A RISK AND FORECASTING ANALYSIS OF WEST TEXAS INTERMEDIATE PRICES

January 2010

By

David E. Allen[1]
School of Accounting, Finance and Economics,
Edith Cowan University

Abhay Kumar Singh
School of Accounting, Finance and Economics,
Edith Cowan University

[1] School of Accounting, Finance and Economics, Edith Cowan University, 270 Joondalup Drive, Joondalup WA 6027. Tel 618630455471. Email d.allen@ecu.edu.au

List of Symbols

$f(\,)$	**Function**
X	**Input Vector**
$h(\,)$	**Hypothesis Function**
$\mathcal{H}$	**Hypothesis Space**
S	**Training Set**
$\mathcal{L}$	**Set of Location Information**
$\mathbf{R}^d$	**Space of d dimension**
λ	**Abstract parameter**
Λ	**Set of abstract parameters**
$L(\,)$	**Loss function**
$R(\,)$	**Risk function**
$\mathbf{R}^+$	**Non-negative real numbers**
$\widetilde{R_n}\,(\,)$	**Empirical risk function**
w, b	**Parameters for hyperplane**
α, β	**Lagrange multipliers**
ξ_i, ξ_i^*	**Slack Variables**
$F_\sigma(\xi)$	**Penalty function**
ε	**Error**

Executive Summary

In this report we engage in two major pieces of analysis which are applied to a sample of logarithmic returns formed from the daily closing prices of TWI oil prices. In the first section we employ CAViaR, a modelling approach formulated by Engle and Manganelli in (2004) which is a Value at Risk (VaR) modelling technique that uses quantile regression, to forecast WTI value at risk. This has not been used in its original state to predict Value at Risk (VaR) for oil prices.

In the second section of analysis we also show the applicability of Support Vector Regression for oil price prediction and compare it with more standard time-series ARIMA modelling. Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. In simple words, given a set of training examples, each marked as belonging to one of two categories, a SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. Intuitively, an SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. A version of SVM for regression was proposed in 1996 by Vapnik, Drucker, Burges, Kaufman and Smola. This method is called support vector regression (SVR).

We apply this technique to predict WTI price levels, and present a comparative analysis of two methods, ARIMA a widely used method based on lag and momentum effect and Support Vector Regression, a more efficient machine learning approach. The results obtained clearly indicate that SVR is more efficient in predicting future oil price levels. The current work can be extended by using more sophisticated optimization routines for tuning the parameters of the SVR and also by changing the frequency of data used.

Contents

1. INTRODUCTION

As one of the biggest traded commodities, the behaviour of crude oil prices has enormous significance and impacts on different parts of the economy; effecting government, enterprises and investors, etc. Therefore, there are increasing demands for a better characterization and prediction of crude oil prices. The behaviour of oil prices has a significant effect on economic activity and state revenue in Western Australia.

The four major benchmarks in the world of international oil trading today are: (1) West Texas Intermediate (WTI), the reference crude for USA, (2) Brent, the reference crude oil for the North Sea, (3) Dubai, the benchmark crude oil for the Middle East and Far East, and (4) Tapis, the benchmark crude oil for the Asia-Pacific region. West Texas Intermediate (WTI) is the major benchmark of crude oil prices in the Australian market.

Due to unpredictable trends in Oil prices, oil price prediction and its risk measurement has been a major research theme in the last decade. Oil plays a key role in economies throughout the world, future price prediction and quantification of the various risks associated with oil are critical for central governments and businesses around the world.

Various time series modelling techniques including ARMA, ARIMA, ARCH, GARCH etc has been used in oil price and volatility prediction in the past. Also some advanced learning algorithms including neural networks, genetic algorithms, machine learning etc are frequently tested. Support Vector Machines (SVMs) are a new generation learning system based on recent advances in statistical learning theory. The Vapnik–Chervonenkis theory (also known as VC theory) was developed during 1960–1990 by Vladimir Vapnik and Alexey Chervonenkis. The concept of the Support vector machine, was developed by Vapnik (1995) and his colleagues. The theory is a form of computational learning theory, which attempts to explain the learning process from a statistical point of view. Established on the unique theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVM is resistant to the over-fitting problem and can model nonlinear relations in an efficient and stable way. Furthermore, SVM is trained as a convex optimization problem resulting in a global solution that in many cases yields unique solutions. Originally, SVMs were developed for classification tasks Burges (1998). With the introduction of Vapnik's ϵ -insensitive loss function. SVMs have been extended to solve nonlinear regression and time series prediction problems, and they exhibit excellent performance, Huang, W., Nakamori, Y., Wang (2005), Muller, K.R., Smola, J.A., Scholkopf (1997).

The widely used risk standard, Value at Risk (VaR) reduces the risk associated with any kind of asset to just a number (an amount in terms of a currency), which can be well understood by regulators, board members, and other interested parties. One of the objectives of this study is to employ a new value at risk methods approach proposed by Engle and Manganelli (2004) to the forecasting of oil price risk. Previous work in oil price risk measurement has employed various other techniques including; ARMA historical simulation approach, Cabedo, J.D.,

Moya (2003) , ARCH based volatility measures, Kuper, G (2001), and other VaR based risk measuring techniques, Sauter, R., Awerbuch (2002).

In this study we employ CAViaR, the value at risk modelling technique that uses quantile regression, to forecast WTI value at risk, which has not been used in its original state to predict Value at risk for oil prices. We also show the applicability of Support Vector Regression for oil price prediction and compare it with ARIMA modelling.

The rest of the report is divided into 6 sections; the following section two gives the background and introduces quantile regression, the CAViaR model, ARIMA and support vector machine regression. Section 3 presents the data and methodology used in the study. Section 4 discusses the results of the CAViaR implementation on WTI oil prices followed by Section 5 which compares the forecasting results obtained from ARIMA and Support Vector Regression techniques. Section 6 provides the conclusion followed by the references.

2. BACKGROUND

2.1 CAViaR and Quantile Regression

The problem in estimating VaR is that it is a particular quantile of potential future portfolio values, conditioned on current available information. However, portfolio returns and risk change over time, so a time-varying forecasting procedure is required. Essentially this involves forecasting a value each period that will be exceeded with a probability of *(1-θ)* by the current portfolio value. In this case $\theta \in (0,1)$ are representative of the confidence level attached to the VaR.

Engle and Manganelli (2004), used the robust technique of quantile regression and proposed another method for calculation of Value at Risk which they termed as, Conditional Autoregressive Value at Risk by Regression Quantiles, or CAViaR. CAViaR, uses quantile regressions and instead of modelling the whole return distribution for calculation of VaR, it models the required quantiles of the return distribution directly. To predict the value at risk by modelling the lower quantiles, the model uses a conditional autoregressive specification, inspired by the fact that the distribution of volatilities over time is auto-correlated, hence the model. In their CAViaR paper, Engle and Manganelli (2004) propose four different specification processes for the calculation of value at risk viz: an Adaptive model, a Symmetric Absolute Value, an Asymmetric Slope and an Indirect GARCH model. We follow suit and test the relative suitability of all the four models on our Australian sample data set in the calculation of VaR.

The first model; an Adaptive model, is a smoothed version of a step function (for finite G), is given by

$$f_t(\beta_1) = f_{t-1}(\beta_1) + \beta_1\{[1 + \exp(G[y_{t-1} - f_{t-1}(\beta_1)])]^{-1} - \theta\}, \quad (1)$$

where G is some positive finite number. The model is a smoothed version of a step function. Engle and Manganelli suggest that the adaptive model incorporates the following rule: whenever VaR is exceeded, you should immediately increase it, but in circumstances where

you do not exceed it, you should decrease it very slightly. This strategy obviously will reduce the probability of sequences of hits and will also make it unlikely that there will never be hits. The structure of this model means it learns little from returns that are either close to the VaR or are extremely positive, when G is large. It increases the VaR by the same amount regardless of whether the returns exceed the VaR by a small margin or a large margin. This model has a unit coefficient on the lagged VaR.

A second model which features symmetric absolute values is set out below:

$$f_t(\beta) = \beta_1 + \beta_2 f_{t-1}(\beta) + \beta_3 |y_{t-1}| \quad (2)$$

A third has an asymmetric Slope:

$$f_t(\beta) = \beta_1 + \beta_2 f_{t-1}(\beta) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^- \quad (3)$$

where, notation $(x)^+ = max(x,0)$, $(x)^- = -min(x,0)$.

Whilst the fourth is an indirect GARCH (1,1):

$$f_t(\beta) = (\beta_1 + \beta_2 f^2{}_{t-1}(\beta) + \beta_3 y^2{}_{t-1})^{1/2} \quad (4)$$

The first and third models respond symmetrically to past returns, whereas the second allows the response to positive and negative returns to be different. All three are mean-reverting in the sense that the coefficient on the lagged VaR is not constrained to be 1.

The indirect GARCH model would be correctly specified if the underlying data were truly a GARCH(1, 1) with an iid error distribution. The symmetric absolute value and asymmetric slope quantile specifications would be correctly specified by a GARCH process in which the standard deviation, rather than the variance, is modelled either symmetrically or asymmetrically with iid errors. The first specification and estimation of this model was by Taylor (1986) and Schwert (1988) and it was subsequently analysed by Engle (2002). A merit of the CAViaR specification, as suggested by Engle and Manganelli (2004), is that it is more general than these GARCH models.

The VaR results from the four methods are tested using a dynamic quantile test, as proposed by Engle and Manganelli (2004). We will omit further details of the methods for the sake of brevity, as further insights can be obtained from their original paper.

Quantile Regression

CAViaR uses quantile regression for estimation of its parameters, quantile regression was first introduced by Koenker and Bassett (1978). Koenker and Bassett showed how to extend the notion of a sample quantile to a linear regression model.

Quantile regression as introduced in Koenker and Bassett (1978) is an extension of classical least squares estimation of conditional mean models to the estimation of an ensemble of models for conditional quantile functions. The central special case is the median regression estimator that minimizes a sum of absolute errors. The remaining conditional quantile

functions are estimated by minimizing an asymmetrically weighted sum of absolute errors. Taken together the ensemble of estimated conditional quantile functions offers a much more complete view of the effect of covariates on the location, scale and shape of the distribution of the response variable.

In linear regression, the regression coefficient represents the change in the response variable produced by a one unit change in the predictor variable associated with that coefficient. The quantile regression parameter estimates the change in a specified quantile of the response variable produced by a one unit change in the predictor variable.

The quantiles, or percentiles, or occasionally fractiles, refer to the general case of dividing a dataset into parts. Quantile regression seeks to extend these ideas to the estimation of conditional quantile functions - models in which quantiles of the conditional distribution of the response variable are expressed as functions of observed covariates.

In quantile regression, the median estimator minimizes the symmetrically weighted sum of absolute errors (where the weight is equal to 0.5) to estimate the conditional median function, other conditional quantile functions are estimated by minimizing an asymmetrically weighted sum of absolute errors, where the weights are functions of the quantile of interest. This makes quantile regression robust to the presence of outliers.

We can define the quantiles through a simple alternative expedient as an optimization problem. Just as we can define the sample mean as the solution to the problem of minimizing a sum of squared residuals, we can define the median as the solution to the problem of minimizing a sum of absolute residuals. The symmetry of the piecewise linear absolute value function implies that the minimization of the sum of absolute residuals must equate the number of positive and negative residuals, thus assuring that there are the same number of observations above and below the median.

In summary, consider a series of observations $y_1,....y_T$ generated by the following model

$$y_t = \mathbf{x}\boldsymbol{\beta}^{0} + \varepsilon_{\theta t}, \quad Quant_{\theta}(\varepsilon_{\theta t} | \mathbf{x}_t) = 0, \qquad (5)$$

Where $\mathbf{x_t}$ is a p-vector of regressors and $Quant_{\theta}(\varepsilon_{\theta t} | \mathbf{x_t})$ is the θ-quantile of $\varepsilon_{\theta t}$ conditional on $\mathbf{x_t}$. Let $f_t(\boldsymbol{\beta}) \equiv \mathbf{x}_t\boldsymbol{\beta}$. Then the θ th regression quantile is defined as any $\hat{\boldsymbol{\beta}}$ that solves:

$$\min_{\beta} \frac{1}{T}\sum_{t=1}^{T}[\theta - \mathbf{I}\boldsymbol{\beta}y_t < f_t(\ ))][\boldsymbol{\beta}y_t - f_t(\ )] \qquad (6$$

We will not discuss further the mathematical details of the regression technique, please refer to Koenker's (2005) monograph for a comprehensive discussion.

2.2 Autoregressive Integrated Moving Average (ARIMA):

Forecasting is the process of estimation in unknown situations from the historical data. For example forecasting weather, stock index values, commodity prices etc. Time series forecasting provides a method to forecast future price levels using the historical price of commodities, here oil.

In statistics, ARIMA models, sometimes called Box-Jenkins models after the iterative Box-Jenkins methodology usually used to estimate them, are typically applied to time series data for forecasting.

Given a time series of data X_{t-1}, X_{t-2},.....X_2, X_1, the ARIMA model is a tool for understanding and, perhaps, predicting future values in this series. The model consists of three parts, an autoregressive (AR) part, a moving average (MA) part and the differencing part. The model is usually then referred to as the ARIMA (p, d, q) model where p is the order of the autoregressive part, d is the order of differencing and q is the order of the moving average part.

If d = 0, the model becomes ARMA, which is linear stationary model. ARIMA (i.e. d > 0) is a linear non-stationary model. If the underlying time series is non-stationary, taking the difference of the series with itself d times makes it stationary, and then ARMA is applied onto the differenced series.

ARIMA(p,d,g) model is given by:

$$\varphi \nabla^d X_t = \theta \epsilon_t \tag{7}$$

Where AR part:

$$\varphi = 1 - \sum_{i=1}^{p} \varphi_i L^i$$

MA part:

$$\theta = 1 + \sum_{j=1}^{q} \theta_j L^j$$

I (difference) part:

$$\nabla = (1 - L^1)$$

Here L is lag operator, i.e. $L^i X_t = X_{t-i}$. φ_i and θ_j are the model parameters which need to be found before applying the model for forecasting. ε_t is a white noise process with zero mean and variance σ^2 .

φ_i are the parameters of the autoregressive part of the model, θ_j are the parameters of the moving average part and the ε_t are error terms. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

2.3 Support Vector Machine (SVM) and Support Vector Regression (SVR)

SVM and SVR are a set of related supervised learning methods used for classification and regression respectively. They belong to a family of generalised linear classifiers. A special property of SVMs is that they simultaneously minimise the empirical classification error and maximise the geometric margin; hence they are also known as maximum margin classifiers. SVMs follow the Structured Risk Management (SRM) principle.

Linear separating functions, however, generalized with an error minimization rate are not suitable for real world applications. The idea of the *Support Vector Machine* (SVM) is to map the input vectors into a *feature space* with higher number of dimensions, and to find an optimal separating hyperplane in the feature space. For example, points in a two- dimensional space $(x_1, x_2) \in R^2$ may be mapped into the 5 dimensional plane $(x_1, x_2, x_1x_2, x_1^2, x_2^2) \in R^5$, a separating hyperplane in this larger space will correspond to a conic separator in R^2.

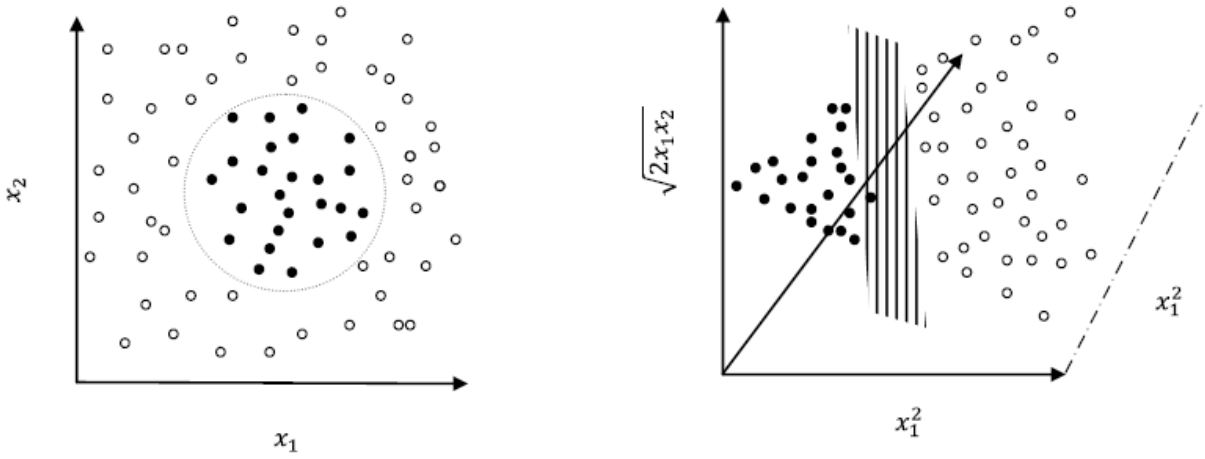

Figure 1: Separating Hyperplane in higher dimension

A kernel is used to construct a mapping into high dimensional feature space by the use of reproducing kernels. The idea of the kernel function is to enable operations to be performed in the input space rather than the potentially high dimensional feature space. Hence, the inner product does not need to be evaluated in the feature space. This provides a way of addressing the curse of dimensionality. However, the computation is still critically dependent upon the number of training patterns and to provide a good data distribution for a high dimensional problem will generally require a large training set.

The kernel function $K\langle .,. \rangle$is a convolution of the canonical inner product in the feature space. Common kernels for use in a SVM are the following:

1. **Dot Product**: *K(x, y) = x .y*; in this case no mapping is performed, and only the optimal separating hyperplane is calculated.
2. **Polynomial functions**: *K(x, y)* = $(x.y + 1)^d$, where the *degree d* is given.
3. **Radial Basis Functions** (RBF): *K(x, y)* = $e^{-\gamma \|x-y\|^2}$ with parameter γ.
4. **Sigmoid (or neural) kernel**: *K(x, y)* = tanh (*ax.y + b*) with parameters *a* and *b*.

5. **ANOVA kernel:** $K(x, y) = (\sum_{i=1}^{n} e^{-\gamma (x_i - y_i)})^d$, with parameters γ and *d.*

Due to prior work with an emphasis on Radial basis functions kernel, it is used in the current study for prediction.

SVM for Regression

The Basic Idea

Suppose we are given training data $\{(x_1, y_1), \ldots\ldots, (x_l, y_l)\} \subset X \times R$ where X denotes the space of input patterns (e.g. $X = R^d$). In $\varepsilon - SV$ regression, our goal is to find a function $f(x)$ such that it has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. In other words, we do not care about errors as long as they are less than ε, but will not accept any deviation larger than this. This may be important if we don't want error in location in context of wireless to be more than ε. We begin by describing the case of linear functions f taking the form

$$f(x) = \langle w, x \rangle + b \; with \; w \in X, b \in R \tag{8.1}$$

Where $\langle ., . \rangle$ denotes the dot product in *X*. *Flatness* in this case (6.1) means that one seeks a small lw. One way to ensure this is to minimize the norm, i.e. $\left|\left|w\right|\right|^2 = \langle w, w \rangle$. We can write this problem as a convex optimization problem:

$$\text{Minimize } \frac{1}{2} \left|\left|w\right|\right|^2$$

$$\text{Subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \tag{8.2}$$

Here, the assumption is there that such a function f actually exists that approximates all pairs (x_i, y_i) with ϵ precision, or in other words, a convex optimization problem is really possible. Sometimes, however, this may not be the case, or we also may want to allow for some errors. Analogously to the soft margin loss function, one can introduce slack variables $\boldsymbol{\xi_i}, \boldsymbol{\xi_i^*}$ to cope with the otherwise infeasible constraints of the optimization problem. Hence, there is a formulation stated by Vapnik:

$$\text{Minimize } \frac{1}{2} ||w||^2 + C \sum_{i=1}^{\ell} (\boldsymbol{\xi_i} + \boldsymbol{\xi_i^*})$$

$$\text{Subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \boldsymbol{\xi_i} \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \boldsymbol{\xi_i^*} \\ \boldsymbol{\xi_i}, \boldsymbol{\xi_i^*} \qquad\qquad \geq 0 \end{cases} \tag{8.3}$$

The constant C > 0 determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated. This corresponds to dealing with a so-called ε - sensitive loss function $|\boldsymbol{\xi}|_\varepsilon$ described by

$$|\boldsymbol{\xi}|_{\varepsilon} \coloneqq \begin{cases} 0 & if\ |\boldsymbol{\xi}| \leq \varepsilon \\ |\boldsymbol{\xi}| - \varepsilon & \text{otherwise} \end{cases} \tag{8.4}$$

It turns out that in most cases the optimization problem can be solved more easily in its dual formulation. The Dual formulation provides the key for extending the SVM to non-linear functions.

The Dual Problem

Now Lagrange function is constructed from the objective function and the corresponding constraints. It can be shown that this function has a saddle point with respect to the primal and dual variables at the solution.

$$L \coloneqq \frac{1}{2}\left||w|\right|^2 + C\sum_{i=1}^{\ell}(\boldsymbol{\xi_i} + \boldsymbol{\xi_i^*}) - \sum_{i=1}^{\ell}(\boldsymbol{\eta_i}\boldsymbol{\xi_i} + \boldsymbol{\eta_i^*}\boldsymbol{\xi_i^*}) - \sum_{i=1}^{\ell}\alpha_i(\varepsilon + \boldsymbol{\xi_i} - y_i + \langle w, x_i\rangle + b)$$

$$-\sum_{i=1}^{\ell}\alpha_i^*\,(\varepsilon + \boldsymbol{\xi_i^*} + y_i - \langle w, x_i\rangle - b) \tag{8.5}$$

Here L is the Lagrangian and $\boldsymbol{\eta_i}, \boldsymbol{\eta_i^*}, \alpha_i, \alpha_i^*$ are Lagrange multipliers. Hence, the dual variables in (6.5) have to satisfy positive constraints

$$\alpha_i^{(*)}, \eta_i^{(*)} \geq 0 \tag{8.6}$$

Note that by $\alpha_i^{(*)}$, we refer to α_i and α_i^*

The Saddle point condition: partial derivatives L of with respect to primal variables (w, b, ξ_i, ξ_i^*) have to vanish for optimality.

$$\partial_b L = \sum_{i=1}^{\ell}(\alpha_i^* - \alpha_i) = 0 \tag{8.7}$$

$$\partial_w L = w - \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)x_i = 0 \tag{8.8}$$

$$\partial_{\boldsymbol{\xi_i^{(*)}}} L = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \tag{8.9}$$

The Dual Optimization problem reduces to

$$maximize \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{\ell}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ -\varepsilon\sum_{i=1}^{\ell}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{\ell} y_i(\alpha_i - \alpha_i^*) \end{cases} \tag{8.10}$$

$$subject\ to \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*) = 0\ and\ \alpha_i, \alpha_i^* \epsilon [0, C]$$

In deriving (8.10), we already eliminated the dual variables $\boldsymbol{\eta_i}, \boldsymbol{\eta_i^*}$ through condition (8.9) which can be reformulated as $\eta_i^{(*)} = C - \alpha_i^{(*)}$.

Equation (8.8) can be rewritten as follows

$$w = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)x_i, thus\ f(x) = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b \tag{8.11}$$

This is support vector expansion, i.e. w can be completely described by a linear combination of the training patterns x_i .In a sense, the complexity of a function's representation by SVs is independent of the dimensionality of the input space X, and depends only on the number of SVs.

3. DATA & METHODOLOGY

The current work analyzes the risk and forecasting of oil prices, the data used is a 15 year sample (21^{st} October 1994- 10^{th} October 2009), of daily closing spot prices of WTI. We use the new efficient model of Engle and Manganelli (2004) CAViaR, which uses quantile regression to model the value at risk of the WTI oil price return. ARIMA and Support Vector Regression techniques are also utilised and compared for analysing the forecasting of oil prices. We make use of percentage daily returns calculated in logarithms, which also makes the time series stationary for forecasting methods.

3.1 Risk Analysis

We apply the four CAViaR methods to WTI spot price data. We use three window sizes to estimate the model, (i) 3000 daily observations with 2500 in sample observations and 500 out of sample observations, (ii) 2000 daily observations with 1600 in sample observations and 400 out of sample observations and, (iii) 1000 daily observations with 700 in sample observations and 300 out of sample observations. We estimate 1% CAViaR using the model, the obvious choice of 1% for VaR comes from the current industry standards.

Engle and Manganelli (2004) discuss the appropriate specifications for testing these time series models based on quantile regressions. They derive both in sample tests and out of sample tests. These essentially feature measuring the proportion of hits of the limiting VaR and having them equal to the target fraction with no correlation with own lagged values, no autocorrelation in the hits and the correct fraction of exceptions. They explain that the in-sample test, or DQ test is a specification test for the particular CAViaR process under study and it can be very useful for model selection purposes. They suggest the parallel DQ out of sample tests could be used by regulators to check that the VaR estimates submitted by a financial institution satisfy some basic model specification requirements such as unbiasedness, independent hits and independence of the quantile estimates. We utilise their tests and Matlab code in this paper. (We are thankful to Simone Manganelli for making available his MATLAB code for the exercise).

3.2 Forecasting

We analyse two different forecasting techniques, ARIMA and Support vector regression to forecast the oil prices. The logarithm percentage daily returns are used for forecasting as it makes the time series stationary which is non stationary otherwise. The time series is tested for unit root using the Augmented Dickey Fuller (ADF) test.

We use last 1000 observations in the data set for prediction purposes with 990 observations for building the model (in both ARIMA and SVR) and 10 for testing purposes. The two models are compared based on the errors of prediction given by the Root mean square error and mean error. The lower the error the better the prediction technique is. GRETL, open source software is used for the ARIMA analysis. The ARIMA model is identified based on the significance of coefficients based on the t-statistics. WEKA java based machine learning software, which uses LIBSVM wrapper class is used for support vector regression prediction.

SVR forecasting involves following steps.

(a) Data sampling. We use WTI daily spot price data for this research.

(b) Data pre-processing. The collected oil price data may need to be transformed into certain appropriate ranges for the network learning by logarithmic transformation, difference or other methods. Here logarithmic return series is used. Then the data should be divided into in-sample data and out-of-sample data. We use first 990 observations (in-sample) for training the SVR model and rest 10 observations (out-of-sample) for testing the built model.

(c) Training and learning. The SVM architecture and parameters are determined in this step by the training results. There are no criteria in deciding the parameters other than by a trial-and-error basis. In this investigation, the RBF kernel is used because the RBF kernel tends to give good performance under general smoothness assumptions. Consequently, it is especially useful if no additional knowledge of the data is available. Finally, a satisfactory SVM-based model for oil price forecasting is reached.

(d) The parameters of SVR are tuned using the trial and error method. The parameters are tuned based on the error received from the forecasting results. We use the last 2 days returns as predictor variables.

(e) Future price forecasting.

Both the methods are used to forecast the next 10 day returns and then are analyzed based on the root mean square error received.

4. ANALYSIS OF RESULTS

4.1 CAViaR Risk Results:

Table-1, gives the results obtained by using 1000 time series returns of WTI spot prices, with 700 in sample and 300 out of sample observations. DQ test results show that while the value at risk is significant for the in-sample period for Symmetric Absolute Value (SAV), Asymmetric Slope (AS), and Adaptive model, its insignificant for the Indirect GARCH model. On the other hand the Indirect GARCH model gives significant out of sample results when rest of the models fail.

Table 1: Estimates for four CAViaR specification with 1000 return series.

	SAV	**AS**	**Indirect GARCH**	**ADAPTIVE**
Beta1	2.8165	2.3795	0.3201	-0.0878
Std Errors	1.94	2.0784	0.607	0.1393
p Values	0.0733	0.1261	0.299	0.2641
Beta2	0.32	0.3957	0.9444	0
Std Errors	0.4209	0.4483	0.0297	0
p Values	0.2235	0.1888	0	0
Beta3	0.3354	0.2994	0.2795	0
Std Errors	0.1465	0.1572	0.2836	0
p Values	0.011	0.0284	0.1621	0
Beta4	0	0.2576	0	0
Std Errors	0	0.1668	0	0
p Values	0	0.0613	0	0
RQ	35.6822	35.1623	35.9985	37.4356
Hits in-sample(%)	0.7143	1.1429	0.8571	0.7143
Hits out-of-sample(%)	8	7.6667	0.6667	42
DQ In Sample (p value)	0.9995	0.9915	0.0051	0.9996
DQ Out of Sample (p value)	0	0	0.9985	0

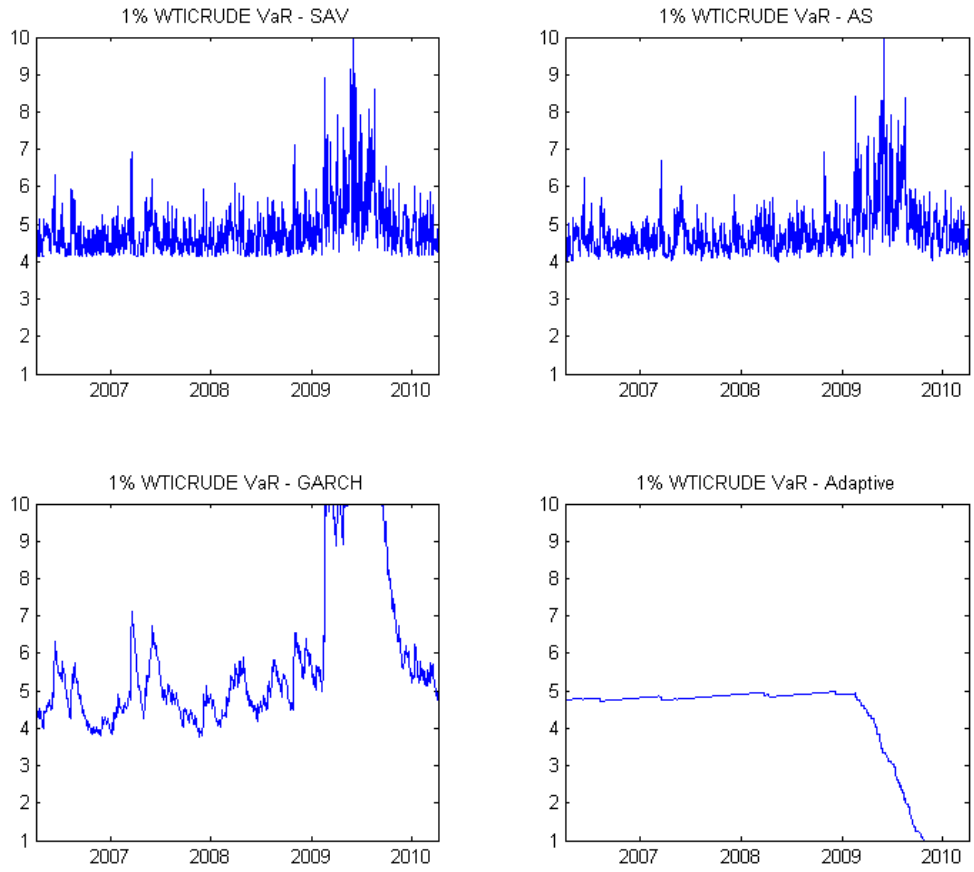

Figure 1: Estimated 1% CAViaR graph for Sample 1.

Table-2, gives the results from last 2000 return series from the experiments sample data with 1600 in-sample and 400 out of sample observations. The results here change with sample size, it is evident from the DQ test results that now all the four models are significant for in-sample results. The out of sample results show that none of the models are significant. This shows that the results from the caviar model do change with sample size.

Table 2: Estimates for four CAViaR specification with 2000 return seris.

	SAV	**AS**	**Indirect GARCH**	**ADAPTIVE**
Beta1	6.3503	7.139	0.0622	0.3498
Std Errors	1.7954	1.8129	0.252	0.2684
p Values	0.0002	0	0.4025	0.0962
Beta2	-0.3402	-0.4942	0.9816	0
Std Errors	0.3565	0.3403	0.0101	0
p Values	0.17	0.0732	0	0
Beta3	0.3621	0.341	0.1055	0
Std Errors	0.0846	0.115	0.0908	0
p Values	0	0.0015	0.1228	0
Beta4	0	0.2654	0	0
Std Errors	0	0.1048	0	0

p Values	0	0.0057	0	0
RQ	102.1509	102.1364	102.7917	104.9777
Hits in-sample(%)	0.9375	1	1.0625	0.75
Hits out-of-sample(%)	7	7.25	1.75	3.75
DQ In Sample (p value)	0.9793	0.9759	0.4681	0.1265
DQ Out of Sample (p value)	0	0	0.0002	0

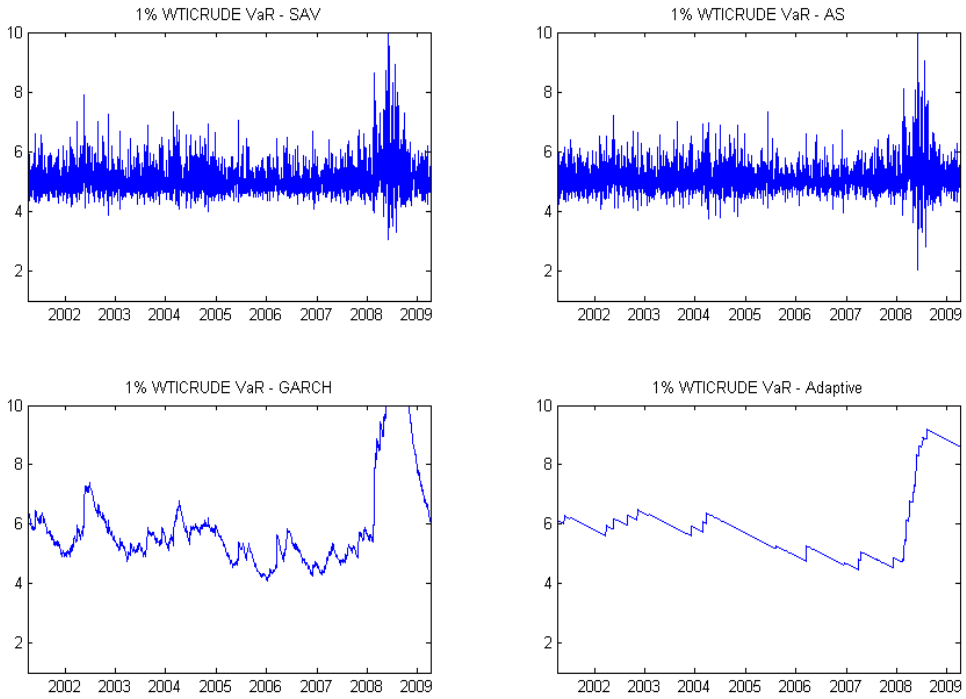

Figure 2: Estimated 1% CAViaR graph for Sample 2.

Finally table -3 gives the results when we increase the sample size further by a 1000 observations. Here the in-sample size is 2500 and the out sample size is 500 which matches the out of sample size used in the original study (engle ref.). The oil price CAViaR predictions show here that the first three models are efficient for both in-sample and out sample observations while, with this sample size the adaptive model fails to give significant results.

The result here shows that with an increase in sample size the first three models become more relevant than the adaptive model.

Table 3: Estimates for four CAViaR specification with 3000 return seris.

	SAV	AS	Indirect GARCH	ADAPTIVE
Beta1	1.5846	1.1258	6.914	0.5981
Std Errors	1.0244	0.6523	5.9457	0.2411
p Values	0.0609	0.0422	0.1224	0.0065
Beta2	0.6348	0.6622	0.6611	0
Std Errors	0.1976	0.1226	0.1721	0
p Values	0.0007	0	0.0001	0
Beta3	0.5567	0.6642	1.3579	0
Std Errors	0.1602	0.1161	1.5042	0
p Values	0.0003	0	0.1833	0
Beta4	0	0.5028	0	0
Std Errors	0	0.1283	0	0
p Values	0	0	0	0
RQ	212.4748	212.3826	213.6665	220.4636
Hits in-sample(%)	1.08	1.04	1.04	0.72
Hits out-of-sample(%)	1	0.8	0.8	2.2
DQ In Sample (p value)	0.9044	0.9361	0.9184	0.0493
DQ Out of Sample (p value)	0.9943	0.9955	0.9798	0

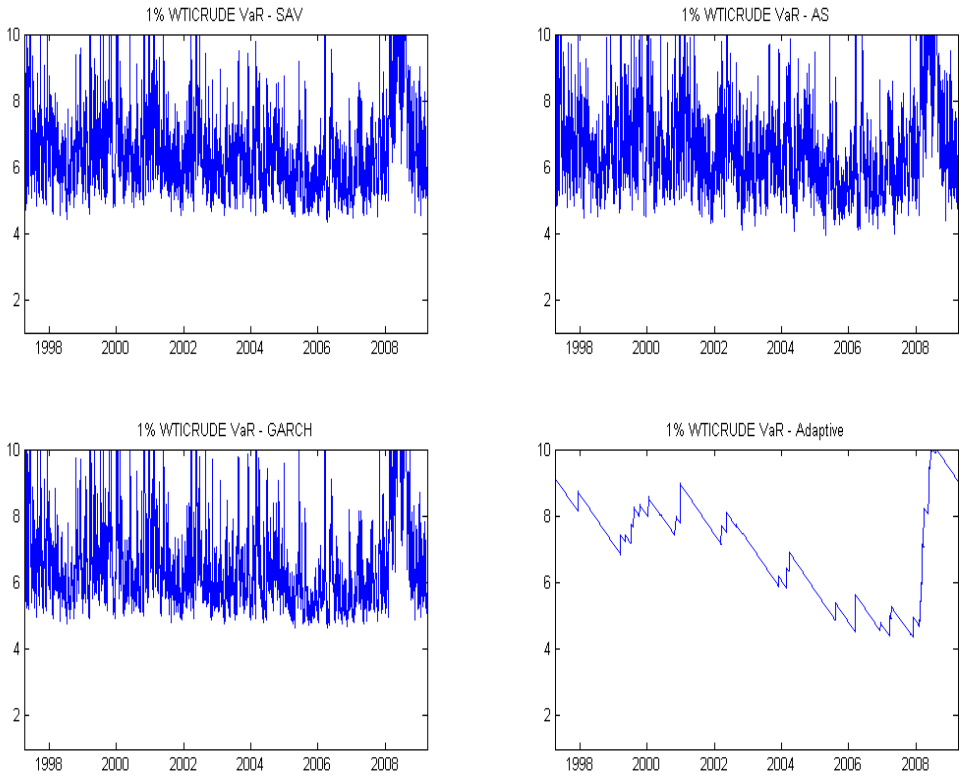

Figure 3: Estimated 1% CAViaR graph for Sample 3.

4.2 Forecasting Result Analysis

For oil price prediction using ARIMA and SVR we use the last 1000 WTI logarithmic return series. The series itself is not stationary and gives a unit root when tested with an Augmented Dickey Fuller (ADF) Test with a constant. The return series is used which is useful as it makes the sample time series stationary and hence can be used for prediction.

ARIMA Forecasting Result Analysis

Based on the significance of factor coefficients, ARIMA(2,0,2), ARIMA(3,0,2) and ARIMA(2,0,3) models are identified. No differencing of the series is needed as we are using the return series which is stationary. We will forecast the time series using all the three models identified and will chose the one having minimum root mean square error (RMSE).

Table-4 gives the error obtained from forecasting using the three ARIMA models. ARIMA(3,0,2) gives the minimum error and hence it is used for the prediction of rest 10 observations after building the model on the first 990 observations.

Table 4: Errors for ARIMA models.

Model	**Root Mean Square Error**
ARIMA(2,0,2)	1.5653
ARIMA(3,0,2)	1.5427
ARIMA (2,0,3)	1.557

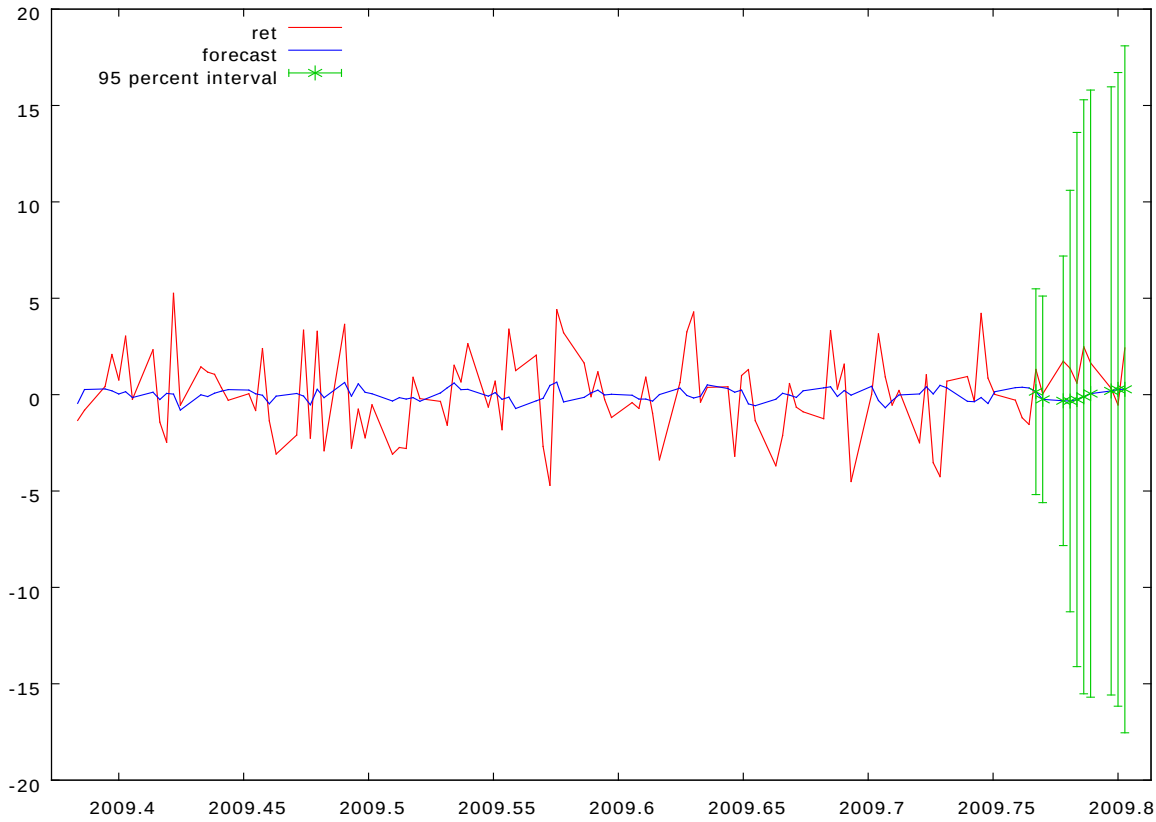

Figure 4: Graph for actual and forecasted return series.

Figure-4 shows the forecasting results using ARIMA(3,0,2) model. The forecast here is a conditional mean forecast as the technique uses the Ordinary Least Squares regression method.

Table- gives the rest of the prediction statistics for the model.

Table 5: Prediction Statistics for ARIMA model.

Mean Error	1.1701
Mean Squared Error	2.3799
Root Mean Squared Error	1.5427
Mean Absolute Error	1.3319
Mean Percentage Error	170.35
Mean Absolute Percentage Error	140.15
Theil's U	1.2044
Bias proportion, UM	0.57532
Regression proportion, UR	0.0452
Disturbance proportion, UD	0.37948

Support Vector Regression Based Forecasting

The main issue in time series forecasting using SVR is the tuning of parameters. Though more computationally intensive methods like grid search are available, we here decided to use a trial and error approach for tuning the parameters to save computational time and complexity.

We tune the first parameter C, the cost using the module CV ParameterSelection. The parameter is tested between 1000<C<2000 with 10 steps, this gives an efficient value of 1500. Finally we tune the other two parameters, gamma γ and epsilon ϵ using trial and error, the final values used for the parameters after trial and error optimization (based on minimum RMSE) are, C= 1500, gamma= 0.333, and epsilon= 0.0001.

The choice of lagged predictor variables here is last two day returns. This choice can be further tested by widening the window size.

Figure-5 shows the graph for actual and predicted out of sample values. This shows that even if the values are not so efficiently predicted they follow the trend better than the trend followed by the ARIMA model. Table- gives the prediction test statistics, the value of RMSE clearly indicates the efficiency of SVR over the ARIMA model. The current model is tuned with a trial and error approach, which is more likely to reduce the error when tuned using more computationally intensive search algorithms.

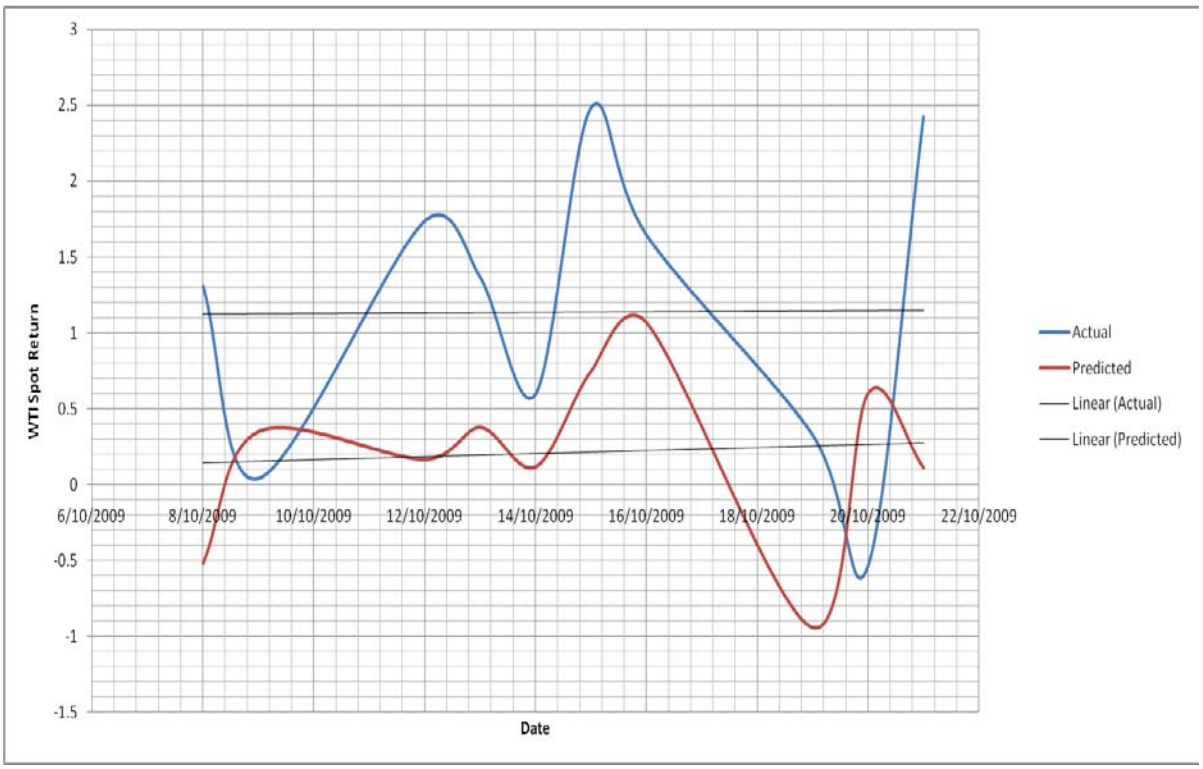

Figure 5: Graph for actual and predicted returns obtained from SVR.

Table 6: SVR forecasting statistics.

Correlation Coefficient	0.2062
Root Mean Squared Error	1.3658
Mean absolute error	1.2194
Relative absolute error	97.7989 %
Root relative squared error	91.8446 %
Total Number of Instances	10

5. CONCLUSION

In this work we applied a new quantile regression approach to modelling VaR (Value at Risk); CAViaR on WTI spot price returns. The analysis shows that the proposed VaR risk modelling technique though efficient, is dependent on the sample size. There is further scope for improving the technique and testing via other sample sizes, thus facilitating a comparative analysis. This research also focussed on the prediction of WTI price levels, and presented a comparative analysis of two methods, ARIMA a widely used method based on lag and momentum effect and Support Vector Regression, a more efficient machine learning approach. The results obtained clearly indicate that SVR is more efficient in predicting the future price levels. The current work can be extended by using more sophisticated optimization routines for tuning the parameters of SVR and also changing the frequency of data used.

This research work also made use of open source software platforms for the forecasting analysis, which are highly capable for financial econometrics and machine learning projects and can be further used by researchers and practitioners to analyse market scenarios and financial time series.

6. REFERENCES

Abramson, B., Finizza, A.: Probabilistic forecasts from probabilistic models: a case study in the oil market. International Journal of Forecasting 11(1995) 63-72

Allen, D.E, McAleer, M. and M. Scarth, (2008) "Realised Volatility Uncertainty" Working Paper, 0807, School of Finance, Economics and Accounting, Edith Cowan University.

Allen, D.E, Singh, Abhay Kumar (2010) "Caviar and The Australia Stock Markets: An Appetiser.", Working Paper Series, School of Finance, Economics and Accounting, Edith Cowan University.

Barone-Adesi. G., Bourgoin, F.. Giannopoulos, K.: Don't look back. Risk August 8 (1998) 100-103

Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2 (1998) 121-167

Cabedo, J.D., Moya, I.: Estimating Oil Price Value-at-Risk using the Historical Simulation Approach. Energy Economics, 25 (2003) 239–253

Davis, R. A. and W. T. M. Dunsmuir (1997), "Least Absolute Deviation Estimation forRegression with ARMA Errors," *Journal of Theoretical Probability*, 10, 481–497.

Edgeworth, F. (1888), "On a New Method of Reducing Observations Relating to Several Quantities," *Philosophical Magazine*, 25, 184–191.

Engle R and Nelson Daniel B, "ARCH Models" in *Handbook of Econometrics*, Vol.IV (eds. Robert F. Engle and Daniel McFadden), Amsterdam: Elsevier Science B.V., 1994. Chapter 49, pp. 2959-3038.

Engle, R.F. and S. Manganelli (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles", *Journal of Business & Economics Statistics*, Vol. 22, No. 4.

George Edward Pelham Box and Gwilym M. Jenkins. Time Series Analysis: Forecasting and Control. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1994.

Glyn Holton, *Value-at-Risk: Theory and Practice*, Academic Press (2003). ISBN 978-0123540102.

Hagen. R.: How is the international price of a particular crude determined? OPEC Review 18(1994)145-158

Huang. W.. Nakamori. Y.. Wang. S.Y.: Forecasting stock market movement direction with support vector machine. Computers & Operations Research 32 (2005) 2513-2522

Huntington. H.G.: Oil price forecasting in the 1980s: what went wrong? The Energy Journal 15 (1994) 1-22

Koenker, R. and Q. Zhao (1996), “Conditional Quantile Estimation and Interence for ARCH Models,”*Econometric Theory*, 12, 793–813.

Kuper, G.: Measuring Oil Price Volatility. CCSO Working Paper, No. 2002/08, Groningen University. http://www.eco.rug.nl/ccso/quarterly/2002q2.html (2002)

Li, W.K., Ling, S. and McAleer, M. (2002) A survey of recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* 16, 245-269.

Morana, C.: A Semiparametric Approach to Short-Term Oil Price Forecasting.Energy Economics, 23(3) (2001) 325-338

Muller. K.R.. Smola. J.A.. Scholkopf. B.: Prediction lime series with support vector machines. Proceedings of In|ternational Conference on Artificial Neural Networks. Lausanne (1997)999-1004

Nelson. Y., S. Stoner, G. Gcmis. H.D. Nix: Results of Delphi VIII survey of oil price forecasts. Energy Report. California Energy Commission (1994)

Panas. E., Ninni. V.: Are oil markets chaotic? A non-linear dynamic analysis. Energy Economics 22 (2000) 549-568

Philippe Jorion, Value *at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. McGraw-Hill (2006). ISBN 978-0071464956

Sauter, R., Awerbuch, S.: Oil Price Volatility and Economic Activity: A Survey and Literature Review. IEA Research Paper, International Energy Agency, Paris, (2002)

Taylor, J. W. (2008) “Using exponentially weighted quantile regression to estimate value at risk and expected shortfall”, *Journal of Financial Econometrics*, 6,3, 382-406

Vapnik. V.N.: The Nature of Statistical Learning Theory. Springer. New York (1995)

Vladimir N. Vapnik. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

Wang. S.Y., L.A. Yu. K.K.Lai: Crude oil price forecasting with TEI@I methodology. Journal of Systems Science and Complexity 18 (2005) 145-166

Xie, W., Yu, L., Xu, S., Wang, S.: A New Method for Crude Oil Price Forecasting Based on Support Vector Machines. Lecture Notes in Computer Science, 3994 (2006) 444-451