CONVENIENCE YIELD MODEL WITH PARTIAL OBSERVATIONS AND EXPONENTIAL UTILITY

RENÉ CARMONA AND MICHAEL LUDKOVSKI

Abstract. We consider the problem of hedging and pricing claims for delivery of crude oil or natural gas to a given location. We work with a three factor model for the asset spot, the convenience yield and the locational basis. The convenience yield is taken to be unobserved and must be filtered. We study the value function corresponding to utility pricing with exponential utility. Assuming the basis is independent from the spot, the partially observed stochastic control problem can be expressed as a closed-form expectation. We show how to numerically compute this expectation using a Kalman or particle filter. The basic model may be generalized to include nonlinear dynamics and further dependencies. We compute a set of numerical statics and we compare our results in the partially observed case to those of the full information case.

convenience yield, filtering, partial observations, stochastic control, utility pricing, HJB equation

1. Introduction

In this paper we analyze hedging of commodity contingent claims at a given geographical location. We work with a partially observed convenience yield model to capture the economic intuition regarding asset dynamics. In addition, we introduce a basis factor relating the local price to the market benchmark. Our approach is based on utility pricing with exponential utility. We recall the linearization described by Lasry and Lions [20] to simplify the resulting stochastic control problem. In the degenerate case when the basis is independent from the rest of the system, it then follows that the additional cost pertaining to partial observations is additive. In the more realistic case where the basis is a function of the current convenience yield, we obtain a reduced-form expression that can be numerically approximated using Monte-Carlo simulation together with a filtering algorithm. After giving a complete implementation, we discuss the effect of our model on the forward curve which is the fundamental object in energy markets.

Date: June 2004.

Our contribution to literature is two-fold. First, our analysis takes into account the unobservability of the convenience yield. While models with partially observed stochastic drift have been extensively studied in the context of predictable equity returns, to our knowledge this is the first application in energy valuation. Second, we present a new framework for pricing locational assets which combines convenience yields with direct modeling of the basis. This allows us to obtain a full numerical solution using filtering algorithms. Combined, this paper is a first step towards a more realistic model for pricing location-specific energy contracts.

Our work is inspired by the brief note of Lasry and Lions [20] who point out that the wealth-invariance property of exponential utility carries over to models with partial observations. However, their report does not mention any applications and does not consider the case where the payoff depends on the unobserved. The closely related problem of indifference pricing with exponential utility and unhedgeable risks is discussed by Musiela and Zariphopoulou [22] and Zariphopoulou [33]. In a similar vein, Henderson [16] looked at hedging non-traded securities using a closely correlated asset.

From an applications point of view, our work continues the sequence of convenience yield models begun by Gibson and Schwartz [13]. More recent generalizations include Schwartz [29] and Bjork and Landen [2]. However, all those papers assume a fully observed setting. In contrast, we insist that the convenience yield is unobserved because it is an abstract concept with no direct analogue in the marketplace. Many authors have applied filtering techniques to recover unobserved factors in financial models. We mention the series of papers by Runggaldier [28] and references therein, who describes the general approach.

Finally, several papers have treated the problem of portfolio optimization with partial observations. Most authors concentrated on the Gaussian case where explicit computations are possible. Using the convex duality approach, Lakner [19] found the optimal final wealth by exhibiting a primal-dual pair. His method relied on the fact that both the final wealth and the Kalman filtering equations lead to certain exponential-type martingales of similar structure. Nagai [23] extended the result to the case of non-zero correlation between the unobserved factor and the spot. Sekine [31] further extended to cover the situation of both observed and unobserved factors.

The rest of the paper is organized as follows. In Section 2 we describe the financial setting of our problem and present the basic convenience yield model of Gibson and Schwartz. Section 3 briefly reviews the complete information case and summarizes the results of Bjork and Landen [2]. In Section

4 we explain our pricing methodology based on indifference prices. Section 5 recalls the filtering results we need which are then specialized to the Gibson-Schwartz model. Section 6 contains our key results on indifference prices with exponential utility. In Section 7 we summarize and compare the numerical results. Finally, Section 8 concludes and outlines possible extensions to consider in the future.

2. Model Setup

2.1. **Financial Setting.** A sample real-life scenario that serves as inspiration for our analysis is the following. Consider a gas fired power plant PP in New Jersey and assume that the plant operator is interested in purchasing a contract for delivery of natural gas to its gate at a future time T. For simplicity, suppose that PP would like to buy a European claim paying out $\phi(S_T^{NJ})$ where S^{NJ} is the spot price of gas at the nearest interconnection node. Unfortunately, no market exists for S^{NJ}-contingent claims. Such contracts can only be bought over-the-counter from market makers, such as large utilities and pipeline companies.

The practical solution is to do pricing and hedging using a similar, *traded* asset. While gas markets have a multitude of trading nodes [1], for trading purposes there exists a clear benchmark, the Henry Hub of Sabine Inc. in Louisiana. Henry Hub contracts are traded on the New York Mercantile Exchange (NYMEX) and provide an extremely liquid and efficient market. Market participants refer to such contracts simply as NYMEX gas. Of course, using the NYMEX spot S^{HH} exposes the power plant to *basis* risk, i.e. the time-varying spread between prices at the New Jersey node and at Henry Hub. The basis is a function of a multitude of parameters, including the transportation cost through the pipelines, demand from other New Jersey customers and operational characteristics of the specific node [10].

Going back to the modeling problem, one approach now would be to write down a joint model for S^{HH} and S^{NJ}. Normally, one takes both processes to be diffusions with high correlation $c \sim 1 - \epsilon$. However, we prefer to model the basis directly as $S^{NJ} = spr(S^{HH}, B)$, where B is a random quantity corresponding to the basis and spr is some transformation function. For example, we will consider $S^{NJ} = S^{HH} + B$. This is preferable for a couple of reasons. Firstly, we can isolate the effect of the basis in our mathematical analysis. Secondly, we can impose simple conditions to guarantee, for instance, that $S^{NJ} - S^{HH}$ is bounded which is economically desirable, but

[1] Platt's Gas Daily© newsletter lists spot natural gas prices at over 60 locations, including most major citygates.

very difficult to achieve in the other approach. Thirdly, this allows for future extensions with more sophisticated modeling of the basis. In any case, we believe that the basis is a more meaningful financial object than the correlation between different location prices. We refer to [4, 9, 21] for more information on modeling the basis and other types of spreads in the energy markets.

2.2. **Convenience Yield Models.** The class of convenience yield models is the most popular choice for modeling the evolution of the spot price of energy assets such as crude oil or natural gas. The physical nature of commodities requires modification to the risk-free rate of return. Indeed, physical ownership of the commodity carries an associated flow of services. On the one hand, the owner enjoys the benefit of direct access which is important if the asset is to be consumed. On the other hand, the decision to postpone consumption implies storage expenses. As a result, the standard assumption is that the risk-neutral rate of return is not the short interest rate r_t, but rather $r_t - \delta_t$. Here δ_t is the *convenience yield* which measures the instantaneous net benefit of holding the physical asset [3]. The accepted approach [13, 17, 29] takes δ_t as a separate stochastic process.

2.3. **The Gibson-Schwartz Model.** Let $\{\Omega, \mathcal{F}, \mathbb{P}\}$ be a complete probability space. We assume the classical Gibson-Schwartz [13] two-factor model for the commodity spot S_t and spot convenience yield δ_t.

$$dS_t = (r_t - \delta_t)S_t\,dt + \sigma S_t\,dW_t^1, \tag{1a}$$

$$d\delta_t = \kappa(\theta - \delta_t)dt + \gamma\,dW_t^2, \tag{1b}$$

with W^1, W^2 1-dimensional Wiener processes satisfying $d\langle W^1, W^2\rangle_t = c\,dt$. In the sequel we will often use an alternative form using $S^0 \equiv \log S_t$:

$$\begin{aligned} dS_t^0 &= \big(r_t - \frac{1}{2}\sigma^2 - \delta_t\big)dt + \sigma dW_t \\ d\delta_t &= \kappa(\theta - \delta_t)dt + c\gamma\,dW_t + \sqrt{1-c^2}\gamma\,dW_t^{\perp}, \end{aligned} \tag{2}$$

with $W_t^{\perp}$ a standard Wiener process independent of W_t. The equations in (2) emphasize the linearity of our model. We let $\mathcal{F}_t = \sigma\big\{(S_s, \delta_s) : 0 \leqslant s \leqslant t\big\}$ denote the natural filtration generated by the entire process.

Strong positive correlation $c \sim 0.3 - 0.7$ between the spot and the convenience yield has been widely documented empirically. According to the theory of storage developed in the fifties, the endogeneous economic link is through inventory levels: when inventories are low, shortages are likely, causing high prices as well as valuable optionality of holding the physical asset. Mean-reversion in the convenience yield is desirable [12], based on economic intuition of long-term equilibrium in the energy market. Since on a long-term timescale energy assets are consumption goods we expect to achieve a supply-demand equilibrium which includes some sort of stationary premium for holding the physical asset. Unlike interest rates, depending on market conditions, convenience yields can be either positive or negative, and so the choice of an Ornstein-Uhlenbeck process for δ_t in (1b) makes sense.

From now on we shall also assume that

Assumption 1. *The short interest rate r_t is deterministic.*

The crucial implication of Assumption 1 is that futures and forward prices both equal the risk neutral expected future spot price. Thus, we disentangle the dynamics of the interest rates from the dynamics of the spot.

We denote by β_t the bank account following the ordinary differential equation

$$d\beta_t = r_t \beta_t dt.$$

Below we will sometimes assume that $r_t \equiv 0$, which is equivalent to working with discounted state variables. Making the above assumption often considerably simplifies the resulting equations and makes more transparent the effect of other parameters.

3. Complete Information Setting

One usually postulates that both S_t and δ_t are observed in the market modulo small noise disturbances (bid-ask spread, etc.) For the convenience yield the standard method is to use the *implied* δ_t via $\delta_t \approx r_t - \log(F(t,T_2)/F(t,T_1))$ where $F(t,T_1)$ and $F(t,T_2)$ are usually the two closest futures contracts [13]. Unfortunately, the implied convenience yield is highly unstable and inconsistent with the forward curve. Different forward contracts generate wildly different estimates and the empirical data rejects the notion of implied δ_t [5].

Nevertheless, assuming both factors are observed, pricing contingent claims is straightforward since we face a 2-dimensional complete market. Furthermore, great simplifications are possible thanks to the special form of (1). The Gibson-Schwartz model belongs to the so-called class of exponential affine models. Using $X_t = [S^0_t,\ \delta_t]'$ as the state variable in (2) we have

$$dX_t = \Big(\begin{bmatrix} \mu - \frac{1}{2}\sigma^2 \\ \kappa\theta \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 0 & -\kappa \end{bmatrix} X_t \Big) dt + \begin{bmatrix} \sigma & 0 \\ \gamma c & \gamma\sqrt{1-c^2} \end{bmatrix} d \begin{bmatrix} W_t \\ W_t^{\perp} \end{bmatrix}. \tag{3}$$

The above dynamics are linear in X_t and hence the present price of the forward with expiration date T can be expressed as

$$F(t,T) = \mathbb{E}^{\mathbb{Q}}[S_T|\mathcal{F}_t] := \exp\big(B(T-t)\cdot X_t + A(T-t)\big),$$

where $A(T-t)$ and $B(T-t) \equiv [B_S, B_\delta]$ solve the following equations

$$A_t(T-t) - \frac{1}{2}\sigma^2 B_S + \kappa\theta B_\delta + \frac{1}{2}B\begin{bmatrix}\sigma^2 & c\gamma\sigma\\ c\gamma\sigma & \gamma^2\end{bmatrix}B' = 0,$$
$$B'_\delta - \kappa B_\delta - B_S = 0,\ B'_S = 0,$$

with terminal conditions $A(0) = 0, B_S(0) = 1, B_\delta(0) = 0$. These can be solved explicitly [2] to obtain

$$\begin{aligned} F(t,T) &= S_t e^{\int_t^T r_s ds} e^{B(T-t)\delta_t + A(T-t)} \quad \text{where} \\ B(t) &= \frac{e^{-\kappa t}-1}{\kappa}, \\ A(t) &= \frac{\kappa\theta + c\sigma\gamma}{\kappa^2}(1 - e^{-\kappa t} - \kappa t) + \frac{\gamma^2}{4\kappa^3}(2\kappa t - 3 + 4e^{-\kappa t} - e^{-2\kappa t}). \end{aligned} \tag{4}$$

Keeping t fixed and letting $T \to \infty$, we see that $B(T-t) \to -1/\kappa$ and $A(T-t)$ is asymptotically linear in T with the sign depending on the value of $\gamma^2/2 - \theta\kappa^2 - c\sigma\gamma\,\kappa$. Thus, the forward price either exponentially explodes or decays with time to maturity (Figure 1). Both cases are unrealistic. In practice the longest-maturity (usually 7 or 10 years ahead) forward price is very stable [12, 30].

4. Utility-based Valuation

We now explain in more detail our pricing methodology. Our major motivation comes from the indifference pricing approach first introduced by Hodges and Neuberger [18] and Davis [8]. The introduction of a non-traded location factor B means that claims of the form $\phi(S_T, B)$ cannot be fully hedged. To avoid the problems associated with super-replication we instead rely directly on the utility function of the agent. More precisely, assuming a subjective utility function for the buyer (seller) of the asset, we hedge energy derivatives based on the wealth-adjusted utility equivalent forgone by the agent. From a modeling point of view this results in a partially observed stochastic control problem.

Besides being exposed to the terminal payoff ϕ, the agent performs portfolio optimization by dynamically rebalancing her asset holdings in the commodity spot and the riskless bank account. At time t, she invests π^i_t dollars into the i-th asset, so that using the self-financing constraint the wealth process w_t must satisfy

$$dw_t = \pi_t^0 \cdot \frac{d\beta_t}{\beta_t} + \pi_t^1 \cdot \frac{dS_t}{S_t}.$$

Consequently, setting $\pi \equiv \pi^1$,

$$dw_t^{x,\pi} = r_t w_t^{x,\pi} dt - \delta_t \pi_t \, dt + \sigma \pi_t \, dW_t, \qquad \text{with} \quad w_0^{x,\pi} = x. \tag{5}$$

Let $\mathcal{G}_t = \sigma\{S_s : 0 \leqslant s \leqslant t\}$. We denote by $\mathcal{A}_t^T$ the set of admissible portfolio strategies $\{\pi_s\}_{t \le s \le T}$ which are square integrable $\mathbb{E}\{\int_t^T \pi_s^2 ds\} < \infty$ $\mathcal{G}_t$-progressively measurable processes.

We work with the exponential utility $U(x) = -\exp(-qx), q > 0$. Since U is defined on the whole real line we do not need to impose any state constraints, such as required positive wealth. Of course, we still want the wealth process to be bounded from below in order to exclude doubling and suicide strategies.

We now define the value function V which is the main object of interest in our analysis. Given a European option with payoff $\phi(S_T, B)$ let

(6)

$$V^\phi(t, s, w, \xi; T) = \sup_{\pi \in \mathcal{A}_t^T} \mathbb{E}\Big\{ \int_{-\infty}^{\infty} U(w_T^{x,\pi} + \phi(S_T, B)) d\mathbb{P}_B \Big| S_t = s, w_t = w, \delta_t \sim \xi \Big\}.$$

Above, the initial value of δ_t is unknown, but we are given some initial distribution ξ. Assumptions regarding the basis factor B will be stated later on. Thus, the value function is the maximum expected utility to be derived from portfolio optimization and the claim ϕ given the specified initial conditions. The optimal hedging strategy is then the π^* achieving the supremum in (6).

Remark 1. Our setting is closely related to the concept of indifference price. The *buyer's* indifference price for claim ϕ, $P = P(\phi(S_T, B), t; s, \xi)$ at time t is defined implicitly via

$$V^\phi(t, s, w - P, \xi) := V^0(t, s, w, \xi). \tag{7}$$

The indifference price generally depends both on the level of wealth and the current spot price, since only the combined process is Markovian. Intuitively, P represents the decrease in initial wealth that balances the increase in terminal utility from buying the derivative $\phi(S_T, B)$. It can be shown that this pricing mechanism assigns a no-arbitrage consistent 'fair value' to the contingent claim [8]. Similarly, we also define the *seller's* price by $V^{-\phi}(t, s, w + P_{sell}, \xi) := V^0(t, s, w, \xi)$. In this paper we concentrate on the buyer's point of view in line with the financial application outlined in Section 2.1.

Remark 2. By itself, stochasticity of the drift for the spot process is irrelevant for pricing S_t-contingent claims. Intuitively, the market defined by the Gibson-Schwartz model is still complete even when δ_t is unobserved. Indeed, through a measure change S_t can be made into a local martingale under some $\widetilde{\mathbb{P}}$. It can be easily checked that there exists a $\widetilde{\mathbb{P}}$-Wiener process $\widetilde{W}_t$ whose natural filtration is equal to the filtration generated by S_t in (1). After invoking the standard martingale representation theorem we conclude that any claim strictly depending only on S_T can be perfectly replicated. In particular, by well-known results [27], the only price of such a claim consistent with no-arbitrage must equal its replication price under $\widetilde{\mathbb{P}}$.

5. Filtering the Convenience Yield

To be able to consider $\mathcal{G}_t$-adapted hedging strategies in (6), we need to replace δ_t by its conditional expectation given $\mathcal{G}_t$. This is known as the filtering problem. We briefly summarize the main results in some generality, following the exposition in [1].

We assume a general correlated model for the n-dimensional observation (traded asset) process Y_t and the d-dimensional unobserved factor X_t.

$$dY_t = h(t, X_t, Y_t)\,dt + \sigma(t, Y_t)\,dW_t, \tag{8a}$$

$$dX_t = g(t, X_t, Y_t)\,dt + \alpha(t, X_t, Y_t)\,dW_t + \gamma(t, X_t, Y_t)\,dW_t^{\perp}, \tag{8b}$$

$$X_0 \sim \xi\ , Y_0 = 0, \quad W^{\perp}\ \text{and}\, W\ \text{independent.} \tag{8c}$$

The Gibson-Schwartz model is a simple version of above, with $Y_t \equiv \log S_t$ the observed log-price of the spot, and $X_t \equiv \delta_t$ the convenience yield. There, α and γ are deterministic and h and g linear and independent of the observations. Note that the diffusion coefficient of the observation process must not depend on X_t. Thus, this setup is inherently different from stochastic volatility models [25]. On the other hand, the unobserved factor drift and volatility may depend on the observed (i.e. the price), which seems in general an important and useful characteristic, even though most known models do not take advantage of it.

We continue with the general case and impose

Assumption 2.
- *$h(t,x,y) : \mathbb{R}^{n+d+1} \mapsto \mathbb{R}^n$ is Lipschitz and of linear growth, $|h(t,x,y)| \leqslant K(1+|x|+|y|)$.*

- *$\sigma(t,y)$ is uniformly continuous and has bounded $\mathcal{C}^3(\mathbb{R}^d)$-norm. Furthermore, σ is uniformly elliptic, that is $\sigma\sigma'(t,y) \geqslant \lambda I$ for all y and t, for some constant $\lambda > 0$.*
- *$\alpha(t,x,y)$ and $\gamma(t,x,y)$ are uniformly continuous, and α is uniformly elliptic.*
- *$g(t,x,y)$ is $\mathcal{C}^2$-bounded and uniformly continuous in x and y.*

These general results apply to model (1) even though the latter does not have bounded drift. A standard localization argument can be used to accommodate the linear growth of the drift.

For notational clarity we suppress from now on all the dependencies on t. Let $\mathcal{Y}_t = \sigma\{Y_s : 0 \leqslant s \leqslant t\}$ be the observable σ-algebra. We use the innovation process to write the dynamics of X_t as a function of the observable dynamics plus independent white noise. Let $D_t = D(t, Y_t) = (\sigma\sigma^t)(t, Y_t)$, which is symmetric and invertible by assumption ($D_t = \sigma^2$ in the Gibson-Schwartz model), and define ζ_t by

$$d\zeta_t = -\zeta_t h'(X_t, Y_t) D_t^{-1/2} dW_t, \quad \zeta_0 = 1, \tag{9}$$

where h' denotes the transpose of the column vector h. By Assumption 2 it follows that ζ_t is an exponential martingale with $\mathbb{E}[\zeta_t] = 1, \forall t \leqslant T$ [1, Lemma 4.1.1], and therefore we can apply Girsanov theorem to define a new probability measure $\widetilde{\mathbb{P}}$ by

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}\Big|_{\mathcal{F}_t} := \zeta_t.$$

Then under $\widetilde{\mathbb{P}}$ there exists a standard Wiener process $\widetilde{W}$ such that,

$$dY_t = \sigma(Y_t)\, d\widetilde{W}_t \qquad \text{and} \tag{10a}$$

$$\begin{aligned} dX_t = \big(g(X_t, Y_t) - \alpha(X_t, Y_t)' \cdot h(X_t, Y_t)' D_t^{-1/2}\big) dt \\ + \alpha(X_t, Y_t)' D_t^{-1/2} dY_t + \gamma(X_t, Y_t)\, dW_t^{\perp}. \end{aligned} \tag{10b}$$

Letting $d\widetilde{Y}_t = D_t^{-1/2} dY_t$, $\widetilde{Y}$ is another wiener process under $\widetilde{\mathbb{P}}$. The crucial observation is that $\widetilde{Y}$ and $W^{\perp}$ are two independent standard $\widetilde{\mathbb{P}}$-Wiener processes. At the same time, since D_t is invertible $\sigma\{\widetilde{Y}_s : 0 \leqslant s \leqslant t\} \equiv \mathcal{Y}_s$. We can write the inverse $\eta_t = \frac{1}{\zeta_t}$ as

$$\begin{aligned} \eta_t &= \exp\Big(\int_0^t h' D_s^{-1/2} dW_s + \frac{1}{2}\int_0^t h' D_s^{-1} h\, ds\Big) \\ &= \exp\Big(\int_0^t h' D_s^{-1}\, dY_s - \frac{1}{2}\int_0^t h' D_s^{-1} h\, ds\Big). \end{aligned}$$

Let $p_t(f) := \widetilde{\mathbb{E}}[f(X_t)\eta_t|\mathcal{Y}_t]$. To compute $\Pi_t(f) := \mathbb{E}[f(X_t)|\mathcal{Y}_t]$ we apply Bayes rule to write

$$\Pi_t(f) = \frac{\widetilde{\mathbb{E}}[f(X_t)\eta_t|\mathcal{Y}_t]}{\widetilde{\mathbb{E}}[\eta_t|\mathcal{Y}_t]} = \frac{p_t(f)}{p_t(1)}. \tag{11}$$

The above is called the Kallianpur-Striebel formula and it demonstrates that it is sufficient just to be able to compute the unnormalized version $p_t(f)$. Suppose that $p_t(\cdot)$ possesses a smooth density $\rho_t(x)dx$, i.e.

$$\forall f \in \mathcal{C}_0^\infty(\mathbb{R}^d), \quad \widetilde{\mathbb{E}}\big[f(X_t)\eta_t \big| \mathcal{Y}_t\big] = p_t(f) = \int_{\mathbb{R}} \rho_t(x) f(x)dx. \tag{12}$$

Then by applying Itô's lemma to $d(\eta_t f(X_t))$ using (10b), taking expectations and integrating by parts we obtain that the L^2-valued process $\rho_t(x)$ must satisfy the adjoint Zakai equation [1]

$$\begin{aligned} d\rho_t(x) =& \Big(\frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\big\{[\gamma\gamma^t + \alpha\alpha^t]_{i,j}\,\rho_t\big\} - \sum_i \frac{\partial}{\partial x_i}(g^i\rho_t)\Big)\,dt \\ & + \Big(h - \sum_i \frac{\partial}{\partial x_i}(\alpha^i\rho_t)\Big)d\widetilde{Y}_t \\ =:& \,\mathcal{L}_X^*(\rho_t)(x)\,dt + \mathcal{S}^*(\rho_t)(x)\,d\widetilde{Y}_t. \end{aligned}$$

Above, the *-s denote formal adjoints, $\mathcal{L}_X$ is the elliptic operator corresponding to the diffusion X_t, and $\mathcal{S}$ is a first-order differential operator to which we shall return later.

5.1. **Technical Results on Spartial differential equations.** The Zakai equation is a stochastic partial differential equation and we must check that it is well-defined. Even though at first glance we work in L^2, it turns out that for technical reasons the weighted Sobolev spaces H^k_β are more convenient [14]. For $\beta \geqslant 0$, define the weighted Sobolev norm

$$\|f\|_{k,\beta} = \sum_{|\alpha|\leqslant k}\Big(\int_{\mathbb{R}^d}\big(\partial^\alpha[(1+|\xi|^2)^{\beta/2}f(\xi)]\big)^2 d\xi\Big)^{1/2}.$$

The Hilbert space $H^k_\beta(\mathbb{R}^d)$ is defined as the completion of $\mathcal{C}_0^\infty(\mathbb{R}^d)$ with respect to the above norm. It can be thought of as the set of all measurable functions $f : \mathbb{R}^d \mapsto \mathbb{R}$ such that $(1+|\xi|^2)^{\beta/2}f(\xi)$ has square-integrable derivatives up to order k. In particular, $H^0_0 = L^2$. We also define H^{-k} as the completion of L^2 under the norm $\|f\|_{-k} = \langle\,(I-\triangle)^k f, f\rangle$. As the notation suggests, H^{-k}_β is the dual of H^k_β. It is also a Hilbert space under the inner product induced by the corresponding norm.

Like SDEs, stochastic partial differential equations must always be interpreted in their integrated form as almost sure equality between the respective random variables. The following is a variant of existence and uniqueness result for Spartial differential equation's in Sobolev spaces.

Lemma 1. [14] *Let $\mathcal{L}$ be the second order linear differential operator on H^2_β corresponding to the generator of X_t. We write $\mathcal{L}$ in divergence form as*

$$\mathcal{L} = -\sum_{i,j} \partial_i(a_{i,j}\partial_j \cdot) + \sum_i \partial_i(g - \partial_j a_{i,j} \cdot), \quad \textit{where} \quad a \equiv (\sigma\sigma^t + \alpha\alpha^t).$$

Let $\mathcal{S}$ be the first order differential operator with domain H^1_β defined in (13). *Suppose the coefficients satisfy Assumption 2, $c < 1$ (strict ellipticity) and $\mathbb{E}\|x\|^2_{1,\beta} < \infty$. Then there exists a unique $Y_t \in L^2([0,T];\Omega,\mathcal{Y}_t,\widetilde{\mathbb{P}})$ satisfying (strong solution of Spartial differential equation)*

$$Y(t) + \int_0^t \mathcal{L}Y_s\, ds = Y_0 + \int_0^t \mathcal{S}Y_s\, d\widetilde{W}_s.$$

Also, we have the energy equality

$$\mathbb{E}\|Y_t\|^2_\beta = \mathbb{E}\|Y_0\|^2 - 2\mathbb{E}\int_0^t \langle \mathcal{L}Y, Y_s\rangle\, ds + \mathbb{E}\int_0^t \|\mathcal{S}Y_s\|^2_\beta\, ds,$$

and for a constant C,

$$\mathbb{E}\|Y_s\|^2_{1,\beta} \leqslant \mathbb{E}\|x\|^2_{1,\beta}(1 + Cs) \quad \textit{with} \quad \mathbb{E}\int_0^T \|Y_s\|^2_{2,\beta}\, ds \leqslant C\mathbb{E}\|x\|^2_{1,\beta}.$$

The result is proven via fixed-point theorems by showing that the corresponding Picard iteration is a contraction in a convenient space. The energy equality is key for establishing estimates of the solution.

5.2. **Filtering Gibson Schwartz.** We now specialize to the linearized version of Gibson-Schwartz (2). The differential operators are

$$\mathcal{L}^*_\delta(f)(x) = \big(\kappa(\theta - x)\, f'(x) - \kappa f(x)\big) + \frac{1}{2}\gamma^2 f''(x), \qquad \text{and}$$
$$\mathcal{S}^*(f)(x) = r - \frac{1}{2}\sigma^2 - x - c\gamma f'(x).$$

We have $d\zeta_t = -\zeta_t\kappa\frac{(r-\frac{1}{2}\sigma^2-\delta_t)}{\sigma}\, dW_t$ and

$$\eta_t = \exp\Big(\int_0^t \frac{r - \frac{1}{2}\sigma^2 - \delta_s}{\sigma^2}\, dS^0_s - \frac{1}{2}\int_0^t \frac{(r - \frac{1}{2}\sigma^2 - \delta_s)^2}{\sigma^2} ds\Big). \tag{13}$$

The un-normalized $\rho_t(x)$ satisfies

$$d\rho_t(\delta) = \Big[\frac{1}{2}\gamma^2\rho_t''(\delta) - \frac{\partial}{\partial\delta}\big(\kappa(\theta-\delta)\rho_t(\delta)\big)\Big]dt + \big(r - \frac{1}{2}\sigma^2 - \delta - c\gamma\rho_t'(\delta)\big)dS_t^0$$

Alternatively, for $f(t,x) \in \mathcal{C}_b^{1,2}(\mathbb{R})$,

$$d\langle\rho_t, f(t,\cdot)\rangle = \langle\rho_t, \frac{\partial f}{\partial t} - \mathcal{L}_\delta f(t,x)\rangle\, dt + \Big\langle\rho_t, (r - \frac{1}{2}\sigma^2 - \delta)f(t,x) + c\sigma^2\gamma\, \partial_x f(t,x)\Big\rangle \frac{1}{\sigma^2}\, dS_t^0.$$

5.3. **Kalman Filtering.** The Gibson-Schwartz model is linear and hence by classical arguments, if the initial distribution δ_0 is Gaussian, we have that $\delta_t|\mathcal{G}_t \sim \mathcal{N}(\hat{\delta}_t, P_t)$ is conditionally Gaussian for all times. The evolution of the conditional mean $\hat{\delta}_t$ and the conditional variance P_t is obtained from the Kalman filter [15].

We can re-write (1) as

$$d\delta_t = \kappa(\theta-\delta_t)dt + c\gamma(\frac{dS_t^0}{\sigma} - \frac{r - \frac{1}{2}\sigma^2 - \delta_t}{\sigma}dt) + \gamma\sqrt{1-c^2}dW_t^{\perp}, \tag{14}$$

so that formally (rigorous justification relies on the innovation process [1])

$$d\hat{\delta}_t = \kappa(\theta - \hat{\delta}_t)dt + \frac{(c\sigma\gamma - P_t)}{\sigma^2}\Big[d(S_t^0) - (r - \frac{1}{2}\sigma^2 - \hat{\delta}_t)dt\Big]. \tag{15}$$

Above P_t is the conditional variance $P_t := \mathbb{E}[(\delta_t - \hat{\delta}_t)^2|\, \mathcal{G}_t]$. To derive the equation for P_t, apply Itô's formula again to obtain a *deterministic* Riccati equation

$$dP_t = \Big[\gamma^2 - 2\kappa P_t - \frac{(c\sigma\gamma - P_t)^2}{\sigma^2}\Big]dt.$$

Summarizing,

Proposition 1. $\forall f \in \mathcal{C}^\infty(\mathbb{R})$,

$$\mathbb{E}[f(\delta_t)|\, \mathcal{G}_t] = \int_{\mathbb{R}} f(\hat{\delta}_t + P_t^{1/2}\xi)\frac{e^{-\frac{1}{2}|\xi|^2}}{\sqrt{2\pi}}\, d\xi. \tag{16}$$

with $\hat{\delta}_t$ and P_t given above, and $\langle\rho_t, f\rangle = \hat{\eta}_t \cdot \mathbb{E}[f(\delta_t)|\, \mathcal{G}_t]$ where

$$\hat{\eta}_t = \exp\Big(\int_0^t \frac{r - \frac{1}{2}\sigma^2 - \hat{\delta}_s}{\sigma^2}dS_s^0 - \frac{1}{2}\int_0^t \frac{(r - \frac{1}{2}\sigma^2 - \hat{\delta}_s)^2}{\sigma^2}ds\Big). \tag{17}$$

The Riccati equation for P_t has been well studied. It can be shown that P_t is monotonic and converges to a limiting value. The speed of convergence to this limit is on the scale of $\frac{1}{\sqrt{\kappa}}$ which is about 6-18 months for estimated parameter values. The Radon-Nikodym density $\hat{\eta}_t$ is an exponential martingale and thus we have $\mathbb{E}[\int \rho_T(x)dx] = \mathbb{E}[\eta_T] = 1$.

5.4. **Expected Value of S_T-Contingent Claims.** Due to explicit expressions in (17), we are able to compute in a reduced form (up to solution of ordinary differential equation) any expectation of the form $\mathbb{E}[\phi(S_T)^\alpha]$. The latter expression can be thought of as α-th power of the price of claim S_T with respect to the *original* $\mathbb{P}$ measure (under which the drift of the spot is $r_t - \delta_t$). As mentioned in Section 4.2, such claims can be perfectly hedged even though δ_t is unobserved. Let h_t be the replicating strategy for ϕ in dollar terms, $\phi(S_T) = \int_0^T h_t \frac{dS_t}{S_t}$, so that

$$\phi(S_T)^\alpha = S_0^\alpha \exp\Big(\int_0^T \Phi_t\, dt + \alpha \int_0^T h_t \sigma\, d\widetilde{W}_t + \frac{\alpha^2}{2} \int_0^T h_t^2 \sigma^2\, dt\Big),$$

where $\Phi_t = -\frac{\alpha}{2}(1+\alpha)h_t^2\sigma^2 + \alpha r + \alpha h_t \hat{\delta}_t$. Then using (16), we just need to compute the following expectation

$$\widetilde{\mathbb{E}}\Big[\exp(\int_0^T \Phi_t\, dt + \int_0^T (r - \hat{\delta}_t - \frac{1}{2}\sigma^2 + \alpha\sigma^2 h_t)\frac{1}{\sigma^2} dS_t^0 \\ - \frac{1}{2}\int_0^T (r - \hat{\delta}_t - \frac{1}{2}\sigma^2 + \alpha\sigma^2 h_t)^2 \frac{1}{\sigma^2}\, dt)\Big].$$

Call the expression inside the expectation L_t. We shall guess that $\widetilde{\mathbb{E}}[L_t]$ is an exponential of a linear function of the current best estimate $\hat{\delta}_0$. Accordingly, let us set $\chi_t = 2\alpha g_t \hat{\delta}_t + \alpha k_t$ for some time-dependent deterministic g_t and k_t. Then we compute

$$d\, e^{\chi_t} = e^{\chi_t}\big\{(2\alpha\hat{\delta}_t\dot{g}_t + \alpha\dot{k}_t)dt + 2\alpha g_t\, d\hat{\delta}_t + \frac{1}{2}(2\alpha g_t U_t)^2 \frac{1}{\sigma^2} dt\big\}.$$

Here $U_t := c\sigma\gamma - P_t$. Using (13) it follows that

$$\begin{aligned} d(L_t e^{\chi_t}) =& L_t e^{\chi_t}\Big(\Phi_t\, dt + (r - \hat{\delta}_t - \frac{1}{2}\sigma^2 + \alpha\sigma^2 h_t)\frac{1}{\sigma^2} dS_t^0 \\ &+ 2\alpha g_t\big\{\frac{U_t}{\sigma^2} dS_t^0 + (\kappa(\theta - \hat{\delta}_t) - (r - \hat{\delta}_t - \frac{1}{2}\sigma^2)\frac{U_t}{\sigma^2})\, dt\big\} \\ &+ (2\alpha\hat{\delta}_t\dot{g}_t + \alpha\dot{k}_t)\, dt + \frac{1}{2}(2\alpha g_t U_t)^2 \frac{1}{\sigma^2}\, dt \\ &+ 2\alpha g_t (r - \hat{\delta}_t - \frac{1}{2}\sigma^2 + \alpha\sigma^2 h_t)\frac{U_t}{\sigma^2}\, dt\Big). \end{aligned} \tag{18}$$

Next we equate powers of $\hat{\delta}_t$ and pick g_t and k_t such that all the drift terms disappear. For boundary conditions we take $g(T) = k(T) = 0$. In this case $L_t e^{\chi_t}$ is a $\widetilde{\mathbb{P}}$-martingale (since P_t is bounded, so is g_t and k_t), so that $\widetilde{\mathbb{E}}[L_T] = \widetilde{\mathbb{E}}[L_T e^{\chi_T}] = e^{\chi_0} = \mathrm{e}^{2\alpha g_0 \hat{\delta}_0 + \alpha k_0}$. The ordinary differential equations satisfied by g_t and k_t are

$$\begin{aligned} dg_t &= -\frac{h_t}{2} + \kappa g_t \\ dk_t &= r + 2\alpha g_t U_t h_t - \frac{1+\alpha}{2}\sigma^2 h_t^2 - 2\kappa\theta g_t + 2\alpha g_t^2 \frac{U_t^2}{\sigma^2}. \end{aligned} \tag{19}$$

Solving,

$$\begin{aligned} g_0 &= \frac{h_0}{2\kappa}(1 - e^{-\kappa T}), \\ k_0 &= \int_0^T -2g_t(\alpha U_t h_t + \kappa\theta) + \frac{1+\alpha}{2}\sigma^2 h_t^2 - r - 2\alpha g_t^2 \frac{U_t^2}{\sigma^2}\, dt. \end{aligned} \tag{20}$$

Figure 1 shows the expected value of the spot $\mathbb{E}[S_T]$ compared to the full information setting.

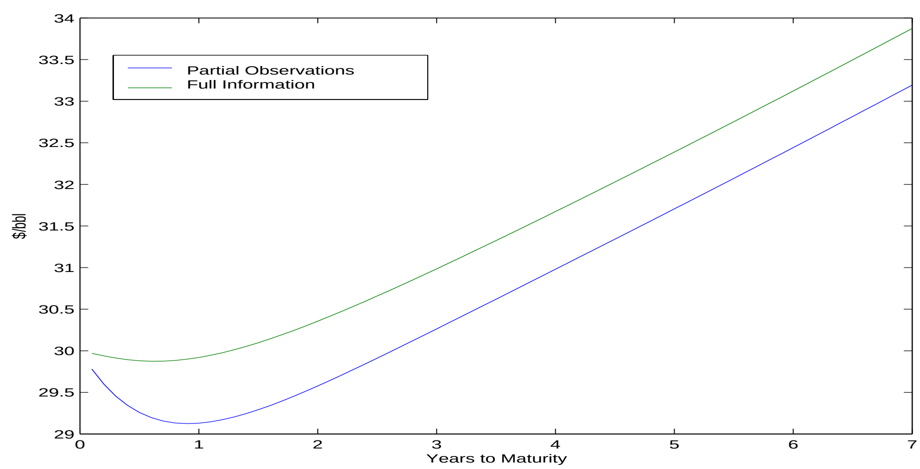

FIGURE 1. Expected spot price with respect to risk-neutral dynamics. Parameter values are from Sch97 in Table 1. Today's price is \$30.

6. Optimal Wealth

We finally turn our attention towards the stochastic control problem with partial information. The general approach we will follow for solving (6) is to setup the dynamic programming (DP) equation. This leads to a second order partial differential equation in an appropriate space. This partial differential equation belongs to the general class of Hamilton-Jacobi-Bellman equations which have been extensively analyzed from a functional-analytic viewpoint. Furthermore, as a general rule it can be shown that the value function is the unique viscosity solution of the HJB equation. In practice, in most cases one can construct a smooth solution and invoke a verification theorem [11] rather than checking for viscosity solution properties.

The HJB equation also provides a method for finding the optimal hedging strategy (i.e. the portfolio weights). According to the maximum principle [32, Ch. 3], the optimal portfolio weights can be obtained by formally computing the supremum in the Hamiltonian of the HJB equation.

Assumption 3. *The basis B is independent of $\mathcal{F}_T$.*

Clearly, Assumption 3 is very strong. Later on, we shall slightly relax and also allow $B = B(\delta_T)$. Since we only look at European claims, without any loss of generality we further assume stationarity that is $B_T \sim B$ for any time T.

In line with market intuition, we restrict our attention to $\mathcal{G}_t$-predictable trading strategies. We rely on the Dynamic Programming principle which states that for any stopping time $\tau : t \leqslant \tau \leqslant T$,

$$V^{\phi}(t, s, x, \xi; T) = \sup_{\pi \in \mathcal{A}_t^{\tau}} \mathbb{E}\big[V^{\phi}(\tau, S_\tau, w_\tau^{x,\pi}, \rho_\tau)\big|\mathcal{G}_t\big]. \tag{21}$$

Combining (12) and (6) then gives

$$\begin{aligned}\mathbb{E}\big[U(w_T^{\pi} - \phi(S_T, B))\big] &= \widetilde{\mathbb{E}}\big[U(w_T^{\pi} - \phi(S_T, B))\eta_T\big] \\ &= \widetilde{\mathbb{E}}\big[\int U(w_T^{\pi} - \phi(S_T, b))d\mathbb{P}_B \int_{\mathbb{R}} \rho_T(x)dx\big] \qquad (22)\end{aligned}$$

since the terminal payoff is independent of δ_T. As we can see the only place the un-normalized conditional density appears is as a scaling factor. This is a degenerate case of the separation principle [1, Ch. 7], where we have been able to separate the problem of estimating the unobserved state from the utility maximization problem. Note that the control only affects the wealth process whose dynamics under $\widetilde{\mathbb{P}}$ are unaffected by ρ_t.

In equation (22) we have succeeded in reducing the partial observation problem to an equivalent problem with full observation, but at the expense of introducing the measure-valued process ρ_t. The full state is now $(S_t, w_t, \rho_t) \in \mathbb{R}_+ \times \mathbb{R} \times H^0_\beta(\mathbb{R}^d)$. We are faced with an infinite-dimensional stochastic control problem which requires delicate handling. However, for smooth parameters the DP intuition still holds [32], and we can use the technique of Hamilton-Jacobi-Bellman equations.

To be able to state results regarding the HJB equation we must require the initial distribution to decrease sufficiently fast, $\rho_0 \sim \xi \in H^0_\beta$. In general, there are also restrictions on the utility function U which must be of polynomial growth at infinity, but this is trivially satisfied by exponential utility.

Let $\mathcal{L}_S$ and $\mathcal{L}_\delta$ be the elliptic operators associated with the state process. For the Gibson-Schwartz model these are given by

$$\mathcal{L}_s = -\delta s\, \partial_s + \frac{1}{2} s^2 \sigma^2 \partial_{ss} \quad \text{and}$$
$$\mathcal{L}_\delta = \kappa(\theta - \delta)\, \partial_\delta + \frac{1}{2}\gamma^2 \partial_{\delta\delta}.$$

By analogy with the finite dimensional case we expect that $V(t,s,w,\xi)$ satisfies the backward parabolic partial differential equation

$$V_t + \langle\, \mathcal{L}^*_\delta \rho_t, V_\rho\rangle + \frac{1}{2}\langle\, V_{\rho\rho}\, S^*\rho, S^*\rho\rangle + \frac{1}{2}\sigma^2 s^2 V_{ss} + \sup_{\pi \in \mathcal{A}^T_t} \Big\{ \sigma^2 \pi s V_{sw} + \frac{1}{2}\sigma^2\pi^2 V_{ww} + \big\langle S^*\rho\,, \sigma s V_{s\rho} + \sigma \pi V_{w\rho} \big\rangle \Big\} = 0, \tag{23}$$

with terminal condition $V(T,s,w,\xi) = \langle \int U(w - \phi(s,b)) d\mathbb{P}_B,\, \xi\rangle$.

Proposition 2. [14, Theorem 5.4] *Let $\overline{\mathcal{A}^T_t}$ be the set of admissible relaxed controls, that is*

$$\overline{\mathcal{A}^T_t} = \{(\Omega, \mathcal{F}, \mathbb{P}, W, \pi),\ \pi \text{ is } \ \ \mathcal{F}^W_t - \text{adapted}\}.$$

Then the value function $\overline{V} \in \mathcal{C}((0,T) \times \mathbb{R}^2 \times H^0_\beta)$ minimizing (22) *over $\overline{\mathcal{A}^T_t}$ is the unique viscosity solution of* (23)*.*

Further growth and continuity estimates on the value function can be made using standard partial differential equation techniques.

Note that in Proposition 2 the Wiener process W_t is not given a priori but together with the set of admissible portfolios, a notion similar to weak solutions of SDEs.

6.1. **Linearization with Exponential Utility.** Assume $r_t \equiv 0$. Lasry and Lions [20] show that the problem (22) inherits the separability property from the complete information setting. Specifically, guess that $V(t,s,w,\rho) = -\exp(-q(w + \psi(t,s,\rho)))$. Formally substituting into (23) we obtain

$$\psi_t + \frac{1}{2}\sigma^2 s^2 [q\psi_s^2 + \psi_{ss}] + \langle\, \mathcal{L}^*_\delta \rho, \psi_\rho\rangle + \frac{1}{2}\Big\langle\, S^*\rho, S^*\rho\big(q(\psi_\rho)^2 + \psi_{\rho\rho}\big)\Big\rangle + \langle\, S^*\rho, q\sigma s \psi_\rho\, \psi_s\rangle - \frac{q}{2}\Big(\sigma s \psi_s + \langle\, S^*\rho, \psi_\rho\rangle\Big)^2 = 0.$$

This equation in fact linearizes to:

(24)
$$\psi_t + \frac{1}{2}\sigma^2 s^2 \psi_{ss} + \langle \mathcal{L}_\delta^* \rho, D_\rho \psi \rangle + \left\langle S^* \rho, \sigma s D_\rho \psi_s \right\rangle + \frac{1}{2}\left\langle S^* \rho, D_{\rho\rho}\psi \; S^* \rho \right\rangle = 0.$$

Here D_ρ is the Fréchet derivative operator on H^0_β. But the above is nothing but the parabolic Kolmogorov Spartial differential equation [26] for the joint diffusion (S_t, ρ_t) with terminal condition $\psi(T, s, \rho) = \frac{1}{q} \log \int e^{-q\phi(s,b)} d\mathbb{P}_B$. Writing in full,

$$-e^{-q(w - \psi(T,s,\rho))} = V(T, s, w, \rho) = -\int e^{-q(w+\phi(s,b))} d\mathbb{P}_B \int_{\mathbb{R}} \rho(x) dx$$

$$\iff \quad -q(w - \psi(T, s, \rho)) = -qw + \log \int e^{-q\phi(s,b)} d\mathbb{P}_B + \log \int_{\mathbb{R}} \rho(x) dx$$

$$\psi(T, s, \rho) = \frac{1}{q} \log \int e^{-q\phi(s,b)} d\mathbb{P}_B + \frac{1}{q} \log \int_{\mathbb{R}} \rho(x) dx.$$

and the second term is zero for any initial density ρ. Therefore,

(25)
$$\psi(t, s, \xi) = \widetilde{\mathbb{E}}\Big[\frac{1}{q} \log \int e^{-q\phi(S_T,b)} d\mathbb{P}_B + \frac{1}{q} \log \int_{\mathbb{R}} \rho_T(x) dx \Big| S_t = s, \rho_t = \xi\Big].$$

The total value separates into the usual "certainty equivalent" price of derivative ϕ plus another cost due to partial observations. We can rewrite the second term as $\log \frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}}$ after which it can be easily seen that its expectation is negative. This also demonstrates that it is square integrable and hence the expectation is well-defined. As expected, the agent is getting a smaller utility from buying ϕ because he cannot observe δ_t.

Because the additional cost imposed by the uncertainty in the convenience yield is independent of the given payoff, the two terms will always cancel each other in the formula (7) for the indifference price. It follows that the indifference price P^ϕ is trivial,

$$P^\phi = \widetilde{\mathbb{E}}[\frac{1}{q} \log \int e^{-q\phi(S_T,b)} d\mathbb{P}_B],$$

which is the same as what one would obtain in a Black-Scholes world given a "totally unhedgeable" factor B [22]. As an example, suppose $\phi(S_T, B) = S_T + B$. This is a forward contract where the basis is assumed to be additive. Then $P^{Fwd} = S_t + const$ and we pay a fixed cost to cover the unhedgeable risk. Thus, up to the time-dependency in B, the forward curve is flat. This result is independent of the postulated model for the spot and the convenience yield, as long as the linearization in (24) occurs. The fact that exponential utility leads to trivial indifference prices for models with stochastic drift seems to be known, but we have been unable to find a clear reference in the existing literature.

Remark 3. The HJB equation linearizes only in the 1-dimensional case, when the entire system is driven by a univariate Wiener process. In particular, this excludes addition of any other factors: stocks, non-traded assets or unobservables. The optimal π is

$$\pi^* = -\frac{sV_{sw} + \langle S^* \rho, V_{w\rho} \rangle}{\sigma V_{ww}} = \psi_s + \langle \frac{1}{\sigma} S^* \rho, D_\rho \psi \rangle. \tag{26}$$

It would be useful to obtain a more computationally amenable expression for the second term which measures the sensitivity with respect to ρ_t.

6.2. **Basis depending on the convenience yield.** The convenience yield is large when there is tight supply in the market. Often tight supply occurs due to limited bandwidth of the pipelines so that the upstream market is unable to quickly respond to increased demand. For instance, unusually cold weather in the Northeast leads to highly increased electricity consumption in the region. To produce extra electricity, peaking power plants that run on natural gas are brought on line. Thus there is also increased demand for gas. However, the Northeast has very limited gas storage facilities, and any gas must be brought through pipelines from the South and the Midwest. Clearly, limited pipeline capacity would then induce high basis between the spot in New Jersey and at Henry Hub.

The above exercise demonstrates that it is reasonable to assume that the basis B might depend on δ_T. A very simple model would be to take $B = a\delta_T + \epsilon$ where a is a scaling constant and ϵ is independent noise with a prescribed distribution. Hence the basis is a linear function of the convenience yield plus some extra noise. To keep the model realistic, we assume that we do not observe the basis at intermediate time points. This is somewhat reasonable, since the market for local gas spot is illiquid and obtaining quotes requires physically contacting various market makers. The prices obtained in such manner are often unreliable or *stale* and it would make sense to discard them altogether rather than determine their accuracy.

The results from the previous section still hold because the HJB equation remains unchanged. We are only modifying the form of the payoff $\phi(S_T, B)$ which corresponds to the terminal condition. Consequently, the linearization goes through. However, now we do NOT have the separability in (25).

Repeating the computation,

$$\psi^{\phi}(t,s,\xi) = \widetilde{\mathbb{E}}\Big[\frac{1}{q}\log\int_{\mathbb{R}}\int e^{-q\phi(S_T,a\delta+\epsilon)}d\mathbb{P}_\epsilon\,\rho_T(\delta)\,d\delta\Big|S_t=s,\rho_t\sim\xi\Big]. \tag{27}$$

For example, for a forward and additive basis $\phi(S_T,B) = S_T + a\delta_t + \epsilon$, $\epsilon\sim\mathcal{N}(0,\sigma_\epsilon^2)$, the indifference price would be

$$\exp(-q(w-P^{Fwd}+\psi^{\phi})) = \exp\big(-q(w-\widetilde{\mathbb{E}}[\frac{1}{q}\log\int\rho_T(x)dx])\big)$$

$$\begin{aligned} P^{Fwd} &= \frac{1}{q}\widetilde{\mathbb{E}}\Big[\log\int\rho_T(x)dx\Big] - \psi^{\phi} \\ &= \widetilde{\mathbb{E}}\Big[S_T + \frac{\sigma_\epsilon^2}{2} + \frac{1}{q}\log\frac{\int\rho_T(\delta)\,d\delta}{\int e^{-qa\delta}\rho_T(\delta)\,d\delta}\Big]. \end{aligned} \tag{28}$$

In Section 7 we will show how to solve for P^{Fwd} using a Monte Carlo approach.

6.3. **Nonlinear Dynamics.** Looking back at (24) we see that the precise dynamics of the spot and the convenience yield did not matter, since we just used the corresponding differential operators. Thus, we can extend our model to include nonlinearities. One interesting case to consider is local volatility for the spot process. As mentioned before, stochastic volatility does not go well with filtering. However, we can use a local volatility function $\sigma=\sigma(S_t)$. An example is the CEV model

$$dS_t = S_t(r-\delta_t)\,dt + \sigma S_t^{1+\beta}dW_t.$$

Our filtering analysis would still go through and in fact everything can still be carried out. The advantage is that we now have the elasticity exponent β as an extra parameter which should facilitate empirical fitting. The spot price now enters the wealth dynamics

$$dw_t = rw_tdt - \pi\delta_tdt + \pi\sigma S_t^{\beta}dW_t,$$

as well as the dynamics of δ_t under $\widetilde{\mathbb{P}}$.

7. Numerical Results

To compute the various expectations obtained in Section 6, we must resort to Monte Carlo (MC) techniques. Because under $\widetilde{\mathbb{P}}$ the spot prices are local martingales these can be simulated independently of everything else. Thus, to perform MC first simulate N paths of the spot process. The simplest method is to use Euler discretization along a fine mesh on $[0,T]$. Then we need to run some sort of filtering algorithm to compute $\rho_T(x)$ along each path. Putting the two together, we can empirically evaluate (28) or (25). Note that we must numerically approximate the stochastic integral in (17) which means that we should simulate the spot on a finer mesh with $\Delta^{fine}t$ and then filter using a larger $\Delta^{filter}t$. For linear models like the basic Gibson-Schwartz (2) we can use the Kalman filter to filter δ_t exactly. However, in other situations, such as filtering δ_t in the Gibson-Schwartz CEV model discussed in the previous section we need a more robust method. Our candidate of choice is the particle filter algorithm also called sequential Monte Carlo.

7.1. **Particle Filtering for the Zakai Equation.** Our account is based on Crisan, Gaines and Lyons [6]. Let $M_F(\mathbb{R}^d)$ be the space of finite measures on $\mathbb{R}^d$ with the topology of weak convergence. At time t, the infinite dimensional random measure ρ_t is approximated by a totally atomic $A^N(t)$, which is an occupation measure of $N(t)$ particles $\{\alpha_t^i\}$. We use the superscript N because at time 0,

$$A^N(0) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\alpha_0^i},$$

where α_0^i are N independent identically distributed random variables with common distribution ξ. The weight of each particle always remains $\frac{1}{N}$ but their number $N(t)$ changes.

Discretize in time by choosing $T_k = k\Delta t$, with $T_0 = 0, T_M = T$. During an interval $[T_k, T_{k+1})$, each particle evolves independently according to the law of δ_t under $\widetilde{\mathbb{P}}$. At time T_{k+1} mutation occurs. Each particle branches such that the mean number of offsprings is given by (11)

$$\mu_k^i = \exp\Big(\int_{T_k}^{T_{k+1}} h' D_t^{-1} dY_t - \frac{1}{2}\int_{T_k}^{T_{k+1}} h' D_t^{-1} h \, dt\Big). \tag{29}$$

The branching of each particle is independent of all the others, and only depends on the behavior of Y_t on $[T_k, T_{k+1})$. The new particles inherit the location of their parent. To control the variance of $A^N(t)$, Crisan et al. [6] suggest using *minimal variance*, so that the number of offspring is either $\lfloor \mu_k^i \rfloor$ or $\lceil \mu_k^i \rceil$. Since $\mathbb{E}[\mu_k^i] = 1$, the expected mean number of particles always remains at N the initial number. It can also be shown that for any f continuous and bounded in $\mathbb{R}^d$, $\langle A^N(t), f\rangle$ is square integrable. Furthermore, if $N \to \infty$ and $\Delta t \to 0$ such that $N\sqrt{\Delta t} \to \infty$ (the number of particles grows quadratically in step size) then $A^N(t)$ weakly converges to a measure

$p(t) \in M_F(\mathbb{R}^d)$ [6, 7] satisfying

$$\langle p(t), f \rangle = \langle \pi_0, f \rangle + \int_0^t \langle p(s), \mathcal{L}_\delta f \rangle \, ds + \int_0^t \langle p(s), \mathcal{S} f \rangle \, dY_s \quad \text{a.s.} \tag{30}$$

By the characterization theorem of Kurtz and Ocone [24], we then have $p(t) = \rho_t$ in $M_F(\mathbb{R}^d)$.

The Zakai particle filter offers an advantage for our problem by directly computing ρ_T. In contrast, if we use the Kalman filter we must first compute the true conditional density of δ_t and then take a second step of approximating the Radon-Nikodym density η_T.

Remark 4. Note that in (25) the $\langle \rho_T, 1 \rangle$ term is just the total mass of the filter, i.e. $\frac{N(T)}{N}$.

7.2. **Parameter Values.** Table 1 summarizes the parameter values from our three references which we call respectively GS90, Sch97 and CL03. Gibson and Schwartz [13] fitted the Jan84- Nov88 time series for crude oil forwards of less than 9 month maturity. Schwartz [29] fitted the Jan90-Feb95 time series for forwards of less than 1 year maturity. In our own recent study [5] we fitted the Jan94-Aug02 time series for the 3-, 6- and 12-month forwards. As we can see the parameters, especially the mean-reversion rate κ are unstable in time and difficult to estimate. The parameter λ refers to the risk premium adjustments for the convenience yield.

Parameters	$GS90$	$Sch97$	$CL03$
κ	16.1	1.488	0.4
θ	0.309	-0.015	-0.15
γ	1.12	0.426	0.5
ρ	0.353	0.922	0.45
σ	0.320	0.358	0.6
λ	-1.796	0.291	0.03

TABLE 1. Empirical Parameter values for the Gibson-Schwartz model

7.3. **Comparative Statics.** We implemented both the Kalman filtering method using (17) and the particle filter using (29) and applied it to (25). The first numerical experiment that we perform is understanding the term-structure of the quantity $\widetilde{\mathbb{E}}[\log \int_{\mathbb{R}} \rho_T(x) dx]$. This is precisely the utility-based cost of being unable to observe δ_t. As Figure 2 illustrates, the price adjustment is almost linear with a slight convexity in the beginning.

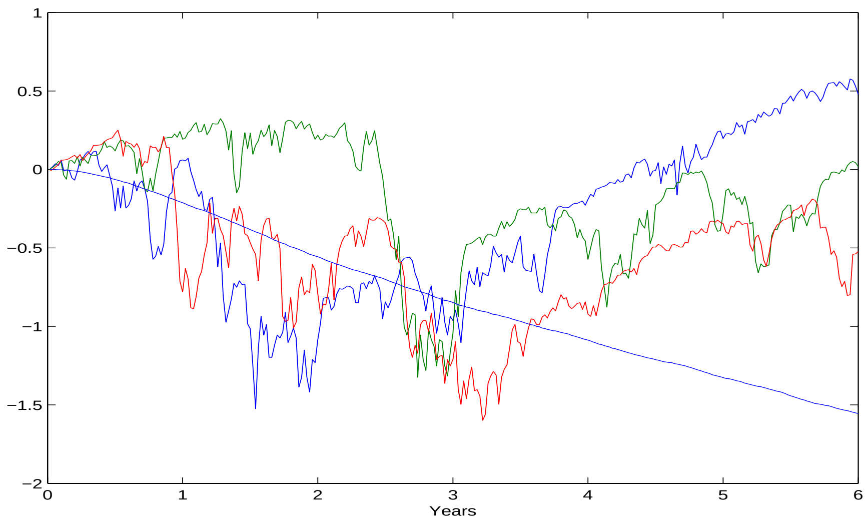

FIGURE 2. Three sample paths and the mean (the almost straight line) showing the term structure of $\log \int \rho_T(x)dx$. We used a particle filter with 500 initial particles. The parameter values are from Schwartz [29].

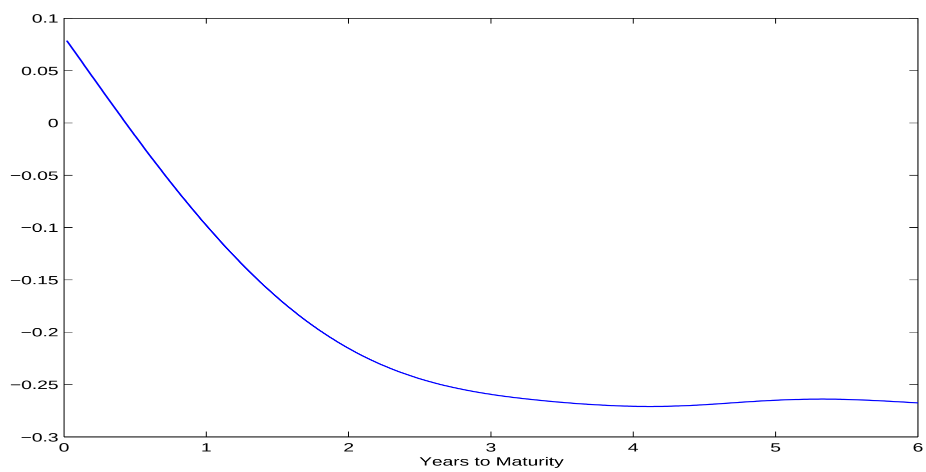

FIGURE 3. Term structure of $\frac{1}{q} \log \frac{\int \rho_T(\delta)\, d\delta}{\int e^{-qa\delta} \rho_T(\delta)\, d\delta}$. 5000 paths using a particle filter, $a = 5, q = \frac{1}{2}$.

We next try to understand the implications of formula (28) for uncertainty cost with basis depending on the convenience yield. This is the time-varying part of the forward curve, since S_T is a $\widetilde{\mathbb{P}}$-local martingale. In Figure 3 we see that this cost stabilizes and approaches a limit as $T \to \infty$. This is in strong contrast to the situation with complete information. We can think of the result as being a "middle ground" between the flat forward curve from (25) and the exponential curve from (4). This is consistent with the stylized empirical fact of "sticky" long end of the forward curve. As Table 2 demonstrates, the results are robust with respect to the model parameters. The general term structure of exponential decay to the long-term limit appears in all cases. The basic shape is monotonically decreasing due to the increasing cost of being unable to observe δ_t. A hump in the middle may occur depending on model parameters. Remember that we are assuming $r_t = 0$. Thus for positive interest rates, the forward curve may be upward sloping if r is sufficiently large.

8. Conclusion & Extensions

This paper demonstrates the feasibility of full treatment of a partially observed convenience yield model. Use of a latent factor model is preferable to relying on implied quantities when it comes to resolving model inconsistencies. While the general approach presents significant technical challenges, in the special case of exponential utility, everything boils down to a computation of a single reduced-form expectation. The latter may be computed by MC methods using filtering techniques.

As opposed to pricing by computing the expected value under a given equivalent martingale measure, utility-based prices have a limiting value as time to maturity increases. Intuitively the cost of unobserved stochastic drift stabilizes as the horizon increases. This is empirically desirable and stands in sharp contrast to full-information models that predict exponential behavior of the forward curve. Our approach essentially corresponds to pricing under the minimal martingale measure when S_t is a local martingale. Because the agent does portfolio optimization in addition to buying the derivative, the return on the spot is irrelevant and the term structure is determined by the risk coming from δ_t.

To obtain a fully satisfactory model for empirical data, further extensions would be necessary. For example, time-dependent parameters would surely be needed as gas prices exhibit high degrees of seasonality. Also, stochastic interest rates must be considered. On a more fundamental level, our model can be extended by presenting a more sophisticated approach for the basis factor. Here we have two choices. Either we model the local spot S^{NJ} as a process in its own right, which then becomes an observed but *non-traded* factor. Or we model the basis itself as a process, for example another Ornstein-Uhlenbeck. In the first situation our payoff depends just on S^{NJ},

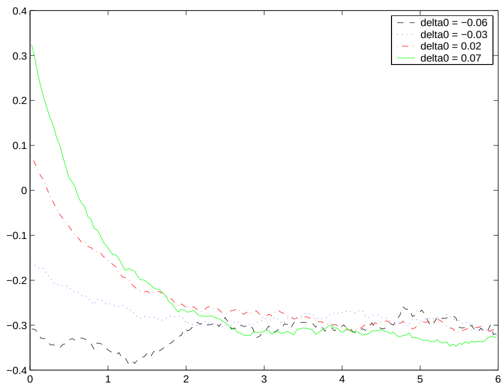

Figure 4: Varying the initial convenience yield mean.

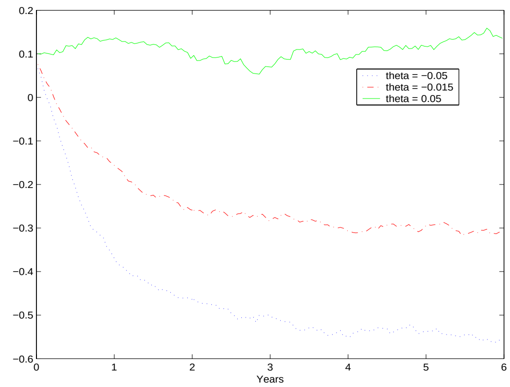

Figure 5: Varying the convenience yield mean-reversion level.

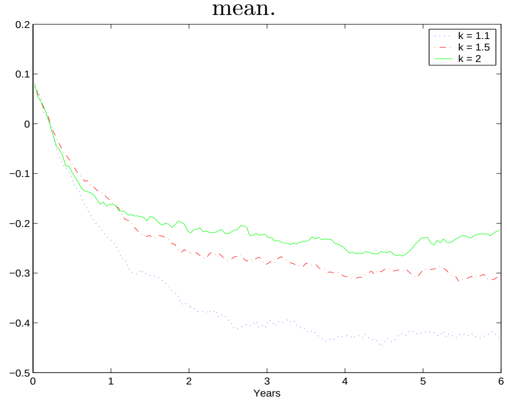

Figure 6: Varying the convenience yield mean-reversion speed.

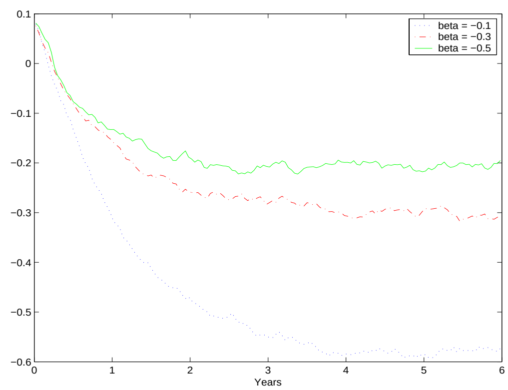

Figure 7: Varying the CEV elasticity.

TABLE 2. Comparative Statics for (28). Parameter values are from Schwartz [29]. We use the Kalman filter in (16), (15) except for the CEV model. $a = 5, q = \frac{1}{2}$.

but the second situation is likely to lead to simpler computations. One could consider a full three-factor correlated model for $[S_t, \delta_t, B_t]$. Unfortunately, as noted before it is not clear how to simplify the HJB equation in the presence of more factors.

A related problem that is very important for practical applications is the case of non-traded spot. In oil and gas markets, the spot market is illiquid and often features unreliable price quotes. More importantly, the inter-temporal transfer is complicated since the commodity must be physically stored. Hence, for practical purposes the concept of holding the spot is undesirable. Thus, an energy trader is likely to hedge derivatives on the spot using liquid instruments, first and foremost the forwards. In particular, the near forwards are highly correlated with the spot and hence can provide a reasonable hedge. One could write down a 3-factor model for the spot, the forward and the convenience yield. Structurally, it would be very similar to the 2-locations model in the previous paragraph. However, the important difference is that a forward is a financial asset that must yield a risk-free rate of return, so the convenience yield would not affect its dynamics.

References

[1] A. Bensoussan. *Stochastic Control of Partially Observable Systems*. Cambridge University Press, 1992.

[2] T. Bjork and C. Landen. On the term structure of futures and forward prices. Technical report, Stockholm School of Economics, 2001.

[3] M. Brennan. The price of convenience and the valuation of commodity contingent claims. In D. Lund and B. Oksendal, editors, *Stochastic Models and Option Values*, pages 33–71. Elsevier Science Publishers B.V., 1991.

[4] R. Carmona and V. Durrleman. Pricing and hedging spread options. *SIAM Review*, 45(4):627–685, 2003.

[5] R. Carmona and M. Ludkovski. Spot convenience yield models for energy assets. In *Mathematics of Finance*, AMS-SIAM Joint Summer Conferences, Providence, RI, 2004.

[6] D. Crišan, J. Gaines, and T. Lyons. Convergence of a branching particle method to the solution of the zakai equation. *SIAM Journal of Applied Mathematics*, 58(5):1568–1590, October 1998.

[7] D. Crišan. Exact rate of convergence for a branching particle approximation to the solution of the zakai equation. *preprint*, 2002.

[8] M. Davis. Option pricing in incomplete markets. In M. Dempster and S. Pliska, editors, *Mathematics of Derivative Securities*. Cambridge University Press, 1997.

[9] J.C. Duan and S. Pliska. Option valuation with co-integrated asset prices. *Journal of Economic Dynamics and Control*, 28:727–754, 2004.

[10] A. Eydeland and K. Wolyniec. *Energy and Power Risk Management: New Developments in Modeling, Pricing and Hedging*. John Wiley&Sons, Hoboken, NJ, 2003.

[11] W. Fleming and M. Soner. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag, New York, 1993.

[12] R. Gibson and E. S. Schwartz. Valuation of long term oil-linked assets. In D. Lund and B. Oksendal, editors, *Stochastic Models and Option Values*, pages 73–103. Elsevier Science Publishers B.V., 1989.

[13] R. Gibson and E. S. Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, XLV(3):959–976, 1990.

[14] F. Gozzi and A. Şwiech. Hamilton jacobi bellman equations for the optimal control of the duncan mortensen zakai equation. *Journal of Functional Analysis*, 172(2):466–510, 2000.

[15] A.C. Harvey. *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, UK, 1989.

[16] V. Henderson. Valuation of claims on nontraded assets using utility maximization. *Mathematical Finance*, 12(4):351–373, 2002.

[17] J.E. Hilliard and J. Reis. Valuation of commodity futures and options under stochastic convenience yields, interest rates, and jump diffusions in the spot. *Journal of Financial and Quantitative Analysis*, 33(1):61–86, March 1998.

[18] S. Hodges and A. Neuberger. Optimal replication of contingent claims under transactions costs. *Review of Futures Markets*, 8:222–239, 1989.

[19] P. Lakner. Optimal trading strategy for an investor: the case of partial information. *Stochastic Processes and their Applications*, 76:77–97, 1998.

[20] J-M. Lasry and P-L. Lions. Contrôle stochastique avec informations partielles et applications à la finance. *C. R. Acad. Sci. Paris, Série I*, 328:1003–1010, 1999.

[21] A. Mbafeno. Co-movement term structure and the valuation of energy spread options. Mathematics of Derivative Securities, pages 88–103. Cambridge University Press, 1997.

[22] M. Musiela and T. Zariphopoulou. A valuation algorithm for indifference prices in incomplete markets. *Finance & Stochastics*, forthcoming, 2004.

[23] H. Nagai. Risk-sensitive dynamic asset management with partial information. In T. Hida and B.S. Rajput et al., editors, *Stochastics in Finite and Infinite Dimension. A Volume in Honor of G. Kallianpur*, pages 321–340. Birkhauser, Boston, 2000.

[24] D. Ocone and T. Kurtz. Unique characterization of conditional distributions in nonlinear filtering. *Annals of Probability*, 16(1):80–107, 1988.

[25] H. Pham and M-C. Quenez. Optimal portfolio in partially observed stochastic volatility models. *Annals of Applied Probability*, 11(1):210–238, 2001.

[26] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, Cambridge, UK, 1992.

[27] R. Rouge and N. El Karoui. Pricing via utility maximization and entropy. *Mathematical Finance*, 10(2):259–276, 2000.

[28] W.J. Runggaldier. Estimation via stochastic filtering in financial market models. In *Mathematics of Finance*, AMS-SIAM Joint Summer Conferences, Providence, RI, 2004.

[29] E. Schwartz. The stochastic behavior of commodity prices:implications for valuation and hedging. *Journal of Finance*, LII(3):922–973, 1997.

[30] E. Schwartz and J. Smith. Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46 (7):893–911, July 2000.

[31] J. Sekine. Power-utility maximization for linear gaussian factor model under partial information. Technical report, Osaka University, 2003.

[32] J. Yong and X.Y. Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer-Verlag, New York, 1999.

[33] T. Zariphopoulou. A solution approach to valuation with unhedgeable risks. *Finance & Stochastics*, 5:61–82, 2001.

Department of Operations Research and Financial Engineering, Princeton, NJ 08544 USA

E-mail address: mludkovs@princeton.edu, rcarmona@princeton.edu