Markov Models for Commodity Futures: Theory and Practice

Leif Andersen
Banc of America Securities*

Abstract

This objective of this paper is to develop a generic, yet practical framework for construction of Markov models for commodity derivatives. We aim for sufficient richness to permit applications to a broad variety of commodity markets, including those that are characterized by seasonality and by spikes in the spot process. In the first, largely theoretical, part of the paper we derive a series of useful results about low-dimensional Markov representation of the dynamics of an entire term structure of futures prices. Extending previous results in the literature, we cover jump-diffusive models with stochastic volatility as well as several classes of regime-switch models. To demonstrate the process of building models for a specific commodity market, the second part of the paper applies a selection of our theoretical results to the exercise of constructing and calibrating derivatives trading models for USD natural gas. Special attention is paid to the incorporation of empirical seasonality effects in futures prices, in implied volatilities and their "smile", and in correlations between futures contracts of different maturities. European option pricing in our proposed gas model is closed-form and of the same complexity as the Black-Scholes formula.

First Version: June 14, 2007. This Version: September 30, 2008.

1 Introduction

The classical approach to derivatives pricing in commodity markets involves the specification of a spot price process for the commodity in question, often of the mean-reverting diffusive type (see [8], [34] and [5], among others, for empirical evidence justifying the usage of mean reversion). Examples of such models include [23], [37], [38],[21], [31], [32], [4] to name just a few. Non-diffusive jump- or spike-behavior is quite common for some commodities, having lead numerous authors to advocate the inclusion of Poisson jumps or regime switch behavior in models for the spot commodity price; some examples in the literature include [15], [22], [3], and [25]. To additionally model the phenomenon of "inverse leverage" where price volatility is inversely related to commodity prices, one can contemplate randomizing volatility and making it positively correlated to the spot price. For models and empirical studies focusing on this idea, see, for instance, [35], [36] and [39].

Commodity spot price modeling is quite natural and can lead to interesting economic insights. From the perspective of the financial engineer interested in contingent claims

*leif.andersen@bofasecurities.com. I am indebted to Jesper Andreasen, Alex Levin, and John Crosby for their comments and suggestions.

pricing, however, the approach is not particularly convenient. First, spot models often rely on assumptions about the behavior and magnitude of the so-called *convenience yield*, an unobservable quantity that complicates estimation of the model. The convenience yield is, of course, reflected in the term structure of futures prices – the *futures curve* – but this curve is endogenous to spot models and may be hard to calibrate to market-observed values. This is particularly true for those commodity markets where futures prices are strongly seasonal, e.g. electricity and natural gas, which will require convenience yields to be oscillating functions of calendar time. In derivatives pricing applications, a more natural way is to build models that treat the entire futures curve as the underlying stochastic object on which dynamics are imposed. In this type of model, the futures curve – seasonality effects and all – is fully exogenous and thereby automatically in calibration with the market. Examples of such full term structure models can be found in [28], [11], [10], and [13], among others. Of course, any futures curve model will uniquely imply dynamics of the spot process, and vice versa – the difference in the two approaches is primarily a matter of perspective and convenience[1].

Despite the substantial body of existing literature on commodities modeling, work still remains in the construction of truly satisfactory models for derivatives pricing purposes. In particular, what is missing is a framework that allows for accurate modeling of the most salient commodity price characteristics, yet is easy to implement numerically and has sufficient analytical tractability to allow for quick calibration to option prices and other market data. Of those "salient characteristics", we highlight spikes and jumps; inverse leverage; and seasonality. We should emphasize that seasonality to us means more than just the ability to perfectly match a possibly oscillating futures curve at time 0: we also require the ability to match seasonality in at-the-money volatilities, in futures curve correlations, and even in the implied volatility smile. A step in this direction was taken by the interesting work of [7] who uses a purely diffusive log-normal futures curve model with 5 state-variables to approximately match the highly seasonal term structure of at-the-money volatilities in USD natural gas. In the second (applied) part of this paper we extend the analysis in [7] in several ways, and in the process tighten up the calibration fit, cut down on the number of state variables, and allow the model to generate a non-flat (and seasonal) implied volatility smile.

The basic model framework employed in this paper involves a multi-factor stochastic volatility model for the entire futures curve, overlaid with Poisson jumps and/or regime-switch spikes. For reasons of practicality, we stick to a setting where the state of the futures curve evolution can be captured uniquely as functions of a finite-dimensional vector of Markov state variables. Sections 2, 3, 4, and 5 of the paper together lay out our generic model framework, with a special emphasis on the parameter conditions required to ensure that the model is Markov, in the sense just described. Along the way, we list a number of supporting results about European option pricing and about the choice of implementation-friendly state variables. The results in Sections 2-5 generalize results from [6] to a non-

[1] A similar split exists in interest derivative applications, where traditional short-rate models have largely been superseded by models that operate directly on forward curve dynamics (e.g. the famous HJM model [27]).

diffusive setting and also extends the recent results in [13] to local volatility, stochastic volatility, regime-switching, and to a broader class of jump-diffusions.

The framework outlined in Sections 2-5 is quite general, and in practice decisions about the number of state-variables and their processes must be made. These decisions will be driven both by the specific application at hand and by the type of commodity in question. To exemplify the process of specific model parameterization, in Sections 6, 7, and 8 we turn to the concrete problem of constructing a derivatives trading model for natural gas. Letting empirical data guide our steps, we construct a practical models for natural gas, gradually increasing model complexity as we go along. Throughout this exercise, we pay particular attention to calibration strategies, to issues of time-stationarity, and to seasonality effects in correlations and implied volatilities. Section 9 concludes the paper and suggests directions for future research.

2 Diffusive Futures Curve Dynamics

2.1 Basic Setup and Results

We consider a probability space $(\Omega, \mathscr{F}, \mathrm{Q})$, equipped with a filtration $\{\mathbb{F}_t\}_{t\geq 0}$ of the sigma-algebra $\mathscr{F}$. Let $F(t,T)$ be the time t futures price for delivery at time T. In the risk-neutral measure Q, if $F(t,T)$ is a martingale it is known that there will be no arbitrage (see [18] for the proof). Assuming first that futures prices are Ito processes, we write

$$dF(t,T)/F(t,T) = \sigma(t,T)^{\top} dW(t), \tag{1}$$

where $W(t)$ is an N-dimensional Brownian motion and $\sigma : \mathbb{R}\times\mathbb{R} \to \mathbb{R}^N$ is a vector-valued, $\mathbb{F}_t$-adapted volatility process for all $T \geq t$. Define $X(t,T) = \ln F(t,T)$, such that, by Ito's Lemma,

$$dX(t,T) = -\frac{1}{2}|\sigma(t,T)|^2 dt + \sigma(t,T)^{\top} dW(t). \tag{2}$$

As discussed above, we are here interested in the case where $X(t,T)$ (and thereby $F(t,T)$) can be reconstituted from a low-dimensional Markov process. Following the idea of [9], we first make the following key assumption.

Assumption 1 *Define an $\mathbb{F}_t$-adapted $N\times N$ square-integrable stochastic process $\alpha : \mathbb{R} \to \mathbb{R}^{N\times N}$ and a deterministic $N\times 1$ function $\beta : \mathbb{R} \to \mathbb{R}^N$. Assume that (1) holds, with*

$$\sigma(t,T) = \alpha(t)\beta(T).$$

If $\sigma(t,T) = \sigma(T-t)$ – that is, the futures volatilities are only a function of time *to* maturity $T-t$, rather than of calendar time t and time *of* maturity T – we say that the volatility term structure is *stationary*. Many commodities display seasonality in implied option volatilities which is inconsistent with stationarity. The ability to separately manipulate the dependence of futures price volatility $\sigma(t,T)$ on t (through α) and T (through β) is, as we shall demonstrate in Section 7, quite useful in the modeling of seasonal behavior.

Applying Assumption 1 to (2) and integrating, it follows that

$$X(t,T) = X(0,T) - \frac{1}{2}\beta(T)^\top y(t)\beta(T) + \beta(T)^\top x(t), \tag{3}$$

where we have defined

$$x(t) \triangleq \int_0^t \alpha(u)^\top dW(u), \quad y(t) \triangleq \int_0^t \alpha(u)^\top \alpha(u)\, du. \tag{4}$$

The following result then immediately follows.

Proposition 1 *Let Assumption 1 hold, and assume further that* $\alpha(t) = f(t,x(t),y(t))$ *for a deterministic function* $f : \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$. *Together,* $x(t)$ *and* $y(t)$ *in (4) then form a Markov SDE system*

$$dx(t) = f(t,x(t),y(t))^\top\, dW(t), \tag{5}$$
$$dy(t) = f(t,x(t),y(t))^\top\, f(t,x(t),y(t))\, dt, \tag{6}$$

where $x(0) = y(0) = 0$ *and the futures curve can be reconstituted as*

$$F(t,T) = F(0,T)\exp\left(-\frac{1}{2}\beta(T)^\top y(t)\beta(T) + \beta(T)^\top x(t)\right), \quad T \geq t. \tag{7}$$

Remark 1 *As* $x(t)$ *contains* N *elements and the symmetric matrix* $y(t)$ *contains* $\frac{1}{2}N(N+1)$ *distinct elements, (5)-(6) generally involves* $\frac{1}{2}N(N+3)$ *different state variables.*

2.2 Mean Reverting State Variables

The primary stochastic driver of the futures curve in Proposition 1 is the $x(t)$ vector, with the locally deterministic matrix $y(t)$ contributing only a "convexity correction" in the reconstitution formula (7). In practical applications, the volatility of $x(t)$ will often exhibit exponential growth in its volatility as a function of t, a feature that may complicate numerical work. It is therefore often useful to change variables such that the primary stochastic driver is mean-reverting. To show one such re-parameterization, we first define two diagonal $N \times N$ matrices,

$$B(t) \triangleq \mathrm{diag}\left(\beta(t)\right), \quad \kappa(t) \triangleq -\frac{dB(t)}{dt}B(t)^{-1}, \tag{8}$$

where we have assumed that $B(t)$ is invertible. Note that $\beta(t) = B(t)\mathbf{1}$, with $\mathbf{1}$ being a column vector of 1's.

Lemma 1 *In the setup of Proposition 1, set (with* $B(t)$ *invertible)*

$$x^*(t) \triangleq -\frac{1}{2}B(t)y(t)\beta(t) + B(t)x(t), \quad y^*(t) \triangleq B(t)y(t)B(t), \tag{9}$$

such that $\alpha(t)B(t) = g(t,x^*(t),y^*(t))$ *for a deterministic function* $g : \mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$. *Together,* $x^*(t)$ *and* $y^*(t)$ *in (9) form a Markov SDE system*

$$dx^*(t) = \left(-\kappa(t)x^*(t) + \frac{1}{2}y^*(t)^\top\kappa(t)\mathbf{1} - \frac{1}{2}G(t)\mathbf{1}\right)dt + g(t,x^*(t),y^*(t))^\top dW(t), \quad (10)$$

$$dy^*(t) = (-2\kappa(t)y^*(t) + G(t))\,dt, \quad (11)$$

where $G(t) \triangleq g(t,x^*(t),y^*(t))^\top g(t,x^*(t),y^*(t))$; B *and* κ *are given in (8); and* $x^*(0) = y^*(0) = 0$. *The futures curve can be reconstituted as (for* $T \geq t$*)*

$$F(t,T) = F(0,T)\exp\left\{\beta(T)^\top B(t)^{-1}\left(\frac{1}{2}y^*(t)\left(\mathbf{1} - B(t)^{-1}\beta(T)\right) + x^*(t)\right)\right\}. \quad (12)$$

Proof: As $B(t)$ is assumed invertible, we can invert (9) to get

$$x(t) = B(t)^{-1}x^*(t) + \frac{1}{2}B(t)^{-1}y^*(t)\mathbf{1}, \quad y(t) = B(t)^{-1}y^*(t)B(t)^{-1}.$$

Insertion into (7) and rearranging yields (12). The dynamics for y^* follow from Ito's Lemma and the fact that B and its derivative are symmetric matrices. As $x^*(t) = -\frac{1}{2}y^*(t)\mathbf{1} + B(t)x(t)$ another application of Ito's Lemma yields the dynamics for x^*. ∎

Remark 2 *The functions* f *and* g *in Proposition 1 and Lemma 1 are related as*

$$f(t,x(t),y(t)) = g\left(t, -\frac{1}{2}B(t)y(t)\beta(t) + B(t)x(t), B(t)y(t)B(t)\right)B(t)^{-1}.$$

The choice of state variables $x^*(t)$ and $y^*(t)$ ensures a particularly simple expression for the spot commodity price $F(t,t)$ (see Section 2.3), but is not necessarily the most convenient choice. One concern is the presence of time-derivatives of B in the drift-term for $x^*(t)$, through the term $-\kappa(t)x^*(t)$. In applications where there is considerable seasonality in futures volatility, $B(t)$ may have oscillating components which, in turn, may cause the mean reversion matrix $\kappa(t)$ to be irregular.

In practical applications, the most reasonable approach often involves estimating an "average" time-independent mean-reversion matrix c, which serves as a proxy for $\kappa(t)$. Specifically, define a matrix E as the matrix exponential of $-ct$, $E(t) = \exp(-ct)$, such that $c = -E'(t)E(t)^{-1}$. Then, rather than working directly with $x(t)$ in Proposition 1, we can use as our preferred state variables $z(t) = E(t)x(t)$ which implies that

$$dz(t) = -cE(t)x(t)\,dt + E(t)\,dx(t) = -cz(t)\,dt + E(t)\alpha(t)^\top dW(t).$$

Writing $x(t) = E(t)^{-1}z(t)$ in (7) restates the futures price reconstitution formula in terms of z; details are left to the reader. When properly dimensioned, the choice of state variables $z(t)$ will transfer seasonality effects from the dynamics of x^*, y^* into the futures curve reconstitution formula, thereby increasing numerical robustness.

2.3 Spot Price Dynamics

As explained earlier, our focus in this paper is on models for the entire futures curve, rather than on spot commodity models. That said, to gain intuition let us quickly look at what type of spot dynamics are implied by the models considered so far.

Lemma 2 *Define the spot price $S(t) = F(t,t)$ and assume that the assumptions of Lemma 1 hold true. Then, with G and x^* defined as in Lemma 1,*

$$S(t) = F(0,t)\exp\left(\mathbf{1}^\top x^*(t)\right), \tag{13}$$

$$dS(t)/S(t) = \left(\frac{\partial \ln F(0,t)}{\partial t} + \frac{1}{2}\mathrm{tr}\left(G\left(t,x^*(t),y^*(t)\right)\right)\right)dt + \mathbf{1}^\top dx^*(t). \tag{14}$$

Proof: (13) follows from (12) and the observation that $B(t)^{-1}\beta(t) = B(t)^{-1}B(t)\mathbf{1} = \mathbf{1}$. The expression for dS is a consequence of Ito's Lemma. ∎

The expression (14) allows us to work out a stochastic convenience yield $y_c(t)$ by equating the drift of (14) with $r(t) - y_c(t)$, with $r(t)$ being the risk-free interest rate.

We note in passing that representation of the dynamics of S in terms of the state variables x and y (or z and y) is, of course, also possible. We leave details to the reader.

2.4 Example: Deterministic Volatility

The simplest case of a model covered by Proposition 1 and Lemma 1 involves setting the functions f and g to be independent of the state variables $x(t)$ (or, equivalently, of $x^*(t)$). In this case, it is easily seen that $y(t)$ (or $y^*(t)$) become deterministic, whereby both f and g depend only on calendar time t. From Assumption 1, it follows that then all futures volatilities $\sigma(t,T)$ are themselves deterministic, and futures prices therefore log-normal in measure Q. This, in turn, allows for significant analytical tractability in the futures curve dynamics. For instance, we highlight the following (trivially proven) Black-Scholes result, useful for calibration of the model to option prices.

Lemma 3 *Let the assumptions in Proposition 1 hold, and assume that $f(t,x,y) = f(t)$, i.e. f depends only on time t. Consider a K-strike European call option on a T-maturity future, paying $(F(T',T) - K)^+$ at the option maturity T', where $T' \leq T$ and $K > 0$. Let the risk-free interest rate be independent of $W(t)$, and let the time 0 discount factor to time T' be $P(0,T')$. Then the time 0 arbitrage-free value of the call option is*

$$C(0) = P(0,T')\left\{F(0,T)\Phi\left(d_+\left(T',T\right)\right) - K\Phi\left(d_-\left(T',T\right)\right)\right\}, \tag{15}$$

with

$$d_\pm(T',T) \triangleq \frac{\ln\left(F(0,T)/K\right) \pm \frac{1}{2}v(T',T)^2}{v(T',T)}, \quad v(T',T)^2 \triangleq \beta(T)^\top \int_0^{T'} f(t)^\top f(t)\,dt\,\beta(T).$$

For the deterministic case, it follows from the results in Section 2.2 that the spot price $S(t)$ can be written in the form

$$S(t) = F(0,t)\exp\left(\sum_{i=1}^{N} x_i^*(t)\right),$$

where the N scalar processes $x_i^*(t)$ all follow Ornstein-Uhlenbeck processes with time-dependent mean reversions, volatilities, and long-term means. We recognize this as an extension of the classical two-factor Gaussian spot model in, say, [38] to N factors and to time-dependence in all parameters. As it turns out, such time-dependence is critical for the model calibration to seasonal markets – although such a calibration is considerably more intuitive and convenient when the model has been cast in futures curve space first. See Section 7 for a calibration example. For development of the N-factor deterministic futures curve model in a setting with correlated Brownian motions[2], see [13].

Remark 3 *For the deterministic case, the state-variables y (or y^*) are non-random and can be stricken from the list of random state variables to be accounted for. The total number of Markov state variables will therefore only be N (compare to Remark 1).*

2.5 Example: Local Volatility

In the setup of, say, Lemma 1, local volatility (that is, the existence of a functional dependence between volatility and the level of the futures curve) can be expressed through the chosen dependence of the $N \times N$ matrix-valued function $g(t,x^*,y^*)$ on the state variables $x^*(t)$ and $y^*(t)$. In the exogenous specification of such dependence, it is often most user-friendly to treat g as a function of a set of N "rolling" futures $F(t,t+\tau_i)$ with distinct relative maturities τ_i, $i = 1,\ldots,N$. That is, we write

$$g(t,x^*(t),y^*(t)) = u(t,F(t,t+\tau_1),\ldots,F(t,t+\tau_N)) \tag{16}$$

for some deterministic function $u : \mathbb{R}\times\mathbb{R}^N \to \mathbb{R}^{N\times N}$. Given u, the function g can be uncovered directly from the reconstitution formula (12) which explicitly links each $F(t,t+\tau_i)$ to $x^*(t)$ and $y^*(t)$.

In practice, there are several meaningful ways of specifying u. To give an example procedure, we can follow the idea (from an interest rate setting) of [1] and specify, for each of the N rolling futures, dedicated local volatility functions $\zeta_i(t,F(t,t+\tau_i))$ and deterministic volatility vectors $\eta_i(t)$. Specifically, we can assume that

$$dF(t,t+\tau_i)/F(t,t+\tau_i) = \mathscr{O}(dt) + \zeta_i(t,F(t,t+\tau_i))\,\eta_i(t)^\top dW(t), \quad i = 1,\ldots,N, \tag{17}$$

where $\zeta_i : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is a smooth local volatility function, $\eta_i(t)$ is an $N\times 1$ deterministic vector, and $\mathscr{O}(dt)$ is a drift-term of order dt. The specification (17) can be used to determine the functions u and g, as follows.

[2] Of course, our usage of a standard (uncorrelated) vector Brownian motion $W(t)$ involves no loss of generality as a suitable rotation (e.g. by a Cholesky matrix) can induce any correlation structure.

Lemma 4 *Let the assumptions of Lemma 1 hold. Define three $N \times N$ matrices*

$$\Psi(t) = \begin{pmatrix} \beta(t+\tau_1)^\top \\ \vdots \\ \beta(t+\tau_N)^\top \end{pmatrix}, \quad \zeta(t) = \operatorname{diag} \begin{pmatrix} \zeta_1(t, F(t,t+\tau_1)) \\ \vdots \\ \zeta_N(t, F(t,t+\tau_N)) \end{pmatrix}, \quad \eta(t) = \begin{pmatrix} \eta_1(t)^\top \\ \vdots \\ \eta_N(t)^\top \end{pmatrix},$$

and assume that $\Psi(t)$ is invertible. If we set the function $g(t,x^*,y^*)$ in Lemma 1 according to (16) with

$$u(t, F(t,t+\tau_1), \ldots, F(t,t+\tau_N))^\top = B(t)\Psi(t)^{-1}\zeta(t)\,\eta(t),$$

then (17) holds for all $i = 1, \ldots, N$.

Proof: Combining Lemma 1 and (16) yields, for $i = 1, \ldots, N$,

$$dF(t,t+\tau_i)/F(t,t+\tau_i) = \mathcal{O}(dt) + \beta(t+\tau_i)^\top B(t)^{-1} u(t, F(t,t+\tau_1), \ldots, F(t,t+\tau_N))^\top \, dW(t).$$

Setting the diffusion term of this equation equal to that of (17) proves the lemma. ∎

For sufficiently simple choice of the ζ_i, there may be tractability – at least approximately – in option pricing problems. We refer to [1] for further details.

Another, somewhat cruder, idea for local volatility is to replace (1) with a simple displaced diffusion form

$$dF(t,T) = (k_1(T)F(t,T) + k_2(T))\,\sigma(t,T)^\top dW(t) \tag{18}$$

where k_1 and k_2 are deterministic scalar functions of T, with $k_1(T) > 0$. With $Y(t,T) \triangleq \ln(k_1(T)F(t,T) + k_2(T))$ we get

$$dY(t,T) = -\frac{1}{2}k_1(T)^2\,|\sigma(t,T)|^2\,dt + k_1(T)\sigma(t,T)^\top dW(t)$$

which is of the same form as (2) and allows for a Markov representation under conditions identical to those of Proposition 1 and Lemma 1. When reconstituting the futures curve from $Y(t,T)$, we should here use the relation

$$F(t,T) = \left(e^{Y(t,T)} - k_2(T)\right)/k_1(T).$$

We note that (18) allows for a closed-form option pricing result when $\sigma(t,T)$ is deterministic, an obvious consequence of the fact that

$$\left(F(T',T) - K\right)^+ = \frac{1}{k_1(T)}\left(e^{Y(T',T)} - K'\right)^+, \quad K' = k_2(T) - Kk_1(T),$$

with $\exp(Y(T',T))$ being log-normal. We leave the details to the reader. While tractability is a plus for (18), note that displaced diffusion implies that $F(t,T) \in [-k_2(T)/k_1(T), \infty)$. Specifically, if k_2 and k_1 have the same signs, $F(t,T)$ can go negative; if they have opposite signs, $F(t,T)$ is bounded by a positive number from below.

2.6 Some Notes on Matrix Dimensions

In our setup (see Assumption 1), the dimension of the Brownian motion N is the same as that of the vector β and of the square matrix α. In some applications, however, the dimension of β and α may be larger than that of the Brownian motion[3]. No additional notation is required for this case, however, which can be handled by a suitable amendment of α with empty rows. Appendix A illustrates the procedure, using the natural gas model from [6] and [7] as an example. Appendix A also discusses various strategies for state-variable selection and demonstrates that the choice in [6] and [7] is not minimal, in the sense that the dimension of their state variable vector is larger than necessary.

3 Stochastic Volatility

3.1 Unspanned Stochastic Volatility

In Section 2, we showed that if $\sigma(t,T)$ is (say) of the form $g(t,x^*,y^*)B(t)^{-1}\beta(T)$, with g being a matrix-valued deterministic function of x^* and y^* in (10)-(11), then a finite-dimensional Markov representation of the futures curve is possible. It is not difficult to extend this result to allow for dependence on additional $\mathscr{F}_t$-adapted random variables in the function g. We list the following obvious result without proof.

Lemma 5 *Let $V(t)$ be an M-dimensional adapted stochastic process, and suppose that Assumption 1 holds. Further, define x^* and y^* as in (9) and assume that $\alpha(t)B(t) = h(t,x^*(t),y^*(t),V(t))$ for a deterministic function $h : \mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^{N\times N}\times\mathbb{R}^M \to \mathbb{R}^{N\times N}$. The dynamics for $x^*(t)$ and $y^*(t)$ are then*

$$\begin{aligned} dx^*(t) &= \left(-\kappa(t)x^*(t) + \frac{1}{2}y^*(t)^\top\kappa(t)\mathbf{1} - \frac{1}{2}H(t)\mathbf{1}\right)dt + h(t,x^*(t),y^*(t),V(t))^\top dW(t), \\ dy^*(t) &= (-2\kappa(t)y^*(t) + H(t))\,dt, \end{aligned} \tag{19}$$

where $H(t) \triangleq h(t,x^*(t),y^*(t),V(t))^\top h(t,x^*(t),y^*(t),V(t))$ and $x^*(0) = y^*(0) = 0$. The futures curve can be reconstituted as (for $T \geq t$)

$$F(t,T) = F(0,T)\exp\left\{\beta(T)^\top B(t)^{-1}\left(\frac{1}{2}y^*(t)\left(\mathbf{1} - B(t)^{-1}\beta(T)\right) + x^*(t)\right)\right\}. \tag{20}$$

If the joint process $(x^*(t),y^*(t),V(t))$ is Markov, then[4] Lemma 5 demonstrates that a finite-dimensional Markov representation of the futures curve is possible. A standard specification of such a Markov model is covered in Section 3.2 below; for now, we just notice that the total number of state-variables will be $M + \frac{1}{2}N(N+3)$.

The astute reader will have noticed that the process $V(t)$ does not figure in the futures curve reconstitution formula (20) and as such represents true *unspanned stochastic volatility* variables, as defined in [2].

[3] Loosely speaking this happens when there are more mean reversion speeds than Brownian motions.

[4] Lemma 5 can, if convenient, be restated using deterministic transformation of the state-variables x^* and y^*, as in Sections 2.1 and 2.2.

3.2 Example: Heston Type Models

While $V(t)$ in Lemma 5 may be any type of well-behaved Markov process, in practical applications of Lemma 5 we often set

$$h(t,x^*(t),y^*(t),V(t)) = \sqrt{V(t)} \cdot \psi(t,x^*(t),y^*(t)),$$

for some $N \times N$ matrix-function ψ, and $V(t)$ a scalar affine diffusion process

$$dV(t) = \kappa_V(t)\,(\vartheta_V(t) - V(t))\,dt + \varepsilon_V(t)\sqrt{V(t)}\,dW_V(t), \tag{21}$$

where $\kappa_V, \vartheta_V, \varepsilon_V$ are deterministic functions of time[5] and W_V is a scalar Brownian motion.

The local volatility function ψ can be set along the lines of Section 2.5. In some cases – e.g. if ψ is chosen to be deterministic – the resulting futures price processes are tractable. To expand on this, let us introduce the characteristic function (i is here the complex unit, $i^2 = -1$)

$$\phi(\omega,T',T) = \mathrm{E}\left(e^{i\omega \ln F(T',T)}\right), \quad \omega \in \mathbb{R}, \quad T' \leq T. \tag{22}$$

Proposition 2 *In Lemma 1, assume that $h(t,x^*,y^*,V) = \sqrt{V} \cdot \psi(t)$, where V satisfies (21) and ψ only depends on time. Let the correlation between W_V and W_i (the ith component of W) be a deterministic function $\rho_i(t)$, $i = 1,\ldots,N$, and let $\rho(t) \triangleq (\rho_1(t),\ldots,\rho_N(t))^\top$. Setting $\sigma_F(t,T) \triangleq \psi(t)B(t)^{-1}\beta(T)$, the characteristic function (22) is then*

$$\phi(\omega,T',T) = e^{a(0;\omega,T',T) + b(0;\omega,T',T)V(0) + i\omega \ln F(0,T)},$$

where $a(t) = a(t;\omega,T',T)$ and $b(t) = b(t;\omega,T',T)$ satisfy Riccati ODEs

$$\frac{da(t)}{dt} = -\kappa_V(t)\vartheta_V(t)b(t),$$
$$\frac{db(t)}{dt} = \frac{1}{2}|\sigma_F(t,T)|^2\,\omega(i+\omega) - \left(\varepsilon_V(t)\rho(t)^\top\sigma_F(t,T)i\omega - \kappa_V(t)\right)b(t) - \frac{1}{2}\varepsilon_V(t)^2 b(t)^2,$$

subject to the terminal condition $a(T';\omega,T',T) = b(T';\omega,T',T) = 0$.

Proof: Under the stated assumptions, the futures price process is

$$dF(t,T)/F(t,T) = \sqrt{V(t)}\psi(t)B(t)^{-1}\beta(T)\,dW(t) = \sqrt{V(t)}\sigma_F(t,T)^\top dW(t),$$

with the co-variation between $\varepsilon_V(t)dW_V(t)$ and $\sigma_F(t,T)^\top dW(t)$ being $\varepsilon_V(t)\rho(t)^\top\sigma_F(t,T)\,dt$. The proposition then follows from standard results in [26] and, more generally, [19]. ∎

Importantly, knowledge of $\phi(\omega,T',T)$ allows us to price put and call options on the futures price $F(T',T)$ through Fourier transform methods. The relevant results are well-known and are omitted in the interest of brevity; see e.g. [30] or [29] for details.

[5] We enable time-dependence to allow for seasonality effects (e.g. that the volatility-of-volatility is higher in winters than in summers, and so forth).

Remark 4 *Even though $\psi(t)$ in Proposition 2 is deterministic, the variables $y^*(t)$ are not, as their drift-terms depend on $V(t)$; see (19). As a result, the number of Markov state variables in the model covered by the proposition is $\frac{1}{2}N(N+3)+1$.*

We should note that similar results for the characteristic function $\phi(u,T',T)$ will also be possible in the case where we add jumps to the affine process (21). This extension – which may be quite relevant for actual commodities, see Figure 7 – is covered in [19] and is omitted here in the interest of brevity.

4 Jump Diffusions

4.1 Basic Idea and Notation

Prices of many commodities exhibit jumps and spikes in their empirical time-series, typically a consequence of the imbalances between supply and demand brought on by unforeseen events (hurricanes, drought, geopolitical events, and so forth). To accommodate this, our first approach is to introduce a vector-valued jump process, $J(t) = (J_1(t),\ldots,J_M(t))^\top$ where each component is a *marked point process* (MPP) with intensity process

$$\lambda_i(dx,t) = w_i(t)q_i(t,x)dx, \quad i = 1,\ldots,M. \tag{23}$$

Here $w_i(t)$ is a deterministic (Poisson) arrival rate and $q_i(t,x)$ is a time-dependent density. We interpret $\lambda_i(dx,t)$ as the probability of having a jump in J_i end up in the interval $[x,x+dx]$ at time t; conditional on a jump in J_i taking place at time t, the distribution of the jump size is $q_i(t,x)$. We assume that all J_i, $i = 1,\ldots,m$ are independent of each other. The form of equation (23) represents a specialization of the general class of MPPs, as discussed, e.g., in [24]. Note that (23) is generally *not* in the Levy class, unless we assume that both h_i and f_i are independent of time[6] t.

To incorporate J into a model for the futures curve, we write

$$dF(t,T)/F(t-,T) = \sigma(t,T)^\top dW(t) + \left(e^{m(t,T)^\top dJ(t)} - 1\right) + \mu(t,T)\,dt, \tag{24}$$

where $W(t)$ and $\sigma(t,T)$ are as in Section 2.1, and where μ and m are deterministic and of dimension 1 and M, respectively. For ease of matrix notation, going forward we shall, in fact, *assume that $N = M$*, i.e. that we have as many jumps as we have Brownian motions. This involves no loss of generality, as we are free to set elements of $\sigma(t,T)$ and $m(t,T)$ to 0 to effectively remove any unwanted Brownian motions and jumps[7].

(24) represents a process where futures prices follow a diffusion process that is occasionally interrupted with jumps of random magnitude. Specifically, if a jump takes place

[6] In many commodity markets, we expect that both the arrival rate and the severity of jumps may exhibit seasonality, i.e. are functions of calendar time.

[7] The use of a vector-valued jump process is primarily useful for applications where we model many commodities simultaneously, as some components of the jump vector can be considered idiosyncratic and some can be shared across commodities; the latter type can be used to create "jump correlation" (see [12]). In most single-commodity applications, one will often use only a small number of jump processes.

at time t in the ith jump process J_i, the futures value will jump to $F(t-,T)e^{m_i(t,T)Y_i}$, where Y_i is a random variable with density $q_i(t,\cdot)$. For (24) to represent a proper martingale, the deterministic drift $\mu(t,T)$ cannot be chosen freely but must be set to

$$\mu(t,T) = -\mathrm{E}\left(e^{m(t,T)^\top dJ(t)} - 1|\mathscr{F}_{t-}\right)/dt = -w(t)^\top \upsilon(t,T), \tag{25}$$

where $w(t) = (w_1(t),\ldots,w_N(t))^\top$ and the vector $\upsilon(t,T)$ has elements

$$\upsilon_i(t,T) = \int_{-\infty}^{\infty} e^{m_i(t,T)x} q_i(t,x)\,dx - 1, \quad i = 1,\ldots,N. \tag{26}$$

In many commodity markets, when a jump in the spot commodity price takes place at time t, the effect on futures prices $F(t,T)$ will often dissipate quite quickly in T; see Figure 3. In general, we would then expect the vector function $m(t,T)$ to decrease in $T-t$, at least in a seasonality-adjusted sense. We return to an actual parameterization example in Section 4.3.

4.2 Markov Representation

Mirroring the ideas that lead to Proposition 1, we now make the following assumption.

Assumption 2 *Let $\sigma(t,T)$ in (24) satisfy Assumption 1, such that $\sigma(t,T) = \alpha(t)\beta(T)$ with α and β as in Assumption 1. Also, introduce an $N\times N$ deterministic matrix $\gamma:\mathbb{R}\to\mathbb{R}^{N\times N}$ and assume that $m(t,T)$ in (24) separates as*

$$m(t,T) = \gamma(t)\beta(T).$$

With $X(t,T) = \ln F(t,T)$, under the terms of Assumption 2 it is easily seen that

$$X(t,T) = X(0,T) - \frac{1}{2}\beta(T)^\top y(t)\,\beta(T) + \beta(T)^\top x(t) + \int_0^t \mu(u,T)du$$

where

$$x(t) \triangleq \int_0^t \alpha(u)^\top dW(u) + \int_0^t \gamma(u)^\top dJ(u), \quad y(t) \triangleq \int_0^t \alpha(u)^\top \alpha(u)du. \tag{27}$$

The following result is thereby obvious.

Lemma 6 *Let Assumption 2 hold, and assume further that $\alpha(t) = f(t,x(t),y(t))$ for a deterministic function $f:\mathbb{R}\times\mathbb{R}^N\times\mathbb{R}^{N\times N}\to\mathbb{R}^{N\times N}$. Together, $x(t)$ and $y(t)$ in (27) then form a Markov jump-SDE system, with $x(0) = y(0) = 0$,*

$$\begin{aligned} dx(t) &= f(t,x(t),y(t))^\top\, dW(t) + \gamma(t)^\top dJ(t),\\ dy(t) &= f(t,x(t),y(t))^\top\, f(t,x(t),y(t))\, dt. \end{aligned}$$

$\mu(u,T)$ is given by (25) and the futures curve can be reconstituted as

$$F(t,T) = F(0,T)\exp\left(-\frac{1}{2}\beta(T)^\top y(t)\beta(T) + \beta(T)^\top x(t) + \int_0^t \mu(u,T)\,du\right). \tag{28}$$

As in Section 2.2, we can elect to transform the state variables in Lemma 6 to a mean-reverting form. For instance, if we set $x^*(t) = -\frac{1}{2}B(t)y(t)\beta(t) + B(t)x(t)$ and $y^*(t) = B(t)y(t)B(t)$, with $B(t) = \text{diag}(\beta(t))$, Lemma 1 remains true after adjusting the process for $dx^*(t)$ to include an additional term $B(t)\gamma(t)^\top dJ(t)$. The obvious details are omitted in the interest of brevity, as is the obvious extension to stochastic volatility.

In the reconstitution formula (28), the integral of $\mu(t,T)$ must generally be computed by numerical integration for all relevant values of t and T. For special jump distributions (most notably exponential jumps), closed-form solutions exist if the jump intensity λ is constant. As discussed in Footnote 6, however, in applications we may wish to use time-dependent λ to model seasonality effects.

4.3 Example

At this point, a small example is in order to clarify matters. Specifically, let us consider adding a single jump process to a stationary Gaussian one-factor model with $\sigma(t,T) = \sigma_0 \exp(-\kappa_0(T-t))$. We wish for the effects of jump to revert to 0 at a speed of κ_J. In Lemma 6, we can set

$$\alpha(t) = \begin{pmatrix} \sigma_0 e^{\kappa_0 t} & 0 \\ 0 & 0 \end{pmatrix}; \quad \gamma(t) = \begin{pmatrix} 0 & 0 \\ 0 & e^{\kappa_J t} \end{pmatrix}; \quad \beta(T) = \begin{pmatrix} e^{-\kappa_0 T} \\ e^{-\kappa_J T} \end{pmatrix}.$$

According to Lemma 6, the state variables for the futures curve are (omitting the deterministic y's)

$$dx_1(t) = \sigma_0 e^{\kappa_0 t}\, dW(t), \quad dx_2(t) = e^{\kappa_J t}\, dJ(t),$$

where both W and J are scalar. Alternatively, if we set $z_1(t) = x_1(t)e^{-\kappa_0 t}$ and $z_2(t) = x_2(t)e^{-\kappa_J t}$, we get the state variable processes

$$dz_1(t) = -\kappa_0 z_1(t) + \sigma_0\, dW(t), \quad dz_2(t) = -\kappa_J z_2(t)dt + dJ(t), \quad z_1(0) = z_2(0) = 0.$$

We are free to exogenously specify the jump arrival rate and jump size density for J.

As the 1st row of γ and 2nd row in α are all zeros, the two state variables z_1 and z_2 above are pure diffusion and pure jump, respectively. In other settings, however, state variables that simultaneously involve both diffusions and jumps may emerge.

For large values of κ_J, the paths of the state variable z_2 will exhibit spike-like behavior, as each jump in J will be dampened away rapidly. [25] use this feature to apply the sample model above to electricity markets, where price spikes are commonplace. See also [13].

4.4 Option Pricing

From the discussion in Section 3.2 we recall that European option pricing can be done analytically, provided that we can establish the characteristic function $\phi(\omega, T', T)$ for $X(T',T) = \ln F(T',T)$, $T' \le T$. Under Assumption 2,

$$dX(t,T) = dX_c(t,T) + \mu(t,T)dt + dX_J(t,T),$$

where we have split $dX(t,T)$ into continuous and pure-jump components:

$$\begin{aligned} dX_J(t,T) &= \beta(T)^\top \gamma(t)^\top dJ(t), \\ dX_c(t,T) &= -\frac{1}{2}\beta(T)^\top \alpha(t)^\top \alpha(t)\beta(T)dt + \beta(T)^\top \alpha(t)^\top dW(t). \end{aligned}$$

Under the terms of Lemma 6, X_J and X_c can be co-dependent as $\alpha(t)$ may depend on the path of $J(t)$. In practice, however, it will often be the case that we can structure the model such that X_J and X_c are independent – see the model in Section 4.3 for an example. Provided that X_J and X_c are independent, we have

$$\phi(\omega,T',T) \triangleq \mathrm{E}\left(e^{i\omega X(T',T)}\right) = \mathrm{E}\left(e^{i\omega X_c(T',T)}\right)\cdot \mathrm{E}\left(e^{i\omega X_J(T',T)}\right)\cdot e^{i\omega\int_0^{T'}\mu(u,T)du}.$$

If independence holds and the characteristic function for $X_c(T,T)$ is known – which is the case for Heston-type stochastic volatility models, say – then $\phi(\omega,T',T)$ can evidently be completed provided that we can determine the characteristic function of X_J. For this, we can use the following result.

Lemma 7 *Let the characteristic function for the magnitude of jump J_i be*

$$\varphi_j(t,\omega) = \int_{-\infty}^{\infty} e^{i\omega x} q_j(t,x)\,dx, \quad j = 1,\ldots,N.$$

For any N-dimensional vector v, let $\varphi(t,v) \triangleq (\varphi_1(t,v_1),\ldots,\varphi_N(t,v_N))^\top$. Then

$$\phi_J(\omega;T',T) \triangleq \mathrm{E}\left(e^{i\omega X_J(T',T)}\right) = \exp\left(\int_0^{T'} w(t)^\top\left(\varphi\left(t,\omega\gamma(t)\beta(T)\right)-\mathbf{1}\right)dt\right).$$

Proof: Set $Z(t,T) = \exp\left(i\omega X_J(t,T)\right) = \exp\left(\int_0^t i\omega\beta(T)^\top\gamma(u)^\top dJ(u)\right)$ and note that

$$dZ(t,T) = Z(t-,T)\left(\exp\left(i\omega\beta(T)^\top\gamma(t)^\top dJ(t)\right)-1\right).$$

Taking expectations, we get

$$\begin{aligned} d\mathrm{E}\left(Z(t,T)\right) &= \mathrm{E}\left(Z(t-,T)\right)\mathrm{E}\left(\exp\left(i\omega\beta(T)^\top\gamma(t)^\top dJ(t)\right)-1\right) \\ &= \mathrm{E}\left(Z(t-,T)\right)w(t)^\top\left(\varphi\left(t,\omega\gamma(t)\beta(T)\right)-\mathbf{1}\right)dt, \end{aligned}$$

and the lemma follows by integration, as $\phi_J(\omega;T',T) = \mathrm{E}(Z(T',T))$. ∎

Remark 5 *In the notation of Lemma 7, in (26) we have $\upsilon_j(t,T) = \varphi_j(t,-i\cdot m_j(t,T))-1$. Also*

$$\mu(t,T) = -w(t)^\top\left(\varphi\left(t,-i\cdot\gamma(t)\beta(T)\right)-\mathbf{1}\right).$$

5 Regime Switching

As discussed in Section 4.3, jump-diffusions with (very) strong mean-reversion can be used to model spike behavior in futures price time-series. A more natural alternative (or supplement) is to introduce Markov chains that can directly govern transitions between "excited" and "normal" states of the futures curve. As we shall see, such *regime switch models* sometimes have significant analytical tractability, allowing one to avoid tedious[8] Fourier space work when pricing options.

5.1 Continuous Time Markov Chain Basics

Let $c(t)$ be a continuous-time Markov chain with m states and generator matrix $Q(t) = \{Q_{ij}(t)\}$. As is common, we take the state space for $c(t)$ to be the basis vectors $\{e_1, e_2, \ldots, e_m\}$ of $\mathbb{R}^m$. That is, we associate state j of the chain with $c(t) = e_j$, where the jth element of the vector e_j is one and all other elements are zero. We recall that when the chain is in state i at time t, the probability of a transition to state j is

$$\Pr\left(c(t+dt) = e_j | c(t) = e_i\right) = Q_{ij}(t)dt, \quad i \neq j. \tag{29}$$

The diagonal elements of $Q(t)$ are given by

$$Q_{ii}(t) = -\sum_{i \neq j} Q_{ij}(t), \tag{30}$$

and, by the Chapman-Kolmogorov forward equation, we have

$$\mathrm{E}\left(c(T)|c(t)\right) = \mathscr{E}_Q(t,T)^\top c(t), \tag{31}$$

where $\mathscr{E}_Q(t,T) \in \mathbb{R}^{m\times m}$ is the *time-ordered exponential* of $\{Q(s)|t \leq s \leq T\}$, satisfying

$$\frac{\partial}{\partial s}\mathscr{E}_Q(t,s) = \mathscr{E}_Q(t,s)Q(s), \quad \mathscr{E}_Q(t,t) = I,$$

with I being the identity matrix. $\mathscr{E}_Q(t,T)$ can be interpreted as a *product-integral* (see [17]) and can be computed numerically by an ODE solver. In the special case of time-independent Q, it can be written as a matrix exponential, $\mathscr{E}_Q(t,T) = \exp\left(Q(T-t)\right)$.

5.2 Jump Process Construction

We can use the Markov chain defined above to produce a rather general jump process $J(t)$, by letting $J(t)$ jump by a random magnitude $H_{ij}(t)$ whenever $c(t)$ transitions from e_i to e_j. We collect the $H_{ij}(t)$ in a random matrix $H(t)$, with elements having known densities

[8] In the expression for ϕ_J in Lemma 7 notice that integrals cannot be reused across option maturities due to the presence of $\beta(T)$ in the integrand. An exception occurs when $\beta(T)$ is constant, including the case where jumps do not mean-revert (a case that is of no interest for markets with spiky price behavior).

$d_{ij}(t,z) = \Pr\left(H_{ij}(t) \in [z, z+dz]\right)/dz$. The elements of H are assumed independent and are further associated with moment generating functions (mgf's)

$$h_{ij}(t,\theta) \triangleq \mathrm{E}\left(e^{\theta H_{ij}(t)}\right) = \int_{-\infty}^{\infty} e^{\theta z} d_{ij}(t,z)\, dz, \quad \theta \in \mathbb{C}. \tag{32}$$

Using the time-ordered matrix exponential introduced in (31), the mgf of J can then be determined as follows.

Proposition 3 *Set*

$$\chi_J(t,T,\theta) \triangleq \mathrm{E}\left(e^{\theta J(T)} \middle| c(t), J(t)\right), \quad \theta \in \mathbb{C}.$$

Define the $m \times m$ matrix $D(t;\theta)$, with diagonal elements $D_{ii}(t;\theta) = Q_{ii}(t)$, $i = 1,\ldots,m$, and $D_{ij}(t;\theta) = Q_{ij}(t)h_{ij}(t,\theta)$, $i \neq j$. Letting $\mathbf{1}$ be a m-dimensional vector of 1's, then

$$\chi_J(t,T,\theta) = e^{\theta J(t)} c(t)^\top \mathscr{E}_D(t,T;\theta)\mathbf{1},$$

where $\mathscr{E}_D(t,T;\theta)$ is the time-ordered exponential of $D(\cdot;\theta)$ on $t,t]$.

Proof: Set $Z(t) = \exp(\theta J(t))c(t)$, and consider the change in $Z(t)$ on the interval $[t, t+dt]$. Supposing that, say, $c(t) = e_i$, $e^{\theta J(t)}$ will increase to $e^{\theta J(t)+\theta H_{ij}(t)}$ with probability $Q_{ij}(t)\,dt$, $j = 1,\ldots,m$. That is, using (30) and (32),

$$\begin{aligned}\mathrm{E}\left(dZ(t)|c(t) = e_i, J(t)\right) &= e^{\theta J(t)}\left(\sum_{j\neq i} Q_{ij}(t)\left(\mathrm{E}\left(e^{\theta H_{ij}(t)}\right)e_j - e_i\right)\right)dt \\ &= e^{\theta J(t)}\left(\sum_{j\neq i} Q_{ij}(t)h_{ij}(t,\theta)e_j + Q_{ii}e_i\right)dt.\end{aligned}$$

The structure of these equations show that, in general,

$$\mathrm{E}\left(dZ(t)|c(t),J(t)\right) = e^{\theta J(t)} D(t;\theta)^\top c(t)\,dt = D(t;\theta)^\top Z(t)\,dt,$$

where the elements of $D(t;\theta)$ are as given above. Integration of this equation yields the result of the proposition, after using $e^{\theta J(T)} = Z(T)^\top \mathbf{1}$ (by definition). ∎

5.3 Futures Curve Model

We proceed to use the process $J(t)$ in an actual model for the evolution of futures prices. As a starting point, let us assume that we have already constructed a Markovian model for the futures curve along the lines in Sections[9] 2 or 3. Let the futures curve generated by this model be $F_c(t,T)$, and consider now the problem of adding regime-switching behavior to

[9] In fact we can also allow for jump-models of the type in Section 4. However, it is unlikely that we in practice would ever use both regime-switching and ordinary jump-diffusions in the same model.

the dynamics of the model. For this, we introduce a regime-switch processes $J(t)$ of the type in Section 5.2 and write

$$F(t,T) = F_c(t,T)F_J(t,T), \tag{33}$$

where $J(t)$ is assumed independent of $F_c(t,T)$, and

$$F_J(t,T) \triangleq \mathrm{E}\left(e^{s(T)J(T)}\middle| J(t),c(t)\right)e^{-G(T)}, \quad G(T) \triangleq \ln \mathrm{E}\left(e^{s(T)\cdot J(T)}\middle| c(0),J(0)\right).$$

We have introduced a deterministic function $s(T)$, the role of which is to incorporate jump-size seasonality (see Section 8). Notice that the spot process $S(t) = F(t,t)$ here becomes

$$S(t) = S_c(t)e^{s(t)\cdot J(t)-G(t)}, \quad S_c(t) = F_c(t,t).$$

In the definition (33), it is clear that F_J is a martingale, such that F will be so as well (F_c is a martingale by assumption). Also note that if $F_c(0,T)$ fits the initial time 0 futures curve, so will $F(0,T)$, as $F_J(0,T) = 1$ by construction. The resulting model for the futures curve will be Markovian in a finite number of state variables, as F_c is so by assumption, and F_J is Markov in two variables $J(t)$ and $c(t)$. The total number of state variables arising solely from the regime-switch mechanism is thereby[10] 2. As for the actual reconstitution of the futures curve from $J(t)$ and $c(t)$, we can use the following result, the proof of which is obvious.

Lemma 8 *In the model (33) we have* $G(T) = \ln \chi_J(0,T,s(T))$, *and*

$$F_J(t,T) = \frac{\chi_J(t,T,s(T))}{\chi_J(0,T,s(T))}, \quad j = 1,\ldots M, \tag{34}$$

where $\chi_J(t,T,s(T))$ *can be computed from* $c(t)$ *and* $J(t)$ *by Proposition 3.*

For the purpose of option pricing, the following result is useful.

Lemma 9 *For* $T' \leq T$, *define* $\phi_J(\omega;T',T) \triangleq \mathrm{E}(e^{i\omega \ln F_J(T',T)})$. *Then, with the matrices D and* $\mathscr{E}_D$ *defined as in Proposition 3,*

$$\phi_J(\omega;T',T) = e^{i\omega(s(T)J(0)-G(T))}c(0)^\top \mathscr{E}_D\left(0,T';i\omega s(T)\right)v(T',T),$$

where $G(T)$ *is given in Lemma 8 and the vector v has components*

$$v_j(T',T) = \left(e_j^\top \mathscr{E}_D\left(T',T;s(T)\right)\mathbf{1}\right)^{i\omega}, \quad j = 1,\ldots,m.$$

[10] We list extensions to multiple jump processes shortly.

Proof: From (34) we have

$$e^{i\omega \ln F_J(T',T)} = \left(\chi_J(T',T,s(T))e^{-G(T)}\right)^{i\omega}$$
$$= e^{i\omega\cdot s(T)J(T')}c(T')^{\top}\left(\mathscr{E}_D\left(T',T;s(T)\right)\mathbf{1}\right)^{i\omega}e^{-i\omega G(T)},$$

where the power function is to be applied component-wise on vectors. From the proof of Proposition 3, we also have

$$\mathrm{E}\left(e^{i\omega\cdot s(T)J(T')}c(T')^{\top}\right) = e^{i\omega\cdot s(T)J(0)}c(0)^{\top}\mathscr{E}_D\left(0,T';i\omega s(T)\right)$$

and the lemma follows. ∎

We can use the result of Lemma to compute European option prices by Fourier techniques (see Section 3.2) provided that the characteristic function of $\ln F_c(T',T)$ is known. If the options are written on the spot price, matters simplify somewhat:

Corollary 1 *In Lemma 9, if $T' = T$, then $\phi_J(\omega;T,T) = \chi_J(0,T,i\omega)e^{-i\omega G(T)}$.*

For the sake of simplicity, we have so far considered the case where a single type of regime-switch jump is added to the continuous futures curve dynamics. It is, however, straightforward to extend the setup to M independent jumps $J_1(t),\ldots,J_M(t)$ associated with M independent Markov chains $c_1(t),\ldots,c_M(t)$. Instead of (33), we would write

$$F(t,T) = F_c(t,T)\prod_{j=1}^{M}F_j(t,T), \quad F_j(t,T) \triangleq \mathrm{E}\left(e^{s_j(T)J_j(T)}\middle| J_j(t),c_j(t)\right)e^{-G_j(T)},$$

with $G_j(T) = \ln \mathrm{E}\left(\exp\left(s_j(T)\cdot J_j(T)\right)\right)$, $j = 1,\ldots,M$. Each of the $F_j(t,T)$ terms can then be reconstituted from $c_j(t)$ and $J_j(t)$ using Lemma 8. The characteristic function for $\ln F(t,T)$ would, due to independence, contain a product of the characteristic functions for the $\ln F_j(t,T)$'s, each to be computed by Lemma 9. We trust that the reader can imagine how this would work, and simply note that the resulting model would involve $2M$ Markov state-variables to be added to those needed to reconstitute $F_c(t,T)$.

5.4 Example: Spike Modeling and Jump Process Simplifications

With Corollary 8 and Lemma 9 we have completed the necessary theoretical characterization of the model outlined so far, the spirit of which is similar to that of the affine spot models in [15]. Application of the given results in practice are generally not trivial due to the need to work with matrix exponentials in the futures reconstitution formula, and the need to operate in Fourier space when pricing options. If, as is likely. the regime switch model is intended to model price spikes a small change in our setup can, however, often reduce the computational burden significantly.

Before we describe what we have in mind, let us first consider a parameterization of the generator matrix Q suitable for spike modeling. Specifically, we consider a chain $c(t)$

with state e_1 being a "base level" from which we can jump into $m-1$ "excited states" $e_2,\ldots,e_m$, at time-dependent arrival rates of $n_j(t)$, $j=2,\ldots,m$. Once in an excited state, the chain must revert back to the base state e_1, at a state-independent arrival rate of $n_1(t)$, before further jumps can take place. In other words,

$$Q_{1j}(t)=n_j(t),\quad Q_{j1}(t)=n_1(t),\quad j=2,\ldots,m, \tag{35}$$

with all other non-diagonal elements of Q set to zero.

The matrix Q defined above defines a setup with $m-1$ different type of spikes, with $n_1(t)$ being a common measure for the average spike duration[11]. In a sense, then, $n_1(t)$ serves the same role as mean reversion does for the models in Section 4. Let us define $p_i^j(t,T)\triangleq \Pr(c(T)=e_j|c(t)=e_i)$ and $p^i(t,T)=(p_i^1(t,T),\ldots,p_i^m(t,T))^\top$; according to (31) then

$$p^i(t,T)=\mathscr{E}_Q(t,T)^\top \mathbf{1}_i, \tag{36}$$

where $\mathbf{1}_i$ is a vector with 1 in the ith position, and zeros everywhere else. While the $p^i(t,T)$ vector can be computed mechanically from (36), a few comments may aid the intuition. First, a symmetry argument shows that, for any j we have

$$p_i^j(t,T)=p_k^j(t,T)\triangleq \eta^j(t,T),\quad i>1,k>1,i\neq j,k\neq j. \tag{37}$$

This follows from the fact that any transition into state e_j from a non-e_j state other than e_1 must pass through e_1 first (and the transition rate into e_1 is independent of the state). Also, for any $j>1$ we have

$$p_j^j(t,T)-\exp\left(-\int_t^T n_1(u)\,du\right)=\eta^j(t,T),\quad j>1, \tag{38}$$

as both sides of this equation equal the likelihood of transitioning from a non-e_1 state to e_1, eventually ending up in state e_j at time T.

Remark 6 *If the n_i, $i=1,\ldots,m$ are all independent of time, $\mathscr{E}_Q(t,T)$ is a regular matrix exponential which can here be computed in closed form by eigenvalue decomposition, since the eigenvalues of Q are 0, $-\sum_{j=1}^m n_j$, and $-n_1$, with the eigenvalue $-n_1$ having multiplicity $m-2$. If the n_i are piecewise constant through time (which is often the case in applications), $\mathscr{E}_Q(t,T)$ can then conveniently be computed by a finite number of multiplications of such exponential matrices.*

Having characterized our "spike" transition matrix, we proceed to introduce a jump process $J(t)$ linked to the Markov chain $c(t)$.

Assumption 3 *Whenever $c(t)$ is in state e_1, $J(t)=0$. When $c(t)$ jumps from state e_1 to state j, $j>1$, $J(t)$ jumps by an amount H_{1j}, the density of which is $d_j(z)$ and independent of time. Also,* $\Pr(J(t)=0|c(t)\neq e_1)=0$.

[11] If we are interested in having multiple average spike durations in the same model, we simply add more Markov chains, as described in Section 5.3 above.

The assumption of $J(t)$ returning to 0 whenever $c(t)$ returns to e_1 is a spike-specific construction inspired by a similar idea in [3], and departs slightly from our previous jump-construction setup[12] in Section 5.2. While we could have retained the structure of that setup (e.g. by having the jump from state e_j back to state e_1 have a "reverse" density $d_j(-z)$, say), Assumption 3 is quite natural for the generation of spikes and ultimately leads to additional tractability. For instance, we have the following simple result.

Proposition 4 *Assume that $J(t)$ is as in Assumption 3, and that Q is given by (35). Let the mgf associated with the jump density $d_j(z)$ be $h_j(\theta)$, $\theta \in \mathbb{C}$, with $h_1(\theta) = 1$. As before, set $\chi_J(t,T,\theta) = \mathrm{E}(e^{\theta J(T)}|c(t),J(t))$ and let $\eta^j(t,T)$ be as defined in (37). Then, with probability one,*

$$\chi_J(t,T,\theta) = 1_{J(t)\neq 0}\left\{e^{\theta J(t)}e^{-\tilde{n}_1(t,T)} + \sum_{j=1}^{m}\eta^j(t,T)h_j(\theta)\right\} + 1_{J(t)=0}\left\{\sum_{j=1}^{m}p_1^j(t,T)h_j(\theta)\right\}, \quad (39)$$

where $\tilde{n}_1(t,T) = \int_t^T n_1(u)\,du$. Equation (39) does not *depend on $c(t)$.*

Proof: If $J(t) \neq 0$ then, with probability one, $c(t) \neq e_1$. Suppose that $c(t) = e_k$, $k > 1$, such that

$$\mathrm{E}\left(e^{\theta J(T)}\middle| c(t) = e_k, J(t) = z\right) = p_k^1(t,T) + e^{\theta z}e^{-\tilde{n}_1(t,T)} + \sum_{j\neq k, j>1} h_j(\theta)p_k^j(t,T) + h_k(\theta)\left(p_k^k(t,T) - e^{-\tilde{n}_1(t,T)}\right).$$

The first term corresponds to the case where $c(T) = e_1$ and $J(T) = 0$, the probability of which is $p_k^1(t,T)$. The second term corresponds to the case where c remains at state e_k with no further excursions (i.e. $J(T) = z$), the probability of which is $e^{-\tilde{n}_1(t,T)}$. The third term corresponds to the case where c ends up in a state other than e_1 or e_k at time T; the expected value of $e^{\theta J(T)}$ is $h_j(\theta)$ if $c(T) = e_j$, the probability of which is $p_k^j(t,T)$. The fourth and final term corresponds to the case where $c(t)$ moves away from e_k at some time during $[t,T]$, yet ends up in state e_k again at time T; the probability of this is $p_k^k(t,T) - e^{-\tilde{n}_1(t,T)}$. Applying (37) and (38) leads to the term on $1_{J(t)\neq 0}$ in (39). The term on $1_{J(t)=0}$ (which applies when $J(t) = 0$ and $c(t) = 0$) is proven the same way. ∎

Retaining the futures curve model of Section 5.3, we note that $F_J(t,T)$ can be reconstituted from Lemma 8 given knowledge *only* of $J(t)$ (and not of $c(t)$). Besides simplifying

[12] In particular, we break the independence of the elements in the matrix H by effectively forcing the jump in J associated with a move from state e_j to state e_1 to be precisely opposite of most recent jump in J (that took place when c transitioned into state e_j).

the result in Proposition 3, evidently Assumption 3 also allows us to cut the number of Markov variables associated with the spike mechanism[13] from 2 to 1.

5.5 Option Pricing in Spike Model

The model outlined in Section 5.4 has several attractive features when it comes to the pricing of options, particularly for the important special case of options on the spot price $S(T) = F(T,T)$. First, it is clear from Corollary 1 that we can compute the characteristic function of $\ln F_J(T,T)$ easily from Proposition 4; notice that the resulting characteristic function $\phi_J(\omega;T,T)$ will *not* involve any ω-dependent time-integrals, which represents a clear computational advantage over the results in Lemma 7. As long as the characteristic function for $\ln F_c(T,T)$ is known, we can price put and call options by Fourier methods.

Second, assuming that i) $F_c(t,T)$ has deterministic volatility, and ii) the jump densities are Gaussian, then put and call options on the spot price in fact requires no Fourier transform work and can be computed in closed form with formulas of the Black-Scholes type. This result is attractive for applied work and is derived below.

Assumption 4 *Let $dF_c(t,T)/F_c(t,T) = \sigma_c(t,T)^\top dW(t)$ for a deterministic N-dimensional vector $\sigma_c(t,T)$, and assume that $J(t)$ is as in Assumption 3 with all jump densities being* Gaussian, *with means μ_j and standard deviations γ_j:*

$$d_j(z) = \frac{1}{\gamma_j\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{z-\mu_j}{\gamma_j}\right)^2\right), \quad j = 2,\ldots,m.$$

For any j we assume that either $\mu_j \neq 0$ or $\gamma_j > 0$.

Proposition 5 *Define a call option $C(0) = P(0,T)\mathrm{E}\left((S(T)-K)^+\right)$, and assume that the model (33) holds with independent components F_c and F_J given as in Assumption 4. Then, if $J(0) \neq 0$,*

$$\begin{aligned}\frac{C(0)}{P(0,T)} = e^{-G(T)}\sum_{j=1}^{m}\mathrm{E}\left(\left(F_c(T,T)e^{s(T)Z_j} - Ke^{G(T)}\right)^+\right)\eta^j(0,T)\\ + e^{-G(T)}\mathrm{E}\left(\left(F_c(T,T)e^{s(T)J(0)} - Ke^{G(T)}\right)^+\right)e^{-\int_0^T n_1(u)\,du},\end{aligned} \tag{40}$$

where $\eta^j(0,T)$ is given in (37) and Z_j is a Gaussian random variable, $Z_j \sim \mathcal{N}(\mu_j,\gamma_j)$, with the convention that $\mu_1 = \gamma_1 = 0$. If $J(0) = 0$, then

$$\frac{C(0)}{P(0,T)} = e^{-G(T)}\sum_{j=1}^{m}\mathrm{E}\left(\left(F_c(T,T)e^{s(T)Z_j} - Ke^{G(T)}\right)^+\right)p_1^j(0,T). \tag{41}$$

[13] If we elect to use M independent jump processes in our model, the number of Markov state variables associated with spike generation will therefore be M.

In both expressions, for $j = 1, \ldots, m$,

$$\mathrm{E}\left(\left(F_c(T,T)e^{s(T)Z_j} - Ke^{G(T)}\right)^+\right) = m_j \Phi\left(\varsigma_j^{(+)}\right) - Ke^{G(T)}\Phi\left(\varsigma_j^{(-)}\right), \tag{42}$$

$$\varsigma_j^{(\pm)} = \frac{\ln\left(\frac{m_j}{Ke^{G(T)}}\right) \pm \frac{1}{2}\ln(u_j m_j^{-2})}{\sqrt{\ln(u_j m_j^{-2})}},$$

with $m_j \triangleq F(0,T)e^{s(T)\mu_j + s(T)^2\gamma_j^2/2}$ *and* $u_j \triangleq F(0,T)^2 e^{\int_0^T |\sigma_c(u,T)|^2\,du} e^{2s(T)\mu_j + 2s(T)^2\gamma_j^2}$.

Proof: Assume first that $J(0) \neq 0$, such that, with probability one, $c(0) > 0$. By a symmetry argument we have, for an arbitrary $i > 1$,

$$\begin{aligned}&\mathrm{E}\left((S(T)-K)^+ | J(0) \neq 0\right)\\ &\quad = e^{-G(T)} \sum_{j=1}^m \mathrm{E}\left(\left(F_c(T,T)e^{s(T)J(T)} - Ke^{G(T)}\right)^+ | c(0) = e_i, c(T) = e_j\right) p_i^j(t,T)\end{aligned}$$

and (40) follows by the same arguments as in the proof of Proposition 4. The result (41) is shown the same way. By standard results, $F_c(T,T)e^{s(T)Z_j}$ is a log-normally distributed variable with non-central first and second moment m_j and u_j respectively. The final result (42) follows from the usual Black-Scholes analysis. ∎

We note that [3] proves a closely related result for a one-factor log-normal spot commodity model combined with a regime-switch jump model with two types of jumps, "up" and "down" (i.e. $N = 1$ and $m = 2$ in our more general setting). The model is shown to yield good empirical fit to Nordpool electricity data.

The discussion above only covers options on spot commodity prices. When pricing options on futures prices, one possibility is to rely on a conditioning argument, where we write (with $T' \leq T$)

$$\begin{aligned}\mathrm{E}\left(\left(F(T',T)-K\right)^+\right) &= \int_{\mathbb{R}} \mathrm{E}\left(\left(F(T',T)-K\right)^+ | J(T') = z\right) d_J(T',z)\,dz\\ &= \int_{\mathbb{R}} e^{-G(T)} \chi_J(T',T,s(T),z)\mathrm{E}\left(\left(F_c(T',T) - K^*(T',z)\right)^+\right) d_J(T',z)\,dz,\end{aligned} \tag{43}$$

where $\chi_J(T',T,s(T),J(T')) = \chi_J(T',T,s(T))$ is given in Proposition 4, $d_J(T',z)$ is the density of $J(T')$, and

$$K^*(z) = \chi_J(T',T,s(T),z)^{-1} e^{G(T)}.$$

Assuming that call option prices on $F_c(T',T)$ are available, (43) can be computed numerically once we establish the density $d_J(T',z)$. But Proposition 4 shows that

$$\begin{aligned}d_J(T',z) = 1_{J(0)\neq 0}&\left\{\delta(z)\eta^1(0,T') + \delta(z-J(0))e^{-\int_0^{T'} n_1(u)\,du} + \sum_{j=2}^m \eta^j(0,T')d_j(z)\right\}\\ &+ 1_{J(0)=0}\left\{\delta(z)p_1^1(0,T') + \sum_{j=2}^m p_1^j(0,T')d_j(z)\right\},\end{aligned} \tag{44}$$

where $\delta(\cdot)$ is Dirac's delta-function. We observe from (44) that $d_J(T',z)$ (not surprisingly) essentially is a mixture density. As an aside, we notice that this observation can sometimes be used to reduce the dimension of the Markov chain; the relevant result is shown in Appendix B for completeness.

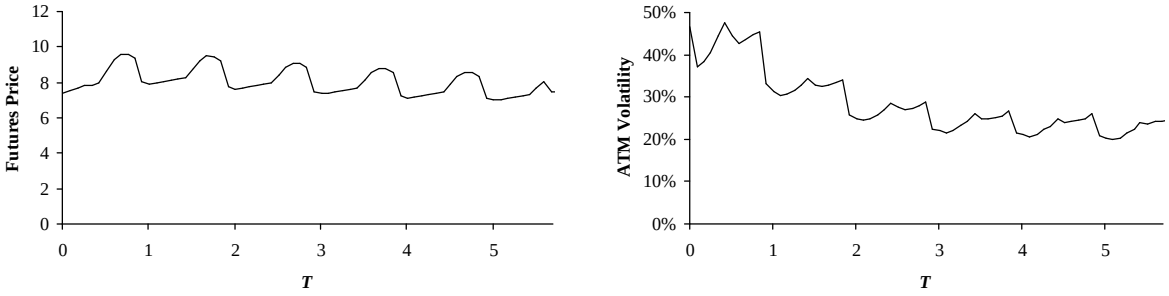
6 Some Empirical Results for Gas Futures

Having now completed the theoretical section of the paper, the next three sections deal with the application of (portions of) the developed model framework to a specific commodity market, USD natural gas. The main intention is here to illustrate generic calibration ideas for seasonal commodities, so we keep the model quite simple in order not to drown out the principles at play with model details. Indeed, it is our intention to limit the model to three state variables or less, which has the added benefit of easing the numerical implementation of the model. As background for later model development, we start out in this section with a brief tour through key empirical results for gas derivatives. We highlight the pervasiveness of seasonality effects.

6.1 Futures and Option Market Snapshot

First, to provide a general feel for the gas derivative markets, in Figure 1 we show the term structure of the USD natural gas futures prices and implied at-the-money (ATM) volatilities for call options on spot, as observed in the market[14] in April 2007.

Figure 1: Futures Prices and ATM Volatilities, USD Gas Market



Notes: The left panel shows the futures curve $F(0,T)$; the right shows implied ATM volatilities for European T-maturity call options on $F(T,T)$ (spot).

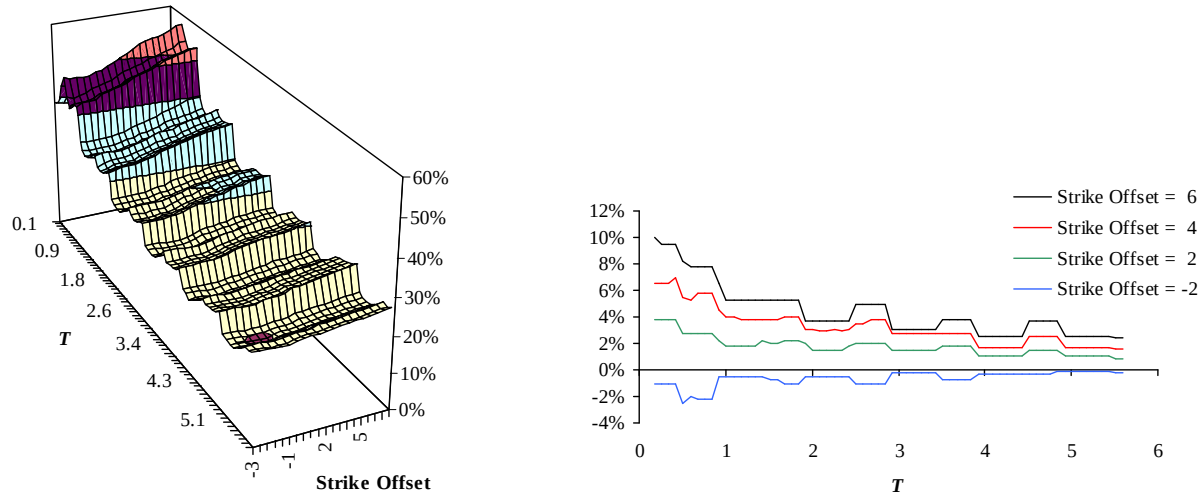
High storage costs combined with increased demand for gas during cold months on average drive gas prices up during the winter, in turn increasing the expected cost of delivery on futures contracts maturing in the winter. This seasonal effect is reflected in the strongly undulating futures term structure in Figure 1, with prices peaking and falling as seasons progress from cold to warm. Increased winter demand also makes prices more

[14] Henry Hub natural gas futures, as marked by the trading desk at Banc of America Securities.

susceptible to deviations and shocks in gas consumption (as triggered by colder-than-expected weather, say), leading to similar seasonal effects for implied options volatilities: high values for winter settlement and low values for summer settlement, *ceteris paribus.* As evidenced in Figure 1, seasonal effects in implied option volatilities move around a "spine" of declining option volatility as a function of maturity. For sufficiently long maturities, this spine appears to approach a constant plateau, roughly 20% in the figure.

Gas options struck away from at-the-money (i.e. the strike differs from the current futures price) will generally trade at implied volatility levels that differ from that of ATM options. Lack of liquidity for non-ATM option makes it difficult to pin down the shape of the volatility smile with great precision, but some data is available. Figure 2 below shows the gas volatility skew surface estimated by BofA's trading desk in April 2007. While the data is somewhat rough, the overall trend is the existence of a "reverse" volatility skew, where volatilities increase in strike. Figure 2 also shows that the volatility skew dies out as the option maturity is increased; this overall skew decay is overlaid by a seasonal component where the skew increases in the winter and decreases in the summer.

Figure 2: Volatility Smile Surface



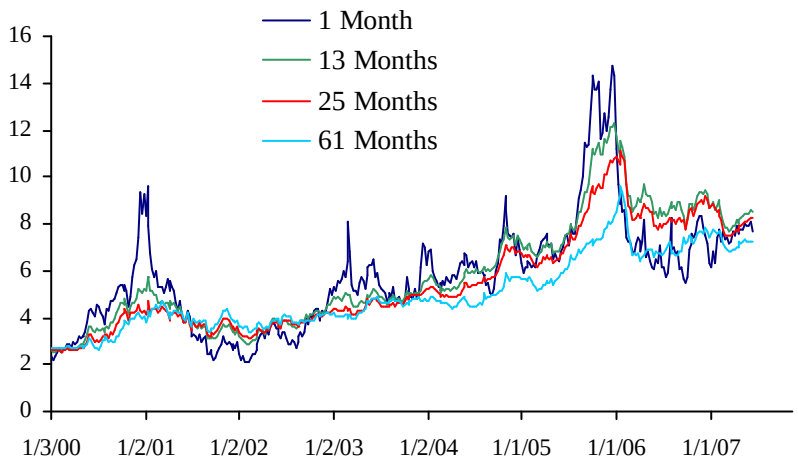
Notes: The left panel shows the volatility smile, as a function of option maturity (T) and strike. The right panel shows the decay with maturity of the "skew" (the implied volatility at a given strike minus the ATM implied volatility). In both panels, the strike is represented as an offset to the ATM strike: strike$= F(0,T)+$offset.

6.2 Time Series Analysis

Having examined a market snapshot, let us move to an informal time-series analysis. In Figure 3, we show "rolling" futures prices $F(t,t+\Delta)$ for various values of time to maturity

Δ, with t moving from January 3, 2000, to June 13, 2007, in weekly steps[15]. For clarity, the Δ's are spaced apart by multiples of a whole year to ensure consistency in the season of delivery across the various futures.

Figure 3: Selected USD Gas Futures



Notes: The figure shows the empirical time series data for $F(t,t+\Delta)$, for Δ = {1 month, 13 months, 25 months, 61 months}.

Casual observation of Figure 3 leads one to conclude that the short end of the futures curve is significantly more volatile than the long end, in qualitative agreement with the maturity decay we observed for implied option volatilities in Figure 1. The coupling between short- and long-dated futures price movements is not particularly strong, and on occasion the short-term futures prices will exhibit rapid upward moves with no corresponding increase in long-term futures. These spikes typically take place in winter months – often as a result of hurricane activity or unusually cold winters – and are followed by a relatively rapid return to a more normal state in the spring. Whenever a spike occurs, the futures curve becomes strongly inverted (i.e. it moves into *backwardation*), with the short-term futures prices substantially exceeding long-term futures prices.

6.2.1 Correlation Structure

To investigate more closely the dependence structure underlying moves in the futures price term structure, set $X(t,T) = \ln F(t,T)$ and define a correlation function

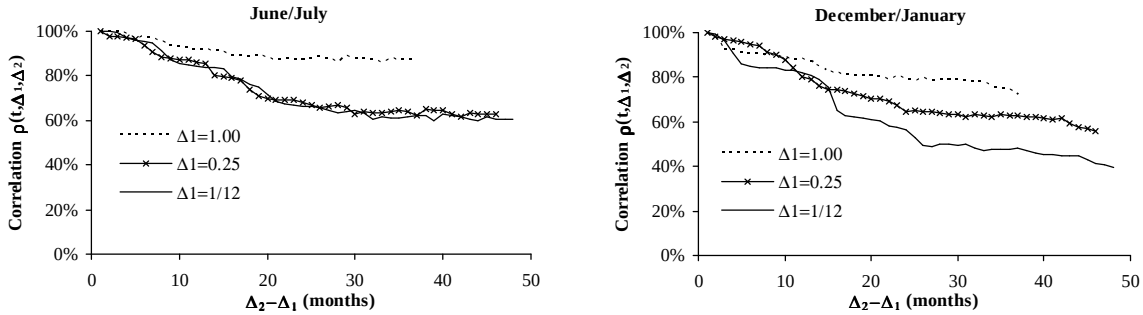
$$\rho(t,\Delta_1,\Delta_2) = \mathrm{corr}\left(dX(t,t+\Delta_1), dX(t,t+\Delta_2)\right). \tag{45}$$

For reasons similar to those described above, one may suspect seasonality in this function, in the sense that $\rho(t,\Delta_1,\Delta_2)$ may depend on t for fixed Δ_1 and Δ_2. To investigate, we

[15]Traded gas futures contracts mature at fixed time *of* maturity dates T spaced one month apart. We have used interpolation between neighboring futures contracts to construct the graph of futures with constant time *to* maturity $T-t$ in Figure 3.

can split our data-sample into observations associated with different calendar-months and compute empirical estimates for ρ. Figure 4 shows results for December/January and June/July, and demonstrates a general trend, namely that correlations between futures observed in the winter are lower than when observed in the summer. Broadly speaking, the correlations $\rho(t,\Delta_1,\Delta_2)$ also tend to decline in $|\Delta_2-\Delta_1|$ and, for fixed $|\Delta_2-\Delta_1|$, increase in $\min(\Delta_1,\Delta_2)$. The precise dependence of correlation on Δ_1 and Δ_2 is, however, season-dependent and overall somewhat "undulating", due to effects associated with $T_1 = t+\Delta_1$ and $T_2 = t+\Delta_2$ moving through the annual seasons.

Figure 4: Empirical Correlation Structure for USD Gas Futures



Notes: The figure shows empirical correlations for log-increments of futures prices observed June/July (left panel) and December/January (right panel). Results were based on daily time-series data covering January 2000 to June 2007.

For later use, and to provide a single concise metric for the dependence of correlations on calendar time t, we define

$$f_\infty(t) = \lim_{\Delta\to\infty} \rho(t,0,\Delta). \tag{46}$$

Using $\rho(t,1/12,4)$ as a rough proxy for this number, Figure 5 below shows empirical estimates for $f_\infty(t)$ for all twelve months of the year.

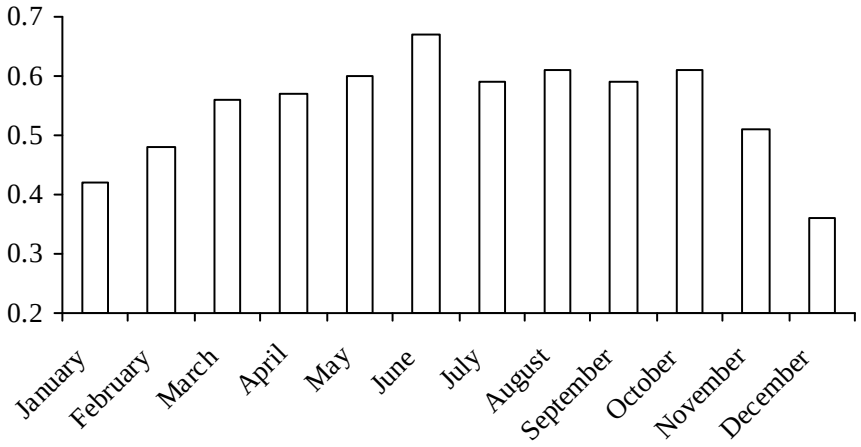
6.2.2 Principal Components Analysis

To investigate how many driving random factors are required for a reasonable model of gas evolution, we now turn to a principal components (PC) analysis of futures price returns. Specifically, we consider the collection of variables

$$X_j(t) = \ln F(t,t+\Delta_j), \quad j = 1 \text{ month}, 2 \text{ months}, \ldots, 48 \text{ months},$$

and are interested in establishing the principal components of daily changes for all 48 X_j's. One important caveat is obvious from Figure 4: due to seasonality effects in the futures price dynamics, standard PC analysis will introduce season-averaging effects that distort the analysis results. To get around this, we proceed as earlier and split the data

Figure 5: Empirical Correlation Asymptotes for USD Gas Futures



Notes: The figure shows empirical estimates of $f_\infty(t)$, as a function of the calendar month to which t belongs. Results were based on daily time-series data covering January 2000 to June 2007.

for the X_j into different t-buckets, with a separate PC analysis performed for daily increments in each bucket. Using the months of January (winter) and July (summer) as two representative buckets, PC analysis results are summarized in Figure 6.

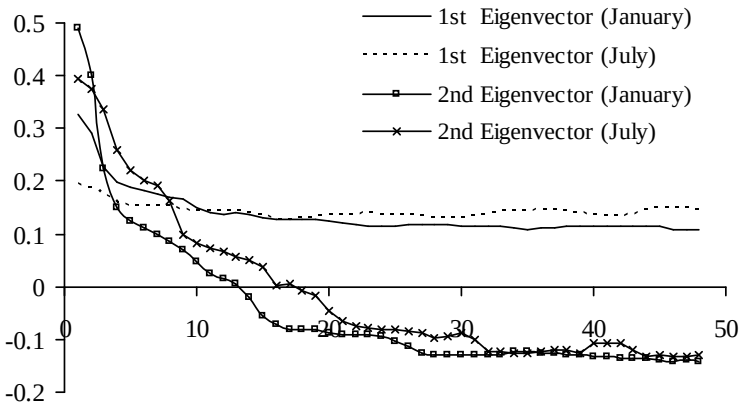
The first principal component of observed futures prices in both summer and winter is a one-directional shift which affects the short end more than the long end. The seasonality effects on this first principal component are somewhat muted, although slight seasonal peaks and valleys (at values of $T_j = t + \Delta_j$ that lie in the winter/summer) can be observed. The first PC alone explains 78% and 84% of the futures curve movements in January and July, respectively. The second PC corresponds to a twist, with the short end of the futures curve moving in opposite direction of the long end. Seasonality is is clearly evident in the second principal component. Together, the first and second principal components explain 95% and 96% of curve movements in January and July, respectively.

We have performed the principal components analysis for all months of the year, but we omit the details since results results for other months are similar to the results for July and January. Across calendar months, a single factor explains about 80% of the variation of the first 48 monthly futures; with two factors, this number goes up to about 95%. The highest percentage of curve variation explained by a single factor is 88% (in June) and the lowest is 76% (in December). As expected from Figure 4, the amount of curve variation explained by the first principal component is higher in summer than in winter.

6.2.3 Implied Volatility

In Figure 7 we have shown time series of implied at-the-money volatility for selected option maturities. Consistent with the decaying term structure of volatility in Figure 1, short-dated options have much higher implied volatility than longer-dated ones. Also,

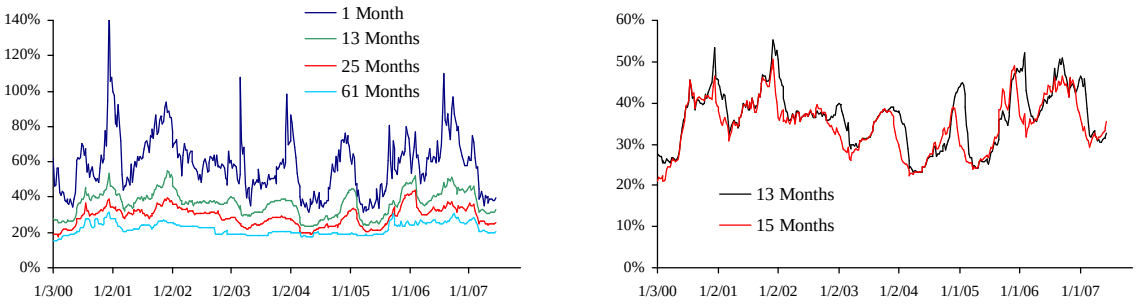
Figure 6: Principal Components Analysis for USD Gas Futures



Notes: The figure shows the first and second eigenvectors computed in a PC analysis of 48 futures price log-increments. The figure contains data from both July (summer) and January (winter). The eigenvalues associated with the first and second eigenvectors were 0.012 and 0.0015 (July) and 0.011 and 0.0023 (January). Results were based on daily time-series data covering January 2000 to June 2007.

there are marked seasonality effects, with winter delivery associated with high implied volatilities.

Figure 7: Selected USD Gas Implied Volatilities



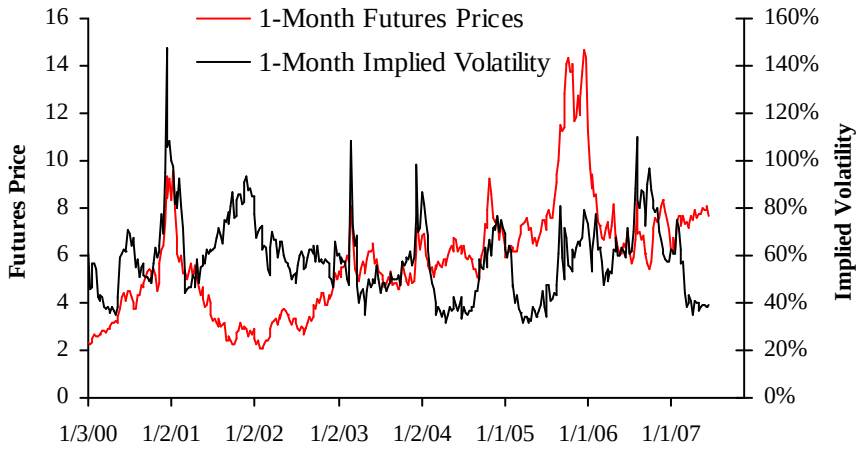
Notes: The figure shows the empirical time series data for implied at-the-money volatilities for options on the spot gas price. Option maturities are as listed in the graphs.

In the left panel of Figure 7, option maturities are spaced one year apart, to ensure that option maturities fall in the same season. In contrast, the right panel examines two options with maturities a few months apart; notice how an increase in maturity causes a "lag" in the time-series, as well as a dampening of the volatility level.

Finally, in Figure 8, we have graphed implied volatility of the 1-month option together with the level of the 1-month futures price. The implied volatility time-series is generally more seasonal than is the futures price, but both have a tendency of being high in the

winter. In particular, large spikes in the futures prices are likely to be accompanied by large movements in the implied volatility. In calmer regimes, the dependency of the two time-series in the graph is more muted, however.

Figure 8: 1-Month USD Gas Implied Volatilities and Futures Prices



Notes: The figure shows the empirical time series data for the one-month gas futures price, along with the one-month implied at-the-money volatility.

7 Markov Diffusion Model

In this section, we temporarily ignore the existence of the volatility skew and aim to build a simple log-normal trading model for derivatives on the gas futures curve. We want the model to calibrate perfectly to observed futures prices and to ATM option prices (see Figure 1), to have (after seasonal corrections) time-stationary dynamics, and overall to broadly comply with the empirical results we presented in Section 6. In particular, we want the option to explicitly capture the seasonal nature of futures price variances and co-variances across the entire term-structure.

7.1 SDE & Futures Curve Reconstitution

For reasons of practicality, we limit ourselves to a setting with two Brownian motions which, as uncovered earlier by a principal components analysis, captures a significant amount of actual gas futures curve movements. Let therefore W_1 and W_2 be independent Brownian motions in the risk-neutral probability measure Q, and write

$$dF(t,T)/F(t,T) = \sigma_1(t,T)dW_1(t) + \sigma_2(t,T)dW_2(t),$$

where σ_1 and σ_2 are deterministic, such that the distribution of all futures prices are log-normal. We further specialize to the case where

$$\sigma_1(t,T) = e^{b(T)} h_1 e^{-\kappa(T-t)} + e^{a(T)} h_\infty; \quad \sigma_2(t,T) = e^{b(T)} h_2 e^{-\kappa(T-t)},$$

where h_1, h_2, h_∞ are constants, $\kappa \geq 0$ is a mean reversion speed, and $a(T)$ and $b(T)$ are seasonality adjustment functions oscillating around 0 at an annual frequency. σ_2 primarily affects the short-end of the futures curve, and decays away at a rate of κ as $T-t$ is increased; σ_1 has a similar decaying term, but also contains a term $(e^{a(T)} h_\infty)$ that persists for long futures maturities. This is consistent with the long-maturity behavior of implied volatilities in Figure 1.

For convenience, let us define $d(T) \triangleq b(T) - a(T)$, such that, in vector notation,

$$dF(t,T)/F(t,T) = e^{a(T)} \begin{pmatrix} h_1 e^{d(T)} e^{-\kappa(T-t)} + h_\infty \\ h_2 e^{d(T)} e^{-\kappa(T-t)} \end{pmatrix}^\top d \begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}. \tag{47}$$

Overall, (47) is consistent with the behavior in Figure 7: volatilities decrease in T due to mean reversion and, through the combined effect of $a(T)$ and $d(T)$, will contain a seasonality "translation" similar to that of the right-hand panel in Figure 7.

We note that the SDE (47) is of the separable type covered in Assumption 1, with

$$\beta(T) = \begin{pmatrix} e^{a(T)+d(T)} e^{-\kappa T} \\ e^{a(T)} \end{pmatrix}, \quad \alpha(t) = \begin{pmatrix} h_1 e^{\kappa t} & h_\infty \\ h_2 e^{\kappa t} & 0 \end{pmatrix}.$$

It follows from results in Section 2.1 that a Markov representation of the futures curve with two state variables is therefore possible. Following the advice of Section 2.2, a good choice of state variables is

$$z_1(t) = e^{-\kappa t} x_1(t), \quad z_2(t) = x_2(t), \tag{48}$$

where $x(t) = (x_1(t), x_2(t))^\top$ is as defined as in Proposition 1. It follows that, with $z_1(0) = z_2(0) = 0$,

$$dz_1(t) = -\kappa z_1(t)\, dt + h_1\, dW_1(t) + h_2\, dW_2(t), \quad dz_2(t) = h_\infty\, dW_1(t). \tag{49}$$

From[16] Proposition 1 we can write $F(t,T) = e^{X(t,T)}$ where

$$\begin{aligned} X(t,T) = {} & \ln F(0,T) + e^{a(T)} \left(z_1(t) e^{-\kappa(T-t)+d(T)} + z_2(t) \right) \\ & - \frac{1}{2} e^{2a(T)} \frac{e^{2d(T)-2\kappa T} \left(e^{2\kappa t} - 1\right) \left(h_1^2 + h_2^2\right) + 4 h_1 h_\infty e^{d(T)-\kappa T} \left(e^{\kappa t} - 1\right) + 2 h_\infty^2 t \kappa}{2\kappa}. \end{aligned} \tag{50}$$

[16] Notice that our model allows for closed-form computation of the deterministic variables $y(t)$ in Proposition 1.

7.2 Some Notes on (49)-(50)

Before proceeding to option pricing and calibration details, let us make a few comments about the model set up so far, First, despite the appearance of the SDEs (49), the results of Section 2.4 imply that the model outlined above generally has time-dependent mean-reversion and volatility when interpreted in a "classical" spot model sense (e.g. as the model in [38]). Indeed, our model implies that $S(t) = F(0,t)\exp(x_1^*(t) + x_2^*(t))$ where

$$dx_1^*(t) = \theta_1(t)\,dt - x_1^*(t)\left(\kappa - a'(t) - d'(t)\right)dt + e^{a(t)+d(t)}\left(h_1 dW_1(t) + h_2 dW_2(t)\right), \quad (51)$$

$$dx_2^*(t) = \theta_2(t)\,dt + x_2^*(t)a'(t)\,dt + e^{a(t)}h_\infty\, dW_1(t), \quad (52)$$

for deterministic functions θ_1 and θ_2 that we, for brevity, do not write out in detail. As discussed in Section 2.2, the choice of state variables z_1 and z_2 ensures that the time-dependence in the dynamics for x_1^* and x_2^* – which originates with the seasonality functions a and d – can be moved into the reconstitution formula (50).

Yet another classical parameterization of the Gaussian two-factor model can be found in [33] (based on [23]) where one state-variable represents the spot price and the other a stochastic convenience yield. As correctly noted by the author[17] calibration to market data will require the introduction of strong time dependence in the mean reversion level and mean reversion speed of the latter process.

As shall be demonstrated shortly, the functions $a(T)$ and $d(T)$ allow for easy calibration of seasonality in volatility and correlations. An alternative way of introducing such seasonality would be to allow for the functions h_1, h_2, h_∞ to depend on calendar time t. It is clear from (51)-(52) that this idea will induce time-dependence only in the volatilities for x_1^* and x_2^* and not in mean reversions; as such, this approach to seasonality modeling will be *weaker* than that employed in the model in Section 7.1 and will not always allow for a perfect fit to ATM volatilities. For demonstration, let us consider the one-factor case where we could contemplate writing (in effect setting $a(T) = d(T) = 0$ for all T)

$$\sigma_F(t,T) = \sigma(t)e^{-\kappa(T-t)}. \quad (53)$$

This model is similar to the classical Hull-White model for interest rates and, at time t, has a T-maturity ATM implied volatility of

$$\sqrt{(T-t)^{-1}e^{-2\kappa T}\int_t^T \sigma(u)^2 e^{2\kappa u}\,du}.$$

Calibration of this model to the market involves finding the function $\sigma(u)$ which equates this expression with the market volatility $\sigma_{impl}(t,T)$. Differentiation with respect to T and rearranging yields

$$\sigma(T)^2 = (T-t)\frac{\partial\left(\sigma_{impl}(t,T)^2\right)}{\partial T} + \sigma_{impl}(t,T)^2\left(2\kappa(T-t)+1\right).$$

[17] [33] is one of the few references in the literature that propose a specific procedure for actual model calibration; his algorithm is less market-oriented (and less convenient) than ours, but shares some of the principles on display in our Section 7.5

As the gradient $\partial\left(\sigma_{impl}(t,T)^2\right)/\partial T$ can often be very negative (see Figure 1), unless κ is (unreasonably) large the right-hand side of this expression will be negative for some values of T, and the model (53) cannot be calibrated. In contrast, we shall see shortly that the model in Section 7.1 can always be calibrated to ATM volatilities, even for the one-factor case (where $h_\infty = h_2 = 0$).

Of course, we could allow for a combined approach where the functions $a(T)$ and $d(T)$ are retained *and* h_1, h_2, h_∞ are allowed depend on calendar time t. Such an extension of our model would still allow for a two-dimensional Markov representation and would also permit one to incorporate non-stationarity volatility behavior, should one desire to do so. Empirical data is somewhat ambiguous on how seasonality on implied volatility should be split into t- and T-dependence in futures price volatilities; our choice in Section 7.1 is not inconsistent with the data[18] and is certainly easy to work with. The reader should nevertheless keep in mind that additional flexibility exists to, say, let h_1, h_2, h_∞ be smooth periodic functions of time.

Finally, let us note that the model in Section 7.1 involves two different mean reversion speeds: one equal to κ and one equal to zero. The second mean reversion does, in fact, not have to be set precisely equal to zero – the main requirement is that it needs to be small relative to κ in order to explain the two separate time-scales evident in the shape of the volatility term structure in Figure 1.

7.3 Term Volatility & Correlation Structure

To characterize the volatility and correlation properties of the model (47), let us first note that the total instantaneous volatility of $F(t,T)$ is

$$\begin{aligned}\sigma_F(t,T) &\triangleq \sqrt{\sigma_1(t,T)^2 + \sigma_2(t,T)^2} \\ &= e^{a(T)}\sqrt{\left(h_1^2 + h_2^2\right)e^{-2\kappa(T-t)+2d(T)} + 2h_\infty h_1 e^{-\kappa(T-t)+d(T)} + h_\infty^2}.\end{aligned}$$

For option pricing purposes, we are less interested in instantaneous volatility than in the net quadratic variation of the futures price on a finite interval. The corresponding average volatility is denoted the *term volatility* for the given interval. For the T-maturity futures contract, the term volatility experienced on the interval $[T_1, T_2]$ is therefore defined by (with $T_1 \leq T_2 \leq T$)

$$\sigma_{term}^2(T_1, T_2; T) = (T_2 - T_1)^{-1}\int_{T_1}^{T_2}\sigma_F(t,T)^2 dt. \tag{54}$$

Under the terms of Lemma 3, $\sigma_{term}(t,T;T)$ is the implied volatility at time t for an option on the *spot* gas price delivered at time T; these volatilities are the ones shown in Figure 1 earlier. If an option contract expiring at time T' calls for delivery of a futures contract $F(T',T)$, $T > T'$, the relevant implied volatility at time t will be[19] $\sigma_{term}(t,T';T)$.

[18] Overall, implied volatilities for gas futures prices with locked time-of-maturity T here displayed no clear-cut seasonality effects.

[19] Notice that in Lemma 3 $v(T',T)^2 = T'\sigma_{term}^2(0,T';T)$.

Lemma 10 *Let $\sigma^2_{term}(T_1,T_2;T)$, $\rho(t,\Delta_1,\Delta_2)$ and $f_\infty(t)$ be as defined in (54), (45), and (46), respectively. For the model (47), we have*

$$\sigma^2_{term}(T_1,T_2;T) = e^{2a(T)}\left\{\left(h_1^2+h_2^2\right)e^{2d(T)}\frac{e^{-2\kappa(T-T_2)}-e^{-2\kappa(T-T_1)}}{2\kappa(T_2-T_1)} + 2h_\infty h_1 e^{d(T)}\frac{e^{-\kappa(T-T_2)}-e^{-\kappa(T-T_1)}}{\kappa(T_2-T_1)} + h_\infty^2\right\};$$

$$f_\infty(t) = \frac{qe^{d(t)}+w}{\sqrt{w}\sqrt{e^{2d(t)}+2qe^{d(t)}+w}};$$

$$\rho(t,\Delta_1,\Delta_2) = \frac{e^{d(T_1)}e^{d(T_2)}e^{-\kappa(\Delta_1+\Delta_2)}+q\left(e^{d(T_1)}e^{-\kappa\Delta_1}+e^{d(T_2)}e^{-\kappa\Delta_2}\right)+w}{\sqrt{e^{2d(T_1)}e^{-2\kappa\Delta_1}+2qe^{d(T_1)}e^{-\kappa\Delta_1}+w}\sqrt{e^{2d(T_2)}e^{-2\kappa\Delta_2}+2qe^{d(T_2)}e^{-\kappa\Delta_2}+w}};$$

where $T_1 = t+\Delta_1$ and $T_2 = t+\Delta_2$, and

$$q \triangleq \frac{h_1h_\infty}{h_1^2+h_2^2}, \quad w \triangleq q\frac{h_\infty}{h_1}.$$

It is interesting to note that $\rho(t,\Delta_1,\Delta_2)$ and $f_\infty(t)$ are *independent* of the seasonality function a and only depend on t through the function $d(t)$. We can use this observation two ways: first, if we wish to use a time-stationary correlation structure we set $d(t)=0$ for all t. Second, if we know the (periodic) function $f_\infty(t)$ we can back out $d(t)$ by inverting the result for $f_\infty(t)$ in Lemma 10; should a solution for $d(t)$ exists, it is given by

$$e^{d(t)} = \frac{q(1-f_\infty(t)^2)+f_\infty(t)\sqrt{(1-f_\infty(t)^2)(w-q^2)}}{f_\infty(t)^2-q^2/w}. \tag{55}$$

7.4 A Useful Reparameterization

Apart from the seasonality functions $a(T)$ and $d(T)$ and the mean reversion parameter κ, parameterization of the model in (47) is done through selection of the three constants h_1,h_2,h_∞. For the sake of intuition, it is often convenient to recast the model in terms of more palatable constants. To do this, assume with no loss of generality that $h_\infty \geq 0$ and momentarily remove seasonality, i.e. set $a(T)=d(T)=0$. In this case, define

$$\sigma_0 \triangleq \sigma_F(t,t) = \sqrt{(h_1+h_\infty)^2+h_2^2}, \tag{56}$$

$$\sigma_\infty \triangleq \sigma_F(t,\infty) = h_\infty. \tag{57}$$

In other words, σ_0 and σ_∞ are the volatilities of the long and the short end of the futures curve, in the absence of seasonality. Also define ρ_∞ as in (58), i.e. as the correlation between the spot price and the long-end of the futures curve. When $d(T)=0$, we have

$$\rho_\infty = \frac{q+w}{\sqrt{1+2q+w}\sqrt{w}} = \frac{h_1+h_\infty}{\sigma_0}. \tag{58}$$

Above, we showed how to compute $\sigma_0, \sigma_\infty, \rho_\infty$ from h_1, h_2, h_∞. These equations can be inverted, allowing us to compute h_1, h_2, h_∞ from given values of $\sigma_0, \sigma_\infty, \rho_\infty$. Specifically, one can use

$$h_\infty = \sigma_\infty, \quad h_1 = -\sigma_\infty + \rho_\infty \sigma_0, \quad h_2 = \sigma_0 \sqrt{1-\rho_\infty^2}. \tag{59}$$

Going forward we will use both parameterizations (56)-(58) and (59) interchangeably.

7.5 Model Calibration

As our model takes the term structure of futures prices as exogenously given quantities, the model (47) is, by construction, in calibration with the term structure of futures prices at time 0. To complete the calibration of the model, it remains to select the parameters $\sigma_0, \sigma_\infty, \rho_\infty, \kappa$, as well as the two functions $a(T)$ and $d(T)$, in such a way that the model has reasonably dynamics and is perfectly calibrated to market-observed implied volatilities for at-the-money options. With "reasonable" dynamics, we here have two criteria in mind. First, the volatility dynamics should be close to time-stationary, in the sense that the term structure of at-the-money volatilities should remain stable over time, when adjusted for seasonality effects. Second, to the extent practical, the correlation $\rho(t, \Delta_1, \Delta_2)$ should be in agreement with empirical data. A possible approach proceeds as follows.

1. Pick a value for ρ_∞, based on empirical data or other considerations.
2. Temporarily set $a(T)$ and $d(T)$ to 0 (i.e. remove seasonality) from the model. Applying Lemma 10 to compute term volatilities $\sigma_{term}(t, T; T)$, use a numerical optimizer to find the values for $\sigma_0, \sigma_\infty, \kappa$ that produce a term structure of at-the-money volatilities volatilities that is as close as possible to market data.
3. Either specify $d(T)$ directly, or determine from empirical data the function $f_\infty(t) = \rho(t, 0, \infty)$, and use (55) to back out d. If we are content with a time-stationary correlation structure, set $d(T) = 0$ (or, equivalently, $f_\infty(t) = \rho_\infty$ for all t).
4. Find the function $a(T)$ by matching the model to the term structure of market-implied at-the-money volatilities.

Several comments are in order here. First, in Step 1 of the algorithm, we can take guidance in the correlation analysis conducted earlier, in Section 6.2.1. Recalling that ρ_∞ denotes the correlation between the long and short end of the futures, a value of $\rho_\infty = 0.5$ appears reasonably consistent with the data in Figure 5.

In Step 2 of the algorithm, the main purpose of the least-squares fit is to produce a model that will be as time-stationary as possible. In particular, the hope is that choosing $\sigma_0, \sigma_\infty, \kappa$ to provide a close option fit on a seasonality-free basis, the seasonality functions $a(T)$ and $d(T)$ will ultimately both end up being largely trend-free, ensuring that the model implies a stable evolution of futures volatilities over time. Whether this goal is realized will depend in part on the choice of $d(T)$ in Step 3, the primary role of which is to provide a mechanism for matching correlation seasonality. If one considers such

seasonality of secondary importance, it is recommended to set $d(T) = 0$; we give a simple example for a non-flat parameterization later in the paper.

The last step of the algorithm, Step 4, follows directly from the result for $\sigma_{term}(t, T; T)$ in Lemma 10 and requires no root-search:

$$a(T) = \ln \sigma_{impl}(t, T; T) - \frac{1}{2} \ln \left\{ \left(h_1^2 + h_2^2\right) e^{2d(T)} \frac{1 - e^{-2\kappa(T-t)}}{2\kappa(T-t)} + 2h_\infty h_1 e^{d(T)} \frac{1 - e^{-\kappa(T-t)}}{\kappa(T-t)} + h_\infty^2 \right\}.$$

Here, $\sigma_{impl}(t, T; T)$ is the time t market-observed volatility for a T-maturity at-the-money option on the spot gas price.

7.6 A Calibration Example

To give an example of application of the algorithm above, we now return to our market data snapshot of April 2007 (see Figure 1). Based on Figure 5, we select $\rho_\infty = 0.5$ (Step 1); a subsequent fitting procedure (Step 2) then gives $\kappa = 1.35$, $\sigma_0 = 0.50$, $\sigma_\infty = 0.17$. From the results in Section 7.4, we can compute that then $h_1 = 0.08$, $h_2 = 0.43$, and $h_\infty = 0.17$. To demonstrate the effects from the selection of $d(T)$, in Step 3 of our calibration algorithm we use two choices for $f_\infty(t)$: i) $f_\infty(t) = \rho_\infty = 0.5$, and ii) $f_\infty(t) = l(t)$, where $l(t)$ is a sine-function loosely fitted to the empirical data in Figure 5:

$$l(t) = 0.5 + 0.1 \sin(2\pi(t - 0.4)). \tag{60}$$

Using the standard convention that $t = 0$ at our snapshot date in April 2007, this specification lets $f_\infty(t)$ vary between 0.4 (winters) and 0.6 (summers) in periodic fashion. The resulting function $d(T)$ is itself close to sinusoidal, with peaks and valleys in the winter and summer months, respectively.

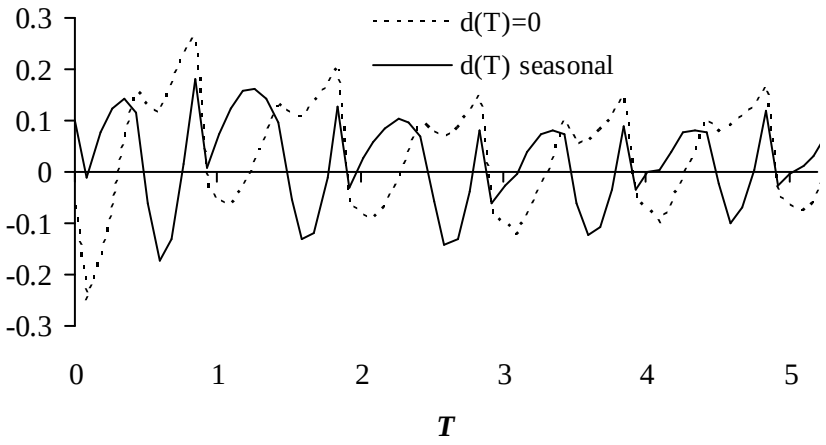
In Figure 9 we show the resulting fit for $a(T)$ to ATM volatilities. For both choices of $d(T)$, $a(T)$ is reasonably close to periodic, although the amplitude has a slight decay in the first year. In Figure 10, for[20] $d(T) = 0$ we show the predicted term volatilities 1.5 and 2 years from today; the model is close to stationary, after appropriate adjustment for the seasonality effect.

Setting $d(T)$ as implied by (60), Figure 11 shows (using Lemma 10), the predicted correlation $\rho(t, \Delta_1, \Delta_2)$ for summer- and winter-settings of t, and for values of Δ_1 and Δ_2 similar to those in Figure 4. Qualitatively, the model does a decent job of replicating empirical data, although seasonality effects in the model appear slightly more pronounced than in the empirical data of Figure 4; a closer fit can, if desired, be accomplished through a more precise specification of $d(T)$.

Finally, in Figure 12 we show $\rho(t, 1/12, \Delta)$ – the asymptote of which is a decent estimate for $f_\infty(t)$ – for summer- and winter-settings of t. As discussed earlier, for $f_\infty(t) =$

[20] For the case where $d(T)$ is found by fitting to $l(t)$ in (60), the stationarity properties of the model are, if anything, better than those in Figure 10, so we omit the graph for this case.

Figure 9: $a(T)$ for USD Gas Options



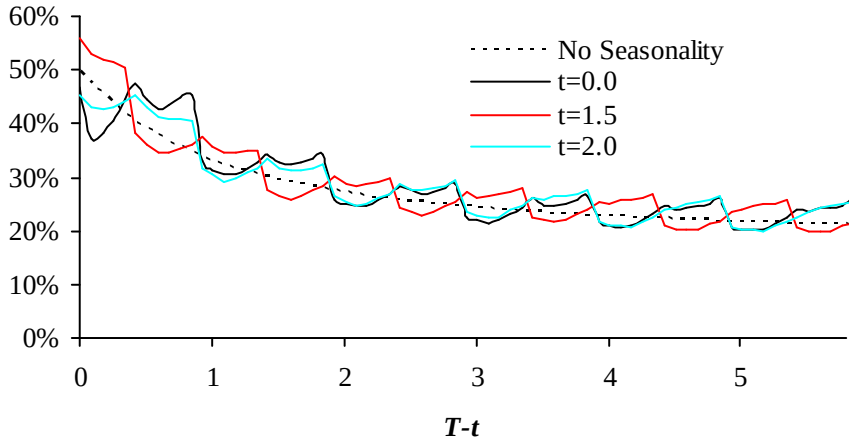
Notes: The figure shows $a(T)$, calibrated to USD gas options in April 2007. Other parameters were set to: $h_1 = 0.08$, $h_2 = 0.43$, $h_\infty = 0.17$, $\kappa = 135\%$; this implies that $\sigma_0 = 0.5$ and $\rho_\infty = 0.5$. The graph shows two cases: i) $d(T) = 0$ and ii) $d(T)$ set to match the correlation seasonality implied by the function $l(t)$ in equation (60).

$l(t)$, $\rho(t,0,\Delta)$ is an oscillating function of Δ, the asymptote of which is t-dependent. This behavior is similar to the empirical observations in Figure 4, as is emphasized in the second panel of Figure 12.

8 Smile Modeling

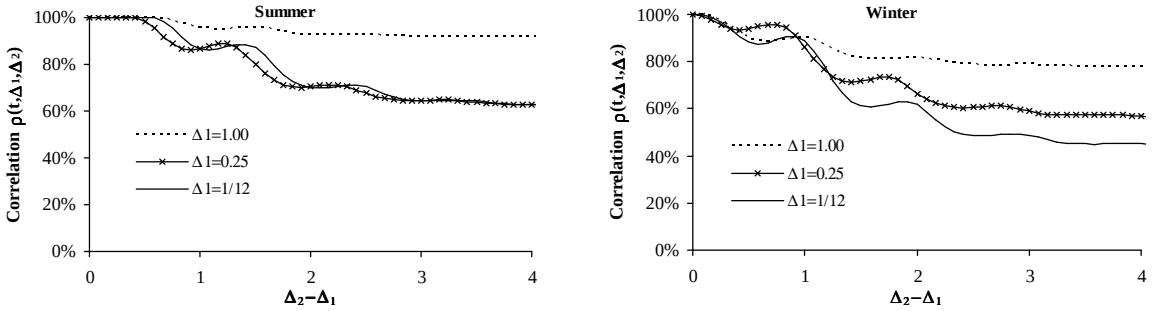
The model introduced so far implies log-normal futures prices and will produce a volatility smile that is constant across strikes. As discussed earlier, however, gas markets display an inverted volatility skew, where high-strike options trade at high implied volatilities, particularly for short-dated options. While there are many ways of modeling this volatility skew, we will here focus on adding regime-switch spikes of the type discussed in Section 5.4. Our reasons for this are threefold. First, spikes are a genuine characteristic of empirical gas data, as evidenced by Figure 3. Second, we have a stated goal of using only three state variables, which will rule out stochastic volatility models (the total number of state variables would balloon to at least 6; see Remark 4). And third, existing literature on stochastic/local volatility and jump-diffusion applications in finance is significantly more developed than for regime-switching, justifying us paying some additional attention to the latter class at the expense of the former two. Finally, we also note that the call option pricing formulas that emerge from the combination of the model (47) with regime-switching are very convenient for practical model calibration. In a truly sophisticated application, it will nevertheless not be a bad idea to add stochastic volatility to the model, in as much as USD natural gas volatility obviously *is* stochastic – see Figures 7 and 8 – and is (mod-

Figure 10: ATM Term Volatilities



Notes: Using the model parameters of Figure 9 with $d(T) = 0$, the figure shows the at-the-money term volatility $\sigma_{term}(t,T;T)$ as function of $T-t$ at two different points in time. For reference, the term volatilities corresponding to $a(T) = 0$ are also shown ("No Seasonality").

Figure 11: Correlation Structure



Notes: Using the model parameters in Figure 9 with $d(T)$ implied from (60), the figure shows model-predicted correlation $\rho(t,\Delta_1,\Delta_2)$ for various values of Δ_1 and Δ_2. In the left panel $t = 0.1$ (summer); in the right panel $t = 0.6$ (winter).

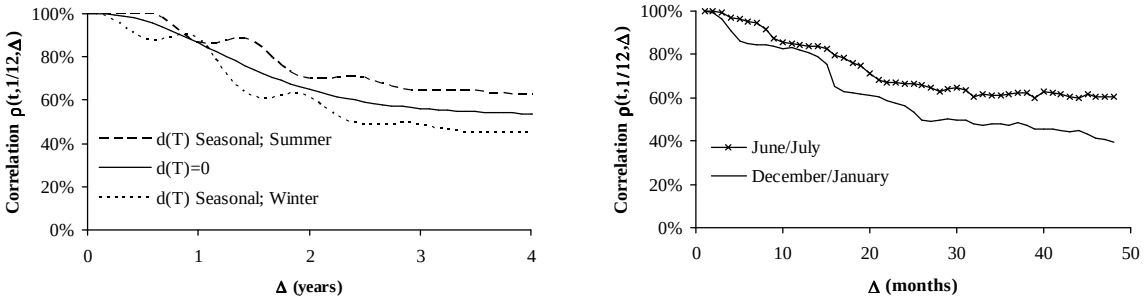
erately) positively correlated to futures prices. We leave detailed examination of such a model to future work.

8.1 Two-State Regime Switch Model

Working in the setup of Section 5.4, we consider a simple two-state regime switch model, where

$$F(t,T) = F_c(t,T)F_J(t,T) = F_c(t,T)\frac{\mathrm{E}_t\left(e^{s(T)J(T)}\right)}{\mathrm{E}\left(e^{s(T)J(T)}\right)}, \tag{61}$$

Figure 12: Correlation Structure



Notes: Using the model parameters in Figure 12, the left panel shows model-predicted values of $\rho(t,1/12,\Delta)$ for $t=0.1$ (summer) and $t=0.6$ (winter). For reference, results for the case $d(T)=0$ are also included. For reference, the right panel shows the relevant empirical data.

where F_c is generated by a two-dimensional diffusion model of the type (47), $s(T)$ is a jump-seasonality function, and J is a regime-switching "spike" process of the type covered by Assumption 4 in Section 5.5. For simplicity, we assume that the Markov chain $c(t)$ driven transitions in J has only two states, $c(t)=e_1$ (base state) and $c(t)=e_2$ (excited state). As in Section 5.4, moves from the base state to the excited state is characterized by a deterministic intensity $n_2(t)$; and moves back to the base state is characterized by a deterministic intensity $n_1(t)$. Equation (36) can be used in general to work out transition probabilities for the chain. If n_1 and n_2 are constant, we can follow the advice in Remark 6 to compute the probabilities analytically. We list the result below for convenience.

Lemma 11 *Assume that n_1 and n_2 are constants, and set $n_{tot}=n_1+n_2$. Then*

$$p_1^1(t,T)=\left(n_1+n_2e^{-n_{tot}(T-t)}\right)n_{tot}^{-1},\quad p_2^1(t,T)=\left(n_2+n_1e^{-n_{tot}(T-t)}\right)n_{tot}^{-1},$$

and $p_1^2(t,T)=1-p_1^1(t,T)$ and $p_2^1(t,T)=1-p_2^2(t,T)$.

If $c(t)$ jumps to e_2 then the $J(t)$ will simultaneously jump to a random value drawn from a Gaussian distribution $\mathcal{N}(\mu_J,\gamma_J)$, where either $\gamma_J>0$ or $\mu_J\neq 0$. From Proposition 4, we get the following result (with $\tilde{n}_1=n_1(T-t)$)

$$\mathrm{E}_t\left(e^{s(T)J(T)}\right)=\begin{cases} p_1^1(t,T)+p_1^2(t,T)e^{\mu_J s(T)+s(T)^2\gamma_J^2/2}, & J(t)=0,\\ p_2^1(t,T)+\left(p_2^2(t,T)-e^{-\tilde{n}_1}\right)e^{\mu_J s(T)+s(T)^2\gamma_J^2/2}+e^{-\tilde{n}_1}e^{s(T)J(t)}, & J(t)\neq 0,\end{cases} \tag{62}$$

which can be used to reconstitute $F_J(t,T)$ from $J(t)$ in accordance with (61). For any value of $T\geq t$, $F(t,T)$ can be computed from knowledge of the *three* Markov state variables $z_1(t),z_2(t),J(t)$, with $z_1(t)$ and $z_2(t)$ as given in (49).

Pricing of call options on the spot gas price can proceed according to Proposition 5; note that the resulting closed-form pricing formula will involve a probability-weighted sum of two terms of the Black-Scholes type. Computation time is obviously negligible.

8.1.1 Some Numerical Tests

The combination of regime-switching with continuous dynamics give us a range of parameters ($n_1(t)$, $n_2(t)$, μ_J and σ_J) that we can use to calibrate the volatility smile surface. On top of this, we now have *three* T-dependent seasonality sources in the model: the functions $a(T)$, $d(T)$, and $s(T)$.[21] In practice, we would likely parameterize two of these functions directly, and solve for the last one to perfectly match at-the-money volatilities.

While a detailed examination of calibration and estimation algorithms for the regime-switch model is outside the scope of this paper, a pragmatic approach is to use some parametric periodic functions for $s(T)$ and, if desired, $d(T)$ (and possibly even for $n_2(t)$) to get a good overall fit to the volatility surface; the function $a(T)$ can then be used subsequently to ensure a perfect match to ATM volatilities. Other calibration approaches are also possible, and it may be possible to replace direct parameterization of $s(T)$ with some type of bootstrap algorithm.

An example of the type of skews that arise from our regime-switch models is shown in Figure 13. The fit was done casually, but still provides a decent match to actual market data, although the short-term skew is generally a bit on the high side and the long-term skew a bit on the low side. We also notice the introduction of small artifacts into the skew from the jagged shape of ATM volatilities. The skew fit can be improved in several ways, including i) introduction of more types of jumps; ii) adding a local volatility component; or iii) adding stochastic volatility. Such model improvements typically come with costs in terms of additional complexity in estimation and computation, and must be weighed carefully against the need to accurately replicate given market skews. In our case, it should obviously be considered that the volatility skew estimates provided by the desk are rather crude, as should be evident from, say, Figure 2.

8.2 A Note on Numerical Implementation

While we have little room to cover numerical model implementation, it should be noted that Monte Carlo implementation of our three-factor model outlined in Section 8 is straightforward, as the three state variables z_1, z_2, J are all easy to simulate through time. In particular, path generation for $J(t)$ involves nothing but simulation of a two-state Markov chain – which is trivial, given the explicit result in Lemma 11 – coupled with draws a Gaussian sample whenever the chain transitions into its excited state.

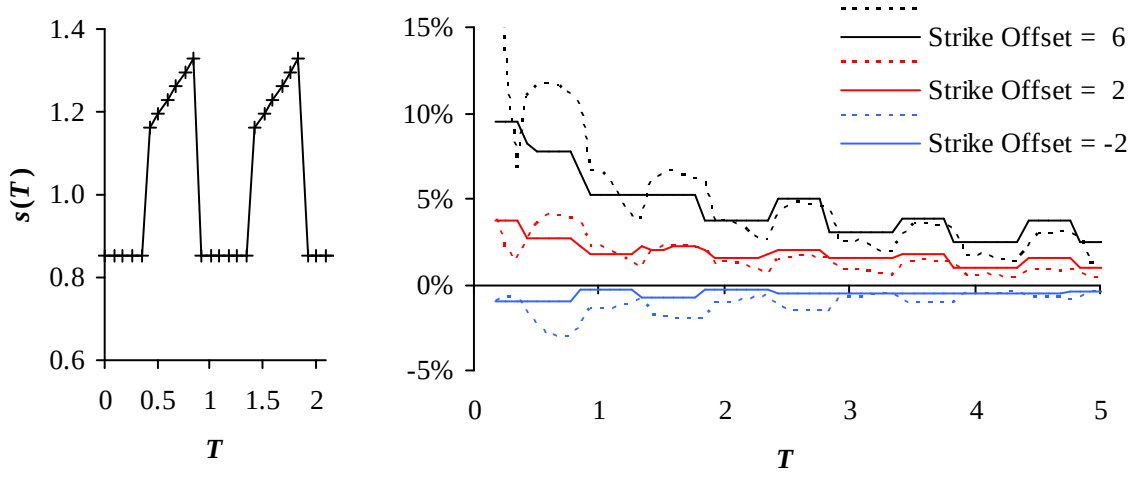
For securities that are most easily priced by backward induction (e.g. American options), a finite difference application is often preferable. The prototypical problem to be solved involves establishing the function

$$A(t, z_1, z_2, J) = \mathrm{E}(k(z_1(T), z_2(T), J(T)) \,|\, z_1(t) = z_1, z_2(t) = z_2, J(t) = J),$$

where k is some payout function at time T, typically involving the reconstitution formulas

[21] In addition, we have the ability to induce further skew seasonality by letting intensities n_1 and n_2 vary periodically with time. In practice, however, the amount of volatility seasonality that can be induced this way is quite limited, except from the short end of the volatility term structure.

Figure 13: Skew Fit for USD Gas Options



Notes: The right panel shows the volatility skew (the difference between the implied volatility minus the ATM volatility) versus option maturity T. Solid lines are desk-estimated market skews; dotted lines are model fits. In the fit, the function $s(T)$ was the simple periodic function in the left panel; and $a(T)$ was set non-parametrically to match ATM volatilities in Figure 1. Other diffusion model parameters were as in Figure 10, with exception of σ_0 which was changed to 40%. Spike parameters were: $\mu_J = 0.5$, $\gamma_J = 0.3$, $n_1 = 2$, $n_2 = 0.3$.

(50) and (61)-(62). Standard methods[22] show that $A(t, z_1, z_2, J)$ then satisfies the PIDE

$$(\mathscr{L}_c + \mathscr{L}_J) A(t, z_1, z_2, J) = 0, \tag{63}$$

with

$$\mathscr{L}_c = \frac{\partial}{\partial t} - \kappa \frac{\partial}{\partial z_1} + \frac{1}{2}\left(h_1^2 + h_2^2\right)\frac{\partial^2}{\partial z_1^2} + \frac{1}{2}h_\infty^2 \frac{\partial^2}{\partial z_2^2} + h_1 h_\infty \frac{\partial^2}{\partial z_1 \partial z_2},$$

$$\mathscr{L}_J A(t, z_1, z_2, J) = 1_{J=0} n_2(t) \left\{ \frac{1}{\gamma_J \sqrt{2\pi}} \int_{\mathbb{R}} A(t, z_1, z_2, q) e^{-\frac{1}{2}\left(\frac{q-\mu_J}{\gamma_J}\right)^2} dq - A(t, z_1, z_2, 0) \right\}$$
$$+ 1_{J\neq 0} n_1(t) \left\{ A(t, z_1, z_2, 0) - A(t, z_1, z_2, J) \right\},$$

subject to the terminal condition $A(T, z_1, z_2, J) = k(z_1, z_2, J)$. Several good finite-difference methods for PIDEs can be found in the literature; see [16] and [3] for details. As PIDEs go, (63) is quite easy to deal with, since only the state $J = 0$ is associated with an integral operation.

[22] Write down an expression for $dA(t)$ and set its expectation equal to zero.

9 Conclusions

In this paper we introduced a broad framework for the construction of Markov models the pricing and risk management of commodity derivatives. We take the term structure of futures prices as exogenously given, and explicitly model a number of empirical effects associated with commodity price behavior, including stochastic volatility, jumps, and spikes. It is our belief that the risk-neutral dynamics of most traded commodities markets can be captured accurately by specific subsets of our general model universe; actual parameterization choices will, however, depend strongly on the specific commodities under consideration, as well as the computational budget available. While an exhaustive analysis of parameterization strategies for a collection of individual commodities markets is beyond the scope of any single paper, we attempted to demonstrate some general parameterization strategies by developing a trading model for USD natural gas. While quite simple, the three-factor model we describe calibrates well to observed option data and to empirical co-variation characteristics of the natural gas futures curve; we believe that the model illustrates well the richness and practical relevance of our broader theoretical setup.

Much work remains, of course, not only in the development of model parameterization strategies for markets other than natural gas (or, for that matter, in the extension of the simple model for gas given in this paper), but also in the development of pricing methods for specific securities. By emphasizing a finite-dimensional Markov structure, our framework is generally quite amenable to numerical methods (see Section 8.2), but the complexity and path-dependence of certain actual commodity derivatives – swing options in the electricity markets come to mind – are challenging in any model setting. Another interesting area of future research concerns itself with extensions to multi-commodity securities, e.g. spread options.

A Appendix: Analysis of Blix Gas Model

In [6] and [7], the following specification is used for $\sigma(t,T) = (\sigma_1(t,T), \sigma_2(t,T))^\top$:

$$\sigma_1(t,T) = c_1 e^{-\kappa_1(T-t)};\ \sigma_2(t,T) = \gamma + c_2 e^{-\kappa_2(T-t)} \sin(\phi + 2\pi t) + c_3 e^{-\kappa_3(T-t)} \sin(\phi + 2\pi T).$$

In the notation of Section 2.1, we can generate this model by setting $N = 4$ and writing

$$\alpha(t) = \begin{pmatrix} c_1 e^{\kappa_1 t} & 0 & 0 & 0 \\ 0 & \gamma & c_2(\sin(\phi + 2\pi t)) e^{\kappa_2 t} & c_3 e^{\kappa_3 t} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};\ \beta(T) = \begin{pmatrix} e^{-\kappa_1 T} \\ 1 \\ e^{-\kappa_2 T} \\ e^{-\kappa_3 T} \sin(\phi + 2\pi T) \end{pmatrix}.$$

With $W = (W_1, W_2, W_3, W_4)^\top$, we emphasize that the Brownian motions W_3, W_4 are superfluous. The martingale state variables for this model are

$$\begin{aligned} dx_1(t) &= c_1 e^{\kappa_1 t}\, dW_1(t), & dx_2(t) &= \gamma\, dW_2(t), \\ dx_3(t) &= c_2 \sin(\phi + 2\pi t)\, e^{\kappa_2 t}\, dW_2(t), & dx_4(t) &= c_3 e^{\kappa_3 t}\, dW_2(t). \end{aligned}$$

All y's are here deterministic, so the effective number of random state variables is 4. To change the martingale representation of state variables to something mean-reverting, we can change variables:

$$z_1(t) = x_1(t)e^{-\kappa_1 t}; \quad z_2(t) = x_2(t), \quad z_3(t) = x_3(t)e^{-\kappa_2 t}; \quad z_4(t) = e^{-\kappa_3 t}x_4(t).$$

Then

$$\begin{aligned} dz_1(t) &= -\kappa_1 z_1(t)\,dt + c_1\,dW_1(t), & dz_2(t) &= \gamma\,dW_2(t), \\ dz_3(t) &= -\kappa_2 z_3(t)\,dt + c_2\sin(\phi + 2\pi t)\,dW_2(t), & dz_4(t) &= -\kappa_3 z_4(t)\,dt + c_3\,dW_2(t). \end{aligned}$$

In [6] the problem of finding a Markov representation for the model above is attacked with a Lie algebraic approach. This method – which is considerably more involved than our direct constructive approach – ultimately produces 5 random Markov state-variables, one of which is locally deterministic. More precisely, their representation is identical to ours, except that our variable z_3 is substituted for the linear system

$$dq_1(t) = (2\pi q_2(t) - \kappa_2 q_1(t))\,dt + c_2 dW_2(t), \tag{64}$$

$$dq_2(t) = -(\kappa_2 q_2 + 2\pi q_1(t))\,dt. \tag{65}$$

Let us try to understand how the representation in [6] relates to ours. First, the linear SDE system above is characterized by a drift matrix of

$$A = \begin{pmatrix} -\kappa_2 & 2\pi \\ -2\pi & -\kappa_2 \end{pmatrix}.$$

A has complex-valued eigenvalues $-\kappa_2 \pm 2\pi i$, and a little work shows that

$$e^{-At} = \begin{pmatrix} e^{\kappa_2 t}\cos(2\pi t) & -2\pi e^{\kappa_2 t}\sin(2\pi t) \\ 2\pi e^{\kappa_2 t}\sin(2\pi t) & e^{\kappa_2 t}\cos(2\pi t) \end{pmatrix}.$$

If we set $v = \left(\sin\phi, (2\pi)^{-1}\cos\phi\right)^\top$ and define $x_3(t) \triangleq v^\top e^{-At}q(t)$, we get

$$dx_3(t) = ce^{\kappa_2 t}\left(\sin\phi\cos(2\pi t) + \cos\phi\sin(2\pi t)\right)dW_2(t) = ce^{\kappa_2 t}\sin(\phi + 2\pi t)\,dW_2(t),$$

where we have used the relation $\sin(x+y) = \sin x\cos y + \cos x\sin y$. Comparison with results above shows that our variable z_3 can be computed uniquely from the variables q_1 and q_2 through a linear sum with (complicated) time-dependent terms. As we have shown above, only $z_3(t)$ (and not the individual components of q_1 and q_2) is, in fact, required for reconstitution of the futures curve, so the parameterization in [6] and [7] is not minimal. On the other hand, the dynamics of all state-variables in [6]–[7] are linear with constant coefficients, whereas our formulation involves a time-dependent term.

B Appendix: Dimension Reduction of Spike Markov Chain

In the spike model of Section 5.4, consider now the special case where the $m-1$ arrival rates $n_j(t)$ $j=2,\ldots,m$ have the form

$$n_j(t) = n(t)r_j, \quad j=2,\ldots,m, \tag{66}$$

where $n(t)$ is a scalar deterministic function and the r_j's are constants. Effectively, we here assume that all types of jumps are characterized by the same shared seasonality in intensities. The transition of the process $J(t)$ in Section 5.4 away from zero is $n_{tot}(t) = n(t)\sum_{j\geq 2} r_j$. The total jump density for states other than $J=0$ can be constructed as a mixture distribution

$$d_{tot}(z) = \sum_{j=2}^{m} \frac{n_j(t)}{n_{tot}(t)} d_j(z) = \left(\sum_{j=2}^{m} r_j\right)^{-1} \sum_{j=2}^{m} r_j d_j(z). \tag{67}$$

But our model has now effectively been reduced to a Markov chain with only *two* states, one state where J is equal to zero and one state where J is equal to "everything else". The latter state has Markov chain arrival rate $n_{tot}(t)$ and, as required for the models covered in Section 5.4, time-independent jump-density $d_{tot}(z)$.

References

[1] Andreasen, Jesper (2005), "Back to the future," *RISK*, September, pp. 104-109.

[2] Andreasen, J., P. Collin-Dufresne, and W. Shi (1997), "An arbitrage model of the term structure of interest rates with stochastic volatility," Proceedings of the French Finance Association AFFI 97, Grenoble.

[3] Andreasen, J. and M. Dahlgren (2006), "At the flick of a switch," *Energy Risk*, February, pp. 71-75.

[4] Beaglehole, D. and A. Chebanier (2002), "A two-factor mean-revering model," *RISK*, July, pp. 65-69.

[5] Bessembinder, J., J. Coughenour, P. Seguin and M. Smoller (1995), "Mean reversion in equilibrium asset prices: evidence from the futures term structure," *Journal of Finance*, 50, pp. 361-375.

[6] Bjork, T., M. Blix and C. Landen (2006), "On finite dimensional realizations for the term structure of futures prices," *International Journal of Theoretical and Applied Finance*, 9, pp. 281-314.

[7] Blix, M. (2003), "A Gas Futures Model," Working Paper, Stockholm School of Economics, Sweden.

[8] Casassus, J. and P. Collin-Dufresne (2005), "Stochastic convenience yield implied from commodity futures and interest rates," *Journal of Finance*, 60, pp. 2283-2331.

[9] Cheyette, O. (1991), "Markov representation of the Heath-Jarrow-Morton model," Working Paper, BARRA.

[10] Clewlow L. and C. Strickland (1999) "A multi-factor Model for Energy Derivatives," Working Paper, University of Technology, Sydney.

[11] Cortazar, G. and E. Schwartz (1994), "The valuation of commodity contingent claims," *Journal of Derivatives*, 1, pp. 27-39.

[12] Crosby, J. (2006), "Pricing a class of exotic commodity options in a multi-factor jump-diffusion model," forthcoming in *Quantitative Finance*.

[13] Crosby, J. (2008), "A multi-factor jump-diffusion model for commodities," *Quantitative Finance*, 8, pp. 181-200.

[14] De Jong, C. (2006), "The nature of power spikes: a regime-switch approach," *Studies in Nonlinear Dynamics & Econometrics*, 10, article 3.

[15] Deng, S. (1998) "Stochastic models of energy commodity prices and their applications: mean reversion with jumps and spikes," Working paper, Georgia Institute of Technology.

[16] Y. d'Halluin, P. Forsyth, and K. Vetzal (2005), "Robust numerical methods for contingent claims under jump diffusion processes," *IMA Journal on Numerical Analysis*, 20, pp. 87-112.

[17] Dollard, J. and C. Friedman (1979), *Product Integration with Applications to Differential Equations*, Encyclopedia of Mathematics and its Applications, 10, Addison-Wesley.

[18] Duffie, D. (2001), *Dynamic Asset Pricing Theory*, 3rd Ed., Princeton Univ. Press.

[19] Duffie, D., J. Pan, and K. Singleton (2000), "Transform Analysis and Asset Pricing for Affine Jump-Diffusion," Econometrica, 68, pp. 1343-1376.

[20] Eydeland, A. and H. Geman (1998), "Pricing Power Derivatives," *RISK*, September.

[21] Geman, H. (2005), *Commodities and Commodity Derivatives*, Wiley Finance.

[22] Geman, H. and A. Roncoroni (2006), "Understanding the fine structure of electricity prices," *Journal of Business*, 79, pp. 1225-1262.

[23] Gibson, R. and E. Schwartz (1990), "Stochastic convenience yield and the price of oil contingent products," *Journal of Finance*, 45, pp. 959-976.

[24] Glasserman, P. and S. Kou (2003), "The term structure of simple forward rates with jump risk," *Mathematical Finance*, 13, pp. 383-410.

[25] Hambley, B., S. Howison, and T. Kluge, "Modelling spikes and pricing swing options in electricity markets," Working Paper.

[26] Heston, S. (1993), "A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies*, 6, 327-344.

[27] Heath, D., R. Jarrow, and A. Morton (1992), "Bond Pricing and the term structure of interest rates: a new methodology for contingent claims valuation," *Econometrica*, 60, pp. 77-105.

[28] Jamshidian, F., "Commodity option valuation in the Gaussian futures term structure model," *Review of Futures Markets*, 10, pp. 61-86.

[29] Lee, R. (2004), "Option pricing by transform methods: extensions, unification and error control," *Journal of Computational Finance*, 7, pp. 51-86.

[30] Lipton, A. (2002), "The volatility smile problem," *RISK*, February, pp. 61-65.

[31] Litzenberger, R. and N. Rabinowitz (1995), "Backwardation in oil futures markets: theory and empirical evidence," *Journal of Finance*, 50, pp. 1517-1545.

[32] Miltersen, K. and E. Schwartz (1998), "Pricing of options on commodity futures with stochastic term structures of convenience yield and interest rates," *Journal of Financial and Quantitative Analysis*, 33, pp. 33-59.

[33] Miltersen, K. (2003), "Commodity price modelling that matches current observables," *Quantitative Finance*, 3, pp. 51-58.

[34] Pindyck, R. (2001), "The dynamics of commodity spot and futures markets: a primer," *Energy Journal*, 22, pp. 1-29.

[35] Ribeiro, D. and S. Hodges (2004), "A two-factor model for commodity prices and futures valuation," In: *EFMA 2004 Basel Meetings Paper*.

[36] Richter, M., and C. Sorensen (2002), "Stochastic volatility and seasonality in commodity futures and options: the case of soybeans," Working paper, Copenhagen Business School.

[37] Schwartz, E. (1997), "The stochastic behavior of commodity prices: implications for pricing and hedging," *Journal of Finance*, 52, pp. 923-973.

[38] Schwartz, E. and J. Smith (2000), "Short-term variations and long-term dynamics in commodity prices," *Management Science*, 46, pp. 893-911.

[39] Schwartz, E. and A. Trolle (2007) "Unspanned stochastic volatility and the pricing of commodity derivatives," Working Paper.