filter for spread prediction

Plan is to calibrate the model first, using linear KF, commodity price as the hidden variables and multiple futures prices as measurements (e.g. both the futures involved in the spread and at least one more futures price). Using multiple futures independently, rather than simply using a spread, gives some extra information and will hopefully give us a better model.

1 State space model and Kalman filter-based calibration

As futures on commodities tend to be more liquid than the commodities themselves in the spot market, they contain more information about the future behaviour of spot prices than the current spot market prices. Scatter[1] and correlation plots on figure **??** illustrate that futures prices are not perfectly correlated with spot prices and different futures contracts contain different information, than the spot price history, regarding the future behaviour of spot prices.

This fact can be exploited to infer the future spot price behaviour using futures prices via a Kalman filter, with the spot price as a latent state variable. Filtering has been used in estimating the spot prices from futures prices in [7], [6] and [5]. A multi-commodity implementation is presented in [1], where the futures prices on different commodities are used simultaneously to forecast the commodity prices.

They [6] introduced a multi-factor model for futures pricing with unobserved underlying commodity prices. The emphasis here is on storable energy commodities with highly liquid futures markets, *viz* crude oil, gasoline and natural gas. The evolution of log-spot price is modelled as a mean-reverting process, resulting in a linear state space system with log futures price vector as observable variable. Unlike [6], which uses a non-parametric seasonality component, the seasonality factor is modelled explicitly in this chapter through a simple parametrized sinusoid. a parametrized seasonality factor, and to a comprehensive empirical evaluation of the forecasting performance of models in subsequent sections. Four different models are described, depending on whether there are one or two-factors and whether or not there is a seasonality component. It is straightforward to extend this work beyond two-factors, although ours numerical experience indicates that models with three of more factors are rather difficult to calibrate reliably and tend to perform poorly out-of-sample, as far as forecasting is concerned.

Assume that a spot price S_t is driven by the process, $x_t = \log S_t : \{x_t \in \Omega\}$ on a probability space $(\Omega, \mathcal{P}, \mathcal{F}_t)$, where Ω is a set of all possible realisations of x_t, $\mathcal{P}$ is the objective probability measure defined on Ω and $\mathcal{F}_t$ is the natural filtration. For representing discrete time, the subscript n is used, ($t_n : n = 0, .., N$, $t_0 < t_1 < \cdots < t_N$, $\Delta := t_n - t_{n-1}$), where N is the total number of time intervals.

The log-spot price is assumed to follow an Ornstein-Ulenbeck type process:

$$dx_t = (\alpha - \kappa x_t)dt + \sigma dW_t^{\mathcal{P}}, \tag{1}$$

where κ and α represent mean-reversion speed and long-run mean of x_t, respectively, and $W_t^{\mathcal{P}}$ is a Wiener process. The fundamental theorem of asset pricing states that the absence of arbitrage opportunities on the market implies an existence of the equivalent martingale measure. Hence, process x_t has the following form if generated by a risk-neutral Wiener process $W_t^{\mathcal{Q}}$:

$$dx_t = (\tilde{\alpha} - \kappa x_t)dt + \sigma dW_t^{\mathcal{Q}}. \tag{2}$$

[1] Natural Gas spot prices were plotted against futures prices with different maturities for an interval of 1500 days, from 29.11.2007 to 12.09.2012.

The drifts are related by $\tilde{\alpha} = \alpha - \lambda_t \sigma$ for some process λ_t, i.e. $dW_t^{\mathcal{Q}} = dW_t^{\mathcal{P}} + \lambda_t dt$; see, e.g. [2] for the exact conditions on λ_t. λ is assumed to be a constant here, which is a commonly used assumption in the literature.

Under the risk-neutral measure, the process x_t is normally distributed. Using Ito's lemma for the function $f(x_t, t) = e^{\kappa t} x_t$, it can be easily shown that x_t has the following mean and variance:

$$\mathbb{E}^{\mathcal{Q}}(x_{t+\Delta} \mid \mathcal{F}_t) = x_t e^{-\kappa\Delta} + \frac{\tilde{\alpha}}{\kappa}(1 - e^{-\kappa\Delta}), \tag{3}$$

$$\mathrm{Var}^{\mathcal{Q}}(x_{t+\Delta} \mid \mathcal{F}_t) = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa\Delta}\right). \tag{4}$$

$$\mathbb{E}^{\mathcal{Q}}(x_{t+\Delta} \mid \mathcal{F}_t) = x_t e^{-\kappa\Delta} + \frac{\alpha - \lambda\sigma}{\kappa}(1 - e^{-\kappa\Delta}), \tag{5}$$

$$\mathrm{Var}^{\mathcal{Q}}(x_{t+\Delta} \mid \mathcal{F}_t) = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa\Delta}\right). \tag{6}$$

Next, let $T = \{T_i : i = 1, ..., m, 0 < T_1 < T_2 < \cdots < T_m\}$ be the collection of the futures maturity dates. Then futures price for maturity T_i for a commodity with log-spot price x_t at time $t < T_i$ can be written as a conditional expectation of the commodity price at the maturity time of the futures contract: $F(t, T_i) = \mathbb{E}^{\mathcal{Q}}(e^{x^i} \mid \mathcal{F}_t)$, $i = 0, .., m$, where the expectation is taken under the $\mathcal{Q}$ measure and $x^i := x_{T_i}$, for brevity of notation. In the case if $T_i > t$, the futures price $F(t, T_i) > 0$, otherwise it is zero. The time to expiry of the i^{th} futures contract is represented by $\Delta_t^i = T_i - t$. Since S_t is log-normally distributed, the futures price is given by:

$$F(t, T_i) = \mathbb{E}^{\mathcal{Q}}(e^{x^i} \mid \mathcal{F}_t) = e^{\mathbb{E}^{\mathcal{Q}}(x^i|\mathcal{F}_t) + \frac{1}{2}\mathrm{Var}^{\mathcal{Q}}(x^i|\mathcal{F}_t)}. \tag{7}$$

This allows us to derive an affine equation for the vector of log futures prices in terms of the log-spot price:

$$vec\{y_t^i\} = x_t e^{-\kappa\Delta_t^i} + \frac{\alpha - \lambda\sigma}{\kappa}(1 - e^{-\kappa\Delta_t^i}) + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa\Delta_t^i}), \tag{8}$$

where $y_t^i = \log F(t, T_i)$ and the vec operator is defined by

$$vec(z_i) = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}^\top.$$

Note that the convenience yield is not modelled explicitly and assume that it is already reflected in the prices of futures contracts. Again, our approach is consistent with the framework followed in [6]. In contrast, convenience yield is explicitly modelled in [4].

1.1 Modelling seasonality

Since energy futures prices depend on the weather conditions, any seasonality pattern needs to be taken into account. In the literature, a variety of seasonality functions for different financial application is used. For example, Manoliu and Tompaidis [6] used a discrete seasonality function with separate parameters representing each month, while Sorensen [8] used a Fourier series to model seasonality. However, a complicated seasonality function makes parameter estimation more difficult and may lead to poorer estimates, especially when the data set is small relative to the number of the parameters. To reduce the parameter estimation complexity, a simple function for seasonality is considered, which is parametrised as follows:

$$f(t) = \exp(c_1 + c_2 \sin(c_3 t + c_4)), \tag{9}$$

where c_1 is a constant level, c_2, c_3 and c_4 are constants representing amplitude, the frequency and the phase of a seasonal pattern respectively. Accordingly, the prices of futures are modified as follows:

$$F(t, T_i) = f(T_i)\mathbb{E}^{\mathcal{Q}}(e^{x^i} \mid \mathcal{F}_t), \tag{10}$$

and

$$vec\{y_t^i\} = \log f(T_i) + x_t e^{-\kappa \Delta_t^i} + \frac{\alpha - \lambda\sigma}{\kappa}(1 - e^{-\kappa \Delta_t^i}) + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa \Delta_t^i}), \tag{11}$$

which denotes a vector of log futures prices, with i^{th} element of the vector denoting log futures price for time to maturity Δ_t^i, as before. In practice, one may parametrise seasonality using multiple sinusoids. However, in our experience, this complicates parameter estimation without necessary improving the quality of out of sample price forecasting.

2 Linear state space representation for latent commodity price models

For the models described in subsections 1.1 and **??**, a state space representation is used, with a measurement equation based on the observable time series of futures prices and a discretized transition equation of log-spot commodity price, which is assumed to be unobservable. This allows us to use the Kalman filter to estimate the parameters by constructing and maximising a likelihood function. Later, we will use this model to forecast the log-spot price and hence indirectly forecast a calendar spread.

The state space equations for one factor with seasonality model in subsection 1.1are provided below.

2.1 One-factor model with seasonality

The state space equations corresponding to the model in section 1.1 can be written as

$$x_{n+1} = Bx_n + g + Rw_{n+1}, \tag{12}$$

$$y_n^l = A_n x_n + d_n + Q^l z_n^l, \tag{13}$$

where the state space model parameters may be expressed in terms of original model parameters as:

$$f(t_n) = c_1 + c_2 \sin(c_3 t_n + c_4), \tag{14}$$

$$B = e^{-\kappa\Delta}, \; g = \frac{\alpha}{\kappa}(1 - e^{-\kappa\Delta}), \tag{15}$$

$$R^2 = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta}), \; A_n = \begin{pmatrix} e^{-\kappa\Delta_n^1} \\ \vdots \\ e^{-\kappa\Delta_n^m} \end{pmatrix}, \tag{16}$$

$$
d_n = \begin{pmatrix} \frac{\alpha-\lambda\sigma}{\kappa}(1-e^{-\kappa\Delta_n^1}) + \frac{\sigma^2}{4\kappa}(1-e^{-2\kappa\Delta_n^1}) + f(T_1) \\ \vdots \\ \frac{\alpha-\lambda\sigma}{\kappa}(1-e^{-\kappa\Delta_n^m}) + \frac{\sigma^2}{4\kappa}(1-e^{-2\kappa\Delta_n^m}) + f(T_m) \end{pmatrix}. \tag{17}
$$

Here, $\Delta_n^i = \Delta_{t_n}^i = T_i - t_n$ for brevity of notation and m is the number of futures prices available at each t_n. $Q^l = \eta^l I_m$, where η^l is a scalar constant indicating the standard deviation of measurements and I_m is an $m \times m$ identity matrix. Superscript l for y, Q and η denotes a log futures price model.

The brief outline on how this state space representation is used along with Kalman filter for parameter estimation can be found in section **??**; see, e.g. [3] for more details on financial time series filtering using state space models.

2.2 Maximum Likelihood(ML) estimation

For the given log futures prices measurements $F = \{y_1, y_2, \ldots, y_N\}$ up to time t_N, Kalman filter can be applied to calibrate parameters of (14)-(17) and (**??**)-(**??**). The joint likelihood function for F can be written as follows:

$$
\hat{L}(F) = p(y_1^l) \prod_{i=2}^{N} p(y_i^l \mid \mathcal{F}_{i-1}), \tag{18}
$$

which, after substituting for joint probabilities and taking logarithms becomes

$$
\log \hat{L}(F) = -\sum_{i=1}^{N} (\log \mid \Sigma_i \mid + v_i^T \Sigma_i^{-1} v_i), \tag{19}
$$

where v_i, Σ_i are the innovations at time t_i and the covariance of innovations at time t_i respectively, and are as defined in section **??**. The constant terms which do not depend on the model parameters are ignored. For a given vector-valued time series $\left\{y_1^l, y_2^l, \ldots, y_N^l\right\}$ and a vector of unknown model parameters Ψ, the optimisation problem can be stated as following:

$$
\hat{\Psi} = \arg\max_{\Psi} \log \hat{L}(F), \tag{20}
$$

$\hat{\Psi}$ is then used for forecasting experiments.

3 Nonlinear filter for forecasting spreads

3.1 Derivation of filtering equations

Given a calibrated model with B, g, A_n, d_n and R, consider a version of model where the measurement is *not* log futures prices, but a single calendar spread between two futures prices:

$$
x_{n+1} = Bx_n + g + Rw_{n+1} \tag{21}
$$

$$
y_n = \left(\exp(A_n^1 x_n + d_n^1) - \exp(A_n^2 x_n + d_n^2)\right) + Qz_n, \tag{22}
$$

where B_n, g are as defined previously; $A_n^i = e^{-\kappa\Delta_n^i}$,

$$
d_n^i = \frac{\alpha - \lambda\sigma}{\kappa}(1 - e^{-\kappa\Delta_n^i}) + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa\Delta_n^i}) + f(T_i),
$$

$i = 1, 2$; w_n, v_n are zero mean unit variance Gaussian noises. Suppose that $\hat{x}_{n-1|n-1}$ and $P_{n-1|n-1}$ are available; further, assume that $\mathbb{E}_{n-1+j}(x_{n-1}) = \hat{x}_{n-1|n-1}$, $j \geq 0$. The recursive equations for a linear minimum variance filter can then be written as follows:

$$v_n = y_n - \mathbb{E}_{n-1}(y_n), \tag{23}$$

$$\Sigma_{n|n-1} = \mathbb{E}_{n-1}(v_n v_n^\top), \tag{24}$$

$$\hat{x}_{n|n} = \hat{x}_{n|n-1} + K_n^* v_n, \text{ where} \tag{25}$$

$$K_n^* = \arg\min_{K_n} \text{ trace } \underbrace{\mathbb{E}_n \left((x_n - \hat{x}_{n|n})(x_n - \hat{x}_{n|n})^\top \right)}_{P_{n|n}}, \tag{26}$$

$$\hat{x}_{n+1|n} = B\hat{x}_{n|n} + g, \tag{27}$$

$$P_{n+1|n} = \mathbb{E}_{n+1} \left((\hat{x}_{n+1|n} - x_{n+1})(\hat{x}_{n+1|n} - x_{n+1})^\top \right). \tag{28}$$

Here, $\mathbb{E}_{n-1}$ is expectation based on information up to time $n-1$. We now proceed to derive closed-form recursive expression for K_n, in terms of parameters. For this purpose, the following fact is crucial: if x is a Gaussian random variable with mean $\bar{x}$ and variance β, $\mathbb{E}(\exp(x)) = \exp\left(\bar{x} + \frac{\beta}{2}\right)$. Note that log commodity price is Gaussian according to our assumption, and hence y_n is a difference of exponential Gaussian random variables. Using this fact and using the fact that variance of x_n based on information up to time $n-1$ is is $P_{n|n-1}$, we get

$$\begin{aligned} \mathbb{E}_{n-1}(y_n) &= \exp(A_n^1 x_{n|n-1} + d_n^1 + \frac{1}{2} A_n^1 P_{n|n-1} (A_n^1)^\top) \\ &\quad - \exp(A_n^2 x_{n|n-1} + d_n^2 + \frac{1}{2} A_n^2 P_{n|n-1} (A_n^2)^\top). \end{aligned} \tag{29}$$

Next, $\Sigma_{n|n-1}$ can be written as

$$\begin{aligned} \Sigma_{n|n-1} &= \mathbb{E}_{n-1}(v_n v_n^\top) = \mathbb{E}_{n-1}(y_n^2) - (\mathbb{E}_{n-1}(y_n))^2, \\ &= \mathbb{E}_{n-1}(\alpha_n^1 - \alpha_n^2 + Q z_n)^2 - (\mathbb{E}_{n-1}(y_n))^2, \end{aligned} \tag{30}$$

where the first step follows since $var(x) = E(x^2) - (E(x))^2$, and $\alpha_n^i = \exp(A_n^i x_n + d_n^i)$, $i = 1, 2$, for brevity. The second term in (30) is given by squaring the right hand side of (29). For the first term, note that

$$\begin{aligned} &\mathbb{E}_{n-1}(\alpha_1 - \alpha_2 + Q z_n)^2 = \mathbb{E}_{n-1}(\alpha_1^2) + \mathbb{E}_{n-1}(\alpha_2^2) - 2\mathbb{E}_{n-1}(\alpha_1 \alpha_2) + Q^2 \\ &= \exp(2A_n^1 x_{n|n-1} + 2d_n^1 + A_n^1 P_{n|n-1}(A_n^1)^\top) + \exp(2A_n^2 x_{n|n-1} + 2d_n^2 + A_n^2 P_{n|n-1}(A_n^2)^\top) \\ &- 2\exp((A_n^1 + A_n^2) x_{n|n-1} + (d_n^1 + d_n^2) + \frac{1}{2}(A_n^1 + A_n^2) P_{n|n-1} (A_n^1 + A_n^2)^\top) + Q^2. \end{aligned} \tag{31}$$

The next step is to find $\mathbb{E}_n \left((x_n - \hat{x}_{n|n})(x_n - \hat{x}_{n|n})^\top \right)$ in terms of K_n, so that we can minimize it over K_n. For this purpose, first note that

$$\begin{aligned} x_n - \hat{x}_{n|n} &= x_n - \left\{ \hat{x}_{n|n-1} + K_n(y_n - \mathbb{E}_{n-1}(y_n)) \right\} \\ &= Bx_{n-1} + g + Rw_n - (B\hat{x}_{n-1|n-1} + g + K_n(y_n - \mathbb{E}_{n-1}(y_n)) \\ &= B \underbrace{(x_{n-1} - \hat{x}_{n-1|n-1})}_{\phi_{n-1}} - K_n(y_n - \mathbb{E}_{n-1}(y_n)) + Rw_n. \end{aligned} \tag{32}$$

Thus, ignoring any terms which do not involve K_n, we have

$$\begin{aligned}\mathbb{E}_n\left((x_n-\hat{x}_{n|n})(x_n-\hat{x}_{n|n})^\top\right) &= K_n\Sigma_{n|n-1}K_n^\top - 2BK_n\mathbb{E}_n\left(\phi_{n-1}(y_n-\mathbb{E}_{n-1}(y_n))\right)\\ &= K_n\Sigma_{n|n-1}K_n^\top - 2BK_n\mathbb{E}_n\left(\phi_{n-1}y_n\right), \qquad (33)\end{aligned}$$

Since $\mathbb{E}_n(\phi_{n-1}\mathbb{E}_{n-1}(y_n))=0$, due to our assumption that $\mathbb{E}_n(x_{n-1})=\hat{x}_{n-1|n-1}$. The next challenge is to evaluate the expectation of the last term in (33). For this, write y_n as

$$\begin{aligned}y_n &= \exp(A_n^1x_n+d_n^1)-\exp(A_n^2x_n+d_n^2)+Qz_n\\ &= \exp(A_n^1(Bx_{n-1}+g+Rw_n)+d_n^1)-\exp(A_n^2(Bx_{n-1}+g+Rw_n)+d_n^2)+Qz_n\\ &= \exp\left(A_n^1(B(x_{n-1}-\hat{x}_{n-1|n-1})+B\hat{x}_{n-1|n-1}+g+Rw_n)+d_n^1\right)\\ &- \exp\left(A_n^2(B(x_{n-1}-\hat{x}_{n-1|n-1})+B\hat{x}_{n-1|n-1}+g+Rw_n)+d_n^2\right)+Qz_n\\ &= \exp\left(A_n^1(B\phi_{n-1}+B\hat{x}_{n-1|n-1}+g+Rw_n)+d_n^1\right)\\ &- \exp\left(A_n^2(B\phi_{n-1}+B\hat{x}_{n-1|n-1}+g+Rw_n)+d_n^2\right)+Qz_n. \qquad (34)\end{aligned}$$

Let us consider $\mathbb{E}(\phi_{n-1}y_n)$ with y_n as in (34), one term at a time. The last term is the easiest: $\mathbb{E}_n(\phi_{n-1}Qz_n)=\mathbb{E}_n(\phi_{n-1})\mathbb{E}_n(Qz_n)=0$. For further brevity of notation, let

$$\zeta_n^i=\exp(A_n^i(B\hat{x}_{n-1|n-1}+g+d_n^i)),\ i=1,2. \qquad (35)$$

For each of the two terms, we have

$$\begin{aligned}&\mathbb{E}\left(\phi_{n-1}\exp\left(A_n^i(B\phi_{n-1}+B\hat{x}_{n-1|n-1}+g+Rw_n)+d_n^i\right)\right)\\ &=\zeta_n^i\mathbb{E}\left(\phi_{n-1}\exp(A_n^iB\phi_{n-1})\right)\mathbb{E}(\exp(Rw_n)) \qquad (36)\end{aligned}$$

Now, we use the following facts:

1. $\mathbb{E}(\exp(Rw_n))=\exp\left(\frac{R^2}{2}\right)$ since w_n is standard normal;

2. Recall that the variance of ϕ_{n-1} is $P_{n-1|n-1}$, and drop the subscripts of P and ϕ for readability, for the moment. Since ϕ is normally distributed random variable with mean zero and variance P, we have

$$\begin{aligned}\mathbb{E}(\phi\exp(A_n^iB\phi)) &= \frac{1}{\sqrt{2\pi P}}\int_{-\infty}^{\infty}\phi\exp(A_n^iB\phi)\exp\left(-\frac{\phi^2}{2P}\right)d\phi\\ &= \frac{1}{\sqrt{2\pi P}}\int_{-\infty}^{\infty}\phi\exp\left(-\frac{(\phi-A_n^iBP)^2}{2P}+\frac{(A_n^iB)^2P}{2}\right)d\phi\\ &= \exp\left(\frac{(A_n^iB)^2P}{2}\right)\frac{1}{\sqrt{2\pi P}}\int_{-\infty}^{\infty}\phi\exp\left(-\frac{(\phi-A_n^iBP)^2}{2P}\right)d\phi\\ &= \exp\left(\frac{(A_n^iB)^2P}{2}\right)A_n^iBP, \qquad (37)\end{aligned}$$

where the second equality is simple completion of squares and the last step follows since the integral simply evaluates the mean of a normally distributed random variable with mean A_n^iBP and variance P.

Returning to (36), we have

$$\begin{aligned}&\mathbb{E}\left(\phi_{n-1}\exp\left(A_n^i(B\phi_{n-1}+B\hat{x}_{n-1|n-1}+g+Rw_n)+d_n^i\right)\right)\\ &=\zeta_n^i\exp\left(\frac{(A_n^iB)^2P_{n-1|n-1}+R^2}{2}\right)A_n^iBP_{n-1|n-1},\end{aligned}\tag{38}$$

where ζ_n^i is as defined in (35). Substituting back in (33),

$$\begin{aligned}&\mathbb{E}_n\left((x_n-\hat{x}_{n|n})(x_n-\hat{x}_{n|n})^\top\right)=K_n\Sigma_{n|n-1}K_n^\top\\ &-2BK_nBP_{n-1|n-1}\left(\zeta_n^1\exp\left(\frac{(A_n^1B)^2P_{n-1|n-1}+R^2}{2}\right)A_n^1-\zeta_n^2\exp\left(\frac{(A_n^2B)^2P_{n-1|n-1}+R^2}{2}\right)A_n^2\right),\end{aligned}\tag{39}$$

where we have still ignored the terms not involving K_n on the right hand side. Differentiating with respect to K_n and setting it to zero gives our optimal filter gain which minimizes variance:

$$K_n^*=\frac{B^2P_{n-1|n-1}\left(\zeta_n^1\exp\left(\frac{(A_n^1B)^2P_{n-1|n-1}+R^2}{2}\right)A_n^1-\zeta_n^2\exp\left(\frac{(A_n^2B)^2P_{n-1|n-1}+R^2}{2}\right)A_n^2\right)}{\Sigma_{n|n-1}}.\tag{40}$$

This is a minimum since the second derivative of the right hand side of (39) is positive. The filter is linear in y_n, but is **very** nonlinear in the past state as there is $\hat{x}_{n-1|n-1}$ in the exponent for ζ_{n-1}^i.

Now that we have K_n^*, we still need two more things to complete the recursion of the filter: explicit expressions for $P_{n|n}$ and $P_{n|n-1}$. Strictly speaking, later is needed only if we want prediction of future spreads. Recall that we had ignored the term which did not involve K_n while computing $P_{n|n}$ earlier. Now writing all the terms in $P_{n|n}$ with $K_n=K_n^*$ in (39),

$$\begin{aligned}&P_{n|n}=\mathbb{E}_n\left((x_n-\hat{x}_{n|n})(x_n-\hat{x}_{n|n})^\top\right)=BP_{n-1|n-1}B^\top+R^2+K_n^*\Sigma_{n|n-1}(K_n^*)^\top\\ &-2BK_n^*BP_{n-1|n-1}\left(\zeta_n^1\exp\left(\frac{(A_n^1B)^2P_{n-1|n-1}+R^2}{2}\right)A_n^1-\zeta_n^2\exp\left(\frac{(A_n^2B)^2P_{n-1|n-1}+R^2}{2}\right)A_n^2\right)\end{aligned}\tag{41}$$

$$=BP_{n-1|n-1}B^\top+R^2-K_n^*\Sigma_{n|n-1}(K_n^*)^\top.\tag{42}$$

Lastly, since

$$(x_{n+1}-\hat{x}_{n+1|n})=(Bx_n+g+Rw_{n+1})-(Bx_{n|n}+g)=B(x_n-x_{n|n})+Rw_{n+1},$$

we have

$$P_{n+1|n}=\mathbb{E}_n\left((x_{n+1}-\hat{x}_{n+1|n})(x_{n+1}-\hat{x}_{n+1|n})^\top\right)=BP_{n|n}B^\top+R^2.\tag{43}$$

3.2 Filtering algorithm

To summarize, the steps for a filter recursion, which maps $(P_{n-1|n-1},P_{n|n-1},\hat{x}_{n|n-1})$ to $(P_{n|n},P_{n+1|n},\hat{x}_{n+1|n})$ once y_n becomes available are as follows:

1. Calculate v_n using (23) and (29).
2. Calculate $\Sigma_{n|n-1}$ using (30), (31) and (29), the last equation being needed to calculate $(\mathbb{E}_{n-1}(y_n))^2$.
3. Calculate K_n^* using (40) and (35).
4. Calculate $\hat{x}_{n|n}$ and $\hat{x}_{n+1|n}$ using (25) and (27), respectively.
5. Calculate $P_{n|n}$ and $P_{n+1|n}$ using (41) and (43), respectively.

Some further notes on this nonlinear filter

It is worth reflecting on our exact assumptions and what we achieve. At time step n, we assume that $\mathbb{E}(x_{n-1}|\mathcal{F}_{n-1}) = \hat{x}_{n-1|n-1}$ and find a linear, minimum variance estimate of x_n. At time step $n+1$, our next linear minimum variance estimate assumes that $\mathbb{E}(x_n\mathcal{F}_n) = \hat{x}_{n|n}$. However, 'conditional expectation = minimum variance estimate' is true for a linear Gaussian system and is not true in our case. The filter is only 'one step optimal', which is still better than linearized filters, *e.g.* extended Kalman filter or unscented filter which are not even one step optimal. We have one more, very big advantage on a conventional filtering set-up; x_n is actually observed. Thus we can simply set $\hat{x}_{n-1|n-1} = x_{n-1}$, *i.e.* use the actual last log commodity price as the estimated last log commodity price at each step for computing ζ_n^i in (35). This makes the assumption for one step prediction correct, and can give decent predictions. The 'output' of filter/ predictor in this case would be $\hat{x}_{n+1|n}$ and $P_{n+1|n}$; $\hat{x}_{n|n}$ is not used again.

The next step would be looking at spread widening or tightening. We can simply generate samples of normally distributed x_{n+1}, with mean $\hat{x}_{n+1|n}$ and variance $P_{n+1|n}$, and correspondingly generate samples from predictive distribution of y_{n+1} (which is a known function of the random variable x_{n+1} - equation 34). As an example: if we generate 100 samples and 60 indicate y_{n+1} is going to be higher than y_n, we go with spread widening as our prediction.

Currently, I have written everything up with mixed-up notation (some $APA^\top$ should be written as A^2P, for example). I was too happy with getting to the end of it, for cleaning it up afterwards. I will do so in due course.

If we are trying to predict futures prices (and not spreads or log futures prices), it is possible to write the whole thing with individual futures prices as measurements (instead of log futures prices). With vector valued measurements, the formulae will look different, but I am sure it will work. Using $x_{n-1|n-1}$= actual last log commodity price, I think this will give pretty good predictions!

References

[1] G. Cortazar, C. Mila, and F. Severino. A multicommodity model of futures prices: Using futures prices of one commodity to estimate the stochastic process of another. *The Journal of Futures Markets*, 28(6):537–560, 2008.

[2] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, 1996.

[3] J. Durbin, S. Koopman, and A. Atkinson. *Time series analysis by state space methods*, volume 15. Oxford University Press Oxford, 2001.

[4] C. Hyndman. Gaussian factor models - futures and forward prices. *IMA Journal of Management Mathematics*, 18(4):353–369, 2007.

[5] D. Lautier and A. Galli. Simple and extended Kalman filters: an application to term structures of commodity prices. *Applied Financial Economics*, 14(13):963–973, 2004.

[6] M. Manoliu and S. Tompaidis. Energy futures prices: term structure models with Kalman filter estimation. *Applied Mathematical Finance*, 9(1):21–43, 2002.

[7] E. Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance New-York*, 52(3):923–974, 1997.

[8] C. Sørensen. Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, 22(5):393–426, 2002.