, \mathbb{P})$ be a stochastic basis satisfying the usual hypotheses that supports a standard two-dimensional Brownian motion $(B_t, W_t)_{t\in\mathbb{R}^+}$.

Let $\rho, \sigma > 0$ be two constants, and $\theta \in \Theta \subset \mathbb{R}^p$ be a vector of real parameters, for some $p \in \mathbb{R}$. Consider also an interval (l, r) with $-\infty \leq l < r \leq \infty$, and two functions $a, b : (l, r) \times \Theta \to \mathbb{R}$ satisfying the following assumption

Assumption 1. *For every $\theta \in \Theta$ fixed, the functions $a(\cdot;\theta)$ and $b(\cdot;\theta)$ are twice continuously differentiable, and there exist constants $q \geq \frac{1}{2}$, $K_\theta > 1$ such that:*

$$
\begin{aligned}
&|a(u;\theta) - a(v;\theta)| \leq K_\theta |u - v|^q \qquad \forall u, v \in (l, r),\\
&|b(u;\theta) - b(v;\theta)| \leq K_\theta |u - v| \qquad \forall u, v \in (l, r),\\
\textit{and} \qquad &a^2(u;\theta) + b^2(u;\theta) \leq K_\theta(1 + u^2) \qquad \forall u \in (l, r).
\end{aligned}
$$

Given two $\mathcal{F}_0$-measurable random variables Y_0 and V_0, both independent of (B, W), define the process $(Y, V) = (Y_t, V_t)_{t\in\mathbb{R}^+}$ as the solution of the stochastic differential equation

$$
\left.\begin{aligned}
dY_t &= \rho(V_t - Y_t)dt + \sigma\, dB_t\\
dV_t &= b(V_t;\theta)dt + a(V_t;\theta)\, dW_t
\end{aligned}\right\} . \tag{1}
$$

Assumption 1 ensures the existence and uniqueness of such a solution. Notice that for the volatility, we need the Hölder condition instead of the Lipschitz condition. This will allow us to consider cases as the CIR-drift model (see Section 6) via the Yamada-Watanabe Theorem. For all the above mentioned existence and uniqueness results we refer to [RW00].

Furthermore, we will make the following two assumptions.

Assumption 2. *For a fixed $v_0 \in (l, r)$, consider the function*

$$s(v) = \exp\left(-2 \int_{v_0}^{v} \frac{b(u)}{a^2(u)}\, du \right)$$

defined for all $v \in \mathbb{R}$. We assume that $\int^r s(u)\, du = \int_l s(u)\, du = \infty$, and that $M = \int_l^r \frac{du}{a^2(u)s(u)} < \infty$.

Assumption 3. *V_0 has distribution $\tilde{\pi}(u)\, du$, where $\tilde{\pi}(u) = \frac{1}{Ma^2(u)s(u)} \mathbf{1}_{(l,r)}(u)$.*

The following result can be found in [GCJL00].

Proposition 2.1. *Under Assumptions 1-3, the process V is strictly stationary and time reversible. Furthermore, the continuous-time process $(V_t)_{t \in \mathbb{R}^+}$ and any of its discrete-time samplings $(V_{nh})_{n \in \mathbb{N}}$ are β-mixing, and hence also α-mixing and ergodic.*

We need to make another two assumptions in order to obtain the ergodicity of the process Y.

Assumption 4. *For some $p > 1$ we have that $\mathbb{E}|V_0|^p < \infty$.*

We will see in Section 4 that the previous four assumptions implies the existence of a unique stationary distribution for the process Y. With that in mind, our last assumption makes sense.

Assumption 5. *Y_0 follows the unique stationary distribution implied by the model (1) under Assumptions 1-4*

We will see in Theorem 4.2, the main result of this paper, that under Assumptions 1-5 the observable process Y is strictly stationary and ergodic.

3 The discretized Model

If Assumptions 1-4 hold, then there exits a constant β such that $\mathbb{E}V_t = \beta$ for every $t \geq 0$. From now on we will denote:

$$y_t = Y_t - \beta, \qquad \text{and} \qquad v_t = V_t - \beta.$$

It is easy to verify that Assumptions 1-4 hold for the processes $V = (V_t)_{t \geq 0}$ if and only if they hold for the process $v = (v_t)_{t \geq 0}$.

Proposition 3.1. *Fix $h \geq 0$. For any $t \geq 0$ we can write:*

$$y_{t+h} = e^{-\rho h} y_t + Z_t^{(h)}, \tag{2}$$

$$\text{with} \qquad Z_t^{(h)} = \mu_t(h) + \Gamma(h) \xi_t^{(h)}, \tag{3}$$

where

$$\mu_t(h) = e^{-\rho h}\int_0^h \rho e^{\rho u} v_{t+u}\, du, \tag{4}$$

$$\Gamma^2(h) = \frac{\sigma^2}{2\rho}(1 - e^{-2\rho h}), \tag{5}$$

$$\text{and} \qquad \xi_t^{(h)} \sim N(0,1).$$

Moreover, if we define $\mathcal{G}_t = \sigma(V_s; s \le t) = \sigma(v_s; s \le t) \subset \mathcal{F}_t$*, then* $\mu_t(h)$ *is* $\mathcal{G}_{t+h}$*-measurable,* $\xi_t^{(h)}$ *is independent of* $\mathbb{G} = (\mathcal{G}_u)_{u\ge 0}$*, and* $\xi_t^{(h)}$ *is independent of* $\xi_s^{(h)}$ *whenever* $|t-s| \ge h$*. Notice also that* $\mathbb{E}Z_t^{(h)} = \mathbb{E}\mu_t(h) = 0$*.*

Proof. Applying Itô's Lemma to $e^{\rho t}(Y_t - \beta)$ we obtain

$$y_{t+h} = e^{-\rho h} y_t + \int_t^{t+h} e^{-\rho(t+h-s)}[\rho(V_s - \beta)\, ds + \sigma\, dB_s].$$

Hence we have (2) where

$$Z_t^{(h)} = \int_t^{t+h} \rho e^{-\rho(t+h-s)} v_s\, ds + \int_t^{t+h} \sigma e^{-\rho(t+h-s)}\, dB_s.$$

The first integral is just (4) after a change of variable. The second integral is a Gaussian random variable with variance

$$\int_t^{t+h} \sigma^2 e^{-2\rho(t+h-s)}\, ds = \Gamma^2(h).$$

so it can be written as $\Gamma(h)\xi_t^{(h)}$ with $\xi_t^{(h)} \sim N(0,1)$. Notice that since the Brownian B is independent of V, and hence of $\mathbb{G}$, we have that $\xi_t^{(h)}$ is independent of $\mathbb{G}$. Also, since B has independent increments, two integrals with respect to dB are independents if the integration intervals do not overlap. Therefore $\xi_t^{(h)}$ is independent of $\xi_s^{(h)}$ whenever $|t-s| \ge h$.

Finally, $\mathbb{E}\mu_t(h) = 0$ since $\mathbb{E}v_t = \mathbb{E}(V_t - \beta) = 0$ for every t. □

If for a fixed $h > 0$ we define the two discrete-time processes $y^{(h)} = (y_{nh})_{n\in\mathbb{N}}$ and $z^{(h)} = (Z_{nh}^{(h)})_{n\in\mathbb{N}}$, then (2) can be rewritten as

$$y_{n+1}^{(h)} = e^{-\rho h} y_n^{(h)} + z_n^{(h)} \qquad n \in \mathbb{N}, \tag{6}$$

$$\text{with} \qquad z_n^{(h)} = \mu_{nh}(h) + \Gamma(h)\xi_n. \tag{7}$$

Here $(\xi_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of standard Gaussian random variables, each independent of the sequence $(\mu_{nh}(h))_{n\in\mathbb{N}}$.

We are interested in the ergodic properties of the discrete-time processes $y^{(h)}$ defined by this equation and the initial state $y_n^{(h)} = y_0$. The first step is to establish the ergodic properties of $z^{(h)}$. The following result is due to [Sau01]. We repeat the proof here for the sake of completeness, and also because our notation is very different.

Theorem 3.2. *The process $z^{(h)}$ is a HMM (see Appendix A) with hidden chain $U^{(h)} = (U_n^{(h)})_{n\in\mathbb{N}}$ which takes values on $\mathbb{R}^2$, defined by $U_n^{(h)} = (\mu_{nh}(h), V_{(n+1)h})$.*

Proof. Denote $U_n = U_n^{(h)}$ and $z_n = z_n^{(h)}$ to simplify the notation. Denote by $C_{\mathbb{R}}([0,h])$ the space of all the continuous functions $\xi : [0,h] \to \mathbb{R}$ endowed with the topology of uniform convergence. If we define the process $S = (S_n)_{n\in\mathbb{N}}$ with values on $C_{\mathbb{R}}([0,h])$ as $S_n = (V_{nh+u})_{u\in[0,h]}$, and the function $F : C_{\mathbb{R}}([0,h]) \to \mathbb{R}^2$

$$F(\xi) = \Big(e^{-\rho h}\int_0^h \rho e^{\rho u}[\xi(u) - \beta]\,du, \xi(h)\Big),$$

then $U_n = F(S_n)$. The strict stationarity of V implies that of S, and since F is continuous, then U is also strictly stationary.

Next, for any bounded Borel-measurable function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ define the operator $T\varphi(x) = \mathbb{E}[\varphi(U_0)|V_0 = x]$. Using the strict stationarity and the Markov property for V we can verify that

$$\begin{aligned}
&\mathbb{E}[\varphi(U_n)|U_0,\dots,U_{n-1}] = \mathbb{E}\Big[\mathbb{E}[\varphi(U_n)|\mathcal{G}_{nh}]\Big|U_0,\dots,U_{n-1}\Big]\\
=&\mathbb{E}\Big[\mathbb{E}[\varphi(U_n)|V_{nh}]\Big|U_0,\dots,U_{n-1}\Big] = \mathbb{E}[T\varphi(V_{nh})|U_0,\dots,U_{n-1}]\\
=&\mathbb{E}[T\varphi(V_{nh})|U_{n-1}] = T\varphi(V_{nh}).
\end{aligned}$$

This shows that

$$\mathbb{E}[\varphi(U_n)|U_0,\dots,U_{n-1}] = \mathbb{E}[\varphi(U_n)|U_{n-1}] = T\varphi(V_{nh}), \tag{8}$$

and therefore U is a Markov process.

Finally, using Proposition 3.1 we have that for any real numbers $c_0,\dots,c_n$

$$\begin{aligned}
&\mathbb{E}\Big[\exp\big(\sum_{k=0}^n ic_k z_k\big)\Big|U_0,\dots,U_n\Big] = \mathbb{E}\Big[\mathbb{E}\big[\exp\big(\sum_{k=0}^n ic_k z_k\big)\big|\mathcal{G}_{(n+1)h}\big]\Big|U_0,\dots,U_n\Big]\\
=&\mathbb{E}\Big[\prod_{k=0}^n \exp\big(ic_k\mu_{kh}(h) - \frac{1}{2}c_k^2\Gamma^2(h)\big)\Big|U_0,\dots,U_n\Big]\\
=&\prod_{k=0}^n \exp\big(ic_k\mu_{kh}(h) - \frac{1}{2}c_k^2\Gamma^2(h)\big).
\end{aligned}$$

This provides all the conditions needed by the definition of an HMM. □

Lemma 3.3. *We have that $c_{U^{(h)}}(n) \le c_V((n-1)h)$ for $c = \alpha$, β or ρ. Hence, if $(V_{nh})_{n\in\mathbb{N}}$ is c-mixing, then $U^{(h)}$ is also c-mixing.*

Proof. We are going to exploit the fact that, from the previous proof, we have $U_n^{(h)} = F(S_n)$, where $S_n = (V_{nh+u})_{u\in[0,h]}$ and F is some continuous function. Then

$$\begin{aligned}
c_{U^{(h)}}(n) &= c\big(\sigma(U_0^{(h)}), \sigma(U_n^{(h)})\big) = c\big(\sigma(F(S_0)), \sigma(F(S_n))\big)\\
&\le c\big(\sigma(V_s; s\le h), \sigma(V_s; s\ge nh)\big) = c_V((n-1)h),
\end{aligned}$$

and the proof is complete. □

By Proposition 2.1 and the previous lemma, $U^{(h)}$ is β-mixing, and hence α-mixing and ergodic. Finally, applying Proposition A.2 we get

Theorem 3.4. *$z^{(h)}$ is a strictly stationary α-mixing process, and hence ergodic. Moreover, if V is ρ-mixing, so is $z^{(h)}$.*

Once we have the strict stationarity and the ergodicity of $z^{(h)}$, the next result is only a rephrase of Theorem 1 in [Bra86b] with our present notation. We need first to extend by stationarity the process $\left(z_n^{(h)}\right)_{n\in\mathbb{N}}$ to $n \in \mathbb{Z}$.

Theorem 3.5. *Fix $h > 0$ and consider the equation*

$$x_{n+1} = e^{-\rho h} x_n + z_n^{(h)} \qquad n \in \mathbb{Z}. \tag{9}$$

If for any $h > 0$ the following condition holds

$$\mathbb{E} \max(0, \ln |z_0^{(h)}|) < \infty \tag{10}$$

then the only stationary solution of (9) is

$$x_n = \sum_{k=1}^{\infty} e^{-(k-1)\rho h} z_{n-k}^{(h)} \qquad n \in \mathbb{Z}, \tag{11}$$

where the sum on the right-hand side converges absolutely a.s. Furthermore, denote by $\pi^{(h)}$ the distribution law of this stationary solution, and let ξ be any $\mathcal{F}_0$-measurable random variable. Then the solution of (9) for $n \in \mathbb{N}$ and $y_0 = \xi$ satisfies

$$x_n \xrightarrow{\mathcal{L}} \pi^{(h)} \qquad \text{as } n \to \infty.$$

Comparing (9) with (6) we obtain as a corollary the most important result of this section.

Theorem 3.6. *If for any given $h > 0$ condition (10) holds, then there exist a unique distribution law $\pi^{(h)}$ such that for any $\mathcal{F}_0$-measurable random variable y_0, the discrete-time process $y^{(h)}$ defined by (6) satisfies:*

$$y_n^{(h)} \xrightarrow{\mathcal{L}} \pi^{(h)} \qquad \text{as } n \to \infty. \tag{12}$$

Moreover, if we take $y_0 \sim \pi^{(h)}$, then $y^{(h)}$ is strictly stationary.

The following proposition shows that taking into account Assumption 4, condition (10) automatically holds. It also gives us a uniform limit condition when $h \to 0$ that we will need in the next section.

Proposition 3.7. *If Assumptions 1-4 hold, then*

(i) *for any $h > 0$ condition (10) holds, and*

(ii) *for any $x > 0$ we have the following limit uniformly in $s \geq 0$:*

$$\lim_{h\to 0} \mathbb{P}[|Z_s^{(h)}| \geq x] = 0.$$

The proof will make use of the following lemma, that bounds the moments of $\mu_t(h)$ from the bounds of the moments of the process V.

Lemma 3.8. *Let $\mu_t(h)$ be defined as in (4). If Assumptions 1-4 hold, then for any $1 \le r \le p$ we have that*

$$\mathbb{E}|\mu_t(h)|^r \le (1-e^{-\rho h})^r \mathbb{E}|v_0|^r \le \mathbb{E}|v_0|^p < \infty \qquad \forall t, h \le 0.$$

Proof. Define the measure η on $\mathbb{R}$ as

$$\eta(du) = \mathbf{1}_{[0,h]}(u) \frac{\rho e^{-\rho(h-u)}}{1-e^{-\rho h}}\, du.$$

Then

$$\mu_t(h) = e^{-\rho h} \int_0^h \rho e^{\rho u} v_{t+u}\, du, = (1-e^{-\rho h}) \int_{\mathbb{R}} v_{t+u}\, \eta(du),$$

and clearly $\int_{\mathbb{R}} \eta(ds) = 1$. By Jensen's inequality

$$\Big| \int_{\mathbb{R}} v_{t+u}\, \eta(du) \Big|^r \le \int_{\mathbb{R}} |v_{t+u}|^r\, \eta(du).$$

The stationarity of v implies that $\mathbb{E}|v_s|^r = \mathbb{E}|v_0|^r$ for all $s \in \mathbb{R}^+$, hence

$$\begin{aligned}\mathbb{E}|\mu_t(h)|^r &= (1-e^{-\rho h})^r \mathbb{E}\Big| \int_{\mathbb{R}} v_{t+u}\, \eta(du) \Big|^r \le (1-e^{-\rho h})^r \int_0^h \mathbb{E}|v_{t+u}|^r\, \eta(du) \\ &= (1-e^{-\rho h})^r \mathbb{E}|v_0|^r \int_R \eta(du) = (1-e^{-\rho h})^r \mathbb{E}|v_0|^r \le \mathbb{E}|v_0|^p.\end{aligned}$$

Minkowski's inequality then finishes the proof:

$$(\mathbb{E}|v_0|^p)^{\frac{1}{p}} = (\mathbb{E}|V_0 - \beta|^p)^{\frac{1}{p}} \le (\mathbb{E}|V_0|^p)^{\frac{1}{p}} + |\beta| < \infty,$$

due to Assumption 4. □

Proof of Proposition 3.7. Using Minkowski's inequality and the previous Lemma

$$\begin{aligned}\big(\mathbb{E}|Z_t^{(h)}|^r\big)^{\frac{1}{r}} &\le \big(\mathbb{E}|\mu_t(h)|^r\big)^{\frac{1}{r}} + \big(\mathbb{E}|\Gamma(h)\xi_t^{(h)}|^r\big)^{\frac{1}{r}} \\ &\le (1-e^{-\rho h})\big(\mathbb{E}|v_0|^r\big)^{\frac{1}{r}} + \Gamma(h)\big(\mathbb{E}|\xi_t^{(h)}|^r\big)^{\frac{1}{r}}.\end{aligned}$$

Since $\lim_{h\to 0}\Gamma(h) = 0$, this clearly implies that $\lim_{h\to 0} \mathbb{E}|Z_s^{(h)}|^r = 0$ uniformly in $s \ge 0$. From Markov's inequality, condition (ii) follows immediately. Also

$$\mathbb{E}|Z_t^{(h)}|^r \le \Big[\big(\mathbb{E}|v_0|^r\big)^{\frac{1}{r}} + \frac{\sigma^2}{2\rho} K_r\Big]^r < \infty,$$

where $K_r = \big(\mathbb{E}|\xi_t^{(h)}|^r\big)^{\frac{1}{r}}$ is a constant. Since $\max(0, \ln|x|) \le |x|$, then

$$\mathbb{E}(\ln|z_0^{(h)}|)^+ \le \mathbb{E}|z_0^{(h)}| = \mathbb{E}|Z_0^{(h)}| < \infty.$$

and condition (i) follows. □

Corollary 3.9. *For any r such that $1 \leq r \leq p$ there exist a constant c_r that does not depend on h or t such that*

$$\mathbb{E}|Z_t^{(h)}|^r \leq c_r < \infty \qquad \forall t \in \mathbb{R}, h > 0.$$

4 Ergodicity of the continuous-time process Y

The idea now is to take the limit as $h \to 0$ in Theorem 3.6, with the goal of obtaining the ergodicity of the continuous-time version of the process Y, or equivalently, the process $y = Y - \beta$. The proof of the following Theorem is heavily inspired on the results found in [GIY04]. We only needed a few refinements in order to adapt it to our present case.

Theorem 4.1. *If Assumptions 1-4 hold, then there exists a unique probability measure π_0 in $\mathbb{R}$ such that if $y_0 \sim \pi_0$, then the continuous-time process $(y_t)_{t\in\mathbb{R}^+}$ and any of its discrete-time samplings $(y_{nh})_{n\in\mathbb{N}}$ are strictly stationary and ergodic.*

Proof. First notice that we can make use of (i) and (ii) from Proposition 3.7. Let y_0 be any $\mathcal{F}_0$-measurable random variable. Consider $m, m' \in \mathbb{N}$ such that $m \leq m'$. If $h = 2^{-m}$ and $h' = 2^{-m'}$, then the sequence $(y_{nh})_{n\in\mathbb{N}}$ is embedded in the sequence $(y_{nh'})_{n\in\mathbb{N}}$. Using (i) and Theorem 3.6 we have that the corresponding stationary laws have to be the same, that is, $\pi^{(h')} = \pi^{(h)}$. Therefore, we can denote by π_0 the common limit distribution for every h of the form $h = 2^{-m}$.

Fix $\varepsilon > 0$, and choose K_ε such that $\pi_0[|x| \geq K_\varepsilon] \leq \frac{\varepsilon}{2}$. Let $s \geq 0$ and $h > 0$. From (2) we have that

$$y_{s+h} - y_s = (e^{-\rho h} - 1)y_s + Z_s^{(h)}.$$

Then

$$\begin{aligned}\mathbb{P}[|y_{s+h} - y_s| \geq \varepsilon] \leq& \mathbb{P}\Big[\big|(e^{-\rho h} - 1)y_s\big| \geq \frac{\varepsilon}{2}\Big] + \mathbb{P}\Big[\big|Z_s^{(h)}\big| \geq \frac{\varepsilon}{2}\Big] \\ =& \mathbb{P}\Big[\big|y_s\big| \geq \frac{\varepsilon}{2(e^{-\rho h} - 1)}\Big] + \mathbb{P}\Big[\big|Z_s^{(h)}\big| \geq \frac{\varepsilon}{2}\Big]\end{aligned}$$

Using that $(e^{-\rho h} - 1) \to 0$ as $h \to 0$ and condition (ii), we can choose $\Delta = 2^{-m}$ independently of s such that

$$\mathbb{P}[|y_{s+h} - y_s| \geq \varepsilon] \leq \mathbb{P}\big[|y_s| \geq K_\varepsilon\big] + \frac{\varepsilon}{2} \qquad \forall h \leq \Delta. \tag{13}$$

Now, for any $t \geq 0$ denote by $s_\Delta(t)$ the largest multiple of Δ smaller than t. Then $s_\Delta(t) < t \leq s_\Delta(t) + \Delta$ and $h = t - s_\Delta(t) \leq \Delta$. By (13)

$$\mathbb{P}[|y_h - y_{s_\Delta(t)}| \geq \varepsilon] \leq \frac{\varepsilon}{2} + \mathbb{P}\big[|y_{s_\Delta(t)}| \geq K_\varepsilon\big].$$

Making $t \to \infty$ and using (12) we obtain

$$\limsup_{t\to\infty} \mathbb{P}[|y_t - y_{s_\Delta(t)}| \geq \varepsilon] \leq \frac{\varepsilon}{2} + \pi_0[|x| \geq K_\varepsilon] \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The next part of the proof follows exactly as in [GIY04]. Let $C(\pi_0)$ be the sets of continuity points of the distribution function $F_0(x) = \pi_0[(-\infty, x]]$. Let $x \in C(\pi_0)$ and choose $\varepsilon > 0$ such that $x \pm \varepsilon \in C(\pi_0)$. Since

$$\begin{aligned} \mathbb{P}[y_h \le x] \le& \mathbb{P}[y_{s_\Delta(t)} \le x + \varepsilon] + \mathbb{P}[|y_h - y_{s_\Delta(t)}| \ge \varepsilon], \\ \text{and} \qquad \mathbb{P}[y_{s_\Delta(t)} \le x - \varepsilon] \le& \mathbb{P}[y_h \le x] + \mathbb{P}[|y_h - y_{s_\Delta(t)}| \ge \varepsilon]. \end{aligned}$$

then

$$F_0(x-\varepsilon) - \varepsilon \le \liminf_{t\to\infty} \mathbb{P}[y_h \le x] \le \limsup_{t\to\infty} \mathbb{P}[y_h \le x] \le F_0(x+\varepsilon) + \varepsilon.$$

Letting $\varepsilon \to 0$ (with $x \pm \varepsilon \in C(\pi_0)$, which is possible since $C(\pi_0)$ is dense and the complement of a countable set), we obtain

$$\lim_{t\to\infty} \mathbb{P}[y_h \le x] = F_0(x) \qquad \forall x \in C(\pi_0).$$

Hence for any y_0 we have shown that $y_h \xrightarrow{\mathcal{L}} \pi_0$.

Finally, making $y_0 \sim \pi_0$ and using the strict stationarity of $(y_{nh})_{n\in\mathbb{N}}$ for every $h > 0$, it is easy to verify the strict stationarity of $(y_t)_{t\in\mathbb{R}^+}$. Since the stationary distribution is unique, the ergodic decomposition theorem by Krylov and Bogolioubov (see [KB37], and [Kal97] for a modern proof) implies the ergodicity of y. □

Combining the previous theorem with Assumption 5, we obtain the main result of this paper.

Theorem 4.2. *If Assumptions 1-5 hold, then both components V and Y of the model defined by model (1) are strictly stationary and ergodic.*

5 Exact moments of the stationary model

From this point on we are going to suppose without mention that all Assumptions 1-5 hold, and hence the observable component Y of the model is strictly stationary and ergodic.

For estimation purposes, we need to compute the moments of the observable component Y exactly. The following result allows us to do that if we know how to compute the moments of $z^{(h)} = (z_n^{(h)})_{n\in\mathbb{Z}} = (Z_{nh}^{(h)})_{n\in\mathbb{Z}}$ exactly for any $h > 0$. The trick is to take the limit as $h \to \infty$ and to exploit the stationarity of Y. Note also that stationarity implies that the relations hold for any $t \in \mathbb{R}^+$.

Theorem 5.1. *Assume that $1 \le r \le p$. Then $\mathbb{E}|Y_t|^r < \infty$ for all $t \in \mathbb{R}^+$. Moreover, if k is an integer such that $1 \le k \le p$, then*

$$\mathbb{E}y_t^k = \mathbb{E}(Y_t - \beta)^r = \lim_{h\to\infty} \mathbb{E}\big[z_0^{(h)}\big]^k. \tag{14}$$

Furthermore, if $p \ge 2$ then for any $h > 0$ we have that

$$\mathbb{E}[y_0 y_h] = \mathbb{E}[(Y_0 - \beta)(Y_h - \beta)] = e^{-\rho h}\mathbb{E}y_0^2 + \sum_{k=1}^{\infty} e^{-\rho(k-1)h}\mathbb{E}\big[z_0^{(h)} z_k^{(h)}\big]. \tag{15}$$

Proof. The first step is to make sure that if the p-moment of V exists (Assumption 4), then all the moments up to p of y also exist. Fix any $h > 0$ (for example, take $h = 1$). From (11) we have that $y_0 = \sum_{k=1}^{\infty} e^{-(k-1)\rho h} z_{-k}^{(h)}$. Using the triangular inequality and Corollary 3.9 we have that

$$\left(\mathbb{E}|y_0|^r\right)^{\frac{1}{r}} \leq \sum_{k=1}^{\infty} \left(\mathbb{E}|e^{-(k-1)\rho h} z_{-k}^{(h)}|^r\right)^{\frac{1}{r}} \leq \sum_{k=1}^{\infty} e^{-(k-1)\rho h} \left(c_r\right)^{\frac{1}{r}} < \infty.$$

The stationarity of y implies that $\mathbb{E}|y_t|^r = \mathbb{E}|y_0|^r < \infty$ for every $t \in \mathbb{R}^+$.

If k is an integer such that $1 \leq k \leq r$, then

$$\begin{aligned} &\mathbb{E}\big[z_0^{(h)}\big]^k = \mathbb{E}\big[y_h - e^{-\rho h} y_0\big]^k \\ =&\mathbb{E}y_h^k + \sum_{i=1}^{k-1} \binom{k}{i} \mathbb{E}\big[(y_h)^{k-i}(e^{-\rho h} y_0)^i\big] + \mathbb{E}\big[e^{-\rho h} y_0\big]^k. \end{aligned} \tag{16}$$

Fix $1 \leq i \leq k-1$, and define $p = \frac{k}{k-i}$ and $q = \frac{k}{i}$. Then by Hölder's inequality

$$\begin{aligned} \Big|\mathbb{E}\big[(y_h)^{k-i} e^{-\rho h} y_0)^i\big]\Big| \leq& \mathbb{E}\big[|y_h|^{k-i} |e^{-\rho h} y_0|^i\big] \leq \big[\mathbb{E}|y_h|^{(k-i)p}\big]^{\frac{1}{p}} \big[\mathbb{E}|e^{-\rho h} y_0|^{iq}\big]^{\frac{1}{q}} \\ =& e^{-\rho h \frac{k}{q}} \big[\mathbb{E}|y_0|^k\big]^{\frac{1}{p}} \big[\mathbb{E}|y_0|^k\big]^{\frac{1}{q}} = e^{-\rho h i} \mathbb{E}|y_0|^k \to 0 \end{aligned}$$

as $h \to \infty$. Finally, taking the limit in (16) we obtain (14).

For the auto-covariance, we can multiply the relation $y_h = e^{-\rho h} y_0 + z_0^{(h)}$ by y_0 and take expectations to get

$$\mathbb{E}[y_0 y_h] = e^{-\rho h} \mathbb{E} y_0^2 + \mathbb{E}\big[y_0 z_0^{(h)}\big].$$

From (11) we have that $y_0 = \sum_{k=1}^{\infty} e^{-(k-1)\rho h} z_{-k}^{(h)}$. Then

$$\mathbb{E}\big[y_0 z_0^{(h)}\big] = \sum_{k=1}^{\infty} e^{-\rho(k-1)h} \mathbb{E}\big[z_{-k}^{(h)} z_0^{(h)}\big].$$

By the strict stationarity of $z^{(h)}$ we have that $\mathbb{E}\big[z_{-k}^{(h)} z_0^{(h)}\big] = \mathbb{E}\big[z_0^{(h)} z_k^{(h)}\big]$. If $p \geq 2$, then Corollary 3.9 implies that

$$2\mathbb{E}\big|z_0^{(h)} z_k^{(h)}\big| \leq \mathbb{E}\big|z_0^{(h)}\big|^2 + \mathbb{E}\big|z_k^{(h)}\big|^2 \leq 2c_2,$$

and hence the series converges absolutely. □

6 Exact moments in some examples

In order to compute the first few moments of Y exactly, we need first to specify the drift a and volatility b of V in (1). From now on we will work with the following model:

$$\left.\begin{aligned} dY_t =& \rho(V_t - Y_t)dt + \sigma dB_t \\ dV_t =& \alpha(\beta - V_t)dt + \nu V_t^{\lambda} dW_t \end{aligned}\right\} \tag{17}$$

where $\alpha, \beta, \nu > 0$. In what follows we will denote $\kappa = \frac{\nu^2}{2\alpha}$.

Specifically, we consider the cases $\lambda = 0, \frac{1}{2}, 1$.

When $\lambda = 0$ the V process in (17) is the Ornstein-Uhlenbeck process. We shall refer to it as the *OU-drift model*. The OU-drift model satisfies Assumptions 1-3 with $(l, r) = (-\infty, +\infty)$, and in that case, the stationary distribution for V is Gaussian with parameters (β, κ).

We refer to the case $\lambda = \frac{1}{2}$ as the *CIR-drift model* (after [JCR85]). The CIR-drift model satisfies Assumptions 1-3 with $(l, r) = (0, +\infty)$ if $\beta \geq \kappa$. The stationary distribution for V is a Gamma with parameters $(\frac{\beta}{\kappa}, \frac{1}{\kappa})$, which has finite moments of any order.

The case $\lambda = 1$ is the *GARCH-drift model* (as in this case V is the diffusion approximation of a GARCH process, see [Nel90]). The GARCH-drift model satisfies Assumptions 1-3 with $(l, r) = (0, +\infty)$, and the stationary distribution for V is an Inverse Gamma with parameters $(1 + \frac{1}{\kappa}, \frac{\beta}{\kappa})$. The moments of order p are finite if $p < 1 + \frac{1}{\kappa}$.

For the details of the distributions of the CIR-drift and GARCH-drift processes, see [GCJL00] (pp. 1072-1073). In Appendix B we give explicit expressions for the moments and auto-covariances of the hidden process V.

It is clear then that once the stationarity Assumption 3 is satisfied, then also Assumption 4 holds for any $p > 1$ in the cases $\lambda = 0, \frac{1}{2}$. In the case $\lambda = 1$ we have that Assumption 4 holds for $p = 3$ if we assume that $\kappa < \frac{1}{2}$. Proposition 3.7 and Theorem 4.1 then allows us to also have Assumption 5. From now on, we are going to suppose that Assumptions 1-5 are all satisfied.

In Lemma B.2 it is shown that $\mathbb{E}V_t = \beta$, hence we can keep the notation from Section 3

$$y_t = Y_t - \beta, \qquad \text{and} \qquad v_t = V_t - \beta.$$

Theorem 6.1. *If $\alpha \neq \rho$, then moments of y are given by the following formulas:*

$$\begin{aligned}
\mathbb{E}y_t =& 0, \\
\mathbb{E}y_t^2 =& \frac{\sigma^2}{2\rho} + \frac{M_2}{1 + \frac{\alpha}{\rho}}, \\
\mathbb{E}y_t^3 =& \frac{M_3}{\left(1 + \frac{\alpha}{\rho}\right)\left(1 + \frac{\alpha}{2\rho}\right)}, \\
\textit{and} \qquad \mathbb{E}[y_0 y_h] =& \left(\frac{\sigma^2}{2\rho} + \frac{M_2}{1 + \frac{\alpha}{\rho}}\right) e^{-\rho h} + \frac{M_2}{1 - (\frac{\alpha}{\rho})^2}(e^{-\alpha h} - e^{-\rho h}), \qquad \forall h > 0.
\end{aligned}$$

Proof. In view of Theorem 5.1 we first need to compute the moments of $z^{(h)}$. By Proposition (7) $z_t^{(h)}$ is conditionally Gaussian, hence

$$\mathbb{E}\big[z_0^{(h)}\big] = \mathbb{E}[\mu_0(h)], \tag{18}$$

$$\mathbb{E}\big[z_0^{(h)}\big]^2 = \mathbb{E}[\mu_0(h)]^2 + \Gamma^2(h), \tag{19}$$

$$\text{and} \qquad \mathbb{E}\big[z_0^{(h)}\big]^3 = \mathbb{E}[\mu_0(h)]^3 + 3\mathbb{E}[\mu_0(h)]\Gamma^2(h). \tag{20}$$

Denote $i_m = \lim_{h\to\infty} \mathbb{E}[\mu_0(h)]^m$ for $m \in \mathbb{N}$. Clearly $i_0 = 1$. For $m \geq 1$ we can use (4) to get

$$i_m = \lim_{h\to\infty} \mathbb{E}\Big(e^{-\rho h} \int_0^h \rho e^{\rho s} v_s \, ds\Big)^m.$$

Using the following identity

$$\Big(\int_0^h f(s)\, ds\Big)^m = m! \int_0^h \int_0^{s_m} \cdots \int_0^{s_2} f(s_1)\cdots f(s_m)\, ds_1 \cdots ds_m,$$

which can be easily proved by induction, yields

$$i_m = m!\rho^m \lim_{h\to\infty} e^{-m\rho h} \int_0^h \int_0^{s_m} \cdots \int_0^{s_2} e^{\rho(s_1+\cdots+s_m)} \mathbb{E}[v_{s_1}\cdots v_{s_m}]\, ds_1 \cdots ds_m.$$

In particular, we are now going to show that

$$i_0 = 1, \qquad i_1 = 0, \qquad i_2 = \frac{M_2}{\big(1+\frac{\alpha}{\rho}\big)}, \qquad \text{and} \qquad i_3 = \frac{M_3}{\big(1+\frac{\alpha}{\rho}\big)\big(1+\frac{\alpha}{2\rho}\big)},$$

where M_2 and M_3 are as in Lemma B.2. Making use of Lemmas B.2 and B.4 we have that

$$i_1 = \rho \lim_{h\to\infty} e^{-\rho h} \int_0^h e^{\rho s_1} \mathbb{E}[v_{s_1}]\, ds_1 = 0$$

since $\mathbb{E}[v_{s_1}] = M_1 = 0$. For $m = 2$ we have

$$\begin{aligned} i_2 &= 2!\rho^2 \lim_{h\to\infty} e^{-2\rho h} \int_0^h \int_0^{s_2} e^{\rho(s_1+s_2)} \mathbb{E}[v_{s_1} v_{s_2}]\, ds_1 ds_2 \\ &= 2\rho^2 M_2 \lim_{h\to\infty} e^{-2\rho h} \int_0^h \int_0^{s_2} e^{\rho(s_1+s_2)} e^{-\alpha(s_2-s_1)}\, ds_1 ds_2 \\ &= \frac{2\rho^2 M_2}{(\rho+\alpha)2\rho}. \end{aligned}$$

And for $m = 3$ we have

$$\begin{aligned} i_3 &= 3!\rho^3 \lim_{h\to\infty} e^{-3\rho h} \int_0^h \int_0^{s_3} \int_0^{s_2} e^{\rho(s_1+s_2+s_3)} \mathbb{E}[v_{s_1} v_{s_2} v_{s_3}]\, ds_1 ds_2 ds_3 \\ &= 6\rho^3 M_3 \lim_{h\to\infty} e^{-3\rho h} \int_0^h \int_0^{s_3} \int_0^{s_2} e^{\rho(s_1+s_2+s_3)} e^{-\alpha(s_3-s_1)}\, ds_1 ds_2 ds_3 \\ &= \frac{6\rho^3 M_3}{(\rho+\alpha)(2\rho+\alpha)3\rho}. \end{aligned}$$

A little algebra yields the desired expressions. Then, using Theorem 5.1, equations (18)-(20), the fact that $\lim_{h\to\infty} \Gamma^2(h) = \frac{\sigma^2}{2\rho}$, and substituting the expressions for i_2 and i_3 we obtain the first three moments of y.

For the auto-correlation, we are going to use (15). Let $k \geq 1$, then by (7) we have that $\mathbb{E}\big[z_0^{(h)} z_k^{(h)}\big] = \mathbb{E}[\mu_0(h)\mu_{kh}(h)]$. Then, by (4) and Lemma B.4 we have that

$$\begin{aligned}
\mathbb{E}[\mu_0(h)\mu_{kh}(h)] =& \mathbb{E}\Big[\Big(e^{-\rho h}\int_0^h \rho e^{\rho s} v_s\, ds\Big) \times \Big(e^{-\rho h}\int_0^h \rho e^{\rho s} v_{kh+s}\, ds\Big)\Big] \\
=& e^{-2\rho h}\rho^2 \int_0^h \int_0^h e^{\rho(s_1+s_2)} \mathbb{E}[v_{s_1} v_{kh+s_2}]\, ds_1 ds_2 \\
=& e^{-2\rho h}\rho^2 \int_0^h \int_0^h e^{\rho(s_1+s_2)} e^{-\alpha(kh+s_2-s_1)} M_2\, ds_1 ds_2 \\
=& e^{-(2\rho h+\alpha k h)}\rho^2 M_2 \int_0^h e^{(\rho+\alpha)s_1}\, ds_1 \times \int_0^h e^{(\rho-\alpha)s_2}\, ds_2 \\
=& \frac{e^{-(2\rho h+\alpha k h)}\rho^2 M_2}{\rho^2-\alpha^2}(e^{(\rho+\alpha)h}-1)(e^{(\rho-\alpha)h}-1),
\end{aligned}$$

Then we can write $\mathbb{E}\big[z_0^{(h)} z_{kh}^{(h)}\big] = \mathbb{E}[\mu_0(h)\mu_{kh}(h)] = n(h)e^{-k\alpha h}$ where

$$n(h) = \frac{M_2}{1-\frac{\alpha^2}{\rho^2}}(1-e^{-(\rho+\alpha)h})(1-e^{-(\rho-\alpha)h}).$$

Finally, by (15) we have that

$$\begin{aligned}
\mathbb{E}[y_0 y_h] &= e^{-\rho h}\mathbb{E}y_0^2 + \sum_{k=1}^{\infty} e^{-\rho(k-1)h}[m(h) + n(h)e^{-k\alpha h}] \\
&= e^{-\rho h}\mathbb{E}Y_0^2 + n(h)\frac{e^{-\alpha h}}{1-e^{-(\rho+\alpha)h}}.
\end{aligned}$$

Substituting the expressions of $\mathbb{E}y_0^2$ and $n(h)$ yields the desired result. □

7 Price Cap and Estimation

Recall that Y_t in the model (1) represents the logarithm of the electricity price at time t. When a price cap of $S_{\max}$ is introduced, instead of Y_t we observe $X_t = \min(Y_t, M)$, where $M = \ln S_{\max}$ is a known constant (see Figure 1 pp. 19).

Since $f_M(y) = \min(y, M)$ is a continuous function, the ergodicity of Y trivially implies the ergodicity of X. Hence, the following Theorem is really a corollary of Theorem 4.2.

Theorem 7.1. *Assume that the following model*

$$\left.\begin{aligned}
dY_t =& \rho(V_t - Y_t)dt + \sigma dB_t \\
dV_t =& b(V_t;\theta)dt + a(V_t;\theta)dW_t \\
X_t =& \min(Y_t, M)
\end{aligned}\right\} \tag{21}$$

satisfies Assumptions 1-5. Then X is a strictly stationary and ergodic process. Moreover, for any $h > 0$ the discrete-time sampling $X^{(h)} = (X_{nh})_{n\in\mathbb{N}}$ is strictly stationary and ergodic.

Remark 7.2. *If we allow M to take the value $+\infty$, then $X_t = \min(Y_t, +\infty) = Y_t$. Therefore, we can see Theorem 4.2 as a particular case of Theorem 7.1.*

Assume that in the model (21) we observe a time discretization of (X_t). That is, for some fixed and known $\Delta > 0$, we observe $(X_{n\Delta})_{n\in\mathbb{N}}$. In the particular case $M = +\infty$, we observe a time discretization of Y. Our goal is to estimate the vector of parameters (ρ, σ, θ) from those observations. Birkhoff's Ergodic Theorem (see for example [Kre85]) implies the following result.

Proposition 7.3. *Assume that the model (21) satisfies Assumptions 1-5. If $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a Borel-measurable functions such that $\mathbb{E}|\varphi(X_0, \dots, X_{(d-1)\Delta})| < \infty$, then as $n \to \infty$*

$$\frac{1}{n}\sum_{i=0}^{n-1} \varphi(X_{i\Delta}, \dots, X_{(i+d-1)\Delta}) \xrightarrow{a.s.} \mathbb{E}\varphi(X_0, \dots, X_{(d-1)\Delta}).$$

In practice, we will observe the process X for a long time and replace the limit in the previous equation with a truncated average to produce the approximate identity for N large:

$$\frac{1}{N}\sum_{i=0}^{N-1} \varphi(X_{i\Delta}, \dots, X_{(i+d-1)\Delta}) \approx f_\varphi(\rho, \sigma, \theta) \tag{22}$$

where $f_\varphi(\rho, \sigma, \theta) = \mathbb{E}\varphi(X_0, \dots, X_{(d-1)\Delta})$. The function f_φ can be evaluated either numerically or analytically, depending on the model. The left hand side of (22) is a number that can be computed explicitly from the data. The right hand side is a function of the parameter values. Therefore, the above relationship gives a nonlinear equation involving the vector of parameters. Repeating the process for different choices of φ (in practice, usually polynomials) yields a system of nonlinear equations, which can then be inverted to obtain estimates $\hat{\rho}, \hat{\sigma}, \hat{\theta}$ of the true parameter values.

The estimation procedure described above is usually referred to as the "Method of Moments". In the computationally intensive case where f_φ needs to be evaluated using Monte-Carlo simulation of the process with the given parameter values it is referred to as the "Simulated Method of Moments". In order to understand better its performance, it is useful to study the error in the approximate equation (22), particularly as it relates to the number of observations N. With additional assumptions on the mixing coefficients of X we can use Ibaragimov's Central Limit Theorem[1] (see [Ibr62] and [HH80]). Given functions $\varphi_1, \dots, \varphi_p$:

$$\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1} \begin{pmatrix} \varphi_1(X_{i\Delta}, \dots, X_{(i+d-1)\Delta}) - f_{\varphi_1}(\rho, \sigma, \theta) \\ \vdots \\ \varphi_p(X_{i\Delta}, \dots, X_{(i+d-1)\Delta}) - f_{\varphi_p}(\rho, \sigma, \theta) \end{pmatrix} \xrightarrow{\mathcal{L}} N_p(0, \Sigma). \tag{23}$$

[1] For example, a sufficient conditions is that X is ρ-mixing (see Appendix A)

That is, as N tends to infinity, the random vector consisting of the errors in (22) *multiplied by* $\sqrt{N}$ converges to a p-dimensional normal random variable with mean 0 and variance-covariance matrix $\Sigma \in \mathbb{R}^{p\times p}$ (the exact expression for Σ is rather complicated). The result therefore gives a (probabilistic) estimate on the size of the error in the equation used for the method of moments estimate. Since the error must be multiplied by $\sqrt{N}$ in order to produce a nontrivial limit, it is common to say that the rate of convergence is $O(1/\sqrt{N})$, or that the error "goes down like $1/\sqrt{N}$". Nonetheless, we stress that, as is common in such problems, we only have a *probabilistic* bound on the error.

In order to estimate the parameters using the Method of Moments, we need as many integrable functions as the dimension of the parametric space, that is $\dim(\Theta) + 2$. One method is to use Assumption 4 ($\mathbb{E}|V_0|^p < \infty$ for some $p > 1$) together with the following lemma.

Lemma 7.4. *Suppose that φ is a Borel-measurable function $\varphi : \mathbb{R}^d \to \mathbb{R}$ and there exist a positive constants K and r such that $1 \leq r \leq p$ and*

$$|\varphi(s_0, s_1, \dots, s_{d-1})| \leq K\Big(1 + \sum_{k=0}^{d-1} |s_k|^r\Big),$$

Then $\mathbb{E}|\varphi(X_0, X_\Delta, \dots, X_{(d-1)\Delta})| < \infty$.

Proof. From Theorem 5.1, Corollary 3.9 and Assumption 4 we have that $\mathbb{E}|Y_t|^r \leq c_r < \infty$. Clearly, for any $0 < M \leq +\infty$ we have that $\mathbb{E}|X_t|^r \leq \mathbb{E}|Y_t|^r < \infty$. Finally

$$\mathbb{E}|\varphi(X_0, X_\Delta, \dots, X_{(d-1)\Delta})| \leq K\Big(1 + \sum_{k=0}^{d-1} \mathbb{E}|X_{i\Delta}|^r\Big) < \infty,$$

which completes the proof. □

In [GCJL00] was proven that the process V in (17) is ρ-mixing. Hence, we can have a Central Limit Theorem using the fact that the process $Z = (Z_n)_{n\in\mathbb{N}}$ defined by

$$Z_n = X_{(n+1)\Delta} - e^{-\rho\Delta} X_{n\Delta}$$

is ρ-mixing (by Theorem 3.4). This has the disadvantage that the parameter ρ has to be known beforehand.

For the OU-drift model a much stronger result is possible, since the joint Gaussianity of V and Y can be used together with the spectral gap inequality (see [Bak02]) to show that Y (and hence X) is ρ-mixing.

8 Numerical results

In this Section we simulate all the models using MATLAB. We verify the convergence of the sample time averages, as predicted by the ergodic theory, and give empirical evidence supporting that the rate of the convergence is of order $\frac{1}{\sqrt{n}}$. Finally, we also also perform the calibration of the parameters.

8.1 Convergence of the moments

Assume the model (17) and define:

$$e_1 = \beta, \qquad e_2 = \beta^2 + \mathrm{Var}(Y_t), \qquad \text{and}$$

$$e_{2+j} = \beta^2 + \frac{v_\lambda}{1-(\frac{\alpha}{\rho})^2} e^{-j\alpha\Delta} + \left(\mathrm{Var}(Y_t) - \frac{v_\lambda}{1-(\frac{\alpha}{\rho})^2}\right) e^{-j\rho\Delta} \quad j = 1, 2, 3,$$

where $\mathrm{Var}(Y_t) = \frac{\sigma^2}{2\rho} + \frac{v_\lambda}{1+\frac{\alpha}{\rho}}$. Also, for $N \in \mathbb{N}$ define:

$$m_N^{(j)} = \frac{1}{N+1} \sum_{n=0}^{N} X_{n\Delta}^j \approx e_j \quad j = 1, 2,$$

$$\text{and} \qquad m_N^{(2+j)} = \frac{1}{N+1-j} \sum_{n=j}^{N} X_{(n-j)\Delta} X_{n\Delta} \approx e_{2+j} \quad j = 1, 2, 3.$$

Tables 1 and 2 shows the results of 1000 simulations of the CIR-drift model ($\lambda = \frac{1}{2}$). In both cases we have 10000 observations, first with $\Delta = 0.1$ on the time interval $[0, 1000]$, and then with $\Delta = 1$ and a longer time interval $[0, 10000]$.

	$e_1 = 10$	$e_2 = 102.46667$	$e_3 = 102.44874$
n	$\bar{m}^{(1)}$ (mean.err.)	$\bar{m}^{(2)}$ (mean.err.)	$\bar{m}^{(3)}$ (mean.err.)
100	9.99549	101.83554	101.82028
	(9.67%)	(18.93%)	(18.96%)
1000	10.02925	102.95713	102.94077
	(6.23%)	(12.23%)	(12.24%)
10000	9.99472	102.34364	102.32580
	(2.24%)	(4.41%)	(4.41%)

Table 1: 1000 simulations with $\Delta = 0.1$ on $t \in [0, 1000]$.

	$e_1 = 10$	$e_2 = 102.46667$	$e_3 = 102.29379$
n	$\bar{m}^{(1)}$ (mean.err.)	$\bar{m}^{(2)}$ (mean.err.)	$\bar{m}^{(3)}$ (mean.err.)
100	9.99681	102.38183	102.20773
	(6.48%)	(12.69%)	(12.77%)
1000	9.98718	102.21245	102.03911
	(2.18%)	(4.29%)	(4.29%)
10000	10.00095	102.48519	102.31227
	(0.69%)	(1.36%)	(1.36%)

Table 2: 1000 simulations with $\Delta = 1$ on $t \in [0, 10000]$.

In both cases we observe the convergence of the moments, as predicted by the Ergodic Theorem. Comparing the results of both tables we see that increasing the period of observations has a better impact on the error than increasing the frequency of the sampling. The 4^{th} and 5^{th} moments behaved very similar to the 3^{rd}.

8.2 Estimation

Using the relations $m_N^{(i)} \approx e_i$ for $i = 1, \ldots, 5$, the estimators $\hat{\alpha}$, $\hat{\beta}$, $\hat{\nu}$, $\hat{\rho}$, and $\hat{\sigma}$ can be computed. The main difficulty is due to the non-linearity of the e_i's as functions of the parameters (specially α and ρ). This implies that we need a numerical method to solve the equations, and finding a good initial guess of the parameters is not trivial. Also, the error incurred in the approximation $m_N^{(i)} \approx e_i$ gets magnified by the non-linearity of the equations.

We ran several simulations and solved the non-linear system numerically. We found that even when we used the real values of the parameters as the initial guess, and the relative errors of the $m^{(i)}$'s were below 1%, we obtained estimators with relative errors above 60%.

All this is because the high dimensionality of the parametric space. Therefore we are going to assume some of the parameters as known, and estimate the others. In doing that simplification, we are going to introduce extra complexity using a price cap.

Assume an OU-drift model ($\lambda = 0$) with a price cap M. Figure 1 shows the result of one simulation for $\alpha = 0.05$, $\beta = 10$, $\nu = 0.1$, $\rho = 0.1$, and $\sigma = 0.6$. The observations were made at intervals of length $\Delta = 1$ from $t = 0$ to $t = 500$.

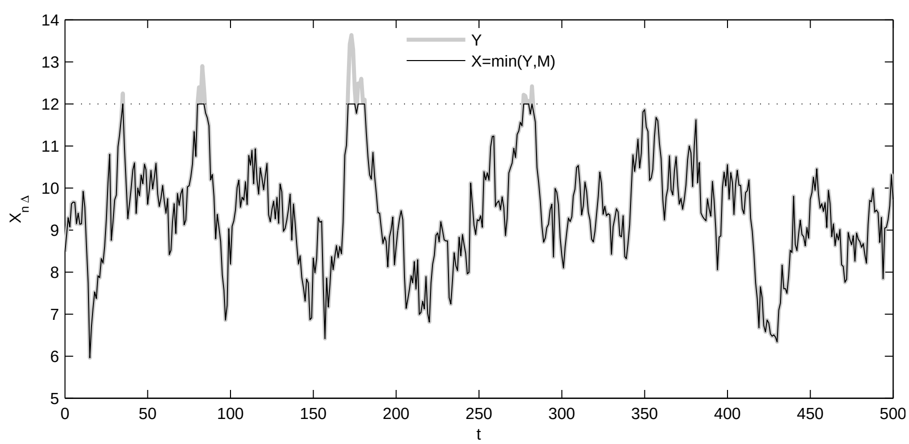

Figure 1: Simulation of X_t ($\alpha = 0.05$, $\beta = 10$, $\nu = 0.1$, $\rho = 0.1$, $\sigma = 0.6$).

Assume that α, ν and σ are known, and we want to estimate β and ρ given the observations $(X_{n\Delta})_{n\in\mathbb{N}}$. Since in this case Y_t is a Gaussian process with:

$$\mathbb{E}Y_t = \beta \qquad \text{and} \qquad \mathrm{Var}(Y_t) = \Gamma^2 := \frac{\sigma^2}{2\rho} + \frac{\frac{\nu^2}{2\alpha}}{1+\frac{\alpha}{\rho}}$$

we can compute $e_1 = \mathbb{E}X_t$ and $e_2 = \mathbb{E}X_t^2$ exactly. Denote:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{and} \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2}\, du.$$

Then:

$$e_1 = M + (\beta - M)\Phi\Big(\frac{M-\beta}{\Gamma}\Big) - \Gamma\varphi\Big(\frac{M-\beta}{\Gamma}\Big),$$

$$\text{and} \qquad e_2 = M^2 + (\beta^2 + \Gamma^2 - M^2)\Phi\Big(\frac{M-\beta}{\Gamma}\Big) - (\beta+M)\Gamma\varphi\Big(\frac{M-\beta}{\Gamma}\Big).$$

Table 3 shows the results for 1000 simulations.

	$e_1 = 9.9565$	$e_2 = 100.7769$	$\beta = 10$	$\rho = 0.1$
n	$\bar{m}^{(1)}$ mean error	$\bar{m}^{(2)}$ mean error	$\hat{\beta}$ mean error	$\hat{\rho}$ mean error
500	2.1%	4.07%	2.25%	20.8%
1000	1.55%	3%	1.66%	14%
5000	0.67%	1.3%	0.719%	5.9%
10000	0.496%	0.961%	0.533%	4.31%

Table 3: 1000 estimations of (β, ρ) with $\Delta = 1$ on $t \in [0, 10000]$.

A Strictly Stationary Processes

A.1 Mixing coefficients

Let $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$ be two σ-algebras. The following three measures of dependence between them can be defined (see for example [Dou94] and [Bra86a])

$$\alpha(\mathcal{G}_1,\mathcal{G}_2) = \sup\big\{|\mathrm{Cov}(U_1,U_2))|; 0 \le U_1, U_2 \le 1, U_i\ \mathcal{G}_i\text{-measurable for } i = 1,2\big\},$$
$$\beta(\mathcal{G}_1,\mathcal{G}_2) = \mathbb{E}\big[\text{ess.sup}\big\{|\mathbb{P}[B|\mathcal{G}_1] - \mathbb{P}[B]|; B \in \mathcal{G}_2\big\}\big],$$
$$\rho(\mathcal{G}_1,\mathcal{G}_2) = \sup\big\{|\mathrm{corr}(X_1,X_2)|; X_1, X_2 \text{ real}, X_1 \in L^2(\mathcal{G}_1), X_2 \in L^2(\mathcal{G}_2)\big\}.$$

These coefficients are related by the inequalities $2\alpha \le \beta \le 1$ and $4\alpha \le \rho \le 1$.

Let X be a process and define $\mathcal{G}_t = \sigma(X_s; s \le t)$ and $\mathcal{G}^t = \sigma(X_s; s \ge t)$. Then $\alpha_X(t)$, $\beta_X(t)$ and $\rho_X(t)$ are defined by $c_X(t) = \sup_{s\ge 0} c(\mathcal{G}_s, \mathcal{G}^{s+t})$, with $c = \alpha$, β or ρ. X is said to be *c-mixing* if $c_X(t) \to 0$ when $t \to \infty$.

If X is a strictly stationary process, then $c_X(t) = c(\mathcal{G}_0, \mathcal{G}^t)$. If X is also a Markov process, then $c_X(t) = c(\sigma(X_0), \sigma(X_t))$, and in this case

$$\begin{aligned}
\alpha_X(t) =& \sup \big\{ |\mathrm{Cov}(f(X_0), g(X_t))|; f, g \text{ are } \mathcal{B}(S)\text{-measurable and } 0 \le f, g \le 1 \big\}, \\
\beta_X(t) =& \mathbb{E}\big[\mathrm{ess.sup} \big\{ |\mathbb{P}[X_t \in B|X_0] - \mathbb{P}[X_t \in B]|; B \in \mathcal{B}(S) \big\} \big], \\
\rho_X(t) =& \sup \big\{ |\mathrm{corr}(f(X_0), g(X_t))|; f, g \in L^2_\pi \big\},
\end{aligned}$$

where π is the stationary distribution of X (see [Bra86a]).

The following result (see [Bra86a]) states that the mixing conditions are stronger than ergodicity. However, some times they are easier to verify.

Proposition A.1. *If a strictly stationary process X is α-mixing, then it is ergodic.*

A.2 Ergodicity and Hidden Markov Models

The following definition is based on [Ler92] (see also [BR96]). Let $(S, \mathbb{B}(S))$ and $(T, \mathbb{B}(T))$ be two Polish spaces equipped with their Borel σ-algebras. A stochastic process $(Z_n)_{n\in\mathbb{N}}$ with state-space S is a *Hidden Markov Model* (HMM) if there exists a strictly stationary Markov process $(U_n)_{n\in\mathbb{N}}$ with state-space T such that:

(i) For all n, $(Z_k)_{k\le n}$ are conditionally independent given $(U_1, U_2, \ldots, U_n)$, and the conditional distribution of Z_k depends only on U_k.

(ii) The conditional distribution of Z_k given $U_k = u$ does not depend on k.

We will refer to U as the *hidden chain* and Z as the *observed chain*.

Proposition A.2. *If Z is a HMM with hidden chain U, then Z is strictly stationary. If U is ergodic, then Z is also ergodic. Moreover, $\alpha_Z(n) \le \alpha_U(n)$ and $\rho_Z(n) \le \rho_U(n)$.*

Proof. The strict stationarity of Z and the inequality $\alpha_Z(n) \le \alpha_U(n)$ are both proven in [GCJL00]. That the ergodicity of U implies the ergodicity of Z is proven in [Ler92]. Here we are going to show the inequality $\rho_Z(n) \le \rho_U(n)$.

First notice that since Z is strictly stationary, the definition of $\rho_Z(n)$ can be rewritten as

$$\begin{aligned}
\rho_Z(n) = \sup \big\{ \big|\mathrm{corr}(\phi(Z_1, \ldots, Z_i), \psi(Z_{i+n+1}, \ldots, Z_{i+n+j}))\big|; \\
\phi : S^i \to \mathbb{R}, \psi : S^j \to \mathbb{R}, \phi, \psi \in L^2_\pi, i, j \in \mathbb{N} \big\}.
\end{aligned}$$

To simplify notation, write $\Phi = \phi(Z_1, \ldots, Z_i)$ and $\Psi = \psi(Z_{i+n+1}, \ldots, Z_{i+n+j})$. Then, with a little abuse of notation

$$\rho_Z(n) = \sup \big\{ \big|\mathbb{E}\Phi\Psi\big|; \mathbb{E}\Phi = \mathbb{E}\Psi = 0, \mathbb{E}\Phi^2 \le 1, \mathbb{E}\Psi^2 \le 1 \big\}.$$

For any L^2_π function $\phi : S^i \to \mathbb{R}$ we can define $H\phi : T^i \to \mathbb{R}$ as

$$H\phi(u_1, \ldots, u_i) = \mathbb{E}[\phi(Z_1, \ldots, Z_i)|U_1 = u_1, \ldots, U_i = u_i].$$

Notice that by Jensen's inequality we have that $(H\phi)^2 \le H(\phi^2)$.

Now assume that $\mathbb{E}\Phi = \mathbb{E}\Psi = 0$, $\mathbb{E}\Phi^2 \le 1$ and $\mathbb{E}\Psi^2 \le 1$. Conditioning with respect to $U_1, \ldots, U_{i+n+j}$ and using the definition of HMM we obtain that

$$\begin{aligned} \mathbb{E}(H\phi)(U_1, \ldots, U_i) &= \mathbb{E}\Phi = 0, \\ \mathbb{E}(H\psi)(U_{i+n+1}, \ldots, U_{i+n+j}) &= \mathbb{E}\Psi = 0, \\ \mathbb{E}(H\phi)^2(U_1, \ldots, U_i) &\le \mathbb{E}\Phi^2 \le 1, \\ \mathbb{E}(H\psi)^2(U_{i+n+1}, \ldots, U_{i+n+j}) &\le \mathbb{E}\Psi^2 \le 1, \\ \text{and} \qquad \mathbb{E}[(H\phi)(U_1, \ldots, U_i)(H\psi)(U_{i+n+1}, \ldots, U_{i+n+j})] &= \mathbb{E}\Phi\Psi. \end{aligned}$$

In that case, $\mathbb{E}[(H\phi)(U_1, \ldots, U_i)(H\psi)(U_{i+n+1}, \ldots, U_{i+n+j})] \le \rho_U(n)$. Thus we have that $\rho_Z(n) \le \rho_U(n)$. □

Remark A.3. *Notice that in a HMM the hidden chain U is a strictly stationary Markov process by definition, but the observable chain Z is not necessarily a Markov process. Nevertheless, the previous proposition asserts that Z is always strictly stationary.*

B Moments of the hidden processes

Assume that V follows the model (17) with $\lambda = 0, \frac{1}{2}, 1$, and V_0 follows its stationary distribution, so that the process V is strictly stationary. Denote $v_t = (V_t - \beta)$. Then

$$dv_t = -\alpha v_t \, dt + \nu(\beta + v_t)^\lambda \, dW_t.$$

Lemma B.1. *For any $0 \le s \le t$, and any $m \in \mathbb{N}$ we have that*

$$e^{r_m t} v_t^m = e^{r_m s} v_s^m + q_m \int_s^t e^{r_m u} [a v_u^{m-2} + b v_u^{m-1}] \, du + [N_t^{(m)} - N_s^{(m)}] \tag{24}$$

where $N_t^{(m)}$ is a $\mathcal{F}_t$-martingale,

$$q_m = \frac{m(m-1)}{2} \nu^2, \qquad r_m = m\alpha - \delta q_m,$$

and the coefficients a, b and δ depend on λ as follows:

λ	a	b	δ
0	1	0	0
$\frac{1}{2}$	β	1	0
1	β^2	2β	1

Proof. Applying Itô's lemma we obtain

$$d[e^{r_m t} v_t^m] = e^{r_m t}\Big[r_m v_t^m\, dt + m v_t^{m-1}\, dv_t + \frac{m(m-1)}{2} v_t^{m-2} d < v >_t \Big]$$
$$= e^{r_m t}\Big[r_m v_t^m\, dt - m\alpha v_t^m\, dt + m\nu v_t^{m-1}(\beta + v_t)^\lambda\, dW_t + q_m v_t^{m-2}(\beta + v_t)^{2\lambda}\, dt\Big].$$

Depending on the value of λ we have

$$v_t^{m-2}(\beta + v_t)^{2\lambda} = \begin{cases} v_t^{m-2} & \text{if } \lambda = 0 \\ \beta v_t^{m-2} + v_t^{m-2} & \text{if } \lambda = \frac{1}{2} \\ \beta^2 v_t^{m-2} + 2\beta v_t^{m-2} + v_t^m & \text{if } \lambda = 1 \end{cases}$$

Regardless the value of λ, all the terms containing v_t^m on the right hand side cancels out. Hence, defining

$$N_t^{(m)} = m\nu \int_0^t e^{r_m u} v_u^{m-1}(\beta + v_u)^\lambda\, dW_u$$

we obtain (24). □

Lemma B.2. *Denote $M_m = \mathbb{E} v_t^m = \mathbb{E}(V_t - \beta)^m$, and $\kappa = \frac{\nu^2}{2\alpha}$. Then*

$$\begin{aligned} M_0 &= 1 \\ M_1 &= 0 \\ M_2 &= \frac{1}{\left(\frac{1}{\kappa} - \delta\right)} a \\ M_3 &= \frac{1}{\left(\frac{1}{2\kappa} - \delta\right)} b M_2 = \frac{1}{\left(\frac{1}{\kappa} - \delta\right)\left(\frac{1}{2\kappa} - \delta\right)} ab. \end{aligned}$$

Proof. Clearly $M_0 = 1$. Taking expectations on (24) and using the stationarity of v we obtain for $m = 1$ that $M_1 = 0$ (since $q_1 = 0$). For $m \geq 2$ we obtain the recursive relation

$$M_m = \frac{q_m}{r_m}[a M_{m-2} + b M_{m-1}] \qquad \forall m \geq 2.$$

Noticing that

$$\frac{q_m}{r_m} = \frac{q_m}{m\alpha - \delta q_m} = \frac{1}{\frac{1}{(m-1)\frac{\nu^2}{2\alpha}} - \delta}$$

we easily obtain M_2 and M_3. □

Corollary B.3. *Depending on the value of λ, the first three moments of v are given by the following table:*

Lemma B.4. *If $0 \leq s_1 \leq s_2 \leq s_3$ then*

$$\mathbb{E}[v_{s_1} v_{s_2}] = e^{-\alpha(s_2 - s_1)} M_2,$$
$$\mathbb{E}[v_{s_1} v_{s_2} v_{s_3}] = e^{-\alpha(s_3 - s_1)} M_3.$$

λ	M_0	M_1	M_2	M_3
0	1	0	$\frac{\nu^2}{2\alpha}$	0
$\frac{1}{2}$	1	0	$\frac{\beta\nu^2}{2\alpha}$	$\frac{\beta^2\nu^4}{2\alpha^2}$
1	1	0	$\frac{\beta^2\nu^2}{2\alpha-\nu^2}$	$\frac{2\beta^3\nu^4}{(2\alpha-\nu^2)(\alpha-\nu^2)}$

Proof. Using that $q_1 = 0$ and $r_1 = \alpha$ we can write (24) for $m = 1$ as

$$e^{\alpha s_2} v_{s_2} = e^{\alpha s_1} v_{s_1} + [N^{(1)}_{s_2} - N^{(1)}_{s_1}]. \tag{25}$$

Multiplying by v_{s_1} and taking expectations we obtain

$$e^{\alpha s_2}\mathbb{E}[v_{s_1}v_{s_2}] = e^{\alpha s_1}\mathbb{E}[v^2_{s_1}] + \mathbb{E}[v_{s_1}(N^{(1)}_{s_2} - N^{(1)}_{s_1})].$$

But $\mathbb{E}[v_{s_1}(N^{(1)}_{s_2} - N^{(1)}_{s_1})] = \mathbb{E}\big[v_{s_1}\mathbb{E}[N^{(1)}_{s_2} - N^{(1)}_{s_1}|\mathcal{F}_{s_1}]\big] = 0$. Hence

$$\mathbb{E}[v_{s_1}v_{s_2}] = e^{-\alpha(s_2-s_1)}M_2.$$

Similarly, rewriting (25) with s_2 and s_3, multiplying by $v_{s_1}v_{s_2}$, and taking expectations we get

$$e^{\alpha s_3}\mathbb{E}[v_{s_1}v_{s_2}v_{s_3}] = e^{\alpha s_2}\mathbb{E}[v_{s_1}v^2_{s_2}] + \mathbb{E}[v_{s_1}v_{s_2}(N^{(1)}_{s_3} - N^{(1)}_{s_2})],$$

thus $\mathbb{E}[v_{s_1}v_{s_2}v_{s_3}] = e^{-\alpha(s_3-s_2)}\mathbb{E}[v_{s_1}v^2_{s_2}]$. To obtain this last expectation we first write (24) for $m = 2$ as

$$e^{r_2 s_2}v^2_{s_2} = e^{r_2 s_1}v^2_{s_1} + q_2\int_{s_1}^{s_2} e^{r_2 u}[av^0_u + bv^1_u]\,du + [N^{(2)}_{s_2} - N^{(2)}_{s_1}].$$

Multiplying by v_{s_1} and taking expectations we obtain

$$e^{r_2 s_2}\mathbb{E}[v_{s_1}v^2_{s_2}] = e^{r_2 s_1}\mathbb{E}[v^3_{s_1}] + q_2\int_{s_1}^{s_2} e^{r_2 u}\big(a\mathbb{E}[v_{s_1}] + b\mathbb{E}[v_{s_1}v_u]\big)\,du$$

Since $\mathbb{E}[v_{s_1}] = M_1 = 0$ and $\mathbb{E}[v_{s_1}v_u] = e^{-\alpha(u-s_1)}M_2$ we have that

$$\begin{aligned} e^{r_2 s_2}\mathbb{E}[v_{s_1}v^2_{s_2}] &= e^{r_2 s_1}M_3 + q_2 bM_2 e^{\alpha s_1}\int_{s_1}^{s_2} e^{(r_2-\alpha)u}\,du \\ &= e^{r_2 s_1}\Big(M_3 - \frac{q_2 bM_2}{r_2-\alpha}\Big) + e^{\alpha s_1 + (r_2-\alpha)s_2}\frac{q_2 bM_2}{r_2-\alpha}. \end{aligned}$$

Lemma B.2 and a little algebra shows that $\frac{q_2 bM_2}{r_2-\alpha} = M_3$. Then

$$\begin{aligned} \mathbb{E}[v_{s_1}v_{s_2}v_{s_3}] &= e^{-\alpha(s_3-s_2)}\mathbb{E}[v_{s_1}v^2_{s_2}] \\ &= e^{-\alpha(s_3-s_2)}e^{-r_2 s_2}e^{\alpha s_1+(r_2-\alpha)s_2}M_3 = e^{-\alpha(s_3-s_1)}M_3. \end{aligned}$$

which concludes the proof. □

References

[Bak02] D. Bakry, *Functional inequalities for Markov semigroups*, Notes de cours du Tata institute, Bombay; preprint, 2002.

[Bar02] M.T. Barlow, *A diffusion model for electricity prices*, Math. Finance **12** (2002), 287–298.

[BR96] Peter J. Bickel and Ya'acov Ritov, *Inference in hidden Markov models I: Local asymptotic normality in the stationary case*, Bernoulli **2** (1996), 199–228.

[Bra86a] R.C. Bradley, *Basic properties of strong mixing conditions*, Dependence in Probability and Statistics, A Survey of Recent Results (Boston) (E. Eberlein and M.S. Taqqu, eds.), Birkhäuser, 1986.

[Bra86b] Andreas Brandt, *The stochastic equation* $Y_{n+1} = A_n Y_n + B_n$ *with stationary coefficients*, Advances in Applied Probability **18** (1986), 211–220.

[Dou94] P. Doukhan, *Mixing: Properties and examples*, Lecture Notes in Statist. 85, Springer-Verlag, Berlin, 1994.

[GCJL00] Valentine Genon-Catalot, Thierry Jeantheau, and Catherine Larédo, *Stochastic volatility models as hidden Markov models and statistical applications*, Bernoulli **6** (2000), no. 6, 1051–1079.

[GIY04] Xavier Guyon, Serge Iovleff, and Jian-Feng Yao, *Linear diffusion with stationary switching regime*, ESAIM Probability and Statistics **8** (2004), 25–35.

[Hes93] S. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Review of Financial Studies **6** (1993), no. 2, 327–343.

[HH80] P. Hall and C. C. Heyde, *Martingale limit theory and its applications*, Academic Press, New York, 1980.

[Ibr62] Ildar A. Ibragimov, *Some limit theorems for stationary processes*, Theory of Probability and its Applications **7** (1962), 349–382.

[JCR85] J.E. Ingersoll J.C. Cox and S.A. Ross, *A theory of the term structure of interest rates*, Econometrica **53** (1985), 385–407.

[Kal97] Olav Kallenberg, *Foundations of modern probability*, Springer, New York, 1997.

[KB37] N. Krylov and N. Bogolioubov, *La théorie générale de la mesure dans son application à létude des systèmes de la mècanique non linéaires*, Annals of Mathematics **38** (1937), 65–113.

[Kre85] Ulrich J. Krengel, *Ergodic theorems*, Walter de Gruyter & Co., Berlin, 1985.

[Ler92] Brian G. Leroux, *Maximum-likelihood estimation for hidden Markov models*, Stochastic Processes and their Applications **40** (1992), 127–143.

[LLSW] Ali Lari-Lavassani, Ali A. Sadeghi, and Anthony Ware, *Modelling and implementing mean reverting price processes in energy markets*, Electronic Publications of the International Energy Credit Association (www.ieca.net).

[LS02] J.J. Lucia and E.S. Schwartz, *Electricity prices and power derivatives: Evidence from the Nordic power exchange*, Review Deriv. Research **5** (2002), 5–50.

[Nel90] D.P. Nelson, *ARCH models as diffusion approximations*, Jounal of Econometrics **45** (1990), 7–38.

[Pil98] Dragana Pilipović, *Energy risk*, McGraw-Hill, New York, 1998.

[RW00] L.C.G. Rogers and D. Williams, *Diffusions, Markov processes and martingales, volume 2: Itô calculus*, second ed., Cambridge University Press, Cambridge, 2000.

[Sau01] D. Saunders, *Mathematical problems in the theory of incomplete markets*, Ph.D. thesis, University of Toronto, 2001.