

**MANAGING OIL PRICE RISK:
AN OBJECTIVE COMPARISON OF
VAR MODELING TECHNIQUES**

**ALEXANDRA COSTELLO
Bachelor of Commerce, University of Victoria, 2004**

**A Research Project
Submitted to the School of Graduate Studies
Of the University of Lethbridge
In Partial Fulfillment of the
Requirements for the Degree**

MASTER OF SCIENCE IN MANAGEMENT

**Faculty of Management
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA**

© Alexandra Costello, 2006

**MANAGING OIL PRICE RISK: AN OBJECTIVE COMPARISON OF
VAR MODELING TECHNIQUES**

ALEXANDRA COSTELLO

APPROVED:

Supervisor: Dr. Ebenezer Asem, PhD

Date

Co-supervisor: Dr. Eldon Gardner, PhD

Date

External Examiner: Dr. Ranjini Jha, PhD
University of Waterloo, Ontario

Date

Chairperson: Dr. Helen Kelley, PhD

Date

Abstract

This study empirically examines the performance of the Historical Simulation with ARMA forecast (C&M) methodology developed by Cabedo and Moya (2003b) vis-à-vis the performance of the Semi-Parametric GARCH methodology developed by Barone-Adesi, Giannopoulos, Kostas, and Vosper (1999). Cabedo and Moya (2003b) suggest that their model outperforms a GARCH model. However, they use an empirical distribution to forecast the future risk structure in the C&M model while they impose a normal distribution on the future risk structure in the GARCH model. This study finds that the GARCH model is not outperformed by the C&M model if the future risk structure is estimated by historical simulation as proposed by Barone-Adesi et al. (1999). Consequently, the study finds that Cabedo and Moya's (2003b) conclusion is mainly driven by the differential in forecasting the future distribution of risk rather than a deficiency in the GARCH model.

Table of Contents

Title Page	i
Signing Page	ii
Abstract	iii
Table of Contents	iv
List of Tables	v
List of Figures	vi
Introduction & Literature Review	1-6
Previous Value-at-Risk Models for Oil Prices	7-13
Methodologies Under Comparison	14-18
Methodology & Data	19-25
Results	26-44
Discussion	45-47
Practical Application	48-49
Limitations	50-52
Conclusion	53-55
References	56-57

List of Tables

1: Descriptive Statistics of Data by Year	23
2: Descriptive Statistics of Oil Prices and Returns	24
3: Ljung-Box Test of In-Sample Data	25
4: AR-HS Mean Equation Coefficients	26
5: Ljung-Box Test of AR-HS Errors	27
6: Ljung-Box Test of AR-HS Squared Errors	27
7: Summary of AR-HS Results	28
8: Likelihood Ratio Statistics for AR-HS Results	29
9: ARMA-HS Mean Equation Coefficients	31
10: Ljung-Box Test of ARMA-HS Errors	32
11: Ljung-Box Test of ARMA-HS Squared Errors	32
12: Summary of ARMA-HS Results	32
13: Likelihood Ratio Statistics for ARMA-HS Results	33
14: AR-GARCH Mean Equation Coefficients	35
15: Ljung-Box Test of AR-GARCH Errors	35
16: Ljung-Box Test of AR-GARCH Squared Errors	36
17: Summary of AR-GARCH Results	36
18: Likelihood Ratio Statistics for AR-GARCH Results	37
19: ARMA-GARCH Mean Equation Coefficients	39
20: Ljung-Box Test of ARMA-GARCH Errors	40
21: Ljung-Box Test of ARMA-GARCH Squared Errors	40
22: Summary of ARMA-GARCH Results	41
23: Likelihood Ratio Statistics for ARMA-GARCH Results	42

List of Figures

1: AR-HS Results – 95% Confidence Interval	30
2: AR-HS Results – 99% Confidence Interval	30
3: ARMA-HS Results – 95% Confidence Interval	33
4: ARMA-HS Results – 99% Confidence Interval	34
5: AR-GARCH Results – 95% Confidence Interval	38
6: AR-GARCH Results – 99% Confidence Interval	38
7: ARMA-GARCH Results – 95% Confidence Interval	42
8: ARMA-GARCH Results – 99% Confidence Interval	43

1.0 Introduction & Literature Review

1.1 Introduction

Managing risk in modern day corporations has become a task of overwhelming magnitude. For many firms, risk management must take into account huge numbers of risk factors of vastly different sources and sizes. In an increasingly complex and uncertain world, new techniques are being developed as an attempt to more effectively and efficiently manage risk. Value-at-Risk (VaR) is one of these recently developed techniques, and has quickly risen to prominence in the financial industry.

Originally the brainchild of Till Guldemann, the head of JP Morgan's global research division during the 1980s, Value-at-Risk was introduced as an alternative to the previously popular "earnings-at-risk" (Jorion, 2001). Once the term was published and seen in the G-30 report in 1993, VaR's popularity increased exponentially. First used in financial institutions to quantify market risk, VaR soon proved to be flexible enough to use in any industry or firm for a number of different purposes. Corporations have quickly progressed beyond using VaR to simply quantify risk, and have begun to control and even manage these risks within the VaR framework. Though it has drawbacks like any other risk management tool, VaR is now the most commonly seen benchmark for measuring financial risks (Jorion, 2001).

Though all companies have risk in their business environments, compared to most conventional firms, oil companies differ in one important way: their companies' major business risk is based almost solely on the price of one, single, solitary commodity: oil. Ironically, the price of oil also happens to be one of the most volatile in the world. The average volatility of crude oil prices is more than 37% a year, which is more than two and a half times that of the average US stock index and more than three times that of most major

world currencies (Jorion, 2001). Another unique property of oil is that changes in oil prices have an impact on economic activity, but economic activity does not have an impact on oil prices (Sadorsky, 1999). Oil price risk is, and should be, the foremost concern of oil companies, and oil prices have unique and often seemingly inexplicable properties.

Evidence from Giot and Laurent (2003) shows that oil prices exhibit excess kurtosis and that their volatility clusters over time. These properties seriously complicate the process of modeling oil prices. Volatility clusters are especially complex in terms of financial modeling, because they do violate the assumption of constant variance that simplifies many models. Volatility clusters imply that the variance changes over time as oil prices go through periods of exceptionally high volatility or exceptionally low volatility. These unique properties of oil prices have led to the development of various techniques to capture these features and forecast VaR more accurately.

1.2 VaR Models

Specifically defined, “VaR summarizes the expected maximum loss (or worst loss) over a target horizon within a given confidence interval” (Jorion, 2001). Hence, the “value at risk” in a specific time period, for example a week, can be summed up with one clear number, for example \$10 million, at a given confidence level. The confidence level refers to the statistical probability that the VaR limit will be exceeded and corresponds to a specific point on the tail of the risk factor’s distribution. For example, at a 99% confidence level, there is only a one in a hundred chance that the losses would exceed the VaR in the relevant period. With this one simple, clear number, a risk manager or senior executive can do a number of important things.

Risk managers can compare alternative investment decisions based on the “worst-case scenario” associated with each. It is important to remember that the VaR number is *not* the worst possible loss; a 99% confidence level is very high, but it does not mean that the VaR cannot be exceeded. Empirical tests of portfolio risk have shown that when the VaR is exceeded, even if it only happens once in a hundred time periods, portfolio losses are on average 30 to 40% higher than predicted by VaR at the 99% confidence level (Hendricks, 1996). At the 95% level, the user would expect the severity of violations to be even larger. Besides using VaR to estimate potential losses, managers can use VaR numbers to compare investments’ trade-offs between risk and return. However, similar to beta measures of investment risks, having a higher VaR is not necessarily worse if the corresponding returns for the investment are significantly higher. Alternatively, senior managers can use VaR to set limits on the amount of risk taken on by various divisions. In all these ways and more, VaR can be used as an objective, unambiguous means of measuring, controlling, and managing a wide variety of risks.

VaR is only one of many available methods of quantifying potential financial losses, and within the VaR framework, there are also various alternative methods of forecasting. It is not enough to find a model and use it. Model risk, described as the danger that a model is mis-specified or that the model parameters are incorrect, can be just as dangerous as not using a model at all (Jorion, 2001). Practitioners often place great amounts of faith in their financial models and make large monetary commitments based on the information that their models provide. If a particular model is flawed, or is not as accurate as it could be, then model risk would expose the asset (or portfolio) to risk levels that are higher than expected. This divergence of actual risk and expected, resulting from model risk, undermines risk

management and control. Consequently, the search for appropriate modeling techniques and model comparisons is an area of active research.

To minimize (or avoid) model risks, back-testing and comparisons of models are commonplace (Bera and Higgins, 1993; Cabedo and Moya, 2003a; Cabedo and Moya, 2003b; Hendricks, 1996; Morana, 2001). However, the results of these empirical tests are only as useful as the analyses used to obtain them. When the analysis or test is flawed, the choice of model can be incorrect, and model risk is once again a danger. In order to minimize model risk as much as possible, the models must be correctly specified and formulated in the first place, and in the second place, the testing of those models must be objectively and systematically carried out. It is also important to properly compare newly developed models to existing, competing models to see if either the existing models remain superior or the newly developed models will outperform them; this is clearly the main objective of this study.

1.3 Extant Literature

Previous studies, though not many, have attempted to model oil prices with varying methodologies, resulting in different degrees of success. Among these, a few studies used linear regression methodologies to attempt to delineate the relationship of oil prices with various other factors, such as stock market indexes (Faff and Brailsford, 1999; Sadorsky, 1999). Other studies use various market models, including the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, to analyze the properties of oil derivatives and their effects on energy markets (Moosa and Al-Loughani, 1994; Panas and Ninni, 2000; Morana, 2001). While all of these studies provide relatively conclusive results and manage to show statistically that their models fit the data sufficiently well, very few studies have

compared VaR models directly (Hendricks, 1996). In fact, only one previous study compares VaR models using oil price data to see which model best fits the unique characteristics of oil prices; Cabedo and Moya (2003b) directly compare the standard parametric GARCH model and their newly developed Historical Simulations with ARMA forecasts (to be called C&M method hereafter) model. Cabedo and Moya find that the GARCH methodology is inferior to their model. This study examines whether their result is driven by the normality assumption imposed on the future structure of oil prices in the GARCH model. An improved comparison between the two models will aid in the quest for more accurate modeling of oil prices.

From the existing literature, it appears that the semi-parametric GARCH methodology developed by Barone-Adesi, Giannopoulos, & Vosper (1999) should have good potential for accurately forecasting oil's VaR. Their methodology not only incorporates more information from the historical data by adjusting for historical and forecasted variance, but it allows the relaxation of the assumption of constant variance that restricts the C&M model. Morana (2001) found that the semi-parametric GARCH(1,1) fits oil price data well, and produces superior forecasts to the standard historical simulation model. As previously mentioned, the C&M methodology surprisingly appears to outperform the Semi-Parametric GARCH methodology in their comparison. This researcher will discover if this result is replicated under a second, more thorough test, as no previous study has yet provided a sufficiently objective and thorough comparison of these two VaR models. To this end, each of the potential flaws in the methods of the existing literature is examined in turn, in order to clearly show the benefits and contribution of this study. This study then attempts to fill in the gaps in the existing research and rectify any inconsistencies in their methodologies and analyses. It is anticipated that a definitive answer will be supplied as to which methodology

can most accurately forecast the VaR distributions of oil. This study contributes to this literature by comparing the VaR forecasting ability of the Semi-Parametric GARCH model to the recently developed C&M under specific and equalized conditions.

2.0 Previous VaR Models For Oil Prices

Forecasting VaR involves important ingredients that the researcher or the practitioner must deal with, including the statistical distribution underlying the risk factor, the process describing the mean and the variance of the risk factor, the sample period, the forecasting range, and the confidence interval. The following subsections describe these important factors in forecasting VaR.

2.1 Forecast Period

In model comparisons, the researcher must decide on the sample period as well as the forecasting period. Data in the sample period is used to estimate the model and the VaR is predicted over the forecasting period. The performances of the VaR's of different models are then compared over the forecasting period to determine if a particular model dominates another. Previous studies, like Cabedo and Moya (2003b) and Morana (2001), which examine oil price, appear to be limited in that most of the available data is used to estimate the model and construct the historical distribution, leaving a very short forecasting period. In Cabedo and Moya, the forecasting period is only a year, and Morana only forecasts two months. These forecasting periods are not long enough to draw valid inferences of VaR performance. For an illustration of this, at a 99% confidence level, the expected number of violations in a year (250 trading days) is only 2.5. Consequently, if the performance of the two models is close to the expected level, it is difficult to draw any reliable conclusions.

This study attempts to perform a more thorough analysis by using historical distributions in both Semi-Parametric GARCH and C&M forecasts, and forecasting for a much longer time period. A longer forecast period allows a better comparison of the performance of the two methodologies. This study uses data covering an 18-year period. The first five years are

used as the initial sample to forecast the VaR for the next observation. After this, the sample period is rolled forward by one observation and the models are re-estimated to forecast the next VaR (also see Barone-Adesi et al. (1999)). The dynamic forecasting performed by Barone-Adesi et al. is more practical and likely more accurate than the static forecast of Cabedo and Moya because a risk manager can update his or her sample daily to forecast the next VaR. In this way, the data used to estimate the mean equation and forecast the VaR is constantly updated and no out of date data is used. Thus, an out-of-sample period of 13 years is available to assess the forecasting performance of the two competing models. This rolling forecast ensures that the five-year historical distribution is as up-to-date and accurate as possible, and with thirteen years of forecasts to compare the two models, a thorough comparison using a range of criteria should provide conclusive results.

2.2 Statistical Distribution

An accurate forecast of VaR requires an accurate forecast of the future risk structure. Thus, the results of VaR models depend largely on how the future distribution of the risk factor is captured or modeled. The simplest way of modeling the distribution of a risk factor is to use a basic normal distribution. However, some researchers have attempted to use alternative distributions and have found that they can increase the accuracy of forecasting models. These alternatives include the Student- t distribution, which partially accounts for the fatter tails of the oil price distribution (Bollerslev, Chou, & Kroner, 1992). Another potential alternative is to use the historical data of the risk factor to model a distribution that directly reflects any unique qualities of the data and avoids misspecifying a forecast model by assigning an arbitrary distribution. The C&M model has good potential to model oil prices due to its use of this historical simulation technique. However, their direct comparison of

this model to the standard GARCH model is inappropriate since under the C&M, the future structure of oil price is historically simulated while in their GARCH model, normality is assumed. Due to the fact that oil prices display characteristics of non-normal distributions, this creates an uneven basis for comparison and may lead to distorted results. As the historically-simulated GARCH model has already been developed by Barone-Adesi et al. in 1999, it is only appropriate to compare these two models using the historically-simulated future structure of oil prices in both cases. This study carries out this task.

An additional contribution of this study is that it is the first study to apply the historically-simulated Semi-Parametric GARCH model to forecasts of VaR for oil prices. Although Morana (2001) applies the Semi-Parametric GARCH to oil prices, his study focuses on evaluating the accuracy of forward oil prices as forecasters of future oil prices. In addition, Morana does not use an ARMA mean equation in the Semi-Parametric GARCH, as Barone-Adesi et al. (1999) recommend. This direct comparison using the historical simulation to predict the future risk structure with ARMA forecasts (C&M method) and historical simulation to predict the future risk structure for the GARCH forecasts (Semi-Parametric GARCH) fills this gap.

2.3 Mean Equations

Another key issue in testing VaR models is the use of mean equations. As stated earlier, Barone-Adesi et al. (1999) advocate the use of Autoregressive Moving Average (ARMA) conditional mean equations in modeling time-series data. Correspondingly, in Cabedo and Moya's (2003b) analysis, ARMA equations are shown to greatly increase the accuracy of standard Historical Simulation forecasts. An ARMA process contains an autoregressive term that incorporates information about the most recent period's return, as well as a moving

average term that contains information about the previous period's error term. The process can be written as:

$$r_t = \alpha + \beta r_{t-1} + \varepsilon_t - \sigma \varepsilon_{t-1} \quad (1)$$

where the term r_t is the return at time t , the α is a constant, ε_t is the error at time t , and β and σ are coefficients of r_{t-1} and ε_{t-1} , respectively. The term "AR" refers to *autoregressive* and is represented by the βr_{t-1} part of the above equation. The "MA" term is the *moving average* term, and is captured by $\sigma \varepsilon_{t-1}$ in the mean equation.

Since Cabedo and Moya (2003b) only used the AR mean equation and not the ARMA mean equation in conjunction with the GARCH model, it is impossible to conclude whether the C&M model outperforms the GARCH model as a result of the advantages of the methodology or the use of different mean equations. Therefore, this analysis conducts tests of the accuracy of the C&M and Semi-Parametric GARCH models using ARMA mean equations in both cases, in order to provide an equitable comparison. In addition, both models are also analyzed using an autoregressive mean process in order to clearly delineate the added benefit of adding the MA term to the mean equation. This alternative and simplified mean equation is shown below:

$$r_t = \alpha + \beta r_{t-1} + \varepsilon_t \quad (2)$$

Using the same mean equations will ensure the differences in the performance of the two models are not driven by differentials in the mean equations.

2.4 Volatility Equation

The key difference of the GARCH model compared to the model developed by Cabedo and Moya (2003b) is that the GARCH model captures time-varying volatility inherent in many financial times series data. As previously mentioned, there is evidence from Giot and Laurent (2003) that the volatility of oil prices clusters over time. Volatility clusters are especially complex in terms of financial modeling, because they do violate the assumption of constant variance that simplifies many models. Volatility clusters imply that the variance changes over time as oil prices go through periods of exceptionally high volatility or exceptionally low volatility. Since the GARCH model accounts for the varying volatility, it may provide superior results to those of the C&M model, which assumes constant variance.

Specifically, in GARCH models, the errors are heteroskedastic over time and are usually represented by Equation (3).

$$\varepsilon_t \sim D(0, h_t) \quad (3)$$

where ε_t has a distribution D with variance h_t , defined in Equation (4) below:

$$h_t = \varpi + \delta \varepsilon_{t-1}^2 + \theta h_{t-1} \quad (4)$$

where ϖ is a constant, δ is a coefficient, ε_{t-1} the lagged error, and θ is the coefficient of the last period's volatility, which is denoted by h_{t-1} . Consequently, GARCH models capture changing volatility through changing h_t , which is influenced by the currently shocks to oil prices ε_{t-1} . As a result, the assumed distribution of the risk factor widens during periods of great volatility and contracts during tranquil periods. By constantly adjusting to the time-varying volatility of the oil price returns, the Semi-Parametric GARCH model makes use of all of the available information contained in the historical returns, as well as emphasizing the information from the most recent return, error, and variance. Thus, provided oil prices

display changing volatility over time, the GARCH technique should deliver more appropriate VaR forecasts than the C&M method.

2.5 Confidence Interval

Another important ingredient in forecasting VaR is the confidence level that the researcher or the practitioner uses in estimating the VaR. In practice, a 95% or 99% confidence level is the most common (Jorion, 2001); the actual confidence level chosen is a matter of preference and would depend on the use and requirements of each individual organization.

In Morana (2001), an 80% confidence level is arbitrarily chosen, which does not reflect the levels used commonly in practice and may not accurately model the unique tails of these time-series distributions. Many previous studies of time-series data note the distinctly fatter tails of time-series distributions (Bera & Higgins, 1993; Bollerslev et al., 1992; Hendricks, 1996), and Panas and Ninni (2000) note the non-normal, leptokurtic characteristics of oil products on the European market. A 99% confidence interval, therefore, may better capture the unique qualities of these tails, and in VaR analyses it is the tails of the distributions that are of vital importance. Cabedo and Moya (2003b) use a 99% confidence level, which takes into account the tails of the distribution and conforms to the level chosen by the Basel Committee for back-testing (Jorion, 2001). However, a 99% level has its own drawbacks, such as the fact that an overly high confidence level does not provide many instances where the VaR is exceeded, as theoretically only one in a hundred data points would exceed the VaR. Consequently, it is difficult to compare VaR models at 99% confidence level since a meaningful comparison will require a large amount of data. A 99% confidence level also results in high VaRs, which is not always desirable to practitioners who must base capital

requirements on these forecasts, and explains why 95% confidence levels are often used in finance (Jorion, 2001). In this study, both the 95% and 99% confidence levels are used, as both levels have different qualities that are useful in a comparison of VaR modeling techniques and both are commonly used in practice.

Although both the upper and lower limits of the confidence intervals are discussed as though they are a pair, in reality the position in an asset (or portfolio) determines whether the upper or lower limit is of interest; an oil producer, for example, would have a long position in oil and would therefore be solely interested in price declines (the lower tail of the price distribution). The lower limit would define how much of the value of his or her inventory is at risk. Alternately, a heavy user of oil, such as a refiner, would have a short position in oil and would be adversely affected by upward price movements. Therefore, a refiner would be exclusively interested in the upper tail of the price distribution. For these reasons, the analysis in this study will look at both the upper and lower portions of the confidence intervals separately.

3.0 Methodologies Under Comparison

3.1 Historical Simulation Methodology

Historical simulation (HS) is often used to model VaR because many researchers and practitioners believe it is potentially one of the most accurate modeling techniques (Hendricks, 1996). HS models are shown in more than one study to outperform almost any other type of VaR modeling technique (Cabedo and Moya, 2003b; Hendricks, 1996). However, historical simulation can be time-consuming, and if the necessary data are not routinely gathered in-house, it can be difficult to obtain enough data from outside sources. On the other hand, if the information is gathered as part of a daily routine, running the simulation and forecasting the VaR for the next period can be done quite quickly.

The basic process to forecast VaR using historical simulation involves the formation of a distribution from historical return. This distribution is formed by ranking the actual returns that have occurred over the in-sample period. Once the historical distribution is formed, the VaR is simply taken as the return that falls at the 1st, 5th, 95th, or 99th percentile, depending on the confidence interval used. The methodology developed by Cabedo and Moya in 2003 is more advanced, and involves first parametrically estimating the ARMA mean equation and generating the residuals from the model. The residuals are ordered, the appropriate percentile is fed into the estimated mean equation to produce the VaR forecast. By including the ARMA mean equation, the C&M method can have high levels of accuracy. However, their method is restricted by the assumption of constant variance. Therefore, the GARCH model will potentially generate better forecasts due to its ability to incorporate time-varying volatility in the forecasts.

Specifically, Cabedo and Moya first estimate the ARMA model using the Box-Jenkins' methodology and generate the errors from the model (ε_t for $t=1, \dots, n$, where n is the sample

size). These errors are ranked and 1st, 5th, 95th, and 99th percentiles (say $\varepsilon_1, \varepsilon_5, \varepsilon_{95}, \varepsilon_{99}$, respectively) of the distribution are used in the estimated mean equation, to predict the returns at the 1st, 5th, 95th, and 99th percentiles, respectively. For instance, the VaR at 5% in the ARMA model is forecasted as:

$$r_5 = \hat{\alpha} + \hat{\beta}r_{t-1} - \hat{\sigma}\varepsilon_{t-1} + \varepsilon_5 \quad (5)$$

where $\hat{\beta}$ denotes the sample estimate of β , and so on. In this way, not only is the historical information incorporated through the constant and the error term, but the most recent information is incorporated through the last day's return and error and its relative importance is adjusted through the parametrically-estimated coefficients.

3.2 Semi-Parametric GARCH Methodology

GARCH models are a generalized version of the ARCH models originally developed by Engle in the early 1980s. Bollerslev (1986) introduced the generalized version of Engle's ARCH model, the GARCH, and many subsequent studies celebrate the GARCH model for its parsimony and flexibility (Moosa and Al-Loughani, 1994; Morana, 2001). In fact, Bera and Higgins show in their 1993 study that a low-order GARCH model (GARCH(1,1)) models time-series data equally as well as a high-order ARCH model (ARCH(6)). Since most model selection criteria take into account the number of factors that must be specified in the model, the GARCH model has a distinct advantage in its simplicity.

GARCH models allow the forecaster to relax the assumption that the variance of the data-generating process is constant. GARCH models incorporate a term that contains information about the variance of the previous period and therefore allow the variance to be adjusted based on what actually happened in the last relevant period. Additionally, the use

of a non-parametric historical distribution allows the forecaster to relax the assumption that the future distribution of the risk factor is normal. Previous studies show that the standard GARCH models can accurately forecast oil price changes (Sadorsky, 1999). Consequently, a GARCH model with a non-parametric historical distribution may have the potential to provide superior VaR forecasts relative to the alternative available techniques, and the C&M model in particular. Barone-Adesi et al. (1999) show that simulating the future distribution in GARCH setups delivers better VaR forecasts than the standard GARCH method.

There are three basic ways to estimate VaR using GARCH models. The standard approach involves imposing a normal distribution on the data, as was done in Cabedo and Moya (2003b). Since oil prices have been shown in previous studies to have a distinctly non-normal distribution, it is not surprising that the standard GARCH model performs comparatively poorly in Cabedo and Moya's analysis. The second possible way to forecast VaR using a GARCH model is by imposing an alternative, non-normal distribution on the data, such as a Student-t distribution. While this method is shown in previous studies to be an improvement on the normal distribution (Bera and Higgins, 1993; Bollerslev et al., 1992), it still imposes an arbitrary distribution on the data that may not accurately capture the empirical characteristics of the data. In contrast, the third method, introduced by Barone-Adesi et al. in 1999, uses a non-parametric distribution that is formed from historical data to forecast the future distribution and the VaR. In this way, a GARCH model using a historical distribution makes use of important sample information to forecast the future risk structure. Consequently, this technique will potentially circumvent one of the most troublesome properties of oil prices, and time-series financial data in general, that Cabedo and Moya's approach cannot account for: the tendency to have volatility clustering (Bera and Higgins, 1993; Bollerslev et al., 1992; Hendricks, 1996; Morana 2001).

The Semi-Parametric GARCH methodology proposed by Barone-Adesi et al. (1999) is an important step in VaR modeling because it alleviates the necessity of imposing an arbitrary, theoretical distribution on the data and does not require the use of correlation matrices, which quickly becomes burdensome when modeling more than one factor. The methodology uses a combination of parametric and non-parametric techniques. The first stage involves the parametric estimation of the coefficients of the ARMA or AR mean equation and the variance equation, shown above as Equations (1) or (2) and Equation (4) respectively.

This variance equation is the key difference between the Semi-Parametric GARCH and the C&M methodologies. In the C&M methodology, the errors are ranked, and the 1st, 5th, 95th, and 99th percentiles of the distribution are fed directly into the mean equation, shown as Equation (1) or (2) above. However, in the Semi-Parametric GARCH, the errors are standardized and adjusted to reflect current volatility condition before they are ranked and the 1st, 5th, 95th, and 99th percentiles are chosen. Since the residuals generated from the estimation of the mean equation are not independently and identically distributed, they are divided by the of the corresponding day's standard deviation, as shown in Equation (6) below, in order to normalize them.

$$e_t = \frac{\varepsilon_t}{\sqrt{h_t}} \sim D(0, h_t) \quad (6)$$

The standardized errors are then independently and identically distributed under the GARCH assumptions. These standardized errors (e_t , $t = 1, \dots, n$, where n is the sample size) are adjusted by next period's standard deviation forecast ($\sqrt{h_{n+1}}$) to adapt them to current volatility condition as follows:

$$s_t = \varepsilon_t \sqrt{h_{n+1}}, \quad \text{for } t = 1, \dots, n \quad (7)$$

The error has been adjusted to the current period's forecasted variance, which incorporates both historical information about the variance over the past five years through the constant ϖ in Equation (4), and more recent information about the most recent volatility and amount of surprise in the market through the terms h_{t-1} and ε_{t-1} in Equation (4).

As with the C&M method, the adjusted standardized errors are ranked and the 1st, 5th, 95th and 99th percentiles (say, s_1 , s_5 , s_{95} , s_{99} , respectively) are used in the estimated mean equation to predict the returns at the 1st, 5th, 95th, and 99th percentiles, respectively. For instance, the VaR at 5% in the ARMA model is forecasted as:

$$r_5 = \hat{\alpha} + \hat{\beta}r_{t-1} - \hat{\sigma}\varepsilon_{t-1} + \varepsilon_5 \quad (8)$$

In sum, the process of adjusting and adapting the VaR forecasts to incorporate additional information about the historical and recent variance of the oil price returns may result in the Semi-Parametric GARCH model providing superior forecasts to those of the C&M model.

4.0 Data & Methodology

4.1 Testing Methodology

Eight comparisons are actually done, in order to form the most accurate picture of the forecast properties of both models. The C&M and Semi-Parametric GARCH forecasts are compared at a 95% confidence level and a 99% confidence level, using both the AR and the ARMA mean equations. The first step of the analysis is to estimate the Semi-Parametric GARCH and C&M model specifications, using the first five years of data. The error terms from the estimation process are saved to form the historical distributions for the second step of the analyses. These non-parametric distributions are used in both forecasts. Once the mean equation and, in the case of the GARCH model the variance equation, are estimated and the distribution is formed, a VaR forecast is estimated for the next day using either the AR or ARMA mean equation. This procedure is reiterated daily using a rolling 5-year sample for the next thirteen years. Five comparison criteria are considered to determine the performance of the models.

4.2 Comparison Criteria

The five comparison criteria used to evaluate the VaR models are:

1. Actual vs. expected number of violations
2. Severity of violations
3. Summed differences between actual and forecasted values
4. Number of consecutive violations
5. Conditional Coverage Likelihood Ratio (Christoffersen, 1998).

The first criterion compares the total number of times that the oil price exceeds its VaR forecast to the number of times that it was expected to do so. For example, 5 out of 100 times the actual price should be greater than the upper limit of the 95% confidence interval, but in actuality it may only be greater 4 times or 7 times. The second criterion deals with the severity of the violations, when they occur. This criterion is calculated by dividing the magnitude of the VaR exceedence into the limit and multiplying the ratio by 100. This criterion is important because although one model may have exactly the expected number of violations, those violations may be severe. The third criterion simply shows how close the VaR are to the actual prices. It would conceivably be possible to draw two straight lines as a confidence interval and have the same number and severity of violations as a continually changing forecast. However, the summed differences would show that the constantly adapting forecasts were much closer over time to the actual values, a result which is extremely important in as it indicates that the forecasts are not being overestimated more than necessary. A banker who is using VaR to set his or her cash requirements will not wish the VaR to be any higher than it must, as holding greater reserves of cash is very costly. The fourth criterion, the number of consecutive violations, is also critical because it shows not only how the methodology deals with periods of greater volatility, but how independent the violations are. The fifth criterion, the conditional coverage likelihood ratio statistic (LR_{CC}), developed by Christoffersen in 1998, also examines independence using statistical inference.

The conditional coverage likelihood ratio is actually made up of two ratios, the unconditional coverage likelihood ratio (LR_{UC}) and the independence likelihood ratio (LR_{IND}). These three likelihood ratios make up a structured system for formally evaluating interval forecasts. The unconditional coverage ratio examines the likelihood that the interval forecast provides sufficient coverage, based on the number of expected and actual violations.

The statistic produced can be compared to a χ^2 distribution with one degree of freedom to see if the ratio is significant at the appropriate level of test. For instance, if the LR_{UC} test statistic is 3.246 and the χ^2 distribution with one degree of freedom at the 5% level is valued at 3.841, then the null hypothesis that the model does not provide statistically significant unconditional coverage can be rejected with 95% confidence. The formula is shown below.

$$LR_{UC} = -2 \log \frac{(1-p)^{n_0} p^{n_1}}{(1-\pi)^{n_0} \pi^{n_1}} \quad (9)$$

where p is the percentage of expected violations, π is the actual percentage of violations, n_0 is the number of times the limit was not violated, and n_1 is the number of times the limit was violated.

In contrast to the unconditional coverage likelihood ratio, the independence ratio statistic tests whether violations occur independently or cluster at some periods. This independence is measured by the number of consecutive violations. A consecutive violation is an instance where a violation on the upper or lower level is immediately followed by another violation. This factor is extremely important because it displays whether or not the model can adapt quickly enough to a large increase in volatility. Like the unconditional coverage statistic, the independence statistic is also compared to a χ^2 distribution with one degree of freedom to see if the ratio is significant at the 5% level. The formula for the independence ratio is shown below.

$$LR_{IND} = -2 \log \frac{(1-\pi_2)^{(n_{00}+n_{10})} \pi_2^{(n_{01}+n_{11})}}{(1-\pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1-\pi_{11})^{n_{10}} \pi_{11}^{n_{11}}} \quad (10)$$

where π_2 is the actual percentage of total violations, π_{01} is the actual percentage of non-consecutive violations, π_{11} is the actual percentage of consecutive violations, n_{00} is the number of times the VaR was not violated non-consecutively or the total number of

forecasts less n_{0i} , n_{10} is the number of times the limit was not violated consecutively, n_{0i} is the number of times the limit was violated non-consecutively, and n_{1i} is the number of times the limit was violated consecutively.

Christoffersen (1998) then states that the two statistics can be summed, as shown below, and compared to a χ^2 distribution with two degrees of freedom to see if the conditional coverage, which includes both the factor of unconditional coverage and independence, is significant at the 5% level.

$$LR_{CC} = LR_{UC} + LR_{IND} \tag{11}$$

These five comparison criteria should provide thorough and multi-faceted means through which the two VaR forecasting methodologies can be compared.

4.3 Data

This study uses data on daily Brent Crude oil prices for the period from May 20th, 1987 to January 18th, 2005. The data are obtained from the Energy Information Administration (EIA) of the US Department of Energy. This time period consists of all available data points from the EIA, in order to perform the most thorough analysis possible. In addition, the use of daily prices ensures that a large number of data points are available; in total, there are 4495 observations. The first five years of data, from May 20th, 1987 to May 19th, 1992, are used to estimate the mean equation, and the errors are saved in order to form the historical distribution from which the forecasts are drawn. After this, the sample is rolled forward by a day and the parameters are re-estimated to forecast the next day's VaR. This process is replicated until the VaR for all the remaining observations are forecasted (in total, the VaRs for 3,206 data points are forecasted from May 20th 1992 to January 18th, 2005). To

estimate the parameters and the VaR, the oil prices are transformed into continuous return calculated as $\log(P_t/P_{t-1})$.

The following table displays the mean, median, and standard deviation of both the daily oil prices and daily returns over the entire in-sample and out-of sample periods. It is evident that some years have much greater volatility than others, and that the average price of oil changes drastically from year to year.

Table 1: Descriptive Statistics of Data by Year

Year	Obs	PRICE			RETURN		
		Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
1987	160	\$18.53	\$18.60	\$0.97	-0.03%	0.00%	1.56%
1988	255	\$14.91	\$15.13	\$1.47	0.03%	0.00%	2.38%
1989	254	\$18.23	\$18.10	\$1.22	0.13%	0.12%	1.66%
1990	256	\$23.76	\$20.57	\$7.73	0.18%	0.17%	3.49%
1991	257	\$20.04	\$19.70	\$1.94	-0.12%	-0.11%	3.38%
1992	257	\$19.32	\$19.48	\$1.13	0.05%	0.04%	1.33%
1993	252	\$17.01	\$17.00	\$1.53	-0.10%	0.00%	1.37%
1994	252	\$15.86	\$16.08	\$1.40	0.03%	0.14%	1.74%
1995	253	\$17.02	\$16.85	\$0.92	0.04%	0.18%	1.27%
1996	254	\$20.64	\$20.05	\$2.27	0.11%	0.08%	2.19%
1997	248	\$19.11	\$18.83	\$1.78	-0.11%	-0.06%	1.73%
1998	253	\$12.76	\$12.61	\$1.57	-0.12%	-0.41%	3.10%
1999	249	\$17.90	\$17.55	\$5.02	0.35%	0.35%	2.45%
2000	253	\$28.66	\$28.86	\$3.39	-0.02%	0.10%	2.76%
2001	257	\$24.46	\$25.44	\$3.41	-0.02%	-0.03%	2.96%
2002	255	\$24.99	\$25.49	\$2.94	0.21%	0.23%	2.18%
2003	258	\$28.85	\$28.80	\$2.48	0.03%	0.15%	2.21%
2004	261	\$38.26	\$37.60	\$5.63	0.14%	0.00%	2.36%
2005	11	\$43.77	\$43.75	\$1.60	1.07%	0.40%	2.47%
Total	4495	\$21.28	\$19.05	\$6.95	0.05%	0.00%	2.35%

Table 2, shown below, summarizes the descriptive statistics of the daily oil prices and daily returns for both the in-sample and out-of-sample periods. It is immediately clear that the volatility of oil prices is high, as both the prices and returns have large standard deviations, and some years were exceptionally volatile, like 1990.

Table 2: Descriptive Statistics of Oil Prices and Returns

Statistics	Price (US\$)		Return (%)	
	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
Mean	19.09394	22.15556	0.04445%	0.04897%
Std Dev	4.66248	7.49915	2.63338%	2.23505%
Skewness	2.24667	1.02756	-0.90636	0.07661
Kurtosis	6.21566	1.12799	20.62335	4.75760
1 st Percentile	12.20000	10.39050	-6.96947%	-5.85795%
5 th Percentile	13.98700	12.35250	-3.41551%	-3.50585%
95 th Percentile	30.39050	36.42000	3.67962%	3.54943%
99 th Percentile	38.36450	45.45200	7.79377%	5.83693%
Jarque-Bera	3156.9191	734.1610	22984.14127	3026.75762

In addition, it seems from Table 1 and Table 2 that the volatility of oil prices is growing slightly over time; as oil prices rise, there is more room for movement and a greater range of prices are possible. Comparing the Jarque-Bera test statistics to a χ^2 distribution with two degrees of freedom at the 5% level, which has a value of 5.991, suggests that the distribution of oil prices or returns are non-normal in both the in-sample and out-of-sample periods. Also, the skewness and kurtosis of the oil prices and returns are extremely non-normal, as both in-sample and out-of-sample prices have skewness and kurtosis with p-values of 0.00. The distributions of oil prices and returns appear to be positively skewed, indicating that there are more likely to be positive returns than negative, a conclusion which is supported by the fact that oil prices are generally rising over time. In addition, the kurtosis of the distribution is high, meaning that it is more “peaked” than a normal distribution. However, this kurtosis may be offset by the potential fat-tails of the distribution and likely does not mean that the distribution has a smaller standard deviation than a normal distribution. This fact adds further weight to the argument that a historical, non-parametric distribution may provide superior forecasts to a normal distribution by incorporating these clues about oil price returns into the projected risk factor’s distribution.

Table 3: Ljung-Box Test of In-Sample Data

Considered Lag	Ljung-Box Test Statistic
Q(12)	741.9094*
Q(24)	823.8779*
Q(36)	950.8926*

*Significance level under 0.05

It is clear from Table 3 above that the in-sample squared returns display statistically significant autocorrelation. This result provides weight to the argument that the GARCH model may be able to provide superior forecasts to the ARMA-HS model since GARCH models capture changing volatility but the ARMA-HS does not. This not only indicates that the GARCH model may better deal with the increasing volatility of oil prices, but that more accurate and flexible forecasting methods will become key as oil markets become even less predictable.

5.0 Results

The results of this analysis are quite consistent with expectations. In all of the models, the 95% and 99% confidence intervals are quite similar; the 99th percentile's band is simply slightly wider than the 95ths. However, there is evidence from the comparison criteria that some of the models forecast more efficiently with either the 95% or 99% confidence interval. The following four sections discuss each of the forecasting methods' results in turn, and the discussion section compares and contrasts the various results.

5.1 Historical Simulation with AR Forecast Results

Table 4 below shows the estimate of the constant and the AR coefficient of the Historical Simulation with Auto-Regressive Forecast (AR-HS) forecast mean equations. The constant is extremely small, and likely indicates that the returns are slightly more likely to be positive than negative, on any given day. The AR coefficient is also quite small, which suggests that the previous day's return gave very little clue as to what today's return would be.

Table 4: AR-HS Mean Equation Coefficients

Coefficient	Mean Value	SE of Mean	T-Stat
Constant	0.0001332194	5.421401e-06	24.57288*
AR Coefficient	0.0512477223	0.000314	163.13092*

*Significance level under 0.05

The Historical Simulation with the Autoregressive mean equation (AR-HS) performs less than adequately in this analysis. The AR-HS model shows statistically significant autocorrelation in both the Ljung-Box test of the errors and that of the squared errors. The results are summarized below in Tables 5 and 6. The raw residuals, shown in Table 5, exhibit autocorrelation, which suggests the AR mean equation does not completely remove

the autocorrelation of the returns. In addition, the information in Table 6 indicates that the square errors, and thus the higher order moments, are also autocorrelated. This fact suggests that the GARCH may be able to provide a superior forecast by dealing more effectively with the autocorrelation of the higher-order moments.

Table 5: Ljung-Box Test of AR-HS Errors

Considered Lag	Ljung-Box Test Statistic	Significance Level
Q(12)	32.0604*	0.00135440
Q(24)	61.0355*	0.00004555
Q(36)	74.6588*	0.00016241

*Significance level under 0.05

Table 6: Ljung-Box Test of AR-HS Squared Errors

Considered Lag	Ljung-Box Test Statistic	Significance Level
Q(12)	152.7191*	0.00000000
Q(24)	165.6505*	0.00000000
Q(36)	192.3672*	0.00000000

*Significance level under 0.05

The results of the Value-at-Risk forecasts are summarized in Table 7, below. The actual number of violations is fairly close to the expected number, although it is interesting to point out that there are significantly more violations of the two lower limits than the two upper limits. Since the summed differences are also smaller for the upper half of the confidence interval than the lower, it seems that the upper half of the forecasted interval is consistently closer to the actual returns. This is a positive facet of the model; a practitioner would prefer to forecast a smaller confidence interval and know that it is less likely to be violated. Not surprisingly, the degree of violation is slightly higher for the two lower percentiles than for the upper two.

Table 7: Summary of AR-HS Results

%tile	Expected # of Violations	# of Violations	Degree of Violation	Consecutive Violations	Summed Differences
1%	32.05	38	74.14%	3	200.45
5%	160.25	182	46.18%	22	111.40
95%	160.25	182	49.28%	15	108.62
99%	32.05	36	39.39%	1	191.55

Christoffersen (1998) provides an excellent means for formally comparing interval forecasts using Likelihood Ratio statistics. These are summarized below in Table 8 for the AR-HS model. These ratios allow the researcher to look at the results in Table 7 and test whether the obtained violations and the conservative violations are significantly different from their expected values. For instance, someone may look at the difference between the expected number of violations and the actual number of violations and think that the number of violations looks too high. However, Christoffersen's Unconditional Coverage Likelihood Ratio (LR_{UC}) offers a formal way to test whether the 182 actual violations are statistically different from the expected violations at some significance level (usually 5%). Similarly, the researcher may be unable to judge whether the number of consecutive violations produced by the AR-HS model is acceptable.

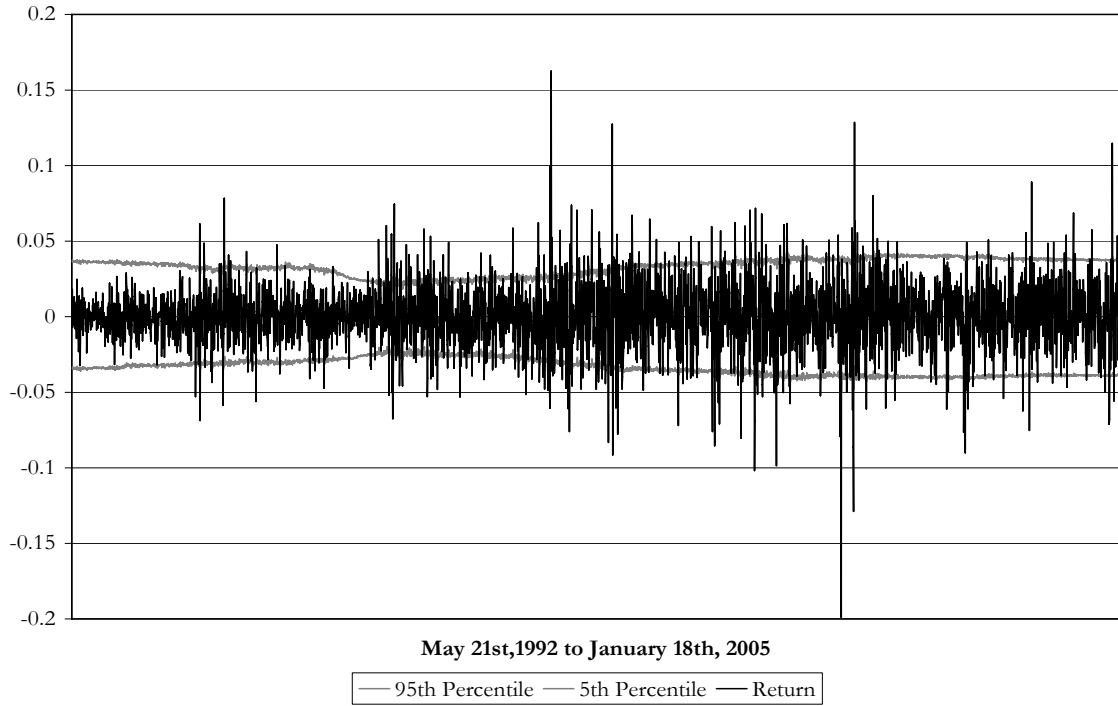
The Independence Likelihood Ratios (LR_{IND}) suggest that the number of consecutive violations at the 1st, 5th, and 99th confidence levels are too high, and that the AR-HS is not producing sufficiently independent forecasts at these confidence levels. Not surprisingly, when the two ratios are summed to form the Conditional Coverage Likelihood Ratio (LR_{CC}), it becomes apparent that the conditional coverage is not sufficient at any of the percentiles but the 95th. Therefore, it is concluded that the AR-HS delivers too many consecutive violations of the VaR on the whole, although the number of actual violations at all four percentiles is acceptably close to the expected number. Overall, the AR-HS model does not forecast the VaR of oil prices adequately.

Table 8: Likelihood Ratio Statistics for AR-HS Results

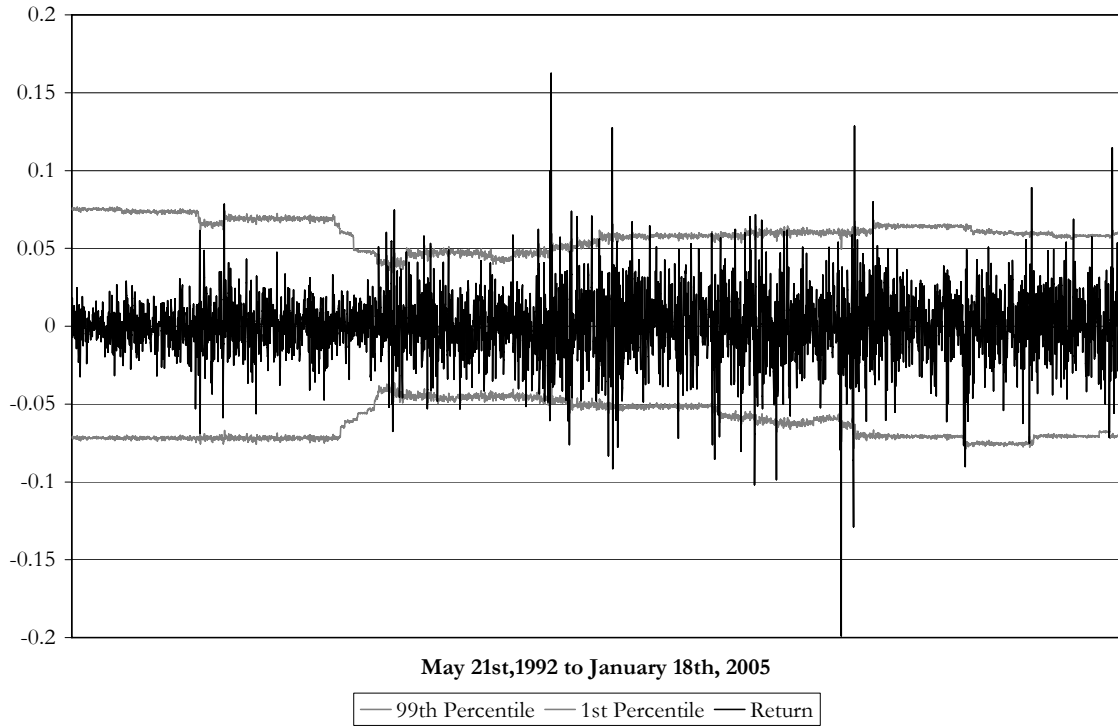
Percentile	LR_{UC}	LR_{IND}	LR_{CC}
1%	1.05313	21.86676*	22.91989*
5%	2.98260	11.58218*	14.56478*
95%	2.98260	2.10882	5.09142
99%	0.47289	7.806798*	8.27968*

*Results show statistically significant differences from expected values at the 0.05 level

As Graphs 1 and 2 show that although the confidence intervals seem to hug the actual returns quite closely, the forecasts are not very flexible. They do vary slightly from day to day, but overall the forecasts do not seem to mirror the true volatility of the oil prices very closely. In addition, both the 95% and, more obviously, the 99% intervals seem to be noticeably wider in the first portion of the out-of-sample period. It seems that the AR-HS model was more likely to overestimate the VaR forecast when the variance of the returns was actually lower; this may be an area for future research. In sum, the forecasts seem to follow the real returns fairly closely, if not flexibly.



Graph 1: AR-HS Results – 95% Confidence Interval



Graph 2: AR-HS Results – 99% Confidence Interval

5.2 Historical Simulation with ARMA Forecast Results

The average coefficients of the mean equations for the Historical Simulation with ARMA forecast (ARMA-HS) are shown below in Table 9. As with the AR-HS model, the constant is quite small, but it is interesting to note that with the addition of the MA term, the coefficient of the AR term becomes not only more heavily weighted, but negative. This suggests that if the previous day's return is large and negative, the mean equation would be strongly weighted with a large, positive AR term. The MA term has a positive and fairly large coefficient, which suggests that the previous day's error has a significant effect on the forecast. It is interesting to note how much the addition of the MA term to the mean equation changes the estimation of the model, yet there seems to be little difference in the resulting forecasts.

Table 9: ARMA-HS Mean Equation Coefficients

Coefficient	Mean Value	SE of Mean	T-Stat
Constant	0.0001319821	5.452439e-06	24.20607*
AR Coefficient	-0.2006616194	0.005191	-38.65555*
MA Coefficient	0.2559199659	0.005385	47.52106*

*Significance level under 0.05

As previously stated, the ARMA-HS model has surprisingly similar results to the AR-HS model, although with a few key differences. Once again, as with the AR-HS model, the ARMA-HS shows statistically significant autocorrelation in the Ljung-Box tests of both the raw residuals and the squared, shown in Tables 10 and 11. This may be one reason that the ARMA-HS model, with the addition of the moving average term in the mean equation, is unable to outperform the AR-HS; it does not appear to deal with the higher-order autocorrelation of the residuals any better than the AR-HS, despite the addition of the moving average error term. The violations and other performance statistics are summarized in Table 12 below.

Table 10: Ljung-Box Test of ARMA-HS Errors

Considered Lag	Ljung-Box Test Statistic	Significance Level
Q(12)	31.9882*	0.00138958
Q(24)	60.7833*	0.00004948
Q(36)	74.4904*	0.00017031

*Significance level under 0.05

Table 11: Ljung-Box Test of ARMA-HS Squared Errors

Considered Lag	Ljung-Box Test Statistic	Significance Level
Q(12)	152.8254*	0.00000000
Q(24)	165.6682*	0.00000000
Q(36)	192.3460*	0.00000000

*Significance level under 0.05

Although the results of the ARMA-HS forecasts are very similar to those of the AR-HS, particularly at the 1% and 5% level, the ARMA mean equation appears to better deal with the upper limits of the interval forecasts than the AR mean equation does. However, although the ARMA-HS has both fewer actual violations and a few less consecutive violations than the AR-HS, it is apparent upon further examination that the degree of violation is slightly higher at all levels in the ARMA-HS forecasts. Since the summed differences are extremely similar between the two models, practitioners would have to decide whether to accept a few more violations if the violations in question were slightly less severe or vice versa. Nevertheless, the *LR* statistics in Table 13 give a more concrete method of assessing the efficiency of the forecast intervals.

Table 12: Summary of ARMA-HS Results

Percentile	Expected # of Violations	# of Violations	Degree of Violation	Consecutive Violations	Summed Differences
1%	32.05	38	75.06%	3	200.23
5%	160.25	183	46.21%	23	111.66
95%	160.25	179	50.55%	12	108.63
99%	32.05	34	42.09%	1	191.06

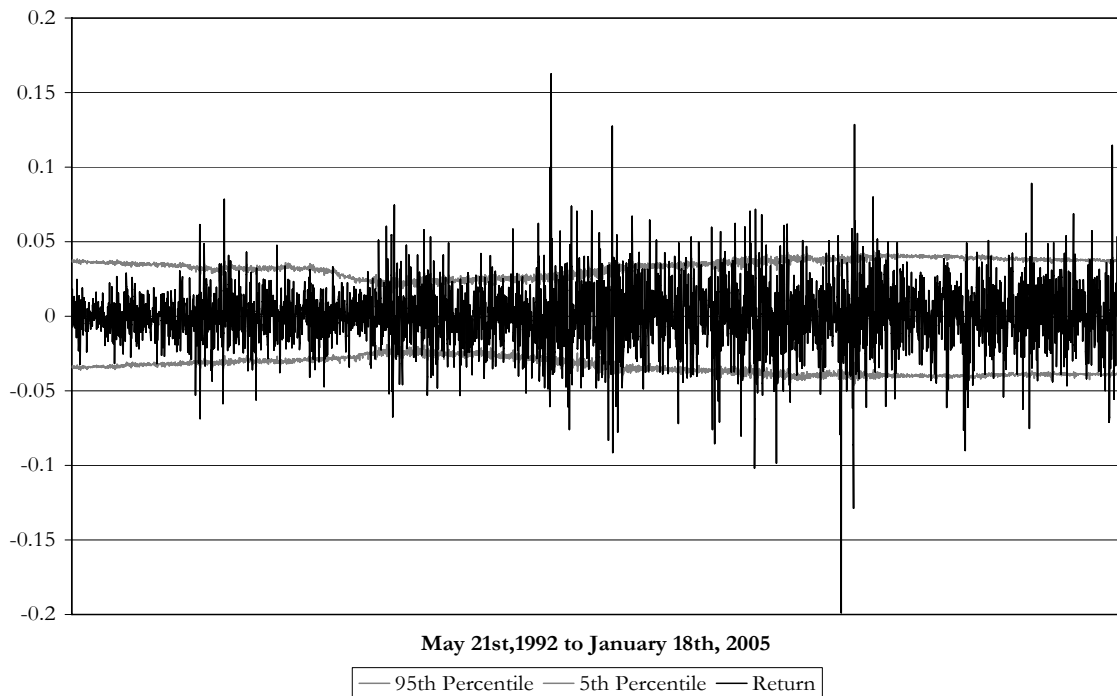
Like the AR-HS model, the ARMA-HS model has only one percentile, the 95th, that provides sufficient conditional coverage. The ARMA-HS provides sufficient unconditional coverage at all four percentiles, just as the AR-HS does. These results suggest there is no significant difference between the two C&M models. Therefore, it appears the MA term does not add anything significant from the VaR perspective. This is evidence of why the five comparison criteria are all crucial to a thorough comparison.

Table 13: Likelihood Ratio Statistics for ARMA-HS Results

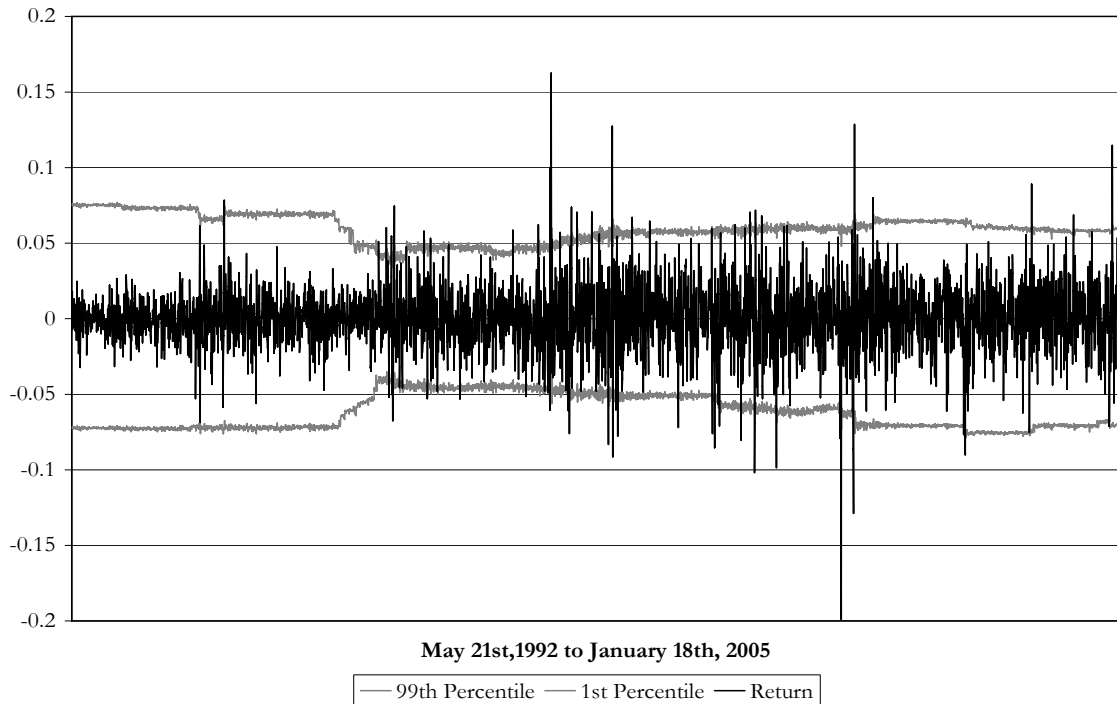
Percentile	LR _{UC}	LR _{IND}	LR _{CC}
1%	1.05313	21.86676*	22.91989*
5%	3.25730	13.11851*	16.37580*
95%	2.22866	0.42590	2.65456
99%	0.11751	7.83817*	7.95568*

*Results show statistically significant differences from expected values at the 0.05 level

Graphs 3 and 4, below, display forecasts almost identical to those of the AR-HS, only with slightly more variance on a daily basis; the ARMA-HS also appears inflexible overall.



Graph 3: ARMA-HS Results – 95% Confidence Interval



Graph 4: ARMA-HS Results – 99% Confidence Interval

In review, the results of the ARMA-HS forecasts are mixed relative to the preceding model. The ARMA-HS does not provide sufficient conditional coverage or independence at the 1%, 5%, or 99% level, but like the AR-HS model has sufficient unconditional coverage at all four percentiles. The similarity in performance between the AR-HS and the ARMA-HS suggests that the addition of the MA term does not significantly improve the specification of the mean equation from the Value-at-Risk perspective. Clearly, neither model is ideal nor do either the AR-HS or the ARMA-HS appear to deal adequately with the interdependence of the oil price return volatility. It may be that the two Semi-Parametric GARCH models, discussed in the following two sections, may be more successful in capturing the volatility clustering and deliver more independent Value-at-Risk violations.

5.3 Semi-Parametric GARCH with AR Forecast Results

Table 14 displays the constant and coefficient estimates of the Semi-Parametric GARCH with AR Forecast (AR-GARCH) mean equation, which are very similar to those of the AR-HS mean equation. Both the constant and AR coefficient are slightly larger than those of the AR-HS; this may indicate that the forecasts are slightly more skewed to the positive side and that the last day's return provides more information to the GARCH forecast than the C&M model.

Table 14: AR-GARCH Mean and Variance Equation Coefficients

Coefficient	Mean Value	SE of Mean	T-Stat
Mean Equation Constant	0.0003883944	9.958302e-06	39.00201*
AR Coefficient	0.0658380959	0.000270	243.51320*
Variance Equation Constant	0.0000207258	9.393505e-07	22.06137*
Variance Error Coefficient	0.0857233801	0.000554	154.66509*
Variance Coefficient	0.8788461594	0.002524	348.19369*

*Significance level under 0.05

The AR-GARCH model performs very well in modeling the VaR for oil prices. As Tables 15 and 16 below display, the AR-GARCH, unlike the C&M models, produces errors and squared errors that are not significantly autocorrelated. This fact reflects the GARCH model's ability to remove dependence in higher order moments of the oil price series. This suggests that the GARCH model captures the conditional volatility of the returns better and, as a consequence, may be able to deliver better VaR forecasts than the method suggested by Cabedo and Moya (2003b). Overall, the AR-GARCH forecasts are superior to those from the C&M models in a number of ways, as shown in the results to follow.

Table 15: Ljung-Box Test of AR-GARCH Errors

Considered Lag	Ljung-Box Test Statistic	Significance Level
Q(12)	11.3216	0.50157894
Q(24)	28.8315	0.22653749
Q(36)	41.7290	0.23575275

*Significance level under 0.05

Table 16: Ljung-Box Test of AR-GARCH Squared Errors

Considered Lag	Ljung-Box Test Statistic	Significance Level
Q(12)	17.5169	0.13116555
Q(24)	26.7525	0.31611938
Q(36)	35.4471	0.49470357

*Significance level under 0.05

The results of the AR-GARCH model are slightly mixed. For example, while the 5th percentile actually has less than the expected number of violations, it has twice as many consecutive violations as the 95th percentile. One very interesting fact to note is that the degree of violations for the 1st percentile is drastically higher than that of the other percentiles; in fact, almost double that of the 99th percentile. This is surprising, because the summed differences of the 1st percentile are higher than the 99th, which would suggest that the forecasts of the 1st percentile are more conservative. However, the large degree of violation suggests that a few very large negative returns were not anticipated, which could be a serious problem in practice. The AR-GARCH, as predicted, displays much fewer consecutive violations than the C&M models, a fact which will be discussed further in the next section.

Table 17: Summary of AR-GARCH Results

Percentile	Expected # of Violations	# of Violations	Degree of Violation	Consecutive Violations	Summed Differences
1%	32.05	35	62.68%	1	180.49
5%	160.25	158	35.93%	12	116.11
95%	160.25	169	39.32%	6	107.50
99%	32.05	34	28.51%	0	173.24

The LR statistics for the AR-GARCH model are assuredly the best out of all four models. The LR_{UC} statistics indicate that the model provides sufficient unconditional coverage at all four percentiles, and the LR_{IND} statistics indicate sufficient independence for all but the 1st percentile. Accordingly, the LR_{CC} statistics for the 5th, 95th, and 99th percentiles

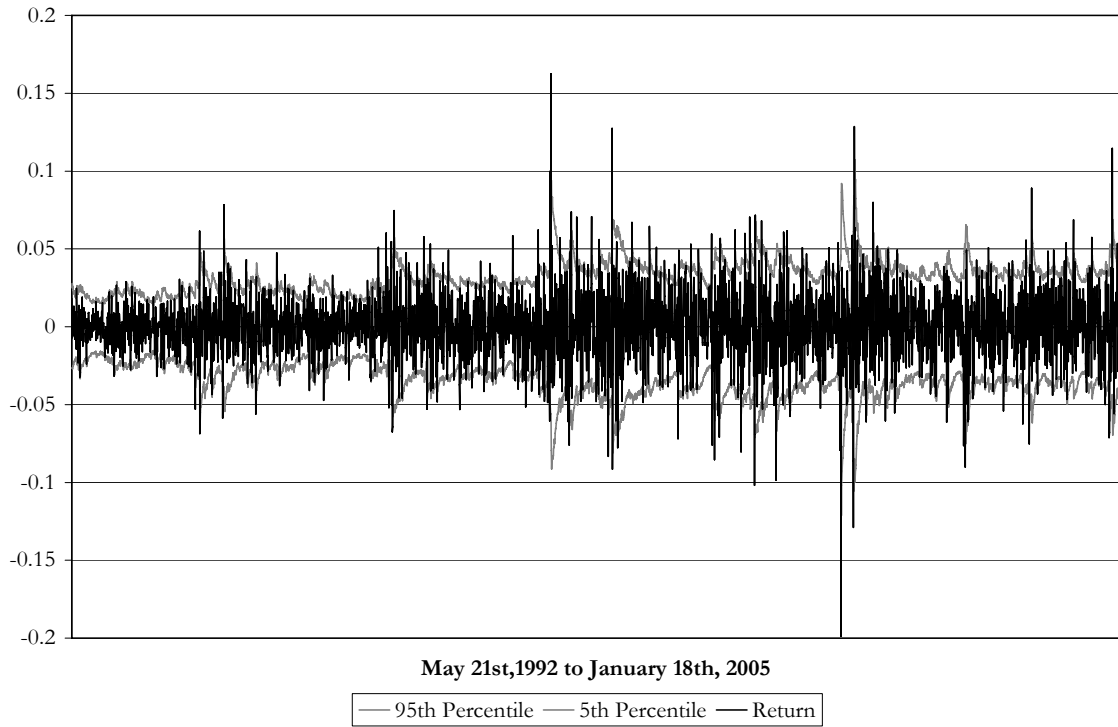
show that the number of actual violations and consecutive violations are not significantly different from their expected values. These are very good results, and although the 1st percentile's LR_{IND} and LR_{CC} values are statistically significantly different from their expected values, this is due to only one single consecutive violation versus the expected value of 0.32 consecutive violations. Overall, the forecasts produced by the AR-GARCH are very good.

Table 18: Likelihood Ratio Statistics for AR-GARCH Results

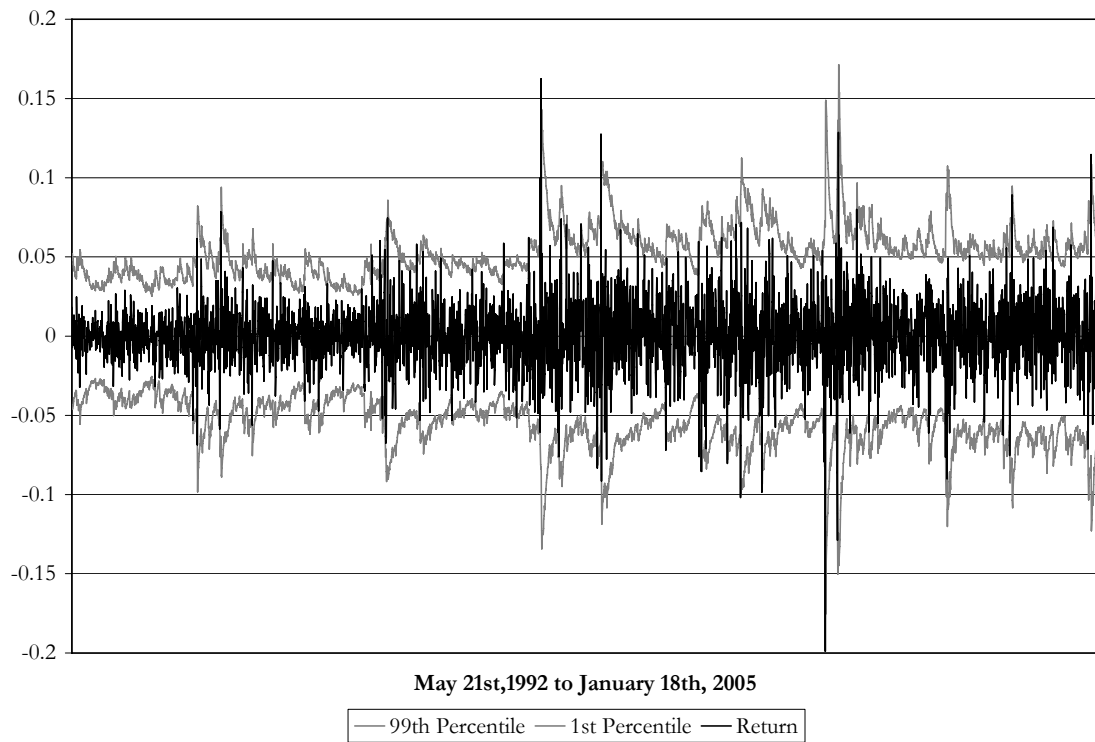
Percentile	LR_{UC}	LR_{IND}	LR_{CC}
1%	0.26631	7.82097*	8.08728*
5%	0.03340	2.19502	2.22843
95%	0.49447	1.18414	1.67861
99%	0.11751	0.72912	0.84663

*Results show statistically significant differences from expected values at the 0.05 level

It is immediately clear from the two graphs below that the GARCH model produces a vastly superior forecast relative to the two C&M models in terms of flexibility of the VaR bands. This additional flexibility is driven by the changing volatility in returns which the Semi-Parametric GARCH captures, while the C&M models are restricted by the assumption of constant variance. From the graphs, it is evident why the summed differences of the AR-GARCH are lower; the intervals seem to adjust much more quickly to changing volatility and the forecasted values follow the actual returns very well.



Graph 5: AR-GARCH Results – 95% Confidence Interval



Graph 6: AR-GARCH Results – 99% Confidence Interval

In sum, it appears that the AR-GARCH provides very good forecasts, but is better able to forecast the upper half of the confidence interval than the lower half. Therefore, this methodology is very sound, and provides superior forecasts compared to the C&M method.

5.4 Semi-Parametric GARCH with ARMA Forecast Results

Interestingly, the coefficients of the Semi-Parametric GARCH with ARMA Forecast (ARMA-GARCH) mean equation are similar to those of the preceding models in different ways. The values, shown below in Table 19, are similar in structure to those of the ARMA-HS; for instance, the constant is small and positive, the AR coefficient is relatively large and negative, and the MA coefficient is relatively large and positive. In all of these ways, the coefficients mirror those of the ARMA-HS model. However, they also mirror those of the AR-GARCH model in that they are all significantly larger than the ARMA-HS values, just as the AR-GARCH values were larger than those of the AR-HS. In this way, it appears that the GARCH models seem to place more weight on the last day's return and error, and also that the GARCH models, through the slightly larger, positive constant, seem to adjust for the fact that returns are more often positive than negative over time (hence the large increase in oil prices over time).

Table 19: ARMA-GARCH Mean Equation Coefficients

Coefficient	Mean Value	SE of Mean	T-Stat
Mean Equation Constant	0.0005257460	1.378129e-05	38.14926*
AR Coefficient	-0.3306019607	0.003941	-83.89695*
MA Coefficient	0.4027207864	0.004026	100.02983*
Variance Equation Constant	0.0000180929	5.343464e-07	33.85992*
Variance Error Coefficient	0.0855338155	0.000546	156.56372*
Variance Coefficient	0.8860899573	0.001665	532.18514*

*Significance level under 0.05

The results of the ARMA-GARCH forecast are almost completely identical to those of the AR-GARCH. The results of the Ljung-Box test of the standardized errors and squared errors generated from the ARMA-GARCH are shown in Tables 20 and 21 below. As with the preceding GARCH model, the squared errors are not significantly autocorrelated. This indicates that the GARCH process is successfully removing autocorrelation from the higher-order moments. This result is encouraging, as the ARMA-GARCH is producing very good forecast data and effectively modeling the time-varying volatility of the returns.

Table 20: Ljung-Box Test of ARMA-GARCH Errors

Considered Lag	Ljung-Box Test Statistic	Significance Level
Q(12)	10.8162	0.54473668
Q(24)	28.6217	0.23469229
Q(36)	42.0406	0.22566554

*Significance level under 0.05

Table 21: Ljung-Box Test of ARMA-GARCH Squared Errors

Considered Lag	Ljung-Box Test Statistic	Significance Level
Q(12)	17.4378	0.13385699
Q(24)	27.1697	0.29660537
Q(36)	35.4200	0.49599104

*Significance level under 0.05

Similar to the AR-GARCH, the ARMA-GARCH seems to deal better with the upper limits of the interval forecast than the lower. Although there are fewer actual violations on the lower level, the degree of violations, the number of consecutive violations, and the summed differences are all better for the two upper percentiles. Once again, just as with the AR-GARCH model, the ARMA-GARCH model has a very large degree of violation for the 1st percentile. Since the summed differences are higher, and the number of violations is so low, it is odd that the degree of violations for the 1% limit is so high. As is the case with the AR-GARCH model, the ARMA-GARCH produces the highest quality results for the two upper limits, and the 95% limit in particular. This indicates, as previously mentioned, that

although the results at all four percentiles were more than adequate, the VaR forecasts generated would be most useful for someone holding a short position in oil.

Table 22: Summary of ARMA-GARCH Results

Percentile	Expected # of Violations	# of Violations	Degree of Violation	Consecutive Violations	Summed Differences
1%	32.05	31	64.39%	1	179.60
5%	160.25	158	35.31%	11	116.32
95%	160.25	168	39.06%	7	108.20
99%	32.05	34	28.45%	1	174.32

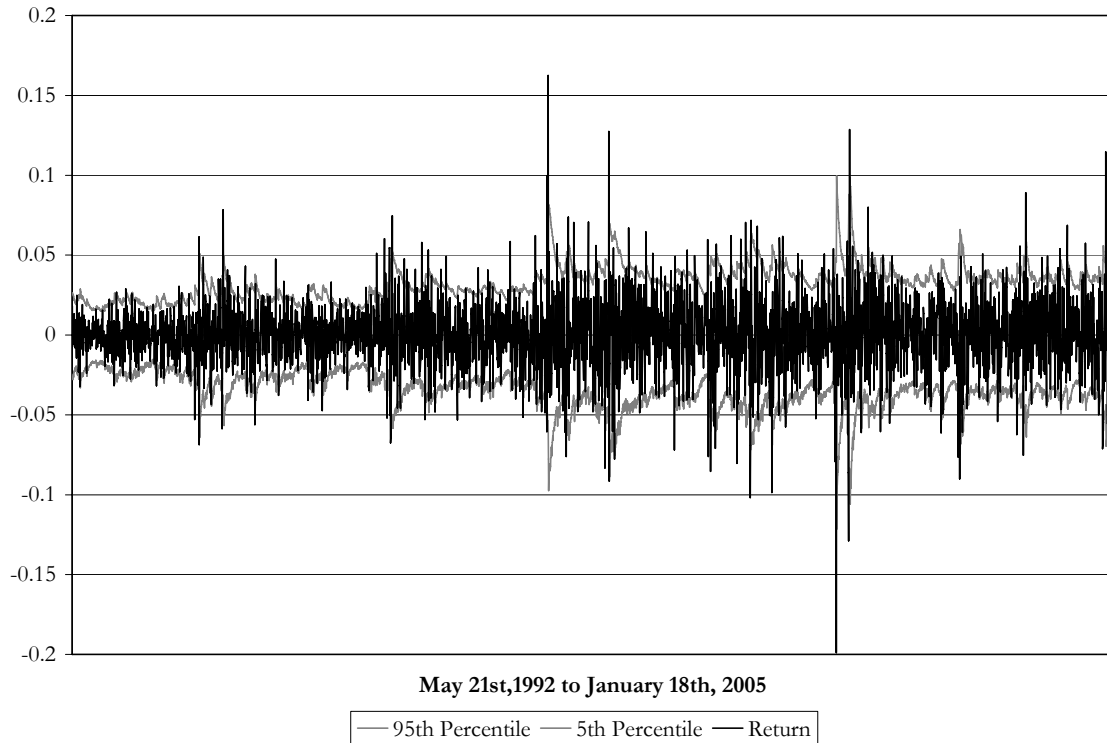
In addition to the difference between the upper and lower limits, the analysis shows a significant difference between the 95% confidence interval and the 99% confidence interval. The LR statistics, shown in Table 23 below, clearly display the fact that the 5th and 95th percentiles have statistically sufficient conditional coverage, while the 1st and 99th percentiles do not. The 1% and 99% interval forecasts have sufficient unconditional coverage, but the fact that both forecasts have even one single consecutive violation is enough to fail to reject the null hypothesis that the forecasts are not independent. As previously mentioned, some judgment can be used to decide whether the variance of the actual number of consecutive violations, in this case 1, from the number of expected consecutive violations, 0.32, is sufficient evidence to reject a model. Since the LR_{IND} statistics display statistically significant variance from the expected values, it is not surprising that the LR_{CC} statistics for the 1st and 99th percentiles are statistically significant as well. These results do not necessarily indicate that the ARMA-GARCH is producing unusable results (as mentioned above the actual variance from the expected values in the case of consecutive violations is extremely small), but they do make it possible to see that the AR-GARCH provides superior results to the ARMA-GARCH, and that both GARCH models provide superior results to the C&M models.

Table 23: Likelihood Ratio Statistics for ARMA-GARCH Results

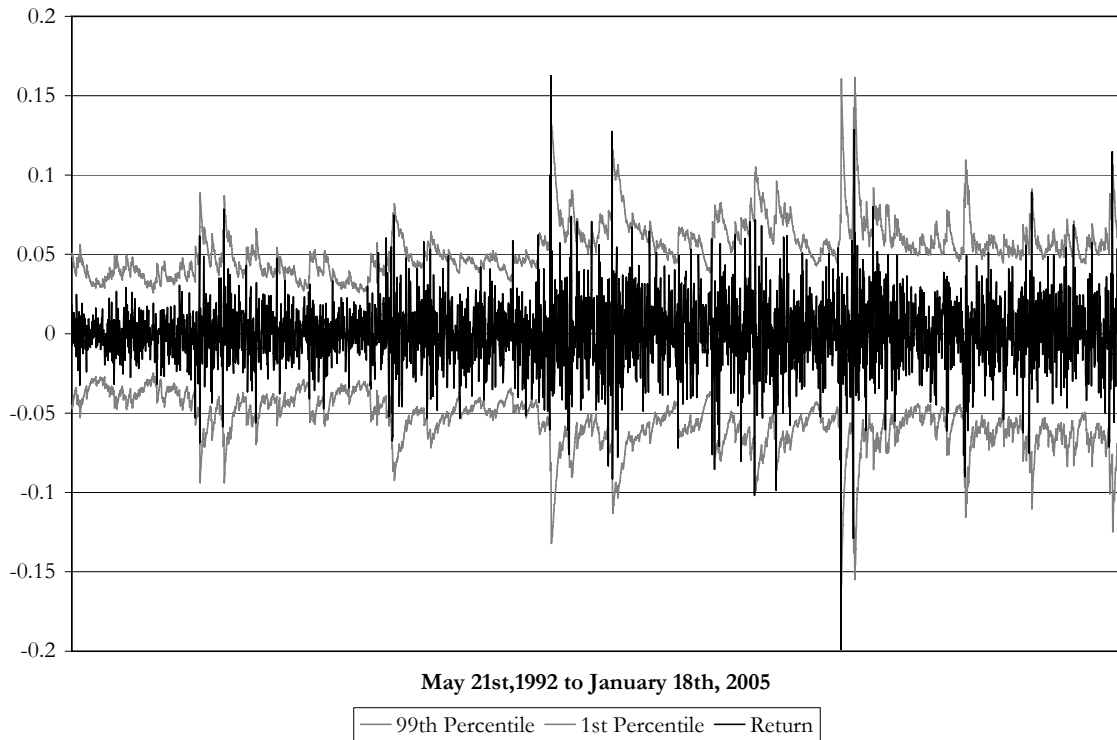
Percentile	LR_{UC}	LR_{IND}	LR_{CC}
1%	0.03513	7.90899*	7.94411*
5%	0.03340	1.31378	1.34718
95%	0.38865	0.44074	0.82939
99%	0.11751	7.83817*	7.95568*

*Results show statistically significant differences from expected values at the 0.05 level

Not surprisingly, the graphs of the ARMA-GARCH forecasts, shown below in Graphs 7 and 8, are very similar to those of the AR-GARCH, simply with more volatility day to day. This indicates that the addition of the MA term introduces more volatility into the forecasts and shows more variance in the forecasts from day to day.



Graph 7: ARMA-GARCH Results – 95% Confidence Interval



Graph 8: ARMA-GARCH Results – 99% Confidence Interval

In sum, the ARMA-GARCH generates good, reliable results, especially at the 5% and 95% level. For an oil producer with a long position in oil, who does not have strict risk management requirements, the ARMA-GARCH would provide very effective results. The same would apply to a heavy user of oil with a short position who simply wants a reliable estimate of how much he or she could expect the cost of his or her inputs to increase. Although the conditional coverage at both the 1st and 99th percentile is not statistically sufficient, it may be that the extremely small numbers of actual and expected consecutive violations simply do not provide enough information to make a firm judgment. Potentially, a much longer forecasting period may provide sufficient data to prove that the Semi-Parametric GARCH models have adequate conditional coverage at the 99% confidence interval as well as the 95%. In any case, the ARMA-GARCH model generates very good

forecasts of VaR values, and appears to deal very well with the changing volatility of oil price returns.

6.0 Discussion

The results of this comparison are somewhat surprising in some respects, and yet perhaps not surprising at all in others. Drawing from the work of Barone-Adesi et al. (1999) and Morana (2001), among others, it is logical to surmise that the GARCH model should be a superior tool in modeling oil price returns. However, from Cabedo and Moya's work in 2003, it is also apparent that in their specific comparison, the Generalized Autoregressive Conditional Heteroskedastic (GARCH) model does not outperform the newly-developed C&M model. Therefore, it is interesting to see whether, in fact, any difference is made by the difference in the mean equations in the Cabedo and Moya's (2003b) analysis, and to see how much better the GARCH model may perform with a non-parametric distribution as opposed to restricting it with the assumption of normality.

In the case of both methodologies, the addition of the MA term not only makes little difference, but in some cases it actually seems to have a negative affect on the forecasts. For instance, in the case of the C&M model the addition of the MA term increases the severity of the violations displayed. In the Semi-Parametric GARCH model, the severity of violations at the 1st percentile increase from 62.68% to 64.39%. In addition, the MA term appears to increase, not decrease, the number of consecutive violations seen in the GARCH model by one at both of the upper limits. Conversely, the MA term decreases the actual number of violations and consecutive violations seen in the C&M model, so its contribution appears to be mixed. All in all, the benefit of the MA term would have to be decided by the end user and their position; for example, a producer with a long position in oil may conclude that the ARMA-GARCH deals better with the lower portion of the interval forecasts. On the other hand, a refiner who is short oil may decide that the AR-GARCH has an acceptable

number of violations and consecutive violations at both the 95% and 99% level, and is therefore more suitable to their needs.

As earlier mentioned, it seems that the inferior performance of the GARCH model may be the result of one particular factor in Cabedo and Moya's analysis. Ultimately, their comparison of these two models is flawed by its use of a historical, non-parametric distribution for the C&M model, and a normal distribution for the GARCH model. This factor, from all the evidence gathered, is the key reason that the GARCH model does not compete strongly in their comparison. Therefore, it is not surprising that the semi-parametric GARCH model performs well against the C&M in this analysis, as it is not restricted by the assumption of normality. In sum, after an extensive and thorough testing process, it appears Cabedo and Moya's conclusion that their model outperforms the GARCH model is driven in large part by the normal distributional assumption imposed on the GARCH model. The simpler C&M model simply cannot adequately deal with the volatility clustering in oil prices, and the Semi-Parametric GARCH models the time-varying volatility extremely well, as evidenced by the removal of the auto-correlation in the squared residuals and the excellent VaR forecasts.

Since the Semi-Parametric GARCH model relaxes the assumption of constant variance on the future risk structure, it offers more flexibility in modeling VaR than the C&M model. In particular, variance equation used in the Semi-Parametric GARCH model allows the forecasts to be adjusted not only by the historical returns and errors, but by the historical and forecasted variance. In this way, the GARCH model provides additional flexibility in the sense that it models changing volatility while the C&M does not. Consequently, as demonstrated in this study, the C&M methodology does not surpass the Semi-Parametric GARCH. In fact, if the series displays significant changing volatility, the Semi-Parametric

GARCH should be deliver superior VaR forecasts. Thus, it appears the C&M conclusion is driven by different assumptions about the future risk structure rather than an inherent flaw in the GARCH model.

7.0 Practical Application

The goal of comparing forecasting models is not only to determine which produces the most accurate forecasts, but also to evaluate the practicality and ease of use of the models. The greater simplicity of the C&M models is a benefit in practical application, as not all users of the VaR methodology will be accomplished in the field of econometrics. If a more parsimonious model is able to produce equal, or superior, results relative to a more complex model, the simpler models would always be preferred. However, since the Semi-Parametric GARCH models produce better results in this analysis, it is also necessary to point out the fact that although the GARCH methodology is slightly more complex than that of the C&M model, once the model is set up and running smoothly, little additional work needs to be done. Therefore, although the initial estimation and programming of the GARCH model may involve slightly more work, once it is set up both methodologies simply need to have another day of data added and to have the in-sample data rolled forward by one day.

In this light, the GARCH model is just as simple to run on a daily basis as the C&M. Along the same vein, the fact that the forecast equation of the AR-GARCH methodology contains fewer estimates than the ARMA-GARCH is a slight point in its favour. Since the AR-GARCH also generates more accurate forecasts than the ARMA-GARCH based on the overall conditional coverage, this indicates that it should be the preferred model. A further advantage of the GARCH methodology over the C&M methodology is that the instability of the oil prices can make it difficult to generate accurate results using the C&M method, and this difficulty increases in periods of higher volatility. Consequently, since oil markets seem to be experiencing greater volatility over time, having a model that can better deal with the increasing variance, as the GARCH does, is a clear benefit.

One attractive quality of the Semi-Parametric GARCH model is that the 99% confidence interval VaR estimates performed well, if not quite as well as the 95% confidence level; as previously mentioned, judging the independence, and consequently the conditional coverage, of the 99% confidence interval is problematic. If the model could produce slightly better forecasts at the 1st percentile, it would have great benefit in practice due to banks' strict requirements. It would be interesting to see if the AR-GARCH performs as well, or better, with less volatile commodities. As the Basel Committee requires that banks hold a cash reserve large enough to cover a VaR at 99% confidence (Jorion, 2001), being able to accurately and flexibly forecast this percentile without overestimating would be of great value to practitioners. Aside from banks, there are many large and small businesses with short positions in oil that have reason to look for a conservative VaR number; many large trusts and publicly-owned companies have very risk-averse profile, and many small companies simply do not have the liquidity or working capital to risk underestimating their potential losses. In these cases, and in many others, having a dependable and flexible VaR model is paramount.

8.0 Limitations

This comparison, like any data-driven study, is limited by a number of factors. The study itself was driven by the need to replicate the comparison performed by Cabedo and Moya (2003b) in a more thorough and equitable manner. Therefore, it is possible that another methodology exists that could outperform the Semi-Parametric GARCH model and the other models within this analysis.

Another limitation of this study is the commodity forecasted; due to the fact that oil prices have unique characteristics like fat-tails and conditional volatility (Bera and Higgins, 1993; Bollerslev et al., 1992; Hendricks, 1996; Panas and Ninni, 2000), the preceding models may perform very differently when forecasting commodities or series other than oil returns. Replication of this comparison on portfolios of stocks or exchange rates would go far in affirming the preceding results.

One factor that does not limit this study is the data. The data is undeniably sufficient and covers a greater length of time than most other comparisons. It is also essentially the complete set of available data; oil has been traded on the New York Mercantile Exchange (NYMEX) since 1983, and the only data not used in this test were the prices from the time study began until today. Data is vital to any back-testing, and this data set is complete, with very few missing observations, and from a reliable source, the Energy Information Administration of the US Department of Energy.

One note of interest, the 1st percentile seems to cause problems for all of the four models. This could either be a limitation of the methodologies examined in the preceding analysis, or an inherent quality of oil prices that should be further explored. Specifically, in September and October of 2001, there are very large negative returns that are very likely related to the bombing of the World Trade Centre on September 11th. These singular and

huge drops in oil prices cause great difficulties for all the models, and likely skew both the independence likelihood ratios and the degree of violations. In total, there are many more consecutive violations on the lower level than the upper level of the intervals. Upon initial examination, it appears that both the C&M and the Semi-Parametric GARCH models are much more likely to underestimate the lower limit of the confidence interval than the upper limit. In terms of oil price returns, this statistic seems to follow logically, as the lower limit of the return is actually bounded; the largest possible drop in oil price is 100%. On the other hand, oil price could hypothetically increase infinitely. However, contrary to this hypothesis, by digging deeper we see that there are actually higher summed differences on the *lower* level than the upper level, counter-intuitively indicating a greater tendency to overstate the downside risk of oil prices. The higher consecutive violations and degree of violations may simply be the result of the few outliers from events like September 11th and would be a fascinating topic for future investigation.

In reference to this quality of the VaR forecasts, Barone-Adesi et al. (1999) and Morana (2001) both incorporate an additional factor into the Semi-Parametric GARCH's variance equation that was not used in this analysis. This additional constant, shown below in the alternative variance equation as γ , is intended to measure the asymmetry of the last observed error.

$$h_t = \varpi + \alpha(\varepsilon_{t-1} + \gamma)^2 + \theta h_{t-1} \quad (12)$$

This model, called the Asymmetric GARCH model (Morana, 2001), is intended to allow an asymmetric reaction to large price movements; essentially, it would allow the model to follow large downward price movements with higher volatility estimates than large upward price movements. This model may better deal with the time-varying volatility and skewness

of oil price returns and the lack of an asymmetric term in the variance equation may partially account for the fact that the models all appeared to deal better with the upper limits of the interval forecasts than the lower limits. Further research may provide interesting results.

Replication should be done to affirm the preceding conclusions, and new models should be tested as they are developed. In terms of this study, the evidence shows that overall the Semi-Parametric GARCH model outperforms the model developed by Cabedo and Moya in 2003.

9.0 Conclusion

By amalgamating the information contained in all of the preceding sections, several facts become obvious. Firstly, the results of this analysis are mixed. No one method stood out as clearly superior to the others, and no one method performed poorly enough that it would not be a useful forecasting tool. Secondly, the two GARCH models overall produce more efficient forecasts than the two C&M models. And thirdly, the addition of the MA term to the mean equation has little benefit, especially in the case of the GARCH models, where the results are almost identical. Therefore, within the scope of this analysis, it is concluded that the Semi-Parametric GARCH model produces the most accurate Value-at-Risk forecasts of oil price returns.

In Cabedo and Moya's (2003b) study, they compare three methodologies for estimating the VaR of oil prices: the standard Historical Simulation model, the parametric GARCH model, and the Historical Simulation with ARMA forecasts model that they develop within the scope of the study. They conclude that their newly-developed model is the superior. As Cabedo and Moya (2003b) do not use formal comparison criteria to assess the efficacy of the VaR forecasting methods, their assertion that their model is superior is simply based on the fact that the model has close to the required percentage of violations, and appears more flexible than the standard GARCH approach which uses a normal distribution and no MA term in the mean equation. However, as is apparent from the preceding analysis, there are many other factors that must be taken into consideration when undertaking a thorough and fair comparison of interval forecasts.

In their closing, Cabedo and Moya state that there are two main advantages to their model: firstly, that an autoregressive mean equation is used to forecast the VaRs, as opposed to the standard Historical Simulation model, which simply takes the values directly from the

historical distribution of returns; and secondly, that their model does not arbitrarily impose a distribution on the structure of the risk factor, as is the case in the parametric GARCH. However, it is quickly obvious that the Semi-Parametric GARCH model analyzed within this study not only shares both of these stated advantages, but surpasses them with a third: the ability to adjust the VaR forecast to the time-varying volatility of oil prices. In doing so, the Semi-Parametric GARCH methodology developed by Barone-Adesi et al. (1999) summarily outperforms Cabedo and Moya's Historical Simulation with ARMA Forecasts methodology and provides superior estimates of oil's Value-at-Risk. The key difference being that the VaR violations are more independent under the Semi-Parametric GARCH model than under the C&M method.

As discussed above, the drastic differences in the conclusions of this study and that of Cabedo and Moya (2003b) stem from the misspecification of the GARCH model by arbitrarily applying a normal distribution to the structure of the risk factor. Logically, the evidence given within their study to support the use of the non-parametric historical distribution in conjunction with the ARMA forecasts would also apply in the case of the GARCH model. By following this logic through and applying the historical distribution in both models, it was possible to see the potential that the GARCH model has in forecasting the VaR of oil prices and conclude that Cabedo and Moya's assertion that their model is superior to the GARCH is based on the difference in the assumed distribution used in the forecasts and not on any true inferiority of the GARCH methodology.

In sum, this study adds to the extant literature by performing a thorough and extensive comparison of two Value-at-Risk methodologies for forecasting oil prices. These results should be replicated, and further study is needed to see how each model deals with different time-series data, but this comparison should provide both researchers and practitioners with

conclusive evidence on the usefulness of both the Cabedo and Moya and Semi-Parametric GARCH methodologies in forecasting the Value-at-Risk of oil prices.

10.0 References

- Barone-Adesi, G., Kostas, G., & Vosper, L. (1999). VaR without Correlations for Portfolios of Derivative Securities. *The Journal of Futures Markets*, 19(5), 583-602.
- Bera, A.K. & Higgins, M.L. (1993). ARCH models: properties, estimation and testing. *Journal of Economics Surveys*, 7(4), 305-366.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T., Chou, R.Y., & Kroner, K.F. (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, 52(1-2), 5-59.
- Cabedo, J.D. & Moya, I. (2003a). Estimating oil price 'Value at Risk' using the historical simulation approach. *Energy Economics*, 25, 239-253.
- Cabedo, J.D. & Moya, I. (2003b). Value at risk calculation through ARCH factor methodology: Proposal and comparative analysis. *European Journal of Operational Research*, 150(3), 516-528.
- Christoffersen, P.F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4), 841-862.
- Duffie, D. & Pan, J. (1997). An Overview of Value at Risk. *Journal of Derivatives*, 4(3), 7-48.
- Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987-1007.
- Faff, R.W. & Brailsford, T.J. (1999). Oil price risk and the Australian stock market. *Journal of Energy Finance and Development*, 4, 69-87.
- Fleming, J. & Ostdiek, B. (1999). The impact of energy derivatives on the crude oil market. *Energy Economics*, 21(2), 135-167.
- Giot, P. & Laurent, S. (2003). Market risk in commodity markets: a VaR approach. *Energy Economics*, 25, 435-457.
- Hendricks, D. (1996). Evaluation of Value-at-Risk models using historical data. *Economic Policy Review – Federal Reserve Bank of New York*, 2(1), 39-70.
- Hopper, G.P. (1996). A new methodology for measuring portfolio risk. *Business Review – Federal Reserve Bank of Philadelphia*, Jul/Aug, 19-32.
- Jorion, P. (2001). *The New Benchmark for Managing Financial Risk: Value at Risk*. New York: McGraw-Hill.

- Moosa, I.A. & Al-Loughani, N.E. (1994). Unbiasedness and time varying risk premia in the crude oil futures market. *Energy Economics*, 16(2), 99-105.
- Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. *Energy Economics*, 23(3), 325-338.
- Panas, E. & Ninni, V. (2000). Are oil markets chaotic? A non-linear dynamic analysis. *Energy Economics*, 22(5), 549-568.
- Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics*, 21(5), 449-469.