School of Economics and Management
Department of Economics
Master Thesis, Spring 2010

Expected Shortfall as a Complement to Value at Risk

A study applied to commodities

Supervisors
Birger Nilsson
Anders Vilhelmsson

Authors
Filippa Gerstädt
Maria Olander

Abstract

Basel II requires Value at Risk (VaR) as a standardized risk measure for calculating market risk. However, the validity of the risk measure has been questioned since it neglects the losses beyond the VaR level. Expected Shortfall (ES) is a response to this limitation, as it is defined as the average of the losses ignored by VaR. This study applies VaR and ES to three commodities; gold, oil and corn by using the models historical simulation, age-weighted HS, volatility-weighted HS, normal distribution, Student-t distribution, log-normal distribution. Also, conditional volatility, structured as a GARCH(1,1) model, is applied to the three distributions. These nine models are evaluated by backtesting procedures for each commodity. Applying conditional variance improves the models radically and we conclude that the models Volatility-weighted HS and Student-t GARCH(1,1) are the most accurate models regarding the three commodities. Additionally, estimating ES adds value to this study, even though it is almost perfect positively correlated with VaR.

Keywords: Value at Risk, Expected Shortfall, historical simulation, age-weighted HS, volatility-weighted HS, normal distribution, Student-t distribution, log-normal distribution, GARCH(1,1), backtesting

Contents

1 Introduction

1.1 Background

Basel II, issued by the Basel Committee on Banking Supervision, regulates and makes recommendations on banking laws. Basel II requires Value at Risk (VaR) as a standardized risk measure which shall capture the market risk of the bank attempting to eliminate the probability of default, as well as to hold capital for default risk.[1]

Yamai's and Yoshiba's article *Value-at-Risk versus expected shortfall: A practical perspective*, written in 2004, highlights the disadvantages of VaR, a risk measure that has turned out to be standardized for risk management. According to Yamai and Yoshiba, VaR does not capture a complete risk profile, and mean that Expected Shortfall (ES) is broader in that sense. The article focuses on market stress applied to a concentrated credit portfolio and foreign exchange rates, and concludes that VaR should be complemented with ES to eliminate the current limitations of using one standardized financial risk measure.[2]

Additionally, Nassim Taleb, famous as the author of the book *The Black Swan*, pinpoints the disadvantage of the standardized risk measure VaR and its simplicity; *"Proponents of VaR will argue that it has its shortcomings but it's better than what you had before*".[3] Taleb's quote and Yamai's and Yoshiba's article therefore raise questions about the adequacy of VaR, and thus implicitly of Basel II.

Büyükşahin, Haigh and Robe wrote *Commodities and Equities: A "Market of One"?* in 2008, where they discuss the increased investments in commodities in the past decade and weather this has led to changed relations between commodities and traditional financial assets. Historically, risk factors that explain cross sectional variations in equity returns have had no forecasting power in commodity markets. However, the financial institutions' increased activity in these markets has reduced the extent of cross-market arbitrage opportunities. As a result, more closely linked equity and commodity markets have appeared. The authors apply dynamic correlation and recursive cointegration techniques to S&P 500 return and GSCI total return data, and find that the relation between the two asset classes has not changed significantly in the last 15 years. They also conclude that commodities have continued to offer benefits regarding portfolio diversification.[4]

---

[1] http://www.bis.org/publ/bcbs140.htm
[2] Yamai and Yoshiba (2005)
[3] Jorion (2001) p. 499
[4] Büyükşahin et. al. (2008)

1.2 Problem discussion

In the book *The Black Swan*, Taleb writes the following sentence, meaning that people tend to defend risk measurements as if they were complete:[5]

"*The problem with experts is that they do not know what they do not know*."

According to the article *The level and quality of Value-at-Risk disclosure by commercial banks*, written by Pérignon and Smith in 2009, commercial banks most frequently use historical simulation when calculating VaR. The second most common used method is Monte Carlo simulation. The authors describe two reasons for historical simulations' popularity. First, since banks are exposed to numerous risks, it is complex to estimate parametric VaR. Second, banks do not want too volatile day-to-day risk as a foundation of their VaR estimation, which is the outcome given by a parametric method relative to non-parametric methods, such as historical simulation. The authors express historical simulation as a "*mechanical disconnection between 1-day VaR and future volatility*" and question the common use of this model when measuring market risk.[6]

Yamai's and Yoshiba's article mentioned in the earlier section, as well as Taleb's expressions, made us curious about the qualities of VaR and ES. Previous studies within this field have primarily focused on traditional stock indices, such as S&P 500, see section 2.11 for an example. Referring to the commodity article in section 1.1 above, and the differences between the traditional stock indices and commodities, there are incentives for studying commodities. Also, ES is less studied and its advantage, compared to VaR, is its consideration of observations beyond VaR.

We have chosen to apply VaR and ES to three commodities; gold, oil and corn based on Standard & Poor's GSCI Spot Index, which are priced in USD. The main reasons for choosing commodities are the limited number of studies applied to them within this area, in combination with their increasing importance as traded assets, as well as their divergence in price patterns compared to traditional asset classes. Gold, oil and corn were chosen to represent commodities due to their high liquidity on the commodity market. These commodities are of different characteristics, adding a broader perspective to the study. This thesis applies VaR and ES to the commodities given non-parametric, semi-parametric and parametric approaches with the purpose to evaluate which model that matches each commodity best.

---

[5] Taleb (2007)
[6] Pérignon and Smith (2009)

1.3 Purpose

The purpose of this thesis is to study Value at Risk and Expected Shortfall for long positions on a daily basis, applied to three commodities with non-parametric, semi-parametric and parametric approaches, nine models totally. We want to study how the two risk measures complement each other and investigate if there is one, or a few, models that correctly estimate the market risk for all commodities at the confidence intervals 95 % and 99 %. The evaluation of the nine different models is based on numerous backtesting tests.

1.4 Delimitation

Due to our limited time horizon in the thesis procedure, delimitations are necessary. Including more than three commodities would have improved the statistical inference in the report. Also, covering other financial markets, stock indices and foreign exchange as examples, would have been interesting since it might have been easier to interpret differences in the results, if any. Though, the commodities chosen are of different characteristics, adding a broader perspective to the study. Since the out-of-sample period is characterized by the recent global financial crisis, it would also have been motivated to add an out-of-sample period that does not cover this crisis.

Furthermore, finding the most accurate model requires many approaches and including more than nine models would therefore have added value to the report. Also, the same argument can be applied to the choice of modeling the conditional volatility. This report only covers long positions and it would have been interesting to include short positions in the study as well. Due to the advantages and disadvantages of the backtesting procedure, more tests would have been motivated.

1.5 Reminder

In chapter 2, the basic theoretical background is presented and will work as a foundation for the remaining part of the study, trying to fulfill the purpose. Chapter 3 describes the methodology to reach the results, which are presented in chapter 4. Finally, the results are analyzed in chapter 5, to be concluded in chapter 6.

2 Theoretical background

2.1 Commodity market patterns

Energy products, metals and agricultural commodities are normally affected by business cycle changes that cause disturbance on the commodities' supply and demand.[7] For instance, investors tend to trade more precious metals when the financial markets are turbulent.[8] Generally, energy commodities are impacted by political changes, and agricultural products are sensitive to unanticipated weather. These components have the largest impact on commodity prices and many of them can not be foreseen. Furthermore, turbulent political issues, such as wars, have proven great effects on commodities. Commodities are also affected by decreased storage and seasonality. Additionally, the prices can be imbalanced if radical short-term trading speculations are made under a short period of time. These factors that impact the prices explain the complexity of the commodity market.[9]

Looking beyond the commodities' exposures presented above, commodity prices are positively correlated with inflation, meaning that commodities can affect the overall price level in an economy. Rogers, the author of the book *Hot Commodities: Anyone Can Invest Profitably in the World's Best market*, explains that commodities have had better return, in combination with lower risk, than equities and bonds since 1959. Therefore, Rogers concludes that commodity investments will increase the diversification of a portfolio containing equities and bonds. He also argues that people investing in equities, without good knowledge of the commodity market, can not make as large returns as investors that are familiar with this market. That is, equities are, to a large extent, impacted by changes in commodity markets.[10] Moreover, it is common that individuals invest directly in commodities instead of indirect investments via equities. Since commodities usually are expressed in USD, they also offer possible hedges against reduction in USD in relation to other currencies.[11]

[7] Giot and Laurent (2003)
[8] https://www.dnbnor.se
[9] Giot and Laurent (2003)
[10] Rogers (2004) p. 208-214
[11] https://www.dnbnor.se

2.2 Market risk

Market risk develops when certain factors in the overall economy changes, such as the level, or volatility,[12] of market prices.[13] An asset class' market risk is typically a risk based on changes in interest rates, in which LIBOR is an example of a reference interest rate that generates the market risk. That is, since balance sheet assets' and liabilities' values are affected by changes in interest rate. Changes within commodities, currencies and equity values are other examples of factors that affect the market risk.[14] Market risk is normally defined as an absolute risk when it is measured in absolute terms of a currency, e.g. USD. Otherwise, it is a relative risk and studies deviations from a benchmark index.[15]

The concept risk factor is defined as influences on cash flow when any market-determined asset price, rate or index value, changes. Furthermore, market risk can also be described by determining how the risk factors affect the values. To illustrate market risk, the Greek letters that describe five different types of market risk will be introduced. Delta is the risk based on a change in an underlying risk factor, all else equal, e.g. an interest risk affecting a bond price. The risk called Gamma is determined by the size of changes in Delta when the underlying risk factor changes. Vega is the risk of volatility changes of the underlying risk factor that impact the return of the asset. The risk factor Theta measures the exposures over time, in which insurance is an example, since it is bought in case something happens, and is useless if no incident occurs. The fifth risk is Rho, which describes changes in interest rates and affects net present value calculations of cash flows, since the rate is used as a hurdle rate.[16] These Greek letters are often divided into two groups; directional risks and non-directional risks. Directional risks are influenced on direct financial market changes, e.g. changes in stock prices, commodities, interest rates and exchange rates, on which Delta is an example. Gamma is an example of a non-directional risk, since it is exposed to second-order effects on the market.[17]

Additional market risks are correlation risk and basis risk. Correlation risk occurs when correlation changes between factors that impact the value of the portfolio. An example of basis risk is the difference in spot price of an asset compared to a future contract with the

[12] For a deeper discussion on the topics of statistic and econometric, the reader is referred to Gujarati, Damodar N. *Essentials of Econometrics* (2006) and Brooks, Chris. *Introductory Econometrics for Finance* (2008).
[13] Jorion (2001) p. 15-16
[14] Culp (2001) p. 16-18
[15] Jorion (2001) p. 15-16
[16] Culp (2001) p. 16-18
[17] Jorion (2001) p. 15-16

same underlying asset.[18] Other market risk measures are duration, used for measuring interest rate risk, as well as Value at Risk and Expected Shortfall.[19]

2.2.1 Basel II

The Internal Models Approach (IMA) to market risk states regulations on how banks must consider their market risk. The Basel Committee on Banking Supervision set the general criteria, which each country's supervision authority revises to fit their exposure. The supervision authorities oversee that each bank has an independent risk control suited for its business. Banks must continuously backtest their market risk estimates, as well as stress test a hypothetical exposure, in relation to the actual movements on the financial market. Their evaluations should be modified on a regular basis, to match the current movements of the financial markets, and it is of importance that their day-to-day risk estimates fit their risk management. Furthermore, the supervision authorities make sure that the board of directors is closely involved with the risk management concerning the market risk exposure.[20]

According to Basel II, banks can decide which model to attempt when estimating their VaR, but mentions historical simulation and Monte Carlo simulation as examples. However, it is a necessity to calculate the risk at a confidence interval of 99 % on a daily basis based on an in-sample period of at least one year.[21]

As of relevance in this study, Basel II requires banks to consider commodity markets in which the banks hold substantial positions. A bank with a limited commodity exposure is only demanded to specify the exposure by each given risk factor, or by each given risk group, in which different types of oil could be grouped as an example. However, if the bank trades commodities frequently, the exposure of their derivatives and cash positions should be detailed.[22]

2.3 Value at Risk

VaR measures the size of an amount at risk of underlying assets or liabilities at a specific probability and within a certain time period. In other words, VaR states the exposed amount

[18] Culp (2001) p. 16-18
[19] Dowd (2005) p. 5, 27, 35
[20] http://www.bis.org/publ/bcbs148.pdf
[21] Ibid.
[22] Ibid.

and how likely it occurs. VaR is specified as follows, where VaR represents the absolute amount given its likelihood, α %:[23]

$$\Pr(x \leq VaR) = \int_{-\infty}^{VaR} f(x)dx = \alpha \quad (1)$$

$f(x)$: probability density function for x
α: confidence level e.g. 5 %

When estimating VaR, a time horizon must be stated, in which market-related losses of the assets or liabilities might occur. This horizon depends on different circumstances; for instance the firm's tolerance in relation to actual risks, how often the underlying assets or liabilities require risk evaluations, and how easily the firm can liquidate or hedge large losses. One day, one week or quarterly are time horizons that are commonly used when calculating VaR.[24]

Furthermore, a relevant confidence level, α, must be assumed for the underlying assets or liabilities, given the probability of a loss. The confidence interval is equivalent to ($1-\alpha$) %, where α is the probability of the left tail of a risk distribution, the normal distribution as an example. More specific, α is called the VaR level or the critical probability, meaning that VaR should not be exceeded more than α % of the time period. The two most common confidence levels are 1 % and 5 %, depending on the probability in which losses occur.[25]

2.4 Expected Shortfall

A coherent risk measure $\rho(.)$ should fulfill the following four axioms:[26]

- Translation invariance. *For all* $X \in \mathcal{G}$ *and all real numbers* α, $\rho(X + \alpha \cdot r) = \rho(X) - \alpha$
- Subadditivity. *For all* X_1 *and all* $X_2 \in \mathcal{G}, \rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$
- Positive homogeneity. *For all* $\lambda \geq 0$ *and all* $X \in \mathcal{G}, \rho(\lambda X) = \lambda\rho(X)$
- Monotonicity. *For all* X *and* $Y \in \mathcal{G}$ *with* $X \leq Y$, $\rho(Y) \leq \rho(X)$

$\mathcal{G}$: all risks
X: random variable
Y: random variable
r: strictly positive price
α: amount invested in risk free asset
λ: constant stating the proportion of the portfolio

[23] Culp (2001) p. 342-344
[24] Ibid.
[25] Ibid.
[26] Artzner et. al. (1999)

VaR does only fulfill the second axiom, subadditivity, when the returns are elliptically distributed. Generally, VaR can not be considered as a coherent risk measure. This second axiom says that the risk of a portfolio made up of sub-portfolios is at most the sum of the individual sub-portfolios' risks, which is an effect that arises from diversification. Acerbi and Tasche (2001) argue that the axiom of subadditivity should not be violated in any case and the risk measure still remains coherent.[27]

As a response to the shortcoming of VaR, the coherent risk measure Expected Shortfall was introduced by Artzner et. al. (1997) and Delbaen (2002). ES is the expected value of the loss given that a VaR violation has occurred and expresses, in the same way as VaR, the expected risk in one figure.[28] Expected Shortfall can be stated as the average of the worst 100α % of losses:[29]

$$ES_{1-\alpha} = \frac{1}{\alpha} \int_{1-\alpha}^{1} q_p dp \qquad (2)$$

α: confidence level, e.g. 5%
q_p: the p-quantile where $p = \alpha$

In the same way as VaR, ES varies with the confidence level and holding period. Except for its role as a coherent risk measure, ES has other advantages over VaR. Value at Risk says nothing about the size of the loss exceeding the value of the estimated VaR, while ES specifies what happens in those bad states. ES is also consistent with expected utility maximization if portfolios can be ranked according to second order stochastic dominance, whereas VaR requires the more restrictive first order stochastic dominance.[30] Another advantage with Expected Shortfall is that the portfolio's risk surface will be convex implied by the subadditivity character of the risk measure. The convexity will make certain that a local optimum also is a global optimum, solving the optimization problem, which is not the fact with VaR.[31]

However, Expected Shortfall is also related to disadvantages. As ES considers the losses beyond VaR, the accuracy of the estimated tail of the distribution is especially important. With conventional estimation methods, such as Monte Carlo simulation, this estimation might be difficult, particularly under extreme market conditions.[32] With fat-tailed

[27] Acerbi and Tasche (2001)
[28] Angelidis and Degiannakis (2007)
[29] Dowd (2005) p. 35
[30] Dowd (2005) p. 35-36
[31] Acerbi and Tasche (2001)
[32] Yamai and Yoshiba (2002)

underlying distributions, the estimation error is much larger for ES than for VaR, and therefore, the sample size of the simulation should increase.[33] Another problem with ES is the backtesting of the accuracy of the model, which requires more data than the backtesting of VaR.[34]

2.5 VaR and ES

2.5.1 Parameters for VaR and ES

VaR and ES are based on two parameters, which have to be chosen with respect to the context in which the risk measure is used:[35]

- Holding period
- Confidence level

Additionally, when calculating ES and VaR, frequencies of the observations must be set. This frequency should be equal to, or lower, than the holding period. As an example, the frequency might be one day if the holding period is one week.[36] The following equations exemplify normal distributed VaR and ES, respectively, over a holding period of h and confidence interval $(1-\alpha)$:[37]

$$VaR(h, 1-\alpha) = -h\mu_R + \sqrt{h}\sigma_R z_{1-\alpha} \quad (3)$$

$$ES(h, 1-\alpha) = -h\mu_R + \sqrt{h}\sigma_R \frac{\phi(z_{1-\alpha})}{\alpha} \quad (4)$$

2.5.1.1 Holding period

Liquidity is one of four main aspects affecting the choice of holding period. According to this factor, the holding period shall be set equal to the time it takes to liquidate the position in that market. For example, thin markets can cause time-consuming liquidation and therefore, a longer holding period should be more suitable. Trading over the counter is assumed to be particularly thin. The second factor comes from a normal approximation which can be more justifiable for non-normally distributed asset returns when shorter holding periods are used. Using shorter holding periods, more observations will be obtained within a time horizon, compared to longer holding periods. VaR typically requires that the composition of a portfolio

[33] Yamai and Yoshiba (2005)
[34] Yamai and Yoshiba (2002)
[35] Dowd (1998) p. 50
[36] Culp (2001) p. 342-344
[37] Dowd (2005) p. 155

of assets remains the same over the holding period, which also motivates the use of short holding periods. Reliable validations require a large data set. Long holding periods will lead to observations collected under a much longer time horizon than for shorter holding periods and might include observations that are too old to be meaningful. This reason speaks for the use of shorter holding periods.[38]

2.5.1.2 Confidence level

The choice of confidence level is impacted by the assumed probability distribution and the purpose of the risk measure. For example, the process of determining the internal capital requirements, and providing inputs for internal risk management, will influence the choice of confidence level in different ways. Generally, a high confidence level can be used for validation, a medium, or low, level for accounting and comparison purposes, and a low confidence level for risk management and capital requirements.[39]

2.5.1.3 Sample

The observations in the in-sample period are used to estimate the models' risk values, which can be evaluated through a comparison to the out-of-sample observations. For instance, with a time period of seven years, the first five years can be defined as the in-sample period and used to estimate values for the remaining two years, the out-of-sample period. The estimates can be tested by comparing them to the out-of-sample observations[40], referred to as backtesting.[41]

A rolling window includes a constant number of observations, moving through the sample as the time passes by. A recursive window keeps the initial starting point and includes one more observation for each time sequence.[42]

2.5.2 Approaches

VaR can be calculated using one of the three approaches:[43]

- Non-parametric
- Parametric
- Semi-parametric

[38] Dowd (1998) p. 51-52
[39] Dowd (1998) p. 52-53
[40] Brooks (2008) p. 245
[41] Dowd (2005) p. 323
[42] Brooks (2008) p. 246
[43] Angelidis and Degiannakis (2007)

2.6 Non-parametric approach

The non-parametric approach estimates the risk measure without assuming central features of the returns' underlying distribution. The recent empirical distribution of the returns is used together with the assumption that the history will repeat itself in the future.[44] For assets which returns are not normally distributed, or when extreme events are likely to occur more frequently than the normal distribution states, the non-parametric approach is well performing.[45] Other appealing advantages with this method are its independence of parametric assumptions and its ability to make improvements and refinements when necessary. However, the method has a high dependence on the data set and does not allow possible events that have not occurred in the history.[46]

2.6.1 Historical simulation

Historical simulation is a histogram-based method and the most popular within the non-parametric approach. The future distribution of the returns is assumed to match the historical distribution, which can be used to simulate the asset's VaR and ES.[47] By plotting the returns in a histogram, Expected Shortfall and Value at Risk can easily be estimated by observing the value from the histogram. As an example, a sample of 10,000 observations at a confidence level of 99 %, VaR will be the 101st highest loss, and ES will be the average of the 100 highest losses.[48] This procedure will directly account for fat tails, skewness and other properties.[49] The approach is, among many things, simple to both implement and explain, it can accommodate non-normal features, avoids model risk, accommodates volatilities and correlations implicitly, and can be used on all types of instruments. However, the historical simulation has disadvantages primarily related to data problems. As it is completely dependent on the data set, only events that are represented in the collected historical data will have a probability of occurring in the future, and therefore, be covered in the risk measure. Another disadvantage is the ghost effect appearing as results of using a rolling window and equal weights for all observations in the in-sample. When the window moves, observations fall out of the sample and lead to "leaps" in the values of the risk measure.[50]

---

[44] Dowd (2005) p. 83
[45] Angelidis and Degiannakis (2007)
[46] Dowd (2005) p. 99-100
[47] Dowd (1998) p. 99-104
[48] Dowd (2005) p. 84
[49] Bodoukh et. al. (1998)
[50] Dowd (1998) p. 99-104

2.7 Semi-parametric approach

The semi-parametric approach combines the parametric and non-parametric methods. Two of the models included are filtered historical simulation and extreme value theory.[51] Weighted historical simulation models can also be referred to as semi-parametric approaches.[52]

2.7.1 Weighted historical simulation

The traditional historical simulation, discussed above, gives all past observations the same weight. An observation will therefore have a constant impact on the risk measure as long as it stays within the sample, and when it is too old and falls out of the sample, it has no effect. The traditional approach also assumes that the observations are IID (Independent and Identically Distributed). This assumption comes from the fact that this method treats the observations as if they were equally likely and independent of each other over time. By weighting the observations, it is possible to overcome some of the problems with the traditional historical simulation.[53]

2.7.1.2 Volatility-weighted historical simulation

The observations can be adjusted according to the changes in volatility and the beliefs about future volatility in relation to its past. This method will, in contrast to the traditional historical approach, not underestimate the future risk in periods of high volatility, or overestimate the risk when the volatility is low. The observations can be weighted with the following formula:[54]

$$r^*_{t,i} = \left(\frac{\sigma_{T,i}}{\sigma_{t,i}}\right) r_{t,i} \quad (5)$$

$\sigma_{t,i}$: historical forecast of the returns' volatility on asset *i* for day t made at day *t*-1
$\sigma_{T,i}$: the most recent forecast of the volatility of asset *i*
$r_{t,i}$: the return in the data set

By using the modified return $r^*_{t,i}$, observations in periods of lower volatility will have lower impact on the risk, and higher impact in periods of higher volatility.[55]

[51] Angelidis and Degiannakis (2007)
[52] Dowd (2005) p. 100
[53] Dowd (2005) p. 92-93
[54] Dowd (2005) p. 94-95
[55] Ibid.

2.7.1.3 Age-weighted historical simulation

The observations can also be adjusted according to their age, and for example, make the newer observations more important than older. The weights used in the traditional historical simulation, $1/n$, is replaced with $w(i)$ from the following formula:[56]

$$w(i) = \frac{\lambda^{i-1}(1-\lambda)}{1-\lambda^n} \tag{6}$$

λ: reflects the exponential rate of decay, $0 < \lambda \leq 1$
i: the figure reflecting the age of the observation, the most recent observation has value 1
n: number of observations

The traditional historical simulation can be expressed with the weight above using zero decay, equivalent to a decay factor of one, $\lambda = 1$. By using a value of λ different from one and a recursive window, there should no longer be any leaps in the sample as old observations no longer falls out. In this way, no information will be thrown away and it is possible to let the sample grow to sizes not handled under the traditional historical simulation approach.[57]

There is no statistical method for estimating the parameter λ. However, Christoffersen (2003) suggests that λ should be set equal to 0.95 and Boudoukh et al. (1998) that 0.97 or 0.99 should be used.[58]

2.8 Parametric approach

Under the parametric approach, a probability curve is fitted to the data. As this method makes use of the additional information, included in the assumed distribution function, it is more powerful than the non-parametric approach. However, the results depend to a large extent on the quality of the parameters estimated and assumed. Fitting the distributions conditionally, rather than unconditionally, is recommended and will, for example, take volatility clustering into account. Another pleasant aspect with this approach is that the formulas provided by the different distributions can be used for any confidence level and holding period.[59]

The continuously compound rate of return from time t–1 to t, y_t, can be expressed as follows:[60]

$$y_t = \ln{(P_t/P_{t-1})} \tag{7}$$

$$y_t = \mu_t + \varepsilon_t \tag{8}$$

[56] Dowd (2005) p. 93-94
[57] Ibid.
[58] Angelidis and Degiannakis (2007)
[59] Dowd (2005) p. 151, 154
[60] Angelidis and Degiannakis (2007)

$$\varepsilon_t = \sigma_t z_t \quad (9)$$

$$\sigma_t = g(\theta | I_{t-1}) \quad (10)$$

$$z_t \overset{i.i.d}{\sim} f(w; 0,1) \quad (11)$$

$\ln(P_t/P_{t-1})$: continuously compound return
μ_t: predictable component
ε_t: unpredictable part
σ_t: conditional standard deviation
g: volatility function
θ: vector of unknown parameters
I_t: information set available at time t
$f(.)$: density function of z_t
z_t ? $N(0,1)$
w: the vector of the parameters of $f(.)$

The conditional mean, μ_t, and conditional standard deviation, σ_t, depend on the information available at time t–1.[61] In this context, it has been shown that a complex arrangement of the conditional mean does not add any value in the prediction process.[62]

In general, the distribution of the returns will depend on the volatility of the unpredictable component ε_t. Returns can be made both skewed and heavy-tailed by a stochastic process, and is therefore important to take into account when using the parametric approach. The stochastic volatility process can, for example, be expressed as a GARCH.[63]

The one-step-ahead VaR and ES forecasts can under any distributions be computed as follows:[64]

$$VaR_{t+1|t}^{(p)} = \mu_{t+1|t} + F_{\alpha}(z_t; \xi^{(t)}, v^{(t)})\sigma_{t+1|t} \quad (12)$$

$$ES_{t+1|t}^{(p)} = E\left[y_{t+1} | \left(y_{t+1} \leq VaR_{t+1|t}^{(p)}\right)\right] \quad (13)$$

$F_a(z_t; \xi^{(t)}, v^{(t)})$: the a^{th} quantile of the assumed distribution, given information available at time t
$\mu_{t+1|t}$: conditional mean forecast at time t+1 based on information available at time t
$\sigma_{t+1|t}$: conditional standard deviation forecast at time t+1 based on information available at time t

For ES, no analytical solutions exist for all distributional assumptions, and the risk measure can therefore be calculated by taking the average of the VaRs in the tail.[65]

2.8.1 Normal distribution

The normal distribution only requires two parameters, the mean and the standard deviation, and is in many contexts well motivated by the central limit theorem. As the mean and standard deviation typically are unknown, it is necessary to work with their estimates. Formulas for VaR and ES are presented below:[66]

---

[61] Angelidis and Degiannakis (2007)
[62] Angelidis and Degiannakis (2004)
[63] Dowd (2005) p. 152-153
[64] Angelidis and Degiannakis (2007)
[65] Ibid.
[66] Dowd (2005) p. 154-157

$$VaR = -\mu_R + \sigma_R z_{1-\alpha} \tag{14}$$

$$ES = -\mu_R + \sigma_R \frac{\phi(z_{1-\alpha})}{\alpha} \tag{15}$$

A drawback with this parametric approach is that most financial returns are not normally distributed. They do rather have excess kurtosis.[67]

2.8.2 Student t-distribution

The Student-t distribution is symmetric around the mean and offers a method to manage the excess kurtosis, the fat tails. The symmetry has made authors to dismiss this approach in favor for asymmetric distributions that can manage the observed skewness of financial time series.[68] The kurtosis is related to the degrees of freedom, v, in the following manner:[69]

$$3(v-2)/(v-4) \tag{16}$$

A high kurtosis is a result of a low value for v, and vice versa. When the degrees of freedom reach a certain point, the Student-t distribution converges to a normal distribution. The formula for VaR given this distribution looks as follows:[70]

$$VaR(1-\alpha) = -\mu_R + \sqrt{\frac{v-2}{v}}\sigma_R t_{(1-\alpha),v} \tag{17}$$

$t_{(1-\alpha),v}$: confidence interval term, refers to a Student-t distribution

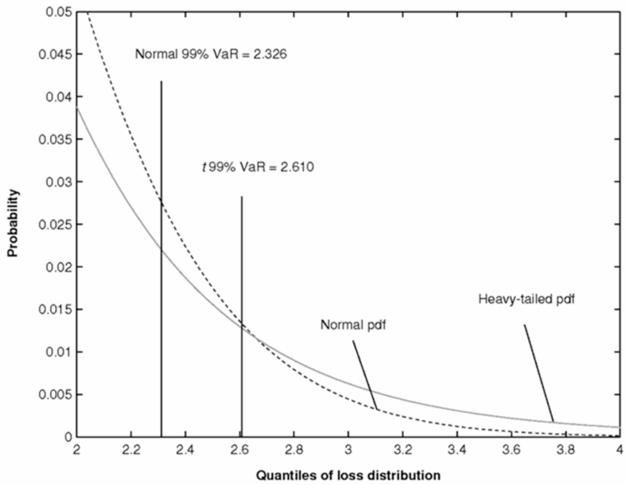


Figure 1. *The difference between the tails of normal and Student-t distribution.*[71]

[67] Dowd (2005) p. 154-157
[68] Angelidis and Degiannakis (2007)
[69] Dowd (2005) p. 160
[70] Dowd (2005) p. 159-160
[71] Dowd (2005) p. 158

2.8.3 Log-normal distribution

Assuming that the geometric returns are normally distributed is the same as assuming that the asset prices are log-normally distributed. A log-normal distribution is asymmetric, which implies that long and short trading positions do not have identical risk exposure. The loss of a long position is tied to the value of the investment, while a short position can lead to much greater losses. Another feature, except the asymmetry, is its consistence with a geometric Brownian motion process for the underlying assets' price. The formulas below describe a long and short position for VaR, respectively:[72]

$$VaR = 1 - Exp(\mu_R - \sigma_R z_{(1-\alpha)}) \tag{18}$$

$$VaR = -\left[1 - Exp\left(\mu_R + \sigma_R z_{(1-\alpha)}\right)\right] \tag{19}$$

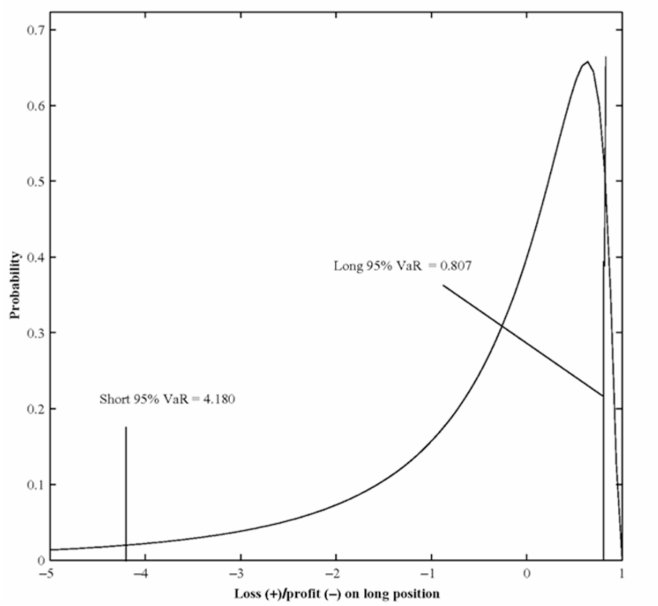


***Figure 2**. Long and short positions of the log-normal distribution.*[73]

[72] Dowd (2005) p. 161-163
[73] Dowd (2005) p. 163

2.9 Volatility modeling

Volatility is measured as the standard deviation of returns and can be estimated by a numerous of different models. A Generalised Autoregressive Conditionally Heteroscedastic model (GARCH) is one of these models and allows the conditional variance to be dependent upon its previous own lags. A GARCH model does not assume constant variance and will take volatility clustering into consideration. Following equation presents a GARCH(1,1):[74]

$$\sigma_{t+1|t}^2 = \alpha_0 + \alpha_1 u_t^2 + \beta \sigma_t^2 \quad (20)$$

$y_t = b_1 + b_2 x_2 + \varepsilon_t$
the residual equation: $\varepsilon_t = v_t \sigma_t$

Hansen and Lunde (2005) presented an article of a comparison of 330 volatility models in terms of the models ability to forecast the conditional variance on data consisting of one exchange rate (DM - $) and one stock price (IBM). Their results show that there is no model that outperforms the GARCH(1,1) model when it comes to either the exchange rate or the stock price.[75] However, studies of volatility models within the area of computing VaR point in different directions. Ané suggests that a GARCH(1,1) is sufficient when measuring the discussed risk,[76] while Billio and Pelizzon are not as convinced about its superiority.[77]

2.10 Backtesting

When a risk model has been constructed and presented, it should be statistically evaluated. One important feature in this context is backtesting, which is a quantitative method of determining whether the estimated results of a model are correct given the assumptions of the model.[78]

VaR forecasts can for example be tested with the Kupiec test and the Christoffersen test. However, those tests are not applicable when evaluating ES models. Instead, methods such as the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), the Generalized Mincer-Zarnowitz regression (GMZ) and ES ratio can be used.

2.10.1 Kupiec test

The Kupiec test, developed by Kupiec (1995), focuses on the property of unconditional coverage. The test examines whether the forecasted VaR is violated more or less than

[74] Brooks (2008) p. 383-392
[75] Hansen and Lunde (2005)
[76] Ané (2006)
[77] Billio and Pelizzon (2000)
[78] Dowd (2005) p. 321

100α % of the time. The accuracy of the risk model is questioned if the number of violations is significantly different from 100α %. That is, if the model systematically underestimates or overestimates VaR.[79] The Kupiec test is formulated as a $\chi^2(1)$ distributed Likelihood Ratio (LR) test. The formula is presented below:[80]

$$LR_{uc} = -2log\frac{(1-p)^{n_0}p^{n_1}}{(1-\pi)^{n_0}\pi^{n_1}} \sim \chi^2(1) \quad (21)$$

p: predicted probability of exceedances
n_1: number of exceedances
n_0: number of non-exceedances
π:empirical frequency of exceedances

Two drawbacks are related to tests of unconditional coverage. First, the Kupiec test has difficulties detecting VaR measures that systematically underestimate and overestimate the risk. Increasing the sample, over which the test is constructed, is one way to reduce the effects of this shortcoming. Second, as the focus only lies on the unconditional coverage property, the independence property is neglected. This means that the test pays no attention to whether the losses that exceed the forecasted VaR appear in clusters. A p-value below the chosen significance level indicates that the model can be rejected.[81]

2.10.2 Christoffersen test

Due to criticism of the Kupiec test, tests that explicitly investigate the independence property of VaR series were developed.[82] The Christoffersen test (1998) examines the independence against an explicit first-order Markov process.[83] This test evaluates whether or not a VaR violation depends on whether a violation took place the earlier day or not. A correct model should imply that the probability of having a violation of VaR is independent of the previous day's outcome.[84] Also, the Christoffersen test is formulated as a LR test, presented below:[85]

$$LR_{ind} = -2log\frac{(1-\pi_2)^{(n_{00}+n_{10})}\pi_2^{(n_{01}+n_{11})}}{(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}} \sim \chi^2(1) \quad (22)$$

n_{ij}: number of days that state *j* occurred after state *i* occurred the previous day, state refers to exceedance/non-exceedance

$\pi_{01} = \frac{n_{01}}{n_{00}+n_{01}}$ $\quad \pi_{11} = \frac{n_{11}}{n_{10}+n_{11}}$ $\quad \pi_2 = \frac{n_{01}+n_{11}}{n_{00}+n_{10}+n_{01}+n_{11}}$

---

[79] Campbell (2005)
[80] Christoffersen (1998)
[81] Campbell (2005)
[82] Ibid.
[83] Christoffersen (1998)
[84] Campbell (2005)
[85] Christoffersen (1998)

The main drawback with independence tests is the specification of the alternative hypothesis the independence property is tested against. As there exist numerous ways, in which the independence property may be violated, an independence test should describe all of these anomalies that the test evaluates. Violations that are not related to the defined anomalies will therefore not be detected. As for the Kupiec test, a p-value below the chosen significance level indicates that the model can be rejected.[86]

2.10.3 Joint test of Kupiec and Christoffersen

The Joint test combines the Kupiec unconditional coverage with the Christoffersen independence into one complete test of conditional coverage. The three tests are related as follows:[87]

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^2(2) \tag{23}$$

This Joint test will detect a VaR model which violates neither the unconditional coverage property, nor the independence property. However, a situation where only one of the two properties is fulfilled might lead to difficulties for the Joint test to detect the insufficiency of the VaR model.[88]

2.10.4 ES Ratio

The RiskMetrics Group, a financial services company, presents a way to evaluate the forecasted ES values. This measure will in this report be referred to as the ES ratio. The process of determining the ES ratio can be explained by the following steps:[89]

- Forecast ES_t for the quantile probability q and observe the actual return, R_t
- Choose the days which losses are within the q-tail, R_t< -VaR_t, and divide those return by the forecasted ES_t, R_t/ES_t
- Calculate the average value of those quotas, (R_t/ES_t)/number of VaR violations

A correct model should give an expected value of R_t equal to the estimated ES_t, and therefore, this model should have an average quota of one.[90] The quota calculated in this

[86] Campbell (2005)
[87] Christoffersen (1998)
[88] Campbell (2005)
[89] http://www.riskmetrics.com
[90] Ibid.

report is then subtracted by one meaning that a adequate model should have an ES ratio equal to zero. Also, as no distribution is stated regarding the test process, we chose to evaluate the values in relation to the properties of the adequate model.

2.10.5 MAE and RMSE

Following Lopez (1999), the previous discussed backtesting approaches of VaR models can come up with more than one adequate risk model, and therefore lead to difficulties in choosing one unique way of measuring risk. Lopez suggested a forecast evaluation framework that should overcome this shortcoming. However, this framework was criticised by Angelidis and Degiannakis in *Backtesting VaR Models: An Expected Shortfall Approach*, where they presented a modification of Lopez's approach. Instead of evaluating VaR, their suggestion is to use MAE and MSE to test ES's accuracy. The equations are presented below:[91]

$$\Psi_{1,t+1}^{(i)} = \begin{cases} \left| y_{t+1} - ES_{t+1|t}^{(i)} \right|, & if\ violations\ occur \\ 0, & else \end{cases} \tag{24}$$

$$MAE = T^{-1} \sum_{t=1}^{T} \Psi_{1,t+1}^{(i)} \tag{25}$$

$$\Psi_{2,t+1}^{(i)} = \begin{cases} (y_{t+1} - ES_{t+1|t}^{(i)})^2, & if\ violations\ occur \\ 0, & else \end{cases} \tag{26}$$

$$MSE = T^{-1} \sum_{t=1}^{T} \Psi_{2,t+1}^{(i)} \tag{27}$$

$$RMSE = \sqrt{MSE} \tag{28}$$

The most accurate model will give the lowest values for MAE and MSE, respectively.[92] In this report, the square root of MSE will be taken in order to simplify comparison between the two measures.

2.10.6 GMZ

Another approach when evaluating Expected Shortfall is by using a Generalized Mincer-Zarnowitz, GMZ, regression. ES is a conditional mean according to the following regression:[93]

$$E_t[ES_{t+1}] = E_t[R_{t+1} | R_{t+1} < -VaR_{t+1}] \tag{29}$$

[91] Angelidis and Degiannakis (2006)
[92] Ibid.
[93] http://www.kevinsheppard.com

Therefore, the GMZ regression can be used to test whether or not this mean is zero. The GMZ regression is presented below:[94]

$$\left(ES_{t+1|t} - R_{t+1}\right)I_{\left(R_{t+1} < -VaR_{t+1|t}\right)} = x_t\gamma \qquad (30)$$

$I_{\left(R_{t+1} < -VaR_{t+1|t}\right)}$: indicates that the return was less than the VaR
note: $\left(ES_{t+1|t} - R_{t+1}\right)$ will be a negative number when the return is less than the negative of ES, and vice versa

If the model for forecasting Expected Shortfall is correct, the regression should not have a significant intercept, $\gamma = 0$, and the model should not be rejected. If $\gamma \neq 0$, the VaR violated returns differ from the forecasted ES values, and the model can be improved. A p-value below the chosen significance level indicates that the model can be rejected.[95]

2.11 Previous research

Acerbi and Tasche wrote *Expected Shortfall: a natural coherent alternative to Value at Risk* in 2001, as a response to VaR's flaws regarding measuring the risk of portfolios. They argue that ES, as a coherent risk measure, captures all advantages of VaR. Moreover, it can be applied to any kind of risk and instrument and it considers the different risks that a portfolio is exposed to. Additionally, the article means that ES easily can replace the banks' present risk management system based on VaR, because of its simplicity and similarities to VaR.[96]

The article *Backtesting VaR Models: An Expected Shortfall Approach*, written by Angelidis and Degiannakis 2006, tests VaR and ES on S&P 500, gold and the exchange rate USD/GBP given four different distributions. The study was made at two confidence intervals, 95 % and 99 %, as well as for long and short trading positions. Angelidis and Degiannakis analyzed the data using a SPA model, among others, and found that the equity index, the commodity, and the exchange rate all fitted different distributions. However, some similarities among the distributions were found; for example, the Student-t and skewed Student-t distributions both overestimated the true VaR. The article concludes that different models can be chosen to explain the risks of S&P 500, gold and USD/GBP.[97]

The study *Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach*, by McNeil and Frey 2000, estimates VaR and other risk measures, such as ES, on heteroscedastic financial returns. The article combines pseudo-maximum-likelihood fitting of GARCH models to approximate volatility and extreme value

[94] http://www.kevinsheppard.com
[95] Ibid.
[96] Acerbi and Tasche (2001)
[97] Angelidis and Degiannakis (2006)

theory. McNeil and Frey state that their approximation is well suited for one-day estimations, since it captures the risks situated in the tails of the distributions. Their method uses multiple-day time horizons for conditional quantiles of returns and the authors mean that it surpasses the method called square-root-of-time scaling. The article concludes that ES is a good risk measure in sense of its theoretical properties compared to VaR.[98]

Market risk in commodity markets: a VaR approach is an article written by Giot and Laurent 2003 that focuses on long and short trading positions for commodity traders with a VaR approach. Six commodities are chosen for estimating the market risk, based on a one-day time horizon; aluminum, copper, nickel, two types of crude oil and cocoa. These commodities are backtested on RiskMetrics, skewed Student-t APARCH and skewed Student-t ARCH models, in which Student-t APARCH is best suited for each commodity. However, the skewed Student-t ARCH gives good approximations and its characteristics are simple.[99]

---

[98] McNeil and Frey (2000)
[99] Giot and Laurent (2003)

3 Methodology

3.1 S&P GSCI

The Standard & Poor's Goldman Sachs Commodity Index (S&P GSCI) is a benchmark for investments in the commodity markets. The index is recognized as a leading measure of inflation in the world economy and general commodity price movements. The index series includes several sub indices, such as those indices comprised of commodity sectors, thematic baskets and single components. The physical commodities included in the indices are traded on active and liquid futures markets. Three criteria have to be fulfilled for a contract to be included in the S&P GSCI:[100]

- General eligibility. The contract must be denominated in USD and traded on a trading service with the principal business place in a country member of OECD.
- Volume and weight. The contract must meet both the minimum reference percentage dollar weights and the total dollar value traded requirements.
- Number. Commodities are considered to constitute a single S&P GSCI commodity based on factors such as physical characteristics, production, trading, use and pricing.

The S&P GSCI Spot Index is based on price levels of the contracts included in the S&P GSCI.[101] The commodities studied in this report are presented in the table below:[102]

Official New Index Name
S&P GSCI Corn Index Spot
S&P GSCI Crude Oil Index Spot
S&P GSCI Gold Index Spot

Table 1. *S&P GSCI*

3.2 Method

The three commodities' daily prices are taken from Datastream in the Finance Lab on the School of Economics and Management at Lund University. The commodity data is based on S&P GSCI, and the returns are calculated as daily log-returns in Excel. The market risk is measured in USD, and since the commodities are priced in the same currency, no translation

[100] http://www.standardandpoors.com
[101] Ibid.
[102] http://www.goldman-sachs.ch

effects occur. The data period covers January 1st 1998 to December 30th 2009, and provides the study with 3,130 observations for each commodity. The in-sample period contains ten years, January 1st 1998 to December 31st 2007, resulting in an out-of-sample period between January 1st 2008 and December 30th 2009. However, the in-sample period varies depending on if the evaluated model is based on a rolling or recursive window. The recursive window is applied to the models Age-weighted HS and Volatility-weighted HS to avoid the loss of observations, whereas the remaining models are calculated with a rolling window.

Regarding VaR, the percentile function, and other regular formulas, have been used in Excel for calculating all VaR models manually, except the Volatility-weighted HS. Volatility-weighted HS has been calculated with a macro in Excel, which can be viewed in Appendix 1. Additionally, the Average VaR, VaR Rate, Kupiec, Christoffersen and Joint tests have been calculated in Excel.

The calculations concerning ES are, just like VaR, generally based on calculations in Excel. However, the Volatility-weighted HS model and the Age-weighted HS model are calculated with an Excel macro, see Appendix 1. Furthermore, the Average ES, ES Ratio, MAE and RMSE tests are calculated manually in Excel. Excel has also been used for calculating correlations between VaR and ES for each model and confidence interval.

EViews has been used for the Jarque-Bera test regarding the normality test of the data sample. Also, all plotted commodity returns are outputs from EViews, as well as estimations of GARCH(1,1), see Appendix 2.

4 Empirical results

In this chapter, the results are presented in the order gold, oil and corn. Furthermore, the results are categorized into each confidence interval. This order has been chosen with respect to the layout of the result tables. However, the analysis is more interesting to discuss for each commodity and risk measure and therefore another sorting is chosen for the next chapter. Note that the GMZ test for ES, in some cases, includes too few observations to give results possible for interpretation, stated as NA. It should also be noted that the power of the various tests are lower at the 99 % confidence interval than at the 95 % confidence interval for both VaR and ES. The results are presented in Table 2 to Table 7, as well as in Appendix 3 to Appendix 6. Note that the scales of the plots are different in Appendix 3 to Appendix 5. The results for the Log-Normal model are risk measures given long positions. The correlations between VaR and ES for all models, respectively, are high and positive, and the majority of the models are perfectly positively correlated. The correlations can be viewed in Appendix 7 to Appendix 9.

A decay factor of 0.999, when calculating Age-weighted HS, has been chosen. When using the recommended 0.95, 0.97 or 0.99, the differences between the observations' weights would have been too large and difficult to motivate for our data.

4.1 Gold

The average degrees of freedom used when calculating the Student-t distribution for gold is 4.945 and the kurtosis varies between 8.190 and 10.413. The daily returns of gold during the time period 1998-01-01 to 2009-12-30 can be seen in the figure below. For a more specific illustration of the returns during the out-of-sample period, see Figure 6 on page 37. The Jarque-Bera test, presented in Appendix 10, shows no normality for the 3,130 observations.

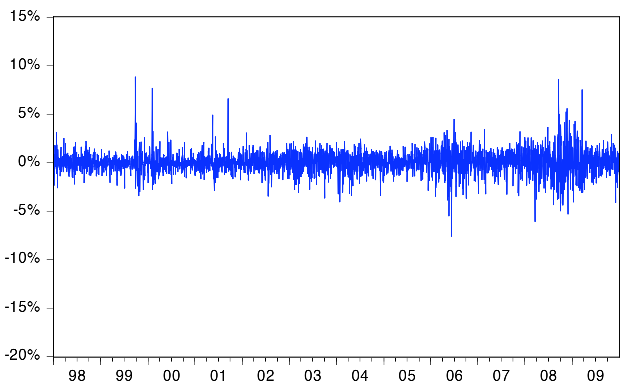


Figure 3. *Gold daily returns during 1998 to 2009.*

4.1.1 Gold 95 %

A summary of the results at the 95 % confidence interval for gold are shown in Table 2, more extensive results are presented in Appendix 3 and Appendix 6. The table describes a great variation of Average VaR values and Average ES values. The VaR Rates all exceed 5 %, which is the value expected at the 95 % confidence interval. At the same confidence interval,

Gold 95 %

	Value at Risk					Expected Shortfall							
Approach	**Av. VaR**	**VaR Rate**	**Kupiec**	**Chris**	**Joint**	**Av. ES**	**ES Ratio**	**MAE**		**RMSE**		**GMZ**
Historical Simulation	1.750	12.835	< 0.01	< 0.01	< 0.01	2.485	0.089	0.356	(7)	1.452	(8)	0.064
Volatility-weighted HS	2.642	5.172	0.857	0.611	0.589	3.732	-0.015	0.102	(1)	0.834	(1)	0.418
Age-weighted HS	2.037	9.579	< 0.01	0.131	< 0.01	2.971	0.008	0.219	(5)	1.227	(5)	0.940
Normal	1.787	12.452	< 0.01	0.064	< 0.01	2.245	0.227	0.427	(8)	1.527	(9)	< 0.01
Normal GARCH(1.1)	2.609	5.172	0.857	0.611	0.589	3.273	0.124	0.175	(4)	1.059	(4)	0.177
Student-t	1.770	12.452	< 0.01	0.064	< 0.01	2.621	0.042	0.324	(6)	1.415	(7)	0.399
Student-t GARCH(1.1)	2.590	5.364	0.706	0.681	0.577	3.834	-0.051	0.103	(2)	0.843	(2)	0.127
Log-Normal	1.771	12.644	< 0.01	0.036	< 0.01	2.219	0.234	0.434	(9)	1.291	(6)	< 0.01
Log-Normal GARCH(1.1)	2.571	5.364	< 0.01	0.681	< 0.01	3.211	0.129	0.174	(3)	1.053	(3)	0.163

***Table 2**. Av. VaR, VaR Rate and Av. ES are stated in percent. Kupiec, Chris. (Christoffersen), Joint and GMZ are stated in p-values and are not rejected when the p-values are higher than 0.01. ES Ratio is expressed in quota scale and MAE and RMSE are ordinal scales, ranked in ascending order.*

the Volatility-weighted HS, Normal GARCH(1,1) and Student-t GARCH(1,1) are the models that can not be rejected in the Kupiec test, whereas the Historical Simulation is the only model that can be rejected in the Christoffersen test. Since Volatility-weighted HS, Normal GARCH(1,1) and Student-t GARCH(1,1) are the only models with p-values higher than 0.01 in the tests Kupiec and Christoffersen, these are also the single models that can not be rejected in the Joint test.

Regarding the ES Ratio, values close to zero can be found for all models. Furthermore, the most preferred MAE and RMSE results are close to zero, stated as a ranking in Table 2. The two models Normal and Log-Normal, have p-values below 0.01 in the GMZ test, meaning that the models can not be rejected.

4.1.2 Gold 99 %

Compared to the 95 % confidence interval, the Average VaR values are, by definition, higher at the 99 % confidence interval, see Table 3. Moreover, the averages show variation. The same pattern is stated for Average ES. The results of the Kupiec test explain four models that can not be rejected; Volatility-weighted HS, Age-weighted HS, Normal GARCH(1,1) and

Student-t GARCH(1,1). As a consequence, these models can not be rejected in the Joint test, solely. Regarding the MAE test, the Volatility-weighted HS model presents the value closest to zero. Additionally, the Volatility-weighted HS model has the best RMSE result. However, the best result in the ES Ratio test is given by Age-weighted HS. The ranking of MAE and RMSE tests are shown in Table 3. The GMZ test gives seven models with p-values above 0.01, indicating that the models can not be rejected.

Gold 99 %

	Value at Risk					Expected Shortfall						
Approach	**Av. VaR**	**VaR Rate**	**Kupiec**	**Chris**	**Joint**	**Av. ES**	**ES Ratio**	**MAE**		**RMSE**		**GMZ**
Historical Simulation	3.104	2.682	< 0.01	0.380	< 0.01	3.761	0.119	0.171	(7)	1.181	(7)	0.091
Volatility-weighted HS	4.537	0.575	0.289	0.852	0.281	5.384	0.273	0.033	(1)	0.539	(1)	0.336
Age-weighted HS	3.507	2.107	0.027	0.491	0.020	4.547	0.001	0.073	(5)	0.832	(5)	0.941
Normal	2.546	5.747	< 0.01	0.345	< 0.01	2.917	0.199	0.240	(8)	1.271	(8)	< 0.01
Normal GARCH(1.1)	3.709	1.533	0.257	0.104	0.047	4.246	0.208	0.055	(3)	0.642	(3)	0.200
Student-t	3.119	2.874	< 0.01	0.346	< 0.01	4.269	-0.027	0.103	(6)	0.967	(6)	0.616
Student-t GARCH(1.1)	4.560	0.575	0.289	0.852	0.281	6.249	0.122	0.035	(2)	0.565	(2)	0.607
Log-Normal	2.514	5.747	< 0.01	0.345	< 0.01	2.874	0.216	0.250	(9)	1.533	(9)	< 0.01
Log-Normal GARCH(1.1)	3.632	1.533	< 0.01	0.104	< 0.01	4.144	0.228	0.069	(4)	0.719	(4)	0.165

Table 3. *Av. VaR, VaR Rate and Av. ES are stated in percent. Kupiec, Chris. (Christoffersen), Joint and GMZ are stated in p-values and are not rejected when the p-values are higher than 0.01. ES Ratio is expressed in quota scale and MAE and RMSE are ordinal scales, ranked in ascending order.*

4.2 Oil

As can be seen in Appendix 11, the Jarque-Bera test for oil illustrates non-normality. The Student-t distribution's average degrees of freedom is 6.050 and the kurtosis interval is between 5.486 and 6.417. Figure 4, illustrating the daily returns, demonstrates higher movements than the returns of gold. Furthermore, the volatility tends to be higher during the study's out of sample period, which is explicitly presented in Figure 8 on page 43.

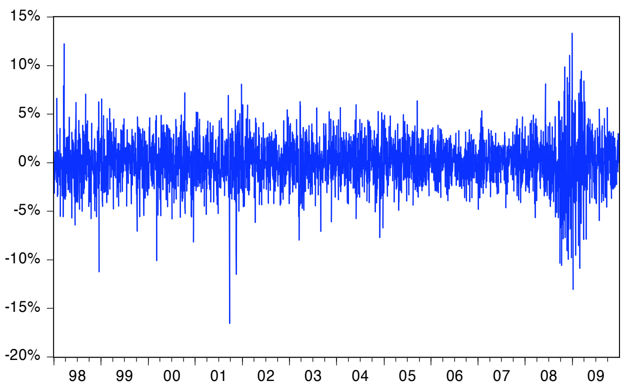


Figure 4. *Oil daily returns during 1998 to 2009.*

4.2.1 Oil 95 %

Oil's Average VaR values are higher than the averages of gold at the confidence interval of 95 %. Moreover, when comparing oil in Table 4 to gold in Table 2, the VaR values for oil are even higher than the Average ES figures for gold at this confidence interval. As regards the VaR Rate, no models fall below the 5 % level. All models have higher p-values than 0.01 under the Christoffersen test. However, the Joint test can not be rejected for Volatility-weighted HS and Normal GARCH(1,1) since these models exclusively have p-values above 0.01 under the Kupiec test.

Oil 95 %

	Value at Risk					Expected Shortfall				
Approach	**Av. VaR**	**VaR Rate**	**Kupiec**	**Chris**	**Joint**	**Av. ES**	**ES Ratio**	**MAE**	**RMSE**	**GMZ**
Historical Simulation	3.592	10.728	< 0.01	0.381	< 0.01	5.054	-0.089	0.778 (7)	3.139 (8)	0.074
Volatility-weighted HS	4.948	6.13	0.252	0.424	0.162	6.712	-0.079	0.270 (1)	1.960 (2)	0.381
Age-weighted HS	3.683	9.962	< 0.01	0.394	< 0.01	5.425	-0.081	0.763 (5)	3.147 (9)	0.167
Normal	3.667	10.536	< 0.01	0.327	< 0.01	4.599	-0.007	0.818 (9)	3.127 (7)	0.877
Normal GARCH(1.1)	4.685	7.471	0.015	0.543	0.013	5.872	0.053	0.431 (3)	2.559 (3)	0.561
Student-t	3.608	10.92	< 0.01	0.439	< 0.01	5.079	-0.086	0.774 (6)	3.126 (6)	0.115
Student-t GARCH(1.1)	4.610	7.663	< 0.01	0.968	< 0.01	6.485	-0.031	0.278 (2)	1.952 (1)	0.743
Log-Normal	3.601	10.92	< 0.01	0.439	< 0.01	4.492	0.016	0.811 (8)	3.062 (5)	0.791
Log-Normal GARCH(1.1)	4.557	7.854	< 0.01	0.893	< 0.01	5.665	0.074	0.506 (4)	2.721 (4)	0.442

Table 4*. Av. VaR, VaR Rate and Av. ES are stated in percent. Kupiec, Chris. (Christoffersen), Joint and GMZ are stated in p-values and are not rejected when the p-values are higher than 0.01. ES Ratio is expressed in quota scale and MAE and RMSE are ordinal scales, ranked in ascending order.*

The ES Ratios show negative numbers overall and the values, together with the Average ES, are presented in Table 4. Similar to the Average VaR values and Average ES values, the MAE and RMSE for oil are relatively high compared to gold. Finally, the GMZ states that no models can be rejected.

4.2.2 Oil 99 %

The Average ES values cover a large interval; from 5.789 to 9.881, which might entail in a variety of interpretations when evaluating the models. Comparable to the 95 % confidence interval of oil, the 99 % interval of ES Ratios are generally negative. The models are ranked according to their respectively values of MAE and RMSE, shown in Table 5. Since five models have too few observations in the regressions, the GMZ can not be stated. The models Normal, Normal GARCH(1,1), Log-Normal and Log-Normal GARCH(1,1) can not be rejected in the GMZ.

	Oil 99 %									
	Value at Risk					Expected Shortfall				
Approach	Av. VaR	VaR Rate	Kupiec	Chris	Joint	Av. ES	ES Ratio	MAE	RMSE	GMZ
Historical Simulation	6.002	4.789	< 0.01	0.846	< 0.01	7.871	-0.135	0.369 (7)	2.552 (7)	NA
Volatility-weighted HS	7.836	1.341	0.457	0.075	0.054	9.433	-0.047	0.038 (1)	0.868 (2)	NA
Age-weighted HS	6.432	4.023	< 0.01	0.864	< 0.01	8.590	-0.047	0.256 (5)	2.219 (5)	NA
Normal	5.211	6.130	< 0.01	0.166	< 0.01	5.966	-0.050	0.663 (8)	3.061 (8)	0.488
Normal GARCH(1.1)	6.651	2.299	0.011	0.269	< 0.01	7.612	0.035	0.206 (4)	1.731 (4)	0.837
Student-t	5.974	4.789	< 0.01	0.846	< 0.01	7.739	-0.136	0.365 (6)	2.524 (6)	NA
Student-t GARCH(1.1)	7.626	1.724	0.132	0.138	0.034	9.881	-0.048	0.038 (1)	0.867 (1)	NA
Log-Normal	5.078	6.130	< 0.01	0.166	< 0.01	5.789	-0.023	0.676 (9)	3.100 (9)	0.742
Log-Normal GARCH(1.1)	6.394	2.490	< 0.01	0.323	< 0.01	7.273	0.061	0.203 (3)	1.702 (3)	0.744

Table 5. *Av. VaR, VaR Rate and Av. ES are stated in percent. Kupiec, Chris. (Christoffersen), Joint and GMZ are stated in p-values and are not rejected when the p-values are higher than 0.01. ES Ratio is expressed in quota scale and MAE and RMSE are ordinal scales, ranked in ascending order.*

As has been seen before, few results for oil show Kupiec values above 0.01 and consequently, few models in the Joint test can not be rejected. However, all models show p-values higher than 0.01 in the Christoffersen test. The Student-t GARCH(1,1) model and Volatility-weighted HS model have VaR Rates close to 1 %. Exactly as at the 95 % confidence interval for oil, the Average VaR values are higher than gold's Average VaR and ES values at the 99 % confidence interval.

4.3 Corn

Figure 5 illustrates the movements of corn's daily returns, which tend to increase the most recent years of the sample and are specifically are shown in Figure 9 on page 46. In Appendix 12, the probability value of the Jarque-Bera test implies that the null hypothesis of normality can be rejected. The average degrees of freedom for oil is 7.388 and is used when calculating the market risk for the Student-t distribution. The kurtosis lies between 4.403 and 5.228.

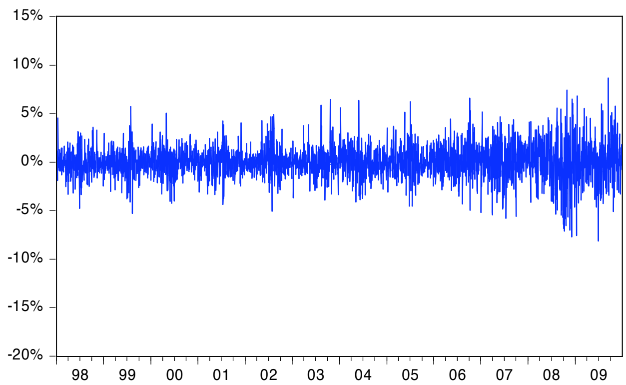


Figure 5. *Corn daily returns during 1998 to 2009.*

4.3.1 Corn 95 %

The Average VaR results, of the models respectively, are generally lower than their respective values for oil and higher than for gold. All VaR Rates are above the confidence level of 5 %, in which the Volatility-weighted HS model has the lowest VaR Rate. This model is also the single model with a Kupiec result indicating that it can not be rejected. The results of the Christoffersen test are low and therefore differ from previous commodities. Since Volatility-weighted HS is the only model with a p–value above 0.01 in the Kupiec test, but can be rejected in the Christoffersen test, the Joint test results are all below 0.01.

The GMZ shows high values for Student-t and Age-weighted HS, but all models have p-values above 0.01, indicating that the models can not be rejected. All models have ES Ratios close to zero. MAE and RMSE are ordered according to their size in Table 6, with the models Volatility-weighted HS and Student-t GARCH(1,1) firstly and secondly ordered, respectively. The average values of ES for corn, at the confidence interval of 95 %, vary between 3.236 and 5.123.

Corn 95 %

	Value at Risk					Expected Shortfall							
Approach	**Av. VaR**	**VaR Rate**	**Kupiec**	**Chris**	**Joint**	**Av. ES**	**ES Ratio**	**MAE**		**RMSE**		**GMZ**
Historical Simulation	2.578	13,602	< 0.01	0.026	< 0.01	3.567	-0.019	0.690	(6)	2.436	(7)	0.657
Volatility-weighted HS	3.882	6,513	0.129	< 0.01	< 0.01	5.123	-0.105	0.264	(1)	1.716	(1)	0.040
Age-weighted HS	3.042	9,962	< 0.01	0.011	< 0.01	4.269	-0.020	0.567	(5)	2.325	(6)	0.706
Normal	2.627	13,410	< 0.01	0.046	< 0.01	3.291	0.082	0.745	(8)	2.465	(9)	0.156
Normal GARCH(1.1)	3.673	7,854	< 0.01	0.042	< 0.01	4.598	-0.058	0.393	(3)	2.013	(3)	0.314
Student-t	2.600	13,602	< 0.01	0.059	< 0.01	3.563	0.001	0.712	(7)	2.456	(8)	0.940
Student-t GARCH(1.1)	3.634	8,046	< 0.01	0.017	< 0.01	4.977	-0.115	0.323	(2)	1.896	(2)	0.613
Log-Normal	2.593	13.602	< 0.01	0.059	< 0.01	3.236	0.100	0.753	(9)	2.207	(5)	0.088
Log-Normal GARCH(1.1)	3.603	8.238	< 0.01	0.022	< 0.01	4.485	-0.037	0.415	(4)	2.035	(4)	0.507

Table 6. *Av. VaR, VaR Rate and Av. ES are stated in percent. Kupiec, Chris. (Christoffersen), Joint and GMZ are stated in p-values and are not rejected when the p-values are higher than 0.01. ES Ratio is expressed in quota scale and MAE and RMSE are ordinal scales, ranked in ascending order.*

4.3.2. Corn 99 %

The Average ES values of Student-t GARCH(1,1) are higher than for Student-t. This relation can also be seen between Normal GARCH(1,1) and Normal, as well as for Log-Normal GARCH(1,1) and Log-Normal. As regards MAE and RMSE, Student-t GARCH(1,1) and Volatility-weighted HS are the best performing models. However, as these models have few observations in their respectively regression, no values can be stated for them in the GMZ test, as can be seen in Table 7. Four models have p-values above 0.01 in the GMZ test. The ES Ratios are generally close to zero.

Corn 99 %

	Value at Risk					Expected Shortfall						
Approach	**Av. VaR**	**VaR Rate**	**Kupiec**	**Chris**	**Joint**	**Av. ES**	**ES Ratio**	**MAE**		**RMSE**		**GMZ**
Historical Simulation	4.308	5,364	< 0.01	0.248	< 0.01	5.146	-0.002	0.344	(6)	1.978	(7)	0.915
Volatility-weighted HS	6.108	1,149	0.738	0.709	0.616	6.686	0.006	0.076	(2)	1.013	(2)	NA
Age-weighted HS	5.049	3,257	< 0.01	0.575	< 0.01	6.185	-0.042	0.267	(5)	1.850	(5)	0.715
Normal	3.727	7,663	< 0.01	0.032	< 0.01	4.264	-0.234	0.502	(9)	2.226	(8)	< 0.01
Normal GARCH(1.1)	5.205	2,299	0.011	0.025	< 0.01	5.954	0.118	0.147	(3)	1.383	(3)	NA
Student-t	4.161	5,747	< 0.01	0.109	< 0.01	5.239	-0.035	0.344	(6)	1.975	(6)	0.577
Student-t GARCH(1.1)	5.810	1,533	0.257	0.618	0.215	7.313	-0.084	0.027	(1)	0.622	(1)	NA
Log-Normal	3.658	7.854	< 0.01	0.042	< 0.01	4.173	0.085	0.497	(8)	2.467	(9)	0.191
Log-Normal GARCH(1.1)	5.065	2.682	< 0.01	0.050	< 0.01	5.769	0.149	0.193	(4)	1.506	(4)	NA

Table 7*. Av. VaR, VaR Rate and Av. ES are stated in percent. Kupiec, Chris. (Christoffersen), Joint and GMZ are stated in p-values and are not rejected when the p-values are higher than 0.01. ES Ratio is expressed in quota scale and MAE and RMSE are ordinal scales, ranked in ascending order.*

At the 99 % confidence interval, corn has three models that can not be rejected in the Kupiec test. All models show p-values above 0.01 in the Christoffersen test. However, only two models can not be rejected in the Joint test, explained by low Kupiec and Christoffersen values for Normal GARCH(1,1). All models have a VaR Rate higher than 1 %. The Average VaR varies between 3.658 and 6.108.

5 Analysis

The analysis will discuss the three commodities individually and evaluate the two confidence intervals simultaneously. This chapter focuses on interpreting the empirical results referring to the purpose of report; to evaluate if ES complements VaR and to find a model that accurate estimates the market risk for each commodity. Following this, the analysis first evaluates VaR in detail, as it can be seen as the foundation of measuring market risk, to later analyze ES as the complementing measure. With respect to the reader, gold, as the first commodity in order, will be discussed in more detail and will serve as reference material for the remaining commodities to avoid unnecessary repetition. Appendix 3 to Appendix 12 delivers detailed tables and figures of VaR and ES.

The in-sample period, which covers the years 1998 to 2007, is relatively to the out-of-sample period, a calm and less volatile period for all three commodities. The out-of-sample period is characterized by the recent financial crisis, which has impacted the financial markets. The use of rolling window, compared to recursive window, makes the different characteristics between the in-sample period and the out-of-sample period less apparent. As stated in section 2.1, commodities, such as gold, oil and corn, are typically affected by business cycles and consequently, so are their returns. An illustration of the increased volatility during the years 2008 and 2009 can be seen in Figure 6, Figure 8 and Figure 9. As the in-sample period is used for forecasting values for the out-of-sample period, the differences in their characteristic might have impacted the results negatively. However, it is possible that a crisis of this caliber will be repeated and therefore, a functioning model should also be applicable when the in-sample period differs from the out-of-sample period. Consequently, our results should not be neglected.

Generally, when analyzing the results, it is important to be critical against the outcome of the different statistical tests. As mentioned in previous sections, the tests in this study should be questioned regarding their advantages and disadvantages, see section 2.10. Moreover, the left tail only contains 5 % and 1 % of the data at the 95 % and 99 % confidence intervals, respectively. As an example, year 2008 contains 262 trading days in our data, meaning that 2.62 observations should, statistically, be found in the normally distributed tail at a 1 % confidence level. To enable conclusive ES analyzes, we chose to study two years instead of the normalized one year.

We believe that it will be interesting to analyze the models Normal, Student-t and Log-Normal. That is, we are critical against the common use of normal distributions in financial modeling, since it forces the data to fit a distribution which might be incorrect. The same critique can be applied to Student-t and Log-Normal.

5.1 Gold

As can be seen in Figure 6, the volatility of gold is lower compared to the other commodities. During quarter three 2008 to quarter one in 2009, the volatility increased in relation to the other quarters. This pattern might be an evidence for volatility clustering, and therefore, motivate better results for models including time varying volatility, Volatility-weighted HS and Log-Normal GARCH(1,1) as examples. The plotted VaR and ES can be found in Appendix 3.

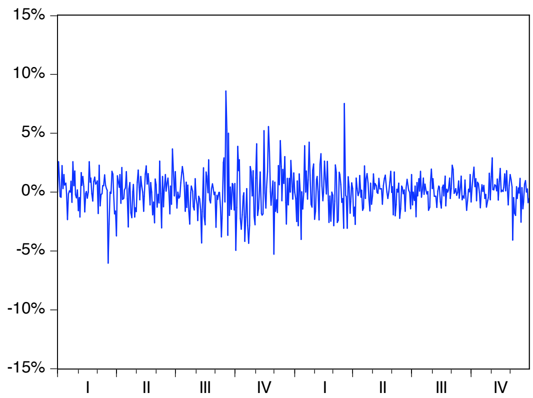


Figure 6. *Gold daily returns during 2008 and 2009.*

5.1.1 Gold – Value at Risk

All models that do not include time varying volatility tend to lag the returns. They do not increase in value at the same speed as the returns increase in volatility. Furthermore, since the VaR remains at a high level, they do not follow the reduction in volatility that started in quarter two 2009. Though, the parametric models including time varying volatility, and the Volatility-weighted HS, share similar shapes and seem to fit the pattern of the returns in Figure 6 better. The lag is no longer apparent and the shapes of the curves demonstrate larger movements. The models with time varying volatility illustrate risk with time varying values, which therefore should capture the volatility clustering better. Overall, they also have higher risk estimates than their corresponding non-time varying models. Comparing the three GARCH(1,1) models to the non-time varying parametrical models, the former give better results, shown in Table 2 and Table 3, suggesting that the returns show volatility clustering. Referring to Figure 6, these results are not inconceivable and are in line with the previous discussion regarding the differences of the in-sample and out-of-sample periods. The discussion above regarding the improvements of applying time varying volatility models is not in line with what banks prefer, stated in section *1.2 Problem discussion*. Banks favor non-parametric methods by reason of too volatile day-to-day risk estimates from parametric methods. The question is then; shall this analysis consider what banks prefer due to cost reasons, as an example, or shall the most adequate model be suggested? We still believe that the latter should be prioritized.

The criticism of Historical Simulation, the ghost effect mentioned in section 2.6.1, can clearly be seen in Appendix 3 as its curve tend to have a "stair" shaped pattern. This pattern appears at both confidence intervals. However, it is even more evident for the confidence interval of 99 %. This effect also occurs in the Age-weighted HS, but not in the Volatility-weighted HS. An advantage with Historical Simulation is that it can be applied to any kind of financial instruments, which is another argument for the banks common use of non-parametrical models. We consider this argument acceptable compared to the argument brought up in the previous paragraph.

The Average VaR values are, as ought to be, higher on the 99 % confidence interval. Though, the differences between the confidence intervals for the models Log-Normal and the Log-Normal GARCH(1,1) are slightly lower than for the rest of the models. At the 95 % level, all VaR Rates are above 5 %, which indicates underestimations of the forecasted VaR values. On the other hand, the 99 % confidence interval shows VaR Rate values below 1 % for the models Volatility-weighted HS and Student-t GARCH(1,1), implying an overestimation of the 99 % VaR. As will be discussed later in this section, the cross point between the Student-t distribution and the normal distribution explains the results regarding the Student-t GARCH(1,1) model. Referring to Figure 5, when analyzing the Volatility-weighted HS model at the 1 % level, the in-sample period contains large negative returns of similar sizes as the out-of-sample period. For instance, the estimated VaR value 2008-01-01 refers to the 27th lowest observation in the in-sample, which is an outlier of larger, negative size compared to what the forecasted VaR should be according to the backtesting. At the 5 % level, the estimated VaR 2008-01-01 refers to the 131st lowest observation in the in-sample, though this observation is too low relative to the accurate estimation of the VaR.

The best models, regarding the VaR Rates, are Volatility-weighted HS, Normal GARCH(1,1), Student-t GARCH(1,1) and Log-Normal GARCH(1,1) at both confidence intervals. The former three models can not be rejected in the Kupiec test at either confidence interval. The Age-weighted HS model only has a p-value above 0.01 in the Kupiec test at the

99 % confidence level and an interpretation of this is that the model should be suited after the confidence interval. The critique against Kupiec, presented in section 2.10.1, is toned down when introducing the Christoffersen test. Regarding this latter test, the results imply that the models, almost exclusively, can not be rejected at either confidence interval. However, the critique mentioned in 2.10.2 must be considered, since only one relation of dependence has been analyzed in this study. Therefore, weekly and monthly effects, as examples, are neglected. Since Table 3 separates the values of Kupiec, Christoffersen and Joint, it is apparent in which direction the values affect the Joint test. This separation eliminates the drawback of this test and makes it clear why only the models that can not be rejected in the Kupiec test, mentioned earlier, have higher p-values than 0.01 in the Joint test.

Considering Historical Simulation as a standard model, and Volatility-weighted HS and Age-weighted HS as modifications of this simple model, it is interesting to see how the results differ among the three models. The Age-weighted HS model can be rejected in the Joint test at the 5 % level, due to a low value in the Kupiec test. Yet, at the 1 % level, the Joint p-value is higher than 0.01. With respect to the characteristics of this model, we see two possible explanations. First, the majority of the violations are covered within the 5 % of the histogram at the 95 % confidence interval, but not within the 1 %. Second, the 5 % level contains older observations than the 1 % level of the histogram. Since older observations are given lower weight, more of these violations can be included in the interval, which will lower the estimated VaR and therefore increase the VaR Rate.

Concerning the normal distribution, the Jarque-Bera test implies that the gold returns are not normally distributed, and therefore, it is interesting to analyze how accurate the Normal model is in this context. Regarding all parametrical models, it is of interest to interpret the shape of the different distributions. The structure of the log-normal distribution implies lower market risk for long positions and higher market risk for short positions, which is a result of the high probability of a value closer to the right part of the curve for the former position, see Figure 2. The models Log-Normal and Log-Normal GARCH(1,1) have slightly lower Average VaR and Average ES values than their corresponding Normal and Normal GARCH(1,1) models. Again, regarding the shape of the distribution, these results are not surprising. Due to the log-normal distribution's asymmetry, the market risk in terms of VaR was also calculated for a short position to easily understand the characteristic of this distribution. At the 95 % confidence interval, the Average VaR for a short position is 1.894 and 2.670 at the 99 % confidence interval. Due to the divergence between long and short positions, we believe that it is important to understand the difference and take it into account in this context. Especially since Angelidis and Degiannakis (2007) demonstrate that this asymmetric model manages the skewness of time series, which is observed in financial markets.

Comparing the normal distribution to the Student-t distribution, the latter has fatter tails, but will converge to a normal distribution as the degrees of freedom reach infinity, see Figure 1. The probabilities of large positive and large negative values are higher with the Student-t distribution, as a result of the fat tails mentioned above. Continuously, the Normal model shows higher Average VaR values than the Student-t model at the 95 % confidence interval. Vice versa, the Student-t model has higher values than the Normal model at the 99 % confidence interval. Three parameters explain the Student-t distribution; the mean, the standard deviation and the degrees of freedom. Our interpretation is that these parameters, estimated from the returns of gold, lead to a shape of the curve, in which the 95 % interval lies at a position between the mean and the point where the Student-t and Normal curves cross each other. Furthermore, the 99 % interval lies to the left of the point where they cross. This interpretation explains the results where the Normal model has higher values at the 95 % interval, in contrast to the Student-t model which have higher results at the 99 % interval. This discussion is also applicable to their corresponding GARCH(1,1) models. The argumentation is valid for oil and corn as well, and will therefore be argued briefly in section 5.2 and 5.3.

5.1.2 Gold – Expected Shortfall

The forecasted values for VaR and ES for the models, except Historical Simulation, Age-weighted HS, Student-t at both confidence intervals, and Volatility-weighted HS at the 99 % confidence interval, are perfectly, positively correlated, see Appendix 7. The lowest correlation, estimated for Age-weighted HS 99 %, is 0.923. Student-t is, together with Normal and Log-Normal, a parametrical model and an exception among these models which is explained by the kurtosis. That is, the additional parameter that explains the Student-t distribution. The Student-t distribution is characterized by its leptokurtic feature, the fat tails, in which the focus lies when calculating ES. For the 522 observations of gold, the kurtosis varies between 8.190 and 10.413. This interval is smaller for both oil and corn. Naturally, the fat tails of the Student-t distribution contains more large negative observations than the normal distribution and the Student-t model shows higher ES values than the Normal model at both confidence intervals. Comparing with the discussion of VaR and the cross point of the two distributions, the same obvious pattern can not be found for ES and the explanation lies in the calculation of ES. That is, as ES is calculated as the average of all observations beyond

VaR, the value will logically be higher for the distribution that contains more large observations in the tail. Especially at high confidence intervals, the tail observations have low probability and values far from the sample mean. This argument will be discussed for each commodity in section 5.2.2 and 5.3.2.

Focusing on the plotted values in Appendix 3, VaR and ES have nearly identical patterns, except for the different scales. Together with the former discussion about the correlation, these facts could suggest that ES calculations add no value to the risk measure, except for higher market risk estimates. These results could easily be attained by scaling the VaR values upwards. As mentioned in the theoretical section, subadditivity is one of the criticisms against VaR which is a property VaR only fulfills when the return distribution is elliptically distributed. Normal distribution is one of these elliptical distributions. Subadditivity encourages portfolio diversification, which is a phenomenon that is widely accepted in the financial theory. If the three commodities were studied in a portfolio perspective, the ES results would probably have added value to the report.

The ghost effects of Historical Simulation and Age-weighted HS appear to be a problem even for ES. However, the effects of Age-weighted HS are illustrated differently, as the movements are sharp, examplified in Figure 7, in comparison to the flatter stair shaped movements for Historical Simulation. The sharpness and flatness for the models respectively are more obvious at the 99 % interval. The ghost effect should be eliminated by using the Age-weighted HS model, which evidently did not occur. We do find one possible explanation for this problem. As mentioned in the theory section 2.7, there is no given decay factor to apply when calculating the age-weighted simulation, and the suggestions are more likely applicable to common stock indices, such as S&P 500. In this study, a factor of 0.999 was chosen. Evaluating commodities, another decay factor could be reasonable, and a possible explanation for the non-satisfying modification of Historical Simulation. Additionally, using a recursive window, the minimum number of observations in the out-of-sample will be ten years. These years have large differences in characteristics leading to a misleading sorting of the largest negative returns. That is, these might occur at different times within the sample period, giving them different weights depending on their age. As exemplified in Figure 7, point 1, a new large negative return is given a high weight. But as the return gets older it will lose weight and the risk measure declines to point 2. At that point, a

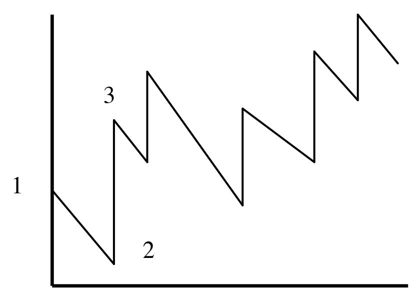


***Figure 7**. Age-weighted HS examplified*

new observation with a large negative return appears and will once again increase the risk measure reaching point 3. Therefore, the returns are given different weights and the sharp movements are created.

According to Figure 6, quarter three 2008 to quarter one in 2009 has high volatility in relation to the other quarters. Comparing this figure with the plots of ES in Appendix 3, it is obvious that it takes time for the non-time varying volatility models to adapt to a higher volatility; they lag the movements in the actual returns. The GARCH(1,1) models and the Volatility-weighted HS model immediately capture the volatility and show a similarity to the curve of the returns as is illustrated in Figure 6. According to these facts, the three time varying volatility parametrical models, as well as the Volatility-weighted model, should give better results. The arguments are relevant even for VaR, and due to the high correlations, this should not be surprisingly.

Additionally, comparing the plots of ES with the highest loss that occurred during 2008-2009, -6.037 %, Volatility-weighted HS model and all GARCH(1,1) models clearly estimate larger, negative values than this at the 99 % confidence interval. Since the results from the four ES tests are satisfying and the two models have ES Ratios close to zero, they seem to fit the returns of gold. This discussion also regards VaR. The average GARCH(1,1) ES values are higher than the corresponding non-time varying figures. That is a result of the time varying volatility models being less static and immediately update to match the movements of the returns better, following the discussion above. At both confidence intervals, the Volatility-weighted HS and Student-t GARCH(1,1) show best results regarding MAE and RMSE. However, the Age-weighted HS model has an ES Ratio closest to zero, but it should be noted that all ES Ratios at both confidence levels are close to zero. GMZ indicates that all models, except the models Normal and Log-Normal, perform well at both confidence levels. The recommendations from the four ES tests are basically built on the same principles, see the equations 24-28 and 30 in section 2.10 and section 2.10.4, which makes us question their relevance.

Similar to VaR, the Log-Normal model and the Log-Normal GARCH(1,1) model have lower Average ES values at both confidence intervals than the models Normal and Normal GARCH(1,1). As illustrated in Figure 2, the asymmetrical shape of the log-normal distribution leads to more observations in the right part of the curve, resulting in a lower average of the observations beyond VaR than the case for the normal distribution. That is, the explanation for lower Average ES is found in the characteristic of the distribution.

5.2 Oil

The plotted daily returns of oil, illustrated in Figure 8 below, exemplify large movements during the years 2008 and 2009. The figure shows increasing volatility during quarter three 2008 to quarter two 2009. Similar to the plot of gold, the movements are high in the middle of the time interval. This plot shows a lowest return of -13.065 % suggesting that the VaR and ES should not go below this value, see Appendix 4. Referring to Figure 4 in chapter 4, the in-sample period contains some exceptional large negative returns implying that the in-sample period has potential to forecast values for the out-of-sample period. However, the volatility clustering that appears in Figure 8 has not occurred previous years.

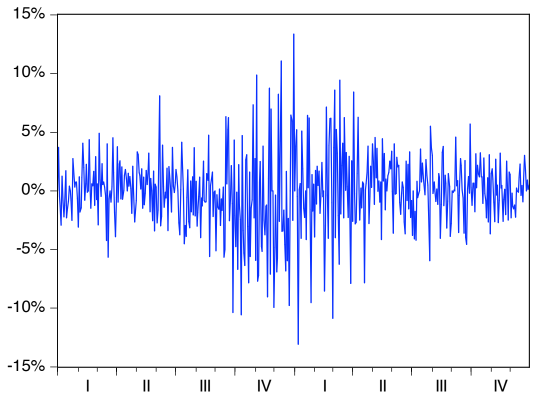


Figure 8*. Oil daily returns during 2008 and 2009.*

5.2.1 Oil – Value at Risk

Referring to the discussion regarding the lowest return above, the VaR exceeds the actual loss several times, as can be seen for Volatility-weighted HS 99 % in Appendix 4. However, the shape of the curve fits the actual returns shown in Figure 8, implying that, except for the overestimation, the model should give good results. The 95 % interval does not exceed this lowest return. The Volatility-weighted HS model presents the highest Average VaR values at both confidence intervals, and is also the best model regarding the VaR Rate and the three tests. Despite that the risk in some cases is overestimated at the 99 % confidence interval, the model explains the market risk of oil particularly well during this two year period. Similar overestimations and corresponding performance are made by the other time varying volatility models. Referring back to *2.11 Previous research*, Angelidis and Degiannakis found that models based on a Student-t distribution overestimated the true VaR in their study. The same conclusion can be made by studying the VaR plots for oil at the 99 % confidence interval, see

Appendix 4. Yet, the results summarized in Table 5 suggest that Student-t is an accurate model in this context.

Following the recommendations from Kupiec, it is easy to understand the superiority of models based on time varying volatility rather than models relying on historical standard deviation. As an example, at the 99 % confidence interval, Student-t GARCH(1,1) has a VaR Rate of 1.72 % whereas Student-t has a rate of 4.79 %. Basing the model on time varying volatility improves the risk measure. This can also be seen comparing their respective plots in Appendix 4. The Student-t model does not follow the pattern of the actual returns during 2008 to 2009, while the Student-t GARCH(1,1) shows this similarity. These two models also pass the Christoffersen independence test.

The Jarque-Bera test indicates no normality, with a value significantly different from zero. The Average VaR values for the Normal model underestimates the market risk at both confidence intervals and these failures correspond with Jarque-Bera. As regards section *2.5.1.1 Holding period*, it is stated that shorter holding periods entitle normal approximation and as VaR is measured on a daily basis, the motivation of using the models based on a normal distribution rises. However, modifying the Normal model with time varying volatility, Normal GARCH(1,1), the results improve. Concerning both confidence intervals, the VaR Rates are low enough to show p-values higher than 0.01 in the Kupiec test.

The discussion regarding the relationship between Student-t and Normal in section 5.1.1, evidently appears for oil as well. Student-t is larger than Normal at the 99 % confidence interval, whereas the opposite relationship is stated at the 95 % interval. In other words, the effect of the fat tails in the Student-t distribution is shown at the 99 % confidence interval. Also, this is the interval required by Basel II when measuring market risk, section 2.2.1, meaning that the Student-t distribution will give higher estimates than the normal distribution.

Not surprisingly, the Historical Simulation shows ghost effects, which is a common criticism against the simple model. This is an effect that should have been eliminated with the Age-weighted HS model, but this is not the case. The decay factor used in this study, 0.999, is too close to 1, which is the decay factor used in Historical Simulation. Perhaps, a better result would have been attained by choosing a lower factor. Though, the scale for the daily VaR values, viewed in Appendix 4, is larger for the Age-weighted HS model than for the Historical Simulation. Since the Age-weighted HS model considers recent observations more valuable, we expected the model to be more updated to the current market situation, compared to the Historical Simulation, which values each observation equally. As discussed, due to the choice of decay factor, this is not the case for oil as the Age-weighted HS model not adapts to the more volatile period, during quarter three 2008 to quarter two 2009.

To summarize, the most interesting models for VaR concerning both oil and gold are the Volatility-weighted HS and Student-t GARCH(1,1). Since they share similarity regarding their shapes, both within the group and with the daily returns of oil and gold, as well as their common feature, time varying volatility, these results are convincible. The results are satisfying and convenient in our aim of finding one or a few reliable models to measure the market risk of the three commodities.

5.2.2 Oil – Expected Shortfall

VaR and ES, when analyzing oil at both confidence intervals, are perfectly correlated for the models Volatility-weighted, Normal, Normal GARCH(1,1), Log-Normal and Log-Normal GARCH(1,1), see Appendix 8. Additionally, the Student-t GARCH(1,1) shows perfect correlation at the 95 % confidence interval. The kurtosis for the Student-t model varies between 5.486 and 6.417, with an average of 5.968. Therefore, the former discussion in section 5.1.2 about the kurtosis and the fact that the ES estimations of the Student-t model exceed those estimations of the Normal model is valid even for oil.

All models, except the models Normal GARCH(1,1), Log-Normal and Log-Normal GARCH(1,1), have negative ES Ratios at the 95 % confidence interval, implying that the forecasted ES overestimates the risk. Still, these models show results close to zero, indicating that the average of the tail is a good measure of the market risk. The best models, mentioned in the VaR section for oil above, Volatility-weighted HS and Student-t GARCH(1,1) have the highest ES values at both confidence intervals. Following the discussion of the VaR analysis for oil, the largest negative return is exceeded by those two models' ES estimations. The highest values of ES are approximately 16 % for both Volatility-weighted HS and Student-t GARCH(1,1) at the 95 % confidence interval. Naturally, those corresponding values are higher at the 99 % confidence interval. Considering what is preferred in this context, models that suggest too high values, due to high risk aversion as an example, to avoid returns violating ES frequently, should not be preferred. It should be noted that no correct model exists, the future is not predictable. Yet, these two models generally show the highest Average ES values, but simultaneous have ES Ratios close to zero as an example, meaning that they still can be referred to as accurate models.

The GMZ test does not show any outcome for five of the nine models at the 99 % confidence interval. Since the study contains 522 observations during the years 2008 and

2009, only five violations at the 99 % confidence interval should be found, approximately. With so few violations, it is not possible to make a regression and therefore, it is not surprising that the GMZ can not make conclusions. Once again, when observing the MAE and RMSE tests, the models Volatility-weighted HS and Student-t GARCH(1,1), show the best results. As mentioned in section 2.4, a drawback with ES is that the backtesting of the models require more data than VaR. This is a general problem for all commodities in this study, especially at the 99 % confidence interval.

Again, each model has identical pattern for VaR and ES, with the exception of the size of the values regarding the two risk measures. Referring to Appendix 4, the y-axis are the only items that differ. Applying GARCH(1,1) models improve the standard models radically by matching the risk values to the actual returns during the out-of-sample period. The same pattern is shown with the Volatility-weighted model. All standard parametrical models, together with the models Historical Simulation and Age-weighted HS, show the same increasing shape of the curves. Though, the parametrical models show more smooth curves than Historical Simulation and Age-weighted HS, which curves are more stair shaped and sharp, respectively.

5.3 Corn

Viewed in Figure 9 below, the daily returns for corn are less volatile than oil, but more volatile than gold. Though, the clustering pattern is not identical to the other two commodities, as the returns became more volatile quarter three 2008, followed by a less volatile period in the beginning of 2009, which thereafter, once again, became volatile. As for the other commodities, this phenomenon indicates improved performance using the time varying volatility models in relation to the remaining models. Plots of VaR and ES for corn are shown in Appendix 5.

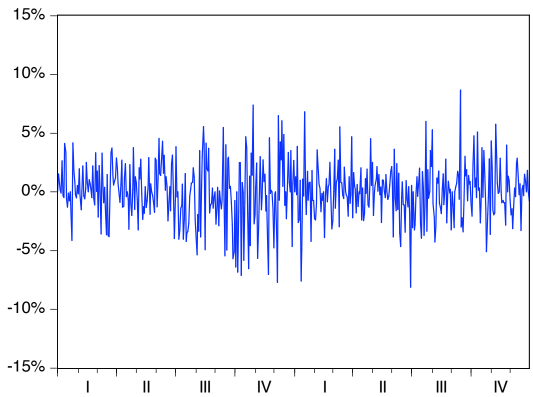


Figure 9. *Corn daily returns during 2008 and 2009.*

5.3.1 Corn – Value at Risk

The volatility clustering mentioned above is clearly captured in the Volatility-weighted HS model, as well as for the time varying models Normal GARCH(1,1) and Student-t GARCH(1,1), see Appendix 5. Furthermore, the reduced movements in quarter one and two 2009 are illustrated in the plots of VaR, indicating that the models immediately update to current market fluctuations. This pattern can also be seen for the Log-Normal GARCH(1,1) model.

The lowest return during 2008 and 2009 was -8.124 % and observing Student-t GARCH(1,1) 99 %, as an example, it forecasts losses larger than this number, meaning that this model might overestimate the risk. In Table 7, the Average VaR of Student-t GARCH(1,1) is the third highest, although the model shows p-values above 0.01 in all tests. As mentioned earlier, a model that constantly overestimates the market risk is not favored, but this does not seem to be a problem for Student-t GARCH(1,1), since the margin of error is relatively low. Analyzing the 95 % confidence interval, this model does not seem to perform as well as for the higher interval, since it only shows p-values above 0.01 for the Christoffersen test. The volatility-weighted HS model is the only model that can not be rejected in the Kupiec test. However, this model is rejected in the Christoffersen test, and therefore, all models fail the Joint test. Corn is the only commodity at the 95 % confidence interval in which no model can be recommended. Although the majority of the models have Christoffersen p-values above 0.01, they are low compared to the other two commodities. As this outcome differs from the other commodities at both confidence intervals, it is difficult to make a conclusion of the reason behind this. A reasonable explanation might be the second volatile period that corn presents during 2009. That is, Christoffersen measures independence in the risk model and corn shows more evidence for volatility clustering than oil and gold, meaning that the observations are more dependent for corn. Since this is an obvious difference between the commodities in the return pattern, this is the only explanation we can propose. Again, we have no foundation to base this hypothesis on. The reason for choosing these three commodities was because of their different properties and trading patterns, and the dispersion in the results this could lead to. The effect of this choice is illustrated in the situation above.

The VaR results are generally better on the 99 % confidence interval. The models Volatility-weighted HS and Student-t GARCH(1,1) can not be rejected in either of the tests. Again, as this interval focuses on the largest negative returns in the 1 % tail, the improvement at this interval might depend on a representation in the in-sample period of returns in the size of the out-of-sample period's returns, which therefore will estimate risk values in the size of the losses during 2008 and 2009.

As stated before, both Historical Simulation and Age-weighted HS illustrate ghost effects and the reason why the latter shows this effect is explained by the decay factor of 0.999. Even though the factor is close to 1, which is the decay factor for Historical Simulation, the plots of the two models are clearly different. This difference indicates that Age-weighted HS has potential to be improved if the decay factor had been adjusted to the returns of corn.

Again, when summarizing, Student-t GARCH(1,1) and Volatility-weighted HS are the most preferred models regarding VaR. The shapes of their plots show similarity compared to the returns of corn and to each other. As can be seen in section *2.8.1 Normal distribution*, the financial markets are seldom normally distributed, they more often appear to have excess kurtosis. This is one motivation for the frequent good results for the Student-t GARCH(1,1) model.

5.3.2 Corn – Expected Shortfall

Similar to the estimation for VaR, Student-t GARCH(1,1) 99 % overestimates ES, suggesting values of 10 % and above several times and should be compared to the lowest return of -8.124 %. Despite this, the model delivers ES Ratios close to zero, and is firstly ranked in both MAE and RMSE at the 99 % confidence interval. Together with Volatility-weighted HS, Normal GARCH(1,1) and Log-Normal GARCH(1,1), the four models have the best ranking. The same pattern can be seen in the ranking at the 95 % confidence interval.

Regarding the shape of the non-time varying volatility models versus the time varying volatility models, the differences are easily exemplified in Appendix 5. This discussion was brought up in section 5.1.2. As discussed about VaR for corn, the observations in the in-sample period showed relatively high movements, implying that a model with historical standard deviation, as for the Normal model, was expected to perform well, or at least better than for the other two commodities. Also for corn, it is obvious that the time varying volatility models, the GARCH(1,1) models and the Volatility-weighted HS model, outperform the remaining models, since it captures the fluctuations over time. There are different opinions of which time varying volatility model that should be preferred in this context, brought up in section 2.9. According to Ané, GARCH(1,1) is sufficient but Billio and Pelizzon have another belief. We chose to follow the former author, and compared to using unconditional variance, the results are radically improved.

The kurtosis explains the difference between the Student-t distribution and normal distribution, since the tails are fatter for the former distribution. For corn, the kurtosis varies between 4.403 and 5.228. The Normal model estimates suggest an Average VaR higher than the Student-t model at the 95 % confidence interval, and at the 99 % interval, the opposite relation is found. For ES, the Student-t model has higher Average ES at both confidence intervals. This is explained by the higher amount of large negative observations in the end of the left tail for a Student-t distribution. Therefore, the average value of these observations will be higher, leading to higher ES than the case for normal distribution. It should be noted that the estimation error is larger for ES than for VaR when fat-tailed distributions are used, which is declared in section *2.4 Expected Shortfall*.

As for the previous commodities, the correlations between VaR and ES are almost perfect, indicating that ES will not add more than higher values to the risk calculation. As stated before, if a portfolio perspective would have been chosen in this study, the difference between the two risk measures would probably be larger and present values of other shapes.

6 Conclusion

The correlations between VaR and ES are almost positively perfect for gold, oil and corn, for all models and at all confidence intervals. The implication of this is that the movements of the two risk measures are almost identical, with the size of the values as the only difference. Comparing VaR at the 99 % confidence interval to ES at the 95 % interval, the former gives higher results for each commodity. This phenomenon regards both the Average VaR and ES values, respectively, and the plotted values viewed in Appendix 3 to Appendix 5. This indicates that ES is not a necessary complement to VaR, as ES, at a particular confidence interval, can be replaced by VaR at a higher interval. However, this relation is only valid when studying each commodity separately. Taking a portfolio perspective, different results would most certainly be attained. Still, it is important to be aware of the difference between the two confidence levels regarding the observations they refer to; 5 % and 1 % respectively. Since the number of observations referred to at the latter level is one fifth of the observations included at the former level, too few observations is a disadvantage when evaluating statistical results. To conclude the discussion above, ES adds value in this context, even when not taking a portfolio perspective. That is, especially when the sample is of a non-satisfying size or when the aim is to take outliners into consideration, due to risk aversion as an example.

The two best models for all commodities, at both confidence intervals, are Volatility-weighted HS and Student-t GARCH(1,1), and if ranking them according to their individual performance in the tests, Volatility-weighted HS is the best performer. This regards both VaR and ES. These models have had the best results throughout the study and simplify the conclusion of choosing one, or two, models that correspond to the three commodities; gold, oil and corn. It is pleasant that the two models turned out well for all commodities, although their characteristics differ. This report has studied a volatile period, during the recent financial crisis in which the prices of gold, oil and corn were affected, and it is therefore not surprising that time varying volatility models are the best performers. Our personal criticism against the three distributions used in this study has not been fulfilled. When modifying these models with time varying volatility, the top ranked models of the study were found.

Basel II does not give recommendations on which model to use and since our findings differ among the nine models when analyzing commodities, we believe that Basel II should oversee the flexibility in its regulation. Pérignon's and Smith's article, presented in the introduction, concludes that the historical simulation is the most used model among banks.

Similar to the authors' doubt, our results imply that the use of this method should be questioned. One implication of Taleb's introducing quote *"Proponents of VaR will argue that it has its shortcomings but it's better than what you had before*" is that VaR should be treated with skepticism. Again, since the future can not be predicted, there are no correct models to estimate future market risk and complementing measures should be considered. Our suggestion is to complement VaR with ES for a more accurate estimation of the market risk to eliminate the limitations of one standard financial risk measure.

References

Printed Sources

Brooks, Chris. *Introductory Econometrics for Finance*. Cambridge University Press. Second edition (2008) Singapore

Culp, Christopher. *The Risk Management Process*. John Wiley & Sons (2001) USA

Dowd, Kevin. *Beyond Value at Risk. The New Science of Risk Management*. John Wiley & Sons (1998) Great Britain

Dowd, Kevin. *Measuring market risk*. John Wiley & Sons. Second edition (2005) Great Britain

Gujarati, Damodar N. *Essentials of Econometrics*. McGraw-Hill/Irwin (2006) USA; Singapore

Jorion, Phillippe. *Value at Risk*. McGraw-Hill. Second edition (2001) USA

Jorion, Philippe. *Vale at risk: the new benchmark for managing financial risk*. McGraw-Hill. Second edition (2001) Great Britain

Rogers, Jim, *Hot Commodities: Anyone Can Invest Profitably in the World's Best market*. Breeland's Interest, Inc. (2004) USA

Taleb, Nassim. *The black swan: the impact of the highly improbable*. Random House. First edition (2007) USA

Articles

Acerbi, Carlo and Tasche Dirk. *Expected Shortfall: a natural coherent alternative to Value at Risk*. (2001)
http://www.bis.org/bcbs/ca/acertasc.pdf (Bank for International Settlements)

Ané, Thierry. *An analysis of the flexibility of Asymmetric Power GARCH models*. Computational Statistics and Data Analysis. (2006) Vol. 51 No. 2 Page 1293-1311

Angelidis, Timotheos; Benos, Alexandros and Degiannakis, Stavros. *The Use of GARCH Models in VaR Estimation* (2004)
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=661261

Angelidis, Timotheos and Degiannakis, Stavros. *Backtesting VaR Models: An Expected Shortfall Approach* (2006)
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=898473

Angelidis, Timotheos and Degiannakis, Stavros. *Econometric Modeling of Value-at-Risk*. New Econometric Modeling Research (2007)
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=933123

Artzner, Philippe; Delbaen, Freddy; Eber, Jean-Marc and Heath, David. *Thinking Coherently*. Risk (1997) Vol. 10 Page 68-71

Artzner, Philippe; Delbaen, Freddy; Eber, Jean-Marc and Heath, David. *Coherent Measures of Risk*. Mathematical Finance (1999) Vol. 9 No. 3 Page 203-228

Billio, Monica and Pelizzon Loriana. Value-at-Risk: a multivariate switching regime approach. Journal of Empirical Finance (2000) Vol. 7 No. 5 Page 531-554

Bodoukh, Jacob, Richardson, Matthew and Whitelaw, Robert F. *The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk*. (1998)
http://www.faculty.idc.ac.il/kobi/thebestrisk.pdf

Büyükşahin, Bahattin, Haigh, Michael S. and Robe, Michael A. *Commodities and Equities: A "Market of One"?* (2008)
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1069862

Campbell, Sean D. *A Review of Backtesting and Backtesting Procedures*. (2005). Board of Governors of the Federal Reserve System.
http://www.federalreserve.gov/PUBS/feds/2005/200521/200521pap.pdf

Christoffersen, Peter F. *Evaluating Interval Forecasts*. (1998). International Economic Review. Vol. 39 No. 4

Delbaen, Freddy. *Coherent Risk Measures on General Probability Spaces*. (2002). Essays in Honour of Dieter Sondermann. In: K. Sandmann and P. J. Schnbucher, (eds.), Advances in Finance and Stochastics, Springer. Page 1-38.

Giot, Pierre and Laurent, Sébastien. *Market risk in commodity markets: a VaR approach*. Energy Economics. (2003) Vol. 25

Hansen, Peter R. and Lunde, Asger. *A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?* (2005) Journal of Applied Econometrics. Vol. 20 No. 7

McNeil, Alexander J. and Frey, Rüdiger. *Estimation of tail-related risk measures for heteroscedasic financial time series: an extreme value approach*. Journal of Emperical Finance. (2000) Vol. 7

Pérignon, Christophe and Smith Daniel R. *The level and quality of Value-at-Risk disclosure by commercial banks*. (2009)
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=952595

Yamai Yasuhiro and Yoshiba Toshianao. *On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall*. Monetary and Economic Studies/January 2002. Bank of Japan
http://www.imes.boj.or.jp/english/publication/mes/2002/me20-1-3.pdf

Yamai, Yasuhiro and Yoshiba Toshianao *On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall*. Monetary and Economic Studies/January 2002. Bank of Japan
http://www.imes.boj.or.jp/english/publication/mes/2002/me20-1-3.pdf

Yamai, Yasuhiro and Yoshiba Toshianao. *Value-at-risk versus expected shortfall: A practical perspective*. Journal of Banking and Finance (2005) Vol. 29 No. 4 Page 997-1015

Webb pages

Bank of International Settlements. *Proposed revisions to the Basel II market risk framework.* July 2008. http://www.bis.org/publ/bcbs140.htm

Bank of International Settlements. *Basel Committee on Banking Supervision, Consultative document – Revisions to the Basel II market risk framework*. January 2009. http://www.bis.org/publ/bcbs148.pdf

DnB NOR.
https://www.dnbnor.se/site/markets/ravaror/investering

Goldman Sachs.
http://www.goldman-sachs.ch/_files/download/GS_CH_Tabelle_commodity_s&p_03.pdf 2010-04-21

Kevin Sheppard.
http://www.kevinsheppard.com/images/2/20/VaR_and_Distributions_2006-2007.pdf

RiskMetrics Group.
https://frontoffice.riskmetrics.com/wiki/index.php/Backtesting_Expected_Shortfall

Standard & Poor's.
http://www.standardandpoors.com/indices/sp-gsci/en/us/?indexId=spgscirg--usd----sp------

Appendix

Appendix 1. Macros Excel

Volatility-weighted HS macro for VaR

```
Sub VWHS()
Sheet9.Select
Dim condvol(3131) As Variant
Dim forcondvol(3131) As Variant
Dim rstar(3131) As Variant
Dim i As Integer
Dim j As Integer
Dim percentilen As Double
For i = 1 To 3131
condvol(i) = Cells(i + 2, 6)
rstar(i) = Cells(i + 2, 7)
forcondvol(i) = Cells(i + 3, 6)
Next i
For i = 2609 To 3131
ReDim rr(i - 1) As Variant
For j = 1 To i - 1
rr(j) = rstar(j) * forcondvol(i - 1) / condvol(j)
Next j
percentilen = Application.WorksheetFunction.Percentile(rr, 0.01)
Cells(i + 2, 13) = percentilen
Next i
End Sub
```

Age-weighted HS macro for ES

```
Sub AWHS()
Blad5.Select
Application.ScreenUpdating = True
Dim Ra As Range
Dim Rb As Range
Dim R As Range
Dim WS As Worksheet ' temporary worksheet
Dim kumsum As Double
Dim lnum As Integer
Dim var(1 To 49) As Double
Dim varlevel As Double
Dim i As Integer
Dim j As Integer
Dim weighta As Variant
Dim l As Integer

For i = 3 To 3132
   ReDim data(i) As Variant
   data(i) = Cells(i + 2, 2)
Next i

 For i = 2611 To 3132
     weighta = ((1 - 0.999) / (1 - 0.999 ^ (i - 3)))
   ReDim data(1 To i - 1) As Variant
   ReDim Weight(1 To i - 1) As Variant
```

```
 For j = 1 To i - 3 ' changed from -1
   data(j) = Cells(j + 2, 2)
   Weight(j) = weighta * 0.999 ^ (i - j - 3)
Next j

   varlevel = 0.01
   Application.ScreenUpdating = False
   Set WS = ThisWorkbook.Worksheets.Add
   Set Ra = WS.Range("A1").Resize(UBound(data) - LBound(data) + 1, 1)
   Ra = Application.Transpose(data)
   Set Rb = WS.Range("b1").Resize(UBound(data) - LBound(data) + 1, 1)
   Rb = Application.Transpose(Weight)
   Set R = WS.Range("a1").Resize(UBound(data) - LBound(data) + 1, 2)
  '  sort the range
   R.Sort key1:=Ra, order1:=xlAscending, MatchCase:=False
   kumsum = WS.Cells(1, 2)
   lnum = 1
   Do While kumsum < varlevel ' 0.01 is the VaR level (or 1 - VaR to be precise)
      lnum = lnum + 1
      kumsum = kumsum + WS.Cells(lnum, 2)
   Loop
   ReDim rets(1 To lnum) As Double

   For k = 1 To lnum
      rets(k) = WS.Cells(k, 1) * WS.Cells(k, 2)
   Next k

   Application.DisplayAlerts = False
   WS.Delete
   Application.DisplayAlerts = True
  Application.ScreenUpdating = True
es = Application.WorksheetFunction.Sum(rets) / varlevel
Cells(i - 2608, 41) = es
Next i
End Sub
```

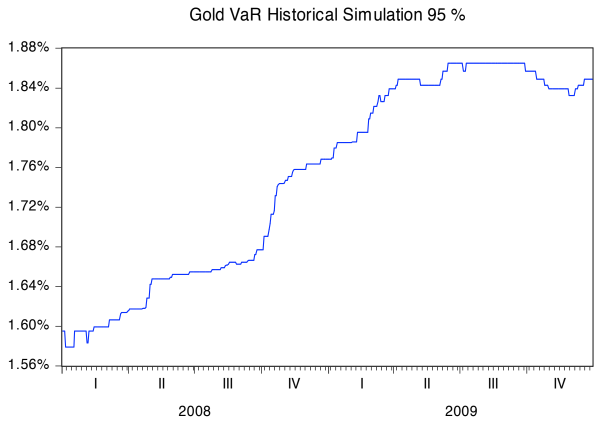
Appendix 2. Programming EViews

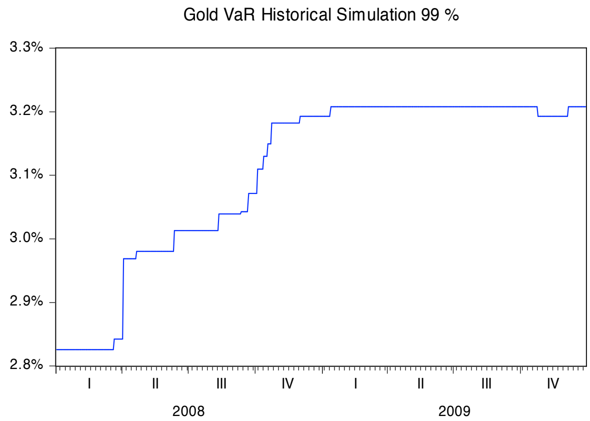
GARCH(1,1) script

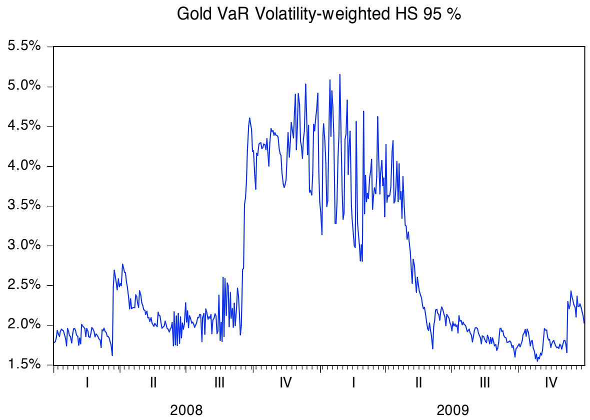
```
smpl @all
scalar noobs=3130
scalar nw=2609
matrix(3131,1) garchforecast
vector(1) forecastresult
for !i=nw to noobs
smpl !i-nw+1 !i
equation garch_n
garch_n.arch(1,1) gold c
garch_n.makegarch garchcondvar
forecastresult(1)=garch_n.@coefs(2)+garch_n.@coefs(3)*resid(!i)+garch_n.@coefs(4)*garchcondvar(!i)
rowplace(garchforecast,@transpose(forecastresult),!i+1)
next
```

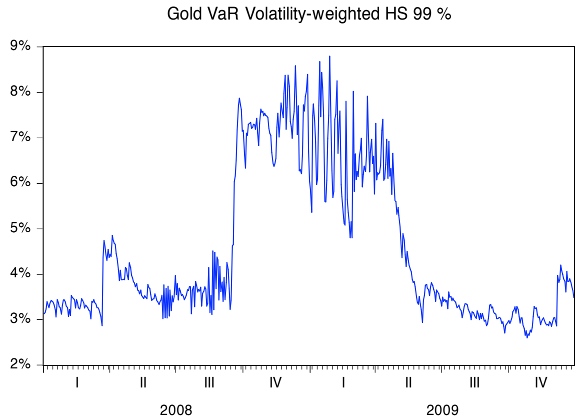
Appendix 3. Gold – VaR and ES

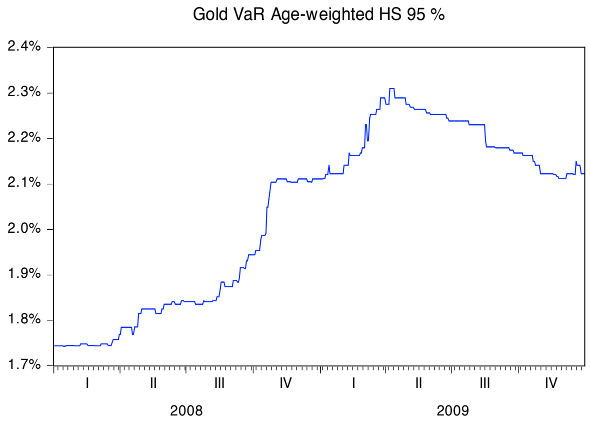
Note that the scales in the plots differ.

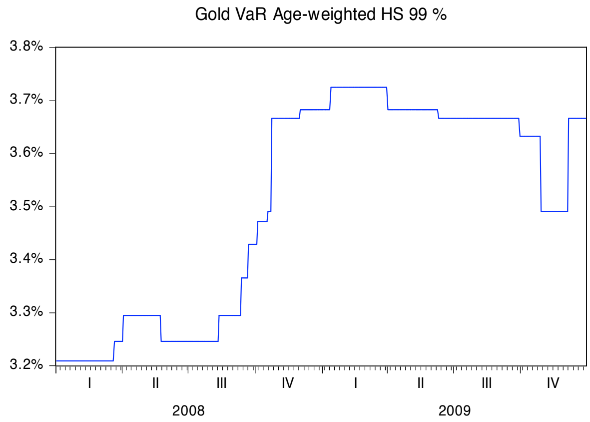




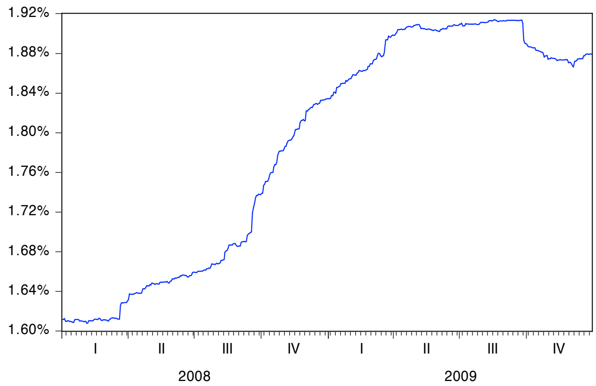




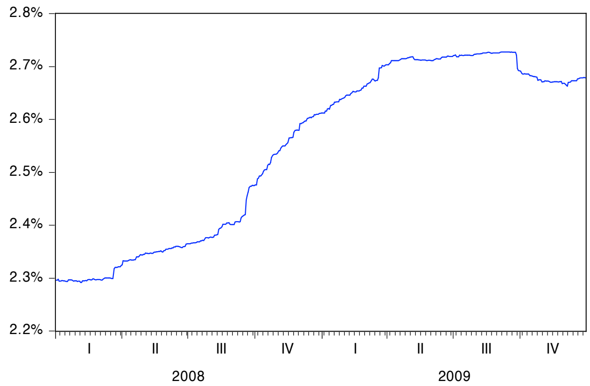




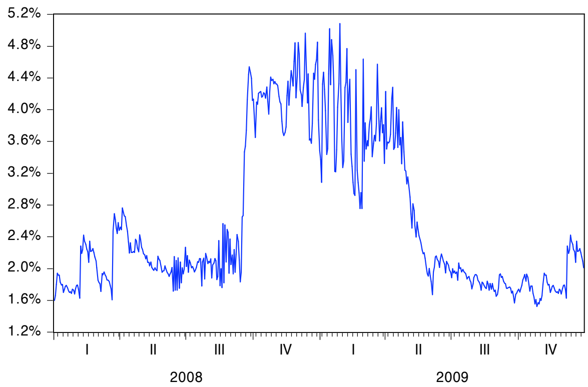
Gold VaR Normal 95 %



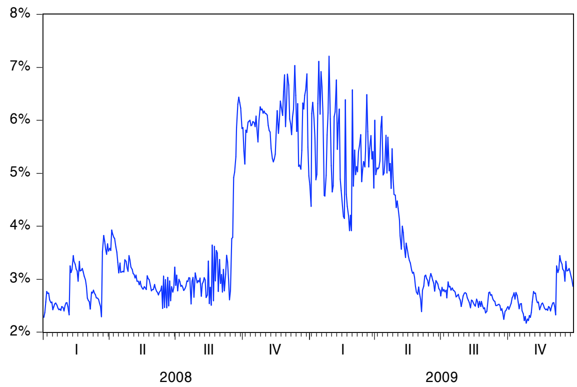
Gold VaR Normal 99 %



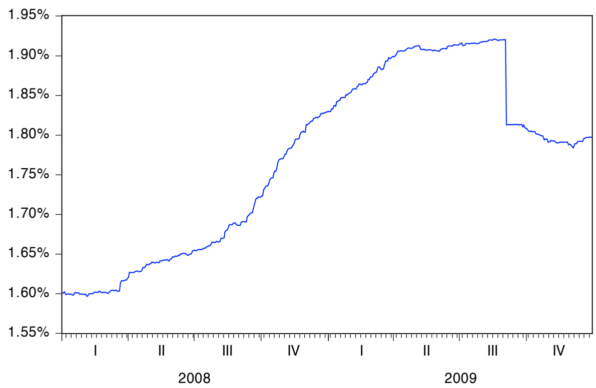
Gold VaR Normal GARCH(1,1) 95 %



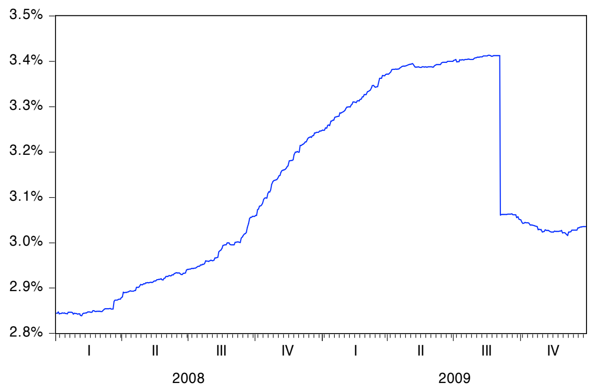
Gold VaR Normal GARCH(1,1) 99 %



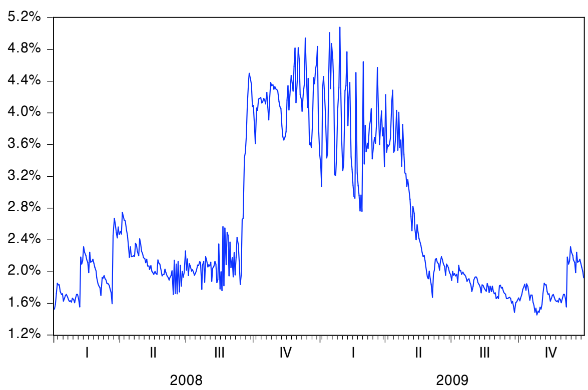
Gold VaR Student-t 95 %



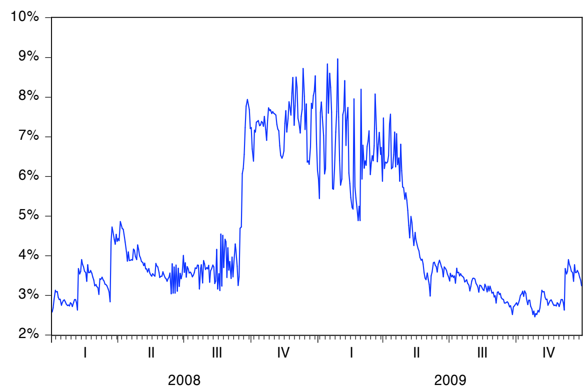
Gold VaR Student-t 99 %



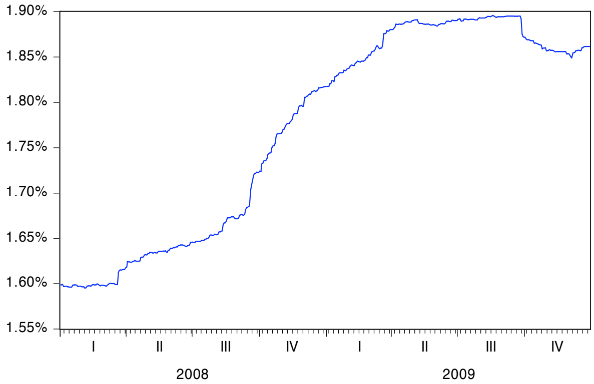
Gold VaR Student-t GARCH(1,1) 95 %



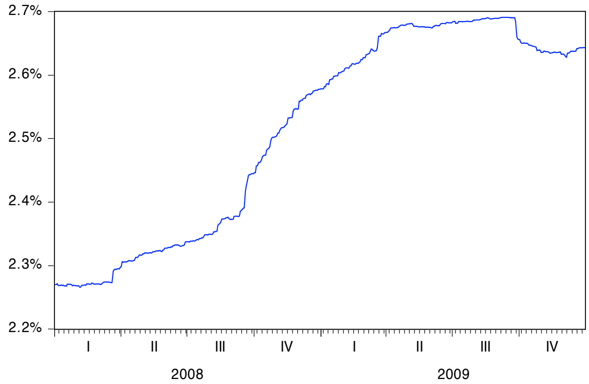
Gold VaR Student-t GARCH(1,1) 99 %



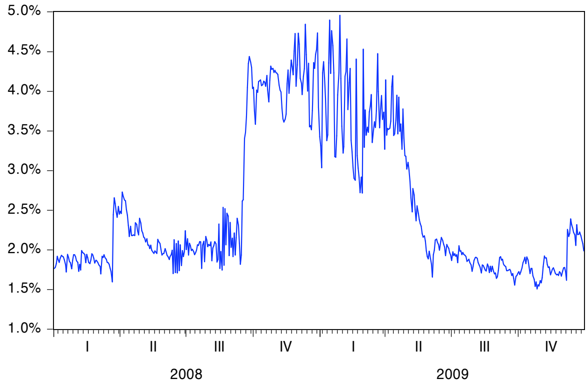
Gold VaR Log-Normal 95 %



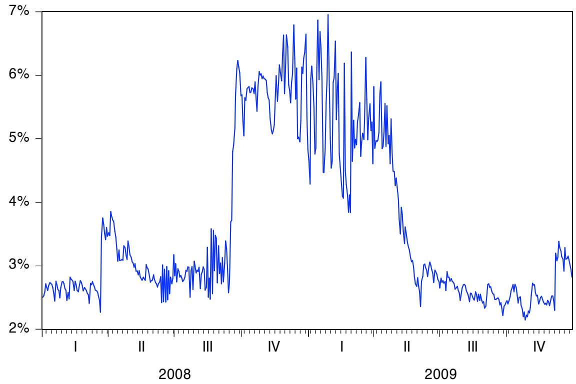
Gold VaR Log-Normal 99 %



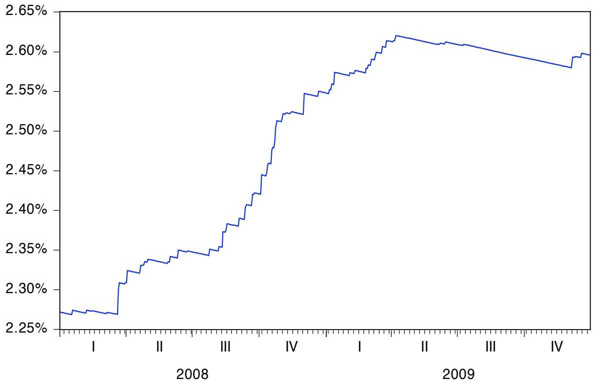
Gold VaR Log-Normal GARCH(1,1) 95 %



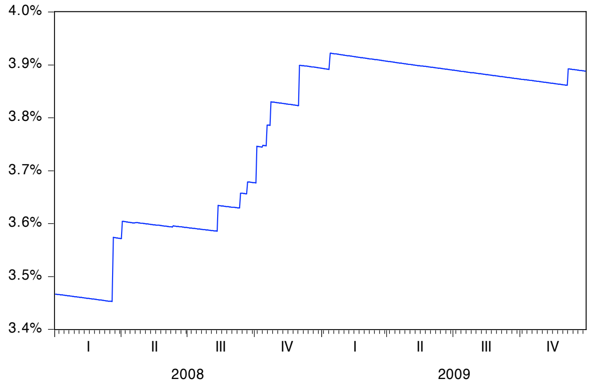
Gold VaR Log-Normal GARCH(1,1) 99 %



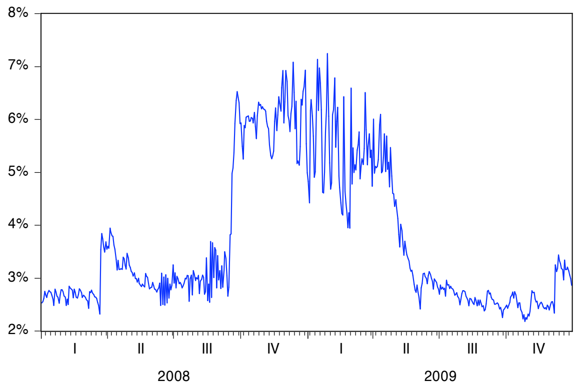
Gold ES Historical Simulation 95 %



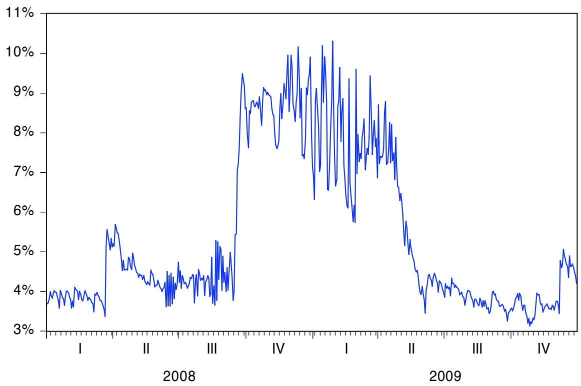
Gold ES Historical Simulation 99 %



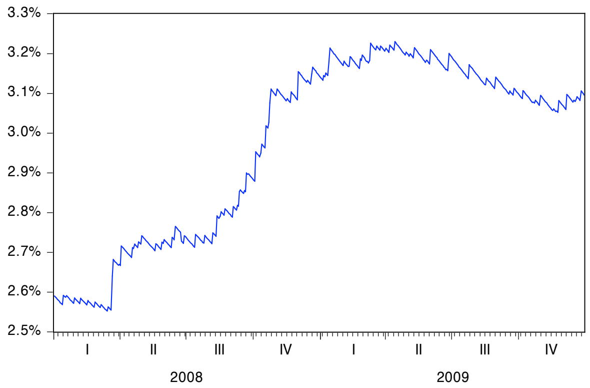
Gold ES Volatility-weighted HS 95 %



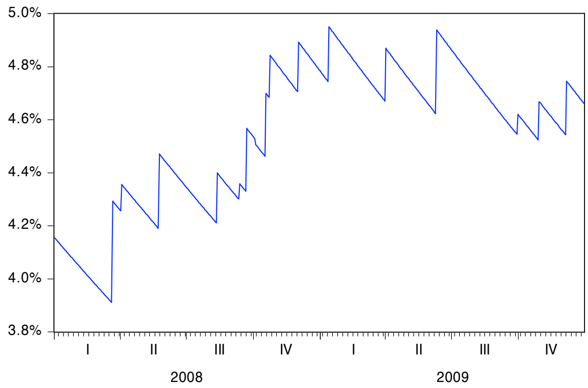
Gold ES Volatility-weighted HS 99 %



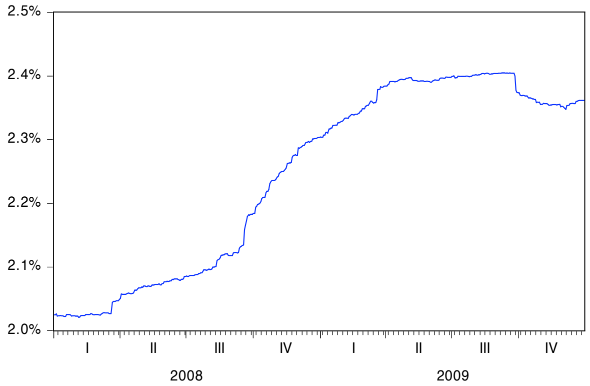
Gold ES Age-weighted HS 95 %



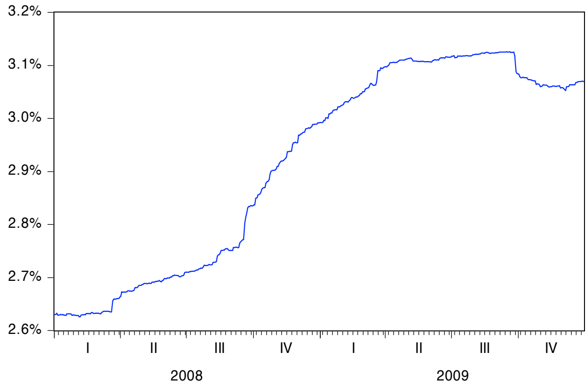
Gold ES Age-weighted HS 99 %



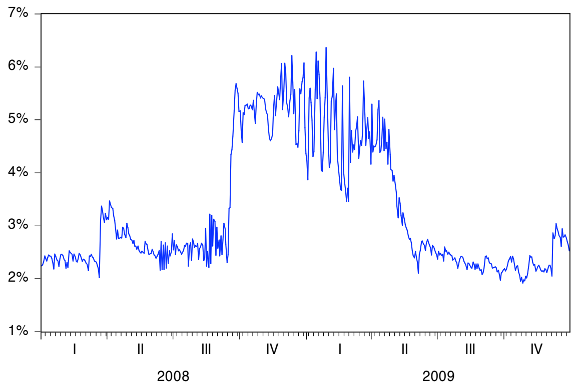
Gold ES Normal 95 %



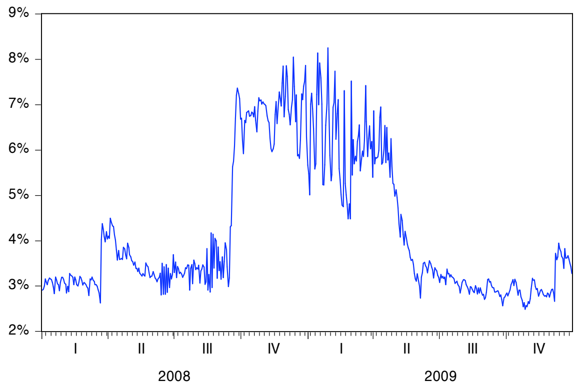
Gold ES Normal 99 %



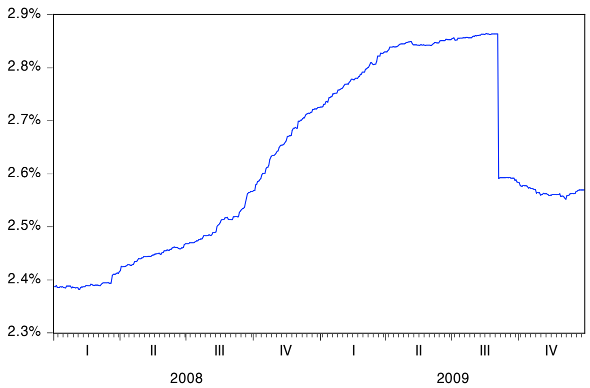
Gold ES Normal GARCH(1,1) 95 %



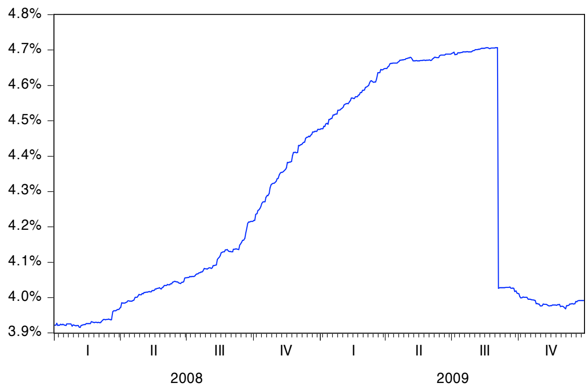
Gold ES Normal GARCH(1,1) 99 %



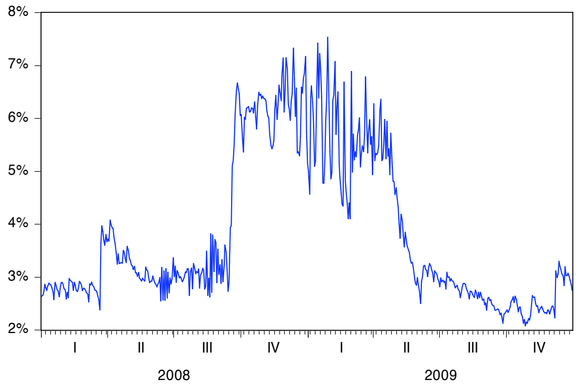
Gold ES Student-t 95 %



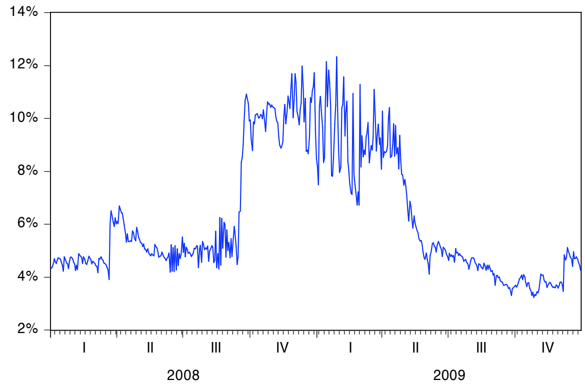
Gold ES Student-t 99 %



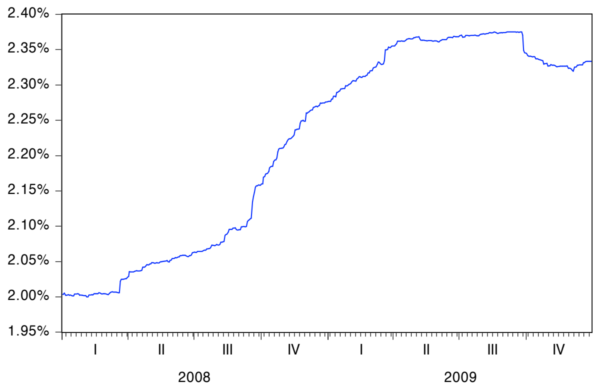
Gold ES Student-t GARCH(1,1) 95 %



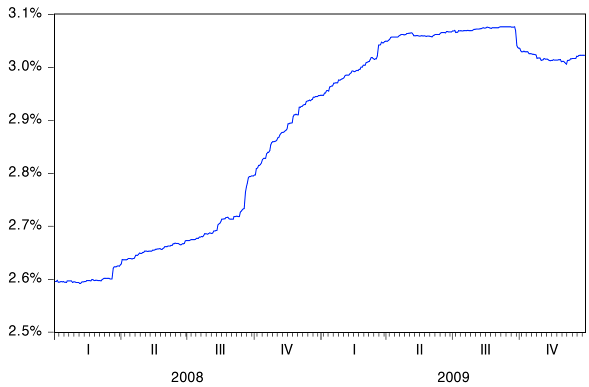
Gold ES Student-t GARCH(1,1) 99 %



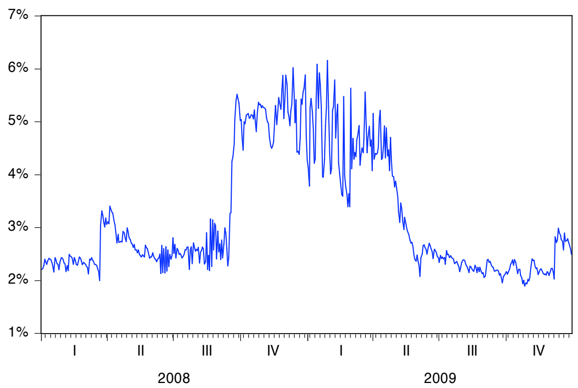
Gold ES Log-Normal 95 %



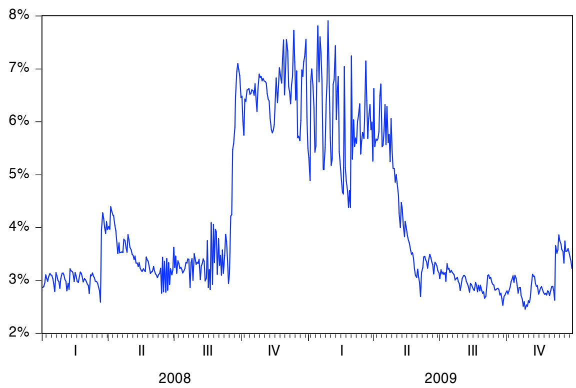
Gold ES Log-Normal 99 %



Gold ES Log-Normal GARCH(1,1) 95 %

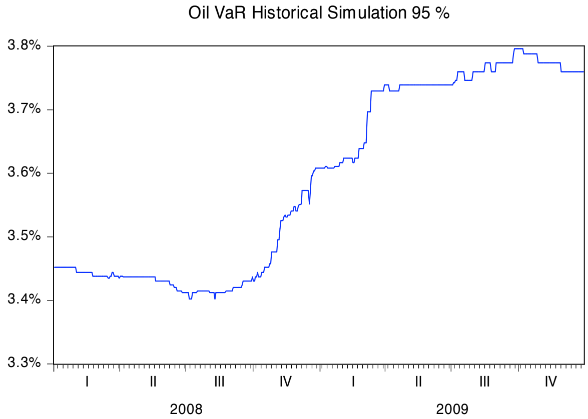


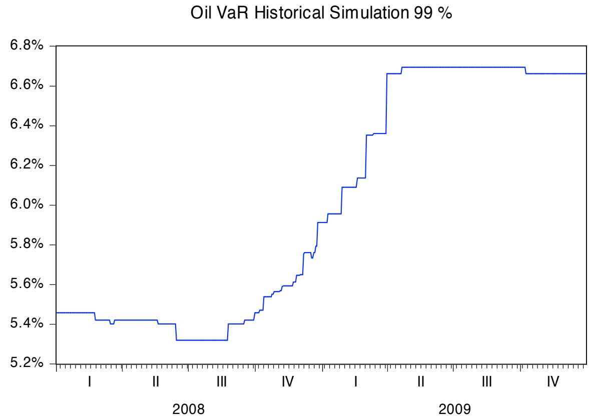
Gold ES Log-Normal GARCH(1,1) 99 %

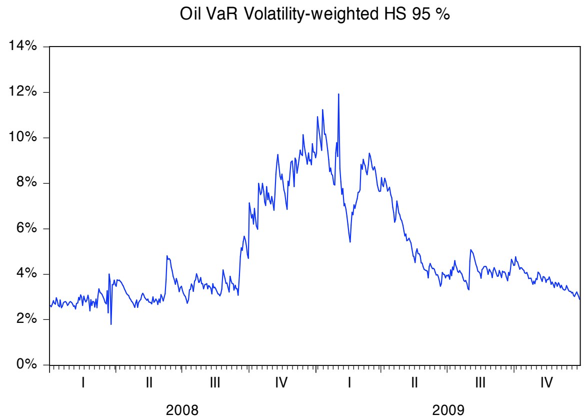


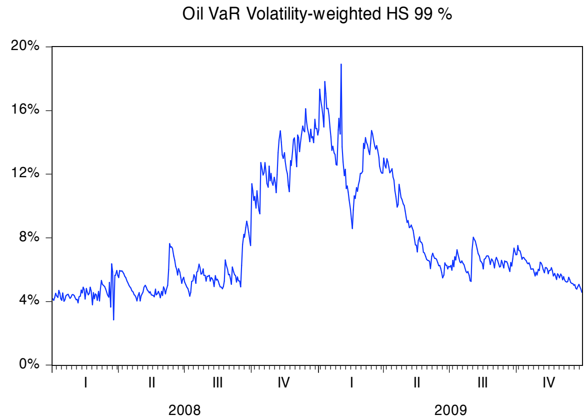
Appendix 4. Oil – VaR and ES

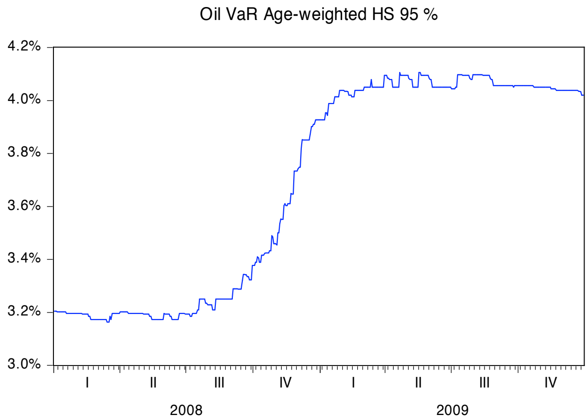
Note that the scales in the plots differ.

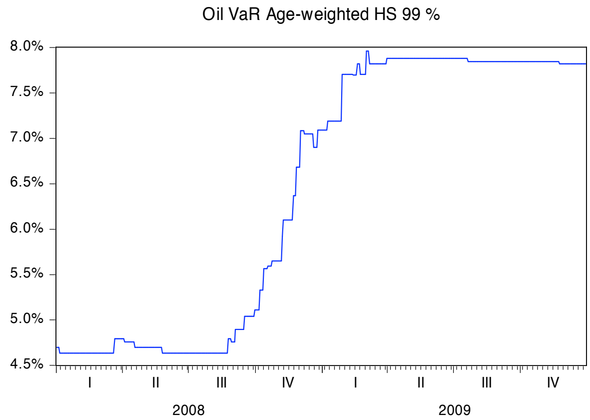




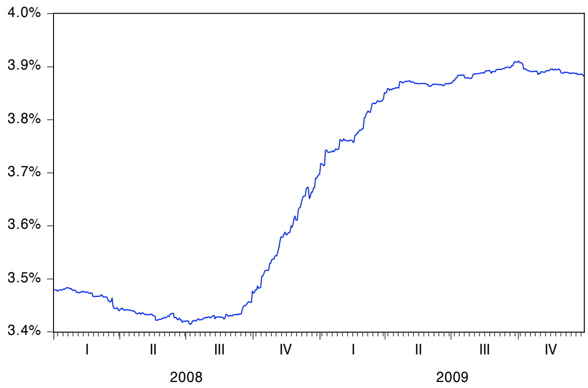




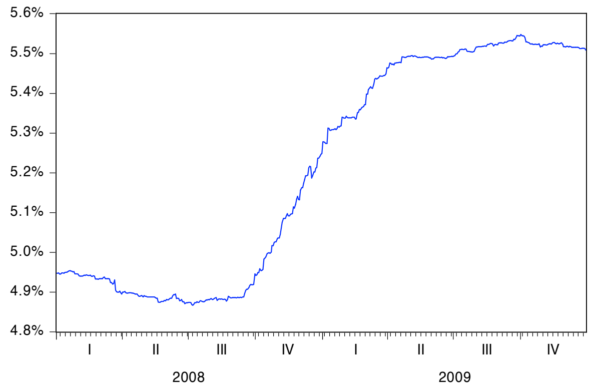




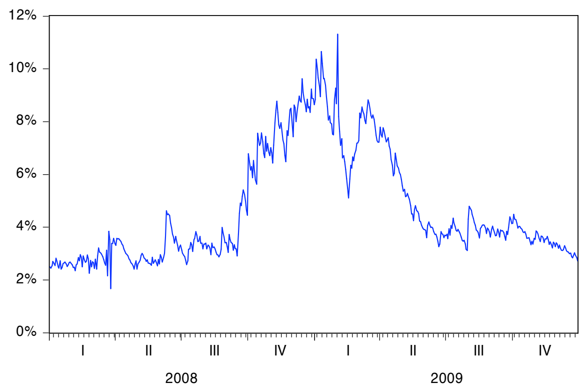
Oil VaR Normal 95 %



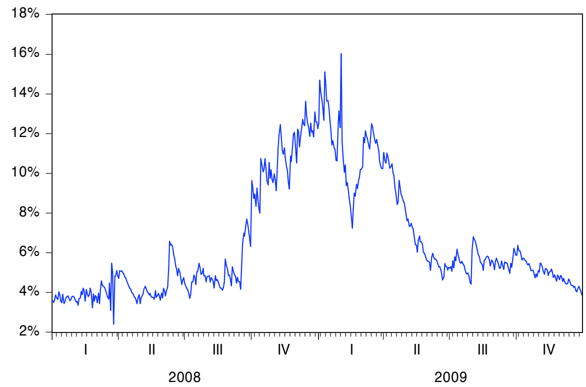
Oil VaR Normal 99 %



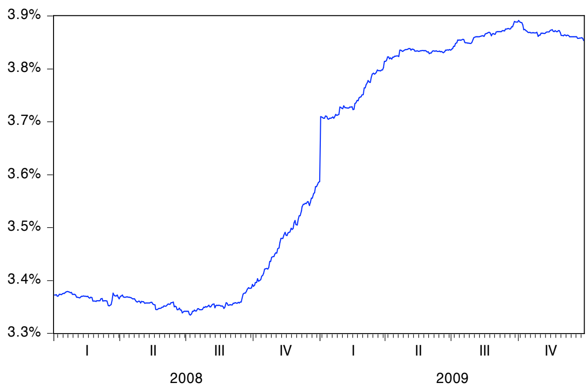
Oil VaR Normal GARCH(1,1) 95 %



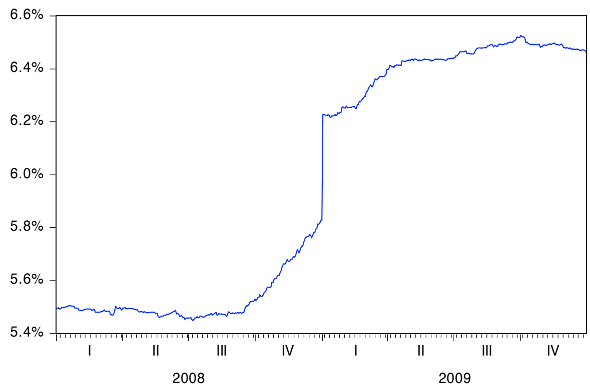
Oil VaR Normal GARCH(1,1) 99 %



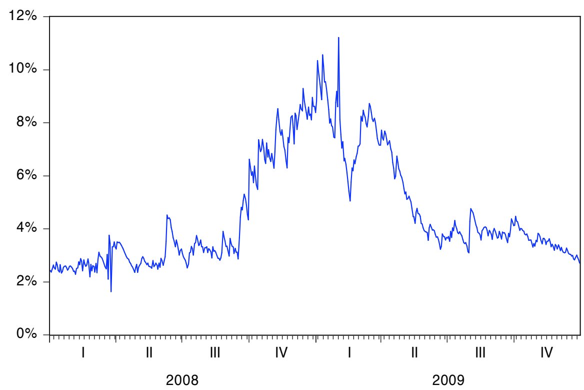
Oil VaR Student-t 95 %



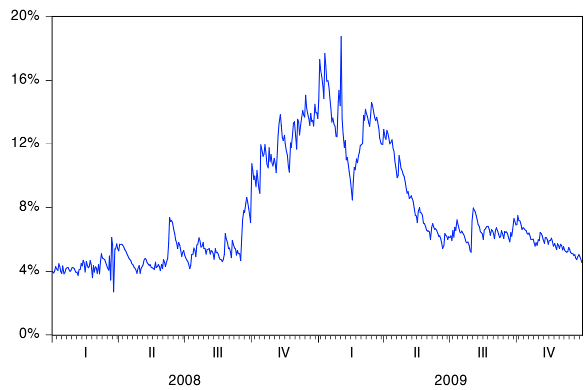
Oil VaR Student-t 99 %



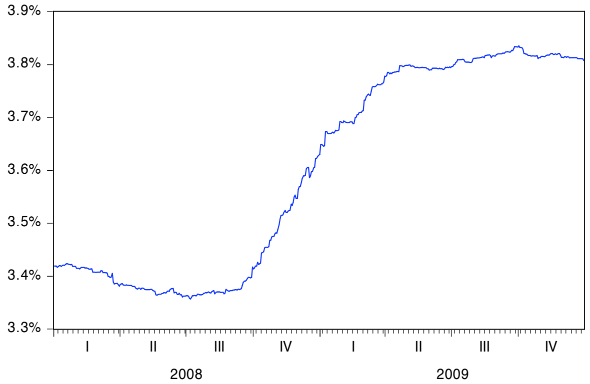
Oil VaR Student-t GARCH(1,1) 95 %



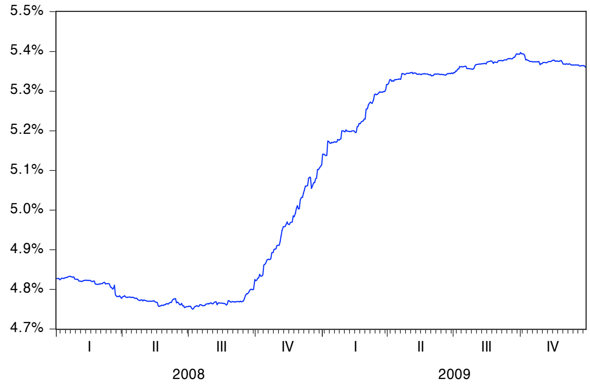
Oil VaR Student-t GARCH(1,1) 99 %



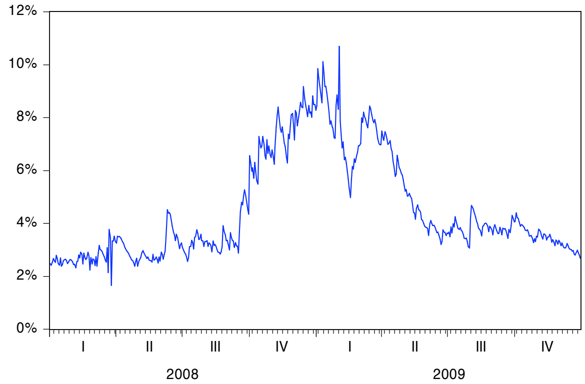
Oil VaR Log-Normal 95 %



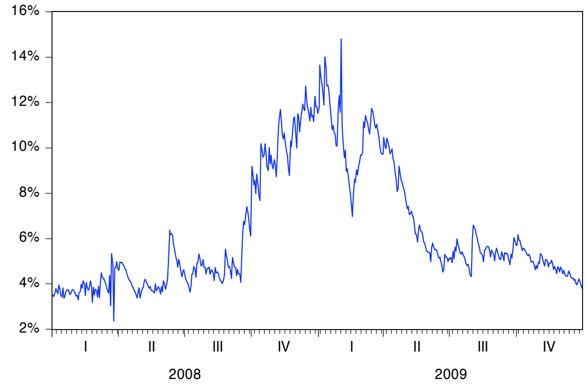
Oil VaR Log-Normal 99 %



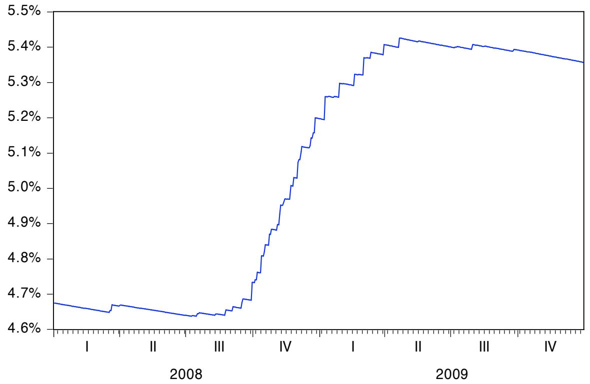
Oil VaR Log-Normal GARCH(1,1) 95 %



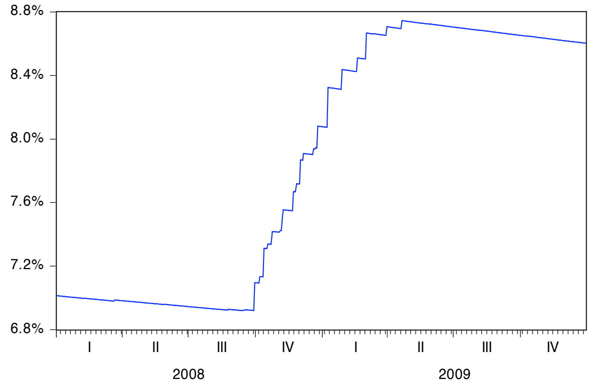
Oil VaR Log-Normal GARCH(1,1) 99 %



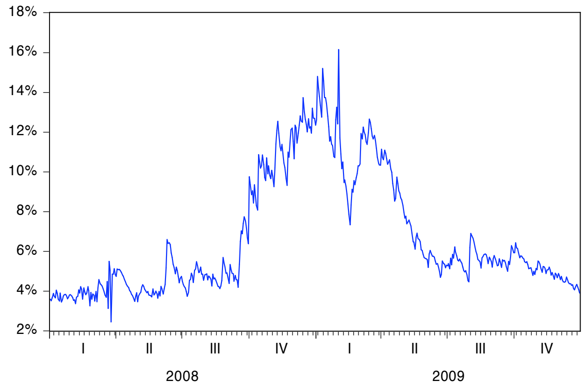
Oil ES Historical Simulation 95 %



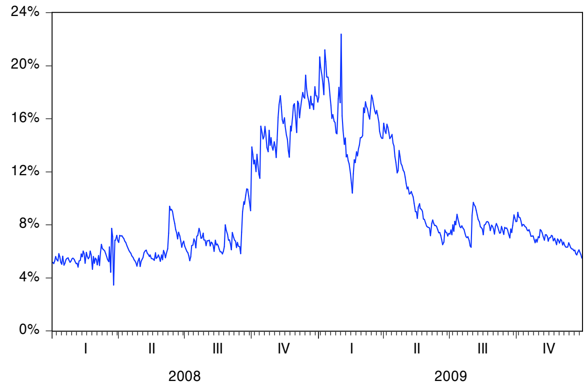
Oil ES Historical Simulation 99 %

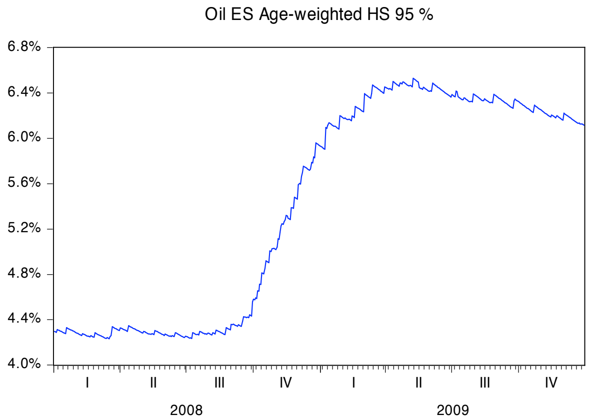


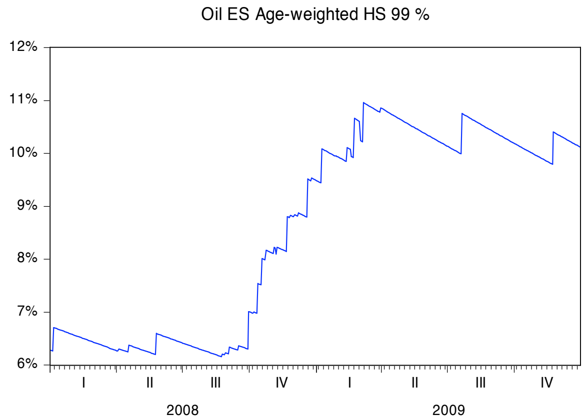
Oil ES Volatility-weighted HS 95 %

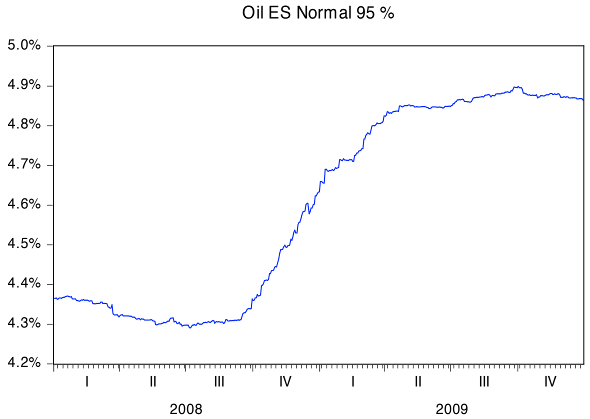


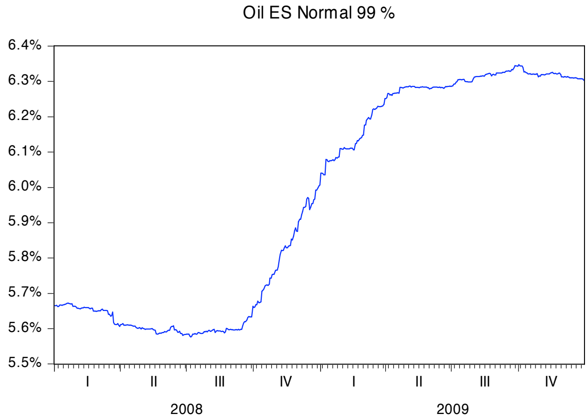
Oil ES Volatility-weighted HS 99 %

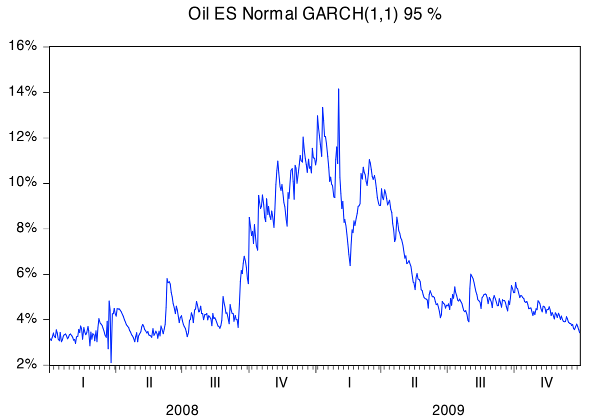


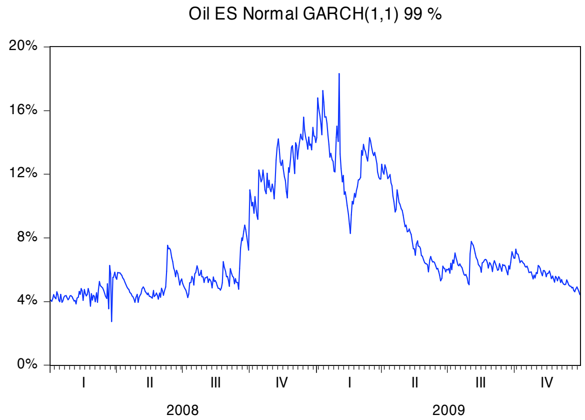

Oil ES Age-weighted HS 95 %
6.8%
6.4%
6.0%
5.6%
5.2%
4.8%
4.4%
4.0%
I II III IV I II III IV
2008
2009

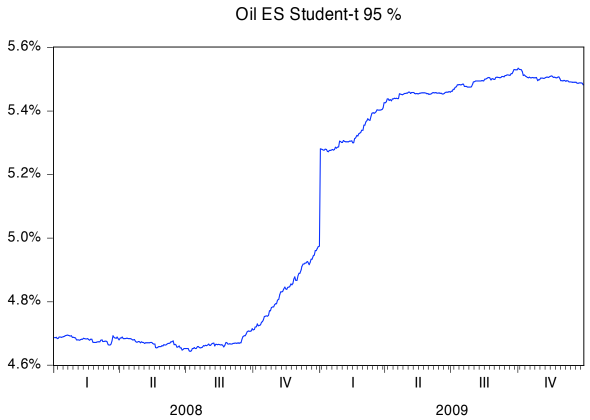

Oil ES Age-weighted HS 99 %
12%
11%
10%
9%
8%
7%
6%
I II III IV I II III IV
2008
2009

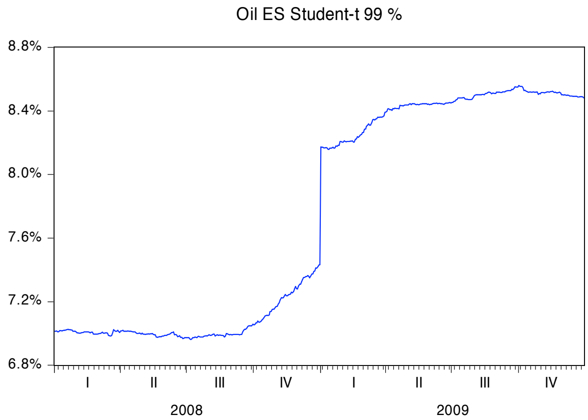

Oil ES Normal 95 %
5.0%
4.9%
4.8%
4.7%
4.6%
4.5%
4.4%
4.3%
4.2%
I II III IV I II III IV
2008
2009

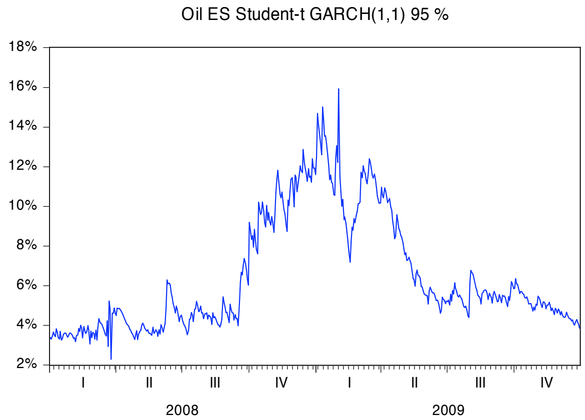

Oil ES Normal 99 %
6.4%
6.3%
6.2%
6.1%
6.0%
5.9%
5.8%
5.7%
5.6%
5.5%
I II III IV I II III IV
2008
2009

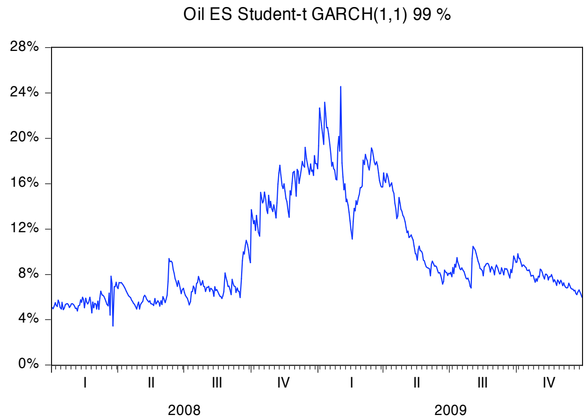

Oil ES Normal GARCH(1,1) 95 %
16%
14%
12%
10%
8%
6%
4%
2%
I II III IV I II III IV
2008
2009

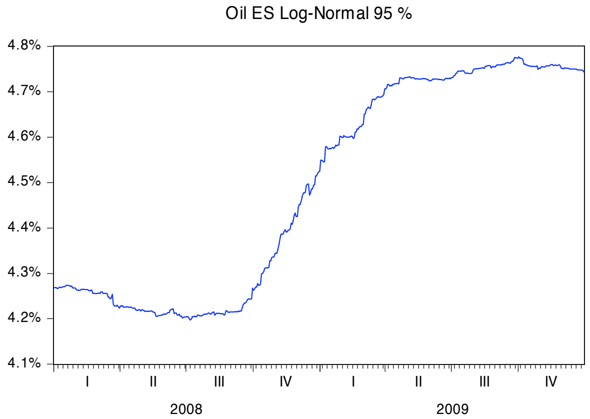

Oil ES Normal GARCH(1,1) 99 %
20%
16%
12%
8%
4%
0%
I II III IV I II III IV
2008
2009

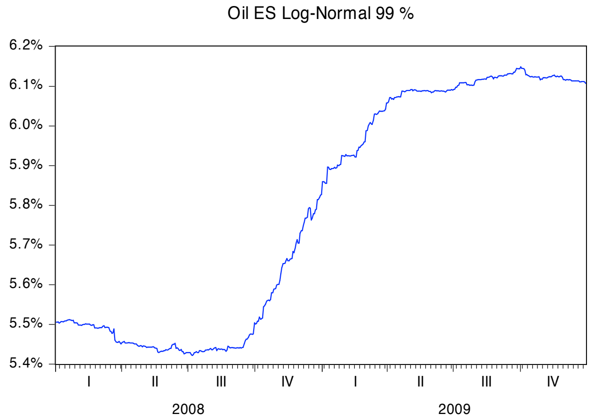

Oil ES Student-t 95 %
5.6%
5.4%
5.2%
5.0%
4.8%
4.6%
I II III IV I II III IV
2008
2009

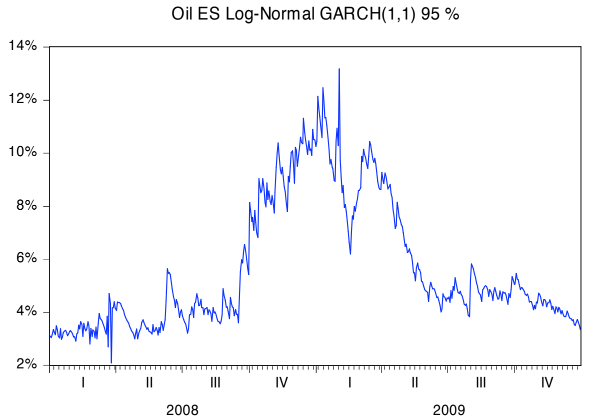

Oil ES Student-t 99 %
8.8%
8.4%
8.0%
7.6%
7.2%
6.8%
I II III IV I II III IV
2008
2009

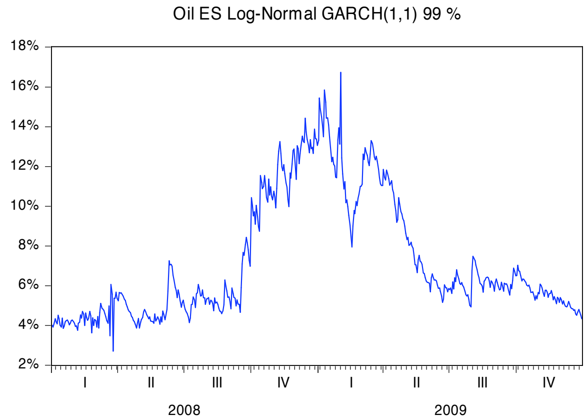

Oil ES Student-t GARCH(1,1) 95 %
18%
16%
14%
12%
10%
8%
6%
4%
2%
I
II
III
IV
I
II
III
IV
2008
2009


Oil ES Student-t GARCH(1,1) 99 %
28%
24%
20%
16%
12%
8%
4%
0%
I
II
III
IV
I
II
III
IV
2008
2009


Oil ES Log-Normal 95 %
4.8%
4.7%
4.6%
4.5%
4.4%
4.3%
4.2%
4.1%
I
II
III
IV
I
II
III
IV
2008
2009

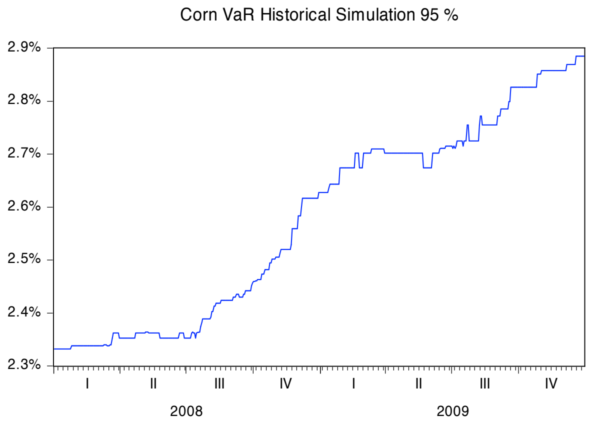

Oil ES Log-Normal 99 %
6.2%
6.1%
6.0%
5.9%
5.8%
5.7%
5.6%
5.5%
5.4%
I
II
III
IV
I
II
III
IV
2008
2009

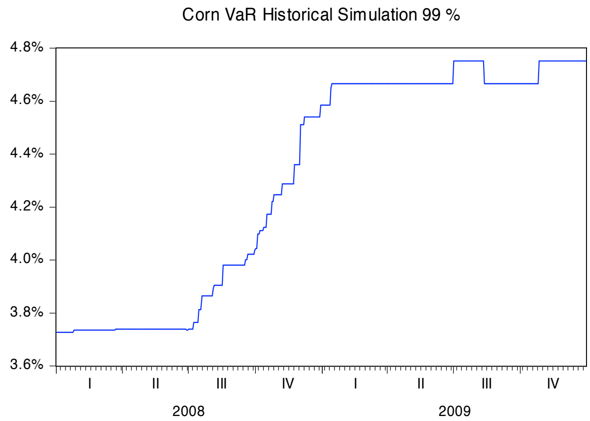

Oil ES Log-Normal GARCH(1,1) 95 %
14%
12%
10%
8%
6%
4%
2%
I
II
III
IV
I
II
III
IV
2008
2009

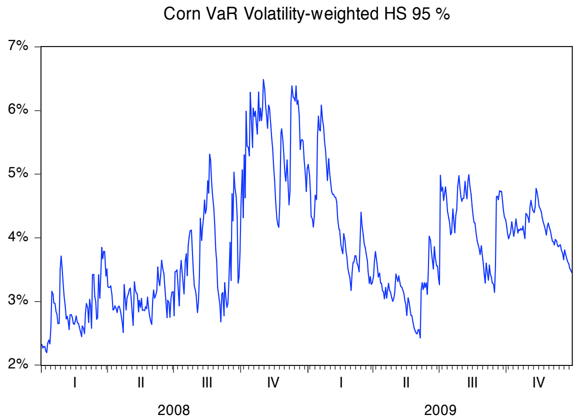

Oil ES Log-Normal GARCH(1,1) 99 %
18%
16%
14%
12%
10%
8%
6%
4%
2%
I
II
III
IV
I
II
III
IV
2008
2009

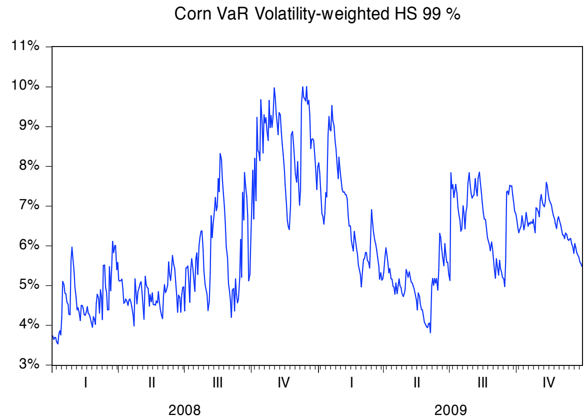
Appendix 5. Corn – VaR and ES

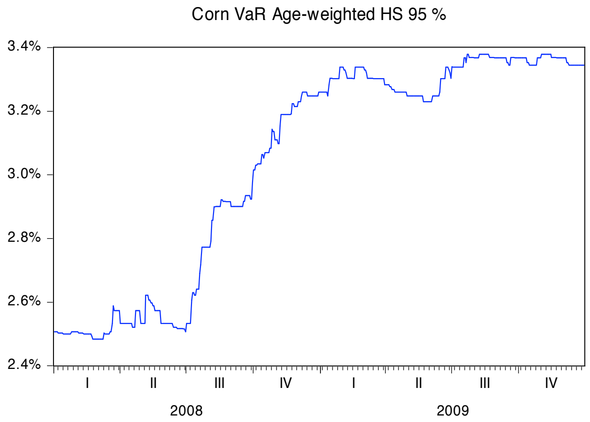
Note that the scales in the plots differ.

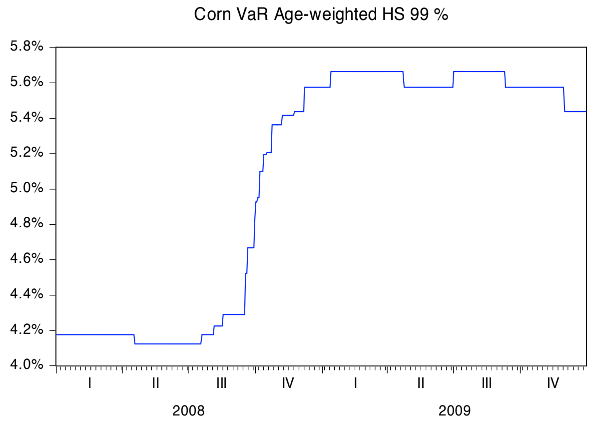




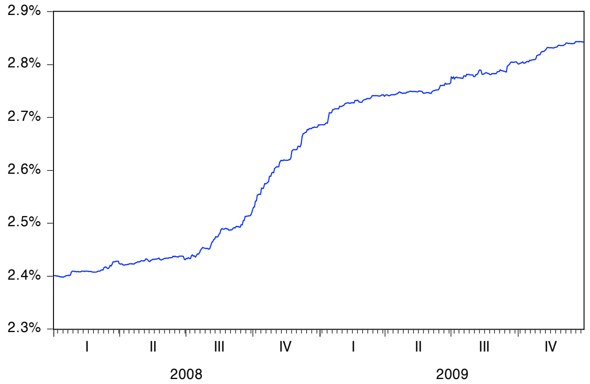




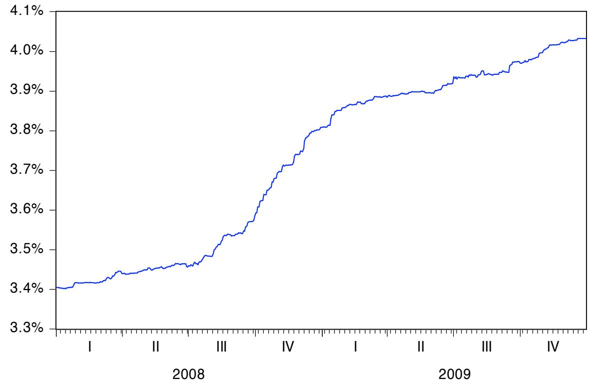




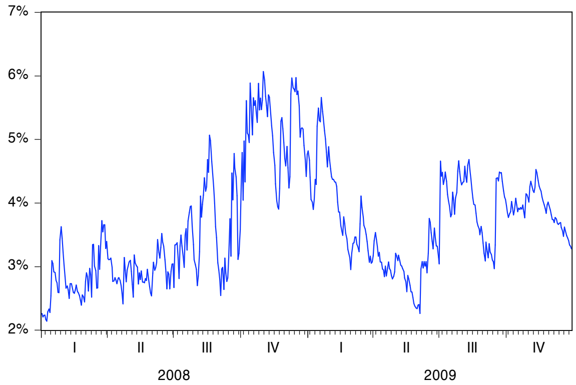
Corn VaR Normal 95 %



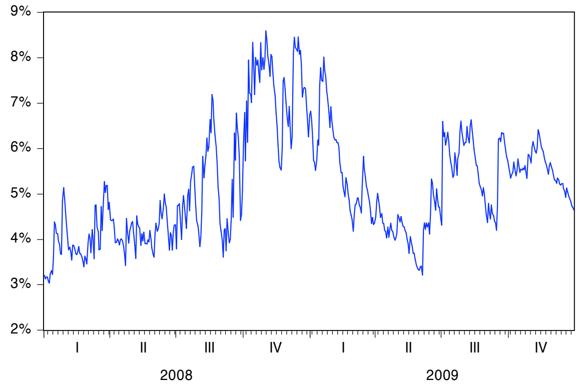
Corn VaR Normal 99 %



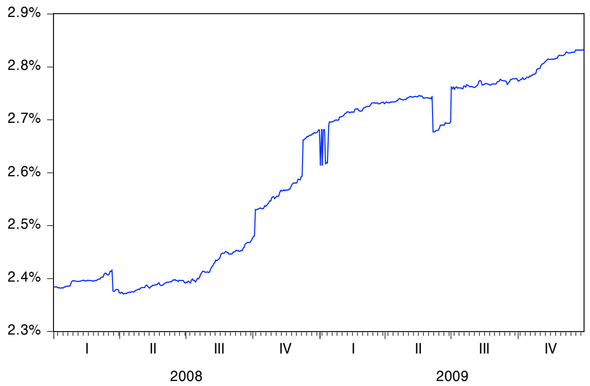
Corn VaR Normal GARCH(1,1) 95 %



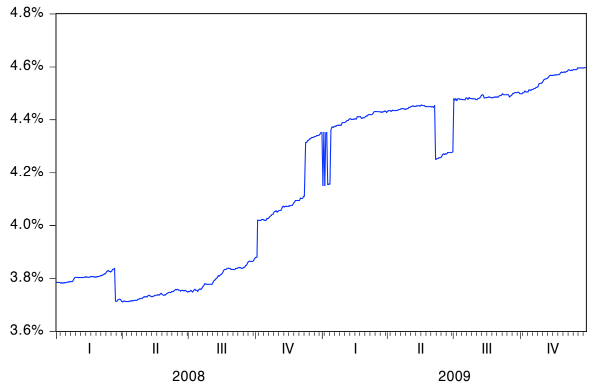
Corn VaR Normal GARCH(1,1) 99 %



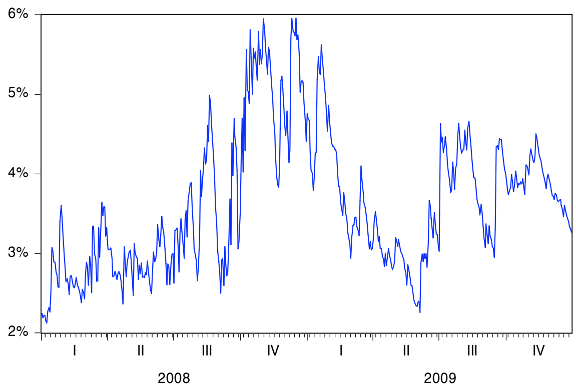
Corn VaR Student-t 95 %



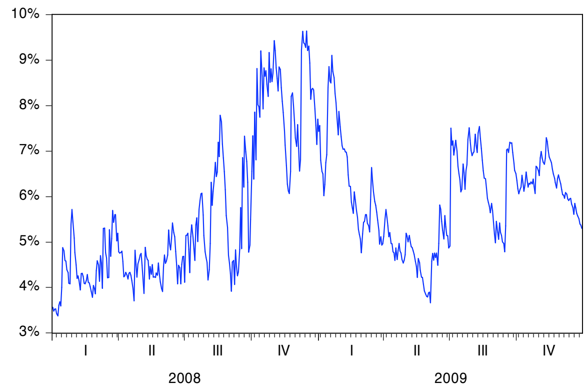
Corn VaR Student-t 99 %



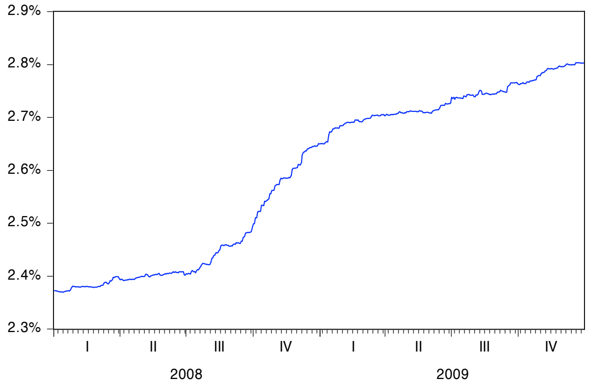
Corn VaR Student-t GARCH(1,1) 95 %



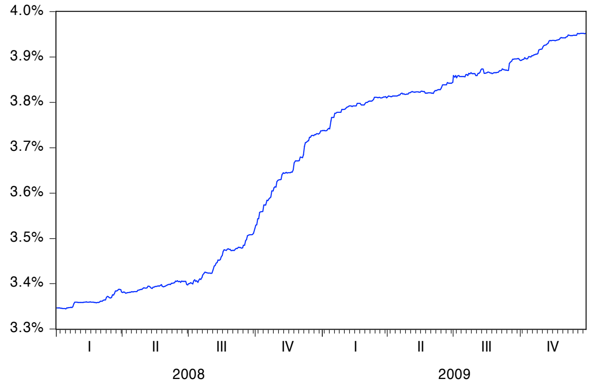
Corn VaR Student-t GARCH(1,1) 99 %



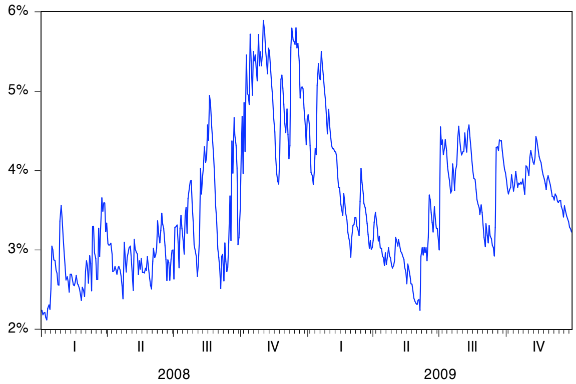
Corn VaR Log-Normal 95 %



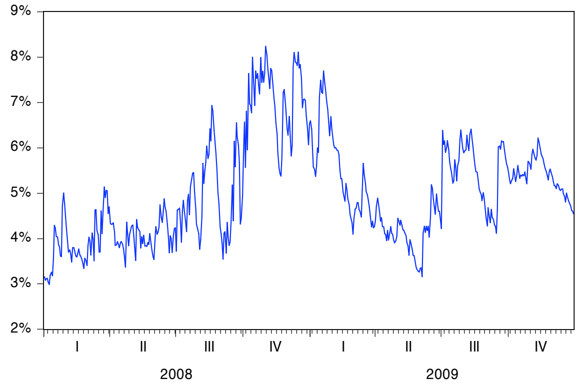
Corn VaR Log-Normal 99 %



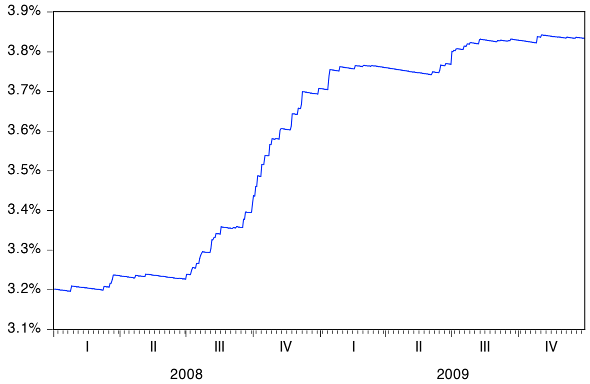
Corn VaR Log-Normal GARCH(1,1) 95 %



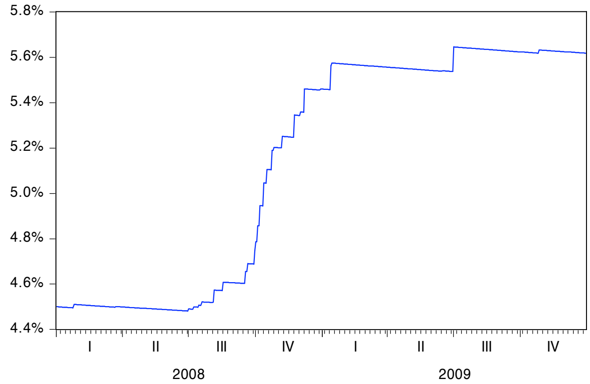
Corn VaR Log-Normal GARCH(1,1) 99 %



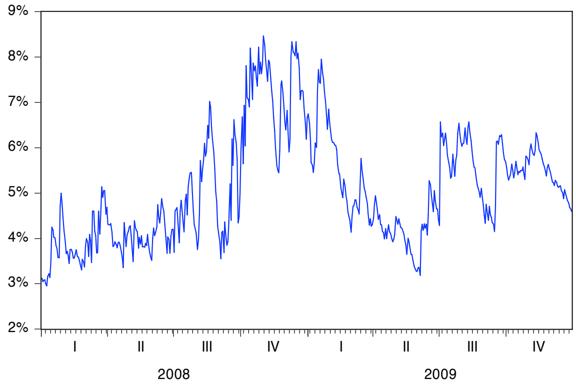
Corn ES Historical Simulation 95 %



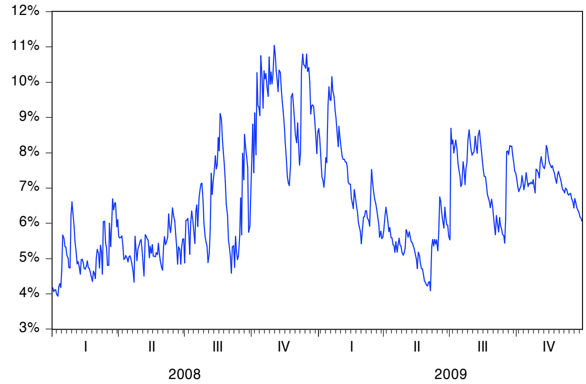
Corn ES Historical Simulation 99 %



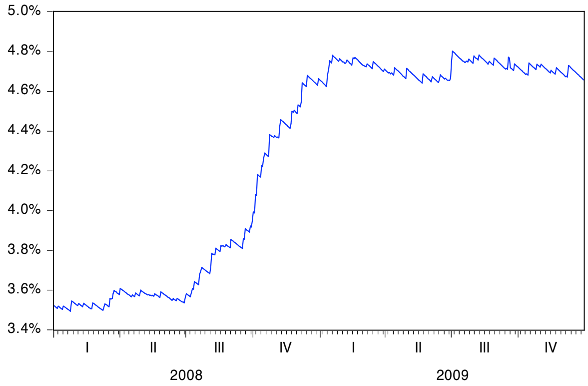
Corn ES Volatility-weighted HS 95 %



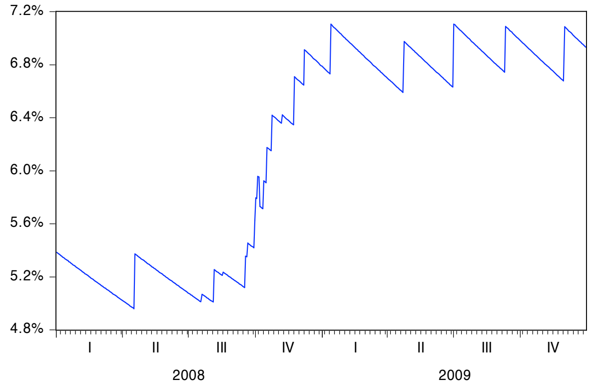
Corn ES Volatility-weighted HS 99 %



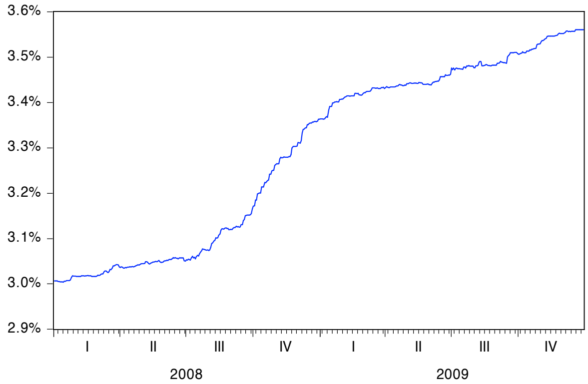
Corn ES Age-weighted HS 95 %



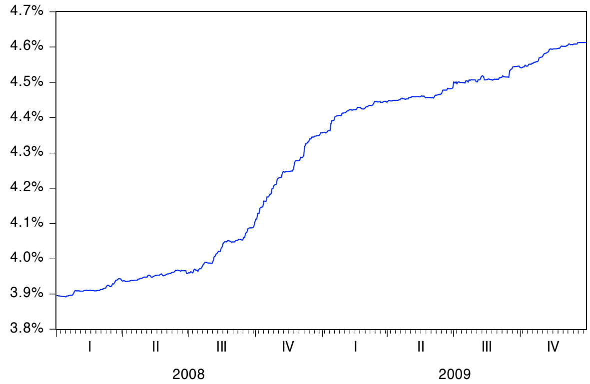
Corn ES Age-weighted HS 99 %



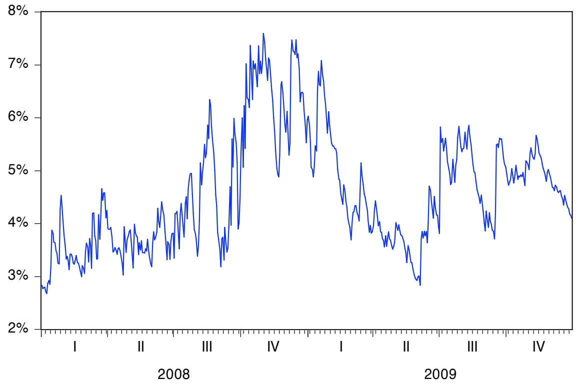
Corn ES Normal 95 %



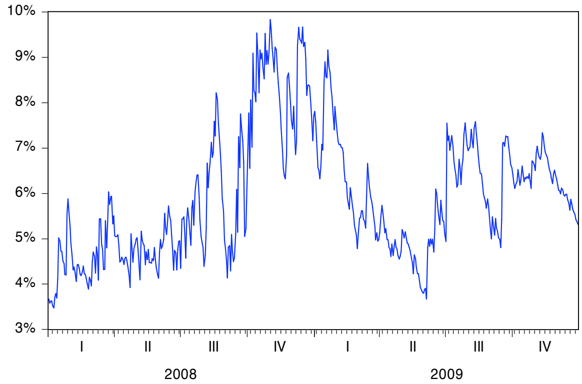
Corn ES Normal 99 %



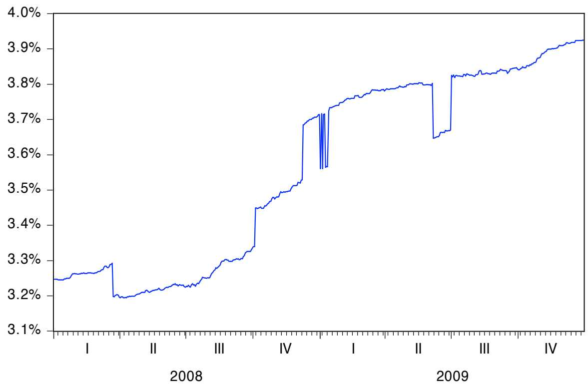
Corn ES Normal GARCH(1,1) 95 %



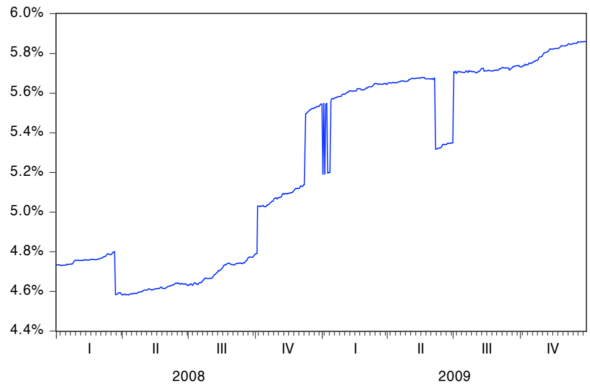
Corn ES Normal GARCH(1,1) 99 %



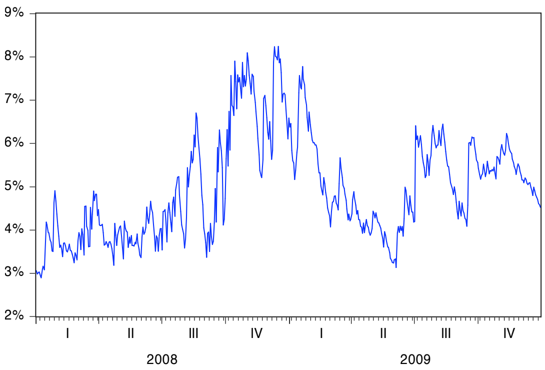
Corn ES Student-t 95 %



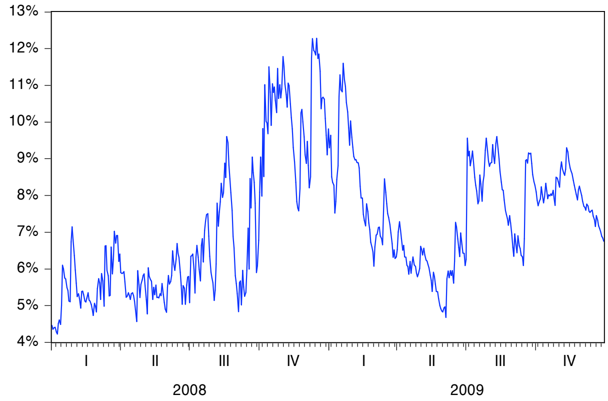
Corn ES Student-t 99 %



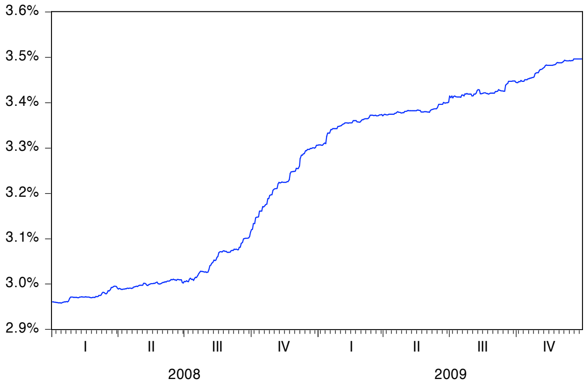
Corn ES Student-t GARCH(1,1) 95 %



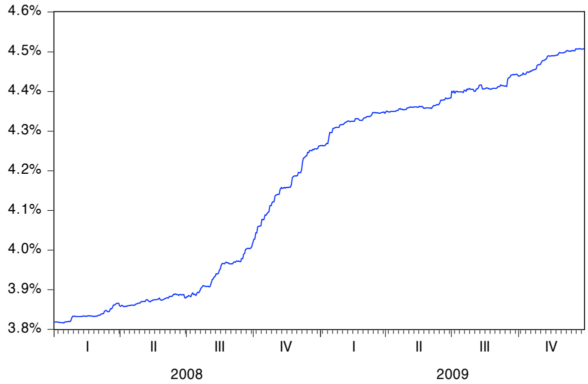
Corn ES Student-t GARCH(1,1) 99 %



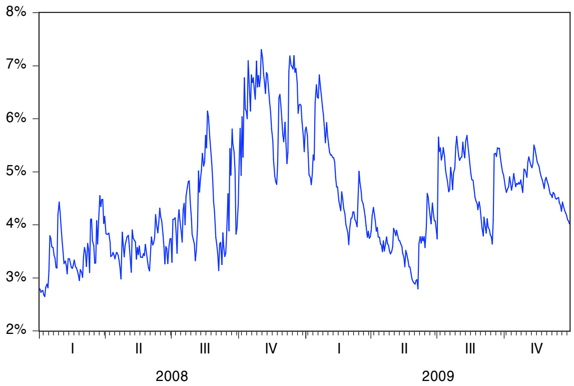
Corn ES Log-Normal 95 %



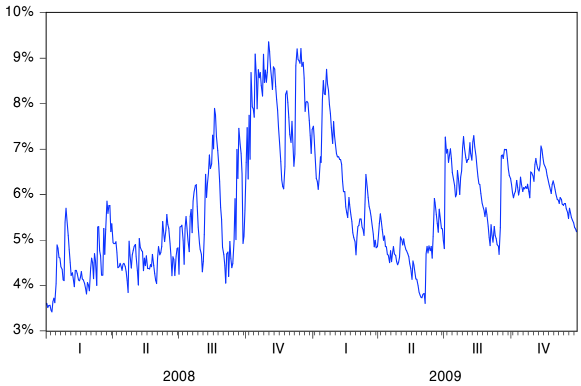
Corn ES Log-Normal 99 %



Corn ES Log-Normal GARCH(1,1) 95 %



Corn ES Log-Normal GARCH(1,1) 99 %



Appendix 6. GMZ test

GMZ test				
		Included observations		
Model		Gold	Oil	Corn
Historical Simulation	95%	67	12	30
	99%	14	1 (NA)	6
Volatility-weighted HS	95%	27	7	3
	99%	3	1 (NA)	0 (NA)
Age-weighted HS	95%	50	11	17
	99%	11	1 (NA)	3
Normal	95%	65	11	28
	99%	30	3	11
Normal GARCH(1,1)	95%	27	7	10
	99%	8	2	1 (NA)
Student-t	95%	65	11	29
	99%	15	1 (NA)	7
Student-t GARCH(1,1)	95%	28	7	10
	99%	3	1 (NA)	1 (NA)
Log-Normal	95%	66	11	28
	99%	30	3	11
Log-Normal GARCH(1,1)	95%	28	7	10
	99%	8	2	1 (NA)

Appendix 7. Gold – Correlation table VaR and ES

Gold Correlations VaR and ES										
Approach		**Historical Simulation**	**Volatility-weighted HS**	**Age-weighted HS**	**Normal**	**Normal GARCH(1,1)**	**Student-t**	**Student-t GARCH(1,1)**	**Log-Normal**	**Log-Normal GARCH(1,1)**
Historical Simulation	95%	0.983								
	99%	0.978								
Volatility-weighted HS	95%		1.000							
	99%		0.999							
Age-weighted HS	95%			0.977						
	99%			0.923						
Normal	95%				1.000					
	99%				1.000					
Normal GARCH(1,1)	95%					1.000				
	99%					1.000				
Student-t	95%						0.970			
	99%						0.971			
Student-t GARCH(1,1)	95%							1.000		
	99%							1.000		
Log-Normal	95%								1.000	
	99%								1.000	
Log-Normal GARCH(1,1)	95%									1.000
	99%									1.000

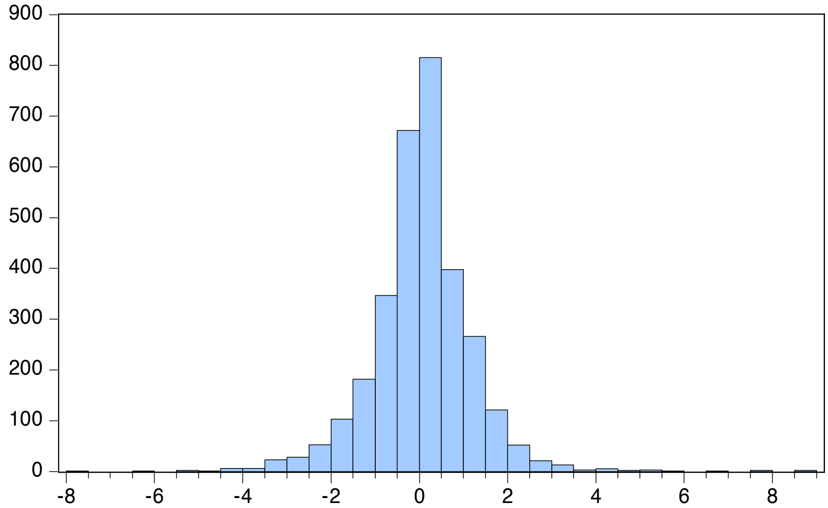
Appendix 8. Oil – Correlation table VaR and ES

Oil Correlations VaR and ES										
Approach		**Historical Simulation**	**Volatility-weighted HS**	**Age-weighted HS**	**Normal**	**Normal GARCH(1,1)**	**Student-t**	**Student-t GARCH(1,1)**	**Log-Normal**	**Log-Normal GARCH(1,1)**
Historical Simulation	95%	0.975								
	99%	0.968								
Volatility-weighted HS	95%		1.000							
	99%		1.000							
Age-weighted HS	95%			0.996						
	99%			0.988						
Normal	95%				1.000					
	99%				1.000					
Normal GARCH(1,1)	95%					1.000				
	99%					1.000				
Student-t	95%						0.999			
	99%						0.999			
Student-t GARCH(1,1)	95%							1.000		
	99%							0.999		
Log-Normal	95%								1.000	
	99%								1.000	
Log-Normal GARCH(1,1)	95%									1.000
	99%									1.000

Appendix 9. Corn – Correlation table VaR and ES

Corn Correlations VaR and ES											
Approach		Historical Simulation	Volatility-weighted HS	Age-weighted HS	Normal	Normal GARCH(1,1)	Student-t	Student-t GARCH(1,1)	Log-Normal	Log-Normal GARCH(1,1)
Historical Simulation	95%	0.975								
	99%	0.968								
Volatility-weighted HS	95%		1.000							
	99%		1.000							
Age-weighted HS	95%			0.996						
	99%			0.988						
Normal	95%				1.000					
	99%				1.000					
Normal GARCH(1,1)	95%					1.000				
	99%					1.000				
Student-t	95%						0.999			
	99%						0.999			
Student-t GARCH(1,1)	95%							1.000		
	99%							0.999		
Log-Normal	95%								1.000	
	99%								1.000	
Log-Normal GARCH(1,1)	95%									0.996
	99%									0.998

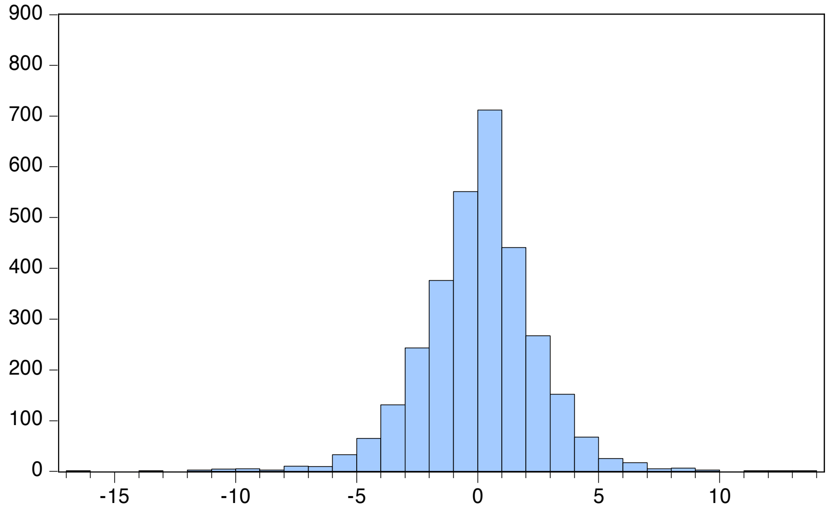
Appendix 10. Gold – Jarque-Bera



Series: GOLD
Sample 1/01/1998 12/30/2009
Observations 3130

Mean	0.042495
Median	0.000000
Maximum	8.828153
Minimum	-7.555714
Std. Dev.	1.137676
Skewness	0.263613
Kurtosis	9.317536
Jarque-Bera	5241.346
Probability	0.000000

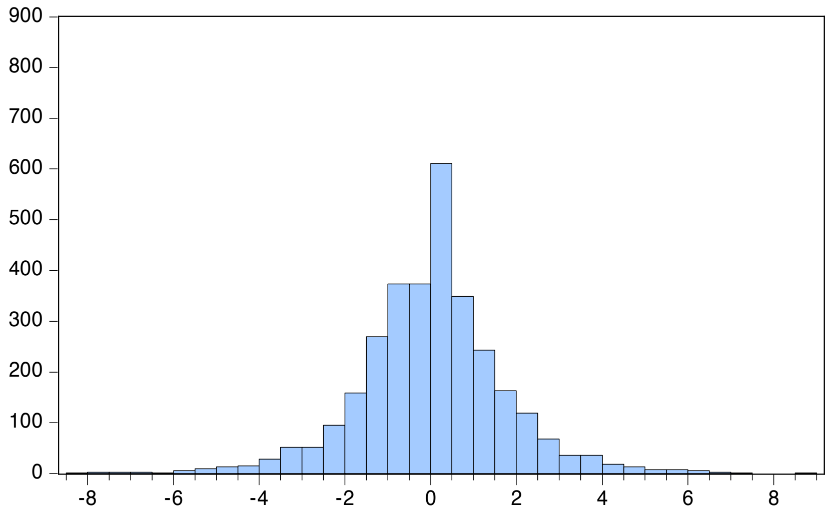
Appendix 11. Oil – Jarque-Bera



Series: OIL
Sample 1/01/1998 12/30/2009
Observations 3130

Mean	0.048045
Median	0.000000
Maximum	13.34145
Minimum	-16.54203
Std. Dev.	2.375633
Skewness	-0.272457
Kurtosis	6.169544
Jarque-Bera	1348.892
Probability	0.000000

Appendix 12. Corn – Jarque-Bera



Series: CORN
Sample 1/01/1998 12/30/2009
Observations 3130

Mean	0.014293
Median	0.000000
Maximum	8.663328
Minimum	-8.124313
Std. Dev.	1.680428
Skewness	0.012768
Kurtosis	5.179741
Jarque-Bera	619.7299
Probability	0.000000