Modeling VaR in Crude Oil Market: A Multi Scale Nonlinear Ensemble Approach Incorporating Wavelet Analysis and ANN

Kin Keung Lai^{1,2}, Kaijian He^{1,2}, and Jerome Yen³

¹ Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue,
Hong Kong

mskklai@cityu.edu.hk

² School of Business Administration, Hunan University, Changsha, Hunan, 410082, China
paulhekj@gmail.com

³ Department of Finance, Hong Kong University of Science and Technology, Hong Kong
risksolution@gmail.com

Abstract. Price fluctuations in the crude oil markets worldwide have attracted significant attentions from both, industries and academics, due to their profound impact on businesses and governments. Proper measurement and management of risks due to unexpected price movements in the markets has been crucial from both, operational and strategic perspectives. However, risk measurements from current approaches offer insufficient explanatory power and performance due to the complicated non-linear nature of risk evolutions. This paper adopts a VaR approach to measure risks and proposes multi-scale non-linear ensemble approaches to model the risk evolutions in WTI crude oil market. The proposed WDNEVaR follows a semi-parametric paradigm, incorporating both, wavelet analysis and artificial neural network techniques. Experiment results from empirical studies suggest that the proposed WDNEVaR is superior to traditional approaches. It provides VaR estimates of higher reliability and accuracy. It also brings significantly more flexibility during the modeling attempts.

Keywords: Value at Risk, Wavelet Analysis, Nonlinear Ensemble Algorithm, Artificial Neural Network.

1 Introduction

Crude oil is a globally important industrial input. On one hand, crude oil price fluctuations affect every sector in regional and global economies. On the other hand, crude oil price forecasting and the allied risk measurement serve as important inputs and form the basis for macro-econometric models, financial markets' risk assessments and derivatives' pricing, etc. Therefore, demand for risk management in the crude oil market is on the rise from both, academics and industries.

Among numerous risk measures, this paper adopts Value at Risk to measure market risks in the crude oil markets. Due to its simplicity, VaR is widely accepted as a risk measure by both, academics and industries [1]. Although different approaches

have been developed to estimate VaRs, the essential multi-scale heterogeneous data structures remain unexplored. Therefore, this paper takes a semi-parametric approach and proposes the novel wavelet based non-linear ensemble approach. Empirical studies are conducted in the representative US West Intermediate Taxes (WTI) crude oil market. The experiment results suggest that the proposed WDNEVaR model outperforms the more traditional ARMA-GARCH based model.

The rest of the paper develops as follows: the second section reviews the relevant theories. The third section lays down the algorithm for wavelet based non-linear ensemble approach for VaR estimates. The fourth section presents and interprets results of empirical studies in WTI crude oil market. The fifth section concludes.

2 Background and Theories

VaR is a single number describing the maximal potential loss of a portfolio of financial instruments during a given period of time for a given probability. Estimation of appropriate VaR is a difficult challenge that has been tackled mainly from three approaches over the years: parametric, non-parametric, and semi-parametric approach [1]. The semi-parametric approach is the emerging methodology that strikes the balance between the parametric and non parametric approaches. These include many inter-disciplinary techniques such as wavelet analysis, extreme value theory, etc [2].

To verify the modeling adequacy, the model's past performance are often taken as the indication of the future performance [1]. Kupiec frequency based backtesting procedure is the most dominant hypothesis based approach, which takes the consistency level of observed frequency of losses as indication of the level of model adequacy. When the sample size is sufficiently large, the Kupiec backtesting procedures would offer an acceptable level of discriminatory power between good and bad models. Therefore, it is adopted in this paper to backtest and evaluate different models.

As the emerging new tool for estimating VaRs, wavelet analysis is the latest development in the harmonic analysis field. It possesses the capability to project data into time-scale domain for analysis [2]. Mathematically, wavelets are continuous functions that satisfy admissibility conditions as in

$$C_\psi = \int_0^\infty \frac{|\Psi(f)|}{f} df < \infty \Leftrightarrow \int_{-\infty}^\infty \psi(t) dt = 0 \text{ and unit energy condition as in } \int_{-\infty}^\infty |\psi(t)|^2 dt = 1,$$

where φ is the Fourier transform of ψ .

There are different families of wavelets designed, each with their own special characteristics [2]. Typical wavelet families include the Haar wavelet, the Daubechies wavelets, and the coiflets wavelets, etc[2].

The wavelet transform is conducted as $W(u, s) = \int_{-\infty}^\infty x(t) \psi_{u,s}(t) dt$ while the inverse (synthesis) operation is performed as in $x(t) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^\infty W(u, s) \psi_{u,s}(t) du \frac{ds}{s^2}$. Thus together

they form the basis for multi-resolution analysis as in $f(t) \approx S_J(t) + D_J(t) + D_{J-1}(t) + \dots + D_1(t)$, where $S_J(t)$ refers to the smooth signals and equals $\sum_k s_{j,k} \phi_{j,k}(t)$. $D_J(t)$ refers to the detail signals and equals $\sum_k d_{j,k} \psi_{j,k}(t)$.

Since Ramsey (1999) introduced wavelet analysis into the mainstream literature, wavelet analysis has been applied in three different areas: firstly, wavelet analysis has been applied for economic and financial relationship identification to reveal the non-linear multi-scale data structure[3]. Secondly, wavelet analysis has been used extensively for denoising financial data, which extracts the maximal level of useful information [4]. Thirdly wavelet analysis has been used extensively in the forecasting model building process to decompose the original data series for further modeling attempts[5]. However, wavelet analysis has received little attention in the risk management field. Fernandez (2006a, 2006b) has used wavelet analysis to investigate the risk distribution across time scales in the Chilean stock markets and found concentration of risk in high frequency parts of data[6, 7]. But, these approaches mainly focus on investigating the distribution of potential market losses embedded in VaR numbers, across the time horizon. Their approach is based on the assumption that wavelet decomposed variances at different scales represent investors' preferences. However, they seem to have ignored the impact of different wavelet families chosen during their analysis, which leaves their findings largely inconclusive.

Another methodology proposed for estimating VaR is the ensemble approach. The Ensemble approach pools the partial information sets employed by different individual forecasters into a unified one as in $\hat{y}_t = \sum_{i=1}^n w_t \hat{x}_{i,t}$, where $\hat{y}_t$ is the

output at time t. w_t is the weight assigned to different forecasts. $\hat{x}_{i,t}$ is the forecast at time t. n is the number of forecasts. Among different techniques introduced to optimize weights, artificial neural network is the most popular one. It is an important connectionist model for estimating non-linear data patterns and has distinguished itself as a universal function approximator for empirical studies. Artificial neural network consists of nodes and connections between the nodes. Neurons are grouped in three different types of layers: input layers, hidden layers and output layers. Depending on the direction of information flow, the neural network structure can be grouped into feed-forward and recurrent types. The network structure takes the matrix form as in (1):

$$y_i = a_0 + \sum_{j=1}^q w_{ij} f(a_j + \sum_{i=1}^p w_{ij} x_i) + \varepsilon_i \quad (1)$$

where a_j refers to the bias on the j^{th} unit, w refers to the connection weight between different layers, f(.) refers to the transfer function of the hidden layer, p is the total number of input nodes while q is the total number of hidden nodes.

3 Wavelet Decomposed Nonlinear Ensemble Value at Risk (WDNEVaR)

When the sample is of sufficient length, VaR can be estimated by the parametric approach as in (2):

$$VaR_t = \mu_t + \sigma_t G^{-1}(\alpha) \quad (2)$$

where $G^{-1}(\alpha)$ refers to the inverse of the cumulative normal distribution.

The estimation process boils down to estimation of conditional mean and conditional standard deviation, which involves three steps:

Firstly, the original data series are decomposed using wavelet analysis based on the chosen wavelet families as in (3):

$$r_t = r_{A^j,t} + \sum_{j=1}^J r_{D^j,t} \quad (3)$$

Secondly, the conditional mean at each scale is assumed to follow ARMA processes and it is forecasted at each individual scales as in (4).

$$\hat{\mu}_{t,s} = a_{0,s} + \sum_{i=1}^r a_{i,s} r_{t-i,s} + \sum_{j=1}^m b_{j,s} \varepsilon_{t-j,s} \quad (4)$$

where $\hat{\mu}_{t,s}$ is the conditional mean at time t , s is the wavelet scales. $r_{t-i,s}$ ($i = 1 \dots r$) is the lag r returns with parameter $a_{i,s}$, and $\varepsilon_{t-j,s}$ ($j = 1 \dots m$) is the lag m residuals in the previous period with parameter $b_{j,s}$.

The conditional standard deviation at each scale is assumed to follow GARCH processes and it is estimated at each individual scales as in (5):

$$\hat{\sigma}_{t,s}^2 = \omega_s + \sum_{i=1}^p a_{i,s} \varepsilon_{t-i,s}^2 + \sum_{j=1}^q \beta_{j,s} \sigma_{t-j,s}^2 \quad (5)$$

where $\hat{\sigma}_{t,s}^2$ is the conditional variance at time t , $\sigma_{t-j,s}^2$ ($j = 1 \dots q$) is the lag q variance with parameter $a_{i,s}$, and $\varepsilon_{t-i,s}^2$ ($i = 1 \dots p$) is the lag p squared errors with parameter $\beta_{j,s}$ in the previous period, during which the mean return is set to be zero.

Thirdly, as wavelet transform decorrelates the decomposed data, the conditional mean can be aggregated from the conditional means estimated at each individual scales as in (6).

$$\hat{\mu}_t = \hat{\mu}_{A^j,t} + \sum_{j=1}^J \hat{\mu}_{D^j,t} \quad (6)$$

The conditional standard deviations at each scale are synthesized, using wavelet analysis, to produce the aggregated estimates as in (7). The synthesis process pools the partial information set at each scale together and retains its authenticity based on the “preservation of energy” property of wavelet analysis.

$$\hat{\sigma}_t^2 = \hat{\sigma}_{A^j,t}^2 + \sum_{j=1}^J \hat{\sigma}_{D^j,t}^2 \quad (7)$$

In the next step, since previous VaRs are estimated based on certain wavelet families, to reduce the biases introduced by arbitrary choice of wavelet families, the non-linear ensemble algorithm is proposed, to combine VaRs estimated based on different wavelet families into the optimal one utilizing artificial neural network, as in (8).

$$\sigma_{WDNEVaR} = \sum_{i \in F} Weight_i * \sigma_i \tag{8}$$

where F refers to the set of wavelet families attempted. Weight refers to the weight of the standard deviation estimated using different wavelet families.

When training artificial neural network to ensemble different volatility forecasts, a historical volatility measure is needed to provide an objective during the training process. There are mainly three dominant volatility measures identified in the literature: squared returns as $Volatility_{squaredReturns} = \sqrt{r_t^2}$, realized volatilities as

$RV = \sqrt{\sum_{t=1}^n r_t^2}$ and historical volatilities $HV = \sqrt{\frac{1}{N} \sum_{t=1}^N (r_t - \bar{r})^2}$.

4 Empirical Studies

Data for this study were obtained from Global Financial Data (GFD) and consists of 5840 daily closing price for US West Intermediate Taxes (WTI) crude oil in dollars/barrel, which transforms to 1212 weekly observations. The date range for the data is from 4th April, 1983 to 30th June, 2006. As for ARMA-GARCH and WDVaR model fitting, 60% of the data, i.e. 727 observations, are reserved as the training set, while the remaining serves as the test set. As for WDNEVaR estimates, 40% of the test set, i.e. 194 observations, is reserved for training the neural network structure. The rolling window method is used to reflect the arrival of new information. Three different ex-post measures of volatilities, i.e. squared returns, historical volatility and realized volatilities are calculated using daily closing prices in the markets. VaRs are estimated using the average weekly prices as calculated in $P_w = \frac{1}{l} \sum_{i=1}^l P_i$, where P_w

is the weekly price. P_i is the daily price within the week. l is the length of the week. Performance of different models is evaluated by both Mean Square Error (MSE) for its accuracy and Backtesting procedures for reliability. The returns are log differenced as $r_t = \ln(P_t / P_{t-1})$. For each experiment, a portfolio of one asset position worth \$1 is assumed.

Several stylized facts about the data series can be summarized from table 1 as follows: price series exhibit significant fluctuations. The standard deviation is significantly positive, which indicates unignorable level of risks in the markets. There are considerable gains, as well as losses, and the market return is asymmetric and loss intending, as suggested by negative skewness levels. There is also high probability of extreme events occurrence as suggested by significant excess kurtosis levels.

Table 1. Descriptive Statistics for crude oil returns series in US WTI crude oil markets

Crude Oil market	Mean	Maximum	Minimum	Median	Standard Deviation	Skewness	Kurtosis	Jarque-Bera Test (p value)	BDS Test (p value)
WTI	0.0007	0.2440	-0.1967	0.0019	0.0412	-0.2107	6.7580	0	0

Rejection of both, Jarque-Bera test of normality and BDS test of independence, suggests that the data deviate significantly from the normal distribution and are non-linear dependent in nature. Therefore, the market is characterized by the leptokurtic (fat-tail and heteroscedasticity) phenomenon, where non-linear dynamics dominate.

Table 2. Experiment results for ARMA-GARCH VaR in US WTI crude oil market

Crude Oil market	Confidence Level	Exceedances	MSE	Kupiec Test Statistics	P-Value	Model Acceptance
US WTI	99%	8	0.0131	1.7281	0.1187	✓
Crude Oil	97.5%	14	0.0099	0.2835	0.5944	✓
	95%	21	0.0076	0.4793	0.4887	✓

Experiment results in table 2 show that the ARMA-GARCH approach offers acceptable levels of reliability in the tested market. The ARMA-GARCH VaR is accepted across all confidence levels. However, under fierce competition and increasingly tight profit margins, the ARMA-GARCH approach no longer suffices due to the following reasons. Firstly, as indicated by the generally low level of failure rates, the ARMA-GARCH provides conservative VaR estimates that achieve reliability at the cost of accuracy. This may result in “more than enough” idle capital and increase operational costs. Secondly, when higher levels of reliability are demanded by market participants due to increasing competition level, this linear combination approach offers little room for further performance.

Table 3. Experiment results for WDVaR(x, 1) in WTI crude oil market

Crude Oil market	Confidence level	WDVaR(Haar, 1)			WDVaR(Db2, 1)			WDVaR(Coif2, 1)		
		Exceedances	MSE	P-Value	Exceedances	MSE	P-Value	Exceedances	MSE	P-Value
US WTI	99.0%	14	0.0123	0.0007	14	0.0148	0.0007	18	0.0137	0.0000
Crude Oil	97.5%	16	0.0093	0.2857	19	0.0111	0.0656	24	0.0104	0.0023
	95.0%	28	0.0072	0.4516	32	0.0085	0.1256	33	0.0079	0.0852

The notion WDVaR(x, i) in table 3 refers to the VaR estimated based on x wavelet families and the decomposition level (scale) i. Experiment results above show that by switching from Coif2 to Db2 and Haar wavelet family, both accuracy and reliability level can be improved. E.g. when switching from Coif2 to Haar wavelet families, both reliability and accuracy level of VaR estimates increase significantly. As experiment results suggest, the performance difference is closely related to the characteristics of the wavelets used as different wavelet families lead to different trade-offs between accuracy and reliability levels in VaR estimates. Clearly the advantage of WDVaR lies in the additional flexibility introduced into the modeling process. By tuning the two parameters, WDVaR provide different perspectives into the underlying risk

evolution and is capable of striking a balance between reliability and accuracy for VaR estimates.

Since VaR estimated based on different wavelet families tracks the underlying risk evolution based on partial information set extraction, to estimate VaR at the highest level of reliability and accuracy, it would be necessary to combine partial information sets during forecasting. The artificial neural network based non-linear ensemble algorithm is used to derive the optimal weights for combining individual forecasts. The back propagation feedforward neural network with structure (3, 6, 1) is used. Through trial and error method, the historical volatility is selected to be the volatility proxy since WDNEVaR using it provides the best performance. Experiment results are listed in table 4.

Table 4. Experiment results for WDNEVAR in WTI crude oil market

Crude Oil Markets	Volatility Proxy	Confidence Level	Exceedances	MSE	Kupiec Test Statistics	P-value	Model Acceptance
US WTI Crude Oil	Historical Volatility	99.0%	3	0.0120	0.0022	0.9626	✓
		97.5%	6	0.0091	0.2525	0.6153	✓
		95.0%	12	0.0069	0.5175	0.4719	✓

Experiment results in table 4 suggest that WDNEVaR achieve higher levels of reliability and accuracy than the ARMA-GARCH based approach. Reliability level of VaR estimated improves as the p value improves in most of the situations, except the slight inferior performance in WTI market at 95% confidence level. More importantly, the accuracy level increases drastically, when measured by significantly lower MSE value uniformly. The improved accuracy in risk tracking implies less idle funds and better allocation of capital resources for financial institutions.

5 Conclusions

The contribution of this paper is two fold. Firstly, this paper incorporates the multi-scale heterogeneous data structure into the modeling process. Separation and integration of different layers in the data are conducted using wavelet analysis. When data are decomposed into the underlying influencing factors in the time scale domain, models are fitted to more stationary data with fewer violations of assumptions and improved goodness of fit. Wavelet analysis has demonstrated its capability to explicitly reveal and track the multi-scale time varying patterns of the data in these experiments. Secondly, in light of the estimation biases resulting from arbitrary selection of different sets of parameters, this paper proposes Artificial Neural Network (ANN) based non-linear ensemble algorithm to reduce estimation bias introduced by different wavelet families. Given the lack of consensus on the suitability and correspondence of certain wavelet families to particular data features in the current literature, non-linear ensemble approaches have proven to be a promising approach to minimize biases introduced by different wavelet families during the forecasting process. It should be noted that the performance of the proposed

WDNEVaR is sensitive to volatility proxies and parameter settings, as suggested by experiment results. More refined volatility proxies could further improve the model performance.

Acknowledgement. The work described in this paper was supported by a grant from City University of Hong Kong (No. 9610058).

References

1. Dowd, K.: *Measuring Market Risk*. John Wiley, Chichester (2005)
2. Gençay, R., Selçuk, F., Whitcher, B.: *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press, San Diego, CA (2002)
3. Ramsey, J.B.: The contribution of wavelets to the analysis of economic and financial data. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* **357** (1999) 2593-2606
4. Aminghafari, M., Cheze, N., Poggi, J.-M.: Multivariate denoising using wavelets and principal component analysis. *Computational Statistics & Data Analysis* **50** (2006) 2381-2398
5. Yamada, H.: Wavelet-based beta estimation and Japanese industrial stock prices. *Appl. Econ. Lett.* **12** (2005) 85-88
6. Fernandez, V.P.: The international CAPM and a wavelet-based decomposition of value at risk. *Stud. Nonlinear Dyn. Econom.* **9** (2005) -
7. Fernandez, V.: The CAPM and value at risk at different time-scales. *International Review of Financial Analysis* **15** (2006) 203-219