Knowledge-Driven Autonomous Commodity Trading Advisor

Yee Pin Lim
School of Information Systems
Singapore Management University
Republic of Singapore
sharon.lim.2011@smu.edu.sg

Shih-Fen Cheng
School of Information Systems
Singapore Management University
Republic of Singapore
sfcheng@smu.edu.sg

Abstract*—The myth that financial trading is an art has been mostly destroyed in the recent decade due to the proliferation of algorithmic trading. In equity markets, algorithmic trading has already bypass human traders in terms of traded volume. This trend seems to be irreversible, and other asset classes are also quickly becoming dominated by the machine traders. However, for asset that requires deeper understanding of physicality, like the trading of commodities, human traders still have significant edge over machines. The primary advantage of human traders in such market is the qualitative expert knowledge that requires traders to consider not just the financial information, but also a wide variety of physical constraints and information. However, due to rapid technology changes and the "invasion" of cash-rich hedge funds, even this traditionally human-centric asset class is crying for help in handling increasingly complicated and volatile environment. In this paper, we propose an adaptive trading support framework that allows us to quantify expert's knowledge to help human traders. Our method is based on a two-state switching Kalman filter, which updates its state estimation continuously with real-time information. We demonstrate the effectiveness of our approach in palm oil trading, which is becoming more and more complicated in recent years due to its new usage in producing biofuel. We show that the two-state switching Kalman filter tuned with expert domain knowledge can effectively reduce prediction errors when compared against traditional single-state econometric models. With a simple back test, we also demonstrate that even a slight decrease in the prediction errors can lead to significant improvement in the trading performance of a naive trading algorithm.

***Keywords*-autonomous trading, commodity trading, switching Kalman filter**

I. Introduction

The technology advancements in finance over the past decade have fueled significant growth in algorithmic trading (AT) across all asset classes and markets. According to the recent report from the Aite Group, AT has already accounted for more than 60% of all trading activities in equity, and for futures and options, the fraction of AT activities are set to climb over 50% in coming years. The US market is a clear leader in the adoption of AT technology, however, other major markets in the Europe and Asia are quickly catching up as well.

The AT technology can be applied in many different aspects of the financial trading, for example, it's already used widely in market-making, trade execution, arbitrage, and speculation. In all these applications, the AT implementations are made possible by increasingly accessible high-quality sources of real-time information. Initially only streams of price quotes on the traded entity were made available, now AT technology developers are able to utilize significant depth of the orderbook, price streams from other markets, other asset classes, and even automatically mining streamed news events.

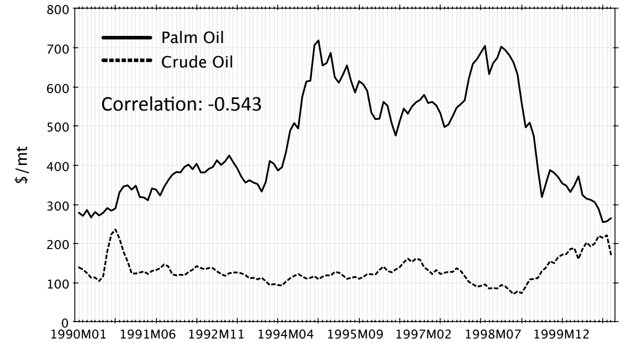
The primary strength of the AT technologies is certainly the speed, both in terms of data processing and execution, however, this does not necessarily mean that human traders will be pushed to extinction anytime soon. Human traders, even without lightening-fast speed advantage, still have an edge over AT bots in areas where qualitative reasoning plays more important role. A good example is in the trading of commodities, where the understanding of physical real-world considerations is critical.

The trading of commodities is special since most commodities have real-world uses (except a few exceptions such as precious metals) and thus the price levels are heavily affected by both the supplies and demands. Because of such physical connections, the prices of important commodities are usually quite stable in the long run when compared to other financial instruments (but can be quite volatile in the short run due to unexpected disruptions in either supplies or demands). Such physical connections use to be the primary reason why it takes years to train a competent trader who specializes in just one commodity (unlike equity or foreign exchange traders who can easily switch between targeted markets, it's rare for commodity traders to switch to a new commodity). Unfortunately, two recent trends have made even experienced commodity traders confused and seeking help at times. First, commodities are increasingly being considered as a promising asset case, and the share of pure speculators has grown significantly, thus making the once-stable commodity prices becoming more volatile. Second, technology advances have created new bondings between commodities or other asset classes that do not exist previously. For example, the use of crops in producing bio-fuel has created strong linkage between crude oil and agriculture commodities (e.g., sugar, corn, soybeans, palm oil). The fact that such relationship might depend on arbitrary numbers of external factors (e.g., the price levels of involved commodities or even weather) further complicates the analysis. It is our intention to propose and implement methodologies that would help human commodity traders making right decisions in face of complicated and constantly changing environment.

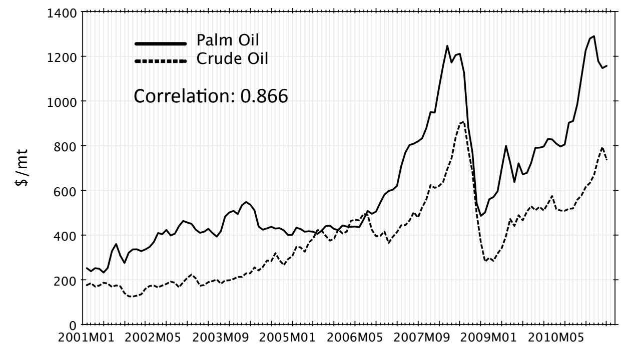
In this paper, we present what we believe is the first attempt in building a platform for supporting commodity trading operations that is capable of capturing expert knowledge on context switch. To concretely present our idea, we will investigate the substitution effects between crude oil and palm oil (one important agriculture commodity that can be used for producing bio-diesel) [1]. Such substitution effects could occur when the cost of using the primary commodity (crude oil) rises significantly higher then its comparable alternative, the secondary commodity (the bio-diesel produced from palm oil). Such a disparity could cause the secondary commodity price to have a provisional dependency with the primary commodity. As markets become increasingly connected and interdependent, the spilling effect of a single commodity price surge on its comparable substitute is increasingly more pronounced and frequent. In most cases, such effects are usually transient as the alternatives may not have the required volume to sustain the required demand. In our implementation, we identify and define two major market states (high and low) based on whether the crude oil price is high enough to trigger the substitution effects. For each of the two scenarios, a price prediction model is built, and we use the switching Kalman filter to detect the context and return a proper price prediction that takes into account how confident we are in being in one of the contexts. We demonstrate the effectiveness of our approach by using a simple back-testing trading system.

II. BACKGROUND AND MODEL SKETCHES

The price correlations between crude oil and major agriculture commodities used to be marginal, with crude oil prices contributing only to factors such as transportation or production. However, in the recent decade, crude oil prices are becoming more and more correlated to a number of agriculture commodities; in particular, the consistent spike in prices for both crude oil and agriculture commodities from 2007 to 2008 has lead experts and academics to rethink the probable impact of biofuel on such previously unseen correlations (for example, see Figure 1 for the palm oil and crude oil correlations in two different decades). After many debates and studies, it's becoming a consensus that the biofuel subsidies introduced at the beginning of 2000 in both the US and the EU (two most important oil consuming regions) to encourage the use of biofuel indeed contributed significantly to such new phenomenon (interested readers can refer to one of the earliest reports that spark the debate [2]).



(a) Monthly prices from 1990 to 2000.



(b) Monthly prices from 2001 to 2010.

Figure 1. The correlation of palm oil and crude oil prices in two different periods with distinctive patterns of correlations.

As discussed in the previous section, the reason why such new correlation emerges is the use of agriculture crops in producing biofuel. A number of crops have been used for producing biofuel around the world, and we choose palm oil in our research for the following reasons: 1) oil palm is perennial, thus making the supply of palm oil much more stable when compared to other annual plants (e.g., production of a particular crop can change significantly if farmers choose to plant other type of crop); 2) the global oil palm plantation is concentrated in only two countries, Malaysia and Indonesia, making it much easier for us to estimate global supply level.

The correlation between the crude oil price and the palm oil price is due to what commodity researchers called the *substitution effect*. The substitution effect occurs when a secondary commodity can be used in replacing primary commodity (in our case, the crude oil is the primary, while the palm oil is the secondary). The substitution is not always feasible, and its feasibility is determined by the utility ratio, U_p/U_s, in which U_p and U_s refer to the utilities of using the primary and the secondary commodities at their current price levels respectively. If the ratio is greater (less) than the equilibrium, users have little incentive to use the secondary (primary) commodity, thus the prices of the primary and the secondary commodities should be decoupled. However, when the ratio is sufficiently close to the equilibrium, the prices of both the primary and the secondary commodities will be closely correlated due to the substitution effect.

The consequence of having substitution effect is that there will be at least two different market structures, one for the close-to-equilibrium utility ratio, and the other one for the out-of-equilibrium utility ratio. What's complicated here is that the so-called equilibrium state is usually not defined on a single price point, but over a price range, and the changes of the market structure will not be discrete, but continuous and gradual over time instead.

Defined formally, the research question we would like to address can be answered in three phases. The first phase is to identify market states and derive models that can be used for decision support in that market state. The second phase is to infer the current market state by continuously receiving updates from external information sources (these information sources may include different price streams or even news events). The third phase is to generate predictions based on our beliefs on the market states.

For the first phase, we propose two market states: 1) close to equilibrium, and 2) out of equilibrium. The training dataset is preprocessed and divided into two chunks, fitting the description of these two market states. Respective econometric models are then created based on the training dataset. For the second phase, we construct a switching Kalman filter with two market states, and use it to infer the likelihood that we will be in any of the market state. Finally, by combining the beliefs and the prediction outputs from the two models, the final prediction will be delivered.

In the next section, we introduce related literature on palm oil pricing and switching Kalman filter.

III. Related Literature

A. Palm Oil Pricing Models

Commodity pricing model are largely based on econometric methods [3]. Recent works by [4] and [5], uses structural econometrics models and VAR models to analyze the structural relationship between palm oil and crude oil prices. Their focus were on the evaluation of possible reasons behind the increase in palm oil prices and the identification of palm oil demand drivers. The models developed were largely meant for short to long term trend predictions and the analysis of structural relationships among the identified price drivers.

In [5], the main objective of the study was to analyze the important factors contributing to the growth of Malaysian palm oil industry in particularly, biodiesel demand. Using a market model with annual data from 1976 to 2008, a system of eight structural equations considering, palm oil production, import/export, domestic demand, world supply were formulated for the analysis. Results from that study suggest that biodiesel demand does have a positive impact on the Malaysian palm oil domestic price. Thus, significant growth in biodiesel demand will affect palm oil price.

In [4], monthly price data of crude oil, palm oil, and soybean oil from January 1982 through February 2011 were used in their analysis. Results from this study indicates that crude oil prices are negatively correlated with palm oil prices in the long run and palm oil prices do not appear to respond to short-run fluctuation in crude oil prices. These results is reflective of a structural model using pricing dynamics dated from January 1982.

However, if we compare the monthly palm oil and crude oil prices in the periods from 1990 to 2000 (Figure 1(a)) and 2001 to 2010 (Figure 1(b)) respectively, we can observed a marked difference in the correlation between prices in the 2 periods. For the period prior to the year 2000, palm oil is weakly negatively correlated with crude oil but for the period after the year 2000, the prices of both commodity exhibited strong correlation. Such observations, could suggest that palm oil price and crude oil price dependency may be transient and triggered only when certain equilibrium conditions were disrupted.

In our study, we assume that such commodity substitution relations could induce transient effects in the pricing of the affected commodities. Hence, our goal in this paper is not to provide statistical evidence of structural relationship between crude and palm oil prices but to explore the feasibility of utilizing such substitution relations among commodity to develop a more robust pricing model for daily price forecasting.

B. Kalman Filter

The Kalman filter [6], [7] provides an efficient means to recursively estimate the state of a process by minimizing the mean of square error. The process is represented as a state-space model. The true value of the variable x_t is hidden and modeled as a Gaussian random variable. The random variable y_t is the noisy observation of x_t. Transition from one state in time t to the next time step $t+1$ is modeled as follows, where $w \in N[0, Q]$ and $v \in N[0, R]$ are the Gaussian noise in the x_t and y_t respectively.

$$x_t = A\, x_{t-1} + w_t, \tag{1}$$

$$y_t = H\, x_t + v_t. \tag{2}$$

With the observed values of y_t, the Kalman filter can compute the value of x_t, given the observation till time t. This is known as filtering. The true state of x_t is filtered using observations of y_1 till y_t. The Kalman filter can also perform prediction by computing the probability of x_{t+k} given observations up till y_t.

C. Switching Kalman Filters

Switching Kalman filter (SKF) [8], also known as Markov-switching model [9], has wide applications in many fields such as motion tracking [10], [11], voice recognition [12] and econometric time series analysis [9]. It allows the use of a weighted combination of more than one linear models or switching between the models across time. Hence, it is often applied in problems where the dynamics of the variable changes and cannot be represented by a single linear model. Through SKF, inference of the state of the variable at each point in time can be made and thus provides a piecewise linear prediction of the variable.

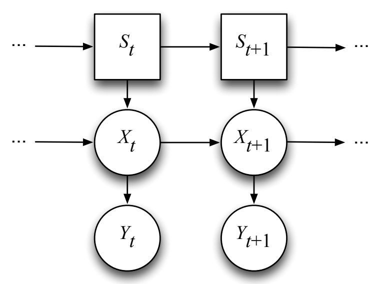


Figure 2. DBN representation of a switching Kalman filter [8].

A *dynamic Bayesian network* (DBN) representation of a SKF is shown in Figure 2. The values of the variable X and the values of the switch variable S are hidden or unknown.

A common technique used in estimating the switch variable S in each time step is the *generalized pseudo Bayesian algorithm* (GPB) [9]. It computes the hidden state value of in each state that are in $t-k$ timesteps. For each state, the probability that the $t-k$ values were generated by the respective state is also computed and is referred to as the weight.

The value of k determines the order of the GPB algorithm. For example, if $k=2$, x^{ij} is the expected value of x if the state at time $t-1$ is i and the state at t is j. Let V^{ij} be the covariance and W^{ij} be the probability that y_{t-1} was generated by state i and y_t was generated by state j.

The means and covariances are then "collapsed" two time steps ago as follows [8], [11]:

$$(x^i, V^j) = \text{Collapse}(x^{ij}, V^{ij}, W^{ij}),$$
$$x^j = \sum_i W^{ij} x^{ij},$$
$$V^j = \sum_i W^{ij} \left(V^{ij} + (x^{ij} - x^j)(x^{ij} - x^j)' \right).$$

IV. Adaptive Commodity Pricing

A. Equilibrium Utility Ratio

We first derive the equilibrium utility ratio of the commodities based on their substitution relationship. In this study, the utility is defined as the energy gained per gallon of fuel, in this case biodiesel versus gasoline. Let P_t and C_t be the per metric ton of palm oil and crude oil prices at time t respectively. The cost function of converting palm oil to biodiesel is estimated as Equation (3). Due to a lack of historical biodiesel price data, the equation is based on the production cost assessments in [13]. In this estimate, we assume that the feedstock, i.e. palm oil, is the only variable cost and all other cost are fix.

The cost function of converting crude oil to gasoline are estimated as Equation (4). This is estimated by regressing historical prices of gasoline and crude oil.

$$P_{\text{Biodiesel}}(t) = P_t + 261.65, \tag{3}$$
$$P_{\text{Gasoline}}(t) = 1.27C_t + 22.58. \tag{4}$$

Given that 0.9 gallon of biodiesel has the energy equivalence of 1 gallon of gasoline [14] and producers receive a tax credit of \$1.00 per gallon of biodiesel [15], i.e., \$358.29 per metric ton; the utility function for palm oil, U_P is defined as :

$$U_P = \frac{1}{0.9(P_{\text{Biodiesel}} - 358.29)}. \tag{5}$$

As there is no subsidy for gasoline, the utility function for crude oil, U_C is defined:

$$U_C = \frac{1}{P_{\text{Gasoline}}}. \tag{6}$$

When $U_p = U_c$, the following utility equilibrium should exist.

$$\frac{U_C}{U_P} = 1. \tag{7}$$

Using Equations (3), (4), and (7), the equilibrium utility ratio can be derived as follows:

$$E = 0.9\left(\frac{P_t + 261.65 - 358.29}{1.27C_t + 22.58}\right) = 1. \tag{8}$$

When $E < 1$, biodiesel would be a more cost effective alternative compared to gasoline and hence trigger the substitution effect. Conversely, when $E > 1$, the substitution effect will be diminished and decouples the price dependency between the primary commodities crude oil and palm oil.

B. Estimate State Processes

Using historical daily price data of palm oil and crude oil from January 2006 till December 2007, we derive the log return, i.e, $r_t = ln(S_t/S_{t-1})$ for each price series. This is to transform the price series to a weakly stationary time series. Let r_p and r_c be the daily log returns of palm oil and crude oil respectively.

State 1: ARMA(2,2) model where returns are dependent only on the individual commodity.

$$r_p(t) = \alpha_1 r_p(t-1) + \alpha_2 r_p(t-2) + \alpha_3 \epsilon_{t-1} + \alpha_4 \epsilon_{t-2} + \epsilon_t$$
$$r_c(t) = \beta_1 r_c(t-1) + \beta_2 r_c(t-2) + \beta_3 \epsilon_{t-1} + \beta_4 \epsilon_{t-2} + \epsilon_t$$

State 2: VAR(2) model where returns are dependent on both palm oil and crude oil prices.

$$\begin{aligned} r_p(t) =& \gamma_1 r_p(t-1) + \gamma_2 r_p(t-2) + \gamma_3 r_c(t-1) + \\ & \gamma_4 r_c(t-2) + \epsilon_t \\ r_c(t) =& \theta_1 r_p(t-1) + \theta_2 r_p(t-2) + \theta_3 r_c(t-1) + \\ & \theta_4 r_c(t-2) + \epsilon_t \end{aligned}$$

C. Prediction

With the estimated models for each state, the time series models are first mapped to the following state-space models in Equations (1) and (2) using the Hamilton form [16]. $\hat{r}_p(t)$ and $\hat{r}_c(t)$ refers to the estimated values, while $r_p(t)$ and $r_c(t)$ are the observed values.

$$x_t = \begin{bmatrix} \hat{r}_p(t) \\ \hat{r}_c(t) \\ \hat{r}_p(t-1) \\ \hat{r}_c(t-1) \\ \hat{r}_p(t-2) \\ \hat{r}_c(t-2) \end{bmatrix} \qquad y_t = \begin{bmatrix} r_p(t) \\ r_c(t) \\ r_p(t-1) \\ r_c(t-1) \\ r_p(t-2) \\ r_c(t-2) \end{bmatrix}$$

State 1: ARMA (2,2) model:

$$A_1 = \begin{bmatrix} \alpha_1 & 0 & \alpha_2 & 0 & 0 & 0 \\ 0 & \beta_1 & 1 & 0 & \beta_2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$H_1 = \begin{bmatrix} 1 & 0 & \alpha_3 & 0 & \alpha_4 & 0 \\ 0 & 1 & 0 & \beta_3 & 0 & \beta_4 \end{bmatrix}$$

State 2: VAR(2) model:

$$A_2 = \begin{bmatrix} \gamma_1 & \gamma_3 & \gamma_2 & \gamma_4 & 0 & 0 \\ \theta_1 & \theta_3 & \theta_2 & \theta_4 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

With the estimated state models, the covariance matrix, Q for each state is derived using past data and it takes the following form, where $var(P_i)$, $var(C_i)$ are the variance measured using model i of palm oil and crude oil prices respectively. The R matrix in both states are null matrices since the observations in this case are the actual price values that we need to predict. Thus, there are no observation noise introduced in the models.

$$Q_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} var(P_i) \\ var(C_i) \\ var(P_i) \\ var(C_i) \\ var(P_i) \\ var(C_i) \end{bmatrix}$$

As the states are hidden, the prior π and transitions matrix Z are assumed to be uniform and initialized as:

$$\pi = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \qquad Z = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

The probabilities of each state at time t are computed using the SKF. The prediction for $r_p(t+1)$ and $r_c(t+1)$ are then made with using the state model with the highest probability.

D. Adaptive Transition Matrix

In this implementation, we attempt to infuse knowledge of the substitution relationship into the SKF through the use of an adaptive transition matrix. We assume that the consumption agents have a natural preference for states that maximize their utility.

Hence, when the equilibrium ratio, $E < 1$, the agents have an increased propensity to switch from the primary commodity to the secondary commodity. In this case, the probability of switching to state 2 would increase. Using the logistic function, we model this propensity, ρ as follows:

$$\rho = 1/(1 + e^{ln(E)})$$

With the defined propensity measure, we derive the following adaptive transition matrix that is used when $E < 1$.

$$Z = \begin{bmatrix} 1-\rho & \rho \\ 1-\rho & \rho \end{bmatrix}$$

When $E >= 1$, the transition matrix remains as a uniform matrix.

V. Results and Discussion

A. State Parameters

Using historical daily prices from January, 2006 till December, 2007, the following parameters were estimated using ordinary least square regression. We specifically use data from 2006 to estimate the VAR(2) model and data from 2007 to estimate the ARMA(2,2) models. This is because, as shown in Figure 1(b), the crude oil and palm oil prices were converging in 2006 and diverging in 2007. By using the respective periods, we can capture a better estimation of the models.

State 1 parameters:

$$\begin{aligned} \alpha_1 &= -0.349194 & \beta_1 &= 1.157400 \\ \alpha_2 &= 0.531266 & \beta_2 &= -0.925547 \\ \alpha_3 &= 0.590707 & \beta_3 &= -1.159660 \\ \alpha_4 &= -0.307434 & \beta_4 &= 0.991747 \end{aligned}$$

$$var(P_1) = 0.000182601 \qquad var(C_1) = 0.00028459$$

State 2 parameters:

$$\begin{aligned} \gamma_1 &= 0.253540 & \theta_1 &= -0.208392 \\ \gamma_2 &= 0.035551 & \theta_2 &= 0.028889 \\ \gamma_3 &= 0.079044 & \theta_3 &= 0.300596 \\ \gamma_4 &= -0.041150 & \theta_4 &= -0.057395 \end{aligned}$$

$$var(P_2) = 0.000104 \qquad var(C_2) = 0.000293$$

B. Experimental Results

Using the state models derived in the previous sections, we ran an 1-step forecast for palm oil daily prices from January 1, 2008 till December 31, 2011 with the following models:

- ARMA(2,2): Autoregressive moving average model (for State 1).
- VAR(2): Multi-variate autoregressive model (for State 2).
- Uniform SKF: SKF with uniform transition matrix
- Adaptive SKF: SKF with adaptive transition matrix

For each of the models we compute the sum of squared error at each time step t till the end of the forecast period (December 31, 2011). As shown in Figure 3, the SKF models predominately yield the lowest sum of squared errors when compared with the single state autoregressive models (i.e., ARMA(2,2) and VAR(2)). This shows that the SKF based models are more robust then the single state models. Table I compares the results from the uniform and adaptive SKF models. It can be seen that the adaptive SKF provides a slightly better forecast then the uniform SKF.

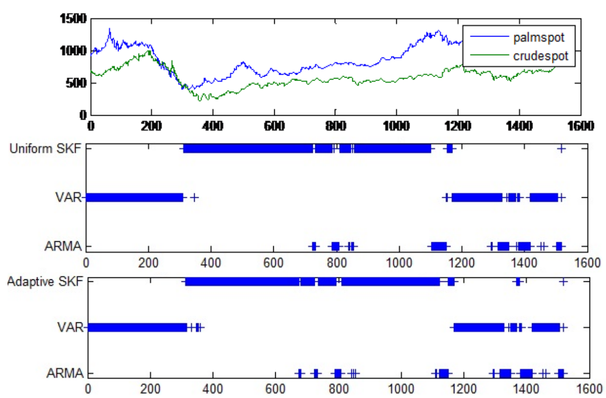


Figure 3. Top chart: Daily spot price of palm oil and crude oil from January 1, 2008 to December 31, 2011. Bottom two charts: Indicate which model has the least MSE in each time period (the SKF in middle and bottom charts refer to uniform and adaptive SKF respectively).

Table I
UNIFORM SKF VS ADAPTIVE SKF

SKF Type	Root MSE	Winning Percentage[1]
Uniform	$3.7258 \cdot 10^{-4}$	50.2
Adaptive	$3.7184 \cdot 10^{-4}$	52.4

[1] Fraction of cases with lowest MSE.

C. Simple Back-Testing Results

In previous analysis, we already see that our proposed SKF models indeed deliver better prediction performance in terms of MSE. However, it's not clear how much real-world impact such difference would generate if the proposed model is indeed implemented in a trading system. To quantify the potential impact of a slightly better MSE, we setup a back-testing environment to observe the actual profits and losses for competing methods using historical palm oil and crude price sequences from January 1, 2008 to Feburary 27, 2012. The back-testing procedure is a commonly used method in validating a proposed trading model by exposing it to historical price streams, assuming that suggested trades from the tested model do not affect the historical price streams.

Since our interest is in comparing different models for commodity traders, we need to define a common trader module that utilizes outputs from different models. For simplicity, we define a rather simple trading algorithm as follows:

1) Suppose we are currently in time t, use the specific prediction model to generate a prediction for time $t+1$.
2) If forecasted return is positive (using price in time t as basis), go long; otherwise, go short.

The profits and losses (PnL) of the same trading algorithm equipped with different prediction model are quantified in Table II. The advantage of both SKF models over single-mode models is significant; the adaptive SKF does have slight advantage over the uniform SKF, however, the difference is not significant (the difference is only $\%0.5$). Do note that the result reported in Table II is the snapshot at the end of the back test. To understand how the PnL evolve over the whole horizon, we can plot the PnL results over all time periods. Since the adaptive SKF works best when there are frequent changes in modes, we should observe the most difference in performance when the market is highly volatile. For both crude oil and palm oil, the markets are the most volatile in the second half of 2008, at the height of the recent financial crisis. To illustrate how adaptive SKF model can cope with such highly volatile market condition, we plot all four PnL curves on the same group from August to December, 2008 in Figure 4.

Table II
PROFITS AND LOSSES FOR DIFFERENT MODELS.

ARMA	VAR	Uniform SKF	Adaptive SKF
3225.76	8675.54	11880.28	12312.94

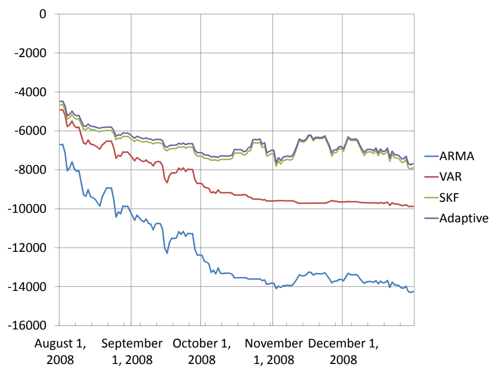


Figure 4. Profits and losses from August to December, 2008.

VI. Conclusions and Future Work

In this paper we have illustrated how an expert's qualitative insights can be incorporated quantitatively. In particular, we show that an effective decision support system can be built for palm oil traders. The built system is special in that it incorporates exact information that an expert would monitor, and the prediction module automatically updates itself to provide a hybrid prediction based on the beliefs on which state the world is in (for palm oil trading, there are two states, indicating whether bio-fuel is economically viable).

Due to the nature of the commodity and the limitation in available information, our state space is made to be very simple (containing only two states). If we have necessary information to expand the state space, the edge of our approach might be even greater. However, we might quickly run into the problem of exploding state space and we might also face the problem of not having enough data to train all derived models, whose number would grow combinatorially in number of state variables. Such practical considerations are not currently considered, and we look to develop our methodology in this direction.

References

[1] I. Tareen, J. Duffield, and M. Wetzstein, "Biodiesel as a substitute for petroleum," *Annual Meeting of the American Agricultural Economics Association*, 1999.

[2] D. Mitchell, "A note on rising food prices," *World Bank Policy Researh Working Paper No 4682*, July 2008.

[3] P. Reissa and F. Wolak, *Structural Econometric Modeling Rationales and Examples From Industrial Organization in Handbook of Econometrics*. Elsevier B.V., 2007.

[4] D. Sanders, J. Balagtas, and G. Gruere, "Revisiting the palm oil boom in southeast Asia," *IFPRI Discussion Paper 01167*, March 2012.

[5] S. Applanaidu and F. Arshad, "An econometric analysis of the link between biodiesel demand and malaysian palm oil market," *International Journal of Business and Management*, 2011.

[6] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, 2010.

[7] G. Welch and G. Bishop, "An introduction to Kalman filter," *University of North Carolina at Chapel Hill, Chapel Hill, NC*, vol. 7, no. 1, 1995.

[8] K. Murphy, "Switching Kalman filters," Tech. Rep., 1998.

[9] C.-J. Kim, "Dynamic linear models with Markov-switching," *Journal of Econometrics*, vol. 60, no. 1–2, pp. 1–22, 1994.

[10] H. Veeraraghavan, P. Schrater, and N. Papanikolopoulos, "Switching Kalman filters-based approach for tracking and event detection at traffic intersections," *Mediterranean Conference on Control and Automation*, 2005.

[11] V. Manfredi, S. Mahadevan, and J. Kurose, "Switching Kalman filters for prediction and tracking in an adaptive meteorological sensing network," *IEEE SECON*, 2005.

[12] M. Fujimoto and K. Ishizuka, "Noise robust voice activity detection based on switching Kalman filter," *INTERSPEECH-2007, 2933-2936*, 2007.

[13] B. B. Canada, "Biofuel costs, technologies and economics in APEC economies," *APEC Project EWG 16/2009*.

[14] Wikipedia, "Gasoline gallon equivalent," http://en.wikipedia.org/wiki/Gasoline_gallon_equivalent.

[15] J. Schumacher, "Oilseed, biodiesel and ethanol subsidies & renewable energy mandates: US federal & selected state initiatives," *Agricultural Marketing Policy Paper No. 16*.

[16] R. Shumway and D. Stoffer, *State-Space Models From Time Series Analysis and Its Applications*. Springer, 2010.