Computers & Operations Research ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

Algorithmic estimation of risk factors in financial markets with stochastic drift

Janko Hernández [a,*,1], David Saunders [b,2], Luis Seco [c]

[a] ITAM School of Business, Mexico
[b] University of Waterloo, Canada
[c] RiskLab, University of Toronto, Canada

ARTICLE INFO

Keywords:
Algorithmic estimation
Stochastic drift
Method of moments
Interest rates
Commodity prices

ABSTRACT

We assume a financial market governed by a diffusion process reverting to a stochastic mean which is itself governed by an unobservable ergodic diffusion, similar to those observed in electricity and other energy markets. We develop a moment method algorithm for the estimation of the parameters of both the observable process and the unobservable stochastic mean. Our approach is contrasted with other methods for parameter estimation of partially observed diffusions, and applications to the modelling of interest rates and commodity prices are discussed.

1. Introduction

Risk analysis in the commodities sector has long been one of the fundamental problems in quantitative finance. At the heart of that problem lies the fact that observed price series do not seem to follow the mathematical assumptions normally used when modelling other asset classes, such as stocks or bonds, which are usually modelled using the simple geometric Brownian motion price model:

$$dY_t = \mu Y_t\,dt + \sigma Y_t\,dB_t, \tag{1}$$

where here Y_t stands for the price, μ stands for the price drift, σ stands for the price volatility and B_t is the Brownian motion.

The reason for the mathematical misfit has been identified to be due to two main factors. On the one hand, price spikes are too extreme compared to the volatility estimations that arise through (1). On the other hand, commodities are usually traded through futures and forward contracts, for delivery and payment at future times, which gives more relevance to the multivariate structure of the entire forward curve, as opposed to their univariate spot price. As a consequence, the price spike problem gets compounded when looking at the entire forward curve, and the inadequacy of normality driven considerations to understand price spikes gets amplified when looking at the entire multivariate structure of the forward curve. This has created the need to develop specific non-Gaussian multivariate models to model commodities markets that are able to explain these phenomena.

In addition, prices in commodities have also been observed to be seasonal; while this seems obvious at an intuitive level (in many parts of the world, energy consumption rises in the winter and summer and drops in other seasons), there are additional seasonality effects created by business cycles as well as supply–demand imbalances that require the modelling of prices with not only price spikes but also with some sort of mean reversion characteristics.

We refer the reader to [23], which highlights the needs for risk management and the challenges in modelling prices, and to [15], which develops a specific model for oil prices that successfully captures price spikes and stochastic volatility.

When modelling difficulties are solved by creating sophisticated mathematical models, invariably those mathematical models give rise to serious computational complexities. In other words, the complexity of the original problem is merely transferred, by the mathematical models, from the world of finance to the world of numerical computations.

Our goal in this article is, in the same vein as [15], to study the following model of relevance in the modelling of electricity prices (although, as we explain below, it is also of relevance in interest rate markets).

Consider the stochastic processes defined as the solution to the following equations:

$$\left.\begin{aligned} dY_t &= \rho(V_t - Y_t)\,dt + \sigma\,dB_t \\ dV_t &= b(V_t;\theta)\,dt + a(V_t;\theta)\,dW_t \end{aligned}\right\}, \tag{2}$$

where $\rho,\sigma > 0$, $\theta \in \mathbb{R}^P$ is a parameter vector, and B and W are independent standard Brownian motions. Only the process Y_t is

* Corresponding author. Tel.: +52 55 5628 4000x3435; fax: +52 55 5628 4049.
E-mail addresses: janko@itam.mx, jankohc@gmail.com (J. Hernández).

[1] The author wants to thank the support received from *Asociación Mexicana de Cultura.*
[2] The author gratefully acknowledges support from an NSERC Discovery grant.

0305-0548/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2010.09.007

observable. We also consider the case where Y_t can only be observed when it is below some level M. The level M represents a cap on the financial variable, e.g. a price ceiling imposed by the regulator of a particular market.

The model bears interesting characteristics: it allows for price spikes by allowing the price drift to be stochastic, and it allows for prices to fluctuate around a level which is not fixed. However, it gives rise to considerable computational challenges. First, of the two variables that make up the system, only one of them, the price Y_t can be observed, and the other one, the drift V_t cannot be observed. Therefore, estimating the parameters of the model from observed data becomes a serious challenge. If the model parameters were known, simulation of the process can easily be done through techniques based on random number generators. But a brute-force approach to simulate sample paths from a parameters space, and then optimizing the chosen parameters from a best-fit objective function is not feasible in practice. One needs direct estimators to express the model parameters from statistical properties of the observed data. In the rest of the article, we will show how the method of moments can be adapted to this situation to circumvent the estimation challenges of the problem.

The method of moments matches sample and population moments to derive a system of nonlinear equations whose solution is the parameter estimate. In our case, the sample moments are time averages of functions of observations of the process Y, which should converge to averages under the stationary distribution. A justification of the method therefore requires two things: first, a proof of the convergence of time averages to averages under the stationary distribution (Ergodic Theorem), and a corresponding theorem giving the distribution of the error of the parameter estimates (Central Limit Theorem); second, a method for generating tractable moment equations. The tractability of moment equations is critical and is the main thrust of this paper: without explicit moment equations, numerical or simulation methods must be employed in order to calculate moments under the stationary distribution, dramatically increasing the complexity and computational cost of the calibration.

The moment equations we will derive rest on issues relating to the convergence of time averages to averages under the stationary distribution, and depend on the ergodic properties of the process Y. Since V is a one-dimensional diffusion, one can easily verify its ergodic properties. It is also easy to show that $X=\max(Y,M)$ is ergodic if Y is. The difficult problem is to obtain the ergodicity of Y from that of V, and to compute the moments of Y once we know those of V. We solve both problems by a discretization, followed by suitable passages to the limit.

The results regarding the ergodicity of the process Y presented in this paper for stochastic drift models were motivated by those of [10] on stochastic volatility models for equity markets, in which they study processes of the form:

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

Here μ, a and b are real functions satisfying some technical conditions. Defining a convenient hidden Markov model, the authors show that if the volatility process σ^2 is ergodic, then the difference process $Z_n=X_{(n+1)h}-X_{nh}$ is ergodic for any step size h. Following similar arguments we show that in our case $Z_n=Y_{(n+1)h}-e^{-\rho h}Y_{nh}$ is ergodic for any step size h. In this paper, we will take the limit as $h\to 0$ to obtain ergodicity of the continuous-time process Y. A similar passage to the limit has been used by [11] for regime switching models. To derive tractable moment equations, we show that by evaluating the limit as $h\to\infty$, the explicit expressions for the moments of Y can be obtained.

Many other methods for estimating the parameters of diffusion processes with hidden variables have been developed. For example [16] employ a Monte-Carlo variant of the EM algorithm and explore applications to neural networks. Many authors have explored other likelihood based methods (see, for example [18] and the references therein). Several authors have also considered methods based on particle filtering (see, for example [9] and the references therein). There is also a large and growing literature covering the case where Y and V are Markov chains, rather than diffusions, see [7].

1.1. Applications

In this section, we briefly describe some applications of our algorithmic model (2). We focus on applications to financial mathematics. Other applications are undoubtedly possible. For example, [22] considered a model of the form (2), with the hidden process V a Markov chain on a periodic lattice, and studied its application to modelling experimental observations of the protein kinesin.

1.1.1. Interest rates

There is an enormous literature on modelling the stochastic behaviour of interest rates (see [6]). Refs. [3,4] used a process of the form (2) to model the evolution of the interest rate curve. In their model, Y_t is the *short rate*, the interest rate for borrowing over an infinitesimal time interval. The mean-reverting form of the model reflects the fact that over time interest rates tend to fluctuate around a long term average level. This "equilibrium level" to which the short rate is reverting is itself a stochastic process, V_t in our notation.

Ref. [4] provides motivation for their specification of the interest rate process in terms of the qualitative mathematical behaviour of rates (leptokurtosis in the short rate, matching that observed in the market) as well as the behaviour of central banks. In particular, in the U.S. market, since the early 1980s the Federal Reserve has emphasized interest rates over monetary aggregates in its management of monetary policy, intervening in the market on an almost daily basis, and shifting the target level of the overnight funds rate every 3–4 weeks on average (see [1,2] for an extensive discussion). Ref. [4] also presents an algorithm for estimating the parameters of the model based on the fact that bond prices are observable functionals of the process (Y_t,V_t), with a known closed form expression. This allows estimation of all the parameters of the process Y_t and partial estimation of the parameters of the process V_t. A mathematical analysis of the estimation method is not given.

1.1.2. Commodity prices

The modelling of commodity prices has been the subject of a dramatic upsurge in interest over the past few years, owing to many factors including the prevailing volatility and the deregulation of markets for some commodities (e.g. electricity). Historically, mean reversion has been viewed as a key component of commodities prices, distinguishing them from, for example, equities. The traditional explanation is that when prices are high there is incentive for high cost producers to enter the market, thus increasing supply and driving prices back down, while when prices are low, less efficient, high cost producers will exit the market, reducing supply and driving prices back up (see [21] and the references therein). Among other mean-reverting diffusion models, [13] study the properties of a model attributed to Pilipović:

$$dY_t=\rho(V_t-Y_t)\,dt+\sigma Y_t^{\gamma}\,dB_t, dV_t=\mu V_t\,dt+\xi V_t^{\delta}\,dW_t, \tag{3}$$

where B_t and W_t are independent Brownian motions, $\gamma,\delta \in \{0,1\}$ and α,μ,σ,ξ are real parameters. Here Y_t is the spot price of electricity, and V_t is the stochastic level to which it reverts. With the choice $\gamma=1,\mu<0$, this model satisfies our assumptions, and the results below apply.

Another property of some energy markets is the existence of *price caps*. These are upper limits on prices imposed by market regulators, resetting the "true" price arising from supply and demand to a maximum allowable price, whenever the market clearing price exceeds that maximum. Mathematically, instead of observing the true price Y_t, what we really observe is $\min(Y_t,M)$. M is the *cap* imposed by regulatory authorities.

1.2. Structure of the paper

The remainder of the paper is structured as follows. Section 2 presents definitions and technical assumptions. Section 3 studies ergodic properties of the model. We establish the ergodicity of Z and Y (by first establishing ergodicity of its discretizations, and then taking a limit). We also obtain explicit bounds for the moments that will be needed later. In Section 4 we show that by letting $h\to\infty$ we can obtain the moments and auto-covariance of Y. In Section 5 we compute the moments of Y explicitly in some important special cases. In Section 6 we discuss price caps and explain how to use the method of moments to construct estimators. Section 7 provides computational examples illustrating the use of the method, and Section 8 presents concluding remarks. Some proofs and technical definitions from the literature are relegated to appendices.

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 conditions and supporting a standard two-dimensional Brownian process $(B_t,W_t)_{t\in\mathbb{R}^+}$. Let $\rho,\sigma>0$, $\theta\in\Theta\subset\mathbb{R}^P$, and consider an interval (l,r) with $-\infty\le l<r\le\infty$, 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 $q\ge\frac{1}{2}$, $K_\theta>1$ such that:

$$|a(u;\theta)-a(v;\theta)|\le K_\theta|u-v|^q,\quad \forall u,v\in(l,r),$$
$$|b(u;\theta)-b(v;\theta)|\le K_\theta|u-v|,\quad \forall u,v\in(l,r),$$
$$\text{and}\quad a^2(u;\theta)+b^2(u;\theta)\le K_\theta(1+u^2),\quad \forall u\in(l,r).$$

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 (2). Assumption 1 ensures the existence and uniqueness of such a solution (see [20]).

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 $M=\int_l^r du/a^2(u)s(u)<\infty$.

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

The following result can be found in [10].

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.*[3]

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

We will see in Section 3 that the previous four assumptions on the hidden process V imply the existence of a unique stationary distribution for the observable process Y.

Assumption 5. Y_0 follows the unique stationary distribution implied by the model (2) under Assumptions 1–4.

3. Ergodicity in discrete and continuous time

If Assumptions 1–4 hold, then there exists a constant β such that $\mathbb{E}V_t=\beta$ for every $t\ge0$. We denote $y_t=Y_t-\beta$, $v_t=V_t-\beta$. Assumptions 1–4 hold for the process $V=(V_t)_{t\ge0}$ if and only if they hold for the process $v=(v_t)_{t\ge0}$.

Proposition 3.1. *Fix $h\ge0$. For any $t\ge0$ we write*

$$Z_t^{(h)}=\mu_t(h)+\Gamma(h)\xi_t^{(h)},\quad y_{t+h}=e^{-\rho h}y_t+Z_t^{(h)}, \tag{4}$$

where

$$\mu_t(h)=e^{-\rho h}\int_0^h \rho e^{\rho u}v_{t+u}\,du, \tag{5}$$

$$\Gamma^2(h)=\frac{\sigma^2}{2\rho}(1-e^{-2\rho h})\quad and\quad \xi_t^{(h)}\sim N(0,1). \tag{6}$$

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 $\mathbb{G}=(\mathcal{G}_u)_{u\ge0}$, 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 (4) 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 (5) after a change of variable. The second integral is a Gaussian random variable with variance $\int_t^{t+h}\sigma^2e^{-2\rho(t+h-s)}\,ds=\Gamma^2(h)$, so it can be written as $\Gamma(h)\xi_t^{(h)}$ with $\xi_t^{(h)}\sim N(0,1)$. Since B is independent of V, and hence independent of $\mathbb{G}$, $\xi_t^{(h)}$ is independent of $\mathbb{G}$. Independence of Brownian increments implies that $\xi_t^{(h)}$ 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. □

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 (4) can be rewritten as

$$y_{n+1}^{(h)}=e^{-\rho h}y_n^{(h)}+z_n^{(h)},\quad n\in\mathbb{N}, \tag{7}$$

with

$$z_n^{(h)}=\mu_{nh}(h)+\Gamma(h)\xi_n. \tag{8}$$

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 $y_0^{(h)}=y_0$. The first step is to establish the ergodic properties of $z^{(h)}$. The following results follow immediately by applying arguments from [10].

[3] For the definitions of mixing, see Appendix A.

Please cite this article as: Hernández J, et al. Algorithmic estimation of risk factors in financial markets with stochastic drift. Computers and Operations Research (2010), doi:10.1016/j.cor.2010.09.007

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* $U_n^{(h)} = (\mu_{nh}(h), V_{(n+1)h})$.

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

By Proposition 2.1 and the previous lemma, $U^{(h)}$ is $\beta-$mixing, and hence $\alpha-$mixing and ergodic.

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

Proof. For ergodicity, see [14], $\rho-$mixing is from [10]. □

We can extend the process $(z_n^{(h)})_{n\in\mathbb{N}}$ by stationarity to $n\in\mathbb{Z}$.

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

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

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*

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

where the sum on the right-hand side converges absolutely a.s. Denote by $\pi^{(h)}$ *the distribution 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* $x_n \xrightarrow{\mathcal{L}} \pi^{(h)}$ *as* $n\to\infty$.

Proof. See [5].

Comparing (9) with (7) we obtain the main result of this section.

Theorem 3.6. *If for any given* $h>0$ *condition* (10) *holds, then there exists a unique distribution* $\pi^{(h)}$ *such that for any* $\mathcal{F}_0$*-measurable random variable* y_0, *the discrete-time process* $y^{(h)}$ *defined by* (7) *satisfies*: $y_n^{(h)} \xrightarrow{\mathcal{L}} \pi^{(h)}$ *as* $n\to\infty$. *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 for any* $h>0$ *condition* (10) *holds, and for any* $x>0$, $\lim_{h\to 0}\mathbb{P}[|Z_s^{(h)}|\ge x]=0$, *uniformly in* $s\ge 0$.

The proof will make use of the following lemma, which 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* (5). *If Assumptions* 1–4 *hold, then for any* $1\le r\le p$ *we have* $\mathbb{E}|\mu_t(h)|^r \le (1-e^{-\rho h})^r\mathbb{E}|v_0|^r \le \mathbb{E}|v_0|^p < \infty, \forall t, h\ge 0$.

Proof. Define the measure η on $\mathbb{R}$ by $\eta(du) = \mathbf{1}_{[0,h]}(u)(\rho e^{-\rho(h-u)}/(1-e^{-\rho h}))\,du$. 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 $\int_{\mathbb{R}}\eta(ds)=1$. But $|\int_{\mathbb{R}} v_{t+u}\eta(du)|^r \le \int_{\mathbb{R}}|v_{t+u}|^r\eta(du)$ due to Jensen's inequality. The stationarity of v implies that $\mathbb{E}|v_s|^r = \mathbb{E}|v_0|^r$ for all $s\in\mathbb{R}^+$, hence

$$\mathbb{E}|\mu_t(h)|^r = (1-e^{-\rho h})^r\mathbb{E}\left|\int_{\mathbb{R}} v_{t+u}\eta(du)\right|^r \le (1-e^{-\rho h})^r\int_0^h \mathbb{E}|v_{t+u}|^r\eta(du)$$

$$= (1-e^{-\rho h})^r\mathbb{E}|v_0|^r\int_R \eta(du) = (1-e^{-\rho h})^r\mathbb{E}|v_0|^r \le \mathbb{E}|v_0|^p.$$

Minkowski's inequality and Assumption 4 then finish the proof. □

Proof of Proposition 3.7. Minkowski's inequality and the previous lemma imply that $(\mathbb{E}|Z_t^{(h)}|^r)^{1/r} \le (1-e^{-\rho h})(\mathbb{E}|v_0|^r)^{1/r} + \Gamma(h)(\mathbb{E}|\xi_t^{(h)}|^r)^{1/r}$.

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 $\mathbb{E}|Z_t^{(h)}|^r \le [(\mathbb{E}|v_0|^r)^{1/r} + (\sigma^2/2\rho)K_r]^r < \infty$ where $K_r = (\mathbb{E}|\xi_t^{(h)}|^r)^{1/r}$ is a constant. Since $\max(0,\ln|x|)\le |x|$, $\mathbb{E}(\ln|z_0^{(h)}|)^+ \le \mathbb{E}|z_0^{(h)}| = \mathbb{E}|Z_0^{(h)}| < \infty$ and condition (i) follows. □

Corollary 3.9. *For any r such that* $1\le r\le p$ *there exists a constant* c_r *that does not depend on h or t such that* $\mathbb{E}|Z_t^{(h)}|^r \le c_r < \infty$, $\forall t\in\mathbb{R}, h>0$.

We now 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. The proof of the next theorem follows [11] closely, and is therefore omitted.

Theorem 3.10. *If Assumptions* 1–4 *hold, there exists a unique probability measure* π_0 *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.*

Combining the previous theorem with Assumption 5, we obtain the following.

Theorem 3.11. *If Assumptions* 1–5 *hold, then both components V and Y of the model defined by model* (2) *are strictly stationary and ergodic.*

4. Exact moments of the stationary model

Henceforth, we suppose that Assumptions 1–5 hold, so that Y is strictly stationary and ergodic. We need to compute the moments of Y. The following result allows us to do so 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. Notice that stationarity implies that the relations hold for any $t\in\mathbb{R}^+$.

Theorem 4.1. *Assume that* $1\le r\le p$. *Then* $\mathbb{E}|Y_t|^r<\infty$ *for all* $t\in\mathbb{R}^+$. *If k is an integer such that* $1\le k\le p$, *then* $\mathbb{E}y_t^k = \mathbb{E}(Y_t-\beta)^r = \lim_{h\to\infty}\mathbb{E}[z_0^{(h)}]^k$. *Furthermore, if* $p\ge 2$ *then for any* $h>0$ *we have that*

$$\mathbb{E}[y_0y_h] = \mathbb{E}[(Y_0-\beta)(Y_h-\beta)] = e^{-\rho h}\mathbb{E}y_0^2 + \sum_{k=1}^{\infty} e^{-\rho(k-1)h}\mathbb{E}[z_0^{(h)}z_k^{(h)}]. \tag{12}$$

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 $h>0$. From (11) we have that $y_0 = \sum_{k=1}^{\infty} e^{-(k-1)\rho h}z_{-k}^{(h)}$. Using the triangle inequality and Corollary 3.9 we have that

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

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 $1\le k\le r$, then $\mathbb{E}[z_0^{(h)}]^k = \mathbb{E}[y_h - e^{-\rho h}y_0]^k = \mathbb{E}y_h^k + \sum_{i=1}^{k-1}\binom{k}{i}\mathbb{E}[(y_h)^{k-i}(e^{-\rho h}y_0)^i] + \mathbb{E}[e^{-\rho h}y_0]^k$. Fix $1\le i\le k-1$,

Please cite this article as: Hernández J, et al. Algorithmic estimation of risk factors in financial markets with stochastic drift. Computers and Operations Research (2010), doi:10.1016/j.cor.2010.09.007

and define $p=k/(k-i)$ and $q=k/i$. Then by Hölder's inequality

$$|\mathbb{E}[(y_h)^{k-i}e^{-\rho h}y_0)^i]| \le \mathbb{E}[|y_h|^{k-i}|e^{-\rho h}y_0|^i] \le [\mathbb{E}|y_h|^{(k-i)p}]^{1/p}[\mathbb{E}|e^{-\rho h}y_0|^{iq}]^{1/q}$$
$$= e^{-\rho h(k/q)}[\mathbb{E}|y_0|^k]^{1/p}[\mathbb{E}|y_0|^k]^{1/q} = e^{-\rho hi}\mathbb{E}|y_0|^k \to 0$$

as $h\to\infty$. Taking the limit in (4), we obtain the desired results. 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 $\mathbb{E}[y_0 y_h] = e^{-\rho h}\mathbb{E}y_0^2 + \mathbb{E}[y_0 z_0^{(h)}]$. From (11) we have that $y_0 = \sum_{k=1}^{\infty} e^{-(k-1)\rho h} z_{-k}^{(h)}$. Then $\mathbb{E}[y_0 z_0^{(h)}] = \sum_{k=1}^{\infty} e^{-\rho(k-1)h}\mathbb{E}[z_{-k}^{(h)} z_0^{(h)}]$. By the strict stationarity of $z^{(h)}$ we have that $\mathbb{E}[z_{-k}^{(h)} z_0^{(h)}] = \mathbb{E}[z_0^{(h)} z_k^{(h)}]$. If $p\ge 2$, then Corollary 3.9 implies that $2\mathbb{E}|z_0^{(h)} z_k^{(h)}| \le \mathbb{E}|z_0^{(h)}|^2 + \mathbb{E}|z_k^{(h)}|^2 \le 2c_2$, and hence the series converges absolutely. □

5. Exact moments in some examples

We consider the special case of model (2) with $a(V_t;\theta) = \alpha(\beta - V_t)$, $b(V_t;\theta) = \nu V_t^{\lambda}$ where $\alpha,\beta,\nu > 0$. In what follows we will denote $q = \nu^2/2\alpha$. Specifically, we consider the cases $\lambda = 0, \frac{1}{2}, 1$. When $\lambda = 0$ the V process is the Ornstein–Uhlenbeck process. We shall refer to it as the *OU-drift model*. It satisfies Assumptions 1–3 with $(l,r) = (-\infty,+\infty)$, and the stationary distribution for V is Gaussian with parameters (β,q). We refer to the case $\lambda = \frac{1}{2}$ as the *CIR-drift model* (after [8]). The CIR-drift model satisfies Assumptions 1–3 with $(l,r) = (0,+\infty)$ if $\beta \ge q$. The stationary distribution for V is a Gamma with parameters $(\beta/q, 1/q)$, 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 [17]). 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 $(1+1/q, \beta/q)$. The moments of order p are finite if $p < 1 + \frac{1}{q}$.

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 $q < \frac{1}{2}$. Proposition 3.7 and Theorem 3.10 then imply that Assumption 5 is satisfied. In the remainder of this section, we assume that Assumptions 1–5 are all satisfied.

The following lemmas provide a recursive equation that can be used to generate all the moments of V, and also results regarding the auto-correlation of V required to derive auto-correlations of Y. The proofs appear in Appendix B.

Lemma 5.1. *Denote* $M_m = \mathbb{E}v_t^m = \mathbb{E}(V_t - \beta)^m$, *and* $q = \nu^2/2\alpha$. *Then* $M_0 = 1$, $M_1 = 0$, *and*

$$M_m = \frac{aM_{m-2} + bM_{m-1}}{\frac{1}{(m-1)q} - \delta} \quad \text{for all } 2 \le m \le p.$$

Corollary 5.2. *Depending on* λ, *the first three moments of* v *are given by the table below. Furthermore,* M_2 *and* M_3 *are only functions of* β *and* q.

λ	M_0	M_1	M_2	M_3
0	1	0	q	0
$\frac{1}{2}$	1	0	βq	$2\beta q^2$
1	1	0	$\frac{\beta^2}{\left(\frac{1}{q}-1\right)}$	$\frac{2\beta^3}{\left(\frac{1}{q}-1\right)\left(\frac{1}{2q}-1\right)}$

Lemma 5.3. *If* $0 \le s_1 \le s_2 \le s_3$ *then*

$$\mathbb{E}[v_{s_1} v_{s_2}] = e^{-\alpha(s_2 - s_1)} M_2, \quad \mathbb{E}[v_{s_1} v_{s_2} v_{s_3}] = e^{-\alpha(s_3 - s_1)} M_3. \tag{13}$$

We can now compute the moments of y explicitly.

Theorem 5.4. *If* $\alpha \ne \rho$, *then moments of* y *are given by the following formulas*:

$$\mathbb{E}y_t = 0, \quad \mathbb{E}y_t^2 = \frac{\sigma^2}{2\rho} + \frac{M_2}{1+\frac{\alpha}{\rho}}, \quad \mathbb{E}y_t^3 = \frac{M_3}{\left(1+\frac{\alpha}{\rho}\right)\left(1+\frac{\alpha}{2\rho}\right)},$$

$$\mathbb{E}[y_0 y_h] = \left(\frac{\sigma^2}{2\rho} + \frac{M_2}{1+\frac{\alpha}{\rho}}\right)e^{-\rho h} + \frac{M_2}{1-\left(\frac{\alpha}{\rho}\right)^2}(e^{-\alpha h} - e^{-\rho h}), \quad \forall h > 0,$$

where M_2 *and* M_3 *are as in Lemma* 5.1.

Proof. In view of Theorem 4.1 we first need to compute the moments of $z^{(h)}$. By Proposition 3.1 $z_t^{(h)}$ is conditionally Gaussian, hence $\mathbb{E}[z_0^{(h)}] = \mathbb{E}[\mu_0(h)]$, $\mathbb{E}[z_0^{(h)}]^2 = \mathbb{E}[\mu_0(h)]^2 + \Gamma^2(h)$, and $\mathbb{E}[z_0^{(h)}]^3 = \mathbb{E}[\mu_0(h)]^3 + 3\mathbb{E}[\mu_0(h)]\Gamma^2(h)$. Denote $i_m = \lim_{h\to\infty}\mathbb{E}[\mu_0(h)]^m$ for $m \in \mathbb{N}$. Clearly $i_0 = 1$. From (5) $i_m = \lim_{h\to\infty}\mathbb{E}(e^{-\rho h}\int_0^h \rho e^{\rho s} v_s\, ds)^m, \forall m \ge 1$. An elementary identity then gives

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

We have that $i_1 = \rho\lim_{h\to\infty} e^{-\rho h}\int_0^h e^{\rho s_1}\mathbb{E}[v_{s_1}]\, ds_1 = 0$ by virtue of Lemmas 5.1 and 5.3, and since $\mathbb{E}[v_{s_1}] = M_1 = 0$. For $m=2$ we have

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

And for $m=3$ we have

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

A little algebra yields the desired expressions. Using Theorem 4.1, the equations for the moments of $z_0^{(h)}$, and the fact that $\lim_{h\to\infty}\Gamma^2(h) = \sigma^2/2\rho$, and substituting the expressions for i_2 and i_3 we obtain the first three moments of y. For the auto-correlation, we use (12). Let $k \ge 1$, then by (4), $\mathbb{E}[z_0^{(h)} z_k^{(h)}] = \mathbb{E}[\mu_0(h)\mu_{kh}(h)]$. By (8) and Lemma 5.3 we have that

$$\mathbb{E}[\mu_0(h)\mu_{kh}(h)] = e^{-2\rho h}\rho^2\int_0^h\int_0^h e^{\rho(s_1+s_2)}\mathbb{E}[v_{s_1}v_{kh+s_2}]\, ds_1\, ds_2$$
$$= \frac{e^{-(2\rho h+\alpha kh)}\rho^2 M_2}{\rho^2-\alpha^2}(e^{(\rho+\alpha)h}-1)(e^{(\rho-\alpha)h}-1).$$

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

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

Finally, by (12) we have that

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

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

6. Censoring and estimation

Recall that in applications of the model (2) to commodity price modelling, the variable Y_t may be subject to a cap M. This censoring procedure implies that we only observe $X_t = \min(Y_t, M)$,

Please cite this article as: Hernández J, et al. Algorithmic estimation of risk factors in financial markets with stochastic drift. Computers and Operations Research (2010), doi:10.1016/j.cor.2010.09.007

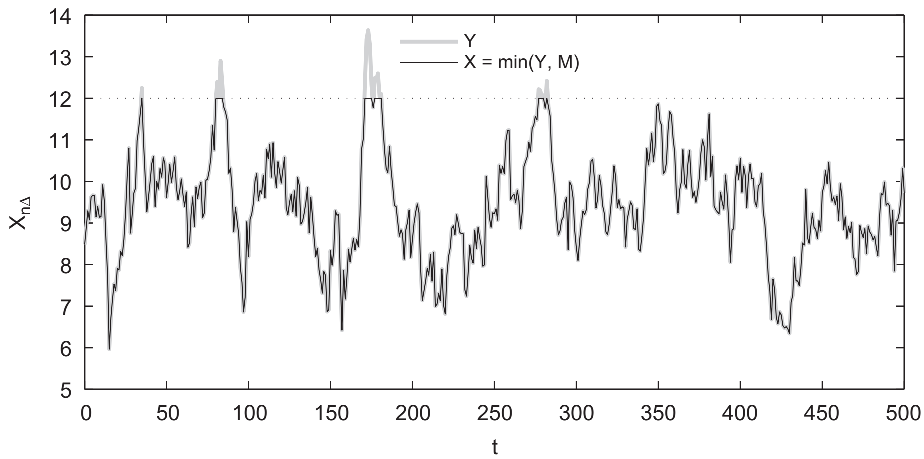

Fig. 1. Simulation of X_t.

where M is a known constant (see Fig. 1). The ergodicity of Y implies that of X, and the following result is thus an immediate corollary of Theorem 3.11.

Theorem 6.1. *Assume that the following model* (2) *satisfies Assumptions* 1–5 *and* $X_t = \min(Y_t, M)$. *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.*

Assume that we observe a time discretization of (X_t). That is, for some fixed and known $h > 0$, we observe $(X_{nh})_{n \in \mathbb{N}}$.[4] Our goal is to estimate the vector of parameters (ρ, σ, θ) from those observations. Birkhoff's Ergodic Theorem (see [12]) implies the following result.

Proposition 6.2. *Assume that the model* (2) *satisfies Assumptions* 1–5. *If* $\varphi : \mathbb{R}^d \to \mathbb{R}$ *is a Borel-measurable function such that* $\mathbb{E}|\varphi(X_0, \ldots, X_{(d-1)h})| < \infty$, *then as* $n \to \infty$

$$\frac{1}{N}\sum_{i=0}^{N-1} \varphi(X_{ih}, \ldots, X_{(i+d-1)h}) \xrightarrow{a.s.} \mathbb{E}\varphi(X_0, \ldots, X_{(d-1)h}).$$

In practice, we will observe the process X for a long time and replace the limit in the previous equation by an approximate identity for N large. The left-hand side can be computed explicitly from the data. The right-hand side is a function of the parameter values. Repeating the process for different choices of φ (in practice, usually polynomials) yields a system of nonlinear equations involving the vector of parameters, which can then be inverted to obtain estimates of the true parameter values.

With additional assumptions on the mixing coefficients of X we can use obtain a Central Limit Theorem.[5] Given functions $\varphi_1, \ldots, \varphi_P$:

$$\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1} \begin{pmatrix} \varphi_1(X_{ih}, \ldots, X_{(i+d-1)h}) - f_{\varphi_1}(\rho, \sigma, \theta) \\ \vdots \\ \varphi_P(X_{ih}, \ldots, X_{(i+d-1)h}) - f_{\varphi_P}(\rho, \sigma, \theta) \end{pmatrix} \xrightarrow{\mathcal{L}} N_P(0, \Sigma). \tag{14}$$

In order to estimate the parameters using the method of moments, we need (at least) 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 6.3. *Suppose that* φ *is a Borel-measurable function* $\varphi : \mathbb{R}^d \to \mathbb{R}$ *and there exists a positive constant* K *and* r *such that* $1 \le r \le p$ *and*

$$|\varphi(s_0, s_1, \ldots, s_{d-1})| \le K\left(1 + \sum_{k=0}^{d-1} |s_k|^r\right).$$

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

Proof. From Theorem 4.1, Corollary 3.9 and Assumption 4 we have that $\mathbb{E}|Y_t|^r \le c_r < \infty$. Clearly, for any $0 < M \le +\infty$ we have that $\mathbb{E}|X_t|^r \le \mathbb{E}|Y_t|^r < \infty$. Finally, $\mathbb{E}|\varphi(X_0, X_h, \ldots, X_{(d-1)h})| \le K(1 + \sum_{k=0}^{d-1} \mathbb{E}|X_{ih}|^r) < \infty$, which completes the proof. □

7. Numerical results

In this section we simulate the model (2) and verify the convergence of time averages in the sample to those under the stationary distribution.[6] We also illustrate the estimation of the model parameters using the method of moments.

7.1. Convergence of the moments

Assume the model given by (2), with $X_t = \min(Y_t, M)$. For a fixed time step h and for $N \in \mathbb{N}$ define:

$$m_N^{(j)} = \frac{1}{N+1}\sum_{n=0}^{N} X_{nh}^j, \quad j = 1,2,3 \quad m_N^{(4)} = \frac{1}{N}\sum_{n=1}^{N} X_{(n-1)h}X_{nh}, \tag{15}$$

$$\overline{m}_N^{(2)} = m_N^{(2)} - [m_N^{(1)}]^2, \quad \overline{m}_N^{(3)} = m_N^{(3)} - 3m_N^{(1)}\overline{m}_N^{(2)} - [m_N^{(1)}]^3, \tag{16}$$

$$\text{and} \quad \overline{m}_N^{(4)} = m_N^{(4)} - [m_N^{(1)}]^2. \tag{17}$$

Here m denotes the sample moment, and $\overline{m}$ denotes the centred sample moment. Proposition 6.2 implies that if N is big enough, then

$$m_N^{(1)} \approx L_1, \quad \overline{m}_N^{(2)} \approx L_2, \quad \overline{m}_N^{(3)} \approx L_3 \quad \text{and} \quad \overline{m}_N^{(4)} \approx L_4, \tag{18}$$

where L_1 is the first moment of Y, L_2, L_3 are the centred moments of order 2 and 3, and L_4 is the auto-covariance. Applying Theorem 5.4 in the case $M = +\infty$ (no cap) we have that

$$L_1 = \beta, \quad L_2 = s + \frac{M_2(\beta, q)}{1+k}, \quad L_3 = \frac{M_3(\beta, q)}{(1+k)(1+\frac{k}{2})} \tag{19}$$

[4] In the particular case $M = +\infty$, we observe a time discretization of Y.

[5] For example, a sufficient condition is that X is ρ-mixing, see [10].

[6] All computations in the paper were carried out using MATLAB.

and

$$L_4 = \left(s - k\frac{M_2(\beta,q)}{1-k^2}\right)e^{-\rho h} + \frac{M_2(\beta,q)}{1-k^2}e^{-k\rho h}, \tag{20}$$

where M_2, M_3 are defined as in Corollary 5.2, and

$$s = \frac{\sigma^2}{2\rho}, \quad q = \frac{v^2}{2\alpha}, \quad k = \frac{\alpha}{\rho}. \tag{21}$$

Tables 1 and 2 show the results of 1000 simulations of the CIR-drift model ($\lambda = \frac{1}{2}$) with $\alpha = 1.5$, $\beta = 2$, $v = 0.9$, $\rho = 6$, and $\sigma = 0.7$. In both cases we have 100,000 observations, first with $h = 0.1$ on $t \in [0,10^4]$, and then with $h = 1$ and a longer time interval $[0,10^5]$. Each entry represents the average value of the estimator over all the simulations, and the average relative error (between brackets) with respect to the true value of the parameter. In both cases we observe the convergence of the empirical moments. Comparing the two tables we see that for a fixed number of observations, increasing the time period yields lower relative errors than increasing the sampling frequency (i.e. decreasing the step h). We also observe that the convergence of the third order moment is considerably slower than that of the first and second order moments. The rate of convergence of the moments of order four is much slower, and poses the main practical obstacle for our estimation method.

7.2. Estimation without a cap

In order to find estimators $\hat{\alpha}$, $\hat{\beta}$, $\hat{v}$, $\hat{\rho}$, and $\hat{\sigma}$ for all the parameters, we need five independent equations involving them. Although it is possible to compute exactly an extra moment L_5, our numerical experiments showed that it is very difficult to obtain accurate estimates. The main difficulty is the slow rate of convergence of moments above order three. Computational experiments show that the numerical problem is more tractable in dimension 4 than in dimension 5. To reduce the dimensionality, we assume that we know the relative speed of reversion of Y with respect to V. That is, the parameter k in (21) is known.

Table 1
1000 simulations with $h = 0.1$ on $t \in [0,10^4]$.

t	$L_1 = 2.0000$	$L_2 = 0.4728$	$L_3 = 0.2074$	$L_4 = 0.4391$
	$m^{(1)}$ (rel.err.)	$\overline{m}^{(2)}$ (rel.err.)	$\overline{m}^{(3)}$ (rel.err.)	$\overline{m}^{(4)}$ (rel.err.)
100	1.9983	0.4560	0.1610	0.4220
	(3.38%)	(12.33%)	(35.57%)	(13.16%)
1000	2.0004	0.4710	0.2003	0.4373
	(1.08%)	(3.92%)	(12.83%)	(4.18%)
10,000	1.9998	0.4724	0.2045	0.4387
	(0.35%)	(1.30%)	(4.51%)	(1.38%)

Table 2
1000 simulations with $h = 1.0$ on $t \in [0,10^5]$.

t	$L_1 = 2.0000$	$L_2 = 0.4728$	$L_3 = 0.2074$	$L_4 = 0.1283$
	$m^{(1)}$ (rel.err.)	$\overline{m}^{(2)}$ (rel.err.)	$\overline{m}^{(3)}$ (rel.err.)	$\overline{m}^{(4)}$ (rel.err.)
1000	2.0004	0.4711	0.1966	0.1271
	(1.13%)	(4.66%)	(15.82%)	(11.96%)
10,000	2.0002	0.4726	0.2041	0.1281
	(0.38%)	(1.39%)	(5.31%)	(3.74%)
100,000	2.0001	0.4728	0.2052	0.1282
	(0.12%)	(0.46%)	(1.92%)	(1.17%)

The following trick further reduces the dimensionality of the problem from 4 to 1. First, we can compute $\hat{\beta}$ and $\hat{M}_3$ in a closed form using (18) and (19):

$$\hat{\beta} = m^{(1)}, \quad \hat{M}_3 = \overline{m}^{(3)}(1+k)\left(1+\frac{k}{2}\right).$$

Then we can obtain $\hat{q}$ by explicitly solving the equation $\hat{M}_3 = M_3(\hat{\beta},q)$ for q. This will work for $\lambda = \frac{1}{2}$ and 1, but not for $\lambda = 0$ where we always have $M_3(\beta,q) = 0$ (we explain why this happens by the end of this subsection). The next step is to use (19) to compute

$$\hat{M}_2 = M_2(\hat{\beta},\hat{q}), \quad \hat{s} = \overline{m}^{(2)} - \frac{\hat{M}_2}{1+k}.$$

The only estimator we cannot obtain directly is $\hat{\rho}$. For that we define

$$F(\rho) = \left(\hat{s} - k\frac{\hat{M}_2}{1-k^2}\right)e^{-\rho h} + \frac{\hat{M}_2}{1-k^2}e^{-k\rho h} - \overline{m}^{(4)}$$

and solve $F(\rho) = 0$ using MATLAB, which employs a variant of Powell's Trust-Region Dogleg method described in [19]. We employ the analytic derivative of $F(\rho)$, avoiding the approximation of the Jacobian by finite differences. We set the initial value arbitrarily at $\rho_0 = 0$. Finally, we obtain the rest of the estimators from (21): $\hat{\alpha} = k\hat{\rho}, \hat{v} = \sqrt{2\hat{\alpha}\hat{q}}, \hat{\sigma} = \sqrt{2\hat{\rho}\hat{s}}$. Notice that, depending on the actual values of the sample moments, we can obtain negative values for $\hat{\beta}$, $\hat{M}_3$, $\hat{s}$ and $\hat{\rho}$ using the previous procedure. Whenever that happens, we set the corresponding estimator to 10^{-5} and continue the process, but report the error at the end of our routine.

Table 3 shows the results of 1000 simulations of the CIR-drift model ($\lambda = \frac{1}{2}$) with $\alpha = 1.5$, $\beta = 2$, $v = 0.9$, $\rho = 6$, and $\sigma = 0.7$. Each entry represents the average value over all the simulations.

Out of these 1000 simulations, in 84 runs we obtained negative estimators for at least one of the parameters, as explained before. The frequency of this problem decreases with the increase in the length of the time interval. We still included the corresponding results in the averages reported. This implies that the average errors for the 916 runs that terminated cleanly are lower than what we are reporting here. The reason why we do not restrict ourselves to those 916 good runs is because with real data we will only have one run (only one set of sample moments) and we wanted to assess how well our procedure handles these corner cases on average.

The case of the OU-drift model ($\lambda = 0$) is particularly interesting because the observable process Y is Gaussian and stationary, hence its law is determined only by its first two moments and the auto-covariance function. This makes it impossible to estimate all the five parameters by the method of moments.

Table 3
1000 simulations with $h = 1.0$ on $t \in [0,10^5]$.

t	$\alpha = 1.5$	$\beta = 2.0$	$v = 0.9$	$\rho = 6.0$	$\sigma = 0.7$
	$\hat{\alpha}$ (rel.err.)	$\hat{\beta}$ (rel.err.)	$\hat{v}$ (rel.err.)	$\hat{\rho}$ (rel.err.)	$\hat{\sigma}$ (rel.err.)
1000	1.4843	2.0004	0.8795	5.9372	0.7519
	(7.00%)	(1.13%)	(5.76%)	(7.00%)	(25.84%)
10,000	1.4938	2.0002	0.8942	5.9752	0.7208
	(2.21%)	(0.38%)	(1.92%)	(2.21%)	(9.60%)
100,000	1.4949	2.0001	0.8960	5.9794	0.7178
	(0.73%)	(0.12%)	(0.70%)	(0.73%)	(3.58%)

Please cite this article as: Hernández J, et al. Algorithmic estimation of risk factors in financial markets with stochastic drift. Computers and Operations Research (2010), doi:10.1016/j.cor.2010.09.007

Table 4
1000 simulations with $h=1.0$ on $t\in[0,10^4]$.

	$L_1=1.9878$	$L_2=3.9769$	$\beta=2.00000$	$\rho=6.00000$
t	$m^{(1)}$ (rel.err.)	$m^{(2)}$ (rel.err.)	$\hat{\beta}$ (rel.err.)	$\hat{\rho}$ (rel.err.)
1000	1.9879 (0.21%)	3.9775 (0.41%)	2.0002 (0.24%)	6.0214 (4.65%)
10,000	1.9878 (0.07%)	3.9770 (0.13%)	2.0000 (0.08%)	6.0084 (1.48%)

7.3. Estimation with a cap

We are going to assume three of the parameters as known, and estimate the other two. In doing that simplification, we are going to introduce extra complexity using a price cap ($M<+\infty$). Assume an OU-drift model with a price cap M for which the parameters α, ν, σ, and hence q, are known. We want to estimate β and ρ given the observations $(X_{nh}=\max(Y_{nh},M))_{n\in\mathbb{N}}$. Since in this case Y_t is a Gaussian process with

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

we can compute $LM_1=\mathbb{E}X_t$ and $LM_2=\mathbb{E}X_t^2$ exactly. Denote

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

Then

$$LM_1=M+(\beta-M)\Phi\left(\frac{M-\beta}{\Gamma}\right)-\Gamma\varphi\left(\frac{M-\beta}{\Gamma}\right)$$

and

$$LM_2=M^2+(\beta^2+\Gamma^2-M^2)\Phi\left(\frac{M-\beta}{\Gamma}\right)-(\beta+M)\Gamma\varphi\left(\frac{M-\beta}{\Gamma}\right).$$

To obtain $\hat{\beta}$ and $\hat{\Gamma}$ we solve the two-dimensional nonlinear equation

$$F(\beta,\Gamma)=\begin{pmatrix} LM_1-m^{(1)} \\ LM_2-\overline{m}^{(2)} \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

using MATLAB's nonlinear solver. We provide MATLAB with the exact Jacobian of $F(\beta,\Gamma)$, and set the initial value arbitrarily at $(\beta_0,\rho_0)=(1,1)$, computing the corresponding (β_0,Γ_0). Finally, we obtain $\hat{\rho}$ transforming

$$\frac{\sigma^2}{2\rho}+\frac{q}{1+\frac{\alpha}{\rho}}-\hat{\Gamma}^2=0$$

into a polynomial equation, and using MATLAB's polynomial solver.[7]

Table 4 shows the results of 1000 simulations of the OU-drift model with a cap, with $\alpha=1.5$, $\beta=2$, $\nu=0.1$, $\rho=6$, $\sigma=0.6$, and $M=2.2$. Each entry represents the average value over all the simulations.

In these 1000 simulations, we did not obtain negative estimators for any of the parameters. Each run terminated cleanly without errors.

[7] In the case that we obtain negative values for $\hat{\beta}$ or $\hat{\rho}$, we set the corresponding estimator to 10^{-5} and continue the process, reporting the error at the end of our routine.

8. Concluding remarks

Diffusion processes reverting to a hidden stochastic mean have recently become important in financial modelling (interest rates and commodity prices). We present a simple method for calibrating such models using the method of moments, and also give a general procedure for deriving tractable moment equations, assuming only tractability of the moments of the hidden process. The case where the observable process is censored so that it is only observed when below a maximum level (important for applications in commodity price modelling) is also considered. Numerical simulations are presented to demonstrate the application of the method in practice.

Appendix A. Mixing coefficients, ergodicity and hidden Markov models

Let $\mathcal{G}_1,\mathcal{G}_2\subset\mathcal{F}$ be two $\sigma-$algebras. The following three measures of dependence between them can be defined (see for example [5]):

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

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 an S-valued 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

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

where π is the stationary distribution of X (see [5]). The following result (see [5]) states that the mixing conditions are stronger than ergodicity.

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

The following definition is based on [14]. Let $(S,\mathbb{B}(S))$ and $(T,\mathbb{B}(T))$ be two Polish spaces equipped with their Borel $\sigma-$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 $(U_1,U_2,\dots,U_n)$, 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*. The following result is from [14,10].

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 $\rho_Z(n)\le\rho_U(n)$.*

Please cite this article as: Hernández J, et al. Algorithmic estimation of risk factors in financial markets with stochastic drift. Computers and Operations Research (2010), doi:10.1016/j.cor.2010.09.007

Appendix B. Proofs of auxiliary results

Lemma B.1. *For any $0 \le s \le t$, and any $0 \le m \le p$ we have that*

$$e^{r_m t} v_t^m = e^{r_m s} v_s^m + p_m \int_s^t e^{r_m u} [a v_u^{m-2} + b v_u^{m-1}]\, du + [N_t^{(m)} - N_s^{(m)}], \qquad (22)$$

where $N_t^{(m)}$ is a $\mathcal{F}_t$-martingale, $p_m = \frac{1}{2} m(m-1)\nu^2$, $r_m = m\alpha - \delta p_m$, 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

$$\begin{aligned} d[e^{r_m t} v_t^m] &= e^{r_m t}\left[r_m v_t^m\, dt + m v_t^{m-1} dv_t + \frac{m(m-1)}{2} v_t^{m-2} d\langle v \rangle_t \right] \\ &= e^{r_m t}[r_m v_t^m\, dt - m\alpha v_t^m\, dt + m\nu v_t^{m-1}(\beta+v_t)^{\lambda}\, dW_t \\ &\quad + p_m v_t^{m-2}(\beta+v_t)^{2\lambda}\, dt]. \end{aligned}$$

For $\lambda = 0, \frac{1}{2}, 1$, all the terms containing v_t^m on the right-hand side cancel 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 (22). □

Proof of Lemma 5.1. Clearly $M_0 = 1$. Taking expectations on (22) and using the stationarity of v we obtain for $m=1$ that $M_1 = 0$ (since $p_1 = 0$). For $m \ge 2$ we obtain the recursive relation $M_m = (p_m/r_m)[aM_{m-2} + bM_{m-1}]$. Noticing that $p_m/r_m = p_m/(m\alpha - \delta p_m) = (1/(m-1)\nu^2/2\alpha - \delta)^{-1}$ we easily obtain the desired expression. □

Proof of Lemma 5.3. Using that $p_1 = 0$ and $r_1 = \alpha$ we can write (22) for $m=1$ as

$$e^{\alpha s_2} v_{s_2} = e^{\alpha s_1} v_{s_1} + [N_{s_2}^{(1)} - N_{s_1}^{(1)}]. \qquad (23)$$

Multiplying by v_{s_1} and taking expectations we obtain

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

But $\mathbb{E}[v_{s_1}(N_{s_2}^{(1)} - N_{s_1}^{(1)})] = \mathbb{E}[v_{s_1} \mathbb{E}[N_{s_2}^{(1)} - N_{s_1}^{(1)} | \mathcal{F}_{s_1}]] = 0$. Hence $\mathbb{E}[v_{s_1} v_{s_2}] = e^{-\alpha(s_2 - s_1)} M_2$. Similarly, rewriting (23) with s_2 and s_3, multiplying by $v_{s_1} v_{s_2}$, and taking expectations we get

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

thus $\mathbb{E}[v_{s_1} v_{s_2} v_{s_3}] = e^{-\alpha(s_3 - s_2)} \mathbb{E}[v_{s_1} v_{s_2}^2]$. To obtain this last expectation we first write (22) for $m=2$ as

$$e^{r_2 s_2} v_{s_2}^2 = e^{r_2 s_1} v_{s_1}^2 + p_2 \int_{s_1}^{s_2} e^{r_2 u} [a v_u^0 + b v_u^1]\, du + [N_{s_2}^{(2)} - N_{s_1}^{(2)}].$$

Multiplying by v_{s_1} and taking expectations we obtain

$$e^{r_2 s_2} \mathbb{E}[v_{s_1} v_{s_2}^2] = e^{r_2 s_1} \mathbb{E}[v_{s_1}^3] + p_2 \int_{s_1}^{s_2} e^{r_2 u} (a \mathbb{E}[v_{s_1}] + b \mathbb{E}[v_{s_1} v_u])\, du$$

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

$$\begin{aligned} e^{r_2 s_2} \mathbb{E}[v_{s_1} v_{s_2}^2] &= e^{r_2 s_1} M_3 + p_2 b M_2 e^{\alpha s_1} \int_{s_1}^{s_2} e^{(r_2 - \alpha) u}\, du \\ &= e^{r_2 s_1}\left(M_3 - \frac{p_2 b M_2}{r_2 - \alpha} \right) + e^{\alpha s_1 + (r_2 - \alpha) s_2} \frac{p_2 b M_2}{r_2 - \alpha}. \end{aligned}$$

Lemma 5.1 and a little algebra show that $p_2 b M_2 / (r_2 - \alpha) = M_3$. Then

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

References

[1] Balduzzi P, Bertola G, Foresi S. A model of target changes and the term structure of interest rates. Journal of Monetary Economics 1997;39:223–49.

[2] Balduzzi P, Bertola G, Foresi S, Klapper L. Interest rate targeting and the dynamics of short-term rates. Journal of Money, Credit, and Banking 1998;30: 26–50.

[3] Balduzzi P, Das S, Foresi S. The central tendency: a second factor in bond yields. Review of Economics and Statistics 1998;80:62–72.

[4] Balduzzi P, Das S, Foresi S, Sundaram R. Stochastic mean models of the term structure of interest rates. In: Tuckman B, Jegadeesh N, editors. Advanced tools for the fixed-income professional. John Wiley and Sons; 2000.

[5] Bradley R. Basic properties of strong mixing conditions. In: Eberlein E, Taqqu M, editors. Dependence in probability and statistics, a survey of recent results. Boston: Birkhäuser; 1986.

[6] Brigo D, Mercurio F. Interest rate models—theory and practice: with smile, inflation and credit. second ed.. Springer; 2007.

[7] Cappé O, Moulines E, Rydén T. Inference in hidden Markov models. Springer; 2007.

[8] Cox J, Ingersoll J, Ross S. A theory of the term structure of interest rates. Econometrica 1985;53:385–407.

[9] Fearnhead P, Papaspiliopoulos O, Roberts G. Particle filters for partially observed diffusions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2008;70(4):755–77.

[10] Genon-Catalot V, Jeantheau T, Larédo C. Stochastic volatility models as hidden Markov models and statistical applications. Bernoulli 2000;6: 1051–79.

[11] Guyon X, Iovleff S, Yao J-F. Linear diffusion with stationary switching regime. ESAIM Probability and Statistics 2004;8:25–35.

[12] Krengel UJ. Ergodic theorems. Berlin: Walter de Gruyter & Co.; 1985.

[13] Lari-Lavassani A, Sadeghi AA, Ware A. Modelling and implementing mean reverting price processes in energy markets. Electronic Publications of the International Energy Credit Association 〈www.ieca.net〉.

[14] Leroux BG. Maximum-likelihood estimation for hidden Markov models. Stochastic Processes and their Applications 1992;40:127–43.

[15] Luo C, Seco L, Wang H, Wu D. Risk modeling in crude oil market: a comparison of Markov switching and GARCH models. Kybernetes 2010;39: 750–69.

[16] Movellan J, Mineiro P, Williams R. A Monte Carlo EM approach for partially observable diffusion processes: theory and applications to neural networks. Neural Computation 2002;14:1507–44.

[17] Nelson D. ARCH models as diffusion approximations. Journal of Econometrics 1990;45:7–38.

[18] Nielsen J, Madsen H, Melgaard H. Estimating parameters in discretely, partially observed stochastic differential equations. Available at 〈http://www2.imm.dtu.dk/documents/ftp/tr00/tr07_00.pdf〉.

[19] Powell MJD. A FORTRAN subroutine for solving systems of nonlinear algebraic equations. In: Rabinowitz P, editor. Numerical methods for nonlinear algebraic equations. Gordon and Breach Science Publishers; 1970. [chapter 7].

[20] Rogers L, Williams D. Diffusions, Markov processes and martingales. Itô calculus, vol. 2. 2nd ed.. Cambridge: Cambridge University Press; 2000.

[21] Schwartz E. The stochastic behavior of commodity prices: implications for valuation and hedging. Journal of Finance 1997;52:923–73.

[22] Walton B. Analysis of single-molecule kinesin assay data by hidden Markov model filtering. PhD thesis, University of Arizona; 2002.

[23] Wu D, Olson D. Introduction to the special section on "Optimizing risk management: methods and tools". Human and Ecological Risk Assessment: An International Journal 2009;15:220–6.

Please cite this article as: Hernández J, et al. Algorithmic estimation of risk factors in financial markets with stochastic drift. Computers and Operations Research (2010), doi:10.1016/j.cor.2010.09.007