Ramex-Forum: Sequential Patterns of Prices in the Petroleum Production Chain

Pedro Tiple[1], Luís Cavique[2], and Nuno C. Marques[3]

[1] GoBusiness Finance
[2] Universidade Aberta
[3] NOVA Laboratory for Computer Science and Informatics, DI-FCT,
Universidade Nova de Lisboa, Portugal
nmm@fct.unl.pt

Abstract. We present a sensibility analysis and new visualizations using an improved version of the Ramex-Forum algorithm applied to the study of the petroleum production chain. Different combinations of parameters and new ways to visualize data will be used. Results will highlight the importance of Ramex-Forum and its proper parameterizations for analyzing relevant relations among price variations in petroleum and other similar markets.

Keywords: Ramex-Forum, financial data analysis, petroleum price, petroleum production chain, business intelligence

1 Introduction

Petroleum is one of the most important resources to the developed world and still is a major variable influencing the Economy and markets. The price of petroleum and its derivatives isn't influenced simply by supply and demand; taxes, speculation, wars, costs in refinement and transportation all contribute in setting prices. Due to its lengthy refinement process, a significant increase in the price of the source material can only reflect in the price of its derivatives after the time it takes to refine (usually within 3-4 weeks [BSoER96]). Moreover, due to its high economic importance and cost, the price of crude oil should always reflect on the final price [Suv09].

This work presents a study on a method to quantify how the price of the crude oil (raw material) can influence the price of manufactured products by using Ramex-Forum. This paper departs from the work of [Tip14], with the original Ramex-forum proposal [MC13]. It analysis how this proposal can be improved and then tunned for finding sequential patterns using the prices of petroleum and derivatives. Section 2 presents the basic method and introduces the main concepts and section 3 presents an evaluation on how the price of derivatives are influenced by the price of the crude oil (the source material). Finally, some conclusions are presented.

2 Counting co-Occurrences in Financial Markets

We assume a crossover strategy to buy and sell financial products. Given the product price index in time 't', denoted I(t), and the moving average of that price index with length of N_{MA} days, calculated by: $MA(t, N_{MA}) = \sum I(t-w)/N_{MA}, \forall w \in \{0 : N_{MA}\}$, the decision is as follows:

– Buy, if $I(t) \cdot (1 + \epsilon) \geq MA(t, N_{MA})$
– Sell, if $I(t) \cdot (1 - \epsilon) \leq MA(t, N_{MA})$

For each moment t, if there is a decision of either Buy or Sell, respective counters ($Counter_B$, $Counter_S$) will be incremented by one unit. If neither of those decisions is made, both counters are reset. This way each counter has the number of consecutive moments where the same decision is made. See example in the Figure 1, the 'B' (Buy) an 'S' (Sell) char illustrate the crossover strategy for a large enough ϵ. In this paper (except when explicitly mentioned otherwise), we will use a standard 1% error, i.e. $\epsilon = 0.01$. Other parameter is also used when defining an influence: parameter δ is the maximum trading period length where a check for relations between two assets is made. Finally we can define $\#Influence(A, B, \delta)$: a cumulative influence counter of a given Buy or Sell decision for a market signal A to a market signal B (denoted $A \rightarrow B$), counts how many times $0 < (|Counter_A| - |Counter_B|) \leq \delta \wedge Counter_B \neq 0$.
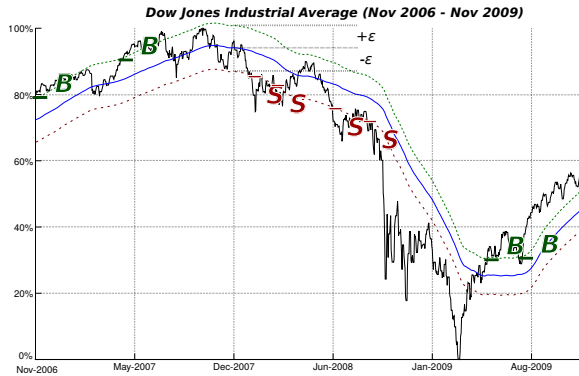

Fig. 1. Financial product (normalized DJI index in black) and respective moving average (blue) and crossover starting *Buy* (green) and *Sell* (red) decision with ϵ confidence.

3 Results on the Petroleum Production Chain

Petroleum is refined into a relatively extensive list [GH01], with each category having hundreds of sub-products. Moreover this division and classification mostly depends on its social usage. This study is based on the repository of publicly

available historical values for a wide range of petroleum related products provided by the U.S. Energy Information Administration[1]. The variations in the prices of these products are also compared with the stock value of the the stock market value of eleven corporations dedicated to extracting, processing, and selling of crude and crude related products. The prices of the 55 products are are separated into retail/bulk price and spot price (for some items the price is taken from retail sellers and for other items it's the security price at that day). The data separated into four categories of known benchmarks [HET08] for crude oil (West Texas Intermediate as OklahomaWTI, European Brent, and the OPEC Basket); Refinery price for Gasoline, RBOB Gasoline, Diesel, KeroseneJet, Propane, and Heating Oil; National, state, and city averages for regular gasoline and diesel; Corporation stock values.

This paper studies the influence and best values for parameters δ, ϵ thresholds, and moving average size. Focus will be put on the *Buy* comparison because in the selected data the increase/decrease of prices is very asymmetrical with a strong lean towards increases. For better parameter comparison an additional measure is used in our results, the average edge weight: a relation between the weight sum of output edges divided by the number of edges in the graph: $AverageEdgeWeight(V,E) = \sum weight(e)/|E|, \forall e \in E$.

Parameter δ was analyzed regarding its effect on the average edge weight changes. The result can be seen in Figure 2A. The chart shows the average edge weight change for each increment in the value of δ. Each line represents the results obtained using different moving average sizes. Several big spikes can be seen every 5 days, this is because gas and diesel prices at the pump are only registered on a weekly basis, so for each 5 day increase in δ the algorithm will pick up another change in value. This makes the analysis somewhat harder but it's still useful as now changes in retail prices are clearly identified. The first thing noted is that at the first week there is already a noticeable increase in the average edge weight, however some of it is due to influences between retail prices and not only from refinery to retail prices. Second, after the fourth week the individual increases in δ barely produce a meaningful increase in value, still the cumulative increases are significant.The parameter δ was fixed at a value of 30 working days (around six weeks: two weeks more than the expected).

Parameter ϵ was studied by trying to find the best combination of parameters. The algorithm was ran several times and the average edge weight value was recorded for each run. The best values for threshold interval and the moving average size are represented in Figure 2B. The graph shows the progression of the average edge weight, in relation to increase in threshold size. The parameters that lead to the highest increase in average weight can be clearly identified as the moving average size of 240 days with a threshold of around 26% of the moving average. However things change when the influence event count is also

---

[1] The used data was downloaded from `http://www.eia.gov/` on June 2014 and ranges form January 2006 to June 2014.

considered: increasing the threshold rapidly decreases the number of detected events (a threshold of 26% will reduce the number of events by about 80%). In this case the starting average is around 130 events and falls to 30, in the 3 years period analyzed: a very low average number of events. For this case study the choice was made to maximize the event count so that a broader spectrum of influences can be detected instead of restricting the analysis to situations where the prices rise or fall sharply (which is what higher threshold values restrict the analysis to). Usually, small increases in the ϵ threshold will raise the average weight while only lowering the event count by small amounts, so going for a threshold of 0% is not the best choice. Nevertheless this trade-off favors the usage of smaller values for ϵ.
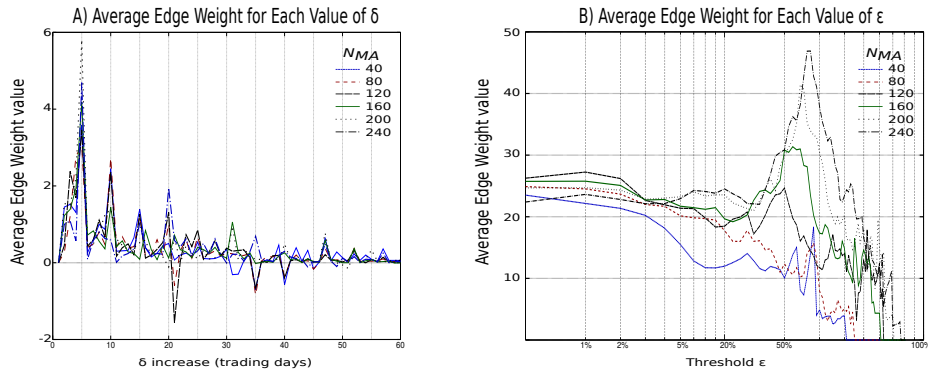

Fig. 2. Graphs showing the change in: (A) average edge weight with each increment of δ using the *Buy* comparison; (B) average edge weight and number of nodes with each $\epsilon_{t+1} = \epsilon_t + 1\%$ increment in the threshold interval for $\delta = 30$ using the *Buy* comparison.

Moving Average Size, N_{MA} The choices for available moving average sizes were based on [MC13] and the graphs show that maximizing this parameter yields the best results and even raises the question of how further increases in the size would fare. The user still needs to take into account of what it means to increase the moving average size, the bigger it is the moving average, the smoother the curve will be and thus it will behave like a noise filter, becoming less and less sensitive to small changes in the behavior of the product. The values overlap for small ϵ values and it is hard to read the effects of the first increments in a linear scale, so Figure 2B uses a logarithmic scale for representing ϵ values, showing that the 240 and 120 moving average sizes have a very similar behavior. The average weight for a moving average of 120 days has a higher starting value than the 240 days one, this means that for a buy signal best thresholds are: $\epsilon = 1\% \wedge \delta = 30 \wedge N_{MA} = 120$. Figure 3 shows that it is possible to find more than just sequential patterns with this paramethers.
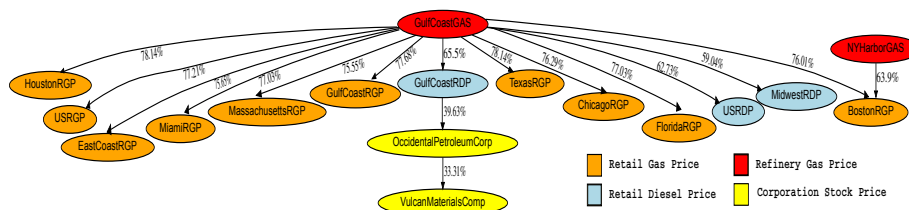
Fig. 3. Part of the graph showing the resulting Buy tree after applying Ramex Forum on the data with the selected parameters.

In the complete result graph (available in [Tip14]) colors were added to each node according to their product type. These colors show a clear grouping of product types, with same color nodes mostly close to each other. This was expected for gas to gas and diesel to diesel influences however even the stock, refinery, and reference benchmark prices tend to group together at least in pairs. Furthermore, refineries are almost exclusively related to the same type of product, gas producing refineries are connected to retail gas prices and diesel producing refineries are connected to diesel retail prices. The Gulf Coast GAS refinery node does not exactly meet the previous observation as it is shown influencing some diesel products, even so, this might be a positive thing as it will alert an attentive analyst to the weight behind the Gulf Coast refinery gas prices. After further analysis Gulf Coast GAS is identified as the most influential node as it has at least one detected event for all other products and its average edge weight is the highest by a margin of 5%, probably due to huge oil production in this area it is mostly the start of oil production chain. It is also interesting to observe that Rocky Mountain retail gas price appears with third level dependences, that could be due to specificities of gas usage in the Rocky Mountain area. Next the most glaring aspect of the graph is how influential specific products are, the tree is not just an assorted web of relations but groups of products aggregating around very influential/influenced products. There are some expected trend setters like the OPECBasket that is used as a benchmark for oil price, the Gulf Coast refineries and then some unexpected like the Minnesota retail gas price. For the other graphs and simulations, equal graphs were done with the color coding and the results were very similar with strong groupings of colors and some few select products influencing groups of others.

4   Conclusions

The presented case study, using real world data and deep analysis aims to provide an illustrative and useful example of Ramex-Forum: the signal-to-noise ratio on the Petroleum production chain analysis already shows that sequential patterns of prices can provide a much deeper description of product dependencies based on events. Moreover the δ and ϵ parameters seem both consistent, intuitive and adaptable alternative for measuring long term dependencies that are not directly

possible with more instantaneous methods. So far only the connections them-
selves have been considered, if the influence weights are also taken into account
the analysis becomes more complex. Future work in this area should extend the
number of products used and their detail. Some related studies [MMC14] show
that less classical hybrid approaches can be used to complement the crossover
event detection approach and are good candidates for future experiments. Of par-
ticular interest will be to include a better characterization of algorithm behavior
during a global market crisis, namely by quantifying the drives and consequences
of the recent crisis in oil prices.

5   Acknowledgements

References

[BSoER96] S. Borenstein, A. Shepard, and National Bureau of Economic Research. *Sticky Prices, Inventories, and Market Power in Wholesale Gasoline Markets*. Number no. 5468 in NBER working paper series. National Bureau of Economic Research, 1996.

[GH01]   J.H. Gary and G.E. Handwerk. *Petroleum Refining*. Institut français du pétrole publications. Taylor & Francis, 2001.

[HET08]   Shawkat Hammoudeh, Bradley T. Ewing, and Mark A. Thompson. Threshold cointegration analysis of crude oil benchmarks. *The Energy Journal*, Volume 29(Number 4):79–96, 2008.

[MC13]   Nuno C. Marques and Luís Cavique. Sequential pattern mining of price interactions. In *Advances in Artificial Intelligence – Proceedings of the Workshop Knowledge Discovery and Business Intelligence, EPIA-KDBI, Portuguese Conference on Artificial Intelligence*, pages 314–325, 2013.

[MMC14]   Diogo Matos, Nuno Marques, and Margarida Cardoso. Stock market series analysis using self-organizing maps. *Revista de Ciências da Computação*, 9(9):79–90, 2014.

[Suv09]   Hannu Suviolahti. The influence of volatile raw material prices on inventory valuation and product costing. Master Thesis, Department of Business Technology, Helsinki School of Economics, 2009.

[Tip14]   P. Tiple. Tool for discovering sequential patterns in financial markets. Master Thesis in Engenharia Informática, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, 2014.