MIT Sloan Finance Research Practicum
Weekly Memorandum - SRR

TO:	Eduardo Roman, Ioannis Michopoulos
FROM:	Jirui Qin, Susan Wang, and Beau Zhao
DATE:	January 23, 2015
SUBJECT:	Multi-factor Models for Forward Curve

I. Introduction

The second part of the project is to construct a multi-factor model to simulate the forward curve in energy markets. This requires us to calibrate the term structure of volatilities and correlations based on historical market quotes, and use this calibration to produce the forward price curve. We decide to start with oil futures as the underlying commodity, and apply the well-known Schwartz and Smith 2-factor model, and 3-factor model by Cortazar. In addition, we want to perform the Seasonal Principal Component Analysis (PCA), a method developed by Financial Engineering Associates (FEA), to capture the seasonal nature for the evolution of forward prices.

II. Multifactor Pricing Models

2-Factor by Schwartz and Smith

The premise of Schwartz and Smith 2-factor model are geometric Brownian motion and mean reversion property of the price of the underlying commodity. The equilibrium price of the commodity, an important component of the model, is assumed to follow a geometric Brownian motion with drift that reflects current expectations of supply, technology, and regulatory policy. Short-term deviation, another important factor, captures the difference between the spot and equilibrium price, and is assumed to revert towards zero, which follows an Ornstein-Uhlenbeck process. Deviations can be a result of short-term fluctuations in demand due to weather changes, or the ability of market forces to adjust inventory in response to changing market conditions.

As this model requires both short-term and long-term spot and future prices, neither of the two factors is directly observable. However, "changes in the equilibrium price, and changes in the difference between near- and long-term futures prices give information about the short-term deviations." (Schwartz, Smith, pp.2) Our goal is to study and replicate this model by fitting the factors with Kalman filtering.

With the geometric Brownian motion and zero-reversion assumptions, the analytic form of the two factors is derived as,

$$\mathrm{E}[(\chi_t,\ \xi_t)] = [e^{-\kappa t}\chi_0,\ \xi_0 + \mu_\xi t] \quad \text{and}$$

$$\mathrm{Cov}[(\chi_t,\ \xi_t)] = \begin{bmatrix} (1-e^{-2\kappa t})\dfrac{\sigma_\chi^2}{2\kappa} & (1-e^{-\kappa t})\dfrac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} \\ (1-e^{-\kappa t})\dfrac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} & \sigma_\xi^2 t \end{bmatrix},$$

where χ_t is the short-term deviation in prices
ξ_t, equilibrium price
κ, short-term deviation zero- reversion coefficient

In order to estimate these two factors, we have to apply Kalman filtering. This procedure uses a recursive formula to compute the unobserved factors based on available observations that depend on the two factors. The details of the Kalman filtering are not included in this memo. Please refer to *Short-Term Variations and Long-Term Dynamics in Commodity Prices* for further information.

3-Factor by Cortazar

Generally, the two factor model is sufficient in modelling the futures prices. However in certain market conditions, the two factor model is unable to fit well the cross section of futures prices and we need an additional factor. The three-factor models are often necessary to explain the variations in commodity futures curve. The Cortazar three factor model are calibrated using only the daily commodity-linked market asset prices. We use the following equation to find the futures price:

$$
\begin{aligned}
F(S,y,v,T) = S\exp\Big[& -y\frac{1-e^{-\kappa T}}{\kappa} + v\frac{1-e^{aT}}{a} + -(\lambda_1)T + \frac{\lambda_2 - \sigma_1\sigma_2\rho_{12}}{\kappa^2} \\
& \times(\kappa T + e^{-\kappa T} - 1) + \frac{\sigma_2^2}{4\kappa^3}(-e^{-2\kappa T} + 4e^{-\kappa T} + 2\kappa T - 3) \\
& + \frac{a\bar{v} - \lambda_3 + \sigma_1\sigma_3\rho_{13}}{a^2}(aT + e^{-aT} - 1) - \frac{\sigma_3^2}{4a^3} \\
& \times(e^{-2aT} - 4e^{-aT} - 2aT + 3) - \frac{\sigma_2\sigma_3\rho_{23}}{\kappa^2 a^2(\kappa + a)}(\kappa^2 e^{-aT} + \kappa a e^{-aT} \\
& + \kappa a^2 T + \kappa a e^{-\kappa T} + a^2 e^{-\kappa T} - \kappa a e^{-(\kappa + a)T} - \kappa^2 - \kappa a - a^2 + \kappa^2 aT)\Big]
\end{aligned}
$$

Seasonal PCA

PCA is a statistical tool used to identify an underlying structure within a dataset of interrelated variables. It has been widely used in trading and risk management to price interest rate models by reducing the number of variables into a set of three factors, typically described as level, slope, and curvature. This approach is more realistic than the conventional methodology that simply collects the volatilities and correlations of each point on the forward curve, which can lead to complex, inefficient calculations when we want to increase the number of maturities on the forward curve. PCA, however, can considerably reduce the number of dimensions to just a few of them (called principal components). Each principal component will have a factor score which explains the significance of contribution of that particular risk to the overall variability of the data. Each principal component factor is also associated with a factor loading that describes the volatility of each maturity with respect to the change in one source of risk. The principal components are constructed in such a way that they are orthogonally independent from each other by first constructing a variance-covariance matrix, and then eigenvalue-decomposing it.

While applying PCA to our data the seasonal behavior of the energy commodity (for example oil futures) should also be incorporated. The solution, proposed by the FEA, is to break down the historical forward price observations into monthly groups, and then perform 12 separate PCAs to each month of data. Once we have completed the Seasonal PCA, we can then use it to simulate the forward curve from thereafter. It is simulated in such a way that preserves the joint changes between spot price movements and forward price changes.

III. Pearson Model Linear Approximation Optimization

After implementing the Pearson linear approximation model in MATLAB, we successfully optimized two constant values *k* and *n* that calculates the constant *h* to form the piecewise linear approximation, where k is a positive constant and 2*n is the number of areas to be calculated under the curve:

$$h = \frac{k\sigma_1\sqrt{T-t}}{\sqrt{n}}$$

For each constant, we use the linear approximation model to derive a matrix of results with varying correlations and strike values. Then, each value in the matrix was subtracted by the corresponding value from Monte Carlo simulation. All the elements in the resulting matrix were squared and summed to derive an error value for each constant value. After optimization, we found the value of *k* to be 1.04 and *n* to be 32, which both produce the least error value of 0.002163794. Below is the 3-D graph with the correlation values from 1 to -1 on the x-axis, the strike values from -20 to -25 on the z-axis and the error values on the y-axis. Thus by optimizing values that gives an approximation closest to the Monte Carlo simulation, we can use these constants to approximate future spread option prices.