University of Sussex

A University of Sussex DPhil thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

Risk Management of Energy Derivatives: Hedging and Margin Requirements

Anannit Sumawong

This thesis is submitted for the degree of Doctor of Philosophy

University of Sussex

September 2014

Declaration

I hereby declare that this thesis has not been submitted, either in the same or different form to this or any other University degree.

Signature:

Anannit Sumawong

Acknowledgements

I would like to thank my main supervisor Carol Alexander for her tireless inputs on my work and development. It has been a true privilege to work under her guidance. To her, I dedicate this Thesis and any success I may be fortunate enough to come across in the future. I thank Marcel Procopczuk for his aid in the first two years of my studies, his sharp critiques has taught me the value of rigour. I thank Ian Davidson for acting as my second supervisor in the third year of my studies. I thank Andreas Kaeck for his guidance in this last year. I thank the academic community whose various inputs have strengthen the quality of my work, especially the two anonymous referees who reviewed my first paper. I thank members of Platts, CME, ICE and the Bank of England whose insight of the industry greatly improved my understanding of the practical world. I thank my friends at the University of Sussex for making these past two years such an enjoyable experience. In particular I thank Xi Chen (Casey) for taking this journey with me from the very beginning. Her companionship is truly invaluable, through laughter, through tears, we have voyaged together to the end and I look forward to many more journeys to come. Finally, I thank Edward O'Reilly and my mother Intranee Sumawong for their love and support throughout the years. This Thesis would not have been possible without them.

Abstract

The recent growth of exchanges has generated large trading platforms for investors. The largest of these institutions, the Intercontinental Exchange and the Chicago Mercantile Exchange group are now responsible for clearing trades for the majority of investors worldwide and are perhaps, as large commercial banks are, too big to fail. This has attracted attention from international regulating bodies to impose strict risk management standards on the exchanges to ensure financial stability. In this thesis, we identify first, that an investor in the market is strongly affected by margins set by the exchanges in determining the transaction costs of a trade. We discuss the possibility that a volatile margin movement would introduce further risks for such an investor causing them to raise more capital to cover possible margin calls which can perhaps lead to procyclicality. We follow this work by addressing how margins can be determined in adherence to the new laws. Exchanges are now required to set margins based on the Value-at-Risk, hence we search for the best Value-at-Risk method for margining use. Here, we find that the simple Orthogonal Exponentially Weighted Moving Average method is sufficient in forecasting the Value-at-Risk, which contradicts a fair body of the literature who suggests that complex developments of GARCH are superior. We then offer methods for setting and evaluating margin requirements upon the Value-at-Risk estimates, concentrating on producing stable margin requirements. The automated methods produced in our work outperform all other methods available in the literature. Furthermore, we are the first to provide methods for assessing margin stability. Our work is timely in addressing the current affairs of the world economy and is among the first to tackle the margin stability issue in detail.

Contents

1. Introduction

The merger wave of futures exchanges in recent years has enhanced market liquidity and facilitated faster transactions for investors. In 2013, the Intercontinental Exchange (ICE) acquired the London International Financial Futures and Options Exchange (LIFFE) and Euronext, creating the largest exchange in Europe, in competition with the Chicago Mercantile Exchange (CME) group in the USA (which now includes the New York Mercantile Exchange (NYMEX), the Chicago Board of Trades (CBOT), the Commodity Exchange (COMEX) and more recently the Kansas City Board of Trade (KCBOT)). Both these exchanges control the majority of the world trade platform. Market participants -hedgers and speculators- are directly affected by margins issued by the exchange. Significant margin changes can perhaps cause these investors to implement risk management procedures on a mass scale, exacerbating the economic cycles which can ultimately lead to major financial distress. It is essential that one understands the impact margins have on the investors and that margin levels are controlled for the well-being of the world economy.

We first illustrate how margins can affect a common investor in the market, more specifically a refinery which buys crude oil and sells refined products. A popular hedging strategy for the refinery is delta-hedging, where one takes the positions on the futures contract to hedge spot price exposures. We compare the effectiveness minimum-variance hedging to the 1:1 naive hedge, and examine the transaction costs generated by each method. We delved into hedging spreads on oil products so that our results -on these complex underlyings- are generalisable to simpler products such as single positions on an equity index futures. Upon this investigation, we find that energy futures margins are difficult to recreate and that historical margins are unavailable. Hence, as a preliminary study, we assume that the margins remain constant over time, which is adequate in comparing hedging strategies. The refinery is directly affected by margin costs and should margins become variable, further risk factors such as margin calls may arise.

We expand our study further to examine how margins can be issued in accordance with new regulatory measures, paying special attention to generating stable margin requirements. Here, the investigation is split into two separate studies: first, we search for the most reliable Value-at-Risk (VaR) model for margin requirement purposes; second, we suggest methods for setting and evaluating the margin requirements.

Detailed historical data for margin requirements is limited to ICE's Brent crude oil futures only hence we apply our margin models on this derivative. We compare the results to historical margins produced by the Standard Portfolio Analysis of Risk (SPAN) software developed in 1988 which is currently the most widely-used software for determining the margin requirements. SPAN generates 16 scenarios for a portfolio's price movements, the most extreme of which (ScanRisk) is used as the margin requirement. Despite its limitations, the popularity of the software has resulted in the industry keeping its name and interface while the methods for generating the 16 scenarios adjust around its existence. The CME group for example now generates the scenarios using different Value-at-Risk (VaR) models, while ICE determines the scenarios heuristically around historical profit and loss series. Note that futures margining methodologies are not applicable to equity margining because stock margin levels are set by regulating bodies (e.g. the Federal Reserves in the USA) which focus only on maintaining the integrity of financial markets. This removes all competitive pricing elements from the exchanges. Moreover, for a long position, leveraging ones investment is a choice, not an obligation hence one cannot view the transaction costs from equity margining in the same light as futures.

In the derivatives market, energy futures is one of the fastest growing sectors worldwide. Its complex movements and richness in data combine to provide an excellent testing ground for our analysis. In this chapter, we provide a short summary of energy spot and futures markets. This is followed by an analysis of how the market players interact. Note that other derivatives, in particular options and swaps, are also actively traded but are not the focal point of our study, although these instruments provide interesting cases for further investigation.

1.1 Energy Markets

There are many complications in the trading of energy products. Unlike financial assets, energy spot prices are difficult to determine. The use of energy spot price data requires meticulous care in evaluating the price generation process, to determine whether or not the spot is truly representative of the transactions in practice. In this section, we present some of the physical aspects of the energy products traded by investors in the spot market. This has implications for our analysis of how a refinery may adopt strategies to hedge their spot positions. This Thesis studies risk management of crude oil and its refined products -gasoline and heating oil. A summary of the physical aspects of these commodities are provided below.

1.1.1 Crude Oil

Almost all petroleum products used for everyday consumption (heating oil, jet fuel, kerosene, bitumen to name a few) are refined from crude oil, making it the most traded energy product worldwide. Crude oil is formed by intense heat and pressure on large conglomerates of buried dead organisms and can be extracted from reservoirs naturally occurring around the world. The highest concentrations of crude oil can be found in Saudi Arabia, Iraq, Iran, Canada and Venezuela whose proved reserves[1] exceed 140 billion barrels in December 2012 (source: `www.eia.gov/countries/index.cfm?view=reserves#allcountries`). As such, political turmoil in these regions can affect crude oil prices and result in large swings in the global financial markets.

The two most heavily traded crude oils are: West Texas Intermediate (WTI) and Brent crude oil. WTI crude refers to a specific grade of crude oil with light, sweet content, which entails a low American Petroleum Institute (API) gravity[2] and low concentration of sulphur (0.24%). This type of crude oil is primarily processed into gasoline. The main trading hub of WTI crude is in Cushing, Oklahoma and is the flagship crude oil product for the CME group. Up until March 2011, WTI was the benchmark crude oil around the world where a barge of crude oil whose grade is undefined will be traded at the WTI crude oil price published by Platts plus a spread.

Brent crude is extracted from the North Sea region and also possesses the light, sweet property. Historically, Brent and WTI spot prices have a small spread but in March 2011 inventory levels in the USA were full and barges of oil were turned away at the port. WTI crude since lost its benchmark status to Brent crude and their prices heavily decoupled for 2 years. *Dated* Brent crude oil is perhaps the most heavily traded commodity to date. The term 'dated' refers to the price of a cargo whose delivery date is specified in advance. This is by all means, not the same as the futures price given the exchange of cash does not necessarily happen on the delivery date. As of May 2011, 60% of crude oil traded throughout the world is priced relative to Dated Brent.

[1] This refers to reserves where 90% of the well can be extracted with certainty
[2] A measure of the heaviness of the product relative to water

1.1.2 Refined products: Gasoline and Heating Oil

Gasoline and Heating oil are two of the major refined products from crude oil. Gasoline consumption as fuel for vehicles makes it one of the most popular petroleum product. Leaded Gasoline was prohibited in the USA in 1995 due to environmental concerns thus making its unleaded counterpart the most heavily traded gasoline contract until 2005 when it was replaced by Reformulated Blendstock for Oxygen Blending (RBOB) gasoline, which is now the main type of gasoline traded in both ICE and CME. Heating oil, (aka heating oil no.2 in the USA) is mostly used for heating homes and is heavily traded in the US and Europe.

Refined products prices are affected by a number of factors, although the price of crude oil is perhaps the most significant. Storage costs, convenience yield, seasonal demands cause their steep term structure and highly volatile nature. Heating oil for example is required for heating homes in the winter, hence its trading activity also rises around the same time. Gasoline can go stale after only one month of storage, while crude oil does not. These contribute to the volatile nature of the spread between crude oil and refined products (*crack spread*, discussed further in chapter 3).

1.2 Energy Futures

Futures and forwards are contractual obligations between two counterparties to trade a product at an agreed-upon price at expiry. Unlike forwards, futures contracts are guaranteed by a central counterparty (CCP), usually a clearing house operating within an exchange, and bears minimal counterparty risk. The exchange fixes the terms of a futures contract, also specifying the amount of commodity to be delivered and the location of the delivery. NYMEX's WTI crude oil futures for example, states that one futures contract denotes a trade of 1,000 barrels of WTI crude oil at expiry, to be delivered at Cushing, Oklahoma.

Traditionally, futures are traded during pit-trading hours in an open-outcry market where traders congregate to bid/ask for the best price. With the advancements of electronic trading, the number of pit-trading venues has been diminishing since. Few open-outcry platforms are present today, of which the London Metal Exchange (LME) is perhaps the most well-known. Electronic systems generate faster transactions and with it, rapidly growing trading volumes of derivatives cleared with a 2.1% rise in the futures and options contract worldwide from 2012 to 2013, see Table 1.1.

The growing trend of derivatives trading is clearly focused on alternative products. The fastest growing markets include commodity indices, credit, fertilizer, housing, inflation, lumber, plastics and weather which experienced a 95.2% increase in number of contracts cleared while energy products are second with a 36.7% growth between 2012 and 2013. Precious metals futures clearing activities have also grown by a significant amount at 34.9%. We choose to study energy futures as their term structures are more convex due to their shorter life span. One cannot generalise results from the prompt-month series on the rest of the term structure. Energy futures are also important drivers to the world economy, thus our research would be of interest to a greater audience.

Among the top ten energy products, futures outnumber options over 20 folds, see Table 1.2. Of which, ICE's Brent crude, CME's light sweet crude (aka WTI crude), RBOB gasoline and no.2 heating oil, which are considered in this study, account for roughly 30% of all energy derivatives trades. ICE's Brent crude has the largest trading volume with 159 million contracts cleared between January and December 2013.

Category	Jan-Dec 2012	Jan-Dec 2013	Change
Individual Equity	6,469,512,853	6,401,526,238	-1.1%
Equity Index	6,048,270,302	5,370,863,386	-11.2%
Interest Rate	2,931,840,769	3,330,719,902	13.6%
Currency	2,434,253,088	2,491,136,321	2.3%
Energy	925,590,232	1,265,568,992	36.7%
Agriculture	1,254,415,150	1,213,244,969	-3.3%
Non-Precious Metals	554,249,054	646,318,570	16.6%
Precious Metals	319,298,665	430,681,757	34.9%
Other	252,686,977	493,359,639	95.2%
Total	21,190,117,450	21,643,419,774	2.1%

Table 1.1.: Number of options and futures contracts traded globally between January 2012 and December 2013. *Other* includes commodity indices, credit, fertilizer, housing, inflation, lumber, plastics and weather. Source: `www.futuresindustry.org/downloads/FIA_Annual_Volume_Survey_2013.pdf`

We identify five major players in the energy futures trading platforms: speculators, hedgers, exchanges and their clearing houses within, brokers (or clearing members) and regulators. In this thesis, we concentrate on the operations of the hedger and the exchange. First, we examine how margins can affect a hedger who implements hedging strategies to reduce the volatility of their positions. Then we examine how margin requirements can be set, which has a direct impact on the traders in the market. The interactions between each of the player is summarised in Figure 1.1.

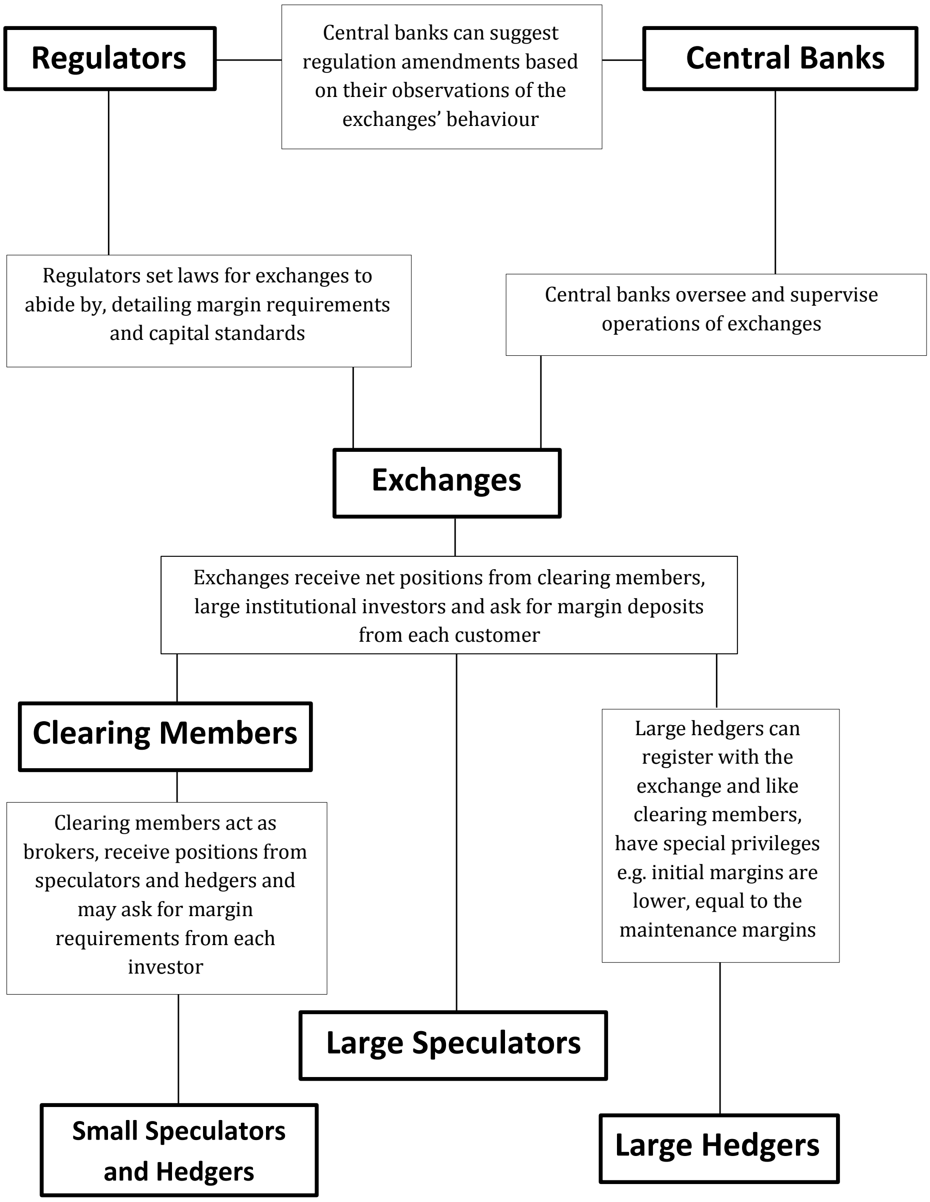

Figure 1.1.: Diagram summarising players in the futures and their interactions

Rank	Contract	Type	Jan-Dec 2012	Jan-Dec 2013	Change
1	Brent Crude, ICE	F	147,385,858	159,093,303	7.9%
2	LS Crude Oil, CME	F	140,531,588	147,690,593	5.1%
3	HH Natural Gas, CME	F	94,799,542	84,282,495	-11.1%
4	Gasoil, ICE	F	63,503,591	63,964,827	0.7%
5	Crude Oil, MCX	F	57,790,229	39,558,169	-31.5%
6	WTI Crude, ICE	F	33,142,089	36,106,788	8.9%
7	NYH RBOB Gasoline, ICE	F	36,603,841	34,470,288	-5.8%
8	No.2 Heating oil, CME	F	36,087,707	32,749,553	-9.3%
9	LS Crude Oil, CME	O	32,525,624	31,478,060	-3.2%
10	Natural Gas, MCX	F	27,886,670	23,828,800	-14.6%

Table 1.2.: Top ten traded energy derivatives and their exchanges, ranked by number of contracts traded between January 2012 and December 2013. Abbreviations: LS - Light Sweet, HH - Henry Hub, WTI - West Texas Intermediate, NYH - New York Harbour, RBOB - Reformulated Gasoline Blenstock for Oxygen Blending, ICE - Intercontinental Exchange, CME - Chicaco Mercantile Exchange, MCX - Multi Commodity Exchange. Derivative types: F - Futures, O - Options. Source: `www.futuresindustry.org/downloads/FIA_Annual_Volume_Survey_2013.pdf`

1.2.1 Hedgers and Speculators

Hedgers are often business entities such as agricultural farmers, power stations, refineries, etc.. For this study, we assume that they concentrate on maximising sell volume rather than generating profit from trades and fixing the profit margin is essential in ensuring effective business operations. For example, a farmer may be interested in selling corn in the next harvest season while they are obligated to pay other fixed costs such as employee wages and utility bills. Therefore, they must know their profit margin in advance so they can plan their expenditures. This can be achieved by taking positions in a futures or forward contract, expiring at the time of sale. It is highly conceivable that the farmer would prefer futures over forwards, given it contains negligible counterparty risk. Speculators on the other hand bear this risk by taking the opposite position with the aim of generating profits from the investment.

Upon entering a futures contract, each trader deposits a margin to the exchange which is then marked to market on a daily basis (mechanisms and rationales explained in section 1.2.2). Here, investors can offset (take the opposite position) at any point in time and retrieve/lose the difference in the futures price since its first undertaking. Hence, unlike the forward contract, the investor is not restricted to gaining the profits and losses only at expiry. Perhaps, It is this mechanism that lured speculators to the futures trading platform, as evident in the large proportion of contracts being offset where only 2% of commodity futures contracts were physically delivered in 2005 (see Alexander and Sheedy (2006)). This introduces a problem for the exchange as they must prepare the right amount of commodity to deliver upon expiry. To manage this risk, the exchange introduced the financial contracts, whereby the value of the commodity at expiry is delivered in cash instead.

Representative Investor: Refinery

Our first study concentrates on the refinery whose exposure to the crack spread has led them to use NYMEX's futures to hedge. Delivery points of gasoline, heating oil are in New York Harbour, while WTI crude can be collected from Cushing, Oklahoma and the refinery is assumed to be located in the USA.

We observe from Figure 1.2 that in 2012, the majority of US refineries are located in the Gulf coast. It is therefore not surprising that the main port for WTI crude oil delivery is in Oklahoma. Other refineries however are scattered around the US and face additional problems when considering hedging using futures due to their geographical location. A refinery based in Alaska for example is likely to take spot positions from producers and retailers/consumers from roughly the same region to reduce transportation costs, while taking positions in the NYMEX's commodities whose delivery points are in Oklahoma (WTI crude) and New York Harbour (Heatin Oil and Gasoline). The prices of each product at the delivery points are not the same as those in Alaska, hence the refinery is also exposed to the *location spread*. Alternatively, the refinery can collect/deliver the products at the contracts' delivery points but then they would be exposed to transportation costs. Further assumptions are discussed in chapter 3.

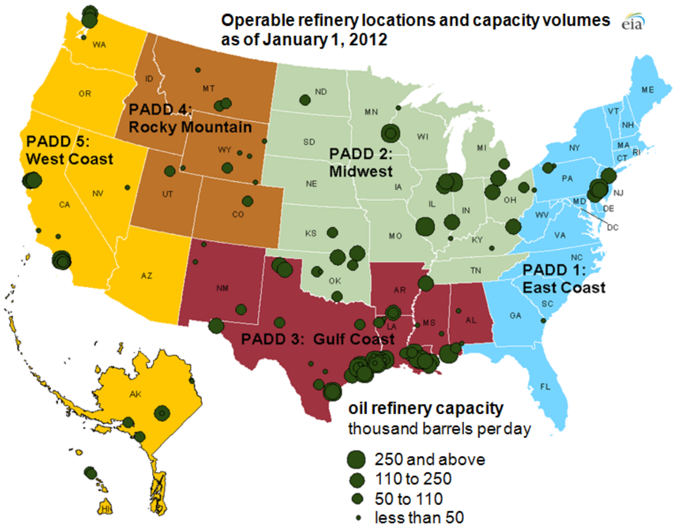

Figure 1.2.: Map of refinery locations and refining capacity in the USA on 1^{st} January 2012. Source: `http://www.eia.gov/todayinenergy/detail.cfm?id=7170`

1.2.2 Clearing House and Clearing Members

Since its introduction in 1853 the clearing house has developed two main roles: acting as a broker, lending to investors who wishes to leverage their investments; and taking on the counterparty risk of the buyer/seller of a futures contract, making sure the product(s) is(are) delivered at the agreed-upon prices at maturity. The clearing house is composed of clearing members, who act as brokers for smaller investors with higher credit risk. On top of this, clearing members do their own proprietary trading and, together with their clients' trades, issue net positions to the clearing house for the settlement process. The clearing house also allows registration from large speculators (other large institutional investors) and hedgers (e.g. refineries, manufacturers) who, contrary to smaller investors, have high credit ratings and are highly unlikely to default.

When entering into a futures contract should the buyer/seller default anytime between purchase/sale and delivery, the clearing house would have to take on the position of the defaulter and immediately look for a new match in the market. To reduce the losses from the price movement between the time of purchase/sale and default, the clearing house marks-to-market the client/clearing member's net positions, where the futures price P&L are transferred to and from the recipients on a daily basis.

The clearing house also asks for collateral in dollar amounts (aka initial margin) to reduce the possible losses between marking-to-market periods. Should the margin account (in dollar amounts) reduce below a predefined level (the maintenance margin), the exchange would ask for a margin call where the investor/clearing member must replenish the margin account back to the initial margin level. Because hedgers and members are generally more risk averse than speculators, exchanges such as ICE and CME set the initial margin at the same level as the maintenance margin. In other words, these investors do not need to raise as much initial margin as speculators but are more exposed to margin calls, which are triggered from any fall in their portfolio value. The amount paid to/received from the exchange is called the *variation margin*.

In addition to the margin requirement the clearing house also places its own capital in a default fund to protect themselves during abnormally volatile market conditions. In an event of a default, should the clearing house take on a loss, it would follow the default waterfall, where the defaulter's margin is first liquidated then losses are further diffused by the risk-sharing pool (additional capital from clearing members) and finally by the default fund. Hence, the initial margin does not necessarily have to reflect extreme events since they are covered by these other risk management strategies. This is also found in historical margin estimates; for instance, there was no reaction from ICE's historical Brent crude futures margin to the price spike on 5^{th} May 2011. Capital levels for clearing houses present an interesting case, however since we are solely dealing with margin requirements, no such attempt has been made to model the rest of the capital structure in the default waterfall (see Shanker and Balakrishnan (2006) for further details on clearing house capital requirements). A summary of world's top 10 exchanges and the number of contracts cleared are listed in Table 1.3.

Rank	Exchange	Jan-Dec 2012	Jan-Dec 2013	% Change
1	CME Group	2,895,125,126	3,161,476,638	9.20%
2	Intercontinental Exchange	2,448,099,505	2,807,970,132	14.70%
3	Eurex	2,291,368,356	2,190,548,148	-4.40%
4	National Stock Exchange of India	2,010,958,057	2,135,637,457	6.20%
5	BM&FBovespa	1,636,327,195	1,603,600,651	-2.00%
6	CBOE Holdings	1,134,329,197	1,187,642,669	4.70%
7	Nasdaq OMX	1,115,078,250	1,142,955,206	2.50%
8	Moscow Exchange	1,062,244,624	1,134,477,258	6.80%
9	Korea Exchange	1,835,938,749	820,664,621	-55.30%
10	Multi Commodity Exchange	960,098,730	794,001,650	-17.30%

Table 1.3.: Number of futures and options contracts cleared by the top 30 exchanges between January 2012 and December 2013. Source: `www.futuresindustry.org/downloads/FIA_Annual_Volume_Survey_2013.pdf`

1.2.3 Regulators

Following the 2008 financial crisis, margin requirements have received considerable attention from international regulating bodies. We focus our discussion on European and US exchange laws as they make up the majority of the clearing operations throughout the world, see Table 1.3 for more details.

The main regulating body for European clearing houses is the European Securities and Markets Authority (ESMA) which replaced the Committee of European Securities Regulators (CESR) in 2011. The main goal of ESMA is to provide European markets with strict supervision with special attention on credit rating agencies and their valuation of the companies. In 2012, ESMA drafted EMIR which contains strict details on clearing houses capital requirements and is currently going through the implementation stage. EMIR suggests strict requirements on the coverage level of the initial margin (two-day, 99% for non-OTC products) whilst requiring that the clearing house must avoid large jumps in the maintenance margins to limit procyclicality.

A variety of EMIR Articles affect margin requirements directly. The EMIR recommendations are suggestions to the central governing body (e.g. Bank of England in the UK) for their implementation in national law. But even before local implementation, adhering to EMIR will portray the clearing house in a positive light, strengthening its reputation for good risk management infrastructure. Subsequently, this could be highly beneficial to its business. As margin changes can cause investors to implement risk management strategies simultaneously, further amplifying economic downturns, EMIR regulations pay particular attention to avoid such an event. With large exchanges affecting more investors, it is expected that this procyclic nature will be prominent now more than ever.

The most notable regulatory reform on clearing house's margin requirement is outlined in the Title VII of the Dodd-Frank act in 2010 which require many over-the-counter (OTC) derivatives, swaps in particular, to be cleared in exchanges. The Act however does not impose any restrictions on futures margin requirements, hence exchanges are free to set margins as they want. The Dodd-Frank act is overseen by the Commodity Futures Trading Commission (CFTC) and the Securities and Exchange Commission (SEC).

The variety of regulatory system causes some divide between margin requirement methodologies. But as financial derivatives are being traded in a global scale, investors are free to move between exchanges as they wish. Unequal regulatory requirements produces varying levels of benefits to investors and with this, possible mass migration between exchanges may become prominent. It is beneficial for regulators to work together in creating laws to avoid such movements as closure of large exchanges may be detrimental to the world economic outlook.

1.3 Summary and Layout of the Thesis

Although crude oil reserves are depleting, the energy derivatives market is expanding rapidly. One would expect that the number of trading venues for such products would increase to facilitate the growing number of contracts being cleared. Instead, large exchanges are merging to facilitate the high demands for the liquidity of these products. Furthermore, in the wake of the 2008 world-wide recession, regulators are imposing stricter risk management standards to ensure the financial stability of these large institutions. These changes entail a need for innovative risk management methods. To this end, we examine the risk management methods for participants in the energy market, paying particular attention to the changing circumstances of the regulatory reforms.

The structure of the rest of the Thesis follows: first, we provide a critical examination of the literature which supports the rationale behind our minimum-variance hedging tests, VaR models and margin rules. Second, we evaluate hedging approaches carried out by a refinery, paying special attention to the transaction costs and how margins may affect such an investor. Third, we search for the best VaR estimation method, which will be applied to margin requirements. Fourth, we identify procedures for setting and evaluating margin requirements, with the view to create stable margins following current regulations. Lastly, we conclude.

2. Literature Review

In this review, we first evaluate the literature on commodity price behaviour, detailing the complexity of their movements. Energy futures volatility, which is essential when deriving minimum-variance hedge ratios, Value-at-Risk (VaR) and margin requirements, is hence difficult to forecast. We found a number of deterministic movements prevalent in the literature, which may help improve the estimation power of volatility models. We then identify gaps and inconsistencies in the literature on hedging, VaR and margin requirements.

2.1 Commodity Price Behaviour

A significant amount of the underlying risk in the energy markets arises from the price process, it is therefore important to understand this in order to model volatility effectively. Commodity prices behave very differently to financial products, this is due to several reasons: commodities are held for consumption purposes as well as investment purposes, commodities require storage which is costly, financial products are actively lent and borrowed whereas most commodity products are not, commodity prices are genuinely more predictable since they are mean-reverting and also display seasonality effects.

2.1.1 Seasonality

The supply and demand of commodities are driven by several external factors, e.g. winter in cold countries causes a demand for more electricity and fuel for heating; droughts cause the supply of agricultural products to diminish. Seasonality effects in the literature are well established with evidences found in Fama and French (1987), Milonas (1991) and Sorensen (2002). The seasonality effects on commodity prices imply deterministic cycles of demand and supply patterns, commodity prices are hence expected to exhibit regime switching behaviours in both price and volatility. Subsequently, hedging models are expected to be responsive to the changing market conditions.

2.1.2 Samuelson Effect

The time-to-maturity effect or the Samuelson effect is when the variance of the futures price increases as the contract approaches maturity. Samuelson (1965) proves that this only occurs when the spot price of the underlying is a mean-reverting process. The findings have influenced a substantial amount of literature on commodity derivatives, Haigh and Holt (2002), Geman and Kharoubi (2008), Back et al. (2013) to name a few.

The Samuelson effect has proved useful in several fields of study in the commodity market (e.g. option pricing and hedging with futures). Integrating the effect for a single asset is straight forward. For multiple assets however, this is not since the behaviour of the covariance between the commodity futures as it approaches maturity is unclear. There is a lack of literature which had properly analysed this feature. Haigh and Holt (2002) for example, model this effect by adding a matrix of so-called 'exogenous' variables. They did not however, explicitly explain how these were determined nor did they justify whether or not the model was successful. There are also works which do not entirely agree with the Samuelson effect. Most notably, Fama and French (1988) show that the effect does not hold at high inventory levels. Routledge and Seppi (2000) introduces a model for futures term structures which at times predict that the variance of the futures prices remain constant towards maturity.

2.1.3 Stochastic Convenience Yield

The difference between the spot and futures price is referred to as the basis. It is common among commodity markets that sometimes the basis is negative, the futures price here is said to be in backwardation. There are several theories which try to explain this behaviour, the most celebrated of which is the convenience-yield theory, as supported by Kaldor (1939), Brennan (1958) and Fama and French (1987). This is simply the premium at which the investor is willing to pay for an immediate replenishment of their inventory. Commodity futures are hence derived not only from the expected spot price and the carrying cost but also the convenience-yield as well. An example of the application of this model for hedging purposes is Schwartz (1997), who applies the Gibson and Schwartz (1990) two-factor model (with stochastic spot price and convenience-yield) and also formulated the three-factor model (also including stochastic interest rates) in the oil futures pricing process for hedging futures positions. More recently, Dempster et al. (2008) also uses the two-factor model in pricing futures spreads directly. To our knowledge there are no application of stochastic convenience yield models to estimating VaR.

2.1.4 Mean-Reversion

The mean-reversion process (aka Ornstein-Uhlenbeck process) also plays a major role in modelling commodity futures prices. The method of its application is well established, the earliest application of this process in the commodity market is from Gibson and Schwartz (1990) where the forward convenience yield is assumed to follow this relationship. It is not until Schwartz (1997) however that this is extended to the log spot prices as well, also providing a better fit to the market, hence allowing for more effective outcomes when hedging futures positions in the oil markets.

2.2 Hedging

When one considers hedging the price risk, one takes the opposite position in a hedging instrument that is highly correlated to their portfolio to ensure the most effective hedge. Such instruments can be proxies which share the same demand and supply patterns, or simply derivative contracts underlying the portfolio itself. In this section, several different types of strategies are explored to give further ideas about what can be used in our research.

According to Working (1953), the futures market in the commodity world is utilised by hedgers and speculators. Hedgers are investors who buy the commodity at the spot position and sell it at a later date. The main source of income for the hedger is not from this operation but from the activities the commodity performs while under the investor's ownership. The hedger hence wishes to bear no price risk in the selling back the commodity. Appropriate hedging strategies can be applied to minimise this risk. Processing spreads however, apply to a slightly different type of investor, a refiner. The investor here will need to replenish their inventory and sell their goods regularly, hence they commit themselves to spot transactions on an equally frequent basis. Since these occur in the future, the investor is exposed to the movement of the spread. There are several derivatives which can be used to manage the price risk: spread futures, spread options, Asian spread options, etc.. Unlike the other derivatives, futures have a linear payoff and are much more straightforward to analyse, making it a popular hedging instrument.

2.2.1 Hedging with Futures

The simplest manoeuvre to hedge an impending long(short) spot position would be to buy(sell) the futures contract whose maturity coincides with the due date for the spot transaction. By doing this, the investor locks the price of the commodity, the portfolio bears no price risk and the spot position is no longer needed. However, this is not always possible given that there may not be a futures contract which exactly fits the required specifications. In this case, the investor would need to adapt an alternative strategy in order to execute the hedge. One method would be to take advantage of the highly correlated nature between the futures and its underlying's price by taking an opposite position in the futures contract. The problem associated with this is that once the futures contract expires, the investor is faced with an obligation to purchase or sell an unwanted amount of the underlying. To avoid this, they must clear the futures position and use the profits from this transaction to hedge the exposed spot position. This strategy can be found in the works of Gagnon et al. (1998), Haigh and Holt (2000) and Haigh and Holt (2002). Another point to consider, especially in the case of the oil market is that, the investor would often have futures contracts with several maturities to choose from. In this case, the entire term structure of futures could be used to hedge a single spot position, this added flexibility could prove beneficial to the hedging outcomes. There are many works associated with hedging the term structure (Driessen et al. (2002) and Bliss (1997)) but none which utilises it as a hedging tool.

After implementing a hedging strategy, the investor needs to determine the weight on the hedging instrument for the most effective outcome. Consider a simple of case of hedging a portfolio of spot positions with futures contracts (generalised from Working (1953)), the hedged portfolio P&L, $\Delta\Pi_t$, can be described by vectors of m spot P&Ls, futures P&Ls and hedge ratios ($\Delta\mathbf{S}'_t$, $\Delta\mathbf{F}_{t,T}$ and $\boldsymbol{\beta}$) such that

$$\Delta\Pi_t = \mathbf{1}'_m\Delta\mathbf{S}_t + \boldsymbol{\beta}'\Delta\mathbf{F}_{t,T} \ , \tag{2.1}$$

where $\mathbf{1}_m$ is a $m \times 1$ vector of ones. The variance, $V[\Delta\Pi_t]$, of this portfolio is given by

$$V[\Delta\Pi_t] = \mathbf{1}'_m V[\Delta\mathbf{S}_t]\mathbf{1}_m + \boldsymbol{\beta}' V[\Delta\mathbf{F}_{t,T}]\boldsymbol{\beta} + 2\boldsymbol{\beta}' Cov[\Delta\mathbf{F}_{t,T}, \mathbf{1}'_m\Delta\mathbf{S}_t] \ , \tag{2.2}$$

where $V[\Delta\Pi_t]$, $V[\Delta\mathbf{S}_t]$ and $V[\Delta\mathbf{F}_{t,T}]$ are the covariance matrices of $\Delta\Pi_t$, $\Delta\mathbf{S}_t$ and $\Delta\mathbf{F}_{t,T}$, respectively, and $Cov[\Delta\mathbf{F}_{t,T}, \mathbf{1}'_m\Delta\mathbf{S}_t]$ is a vector of covariances between $\mathbf{1}'_m\Delta\mathbf{S}_t$ and the individual elements of $\Delta\mathbf{F}_{t,T}$. Up until Johnson (1960), hedging involved the investor taking an equal and opposite position in the futures contract compared to the spot, i.e. $\boldsymbol{\beta} = -\mathbf{1}_m$ (also referred to as naive hedging). The intuition behind this is: given that spot and futures are perfectly correlated, if the spot price moves, the futures price would also move in the same direction. In the case where both the spot and futures move by the same magnitude, the portfolio payoff $\Delta\Pi_t$ would always be zero regardless of the direction of the spot movement, hence locking in the payoff of the portfolio. In practice however, futures and spot prices rarely move with the same magnitude hence the need for the formulation of the hedge ratio.

There are two main approaches of calculating the hedge ratio: the minimum-variance approach and the mean-variance approach (aka 'Optimal' hedging). For the former, the investor is only concerned about reducing the variance of the portfolio. For the latter, the investor is also concerned with maximising the profit of the portfolio as well as reducing the variance. The introduction of mean-variance hedging started with Working (1953), arguing that when performing a naive hedge, hedgers are more concerned about utilising the futures for speculation purposes as well as hedging. Following this, Johnson (1960) points out that hedgers are interested in maximising the expected returns and reducing the variance of the portfolio at the same time. From this, Johnson (1960) develops two methods for calculating the hedge ratio: one based on the utility function and another for minimising the price risk only. The former approach follows a mean-variance framework, the second is the minimum-variance model. [1]

2.2.2 Minimum-Variance Approach

The minimum-variance hedge ratio is derived by minimising the portfolio variance under first-order conditions with respect to the investment weight on the futures contract. Minimising the hedged-portfolio variance yields the optimal hedge ratio vector

$$\boldsymbol{\beta} = -V[\Delta\mathbf{F}_{t,T}]^{-1}Cov[\Delta\mathbf{F}_{t,T}, \mathbf{1}_m'\Delta\mathbf{S}_t] \ , \tag{2.3}$$

which is the global minimum since $\frac{\partial^2 V[\Delta\Pi_t]}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}} = V[\Delta\mathbf{F}_{t,T}]$ is always positive definite. The hedge ratio is dependent upon $Cov[\Delta\mathbf{F}_{t,T}, \mathbf{1}_m'\Delta\mathbf{S}_t]$. This however introduces a problem of measuring these terms accurately. The simplest way to calculate these would be to take a sample of returns or profits and losses (P&L's) on each asset from historical data and then calculate the sample unconditional variance and covariance (aka OLS hedge ratio). However, this may not be responsive enough to the volatile market conditions. Recently, there has been a number of studies which have used conditional variance estimation techniques belonging in the GARCH family to estimate

[1] Johnson (1960) is often revered as the first work on the minimum-variance framework. However, conclusions of the work clearly favour the idea that hedgers also display speculative behaviours, thus supporting Working (1953). This has created some confusion among academics in the past for example, Kahl (1986) heavily criticises Brown (1985) on the misinterpretation of the approaches of Johnson (1960) and Stein (1961). In this review, Johnson (1960) is considered as a follower of the mean-variance approach

time-varying hedge ratios, some works in the minimum-variance framework include: Baillie and Myers (1991) who implements the bivariate GARCH model in estimating hedge ratios for multiple commodities at the same time; Moschini and Myers (2002) supports the use of GARCH variance to calculate the hedge ratios in the corn market; Chan and Young (2006) incorporates jump diffusion in the GARCH framework and find that this is beneficial for hedging copper prices; Lee and Yorder (2007) uses a Markov regime-switching GARCH model to test hedging results in the corn and nickel markets. Alexander and Barbosa (2007) reject the usefulness of this and found that simple constant variance or naive hedges outperform the GARCH hedge ratios in several stock markets.

The effectiveness of the minimum-variance hedge is measured by the Ederington Effectiveness (Ederington (1979)). This is simply the percentage reduction in the variance gained by the hedging strategy, i.e. $(\sigma_u^2 - \sigma_\pi^2)/\sigma_u^2$, where σ_u is the variance of the unhedged portfolio. Works which employ this measure, using the unconditional variance and covariances include: Dale (1981) in the foreign exchange markets, Lence et al. (1993) in the commodity markets and Herbst et al. (1989) in both. Lien (2005) however, points out that the Ederington Effectiveness is biased towards unconditional variance models over conditional variance models. Alexander and Barbosa (2007) hence applies the use of a dynamic measure of hedging effectiveness by measuring percentage changes in conditional variances instead.

2.2.3 Mean-Variance Approach

Most works in the mean-variance hedging school uses the exponential utility. The hedge ratio under this assumption can be determined by maximising the certainty equivalent income (CEI) with respect to the investment weight on the hedging instrument. Following Working (1953), the investor's expected utility value of the hedged portfolio is measured by the certainty equivalent income, $\mathrm{CEI}[\Delta\Pi_t]$. Under certain restrictive conditions this can be expressed analytically as

$$\mathrm{CEI}[\Delta\Pi_t] = E[\Delta\Pi_t] - \frac{1}{2\gamma} V[\Delta\Pi_t] \ , \tag{2.4}$$

where γ is the coefficient of absolute risk tolerance and $E[\Delta\Pi_t]$ and $V[\Delta\Pi_t]$ denote the expectation and variance of $\Delta\Pi_t$ respectively. Maximising the CEI gives the optimal mean-variance hedge ratio vector as

$$\boldsymbol{\beta}^* = V[\Delta\mathbf{F}_{t,T}]^{-1} \left(\gamma E[\Delta\mathbf{F}_{t,T}] - Cov[\Delta\mathbf{F}_{t,T}, \mathbf{1}'_m \Delta\mathbf{S}_t]\right) \ . \tag{2.5}$$

Unlike the minimum-variance approach, the hedge ratio here is not only governed by the variance and covariance terms but also the expected P&L of the hedging instrument. Most works however, assume that the futures contracts used as a hedging instrument follows a martingale process with zero expected return, which is identical to the minimum-variance framework. [2] Works which apply this, using the OLS method include: Brown (1985) who applies portfolio maximisation theories to the wheat, corn and soybean markets, Stulz (1984), on the foreign exchange markets and Myers and Thompson (1989) on the corn, soybean and wheat markets. Those who use conditional variances and covariances include: Kroner and Sultan (1993) and Gagnon et al. (1998) with their works on the foreign exchange market; more recently, Lee (2009) utilises jump-switching dynamics in the Generalised Orthogonal GARCH model in hedging the FTSE100. On a conditional framework however, applying the martingale property of futures is not entirely appropriate. The intuition of Working (1953) follows that the expected return also influences the hedger; by ignoring this, the works almost entirely defeat the ideology and may as well have followed the minimum-variance approach instead. Works which do include the expected return in their estimations include: Haigh and Holt (2000), Haigh and Holt (2002) and Lee (2010). All of these tested both the minimum-variance and mean-variance hedge ratios. The measure of hedging effectiveness in the mean-variance approach is simply the increase in CEI of the hedged portfolio from the unhedged. Possible extensions in the mean-variance framework include: exploring more utility functions or include more moments in the exponential utility functions when calculating the CEI to also take account of the skewness and kurtosis of the portfolio payoff as seen in Brooks et al. (2002). For the latter extension however, one also needs to estimate the GARCH skewness and kurtosis. So far, most works in the hedging area favours the use of time-varying volatility estimation methods for computing the hedge ratios. The GARCH model has been thoroughly analysed in both mean-variance and minimum-variance hedging literature. Adding further extensions to this may complicate the model by many folds, only to receive little gain in comparison. The best research strategy would be to focus on simpler models which can provide hedge ratios that are just as effective.

2.2.4 Other Approaches: Cross and Composite Hedging

Cross hedging stems from Anderson and Danthine (1981) where prices from one market is used as a hedging instrument against another. For example given that cotton-

[2] Although the hedge ratio between the two approaches are equivalent, we have chosen to class these works in the mean-variance category since the approach taken involves maximising the certainty equivalent income.

seed and soybean share the same demand patterns and are highly correlated, taking a short position in soybean futures would provide a hedge to a long position in cottonseed. Composite hedging is when a single position is used to hedge several other positions at the same time. There is a limited amount of literature on this, largely stemming from Herbst and Marshall (1994). For commodity spread positions, hedging strategies require using multiple futures contracts. The hedged portfolio in this case, consists of futures positions from each commodity in the spread. When calculating the hedge ratio of a portfolio of commodities, one takes account of the cross market linkages between them. When the markets are highly linked, it is inherent that the commodities are acting as composite and cross hedges to each other. For example, when hedging a crack spread, natural gas futures in the portfolio also provide a hedge to the crude oil spot position as well as crude oil futures.

To enhance the hedging effectiveness via cross hedging, the investor could add futures from a different market to the hedged portfolio, some examples of cross market hedges include: Witt et al. (1987), Miller (1985), Bennet (1990) , Rahman et al. (2001) and Tanlapco et al. (2002). Given that most of these works use data which are now outdated, most of the inter-market linkages (or lack of) established here cannot be applied. A possible further expansion to our research could be to identify inter-market linkages using the most up to date data for cross hedging the processing spread.

To implement composite hedging, the investor could try to reduce the number of assets in the portfolio by taking advantage of the high correlation between the assets.[3] For example, both gasoline and heating oil positions in the crack spread can be hedged by takings positions in say gasoline futures only. A test could be carried out to determine the extent of which a larger hedged portfolio could provide a better hedge than a smaller one if at all. A portfolio with a small number of assets is simpler to analyse and will also be more attractive to investors (providing the hedging effectiveness of both are on the same level).

2.2.5 Transaction Costs

A collection of different types of transaction costs from futures trades can be found in Marshall et al. (2012), these consist of:

- Spread: aka tightness, breadth, and width - see Kyle (1985) is measured as the difference between the futures traded price and the corresponding midpoint (see

[3] This only applies to the crack spread given evidences of high correlation between the commodities from Paschke and Prokopczuk (2009). There are currently no sources that confirm the same behaviour in crush spreads but this can also be investigated.

further evaluations in Dunis et al. (2008))

- Depth: the number of contracts which can be bought/sold at the current best bid-ask price
- Immediacy by trade size: whether or not an individual can execute an order immediately
- Resilience: the length of time for spread and depth to return to normal conditions following large trades which may cause liquidity to dry up

Works in the minimum-variance hedging literature include other types of transaction costs such as round trip commission costs, see Haigh and Holt (2002). The margin requirement literature includes costs in raising the margin but has not yet been implemented in evaluating trading strategies. It is also possible to model transaction costs using deterministic functions, though applications have been implemented in derivative pricing, see Leland (1985) and and more recently Pennanen and Penner (2010) for example. The same methods can be carried out for hedging transaction costs although we evaluate popular hedging strategies and do not focus on this aspect.

Transaction costs play a vital role in determining whether or not a trading strategy is worth carrying out. This is particularly important in the mean-variance framework since the transaction cost diminishes the expected return. In a minimum-variance case however, one may doubt the importance of transaction costs as the hedger is indifferent about the profitability of the portfolio. For evaluation purposes, one can assume that the investor first considers the variance reduction level from the hedging model. However, when variances of different hedged portfolios are not significantly different from one another, the investor would be indifferent about each hedging models and instead must choose the models via other criteria, such as minimising transaction costs.

2.3 Value-at-Risk Models

VaR is used for a variety of risk management methods: bank capital requirements, setting clearing house margins, investors capital allocation for portfolio management to name a few. Its most attractive feature is its simplicity whereby it is a single number which describes the worse possible loss for a fixed probability over a fixed investment horizon (although, forecasting the P&L or return distribution may be difficult). There are three main methods for estimate VaR:

- Parametric: the VaR is linearly related to the standard deviation of the distribution function

- Quantile-based: the evolution of the VaR is described using an econometric model
- Non(or semi)-parametric: the VaR is derived from the empirical distribution, often through historical simulations

We present an examination of the literature from each strand below.

2.3.1 Parametric VaR

Since Fama (1965) showed that volatility in financial markets are time varying and Engle (1982) and Bollerslev (1986) introduced GARCH models, the VaR literature has been dominated by volatility models from the GARCH family. Estimation types vary across markets, some recent examples include: Su et al. (2011) who suggests the GJR-GARCH model is superior when estimating one-day-ahead downside VaR forecasts and Chen et al. (2012) proposes that the GJR-GARCH model with Laplace distribution innovations would perform best for the Hang Seng Index.

The VaR literature is dominated by parametric methods using the GARCH family of volatility modelling. As multivariate GARCH modelling can be extended to estimating term structure VaR, we evaluate the literature on multivariate volatility models here also. The most widely-implemented multivariate GARCH model is Engle and Kroner (1995)'s BEKK GARCH, also prevalent in the portfolio management and minimum-variance hedging literature, which details the specification for the conditional covariance matrix estimate (see Grier et al. (2004), Kawakatsu (2006) for examples of the model extensions and Bauwens et al. (2006) for a survey of multivariate GARCH models). Implementing such a model for futures term structures however would require estimation of a large number of parameters which can undermine the stability of the VaR estimate.

As term structures are highly correlated systems which share common risk factors, Principal Component Analysis (PCA) can be applied to isolate orthogonal movements, reducing the number of the parameters to be estimated, thus reducing estimation error in the process. See Alexander (2001) for estimation procedures on interest rate and crude oil futures term structures and Van der Weide (2002) for cross-equity indices systems - though the latter is more suitable for smaller, less-correlated systems. With changing market circumstances to date, and when considering more convex commodity term structures such those of natural gas, it is possible that adaptation of these models can enhance estimation power of the VaR.

The literature for VaR on futures term structures is scarce, all of which utilises parametric methods for estimating VaR. To our knowledge, there are only three works which vaguely mention VaR on a term structural level: Tolmasky and Hindanov (2002) includes estimations of VaR for crack spread futures, although parameters are calibrated under the risk-neutral measure which does not entail that the model is still accurate under the physical measure; more recently Bauwens et al. (2013) examines volatility modelling for electricity term structures and briefly mentions the possibility of applying the model to calculate VaR, although they do not include any formal backtests in the work. The only work to have thoroughly tackled VaR on a term structural level is Nomikos and Pouliasis (2014). The work provides elaborate details on energy futures term structure dynamics and introduces a model for modelling the term structure. The proposed model is a development of Tolmasky and Hindanov (2002) to include Markov regime-switching volatilities. Several analyses are presented, however we concentrate on the final part of the work with reference to applications for calculating VaR as this is the only part relevant to our study. Nomikos and Pouliasis (2014) assume that an investor takes a set of weights (both long and short, generated at random) along the term structure of energy spreads and holds this constant. The model is calibrated via a maximum likelihood criterion via numerical optimization methods. Although the individual futures are assumed to follow a term structural GARCH process with normally distributed return innovations, their portfolios may not have the same type of innovations hence, a series of portfolio distribution forecasts is generated via filtered historical simulation. The series of VaR estimates is extracted from this and is backtested via Kupiec (1995)'s unconditional coverage method and the quantile loss method -introduced in the paper- which measures how accurately the tail of the distribution is forecasted. This is repeated 3,000 times, each with a different set of portfolio weights and compared to a benchmark - the Dynamic Conditional Correlation (DCC) model of Engle (2004) - again, quantiles are estimated via filtered historical simulation.

The proposed model performs well on average in comparison to the DCC model. However, we find a number of factors in the study which indicates that their model may not be appropriate for margin requirements. First, there is no mention of how well the model performs on a conditional level. As exceedances may cluster, VaR may not be reactive to changing market conditions leading to biased estimates. Second, when clearing houses issue margin requirements, it is required that all futures along the term structure must contain such estimates. The work focuses on portfolio management purposes only. Lastly, as previously mentioned, backtesting for margin models must be carried out at both long and short tails simultaneously as the clearing house is exposed to both at the same time which is not carried out here. The study also lacks detailed information on how models' performance may vary through the term structure and only provides average test statistics. This may be biased as performance may vary across the term structure and with different portfolio weights. Insight of such characteristics is essential as the margin is directly effected by the shape of the term structure of VaR estimates. The study's time period is out of date, with data ending on the 30^{th} December 2009. Further testing is required to determine the best VaR model for margin requirement purposes.

Although the recent VaR literature for commodity derivatives may vary in GARCH specification, error term innovations and calibration methods, they all share the same view -that one must account for the negative skewness and leptokurtosis in the return or P&L distribution to forecast VaR accurately (see Fan et al. (2008), Huang et al. (2009), Aloui and Mabrouk (2010) and Hammoudeh et al. (2011) for example). This behaviour is not only prevalent in the first-to-mature series but extends throughout the term structure hence a natural extension to Alexander (2001) would be to include similar model specifications when modelling principal component volatilities.

2.3.2 Quantile-based VaR

The most prolific quantile-based VaR method is Engle and Manganelli (2004), where the VaR is described as a conditional quantile. Parameters are estimated via quantile regression and the properties of their CAViaR model parameters are derived in the same work. The literature in support of quantile-based VaR is less extensive compared to the parametric counterpart. Bao et al. (2006) for example, show support of the CAViaR model in equity indices using White (2000)s Reality Check for data snooping bias. For other works which also support this model, see Huang et al. (2009) on crude oil futures and Huang (2010) for applications to equity indices. For a term structure, one may require VaR on a multivariate setting, to which Embrechts and Puccetti (2006) and Cousin and Bernardino (2013) have developed methods for such a task. The formulations however are highly complex and are likely to yield inaccurate forecasts in an out-of-sample basis.

2.3.3 Semi and Non-Parametric VaR

Unlike its parametric counterpart, non-parametric VaR methods do not assume any probability distribution function for the returns. Instead, they focus on the empirical distribution derived from historical returns whose shape is often more elaborate: negatively skewed and leptokurtic.

The simplest methods for estimating non-parametric VaR is by interpolating the relevant quantile from the empirical distribution of historical returns. The problem here is, for a VaR estimate which is deep in the tail, one requires a large window of historical data. This takes data which are too far in the past and may not reflect current market conditions. A number of techniques can be employed to side-step this issue: exponential weighting, volatility weighting, fitting kernel densities to data with shorter window lengths, etc..

The most prevalent method in this category is Barone-Adesi et al. (2002)'s Filtered Historical Simulation (FHS) model which combines time series modelling with historical simulation (hence the term *semi-parametric* VaR). Here, historical returns are described using a time series model (e.g. ARGARCH), a forecast of the distribution of the error term is generated via bootstrapping. This is then multiplied by the current volatility forecast to generate the empirical distribution forecast. This method retains 1) the elaborate shape of the empirical distribution via bootstrapping and sensitivity to market condition via conditional volatility models. FHS has proved popular in a number of works including one in the commodity VaR literature, see Kuester et al. (2006) for example.

However, we find that this method may yet be unrefined. When applying for example a GARCH filter, one calibrates the parameters using a maximum likelihood criterion. Here, one must assume some parametric distribution for the innovation of the error term, such as the normal distribution. In this case, to assume the empirical distribution takes on any shape other than the normal distribution would not be consistent with the calibration process. Further the works in support of this model uses not advanced backtested methodology and may have retained their favourable findings due to data snooping.

2.3.4 Backtesting Methodologies

Although backtesting methodologies have developed fruitfully over the past decade, the majority of the studies aforementioned are somewhat stagnant when it comes to evaluation methods; they choose to persist with traditional procedures from Kupiec (1995) and Christoffersen (1998). The conditional test from Christoffersen (1998) however only examine independence up to order 1 and is not particularly powerful when evaluating VaR which produces few exceedances. Other methods developed in Christoffersen and Pelletier (2004) based on duration between exceedances and Engle and Manganelli (2004) based on conditional quantiles are hence preferred and already implemented in some studies, see Diamandis et al. (2011) for example.

It is clear that different VaR models are suitable for different tasks. Short position holders are interested in the upper tail of futures returns, while long position holders are interested in the lower tail. Some works have already taken this into account, see Giot and Laurent (2003) for example. Clearing houses on the other hand, require accurate VaR forecasts for *both* tails as they are exposed to the short and long positions simultaneously. Though we find no works which address the need for a two-tail risk estimate for the margin requirements problem. One must also consider the appropriate level of protection as well. Although most regulations on VaR are based on the 99% level, see the 1996 Basel I amendment which requires banks to base their capital on a 10-day 99% VaR on market risk exposures and EMIR regulations require margin requirements to cover at least a 2-day, 99% VaR for exchange-traded derivatives for example. It is essential to backtest VaR models according to these different requirements to ensure that the VaR is generalisable to several risk management methods, a note often ignored in the literature.

2.4 Margin Requirements

There are two main schools of thought for setting margin requirements: the prudential approach of Figlewski (1984b) and Gay et al. (1986) which argues that the main purpose of margins is to cover the clearing house's loss when participants default; and the effcient contract design of Brennan (1986), which examines how margins and price limits can be set to make the contracts self-enforcing.[4]

These pioneering works have lead extant literature to conclude that the optimal margin level should be: (1) high enough to cover the default risk faced by the clearing house when taking on defaulters' positions; (2) low enough to limit investors' opportunity costs and maintain liquidity in the market; and (3) stable enough to reduce investors' additional opportunity costs when margin changes. We find however, that the stable-margin problem is poorly addressed in the academic literature. To this end, our work introduces a model for a margin level that is stable yet also an accurate reflection of dynamic market volatility. We ensure an optimal balance of such criteria by calibrating the model in-sample and then employing fuzzy goal programming to allow for stability out-of-sample.

[4] Day and Lewis (2004) suggests that margin calls can be hedged perfectly using binary options. They derive a no-arbitrage relationship between the two and suggest how historical binary option prices can be used to calculate a no-arbitrage margin level. however, this method is not yet applicable to many markets as binary options are not always liquid.

2.4.1 Prudential Margin Requirements

The prudentiality argument requires that margins must cover as many of the clearing house's losses as possible in an event of a client's default. The question here lies in the appropriate percentile one should set the margin at, i.e. how much of the loss distribution should the margin cover? Most studies simply set the margin equal to a quantile of the return distribution. We find a substantial amount of literature which belong to this school of thought and follow Broussard and Booth (1998)'s view that, given clients tend to default on extreme returns, margins should be high enough to protect the exchange against movements that are described using Generalised Extreme Value (GEV) distributions. Methods for fitting GEV can be found in Longin (1999) who analyses COMEX silver futures return distributions between 1975-1994. A return series is segmented into equal sections, minimum and maximum values of each window are extracted and categorised as extreme losses for long and short positions respectively. The GEV is fitted to this sample via least squares methods and the margin is then set as a quantile of the GEV distribution. This simple margin rule can then be evaluated via coverage tests from Christoffersen (1998), see Chiu et al. (2006) for example. Alternatively, margins can be evaluated by examining the historical its coverage level, see Booth et al. (1997) and Cotter (2001) for GEV distributions. The latter method is also used to evaluate margins which are set using different risk measures all together. Ma et al. (1993) suggests that margins should reflect the aggregate utility of the clearing members since distribution forecasts are subject to risk tolerance levels. The margin levels are evaluated via a test for difference in confidence interval to examine if one is significantly larger than the other. Similarly, Cotter and Dowd (2006) suggests using Spectral Risk Measures which too takes the exchange's risk attitudes into consideration.

GEV distributions are also useful in setting circuit breakers, where trading is halted should the futures price movement breach a predetermined level, see Longin (1999) and Broussard (2001) for example. The risk measures are not backtested in these tests because they are purely subjective to the clearing house's outlook on the distribution. We do not encourage the use of a utility as above because the exchange is not the only player affected by the margin requirements. The clearing house's loss at clientale default diffuses further to the risk-sharing pool which are collected from other investors and clearing members as well. Regulators who wish to protect the liquidity of the financial markets should also ensure that a margin provides an adequate level of protection of clearing house losses during client's default. Not only this, even the utility of other exchanges in competition can also be a significant factor of the margin level. To achieve a realistic model, margin requirements must take account of the interaction between the utilities of all players in the market and not just the exchange itself. This, however, requires an elaborate equilibrium model which provides an interesting case for further investigation.

2.4.2 Efficient Contract Design and Opportunity Cost

Initiated by Brennan (1986), the framework focuses on setting margins and price limits to minimise the probability of default by investors of a futures contract, i.e. when there is no costly legal action required by each counterparty in the event of a default. Here, when entering an agreement, the counterparties observe the market futures price $F_{t,T}$ and put up the initial margin M_T , to the central counterparty. The futures contract is to be signed at time $t+1$ but either investor can choose to renege should their total loss, i.e. $F_{t+1,T} - F_{t-T} - M_T$ be greater than the legal costs. Now let the movements be subject to the price limit P_T, if $F_{t+1,T} - F_{t,T}$ exceeds P_T and activates the circuit breaker, the counterparties will have no choice but to enter into the agreement and risk gaining/losing $F_{t+1,T} - F_{t,T}$. Brennan (1986) concluded that the absolute minimum level of the seller's expected loss, assuming futures follow a martingale, with no incentive to default can be described via the relationship

$$E[F_{t+1,T} - F_{t,T}|F_{t+1,T} - F_{t,T} \geq P_T] = M_T \ . \tag{2.6}$$

In other words, the margin is the expected tail loss bounded by P_T . Brennan (1986) imposes a further restriction, which is to set P_T and M_T to reduce investors transaction costs evaluated as the sum of: the opportunity cost of raising the initial margin, OM_T ; the cost of price limits which can halt trading and decrease liquidity, CL_T and the probability of renege Pr_T defined as

$$OM_T = \kappa M_T \ , \tag{2.7}$$

$$CL_T = \frac{Pr(|\Delta F_{t,T}| > P_T)}{Pr(|\Delta F_{t,T}| < P_T)} \ , \tag{2.8}$$

$$Pr_T = 2\beta \mathcal{F}(\Delta F_{t,T} > P_T) \ , \tag{2.9}$$

where $\mathcal{F}(\cdot)$ denotes the probability distribution function for futures prices. Brennan (1986) also provides closed-form solutions for the margins and price limits under normally-distributed and uniformly-distributed futures returns.

Further developments of the model include: Fenn and Kupiec (1993) who incorporate settlement frequency; Shanker and Balakrishnan (2005) who extend the model to differentiate between short and long positions and; Chou et al. (2006) who include price limits from the underlying as well as the futures. These works however do not address how Brennan (1986)'s assumptions have changed over time, let alone accounting for impending regulatory changes. Electronic platforms allow traders to enter futures contracts immediately without signing. Circuit breakers do not stop trading for an entire day; historically, trading halts have lasted as little as five minutes leaving scarce time to renege. Although a natural extension to the framework would be to relax the normally-distributed-return assumptions and impose more flexible return evolutions (such as stochastic volatility models), we do not concentrate on expanding this literature as it requires major reformulation to adapt to changing clearing house mechanics.

Yet, the intuitions here are still applicable. For example, given investors would only default on margin calls, initial/maintenance margin ratios and price limits can be set to avoid successive margin calls to maintain market stability. Some works in progress have implemented this already, see Huang et al. (2011).

2.4.3 Rules for Margin Requirements

Some of the literature have suggested heuristic rules to help stabilise margins. Chiu et al. (2006)[5] imposes the tier-adjustment rule to Hang-Seng Index Futures, where the margin $M_{t,T}$ interacts with the $VaR_{t-1,0.997,1,T}$ via the relationship:

$$M_{t,T} = \begin{cases} M_{t-1,T}, \text{for} 0.85 M_{t-1,T} < VaR_{t-1,0.997,1,T} < 1.15 M_{t-1,T} & \\ VaR^{*}_{t-1,0:997,1,T} & \text{otherwise} , \end{cases} \tag{2.10}$$

where $VaR_{t-1,0.997,1,T}$ is the $VaR_{t-1,0.997,1,T}$ rounded to the nearest 10,000 NTD. They find that the approach dampens the accuracy of the margin coverage level and hence do not encourage the use of such rule. More recently, Lam et al. (2010)'s rules entail that should percentile forecast fall outside a certain range i.e. the margin band b where $M_{t,T}$ follows:

$$M_{t,T} = \begin{cases} M_{t-1,T} & \text{for} F_{t-1,T} k \hat{\sigma}_{t,T} - b < F_{t-1,T} k \hat{\sigma}_{t,T} < F_{t-1,T} k \hat{\sigma}_{t,T} + b \\ F_{t-1,T} k \hat{\sigma}_{t,T} & \text{otherwise} , \end{cases} \tag{2.11}$$

[5] Also the only work we have come across to illustrate such margin levels graphically

where $\hat{\sigma}_{t,T}$ denotes the estimate of the volatility at time t, $F_{t-1,T}$ is the futures price and k and b are parameters to be optimised by minimising the overcharge rate and maximising the margin level. Note that k and b do not change with T as Lam et al. (2010) applies their model only on the first-to-mature series. The optimisation is carried out in-sample on Hang Seng Index futures between 1996 and 2006. We find some ineffcient methods from this study. First, the rule itself produces very large margin changes; this does not agree with impending regulations which demand small margin step sizes to limit procyclicality. Second, Lam et al. (2010) selects the volatility models in view of minimising the overcharge rates and maximising the margin level, which does not indicate that the volatility is representative of the market condition. Lastly, the study does not demonstrate how the rule can be executed out-of-sample, hence it may not be applicable in practice. Moreover, the studies above do not describe how margin requirements can change as futures contracts time to maturity diminishes. Since futures return VaR increases with decreasing time to maturity, so too should the margins. Yang and Yan (2008) is the closest study to address this problem, they examine margins for calendar spreads which are set equal to their GARCH volatilities but does not elaborate on how such a model can be impractical.

2.4.4 Standard Portfolio Analysis of Risk (SPAN)

Since its introduction in 1988 by the CME, SPAN has developed into a complex tool for setting margin requirements as well as for general risk management purposes. It is widely used in a number of exchanges, including both the CME group and ICE, with varying methodologies across different products. The programme generates 16 possible price movement scenarios; the scenario with the greatest loss is taken as the margin. The methods are based on risk measures (VaR, ETL for example), which again, vary among different products. The most detailed technical document, published by the CME group, indicate that margins are calculated using 4 different VaR estimation technique including: normal-mixture, EVT, EWMA and implied volatility where the margins are set at least to cover 99% of the price movements. There are however no specifications to which VaR model should be implemented on which product or the window size of historical data needed for the computation. For ICE however, SPAN estimates margins via heuristic rules based on the historical futures P&L series.

The methodologies for risk estimation has also evolved significantly over time, it is therefore inherently difficult to test SPAN as parameters are time-varying. CME's technical documents issued in 2010 for example impose that the margin coverage level are between the 95-99% VaR while current documents state that all margins cover more than the 99% VaR. The time at which such an estimate come into play is not mentioned in any documents. Given the lack of clarity in the methodologies, SPAN is difficult to recreate, see Kupiec (1994) and Kupiec and White (1996) who mimics and test SPAN against regulation T (a strategy-based margining system for setting margins on equities and equity options) and find that SPAN provides lower margins on average and also covers the necessary amount of daily return movements. More recently, Abruzzo and Park (2013) provide detailed accounts on how exchanges have changed futures margins in the past and find that the strain of competition forces the CME and ICE to alter margins to out price each other. Also, given parameters and VaR estimation models for the margins have changed so frequently in the past, we conclude that SPAN margin changes are heuristic by nature and make no attempt to recreate such a system econometrically. Our examination concentrates on using historical margin levels obtained from `http://www.cmegroup.com/clearing/risk-management/historical-margins.html` and `https://www.theice.com/clear_europe_span.jhtml` instead. Abruzzo and Park (2013) also find that the CME's decreases in margin requirements are more cautious than increases. This is surprising given the race-to-bottom attitude and chances to boost liquidity (see Hardouvelis and Theodossiou (2002)) should encourage central counterparties to be prompt when decreasing margins.

2.4.5 Evaluating the Margin Model

The measurement of the overcharge/margin level optimality was introduced by Lam et al. (2004). The prudentiality index for the margin M_T and losses $L_{t,T}$ is measured using the expected shortfall $ESF_{t,T} = -E[(L_{t,T} - M_T)^+]$ while the expected overcharge $EOC_{t,T}$ is measured as the average losses which do not exceed the margin: $EOC_{t,T} = E[(M_T - L_{t,T})^+]$. Combining the two criteria yields a straightforward result to minimise

$$EOC_{t,T} - ESF_{t,T} = E[|M_T - L_{t,T}|] \ . \tag{2.12}$$

Although not mentioned in Lam et al. (2004), it is intuitive that the first-order-optimal margin level is $M_T = E[L_{t,T}]$ (i.e. equivalent to minimising $E[(M_T - L_{t,T})^2]$); the margin should be at the mean of the losses to satisfy both conditions. The lower partial moment however, can be difficult to estimate, hence Lam et al. (2004) also introduces a function of the current futures price $F_{t-1,T}$, returns $R_{t,T}$, its mean and volatility estimates $\hat{\mu}_{t,T}$, $\hat{\sigma}_{t,T}$ and parameter $k_{t,T}$ which governs the percentile of estimate: the daily shortfall

$$SF_{t,T} = -F_{t-1,T}(|R_{t,T}| - |\hat{\mu}_{t,T} + k_{t,T}\hat{\sigma}_{t,T}|)^+ \ , \tag{2.13}$$

and the daily overcharge

$$OC_{t,T} = F_{t-1,T}(-|R_{t,T}| + |\hat{\mu}_{t,T} + k_{t,T}\hat{\sigma}_{t,T}|)^+ \ . \tag{2.14}$$

The criteria are then optimised to find the the optimal value for $k_{t,T}$ which, in turn, can be used to calculate the optimal margin level

$$M^*_{t,T} = F_{t-1,T}(\hat{\mu}_{t,T} + k_{t,T}\hat{\sigma}_{t,T}) \ , \tag{2.15}$$

assuming $\hat{\mu}_{t,T} + k_{t,T}\hat{\sigma}_{t,T} > 0$. Under Lam et al. (2004)'s redefinition, $M^*_{t,T}$ is dynamic which is inconsistent with minimising relationship 2.12. The equivalent in-sample (constant) margin estimate can be found using the relationship $M_T = E[F_{t-1,T}(\hat{\mu}_{t,T} + k_{t,T}\hat{\sigma}_{t,T})]$. Although the criterion is now on the absolute returns as opposed to the losses, consistent with the first criterion, the optimal percentile level can be calculated by minimising

$$OC^*_{t,T} + SF^*_{t,T} = |F_{t-1,T}(\hat{\mu}_{t,T} + k_{t,T}\hat{\sigma}_{t,T} - |R_{t,T}|)| \ , \tag{2.16}$$

with respect to $k_{t,T}$. Again, the solution is analogous to minimising $F^2_{t,T}(\hat{\mu}_{t,T} + k_{t,T}\hat{\sigma}_{t,T} - |R_{t,T}|)^2$, that is $k_{t,T} = \frac{|R_{t,T}|-\hat{\mu}_{t,T}}{\hat{\sigma}_{t,T}}$. Substituting this into 2.15, we obtain the margin estimate $M^*_{t,T} = F_{t-1,T}|R_{t,T}|$, with the in-sample (constant) margin $M_T = E[|\Delta F_{t,T}|]$. In other words, the optimal $k_{t,T}$ is simply the percentile where the first moment of absolute profit and losses can be found, which in the case of a normal distribution is roughly at the 80th percentile. Given forthcoming EMIR regulations require the margin to cover at least 99% of the distribution, we find that the lower/upper partial moment may be too low, even for a highly leptokurtic distribution. Indeed, should we subject the optimality criteria of Lam et al. (2004) or Lam et al. (2010) to the prudential evaluation methods of Booth et al. (1997), we would find that the margin coverage level is inadequate. Hence we do not use these criteria in our study. The margin model is evaluated in two folds: backtesting of the VaR model to ensure the accuracy of the percentile estimates; and examining whether or not the out-of-sample margin requirements produce the optimal trade-offs between the criteria previously outlined.

We find no works which contain methods of measuring margin stability. This is essential when comparing different margin models.

2.5 Summary

In this literature review, due to storage costs and convenience yields, energy futures returns and volatility are seasonal. Unlike other easy-to-store commodities such as precious metals, energy futures term structures are more curved. These behaviour are thoroughly analysed in the literature and there exists a large number of models which can be applied in our work. We find though, that the majority of the literature has not taken care when examining how these models can be used in practice, i.e. they fail to account for parameter recalibration frequency and transactions. Hence, rather than focusing on improving the models to account for further deterministic movements, we first examine how the most popular models in the literature, i.e. GARCH and EWMA can be used in practice.

We choose to concentrate on evaluating the four main gaps in the literature: first, we evaluate minimum-variance hedging for a refinery, paying particular attention to transaction costs and whether or not variance reduction from more complex models are significantly different from simpler models. Second, we extend previous works on VaR by extending backtesting procedures to test both long and short position VaR simultaneously on the rest of the term structure. Lastly, we re-define previous optimality criteria for futures margining to incorporate stability following recent regulatory changes and we derive rules to allow under such criteria to determine the best margin requirement method. In this work, we also introduce a method for assessing margin stability.

3. (De)Merits of Minimum-Variance Hedging: Application to the Crack Spread

3.1 Introduction

There exists a substantial literature on minimum-variance hedging of spot positions using futures contracts in which sophisticated econometric models are applied for estimating the hedge ratios. The majority of these studies conclude that advanced econometric tools improve the hedging performance over the naïve hedging strategy of shorting one futures contract per unit of spot exposure. However, majority of studies ignore margin and transaction costs, and/or does not evaluate the improvement in a statistically meaningful way. Even, in some cases, insufficient care is taken to pre-filter the data for use in the analysis. Our contribution is to conduct an extensive out-of-sample study of minimum-variance hedging for a complex underlying position, with meticulous processing of the relevant data. We compare several popular hedging approaches and covariance estimation techniques with the simple naïve hedge, explicitly taking margin and transaction costs into account. In contrast to the majority of extant literature we find that none of the sophisticated methods are able to outperform the naïve hedge. Furthermore, our discussion uncovers how variable margin levels could be detrimental for hedgers in the market and that they would benefit from moving to an exchange whose margins are stable.

Minimum-variance hedging has been pioneered by Johnson (1960) and Stein (1961), and further refined by Ederington (1979), Hill and Schneeweis (1982), Figlewski (1984a) and Herbst et al. (1989) amongst many others. Since Fama (1965) found that asset covariance structures are time-varying and Bollerslev (1986) introduced the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) method of estimating conditional variance the application of the GARCH family for estimating hedge ratios has been rapidly growing in popularity. Baillie and Myers (1991) first derived hedge ratios using the bivariate GARCH model. Kroner and Sultan (1993) utilise the CCC GARCH model in the foreign-exchange market and Gagnon et al. (1998) expand the study for multi-asset portfolios using the BEKK GARCH model. Haigh and Holt (2000) and Haigh and Holt (2002) analyse hedging in the freight and crack spread markets using a modified BEKK GARCH model. Further work on GARCH-based hedging includes Lee and Yorder (2007), Lien (2008), Lee (2009), Lee (2010), Chang et al. (2011), and Ji and Fan (2011). All these works conclude that a GARCH-based strategy is superior to other static hedges.

Supporters of GARCH hedge ratios argue in unison, that the implementation of GARCH is necessary in order to capture the time-varying asset covariance structure. This should allow GARCH-based minimum-variance hedging to provide greater variance reduction than naïve hedging. However, due to uncertainty in the GARCH process specification and in its parameter estimates, this may not be the case in practice. Moreover, typically, the hedge ratios derived from GARCH-type models are extremely volatile, suggesting unrealistically frequent re-balancing and hence very large transaction costs for the hedged portfolio.

Most previous papers utilise weekly log returns in the analysis, but log returns are not realised and, for assets with prices that can jump, log returns can be highly inaccurate proxies for percentage returns even when measured at the daily frequency. Additionally, since the hedged portfolio can have zero value, even its percentage return may be undefined. Thus, our hedging analysis is based on profit and loss (P&L) rather than on log or percentage. Also, in most previous papers the estimation of GARCH and OLS parameters is based on a very large sample size. This choice can bias results towards the GARCH approach because OLS regression attributes equal weight to all observations, including outdated information at the beginning of the sample. Moreover, it will typically result in a relatively small out-of-sample period, which consequently yields test results having relatively large standard errors.

However, the most important difference between our methodology and that employed by many of the papers cited above is the use of constant-maturity versus rollover futures series. Recently, Nguyen et al. (2011) have highlighted the pitfalls of using futures series which simply roll over into the next contract as expiry approaches. This practice creates a saw-tooth pattern in the basis which has the unintended effect of biasing the OLS minimum-variance hedge ratios. To avoid this bias, our analysis is based on constant-maturity futures P&Ls.

We base our study on the problem of hedging crack spread positions. Although one might argue that most of the previous literature has studied the problem of hedging equity or pure commodity positions, we decided to use this underlying as it is a more complicated hedging problem, where prices are highly variable and subject to frequent jumps. As such, more advanced methods have a greater chance to improve the performance. Indeed, as mentioned above, Haigh and Holt (2002) conclude that multivariate GARCH models are superior for hedging the crack spread, although they use a mean-variance rather than a minimum-variance framework.

The rest of this chapter is structured as follows: section 3.2 describes the crack spread and hedging implications; sections 3.3 and 3.4 presents the methodological framework; section 4.3 describes the data; section 3.6 presents the results; section 3.7 concludes.

3.2 The Crack Spread

Like other processing spreads such as the crush spread and spark spread, it is heavily traded by both large scale industrial refiners and speculators. In addition to the characteristics mentioned previously, these spread positions display other unique relationships which are useful when considering hedging strategies.

The crack spread represents the profits from a simultaneous purchase and sale of crude oil and its refined products, mainly consisting of gasoline and heating oil. An $a : b : c$ crack spread is defined as buying a units of crude oil, and selling b and c units of gasoline and heating oil, respectively. These ratios are set according to the refineries production technologies.

There are many types of crack spread positions available on the NYMEX, which can be identified by the ratios between the position on crude oil and the refined products [1] i.e. going long an $i : j : k$ crack spread would refer selling i units of crude oil, buying j units of gasoline and buying k units of heating oil. The current technology allows for twice as much production of gasoline to that of heating oil, hence the most commonly traded crack spread position is the $3 : 2 : 1$ crack spread. For this reason, most works on crack spreads also concentrate on this particular ratio.

In 1994 the New York Mercantile Exchange (NYMEX), which offers the highest trading volume on oil-related futures amongst all exchanges worldwide, introduced the possibility for refineries to put up a single margin for the 3:2:1 crack spread. Thus, if refineries hedge this position as a whole using futures contracts on crude oil, gasoline and heating oil in this fixed ratio, margin costs are reduced and maintaining

[1] Often heating oil and gasoline since these are the most abundant refined products. Together, they represent almost the entire petroleum market in the US, see Tolmasky and Hindanov (2002).

the account is simplified for both parties. The popularity of this product led NYMEX to introduce single margins for any $a : b : c$ crack spread position.

The most interesting feature of the crack spread is the high level of co-integration between the commodities within the spread. This is when they share similar co-movements in the long-run. Some evidences of this can be found in Girma and Paulson (1999) and their analysis on the 3 : 2 : 1 crack spread from the NYMEX between 1983 and 1994. More recently, Paschke and Prokopczuk (2009) finds that crack spread commodities are both highly correlated and co-integrated and included this behaviour in their pricing model. In this particular case, more effective hedging strategies need to be implemented. For highly correlated series, one could be used as a cross hedge to the other. This behaviour however, only reflects how two variables move together in the short-run. Their movements in the long run may appear to be much less synchronised therefore only short-term hedging strategies are appropriate for such assets.

For long-term hedges, the investor could utilise the co-integrated nature of the crack spread. There are several hedging methods which can be applied: spot-futures arbitrage, yield curve modelling, index tracking and spread trading are some of tricks that could be performed in this instance. Descriptions of the application of these can be found in Alexander (1999). Our research will mainly concentrate on short-term hedging, the literature on hedging co-integrated series are not explored here. This however provides an interesting case for further expansion of our problem.

A more recent development of in the oil market is the invention of the Reformulated Gasoline Blendstock for Oxygen Blending (RBOB gasoline) which replaced unleaded gasoline in 2005. To our knowledge, all of the existing literature on crack spread positions have concentrated on using unleaded gasoline since there is a richer data source, expanding over two decades in the NYMEX (1984-2005). Unleaded gasoline however, no longer exist in the NYMEX, an analysis on this will be less useful for investors. Our work however will mainly concentrate on using RBOB gasoline instead. This presents an opportunity for our research to be among the first to analyse the behaviour of this type of gasoline in comparison to its predecessor.

3.3 Minimum-Variance Hedging

3.3.1 Hedging Models

Let the $a:b:c$ crack spread spot and futures prices, S_t^z and F_t^z, be given by

$$S_t^z = -aS_t^c + bS_t^g + cS_t^h \ , \qquad F_t^z = -aF_t^c + bF_t^g + cF_t^h \ ,$$

where $S_t^c,\ S_t^g,\ S_t^h,\ F_t^c,\ F_t^g$ and F_t^h denote the spot and futures prices for crude oil, gasoline and heating oil, respectively. The realised hedged portfolio P&L, $\Delta\Pi_t = \Pi_t - \Pi_{t-1}$, is given by

$$\Delta\Pi_t = \Delta S_t^z + a\beta^c \Delta F_t^c - b\beta^g \Delta F_t^g - c\beta^h \Delta F_t^h \ , \tag{3.1}$$

where $\beta^c,\ \beta^g,\ \beta^h$ are the hedge ratios. For the naïve hedge, $\beta^c = \beta^g = \beta^h = 1$.

The hedge ratios that minimise the variance of (3.1) can be obtained by solving the first-order conditions

$$\begin{bmatrix} a\beta^c \\ -b\beta^g \\ -c\beta^h \end{bmatrix} = \begin{bmatrix} a^2\sigma_{\Delta F_t^c \Delta F_t^c} & -ab\sigma_{\Delta F_t^c \Delta F_t^g} & -ac\sigma_{\Delta F_t^c \Delta F_t^h} \\ -ab\sigma_{\Delta F_t^g \Delta F_t^c} & b^2\sigma_{\Delta F_t^g \Delta F_t^g} & bc\sigma_{\Delta F_t^g \Delta F_t^h} \\ -ac\sigma_{\Delta F_t^h \Delta F_t^c} & bc\sigma_{\Delta F_t^h \Delta F_t^g} & c^2\sigma_{\Delta F_t^h \Delta F_t^h} \end{bmatrix}^{-1} \begin{bmatrix} a\sigma_{\Delta S_t^z \Delta F_t^c} \\ -b\sigma_{\Delta S_t^z \Delta F_t^g} \\ -c\sigma_{\Delta S_t^z \Delta F_t^h} \end{bmatrix} , \tag{3.2}$$

where σ_{ij} denotes the covariance between i and j. This method is analogous to Ordinary Least Squares (OLS) regressions of the spot P&L on the hedging instrument(s). In other words, finding the hedge ratios in (3.2) is analogous to performing the single-equation, multiple-variable regression

$$\Delta S_t^z = \alpha - a\beta^c \Delta F_t^c + b\beta^g \Delta F_t^g + c\beta^h \Delta F_t^h + \epsilon_t \ , \tag{3.3}$$

where ϵ_t denotes the regression residuals. We refer to this hedging model as the *single-equation, multiple-variable model.* Since each commodity is exposed to closely related risk factors, it is expected that multicollinearity is present in this setting. Consequently, hedge ratios derived from (3.3) may have biased standard errors yielding to imprecise hedge ratio estimates.

Alternatively, one might employ a *multiple-equation, single-variable model* in which the hedge ratios are estimated via three asset-by-asset regressions, as follows:

$$\begin{bmatrix} a\Delta S_t^c \\ b\Delta S_t^g \\ c\Delta S_t^h \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} a\beta^c\Delta F_t^c \\ b\beta^g\Delta F_t^g \\ c\beta^h\Delta F_t^h \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix} .$$

This model does not account for the covariances between the futures P&Ls at all. In other words, the hedge ratio calculation is the same as in (3.2), but the off-diagonal elements in the square matrix are assumed to be zero.[2]

The first model requires estimates of several covariances and variances and each are prone to estimation errors. In contrast, the second model assumes all futures covariances to be zero. A third possibility is to simply impose the constraint $\beta^z := \beta^c = \beta^g = \beta^h$, whereby one obtains the parsimonious *single-equation, single-variable model* for which the estimation errors might be significantly reduced. This model, which is nested in (3.3), is given by

$$\Delta S_t^z = \alpha + \beta^z \Delta F_t^z + \epsilon_t \ , \tag{3.4}$$

with optimal hedge ratio given by

$$\beta^z = \frac{\sigma_{\Delta S_t^z \Delta F_t^z}}{\sigma^2_{\Delta F_t^z}}.$$

The price to pay for a parsimonious model is the implicit assumption of constant correlations between the futures P&Ls. This model has not previously been considered in the literature, but when the correlations between the components of a multiple hedge portfolio are high, then so are the estimation errors in the covariances of the futures P&Ls in (3.3). Hence, one might expect a superior performance from the *single-equation, single-variable model* despite its restrictive assumptions on correlation.

3.3.2 Estimation Method

We now turn to the econometric methods used to estimate the variances and covariances in the hedging models. We employ four different popular estimation methods: OLS; exponentially weighted moving averages (EWMA); the standard symmetric GARCH; and an asymmetric GARCH model. To conduct an out-of-sample study we

[2] We also estimated hedge ratios using generalised least squares (GLS) in a seemingly unrelated regression equations (SURE) system for this model. As the hedging effectiveness results were indistinguishable; we do not report them.

re-estimate all parameters of the OLS and GARCH models using a rolling window of length n. The parameter of the EWMA model is fixed, a priori.

With OLS, variances and covariances of two assets Y_1 and Y_2 are simply estimated by their sample counterparts

$$\hat{\sigma}^2_{\Delta Y_1,t} = \frac{1}{n-1}\sum_{i=0}^{n}(\Delta Y_{1,t-i} - \Delta\bar{Y}_{1,t})^2 \ ,$$

and

$$\hat{\sigma}_{\Delta Y_1 \Delta Y_2,t} = \frac{1}{n-1}\sum_{i=0}^{n}(\Delta Y_{1,t-i} - \Delta\bar{Y}_{1t})(\Delta Y_{2,t-i} - \Delta\bar{Y}_{2,t}) \ ,$$

respectively. EWMA variances and covariances are estimated via the recursions

$$\hat{\sigma}^2_{\Delta Y_1,t} = (1-\lambda)\Delta Y^2_{1,t-1} + \lambda\hat{\sigma}^2_{\Delta Y_1,t-1} \ ,$$

and

$$\hat{\sigma}_{\Delta Y_1 \Delta Y_2,t} = (1-\lambda)\Delta Y_{1,t-1}\Delta Y_{2,t-1} + \lambda\hat{\sigma}_{\Delta Y_1 \Delta Y_2,t-1} \ ,$$

where λ is the EWMA decay coefficient which takes a value between 0 and 1. With a lower λ more emphasis is placed on the most recent observations and the model hence becomes more reactive to changing market conditions.

GARCH variances and covariances are obtained using the BEKK model specification of Engle and Kroner (1995). For a vector of zero mean P&Ls $\mathbf{\Delta Y}_t$, the multivariate GARCH covariance matrix estimate $\mathbf{H}_t$ is based on the dynamics

$$\mathbf{H}_t = \mathbf{A}'\mathbf{A} + (\mathbf{B}'\mathbf{\Delta Y}_{t-1})(\mathbf{B}'\mathbf{\Delta Y}_{t-1})' + \mathbf{C}'\mathbf{H}_{t-1}\mathbf{C} \ ,$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are $m \times m$ matrices of the BEKK parameters for m assets. The parameter estimates are obtained by maximising the log-likelihood function[3]

$$\ln L(\theta) = -\frac{1}{2}\sum_{t=1}^{n}(\ln(|\mathbf{H}_t|) + \mathbf{\Delta Y}_t'\mathbf{H}_t^{-1}\mathbf{\Delta Y}_t) \ .$$

As it is well known that the symmetric GARCH specification can be improved by allowing for an asymmetric variance response to shocks we also employ the asymmetric GARCH BEKK specification (AGARCH) of Grier et al. (2004). Here the variances

[3]We use Kevin Sheppard's UCSD GARCH toolbox for the estimation, available at `http://www.kevinsheppard.com/wiki/UCSD_GARCH`.

dynamics are specified as

$$\hat{\mathbf{H}}_t = \mathbf{A}'\mathbf{A} + (\mathbf{B}'\boldsymbol{\Delta}\mathbf{Y}_{t-1})(\mathbf{B}'\boldsymbol{\Delta}\mathbf{Y}_{t-1})' + \mathbf{C}'\hat{\mathbf{H}}_{t-1}\mathbf{C} + (\mathbf{D}'\boldsymbol{\Delta}\mathbf{Y}^*_{t-1})(\mathbf{D}'\boldsymbol{\Delta}\mathbf{Y}^*_{t-1})' \ ,$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ are $m \times m$ matrices of the asymmetric BEKK parameters for m assets and $\mathbf{Y}^*_t$ is a vector of $\max\{Y_t, 0\}$ for a positively skewed sample or $\min\{Y_t, 0\}$ for a negatively skewed sample.

For ease of presentation, we abbreviate the hedging models and estimation techniques by $model_{ij}$ where $model$ denotes the estimation method, i.e. $model = \{OLS, EWMA, GARCH, AGARCH\}$, and $i = 1, 3$ and $j = 1, 3$ denote the number of equations and variables in the regression system respectively. For example, EWMA_{13} refers to the single-equation, multiple-variable model as specified in (3.3) where the variances and the covariances are estimated using the EWMA method, etc.

In total, seven hedging models are analysed: naïve, OLS_{31}, OLS_{13}, OLS_{11}, EWMA_{11}, GARCH_{11} and AGARCH_{11}. For the EWMA, GARCH and AGARCH estimation methods we omit results for multiple-equation or multiple-variable models because preliminary results, based only on the OLS models, show that the three regression configurations are more or less equally effective. Moreover, the proliferation of parameters when GARCH and AGARCH models are applied to multiple-equation or multiple-variable models exacerbates the problem of parameter estimate instability, which is discussed later on with reference to Figure 3.4.

3.3.3 Transaction Costs

When trading futures on the NYMEX, transaction costs arise from the round trip commission charged by the exchange and from the bid-ask spread. Since the early 2000's, the NYMEX has reduced the round trip commission costs from \$ 15.00 to \$ 1.45 per futures contract bought and sold. Although the NYMEX is an open-outcry market (which allows limit orders) the hedger is assumed to place market orders, to prioritise the variance reduction over possible gains from trading. The bid-ask spreads of the three considered commodity futures, x, y and z are defined as the average spread between the bid and the ask price divided by the average mid price of each commodity futures. The dollar value of the bid-ask spreads, TC_t, arises from re-balancing the portfolio and is given by:

$$TC_t = aF^c_t|\Delta\beta^c_t|x + bF^g_t|\Delta\beta^g_t|y + cF^h_t|\Delta\beta^h_t|z \ .$$

The modulus signs are placed to indicate how the hedger loses the spread regardless of the direction of trade. We follow Dunis et al. (2008) and set x, y and z to be 1 bps, 10 bps and 12 bps, respectively. Although these are bid-ask spreads of first-to-mature rollover series, we assume that these are constant through the first two months of the term structure and hence use the same for our constant-maturity futures.

Margin costs arise from raising the initial margin and from marking-to-market the maintenance margins. In the past decade there have been several changes to the NYMEX margin requirement rules. When trading an $a:b:c$ crack spread, NYMEX calculates the initial margins based on the portfolio VaR. For a hedger, who shorts a crack spread expiring in 1 month, the initial margin is approximately \$ 11, \$ 18 and \$ 7 per 3:2:1, 5:3:2 and 2:1:1 crack spread bundles respectively (as opposed to \$15, \$25 and \$10 for a speculator).[4] We shall focus on the costs incurred by refineries, which are generally treated as hedgers by the clearing house. The total cost m_t^i from raising the initial margin is

$$m_t^i = |\beta_t^z| M(r_t^d - r_t^f) \ ,$$

where r_t^d is the cost of raising the initial margin, r_t^f is the risk-free rate of return gained from depositing in the margin account, M is the initial margin required per crack spread bundle and β_t^z is the number of crack spread bundles purchased. In the cases where the hedge ratios do not allow for exact transaction of the bundles (i.e. $\beta^c \neq \beta^g \neq \beta^h$), the approximation $\beta^z \approx \frac{a\beta^c + b\beta^g + c\beta^h}{a+b+c}$ is taken instead. The refinery is assumed to raise debt for the initial margin, r_t^d is set as the average cost of debt in the industry. The top ten US refineries are currently, on average, rated AA by Moody's. Hence Moody's AA bond index is chosen as a proxy for the cost of debt r_t^d. The initial margin is set at \$ 11, \$ 18 and \$ 7 for the 3:2:1, 5:3:2 and 2:1:1 bundles, respectively. These were the values quoted by NYMEX on 06/06/2011. Three-months US T-bill rates are used as a proxy for r_t^f.

The gains and losses from the maintenance margin arise from the movement in the futures prices every day. These are marked-to-market daily but as we work with weekly data, we employ a linear approximation of the daily changes in the margin account. See Figure 3.1 for an illustration of this process, the thick black line represent the linear approximation to the actual movements of the futures price from t to $t+1$ (blue line). The shaded pink area represents the total interest earned, which can be calculated as $\frac{1}{2}(F_{t+1}^z - F_t^z)((t+1) - t)$. The weekly interest on the margin account

[4] For an $a:b:c$ crack spread, a "bundle" indicates simultaneously going long a barrels of crude oil, short b barrels of gasoline and short c barrels of heating oil.

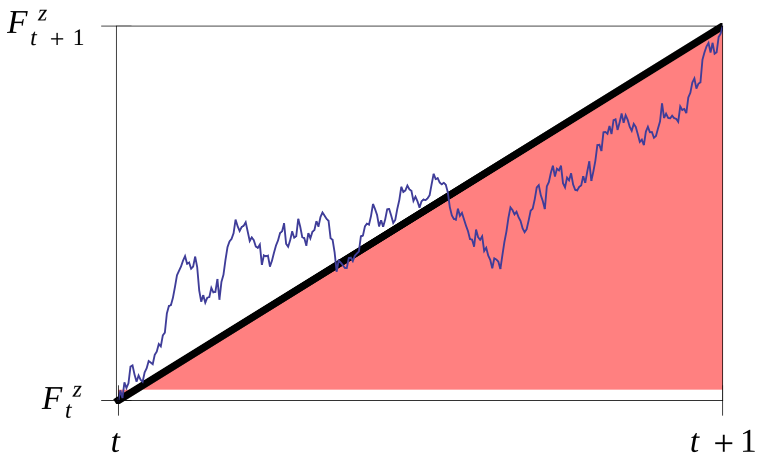

Figure 3.1.: Illustration of the one-half approximation to interest rates earned on a $P\&L$ over the period t to $t+1$. F_t^z denotes the futures price at time t.

m_t^m is therefore approximated as

$$m_t^m = \frac{1}{2}\left(-a\beta_t^c \Delta F_t^c + b\beta_t^g \Delta F_t^g + c\beta_t^h \Delta F_t^h\right) r_t^f \ . \tag{3.5}$$

The total hedged portfolio P&L including margin and transaction costs $\Delta\Pi_t^*$ may now be expressed as

$$\Delta\Pi_t^* = \Delta\Pi_t + m_t^m - m_t^i - TC_t \ . \tag{3.6}$$

3.4 Evaluating Hedging Effectiveness

Hedging effectiveness is measured by the Ederington Effectiveness (EE) calculated as

$$EE = \frac{\sigma_u^2 - \sigma_h^2}{\sigma_u^2} \ ,$$

where σ_u^2 and σ_h^2 are the variances of the unhedged and the hedged portfolios, respectively. We compute the EE for each model in two ways: (i) using unconditional variances over the whole sample period and (ii) using a rolling window of EWMA variances with $\lambda = 0.97$, i.e. we use the conditional EE measures employed by Alexander and Barbosa (2007). The EWMA method is preferred to a rolling window of unconditional variances because the latter produces ghost features where the variances are augmented as long as a spike in the P&L remains inside the window. This is also to avoid any possible bias the unconditional EE may have over the conditional EE, as highlighted by Lien (2005) and Lien (2009). The conditional EE also allows us to examine the how the models' performance changes, as price series move through volatile and non-volatile time periods. To test whether the variance reduction from each model is significantly different from the variance reduction obtained using the naïve hedge, we apply the standard F-test for equality of variances.

3.5 Data

3.5.1 Spot Prices

Wednesday spot prices from 30/12/1992 to 23/02/2011 of Cushing WTI light-sweet crude oil, New York Harbour heating oil no.2, unleaded gasoline and RBOB gasoline barges are taken from Platts. In the rare cases where Wednesday is not a trading day, the price on Tuesday is taken instead.[5] The delivery location of the spot prices is the same as their corresponding NYMEX futures. We use Platts prices as these are collected from a window of physical commodity buyers which truly reflect the spot of the physical commodity trades. Energy products are primarily traded in barges and cargoes. Upon arrival, an investigator takes a sample from the freight and assesses its purity to determine a suitable price for the rest of the shipment. As such, prices may vary from barge to barge. To determine the spot price, Platts takes actual traded price from a window of energy traders and apply algorithms to determine the most likely price a particular barge would be traded at. Platts' prices are used as a benchmark for energy prices around the world.

Platts prices are determined at 4:30pm GMT as opposed to the NYMEX futures prices which are determined at 5:00pm GMT-5/6 (depending on summer/winter time zones), posing a non-synchroneity problem between the two sources. This may invoke a downward bias on the daily correlation between the spot and futures prices, but our analysis is on weekly data with weekly hedging horizons. As such, this relatively minor time difference will have negligible effect on the empirical results.

3.5.2 Futures Prices

Wednesday NYMEX futures prices of crude oil, heating oil, and gasoline from 30/12/1992 to 23/02/2011 are based on the NYMEX closing price. Among these three commodities, gasoline production has undergone some changes over time and therefore, since

[5] And in the case where Tuesday is not a trading day as well as Wednesday, Monday's price is taken instead. In the circumstance where none of those days are trading days, the week is omitted entirely.

2006, the NYMEX has no longer offered the original unleaded gasoline futures, replacing them by Reformulated Blendstock for Oxygen Blending (RBOB) gasoline futures. Due to data availability and low liquidity in the early years of the RBOB futures market, we switch from unleaded to RBOB gasoline in different years in the spot (2003) and futures markets (2006). This problem is of limited importance as both types of gasoline face the same demand and supply trends so that the prices are extremely highly correlated.

There are two ways to create a continuous series of futures P&Ls: the rollover method and the constant-maturity method. A standard rollover series is constructed by taking a futures price series up to a rollover date, the price series then jumps to the prompt futures series which is taken up to the next rollover date and so on. Often, the rollover dates are roughly a week before maturity to avoid thin market trading but for the commodities we study there is no need for this adjustment since trading continues in high volumes right up to the maturity date.

However, there are problems associated with the rollover futures series. As explained by Nguyen et al. (2011) where unlike constant-maturity series, any regression relating spot data to futures data will be contaminated by the "saw-tooth" pattern in the basis.

Galai (1979) compares two methods for creating constant-maturity futures which he terms the value index (interpolation between prices) and the return index (interpolation between returns). Galai shows that the return index method is the only one that provides realisable investments. As we require realisable investments to implement the optimal hedge ratios in practice, but our analysis must also be based on P&L rather than returns, we adapt Galai's return index method to the P&L as follows:

$$\Delta F_{t,T} = \eta_t \Delta F_{t,T_1} + (1 - \eta_t)\Delta F_{t,T_2} \ , \quad 0 \leq \eta_t \leq 1,$$

where $\Delta F_{t,T}$ is the constant-maturity futures P&L expiring in T days, $\Delta F_{t,T_1}$ and $\Delta F_{t,T_2}$ are the futures P&Ls expiring at T_1 and T_2 respectively, and

$$\eta_t = \frac{T_2 - (t + T)}{T_2 - T_1} \ , \quad T_1 < T < T_2.$$

A reasonable choice for T is 44 calendar days, i.e. approximately 1.5 months. With this choice there will always be two maturities straddling the constant-maturity. Of course, to maintain a constant-maturity series of ΔF_t^z for the regression (3.4), all futures' time to maturities must be the same.

3.5.3 Summary Statistics

Tables 4.3 and 3.2 report summary statistics and correlations of the weekly spot and constant-maturity futures P&L distributions based on the entire sample period. Crude oil spot and futures are less volatile than gasoline and heating oil spot and futures, and in each case the spot is more volatile than the futures. Each P&L except spot heating oil is slightly negatively skewed and all series are highly leptokurtic.

	ΔF^c	ΔF^g	ΔF^h	ΔS^c	ΔS^g	ΔS^h
μ	0.0324	0.1100	0.0569	0.0820	0.0964	0.1045
σ	2.3609	2.7426	2.7518	2.4586	3.1717	2.9239
τ	-0.2804	-0.3470	-0.0239	-0.1122	-0.3049	0.0431
$\varkappa$	6.3410	3.3010	7.0075	6.1150	3.2928	5.8725

Table 3.1.: Summary statistics for weekly constant-maturity futures and spot P&Ls for the sample period 30^{th} December 1992 - 23^{rd} February 2011. The total number of observations is 939 for each series. μ, σ, τ and $\varkappa$ denote the mean, standard deviation, skewness and excess kurtosis, respectively.

	ΔF^c	ΔF^g	ΔF^h	ΔS^c	ΔS^g	ΔS^h
ΔF^c	1	-	-	-	-	-
ΔF^g	0.8539	1	-	-	-	-
ΔF^h	0.9006	0.8395	1	-	-	-
ΔS^c	0.9718	0.8268	0.8683	1	-	-
ΔS^g	0.7357	0.9334	0.7423	0.7106	1	-
ΔS^h	0.8385	0.7859	0.9507	0.8128	0.6981	1

Table 3.2.: Correlation matrix between spot and futures P&Ls for the sample period 30^{th} December 1992 - 23^{rd} February 2011. The total number of observations is 939 for each series.

Figure 3.2 displays the P&L time series for all six variables. We observe that all series show rising volatility from the year 2000 onwards. Surges in prices produced by unexpected supply shortages result in frequent jumps in all the series. In many cases a decoupling of spot and futures prices results in a jump in the basis which is difficult to hedge effectively with the one-for-one ratio, and possibly also with a minimum-variance hedge ratio. Only one, very extreme spike in the data was removed. This was during the week of Hurricane Katrina, during which we assume no trades were made.

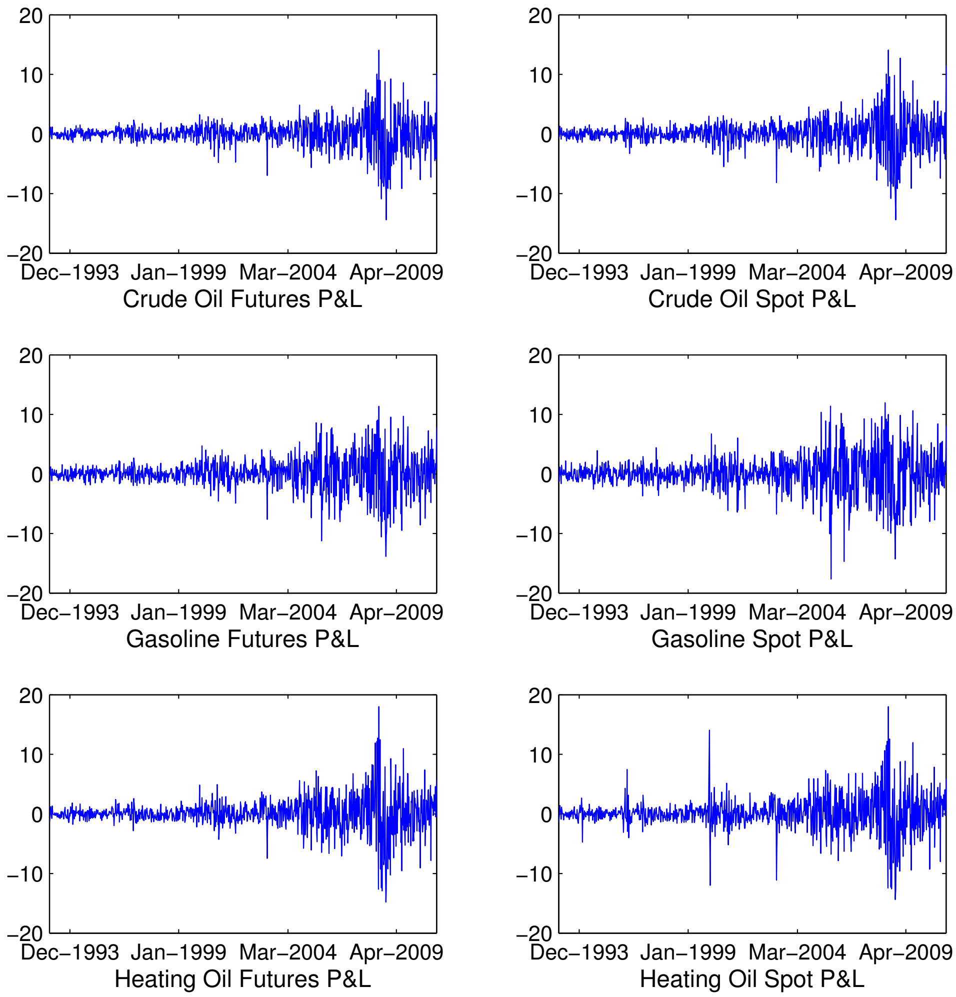

Figure 3.2.: Spot and constant-maturity futures P&L series for each commodity. Period: 30^{th} December 1992 - 23^{rd} February 2011. Prices for the week of 28^{th} August 2005 - 2^{nd} February 2005 have been removed due to abnormal market conditions caused by hurricane Katrina. The investor is assumed to make no trades on this week.

3.6 Empirical Results

We study the hedging performance of seven different models: naïve, OLS_{31}, OLS_{13}, OLS_{11}, $EWMA_{11}$, $GARCH_{11}$ and $AGARCH_{11}$ both, in-sample and out-of-sample. For the in-sample analysis, parameters are estimated using the entire data set, i.e. 939 weekly observations. Hedge ratios are then calculated based on these parameters and held constant for computing the hedge performance. But clearly, the in-sample analysis is just a data-fitting exercise – it is the out-of-sample analysis that matters for practical purposes. Here, the parameters are estimated using a rolling window of 260 weeks.[6] The hedge ratios estimated at time t are then applied to the one step ahead P&L. The hedger is assumed to re-estimate the parameters every week. Since the EWMA parameter λ is always constant, EWMA results are the same both in-sample and out-of-sample.

All empirical results presented are for the 3 : 2 : 1 crack spread bundle, as many refineries have this approximate crack spread and the original NYMEX margin bundles were also based on this spread.

3.6.1 Hedge Ratios

Table 3.3 reports the average hedge ratios for each model and their standard deviations. In-sample hedge ratios are reported for completeness, we focus the following discussion on the out-of-sample hedge ratios. The multiple-equation model OLS_{31} yields hedge ratios closer to 1.0, yet the single-equation models produce hedge ratios nearer to 1.3. It is tempting to conclude that the OLS_{13} model produced these higher hedge ratios because of multicollinearity. However, both OLS_{13} and OLS_{11} produce hedge ratios of roughly the same magnitude, which brings into question any such conclusion.

The OLS_{31} model produces smaller hedge ratios, closer to 1, because all cross-market correlations are assumed to be zero. As they are certainly not (see Table 3.2) this produces a substantial bias. On the other hand, the OLS_{13} assumes an equal cross-market correlation across all commodities – an assumption that seems reasonable in light of Table 3.2.

Figure 3.3 displays the evolution of the OLS models' out-of-sample hedge ratios over

[6]For the OLS methods we have also employed windows of length 104, 156, and 208 to ensure that this choice is not the driver of our results. No significant differences were found. For the GARCH models, shorter windows were not feasible due to the number of parameters to be estimated. In some few instances, the optimisation of the GARCH parameters failed to converge. We then used the estimates from the previous week.

in-sample	OLS_{13}	OLS_{31}	OLS_{11}	$EWMA_{11}$	$GARCH_{11}$	$AGARCH_{11}$
β^c	1.285	1.012	1.277	1.381	1.363	1.381
	-	-	-	(0.215)	(0.262)	(0.205)
β^g	1.296	1.079	1.277	1.381	1.363	1.381
	-	-	-	(0.215)	(0.262)	(0.205)
β^h	1.202	1.010	1.277	1.381	1.363	1.381
	-	-	-	(0.215)	(0.262)	(0.205)
out-of-sample						
β^c	1.417	1.035	1.412	1.381	1.392	1.368
	(0.162)	(0.021)	(0.135)	(0.215)	(0.343)	(0.326)
β^g	1.432	1.147	1.412	1.381	1.392	1.368
	(0.109)	(0.039)	(0.135)	(0.215)	(0.343)	(0.326)
β^h	1.349	1.084	1.412	1.381	1.392	1.368
	(0.363)	(0.061)	(0.135)	(0.215)	(0.343)	(0.326)

Table 3.3.: Average in-sample and out-of-sample hedge ratios with standard deviations in parentheses. In-sample ratios are estimated using the entire sample period: 30^{th} December 1992 - 23^{rd} February 2011. Out-of-sample hedge ratios are estimated using a moving window of 260 weeks.

time. One can observe that OLS_{31} and OLS_{11} are relatively stable. In contrast, the hedge ratio for the heating oil contract of OLS_{13} in particular exhibits some substantial transitions over time. This is also reflected by the relatively high standard deviation in Table 3.3. The $GARCH_{11}$ estimation method produced the most volatile out-of-sample hedge ratios. Although this characteristic is expected given that GARCH parameters are generally more sensitive with respect to innovations in the data, the volatility of the hedge ratios should roughly be of the same magnitude as the $EWMA_{11}$ hedge ratios.[7] According to Table 3.3 however, the out-of-sample $GARCH_{11}$ hedge ratios are roughly 33% more volatile than the $EWMA_{11}$ hedge ratios.

This is also shown in Figure 3.4, which compares the behaviour of the hedge ratios derived for the single-equation, single-variable models over time. Note how volatile the GARCH hedge ratios are over time. Would a serious risk manager implement a hedging strategy that involved re-balancing more than 100 % of the hedging portfolio from week to week? This casts serious doubts on the merits of GARCH-based hedge ratios.

[7] For some periods, the GARCH hedge ratios are unmanageably large for the investor (e.g. up to ± 5 times the spot investment). To control these, when the absolute value of the GARCH hedge ratios exceed twice the absolute value of the EWMA hedge ratio, the GARCH hedge ratio from the previous time step is used instead.

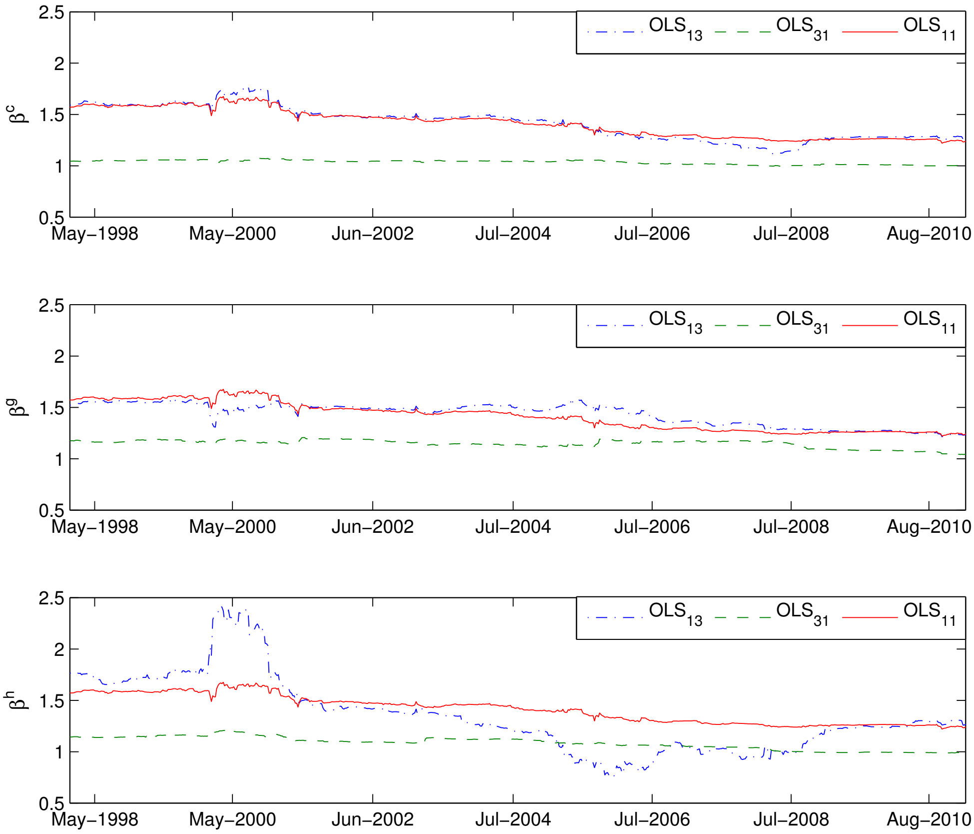

Figure 3.3.: OLS hedge ratios calculated using a moving window of 260 weeks. Period: 14^{th} January 1998 - 23^{rd} February 2011.

The GARCH model parameter estimates are highly volatile over time. Table 3.4 displays the means of the estimated GARCH parameters, and their standard deviations measured over the entire out-of-sample period. Clearly, the estimates are far from being stable. Extending the length of the estimation windows up to 8 years did not produce substantially more stable estimates. Hence, the problem is not one of convergence to local optima instead of a global optimum, but rather an intrinsic problem with applying GARCH models for hedging when there are frequent jumps in a highly volatile basis. In this situation, large changes in the conditional variance parameter estimates are only to be expected. Indeed, the finding of highly unstable GARCH hedge ratios is nothing peculiar for our data set. Previous studies, e.g. Lee and Yorder (2007) and Lee (2010), have found similar results.

Another problem concerns transaction costs. Re-balancing a hedge with such extreme swings will amount to much higher transaction costs in comparison to the other

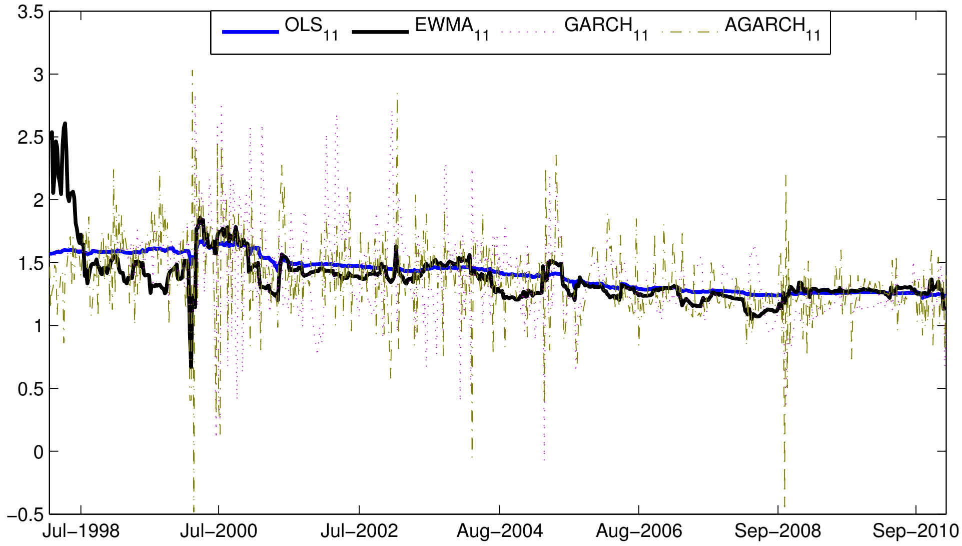

Figure 3.4.: Comparisons between out-of-sample hedge ratio estimates of the OLS_{11}, $EWMA_{11}$, $GARCH_{11}$ and $AGARCH_{11}$ models. $EWMA_{11}$ hedge ratios estimated with $\lambda = 0.97$. Period: 14^{th} January 1998 - 23^{rd} February 2011.

methods having more stable hedge ratios. Table 3.5 presents the average transaction costs (including margin costs) of the seven hedging strategies. One can see that the $GARCH_{11}$ and $AGARCH_{11}$ models produce average transaction costs of $ 0.040 and $ 0.059 per bundle. A refinery that purchases 50,000 3 : 2 : 1 crack spread bundles per week for example, would be paying $ 156,000 per year only to implement their hedging strategy. This is very large in comparison to the other models, especially the naïve strategy, where hedging does not require re-balancing and the associated margin and transaction costs are much smaller.

3.6.2 Hedging Effectiveness

We now consider the hedging effectiveness of each model. The main question is whether the effort to implement more advanced models and the associated transaction costs pay off in a superior hedging performance? Table 3.6 shows the overall hedging performance measured by the unconditional EE of each model both, in- and out-of-sample. In the more relevant out-of-sample test, all models produce variance reductions in the range of 64-71% with the OLS_{11} as the most effective model. The $AGARCH_{11}$ model performs worst, with an EE of 64.81%.

Figure 3.5 displays the out-of-sample conditional EE for each model over time.

	$GARCH_{11}$		$AGARCH_{11}$	
Parameter	Mean	St. Dev.	Mean	St. Dev.
A_{11}	-0.423	(0.791)	-0.292	(0.910)
A_{21}	-0.785	(1.790)	-0.887	(2.054)
A_{22}	0.311	(0.381)	0.163	(0.203)
B_{11}	-0.271	(0.257)	0.058	(0.393)
B_{12}	-1.039	(0.548)	-0.060	(0.808)
B_{21}	0.022	(0.131)	0.031	(0.186)
B_{22}	0.540	(0.194)	-0.466	(0.685)
C_{11}	0.818	(0.274)	0.865	(0.354)
C_{12}	0.158	(0.191)	1.599	(0.668)
C_{21}	0.034	(0.148)	-0.023	(0.182)
C_{22}	0.728	(0.136)	-0.244	(0.423)
D_{11}	-	-	0.139	(0.224)
D_{12}	-	-	0.271	(0.669)
D_{21}	-	-	0.017	(0.209)
D_{22}	-	-	-0.0.21	(0.421)
log-likelihood	-1170.116	(240.992)	-1161.063	(236.468)

Table 3.4.: Out-of-sample mean and standard deviation of $GARCH_{11}$ and $AGARCH_{11}$ parameter estimates using a 260 weeks rolling window. Period: 14^{th} January 1998 - 23^{rd} February 2011.

	naïve	OLS_{13}	OLS_{31}	OLS_{11}	$EWMA_{11}$	$GARCH_{11}$	$AGARCH_{11}$
In-sample	0.008	0.010	0.008	0.009	0.014	0.024	0.050
	(0.004)	(0.005)	(0.005)	(0.005)	(0.011)	(0.027)	(0.050)
Out-of-sample	0.008	0.013	0.009	0.011	0.014	0.040	0.059
	(0.004)	(0.005)	(0.006)	(0.007)	(0.011)	(0.047)	(0.070)

Table 3.5.: Average margin and transactions costs in $ per spot bundle, numbers in parentheses represent standard deviations. Period 14^{th} January 1998 - 23^{rd} February 2011.

It is variable throughout the sample period and occasionally reacts to the jumps in the basis. For instance, during the first quarter of 2000 the hedging effectiveness of all models drops below 0% but then rises to about 40% after about 3 months. This is due to the surge in heating oil prices (note the spike at this time in the bottom, right-hand graph in Figure 3.2). We have not excluded data from this event because the price shift occurred over a period of two months, and hedging would have been necessary over such a long period. Although the GARCH models are expected to perform better under these conditions since they are more capable to react to changing

	naïve	OLS_{13}	OLS_{31}	OLS_{11}	$EWMA_{11}$	$GARCH_{11}$	$AGARCH_{11}$
In-sample	67.28%	69.98%	68.01%	70.47%	69.71%	67.33%	68.58%
Out-of-sample	67.28%	69.70%	67.78%	69.98%	69.71%	67.40%	64.81%

Table 3.6.: Whole-sample EE of each model in percentage points. Hedge ratios calculated using a 260-week rolling window. Whole-sample data points: 679. Period 14^{th} January 1998 - 23^{rd} February 2011. Whole-sample unhedged portfolio variance: 44.34 $\$^2$. Hedged portfolio includes margin and transaction costs.

market conditions, here they produce roughly the same hedging effectiveness as all the other models.

From Figure 3.5 we can conclude that all models have similar effectiveness throughout the entire sample period. To test this more formally, we perform a standard F-test, for equality of variances: between the variance of P&L resulting from the naïve hedge and the P&L variance from each of the models. We use the out-of-sample P&L and evaluate the F-statistic using a rolling window to calculate the individual variances.[8] Figure 3.6 depicts these F-statistics together with lines showing the critical values at the 90% and 95% confidence level. We fail to reject the null hypothesis that the hedge portfolio variance produced by more advanced models is significantly smaller than the naïve strategy in every instance. No model is able to improve upon the naïve hedge, utilising the 3:2:1 bundle offered by NYMEX. The same conclusion is reached for all the $a : b : c$ crack spreads considered, although those results have not been reported for brevity.

A further robustness check was carried out to assess the dependence between the EE and different volatility regimes. The results are presented in Tables 3.7 and 3.8. We re-estimated the average hedging effectiveness of each model: during high volatility and low volatility time periods. As a preliminary test, the first half of the sample (sample A) was taken as the low volatility regime (30^{th} December 1992- 11^{th} August 2004, average underlying variance 10.12 $\$^2$), and the second half of the sample (sample B) was the high volatility regime (11^{th} August 2004 - 23^{rd} February 2011, average underlying variance 58.24 $\$^2$). Empirical results confirm that all models again produce roughly the same EE, as for the full-sample results. We also find that the models perform better in sample B than in sample A: in sample B the EE ranges from 67-74 %, compared with 53-58 % in sample A. One might suppose that this was because the correlation between the spot and futures crack spread increases during

[8] We have used a one year window (52 weeks) in order to obtain a long out-of-sample period. Results for longer windows (260 weeks) yield identical conclusions.

volatile times. However, upon further investigation we find that this is not the case. The GARCH correlation levels remain roughly 0.80 throughout both samples, with the exception of the period 26^{th} January 2000 - 9^{th} February 2000, where the futures-spot correlation drops to -0.12 (attributed to the heating oil surge in January-March 2000). The downward jump in the EE that is evident in figure 3.5 is due to this momentary shock alone. It so happens that the shock occurred during sample A, and this is the reason why models perform better during sample B.

	In-sample	Out-of-sample		
	Whole Sample	Whole Sample	Sample A	Sample B
naïve	67.28%	67.28%	53.38%	70.68%
OLS_{13}	69.98%	69.70%	54.83%	70.95%
OLS_{31}	68.01%	67.78%	57.16%	72.88%
OLS_{11}	70.47%	69.98%	57.83%	73.47%
$EWMA_{11}$	69.71%	69.71%	56.46%	73.05%
$GARCH_{11}$	67.33%	67.40%	56.16%	70.26%
$AGARCH_{11}$	68.58%	64.81%	54.54%	67.53%

Table 3.7.: Whole-sample EE of each model in percentage points. Hedge ratios calculated using a 260-week rolling window. Whole-sample data points: 679. Whole-sample period 4^{th} February 1998 - 23^{rd} February 2011, Sample A period 3^{rd} February 1999 - 9^{th} February 2005, Sample B period 9^{th} February 2005 - 23^{rd} February 2011. Whole-sample, Sample A, Sample B unhedged portfolio variances: 44.34 $\$^2$, 10.12 $\$^2$, 58.24 $\$^2$ respectively. Hedged portfolio includes margin and transaction costs.

	In-sample	Out-of-sample		
	Whole Sample	Whole Sample	Sample A	Sample B
naïve	62.02%	62.02%	59.35%	66.08%
OLS_{13}	65.45%	63.99%	61.94%	68.44%
OLS_{31}	63.06%	62.81%	60.53%	66.50%
OLS_{11}	65.35%	64.54%	62.47%	68.96%
$EWMA_{11}$	64.06%	64.06%	61.99%	68.30%
$GARCH_{11}$	61.86%	62.34%	60.87%	66.07%
$AGARCH_{11}$	62.77%	60.53%	59.67%	63.02%

Table 3.8.: Average EWMA EE of each model in percentage points. Hedge ratios calculated using a 260-week rolling window. Hedged portfolio sample period 13^{th} January 1999 - $23rd$ February 2011. Hedged portfolio includes margin and transaction costs.

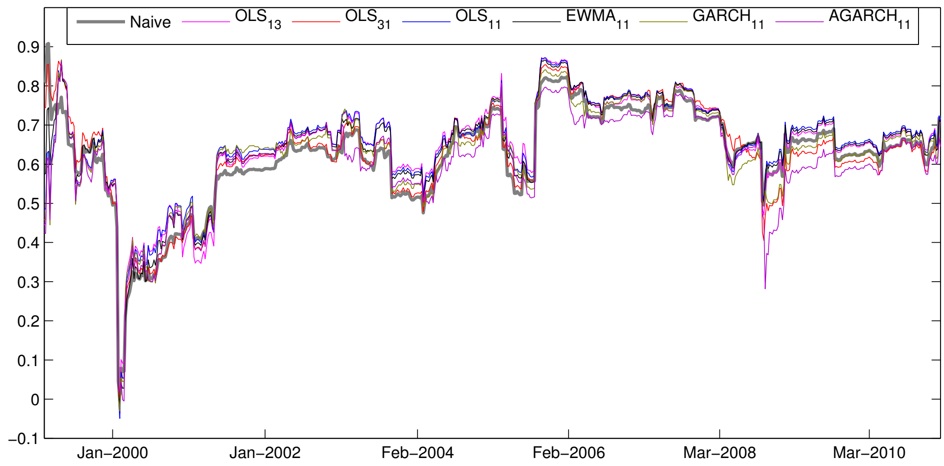

Figure 3.5.: Out-of-sample analysis: EWMA *EE* of each model. Period: 13^{th} January 1999 - $23rd$ February 2011. Hedge ratios calculated using a 260-week rolling window.

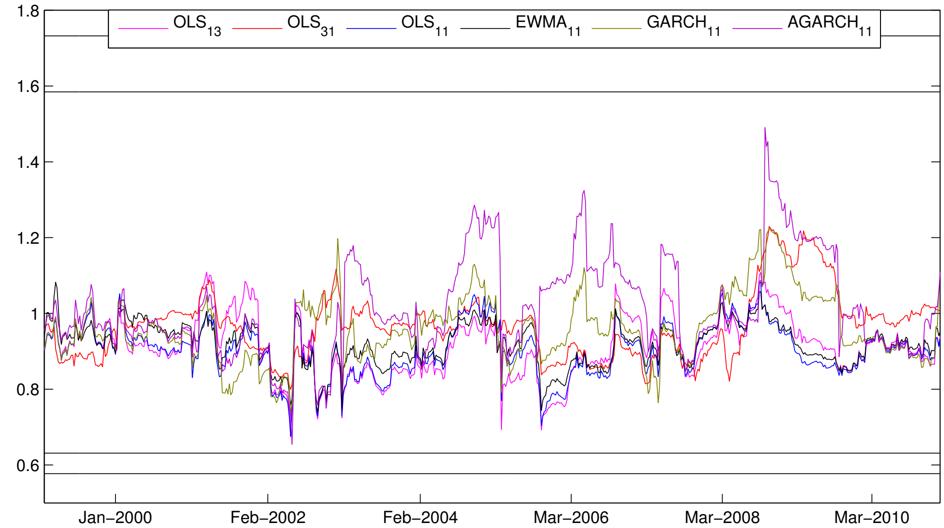

Figure 3.6.: Rolling-moving F-statistic for testing the equality of variances between each hedging model and the naïve hedged portfolio. Rolling-moving variances are calculated using a 52-week rolling window. Hedge ratios calculated using a 260-week rolling window. Period: 13^{th} January 1999 - $23rd$ February 2011. Horizontal lines indicate two-sided critical values at 5% and 10% significance levels, respectively.

To ensure that the hedging effectiveness is not driven by our assumption regarding the size of margin and transaction costs, we repeat the analysis, this time ignoring those costs. The change in EE resulting from excluding margin and transaction costs is found to be very small and mostly positive, the largest being 0.14% for the AGARCH model. One could also account for the correlation between the margin costs, or transactions costs, and the spot and futures P&L when minimising the portfolio variance, i.e. minimising the variance of $\Delta\Pi_t^*$ in equation 3.6 as opposed to the variance of $\Delta\Pi_t$ in equation 3.1. However, the correlations between the hedged portfolio P&L and the margin costs, or transaction costs, are very small (in the region of -0.09 to 0.03). Thus, we do not account for these correlations; it would have only minimal effect on the empirical results.

3.6.3 Margin Calls

We now discuss the case where the refinery has to raise capital to cover the margin call. While the variation in the transaction costs does not significantly undermine the hedging effectiveness, raising the margin call poses a problem for the refinery.

Upon entering the bundle futures contract, the refinery deposits a margin which is marked to market daily. We assume that the refinery has privileges as any exchange member and are not subjected to margin calls intraday. Given the initial and maintenance margins are equal, should the value of the positions on the bundle fall at the end of the next day, the refinery would receive a margin call. There are some studies which suggest complex hedges for margin calls using binary options for speculators (see Day and Lewis (2004)) but as the refinery is a hedger, we assume for now that they are left with one option -to prepare capital to cover possible margin calls. The amount of capital can equate to, say, the 99%, one-day VaR of the bundle's movement. Whilst this is exhausted upon one margin call, the refinery would have to replenish this capital level back to its full amount to guard against the next. Note that, the only risk under consideration here belongs purely to the futures bundle's price movements.

Now, if the margin level is variable, the refinery faces two sources of uncertainty: the changes in the bundle futures price and the changes in the margin level itself. The refinery must then prepare a pool of capital to cover both sources of risk, i.e. the 99%, one-day VaR of $-\Delta F_t^z - \Delta M_t$. From this relationship, it is clear that the capital level needed is highly dependent on the variability of ΔM_t. The larger the variability, the greater the capital required and should this be excessive, the refinery may abandon this strategy all together and bear the volatility of the spot market without hedging. Furthermore, formulating a VaR estimate on such a portfolio is not straightforward. Perhaps a closed-form solution is possible if M_t was set to a constant proportion of

σ_t^z (e.g. any parametric VaR) but given the variability of σ_t^z is often large, especially for energy portfolios, there remains a hurdle for the refinery to raise such capital.

This problem will not only be prevalent among hedgers but all investors in the market. Variable ΔM_t is difficult to manage and adds to investors' transaction costs. It would hence be in the investors' best interest to trade in an exchange who offers stable margin requirements.

3.7 Conclusions

We have compared seven different models for estimating hedge ratios for crack spread delta hedging. Although all models are found to produce a healthy amount of variance reduction (roughly 68% on average). The most complex models (i.e. GARCH models) deliver the worst hedging results. The hedging strategies derived from GARCH models are not only more complicated to implement, they also generate the highest transaction costs. Moreover, instability in parameter estimates is another problem which can lead to unrealistically high hedge ratios associated with the most complex models.

Our findings contradict a fair body of existing literature which concludes that model-based minimum-variance hedging is superior, particularly when GARCH models are employed. In contrast, we find that the hedging effectiveness is statistically indistinguishable between all the models considered. This finding is based on a very long out-of-sample period, but we would have reached the same conclusion had we used much shorter sub-periods or, indeed, had we based conclusions on in-sample analysis alone.

We have taken much more care with the data than the previous studies that have analysed the hedging of the crack spread. We use the best (Platts) spot prices and we replace the rollover log return series, which are typically used in studies of this type and are affected by the saw-tooth pattern in the basis that biases the OLS hedge ratios. Moreover, we take meticulous care to account for all the costs involved in hedging. The margin and transaction costs of minimum-variance hedging have a very small effect on the hedging effectiveness, even for the excessively variable hedge ratios prescribed by GARCH models. However, these costs are important to analyse, because they reveal that GARCH hedging models would be too expensive to implement in practice, even if they did provide statistically significant superior performance (which they do not). Our discussion on margin calls further reveals that more variable margins imposes more costs for investors since they need to raise additional capital to cover movements in the initial margin. It would be in the investors' best interest to move to an exchange which issues stable margin requirements.

The main point for end-users to take away from our study is that, even for complex underlyings such as spreads on oil-related commodities which produce a basis that is extremely variable and jumpy, the maturity mismatch justification for minimum-variance hedging is simply not viable. The naïve hedge ratio performs as well as any other model, and it requires the least re-balancing of all. It may be that minimum-variance hedging can improve on the so-called "naïve" hedge when a proxy futures contract must be used – but even this remains an open question waiting for a thorough empirical analysis.

4. Value-at-Risk for Energy Futures Term Structures Margin Requirements

4.1 Introduction

Since the financial crash in October 1987 and the development of advanced risk management tools, Value-at-Risk (VaR) has become the most popular market risk quantification method worldwide This is evident in its implementation by international regulating bodies: the Basel amendments in 1996, the Dodd-Frank act in 2010 and more recently European Market Infrastructure Regulations (EMIR) coming in place this year. With a significant proportion of investors migrating to the energy futures market in search of alternative investments (especially crude oil), advancements in VaR estimation for energy futures is essential for protecting financial institutions and investors from unexpected loss.

In this study, we search for the best VaR model for margin requirement purposes which includes two unique features not yet addressed in the literature. First, as exchanges are faced with long and short positions simultaneously, the VaR model must be able to accurately quantify both the upper and lower tail of the distribution at the same time. Second, exchanges like to publish margin levels for individual futures online hence all futures along the term structure must have margins assigned. To this, the exchange requires a VaR model that can accurately describe, not only the first-to-mature futures series, as is studied by the majority of works in the literature, but for all available futures contracts along the term structure. We find few works which examine VaR on a term structural basis, Nomikos and Pouliasis (2014) is perhaps the most closely-related work in this respect. Their study however, analyses the use of VaR models for portfolio management and not for margin requirements. Furthermore, when backtesting, they do not consider clustering of exceedances which is essential when examining the risk sensitivity of the VaR; their sample period also stretches up to 2009 only.

Here, we carry out rigorous examinations on VaR, where we use a rolling log-likelihood statistic for the entire term structure. We also address some unanswered questions on whether or not the use of returns or profit and losses (P&L's) based on constant-maturity or rollover futures are more appropriate. As discussed in the chapter 3, the use of log returns is in appropriate. Log returns can be highly inaccurate, even at the daily frequency and inappropriate when applied to constant-maturity futures.

The rest of the chapter is structured as follows: first, we propose a rigorous methodology for examining VaR models for margin requirement purposes; second we present the data; third, we discuss our results; and finally we conclude.

4.2 Methodology

We concentrate mainly on parametric VaR estimation methods as these are the most prevalent in the literature and include one semi-parametric estimation method for comparison. We have not considered more complex models here, as our study is perhaps the first to address backtesting 2-tails VaR models for Brent crude oil futures, further investigation could be to explore more advanced models.

4.2.1 Parametric VaR Models

For a parametric VaR model, one assumes the P&L's or returns follow a parametric distribution. The parameters of the distribution are estimated and the VaR is a percentile on such distribution. Here, we assume two different types of innovations: normal and student t, with cumulative density function denoted as $\Phi[\cdot]$ and $t_\nu[\cdot]$ respectively. The corresponding α percent, h-day VaR can then be estimated according to the relationships:

$$\begin{aligned} VaR_{t,T}^{\alpha,h} &= \Phi^{-1}[\alpha]\hat{\sigma}_{t,T} \ , \\ VaR_{t,T}^{\alpha,h} &= t_\nu^{-1}[\alpha]\hat{\sigma}_{t,T} \ , \end{aligned} \tag{4.1}$$

where $\hat{\sigma}_{t,T}$ is the time-t volatility estimate of the futures return of P&L with maturity T. We consider a number of volatility estimation methods used in practice and the literature as parametric VaR is heavily reliant on this procedure.

Exponentially Weighted Moving Average

Exponentially Weighted Moving Average (EWMA), popularised by JP Morgan, is one of the models used by the CME. Its specification follows

$$\hat{\sigma}^2_{t,T} = \lambda \hat{\sigma}^2_{t-1,T} + (1-\lambda)\Delta F^2_{t-1,T} \ , \tag{4.2}$$

where $F_{t,T}$ is the futures price at time t maturing at time T, assuming $E[\Delta F_{t-1,T}] = 0$. The same relationship can be derived for returns by replacing $\Delta F_{t,T}$ with $\frac{\Delta Ft,T}{F_{t-1,T+1}}$ and the corresponding dollar-VaR is found by multiplying the return VaR forecast to the current futures price. All volatility processes described in this section can be derived using both alternatives. In this chapter, we only show the P&L alternative for brevity. Note that the time $t+1$ variance forecast for the EWMA specification is the same as the variance estimate at time t. For the daily frequency, we choose the decay parameter $\lambda = 0.94$ as recommended by JPMorgan's RiskMetrics technical documents. Although ν and λ can be calibrated to fit the data via a maximum-likelihood criterion, as is carried out in any conventional GARCH calibration procedure, we do not encourage such a procedure for the EWMA model as this would detract from its simplicity in using *ad hoc* parameters. Note that, the excess kurtosis of the t distribution is $\frac{6}{\nu-4}$, hence for a positive and finite kurtosis, ν must be greater than 4. Here, we estimate the VaR using $\nu = 6$ and 12 as these can imitate the leptokurtic nature of futures P&L's well.

Generalised Auto-Regressive Conditional Heteroskedasticity

As the literature tend to favour student-t-asymmetric GARCH models (such as the GED-GARCH model from Fan et al. (2008) or the skewed-t-APARCH model from Giot and Laurent (2003)), we include the GJR-GARCH model in our analysis. The specification is given by:

$$\hat{\sigma}^2_{t,T} = \hat{\beta}_0 + \hat{\beta}_1 \Delta F^2_{t-1,T} + \hat{\beta}_2 \hat{\sigma}^2_{t-1,T} + \hat{\beta}_3 \mathbb{1}_{\Delta F_{t-1,T}<0} \Delta F^2_{t-1,T} \ , \tag{4.3}$$

Note that the GJR GARCH configuration allows for different reaction in P&L depending on its direction. What is not addressed in the literature (see Su et al. (2011) for example), is that at time $t-1$, one does not know the direction of the P&L at time t. The volatility for the upward movement would be different to that of a downward movement. Hence at every time step $t-1$ we need two separate VaR estimations, one based on the lower tail of the distribution and one on the upper tail. The variance forecast of $\Delta F_{t,T}$ is given by

$$\hat{\sigma}^2_{t,T} = \hat{\beta}_0 + (\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 \mathbb{1}_{\Delta F_{t-1,T}<0})\hat{\sigma}^2_{t-1,T} \ , \tag{4.4}$$

and the $VaR^{\alpha,1}_{t,T} = t^{-1}_{\nu}[\alpha]\hat{\sigma}_{t,T}$. In the case of normal distribution innovations, parameter estimates $\hat{\beta}_i$ for $i = 0, 1, 2, 3$ can be found by maximising the log-likelihood function specified as:

$$LL_t = \sum_{i=t-n+1}^{t} \log f(\mathbf{\Theta}_i) \ , \tag{4.5}$$

where n is the length of the estimation window and f probability density function of the innovations with parameters $\mathbf{\Theta}$, i.e. the normal and student t distributions. Although more complex calibration methods such as Markov Chain Monte Carlo can improve GARCH parameter stability, our work is not primarily about calibration methods and they will not be considered here.

Orthogonal Volatility Models

With regards to estimating volatility on a term structural basis in a similar spirit to Tolmasky and Hindanov (2002) and Nomikos and Pouliasis (2014) - we test the models using Principal Component Analysis (PCA). First, we model the orthogonal movements (the first three are: shift, tilt and convexity) of the term structure instead of all the individual series. The variance estimate can be described as:

$$\hat{\sigma}^2_{t,T} = \sum_{i=1}^{k} \hat{\sigma}^{*2}_{t,i} W^2_{i,T} \ , \tag{4.6}$$

where $\hat{\sigma}^{*2}_{t,i}$ is the variance of the i^{th} eigenvector, $W_{i,T}$ is the i^{th} element of the T^{th} eigenvector of the unconditional covariance matrix which is estimated using a 1250-day rolling window and k is number of eigenvectors. The equivalent Orthogonal VaR estimate can then be calculated via the relationship 4.1. Following Nomikos and Pouliasis (2014), we choose $k = 3$ as this can explain the majority of the movements in the term structure. The covariance matrix is preferred to the correlation matrix in this case because the Samuelson effect is embedded in its estimation procedure, which also produces smoothing-decaying VaR estimates along the term structure.

For the EWMA model, we first estimate the volatilities on the principal components. We make no assumption on the P&L distribution at this stage and we transform this to the term structure volatilities according to the relationship 4.6. Then, we assume that each futures P&L's in the term structure are either t-distributed with $\nu = 6$ or 12 or normally distributed and calculate the VaR. On the other hand, we cannot apply the GJRt model on an orthogonal basis. This is because we calibrate the GARCH parameters on the principal components which include the degrees of freedom of the innovations. Although, we can use relationship 4.6 to calculate the futures volatility, we cannot determine the degrees of freedom the t distributions of the futures term structures. Therefore, we employ orthogonalisation assuming normal distribution innovations for the GARCH model only.

4.2.2 Semi-Parametric VaR Models

We include one semi-parametric VaR model, i.e. the Kernel-fitted VaR model, to compare with the parametric VaR models. Although this model is not prevalent in the literature, we find in our test samples that it can capture the skewness of the return distribution well. We generate probability distributions using a normal-kernel smoothing function based on 60 and 120-day (approximately 3 and 6 months) rolling windows of historical empirical distribution of P&L's and use the α percentile of the kernel-fitted distribution as the estimate of $VaR_{t,T}^{\alpha,h}$. We have chosen relatively small windows to allow the VaR to be reactive to market conditions. The time t VaR forecast is assumed to be the same as its estimate at time $t-1$.

4.2.3 Model Summary

In total, we test 11 different VaR models, including 2 semi-parametric models, see Table 4.1 for more details.

4.2.4 Returns or Profit and Losses?

Previous literature concentrates on estimating VaR using returns and multiplying this to the current futures price to calculate the dollar-VaR. These results are different for calculating VaR on the P&L series. Although estimation of VaR on returns can be more stable as these are more stationary than P&L's, the corresponding dollar-VaR estimation can be highly variable. This can be highly problematic for risk management purposes; margin requirements in particular retain many advantages for reducing procyclicality should it be based on more stable VaR estimates.

Here, we find that with volatility models which are adaptable to changing market conditions such EWMA and GARCH, estimation on the P&L series can be just as powerful as those on the return series. We estimate and backtest the VaR on both P&L and returns and evaluate which alternative can produce more accurate VaR

Model Name	Volatility Estimation Method	Distribution
EWMAt-6	EWMA, $\lambda = 0.94$	Student-t, $\nu = 6$
EWMAt-12	EWMA, $\lambda = 0.94$	Student-t, $\nu = 12$
EWMA	EWMA, $\lambda = 0.94$	Normal
GJRt	GJR	Student-t
GJR	GJR	Normal
K-60	-	Fitted Gaussian Kernel using a 60-day rolling window
K-120	-	Fitted Gaussian Kernel using a 120-day rolling window
OEWMAt-6	Orthogonal EWMA, $\lambda = 0.94$, covariance matrix estimated using a 1250-day rolling window	Student-t, $\nu = 6$
OEWMAt-12	Orthogonal EWMA, $\lambda = 0.94$, covariance matrix estimated using a 1250-day rolling window	Student-t, $\nu = 12$
OEWMA	Orthogonal EWMA, $\lambda = 0.94$, covariance matrix estimated using a 1250-day rolling window	Normal
OGJR	Orthogonal GJR, covariance matrix estimated using a 1250-day rolling window	Normal

Table 4.1.: Summary of all tested models in this chapter

forecasts. To the best of our knowledge, no previous study addresses this important, yet, straightforward issue.

4.2.5 Constant-Maturity or Rollover Futures?

Previous studies tend to focus their analysis on the rollover series only. In the presence of the Samuelson effect however, estimating the volatility on the rollover series can lead to biased results. Nomikos and Pouliasis (2014) carried out their analysis on constant-maturity data which are void of these effects. In the context of margin requirements however, one requires the margin of all futures positions along the term structure, which do not have constant maturity.

Methods for generating constant-maturity series is outlined in the previous chapter. There is however one drawback in using constant-maturity series in estimating VaR for margin requirements. To elaborate, consider an investor wishes to purchase a contract with 15 days to maturity, the exchange therefore needs the VaR of this contract to issue the margin. If the VaR term structure was estimated using constant-maturity futures, the exchange would need to fit a spline through the term structure which are at say 30,60, ..., 360 days to maturity and use the same function to estimate the VaR at 15 days to maturity. Intuitively speaking, the error involved in the spline-(or curve-) fitting procedure will ultimately reduce the accuracy of the VaR forecast.

After selecting the most accurate VaR models using constant-maturity series, we also fit a Hermite spline to the term structure at the 30,60, ..., 360 days to maturity. We compare the VaR estimates along the spline with the existing futures contracts' P&L's on the appropriate point in the term structure on a daily basis. We apply conventional back testing methods (from Christoffersen (1998)) to evaluate the accuracy of the VaR predictions (these models are hereafter denoted as *CM-spline* models). We then compare this to results obtained from backtesting VaR calculated via the rollover series.

4.2.6 Model Selection Methods

Here, we outline a rigorous backtesting procedures for the VaR, paying particular attention to the nature of margin requirements and consistency of the VaR models' performance. EMIR laws also require exchanges to publish backtesting reports to ensure their estimations are statistically correct. In particular, Article 49 from EMIR (European Union (2013)) reads

> *A CCP shall assess its margin coverage by performing an ex-post com-*

> *parison of observed outcomes with expected outcomes derived from the use of margin models. Such backtesting analysis shall be performed each day in order to evaluate whether there are any testing exceptions to margin coverage.*

The relationship between *the use margin models* and the VaR is discussed in the next chapter.

We employ Christoffersen (1998)'s backtesting techniques to evaluate the models as this is the most prevalent in the literature. As different interval forecasts are suitable for different risk management purposes, especially for margin requirements, we require backtesting at two intervals simultaneously according to reasons previously discussed in section 4.1 and chapter 2.

For a series of VaR estimates on any P&L series length n, we divide the distribution forecast into 3 sections: 1) $\Delta F_{t,T} < -VaR_{t,T}^{0.99,1}$ 2) $-VaR_{t,T}^{0.99,1} < \Delta F_{t,T} < VaR_{t,T}^{0.01,1}$ and 3) $\Delta F_{t,T} > VaR_{t,T}^{0.01,1}$. Denote the number of observed P&L's in sections $\{1,2,3\}$ as n_1, n_2, n_3 respectively and denote any P&L's in section $i = \{1,2,3\}$ immediately followed by the P&L in section $j = \{1,2,3\}$ as n_{ij}. Also, denote the expected proportion of P&L's in section i as p_i. The unconditional and conditional log-likelihood-ratio statistics evaluated at time t on $\Delta F_{t,T}$, $VaR_{t,T}^{0.99,1}$ and $VaR_{t,T}^{0.01,1}$, $LR_{t,T}^{uc}$ and $LR_{t,T}^{cc}$, can be calculated using the relationships

$$LR_{t,T}^{uc} = -2(L(\Pi_{t,T}^{0}) - L(\hat{\Pi}_{t,T}^{2})) \ , \tag{4.7}$$

$$LR_{t,T}^{cc} = -2(L(\Pi_{t,T}^{0}) - L(\hat{\Pi}_{t,T}^{1})) \ , \tag{4.8}$$

where

$$L(\Pi_{t,T}^{0}) = \sum_{i=1}^{3} \log(p_i) \ , \tag{4.9}$$

$$L(\hat{\Pi}_{t,T}^{1}) = \sum_{i=1}^{3}\sum_{j=1}^{3} n_{ij} \log\left(\frac{n_{ij}}{n_i}\right) \ , \tag{4.10}$$

$$L(\hat{\Pi}_{t,T}^{2}) = \sum_{i=1}^{3} n_i \log\left(\frac{n_i}{n}\right) \ , \tag{4.11}$$

and $LR_{t,T}^{uc} \sim \chi_3^2$ and $LR_{t,T}^{cc} \sim \chi_6^2$ respectively as specified in Christoffersen (1998).

When applying this test, the majority of the literature assumes that the model which produces the lowest log-likelihood ratios is the most accurate model, although this is often not carried out in a statistically meaningful way. In fact, as pointed out by Christoffersen et al. (2001), the log-likelihood ratios are *nested*. To elaborate, consider a one-tail, unconditional test with $LR^{uc}_{t,T} \sim \chi^2_1$ as specified in 4.8. The degrees of freedom of χ^2_1 arises from the difference between the degrees of freedom of the distributions of $-2L(\Pi^0_{t,T})$ and $-2L(\hat{\Pi}^2_{t,T})$. That is, the former term has fewer restrictions than the latter. However, when comparing models we are presented with two log-likelihood statistics whose distributions contain the same degrees of freedom hence the corresponding distribution of the log likelihood ratio is unknown. Christoffersen et al. (2001) outlines the procedures to side step this issue but this is not considered here as the methods are only applicable to parametric models, while we include a semi-parametric model.

Instead, we judge the performance of a model based the consistency of its performance on a 2500-day rolling-window for the whole term structure. Note that EMIR's Article 49 requires backtesting to be carried out *each day* hence our tests also mimic what institutions will go through in practice. To our knowledge, this simple, yet essential extension to the valuation method has not been implemented before in the literature.

In this out-of-sample analysis, we recalibrate the GARCH parameters on a daily basis on both P&L and returns over a one-day horizon ($h = 1$). We do not perform backtesting for longer time horizons as VaR scaling can be highly inaccurate. In addition, calculating VaR on data with lower frequency (even for a two-day horizon) also significantly shortens the out-of-sample period. An accurate examination at 1% or 99% levels however requires a large windows for a healthy amount of expected exceedances.

Finally, we further test the models on their ability to predict the entire distribution via methods introduced by Berkowitz (2001). The intuition here is as follows: denote the estimate of the probability mass and cumulative density functions as $\hat{f}(\cdot)$ and $\hat{F}(\cdot)$, belonging to the process y_t is $x_t = \int_{-\infty}^{y_t} \hat{f}(u)du = \hat{F}(y_t)$; for an accurate forecast, x_t must be *iid* and uniformly distributed on [0,1]. To carry out this test, we compute a forecast of the distribution at every time step and make an *ex post* comparison to the observed returns and P&L's. For each period t, we compute the equivalent $x_{t,T}$ and test whether or not $z_{t,T} = \Phi^{-1}(x_{t,T}) \sim N(0, 1)$, *iid.* This can be achieved using, for example, the Jarque-Bera test for normality, the Durbin-Watson test for autocorrelation and the White test for Heteroskedasticity. In this study, we follow Berkowitz (2001)'s procedure which encompasses all of the above via one log-likelihood statistic. For this, if $z_{t,T}$ follows

$$z_{t,T} - \mu = \rho(z_{t-1,T} - \mu) + \Delta F_{t,T} \ , \tag{4.12}$$

for $z_{t,T} \sim N(0,1)$, *iid*, then one requires $\mu = 0$, $\rho = 0$ and $\sigma^2_{\Delta F_{t,T}} = 1$. The log-likelihood function of this relationship is given by

$$\begin{aligned} L^*(\mu, \sigma^2_{t,T}, \rho) = &-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log[\sigma^2_{t,T}/(1-\rho^2)] - \frac{(z_{1,T} - \mu/(1-\rho))^2}{2\sigma^2_{t,T}/(1-\rho^2)} \\ &- \frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2_{t,T}) - \sum_{t=2}^{T}\left(\frac{(z_{t,T} - \mu - \rho z_{t-1,T})^2}{2\sigma^2_{\Delta F_{t,T}}}\right) \ , \end{aligned} \tag{4.13}$$

where the independence (to order 1) and conditional log-likelihood ratios can be computed using the relationship

$$LR^{in*}_T = -2(\log L^*(\hat{\mu}, \hat{\sigma}^2_{t,T}, 0) - \log L^*(\hat{\mu}, \hat{\sigma}^2_{t,T}, \hat{\rho})) \sim \chi^2_1 \ , \tag{4.14}$$

$$LR^{cc*}_T = -2(\log L^*(0,1,0) - \log L^*(\hat{\mu}, \hat{\sigma}^2_{t,T}, \hat{\rho})) \sim \chi^2_2 \ . \tag{4.15}$$

We carry out this procedure on the whole 20-year period only.

The evaluation process thus follows:

1. We determine whether to use the returns or P&L series, this first study is carried out on constant-maturity series only and the models are evaluated via Christoffersen (1998)'s tests
2. We also determine whether to use constant-maturity or rollover series using Christoffersen (1998)'s tests, this study is based on the best series obtained in step 1)
3. We determine the best VaR model based on the results from steps 1) and 2), using rolling $LR^{cc}_{t,T}$ statistics
4. We verify the estimation of the models in forecasting the entire distribution using Berkowitz (2001)'s tests

4.3 Data

We show summary statistics of Brent crude oil futures in detail as the focal point of our discussion is based on this commodity. We include futures with up to 10 months

Months to mature	$\hat{\mu} \times 100$	$\hat{\sigma}$	$\hat{\tau}$	$\hat{\varkappa}$
			Return-index	
1	0.071	0.020	-0.057	6.082
2	0.066	0.020	-0.043	6.274
3	0.069	0.018	-0.034	6.347
4	0.070	0.018	-0.059	6.311
5	0.068	0.017	-0.068	6.283
6	0.067	0.017	-0.068	6.299
7	0.067	0.016	-0.066	6.299
8	0.066	0.016	-0.063	6.323
9	0.065	0.016	-0.069	6.277
10	0.065	0.016	-0.078	6.218
			P&L-index	
1	2.026	1.160	-0.357	11.468
2	1.974	1.133	-0.348	11.890
3	2.132	1.113	-0.352	12.244
4	2.225	1.093	-0.360	12.598
5	2.241	1.076	-0.367	12.859
6	2.284	1.060	-0.370	13.101
7	2.311	1.046	-0.368	13.327
8	2.299	1.032	-0.374	13.589
9	2.316	1.021	-0.376	13.757
10	2.340	1.010	-0.382	13.926

Table 4.2.: Summary statistics for first 10-month to mature constant-maturity P&L and returns Brent crude oil indices. Total number of observations: 4971. $\hat{\mu}, \hat{\sigma}, \hat{\tau}$ and $\hat{\varkappa}$ denote the mean, standard deviation, skewness and excess kurtosis, respectively. Period 18^{th} April 1994 - 30^{th} December 2013.

to maturity. [1] Our data stretches a approximately 20 years, between: April 1994 - December 2013. We have chosen this starting point since Brent crude oil futures started experiencing higher trading volumes in April 1994.

Rollover	$\hat{\mu} \times 100$	$\hat{\sigma}$	$\hat{\tau}$	$\hat{\varkappa}$
			Returns	
1^{st}	0.079	0.021	-0.020	6.247
2^{nd}	0.066	0.020	-0.050	6.288
3^{rd}	0.067	0.019	-0.015	6.441
4^{th}	0.068	0.018	-0.042	6.388
5^{th}	0.068	0.018	-0.061	6.365
6^{th}	0.067	0.017	-0.065	6.358
7^{th}	0.067	0.017	-0.062	6.340
8^{th}	0.065	0.016	-0.057	6.359
9^{th}	0.064	0.016	-0.063	6.336
10^{th}	0.064	0.016	-0.070	6.259
			P&L	
1^{st}	2.429	1.162	-0.320	11.374
2^{nd}	1.883	1.144	-0.356	11.679
3^{rd}	1.993	1.120	-0.353	12.095
4^{th}	2.103	1.102	-0.361	12.435
5^{th}	2.167	1.082	-0.364	12.774
6^{th}	2.198	1.065	-0.373	13.025
7^{th}	2.257	1.050	-0.372	13.237
8^{th}	2.213	1.037	-0.374	13.477
9^{th}	2.200	1.024	-0.378	13.731
10^{th}	2.204	1.012	-0.379	13.881

Table 4.3.: Summary statistics for Brent crude oil P&L and returns for the first 10-month to mature rollover series. Total number of observations: 4971. $\hat{\mu}, \hat{\sigma}, \hat{\tau}$ and $\hat{\varkappa}$ denote the mean, standard deviation, skewness and excess kurtosis, respectively. Period 18^{th} April 1994 - 30^{th} December 2013.

Months to Mature	EWMAt-6		GJRt		K-60		OEWMAt-6	
	P&L	Return	P&L	Return	P&L	Return	P&L	Return
				$LR^{uc}_{t,T}$				
1	0.87**	1.94**	1.47**	1.80**	2.08**	1.57**	1.36**	1.52**
2	1.14**	0.34**	1.18**	2.23**	1.77**	2.64**	0.87**	0.30**
3	1.09**	1.00**	1.69**	3.30**	1.22**	3.04**	0.87**	0.73**
4	1.46**	1.57**	1.15**	2.11**	2.08**	2.86**	0.87**	1.19**
5	1.00**	2.41**	2.44**	1.51**	3.54**	2.08**	0.64**	1.98**
6	0.87**	1.98**	2.23**	1.70**	4.08**	2.81**	1.00**	3.01**
7	1.57**	2.06**	1.94**	1.87**	3.85**	3.04**	1.59**	3.47**
8	1.68**	2.43**	2.43**	0.51**	1.65**	3.01**	1.94**	4.64*
9	0.73**	3.01**	0.87**	0.78**	0.97**	2.51**	1.94**	3.44**
10	1.15**	3.44**	1.99**	1.58**	1.09**	1.65**	1.69**	4.02**
				$LR^{cc}_{t,T}$				
1	6.00**	6.56**	10.21**	17.94	11.70*	8.30**	6.30**	7.17**
2	6.88**	5.68**	10.50**	15.02	14.16	12.19*	6.00**	6.00**
3	13.13	5.84**	10.91*	16.57	16.70	13.08	11.96*	5.73**
4	12.70	8.66**	9.22**	23.34	16.49	12.56*	11.96*	8.53**
5	8.91**	9.27**	14.04	21.29	12.26*	9.98**	8.61**	8.88**
6	6.17**	8.50**	9.71**	25.38	13.22	9.98**	6.47**	11.93*
7	6.14**	8.39**	10.15**	19.72	10.68*	9.99**	8.72**	12.27*
8	7.06**	11.85*	9.16**	27.34	16.97	10.44**	10.15**	13.06
9	6.43**	11.93*	8.49**	28.07	16.68	10.12**	10.15**	9.98**
10	6.89**	12.44*	9.17**	16.34	16.46	16.34	5.00**	10.32**
				$LR^{in}_{t,T}$				
1	5.13**	4.62**	8.74*	16.14	9.62	6.73**	4.94**	5.65**
2	5.74**	5.34**	9.32*	12.79	12.39	9.55	5.13**	5.70**
3	12.04	4.84**	9.22*	13.27	15.48	10.04	11.09	5.00**
4	11.24	7.09**	8.07*	21.23	14.41	9.70	11.09	7.34**
5	7.91*	6.86**	11.6	19.78	8.72*	7.90*	7.97*	6.90**
6	5.30**	6.52**	7.48**	23.68	9.14*	7.17**	5.47**	8.92*
7	4.57**	6.33**	8.21*	17.85	6.83**	6.95**	7.13**	8.80*
8	5.38**	9.42*	6.73**	26.83	15.32	7.43**	8.21*	8.42*
9	5.70**	8.92*	7.62**	27.29	15.71	7.61**	8.21*	6.54**
10	5.74**	9.00*	7.18**	14.76	15.37	14.69	3.31**	6.30**

Table 4.4.: $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$ and $LR^{in}_{t,T}$ statistics comparison between best performing VaR models (the full selection is shown in the Appendix) at 1% on both tails for the first 10-months-to-mature P&L- and return-index constant-maturity futures. Out-of-sample test period July 1999-December 2013.

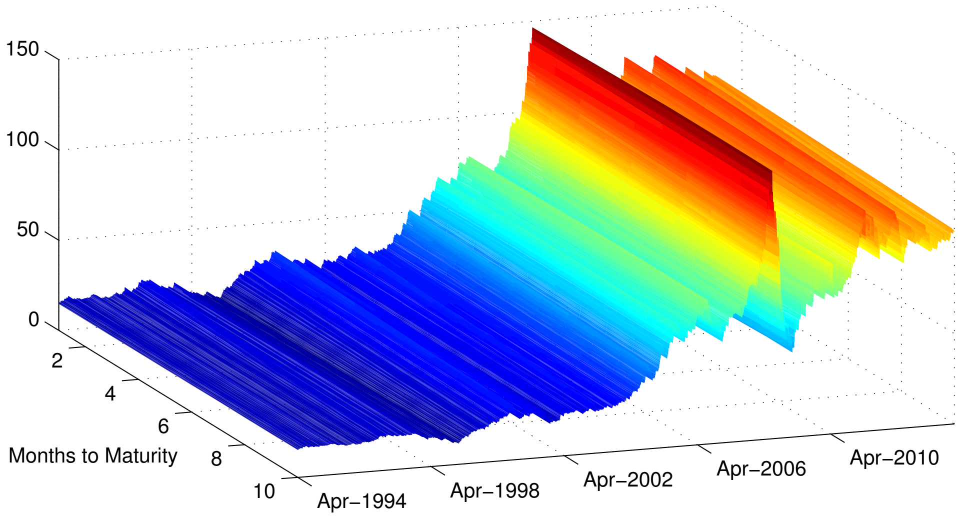

Figure 4.1.: 1 - 10 constant-maturity value-index Brent crude oil futures term structure. Period: 18^{th} April 1994 - 30^{th} December 2013.

4.4 Results and Discussions

4.4.1 Constant-Maturity VaR: Return Versus P&L

We find that the dollar-VaR forecasts based on the return series are sometimes upward sloping which disobeys the Samuelson effect (see figure 4.2 for an illustration using the OEWMAt-6 model). Although VaR estimates on returns are downward sloping, once multiplied by the futures prices which are at times steep in contango due to jumps, the resulting dollar-VaR term structure is upward sloping. This occurred 36 times through the 20 years out-of-sample period, most of which are concentrated between November 2008 to January 2009 due to the world recession in 2008. VaR models based on returns are not sensitive enough to fully capture this change in dynamics and hence produce dollar-VaR term structures which are upward sloping. The VaR calculated using P&L's on the other hand are always downward sloping, regardless of the economic conditions. When calculating margin requirements, investors would expect front month contracts to consistently have higher margin requirements. The violation of this rule can be costly to investors who rollover from one futures contract to the next, indicating that they have to raise more capital for a contract which

[1]We assume that futures further out in the term structure will experience liquidity issues, the corresponding margin requirement will require a premium since the liquidation period may be larger. We do not consider such a case in our work, although this is an interesting point for further study.

should inherently be less risky. Hence, we focus our analysis on using the P&L series only. We would also advise exchanges to favour VaR based on the P&L series as the Samuelson effect is always observable.

From figure 4.2, one may find that VaR based on returns are lower than those generated by the P&L series. This may be an attractive feature for the exchange, as lower VaR term structures will allow for lower margins and consequently lower transaction costs for investors. After further examination on the first to mature series however (see figure 4.3), we find that this is not always the case. In fact, VaR estimates from both series are roughly on par with each other throughout the sample. We find that the VaR generated using the P&L is below its return counterpart 66% of the time. In general, one can be larger than the other for 3-4 months, however these instances are random.

To examine the tests in more detail, we analyse Table 4.4 which shows whole-sample $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998)'s tests for the 99% and 1%, two-tail coverage level on both tails for top-performing models. There is no clear relationship between the log-likelihood ratios and time to maturity, confirming that one cannot generalise results used for the 1-month constant-maturity series on the rest of the term structure. Note that for parametric VaR models, VaR of individual futures series tend to produce lower $LR^{uc}_{t,T}$ and $LR^{cc}_{t,T}$ than their orthogonal counterparts.[2] This is not surprising given that orthogonalisation require estimation of the covariance matrix using a rolling-window which may not be reactive to changing market conditions and hence hinder the performance of the VaR model. The advantage of using orthogonal VaR models is that they are less affected by noise along the term structure and produces smooth-decaying VaR estimates. Although not shown in Table 4.4, at times, performance of the model is more uniform across the term structure (see A.8 and A.9 from the Appendix).

We find that only the EWMAt-6 model produces $LR^{uc}_{t,T}$ and $LR^{cc}_{t,T}$ statistics which are below the critical χ^2_2 and χ^2_6 statistics for both returns and P&L. Note that the GARCH models (with both t and normal distribution innovations) considered in this case tend to perform badly throughout except for the GJRt model estimated on the P&L series which too produces coverage testing statistics under the 95% critical statistic with the exception of the 5-month series. As our work employ GARCH re-calibration on a daily basis, unlike most other works which hold the parameters constant, it is possible that the inaccuracy of the GARCH models in VaR estimation is caused by parameter uncertainty. The high accuracy of the GJRt model using

[2]The full set of results for all models are presented in the Appendix

P&L's may only be specific to this estimation window. Robustness tests are further discussed in the section 4.4.3.

From these tests, in support of the literature (see Su et al. (2011) for example), we find strong evidence against assuming the distributions are normal as this always underestimates the VaR. We find that, with the exception of the GJR model on the 2^{nd}- and 3^{rd}-to-mature series, all four models which assume normal distributions generate inaccurate forecasts at all points along the term structure. Comparing these models' performances from Tables A.4-A.5 to A.2-A.3, we find VaR estimates from both returns and P&L series for non-GARCH models produce $LR^{uc}_{t,T}$, and $LR^{cc}_{t,T}$ statistics which are roughly on par with each other.

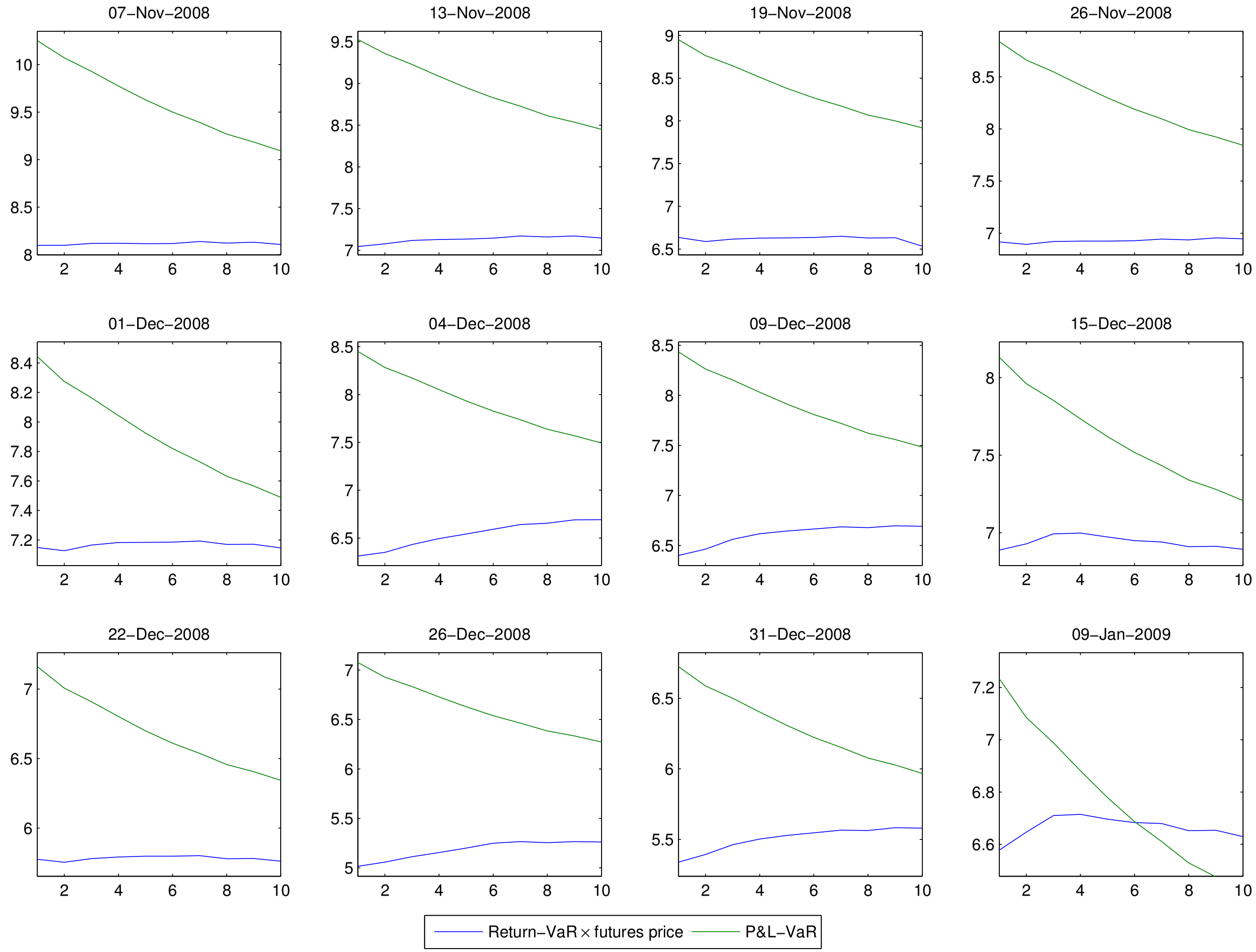

Figure 4.2.: Snapshots of the one-day, 99% OEWMAt-6 dollar-VaR (equivalent to the 1% for a symmetric VaR model) on the term structure constant-maturity return series multiplied by the value-index futures price (green) and that of the 1-month constant-maturity P&L series (blue). Period: November 2008 to January 2009.

4.4.2 Constant-Maturity Versus Rollover VaR

Table 4.5 shows $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$ and $LR^{in}_{t,T}$ calculated using the CM-spline method for constant-maturity futures and rollover futures for the EWMAt-6 and OEWMAt-6 models. The log-likelihood ratios produced by the CM-spline methods are higher than those produced when estimating the VaR using the rollover series, where the $LR^{uc}_{t,T}$ statistics for both models mostly confirm a rejection of the null hypothesis of a good model at 95% for the entire term structure (with the exception of the 7th-to-mature series for the EWMAt-6 model using the CM-spline method). When comparing these results with Tables A.4 and A.5, we also observe that the VaR estimation on constant-maturity P&L's are almost always better than those on the rollover series. It is hence clear that the VaR estimated using the rollover series is superior to those estimated via constant-maturity futures for the purposes of margin requirements. This result however is not generalisable to other risk management practices such as portfolio management, as is demonstrated in Nomikos and Pouliasis (2014) who assumes that investors take positions only on constant-maturity futures.

Thus far, we find that the OEWMAt-6 is the best model for margin requirement purposes given this is the only model which can produce accurate VaR estimates for the most number of points along the term structure. We apply further robustness tests on some of the best performing models to further search for the most suitable VaR model in this case.

4.4.3 Rolling Christoffersen (1998) Log-Likelihood Ratio Statistics

First we observe that the rolling $LR^{in}_{t,T}$ statistic is affected by noise. As exceedances enter the estimation window, the statistic jumps up by roughly 10 points (see Figure 4.4). We find that the jump occurs from the counting convention according to the Christoffersen (1998) methods. Consider a one-tail backtesting scenario with n_1 number of exceedances, n_{11} exceedance clusters, n of observations with the last P&L at time t. If the last P&L value is a VaR exceedance, the counting convention automatically assumes that the P&L at time $t+1$ will not be an exceedance. In other words, the correct specification for this scenario would be to have $n+1$ observations, not n, while all other inputs remain the same. It is in this bias that one witnesses an augmentation in the $LR^{in}_{t,T}$ statistic.

To avoid this problem, one could simply readjust the observations to $n+1$ instead of n in these instances. However, when examining the $LR^{cc}_{t,T}$ statistics on a rolling window basis, the window lengths would vary between n and $n+2$ (should exceedances

	EWMAt-6		OEWMAt-6	
Rollover series	CM-spline	Rollover	CM-spline	Rollover
		$LR^{uc}_{t,T}$		
1^{st}	9.25	2.74**	10.14	2.81**
2^{nd}	9.40	1.56**	11.39	2.60**
3^{rd}	8.54	1.98**	10.28	3.82**
4^{th}	11.65	3.89**	11.19	4.91*
5^{th}	9.60	4.31*	10.97	4.91*
6^{th}	8.55	5.27*	10.05	6.12
7^{th}	6.75*	3.20**	6.75*	4.91**
8^{th}	6.24	1.91**	6.71*	3.52**
9^{th}	9.40	3.71**	8.76	4.91**
10^{th}	6.45	4.91*	7.44	4.78*
		$LR^{cc}_{t,T}$		
1^{st}	12.68	5.06**	12.28*	5.74**
2^{nd}	12.59	4.19**	14.40	7.10**
3^{rd}	12.55*	5.27**	13.39	6.08**
4^{th}	17.55	12.39*	16.78	12.30*
5^{th}	15.27	12.62	18.13	12.30*
6^{th}	13.29	13.23	15.47	13.19
7^{th}	11.30*	8.50**	12.19*	12.19*
8^{th}	11.14*	8.01**	12.22*	12.22*
9^{th}	20.46	21.85	15.15	15.15
10^{th}	15.50	10.58**	12.53**	12.53*
		$LR^{in}_{t,T}$		
1^{st}	3.43**	2.32**	2.14**	2.93**
2^{nd}	3.19**	2.63**	3.01**	4.50**
3^{rd}	4.01**	3.29**	3.11**	2.26**
4^{th}	5.09**	8.50*	5.59**	7.39*
5^{th}	5.97**	8.31*	7.16**	7.39*
6^{th}	4.74**	7.96*	5.42**	7.28*
7^{th}	4.55**	5.30**	3.90**	8.31*
8^{th}	4.90**	6.10**	5.22**	8.70
9^{th}	11.06	18.14	7.94*	10.24
10^{th}	9.05*	5.67**	2.69**	7.75*

Table 4.5.: $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$ and $LR^{in}_{t,T}$ statistics using the CM-spline method and rollover futures P&L series. Models considered EWMAt-6 and OEWMAt-12. Term structure length: constant maturity futures: 1 to 10 months to maturity; rollover - 1^{st} to 10^{th} rollover series. Out-of-sample period: July 1999 - December 2013.

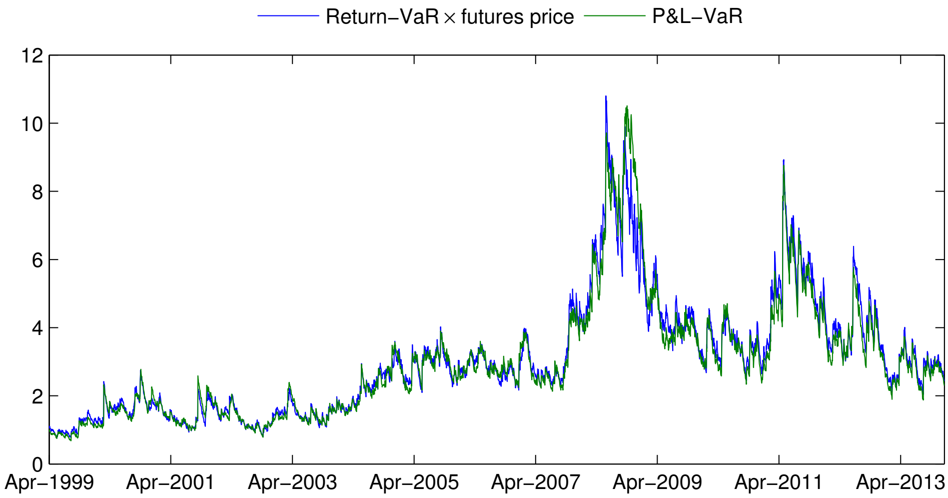

Figure 4.3.: One-day, 99% OEWMAt-6 dollar-VaR (equivalent to the 1% for a symmetric VaR model) on the 1-month constant-maturity return series multiplied by the value-index futures price (green) and that of the 1-month constant-maturity P&L series (blue). Period: April 1999 to December 2013.

appear both at the beginning and at the end of the window) which does not allow for a controlled testing environment. Hence, we use a different approach in avoiding this problem where should exceedances appear at the start or the end of the window, we take the most recent log-likelihood ratios where this is not the case, i.e. should $LR^{cc}_{t,T}$ be affected by starting/ending exceedances, we take $LR^{cc}_{t-1,T}$, $LR^{uc}_{t-1,T}$ and $LR^{in}_{t-1,T}$ instead and should this be affected also, then we take $LR^{cc}_{t-2,T}$, $LR^{uc}_{t-2,T}$, $LR^{in}_{t-2,T}$ and so on.

Figure 4.5 shows rolling $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$ and $LR^{in}_{t,T}$ of the top performing models for the first-to-mature P&L series, cleaned as aforementioned. From the selection of models considered, we find 4 models which are consistently accurate. These are EWMAt-6, GJRt, K-60 and OEWMAt-6 (see figure 4.5 for more details). We observe that $LR^{uc}_{t,T}$ statistics display step functions where jumps occur when exceedances enter and leave the test window. Now we examine rollover $LR^{cc}_{t,T}$ statistics for the term structure, we find that the OEWMAt-6 model remains the best model (see figure 4.6, all other models' performance can be found in the Appendix) as this is accurate for over 94% of all instances along the term structure *and* out-of-sample periods, as the log-likelihood ratios are rolled over. All other rollover $LR^{cc}_{t,T}$ surfaces and the number of instances above the 95%, χ^2_6 critical statistic can be found on Figure A.1 and Table A.1 the Appendix.

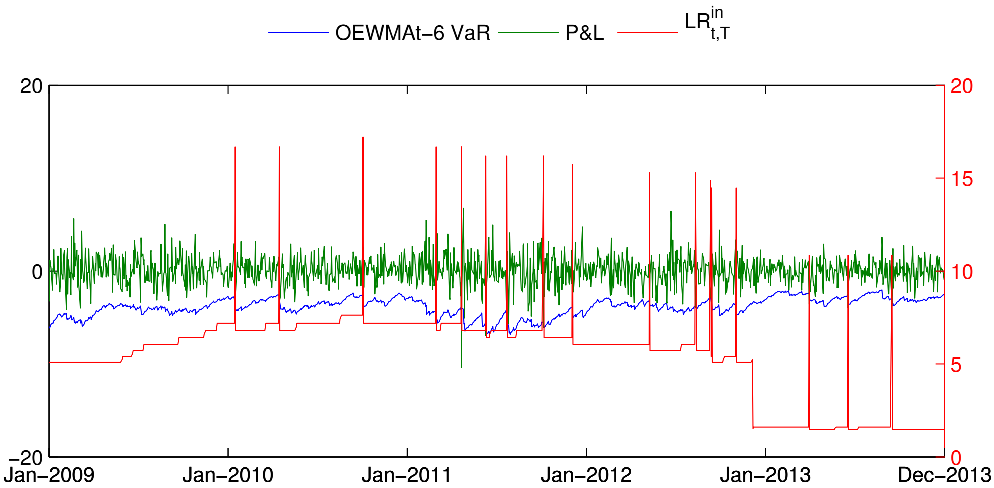

Figure 4.4.: $LR^{in}_{t,T}$ statistics based for the first-to-mature Brent crude oil P&L series, with corresponding OEWMAt-6 VaR and P&L. Period: 23^{rd} January 2009 - 30^{th} December 2013.

4.4.4 Robustness Test: Berkowitz (2001)

Figures 4.6 and 4.7 shows LR^{cc*}_T, LR^{in*}_T based on Berkowitz (2001) coverage tests. We find that for all instances along the term structure, none of the models are able to forecast the entire distribution of futures P&L well. To our surprise, the Kernel-fitted distribution forecast models performed badly, although these are more flexible in terms of the shape of the probability distribution functions. We observe that the kernel-fitted VaR model with a 120-days rolling window performed worse than that of the 60-day rolling window, indicating that even with a short window of 120-days, the sensitivity to the market movements is insufficient, placing too much importance to data too far in the past. The main source of error in this case is possibly due to the fact that each window does not contain enough data to produce a well-defined distribution. This is all the more surprising given that the Kernel-VaR models performed well on the Christoffersen (1998) coverage tests.

We also observe that most models' performances tend to degrade with increasing time-to-maturity. This is perhaps due to increasing excess kurtosis along the P&L term structure. Note that at $\nu = 6$, the excess kurtosis is 3, which compared to the P&L's summary statistics, with excess kurtosis roughly equal to 13, is inadequate. Hence as the excess kurtosis increases beyond this point, so too does the inaccuracy of the models. This is confirmed by LR^{cc*}_T statistics of the GJR-t model where ν is now

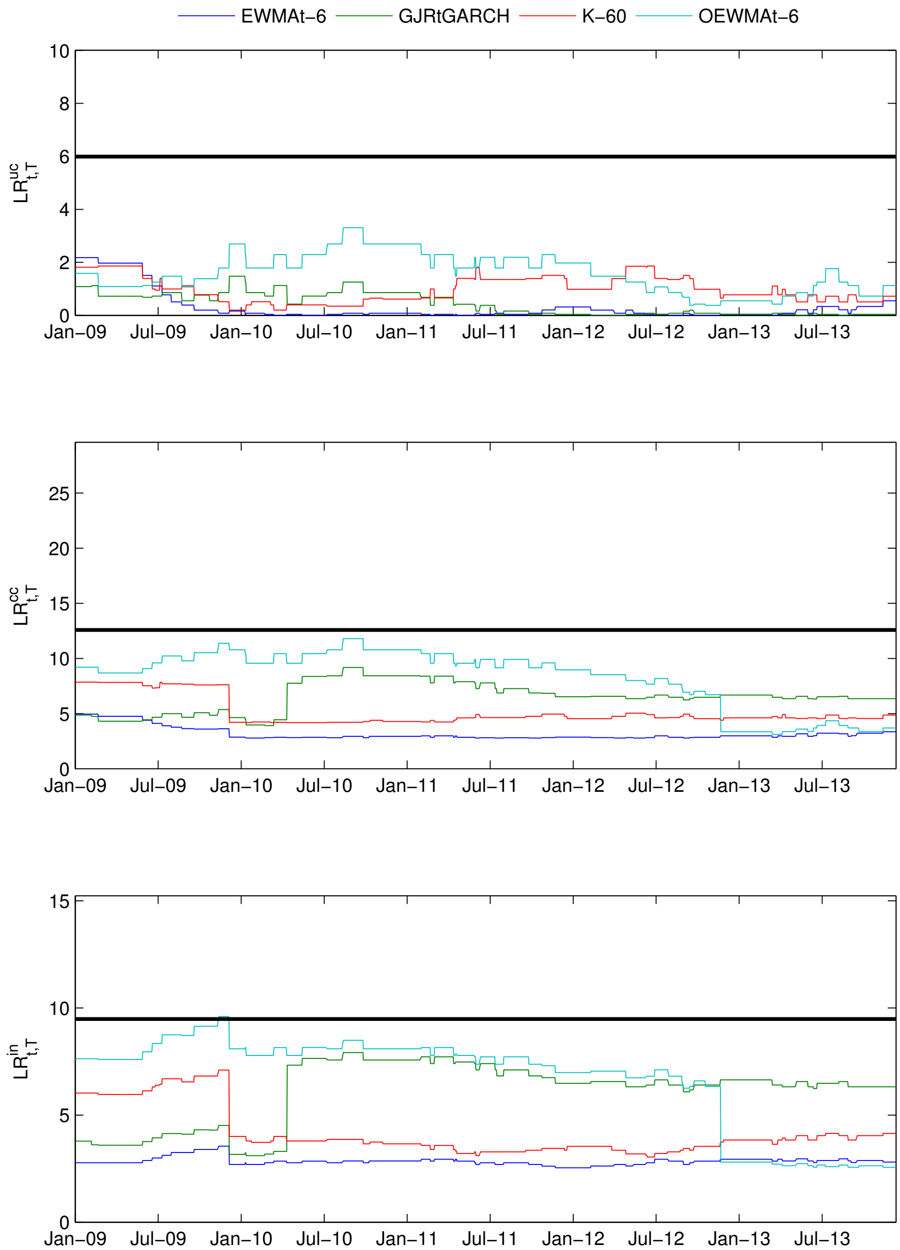

Figure 4.5.: Top: rolling $LR^{uc}_{t,T}$, middle: $LR^{cc}_{t,T}$, bottom: rolling $LR^{in}_{t,T}$ for top performing models on the first-to-mature P&L series (EWMAt-6, GJRt, K-60,k-120 and OEWMAt-6). Coverage test levels at 0.01 and 0.99 simultaneously. Horizontal lines indicate the critical 95% χ^2_2, χ^2_6, χ^2_4 (from top to bottom). Includes 1274 sets of 2500 days out-of-sample periods: rolling from 25th March 1999 -26th January 2009 to 8th April 2004 - 26th December 2013.

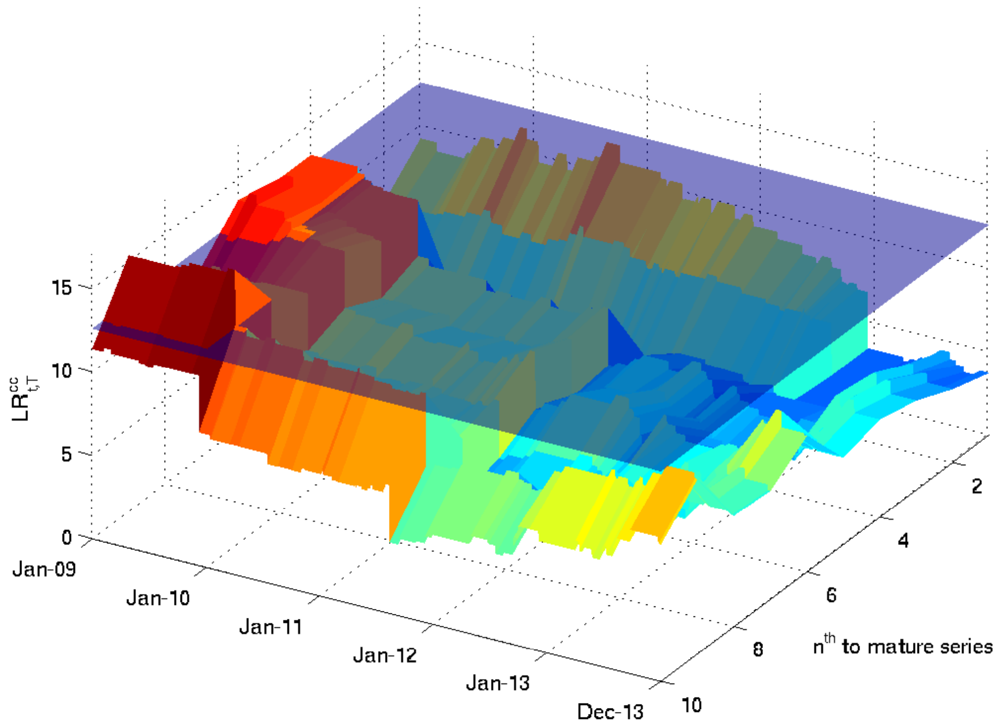

Figure 4.6.: Rolling $LR^{cc}_{t,T}$ statistics for the same model. The transparent surface represents the 95% critical statistics for χ^2_6 distributions for the conditional test.

Rollover series	EWMAt-6	EWMAt-12	EWMA	GJRt	GJR
			LR_T^{cc*}		
1^{st}	15.68	14.48	28.97	51.18	19.10
2^{nd}	15.05	13.93	27.03	45.83	19.17
3^{rd}	14.04	12.61	26.65	47.08	17.31
4^{th}	16.20	14.45	28.70	49.17	20.39
5^{th}	15.47	13.86	28.22	49.71	18.90
6^{th}	15.86	14.35	28.79	42.38	18.68
7^{th}	17.85	16.27	30.41	46.64	19.74
8^{th}	21.02	19.39	32.81	45.30	21.77
9^{th}	21.71	19.79	33.42	49.28	21.16
10^{th}	25.31	23.27	37.05	52.22	26.40
			LR_T^{in*}		
1^{st}	5.02	4.77	4.28	5.79	4.11
2^{nd}	5.65	5.28	4.60	5.98	4.24
3^{rd}	4.97	4.60	3.89	5.35	3.95
4^{th}	6.62	6.15	5.24	6.93	5.35
5^{th}	6.58	6.13	5.26	6.85	5.35
6^{th}	7.06	6.62	5.75	8.08	5.70
7^{th}	9.07	8.56	7.53	9.33	7.29
8^{th}	12.08	11.46	10.17	11.74	9.72
9^{th}	12.92	12.23	10.76	13.171	10.38
10^{th}	16.22	15.54	14.02	15.92	14.38

Table 4.6.: Whole-sample, two-tailed, 1% LR_T^{cc*}, LR_T^{in*} based on Berkowitz (2001) coverage tests on the Brent futures P&L series for the first 10-to-mature series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29th March 1999 - 30th December 2013. Starting and final P&L values are not exceedances of the VaR. Models considered: EWMAt-6, EWMAt-12, EWMA, GJRt, GJR.

Rollover series	K-60	K-120	OEWMAt-6	OEWMAt-12	OEWMA	OGJR
				LR_T^{cc*}		
1^{st}	711.39	780.50	24.67	42.19	13.89	10.98
2^{nd}	489.24	540.05	18.19	31.72	19.12	15.18
3^{rd}	327.08	361.53	14.10	24.26	25.39	19.87
4^{th}	226.65	244.65	14.52	22.12	33.70	27.75
5^{th}	142.25	155.79	13.38	19.82	37.27	32.65
6^{th}	83.90	95.74	13.83	19.80	39.35	36.03
7^{th}	51.35	60.12	15.93	21.75	41.46	39.27
8^{th}	32.43	39.61	19.37	25.73	41.90	42.03
9^{th}	23.51	28.30	20.23	27.05	40.48	41.78
10^{th}	26.09	28.36	24.32	31.71	42.85	45.37
				LR_T^{in*}		
1^{st}	7.75	8.86	4.77	4.52	4.11	5.14
2^{nd}	10.41	11.43	5.30	4.85	4.20	5.09
3^{rd}	10.47	10.90	4.69	4.22	3.54*	4.20
4^{th}	14.26	14.76	6.52	5.96	5.10	6.08
5^{th}	14.69	15.58	6.68	6.14	5.30	6.66
6^{th}	15.98	17.27	7.41	6.90	6.09	7.78
7^{th}	18.46	19.70	9.45	8.90	7.99	10.07
8^{th}	21.91	23.50	12.47	11.81	10.68	13.48
9^{th}	22.15	23.49	13.13	12.34	11.02	14.13
10^{th}	25.75	28.13	16.64	15.91	14.62	18.32

Table 4.7.: Whole-sample LR_T^{cc*}, LR_T^{in*} based on Berkowitz (2001) coverage tests on the Brent futures P&L series for the first 10-to-mature series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29^{th} March 1999 - 30^{th} December 2013. Starting and final P&L values are not exceedances of the VaR. Models considered: K-60, K-120, OEWMAt-6, OEWMAt-12, OEWMA, OGJR.

calibrated. Although the performance of this model is much worse than the EWMA models, the performance is much more uniform along the term structure.

Perhaps more complex models, such as regime-switching GARCH will flourish when forecasting the entire distribution. For the purpose of margin requirement however, the simple OEWMAt-6 model is adequate and is thus selected for further analysis in the next chapter.

4.5 Conclusions

In this study, we have uncovered a number of issues for testing VaR using the traditional Christoffersen (1998) coverage tests. First, we support the use of the P&L series over returns; although VaR performance between the two tend to be similar, as P&L-based VaR consistently produces downward-sloping dollar-VaR series, margins based on such a model would also be downward sloping. Investors will not have to pay more for contracts which should inherently be less volatile. Second, although the use of constant-maturity data has been gaining momentum in the VaR literature (see Nomikos and Pouliasis (2014) for example), we find that this is not appropriate for margin requirements given that the extrapolation of the VaR term structure hinders the accuracy of the forecast.

Of the 11 VaR models considered, for Brent crude oil futures rollover P&L's, we find the OEWMAt-6 to be the best model, where it produces $LR^{cc}_{t,T}$ and $LR^{uc}_{t,T}$ statistics on the 2-tail estimation which fails to reject the null hypothesis of an accurate model at 95% over 94% of all instances. This model is hence selected for setting margin requirements in the next chapter.

Finally, we would like to raise attention to the current trend of VaR modelling, especially in the growing complexity of the models presented in current literature. We have thoroughly shown, with results more robust than any current work known, that the simplest VaR models perform just as well, if not better.

5. Optimality Criteria and Rules for Brent Crude Oil Futures Margin Requirements

5.1 Introduction

The introduction of EMIR (European Union (2013)) limitations to margin requirements for clearing houses in 2012 raises concerns over the coverage levels and stability of margin requirements. The most prevalent method for setting margin requirement in the industry - Standard Portfolio Analysis of Risk (SPAN) produces margins whose coverage levels are inadequate. The literature on margin requirements is extensive, we discover two main branches including: the efficient contract design, pioneered by Brennan (1986) and prudential margin requirements, pioneered by Booth et al. (1997). Although their intuitions vary, their objectives are essentially the same - to determine the ideal the coverage levels for margin requirements. We find however, that the coverage level itself is volatile. A margin that is exactly equal to such a level is inapplicable in practice as investors would face inappropriately high risks of margin calls. Both branches do not address the need for stable margin requirements which is one of the focal points of EMIR.

Our work is related to a much smaller branch of the literature, which concentrates on formulating optimal decision processes to produce stable margin requirements. The most closely related work, Lam et al. (2010) presents methods which are outdated and requires further improvements to allow for margins which adhere to the news laws. Our main contributions in this chapter are: 1) we introduce optimality criteria which concentrate on achieving stable margin requirements following EMIR's Article 28; 2) using simple parameterisation, we mathematically describe how clearing houses can change margins and 3) we provide methods for assessing margin stability.

The rest of the chapter is laid out as follows: first, we present the methodology which includes the new criteria, rules for how margins can change with time, optimisation methods and out-of-sample testing procedures. We then show the calibration results and discuss each method.

5.2 Methodology

We first argue what the optimal margin level should be and formulate criteria surrounding our reasoning. We then derive heuristic rules to fit the criteria introduced. Our arguments are developed not only after we examine the literature but also through observing historical Brent crude oil margins issued by the Intercontinental Exchange (ICE). Hence, we first present and analyse SPAN margins, which eventually leads to the rationales for the optimality criteria and margin rules.

5.2.1 Data: The SPAN Term Structure

Summary statistics of Brent crude oil futures used in this chapter can be found in Table 4.3 in chapter 4. Value-at-Risk (VaR) and margin estimations do not take account of Brent futures P&L spikes on the 5^{th} June 2011 and 29^{th} June 2012 as these were caused by rare events which only affected the market temporarily, see `http://www.reuters.com/article/2012/06/29/us-markets-oil-idUSBRE83H17O20120629` and `http://www.reuters.com/article/2011/05/05/us-markets-oil-idUSTRE7446BH20110505` for news stories linking to each events. SPAN Brent crude oil initial margins for members/hedgers are computed using the SPAN software, hereafter denoted $M^S_{t,T}$. The corresponding margin estimates for the first 10-to-mature series are shown below:

5.2.2 Margin Optimality Criteria and Motivations

First, consider EMIR's Article 24 which identifies the required coverage level of the margins:

> *For the calculation of initial margins the CCP shall at least respect the following confidence intervals: (a) for OTC derivatives, 99,5 %;*
>
> *(b) for financial instruments other than OTC derivatives, 99 %.*

where the time horizon is defined in Article 26:

> *A CCP shall define the time horizons for the liquidation period taking into account the characteristics of the financial instrument cleared, the market where it is traded, and the period for the calculation and collection of the margins. These liquidation periods shall be at least:*

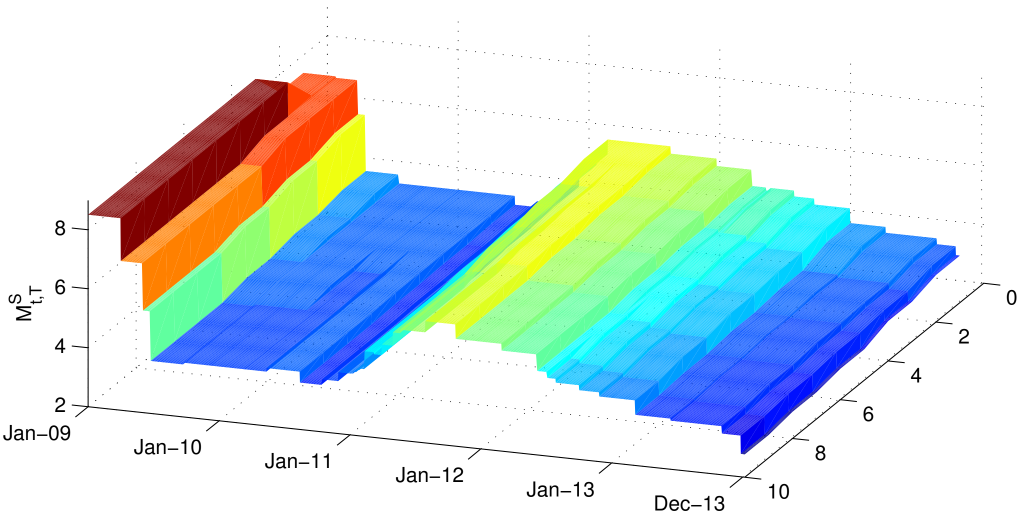

Figure 5.1.: The first 10-to-mature SPAN Brent crude oil futures initial margins for members/hedgers in USD per barrel. Period: 1^{st} July 2009 - 30^{th} December 2013.

(a) five business days for OTC derivatives;

(b) two business days for financial instruments other than OTC derivatives.

Brent crude oil futures is an exchange-traded product with high trading volume. The required liquidation period and coverage level according to this Article is the two-day, 99% horizon. We contemplate, however that this may be too high due to two reasons.

First, taking a two-day time horizon would require forecasting two days in advance which can be inaccurate given that the one-day VaR needs to be scaled to a two-day VaR. We do not encourage finding the VaR on data with greater horizons as this can greatly shortens the out of sample period, even for a two-day horizon.[1] Second, ICE's margin requirements have historically been on par with the one-day VaR but fails to cover $VaR^{0.99,2}_{t,T}$ the majority of the time. To confirm this, we examine the historical SPAN margins, the one-day and two-day 99% VaR of the front-month Brent crude oil futures using the Orthogonal EWMA method, assuming student-t distribution innovations with 6 degrees of freedom (hereafter denoted $M^S_{t,T}$, $VaR^{0.99,1}_{t,T}$ and $VaR^{0.99,2}_{t,T}$ for a futures contract expiring at time T respectively).[2], see Figure 5.2. Exchanges

[1] The one-day ahead VaR also has a concise formulation, while the two-day forecast however is harder to obtain.

[2] Following the results from chapter 4, we calculate the margin requirements using the Orthogonal EWMAt-6 VaR and decay factor 0.94 based on the rollover P&L series.

(and its clearing houses within) rely on high trading volume to attract profit and competition on costs is growing as more exchanges merge into huge conglomerates - reducing bid-ask spreads and margins is more important today than ever before. Such strain of competition (see Abruzzo and Park (2013)) has caused the CME to calculate VaR using roughly the same time horizon. While Dodd-Frank regulations do not prohibit the CME from using a one-day VaR, should ICE enforce the two-day VaR recommended in Article 26, they would fail to generate enough profit. One solution to this is to encourage international regulating bodies to work together when imposing margin requirements. Santos and Scheinkman (2001) suggests competition between exchanges lead to lower margin levels which can help sustain fruitful market liquidity. Providing a platform for exchange competition would be beneficial to the financial markets worldwide. Another solution is to follow Duffie and Zhu (2011)'s suggestions of unionising central clearing for different classes of derivatives to reduce counterparty risk. The element of competition is completely removed hence allowing clearing houses to survive under different regulations. Such regulatory issues present another interesting case for discussion. In this study, we assume that there exist an environment suitable for competition, where EMIR has allowed for the margin to cover *at least* $VaR_{t,T}^{0.99,1}$. We base our optimality criteria (and subsequent margin rules) on $VaR_{t,T}^{0.99,1}$.

We now address Articles relating to margin stability. First, it is important to distinguish between margins for leveraging stocks and futures as these may have different effects on procyclicality. For stocks, the margin account is composed purely of the investor's equity/debt levels; both the initial and maintenance margins are set as gearing ratios denominated in percentages. Historically, stock margin levels have fluctuated wildly on a year-to-year basis; changing to/from 50 - 100% at times, see Largay and West (1973) and Eckardt and Rogoff (1976) for example. Such extreme movements are not possible for futures margins, since these are obligatory, and market volatility is hence presumed to be more sensitive to changes in margin. While regulations pay particular attention to limiting procyclicality, the effects of changes in margin requirements on subsequent market conditions is difficult to analyse because one requires a large sample of margins, which is hard to find. For works which suggest margin changes are procyclical see Telser (1981), Hsieh and Miller (1990), Kupiec (1993).[3] Hardouvelis (1990) is the only work to find margins are countercyclical while others such as Kumar et al. (1991), Day and Lewis (1997) and Phylaktis and Aristidou (2013) finds margin changes do not exacerbate market volatility. However,

[3]Defined differently for short and long positions. Note that when holding a short position, the investor must leverage their position and is hence always susceptible to margin calls.

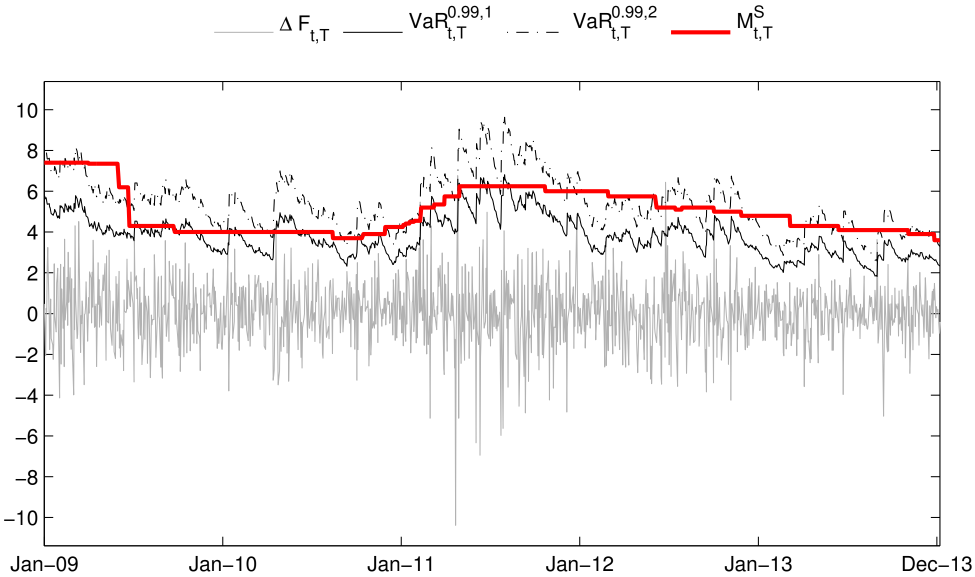

Figure 5.2.: Margin requirements, $VaR_{t,T}^{0.99,1}$, $VaR_{t,T}^{0.99,2}$ and $M_{t,T}^{S}$ of first-to-mature Brent crude oil futures. Period 28^{th} January 2009 to 26^{th} December 2013. VaR calculation method: Orthogonal EWMA, using student-t distribution innovations with 6 degrees of freedom.

it is highly likely that such margin changes were carried out cautiously and have not significantly contributed to procyclicality; whether or not these were set by the Feds or the exchanges themselves.

In this study, we concentrate on futures margining. Let us assume the simplest setting for the margin requirements, i.e. equal to $VaR_{t,T}^{0.99,1}$. Here, the margin would change daily as $VaR_{t,T}^{0.99,1}$ does, which is not applicable in practice. While daily margins would allow for smaller step changes and better reflections of current market volatility, we find two arguments against such a practice. First, exchanges announce margin changes 2 or 3 business days in advance which allows enough time for investors to raise the needed capital. Should the margin change on a daily basis, it would not reflect current market volatility but that from 2 or 3 days in the past. Second, as found in Brunnermeier and Pedersen (2009), where stock margins are set equal to a pseudo-ARCH-type VaR, increases in margins can cause increases in volatility which in turn increases margins further. The series of knock-on effects can cause liquidity to completely dry out.[4] We discourage implementing daily margin changes for these

[4] We also find that much of the asset-pricing literature employs the margin requirement to explain investors behaviour. Garleanu and Pedersen (2011) for example incorporate the margin to the CAPM to explain asset returns behaviour. By describing the margin requirement in a practical

reasons.

We observe that from the investor's point of view, margin falls are always beneficial while margin rises can be detrimental. To understand this, consider a scenario where an investor who is long 1 futures contract with initial/maintenance margins 100/90 USD. Say the exchange announces that initial/maintenance margin will decrease to 90/80 USD in the next 3 business days. The investor now has a choice of:

1. Do nothing
2. Wait 3 days, offset and immediately retake the position to obtain 10 USD back from the margin for reinvestment

If the reinvestment rate overrides the rebalancing costs, then the latter is preferable and will always be carried out by the investor; a margin decrease is thus always beneficial to investors. Now consider a margin increase from 100/90 USD to 110/100 USD. The choices now comprise of:

1. Do nothing
2. Raise capital to cover a possible margin call
3. Clear positions

Although it is clear that strategies (2) and (3) are preferable in reducing the risk of margin call, it is conceivable that investors may adopt strategy (1) due to higher risk tolerance. In other words, should margins change on a daily basis -as $VaR_{t,T}^{0.99,1}$ does- the investor would always face the risk of either: not knowing how much capital to raise for the next increase in margin; or bear the margin call risk which comprises of both the volatility of the futures and the volatility of the VaR.[5] It is clear that, variable margins lead to higher the risk faced by investors. Therefore, it is not surprising that regulators prefer stable margins, as a significant rise can cause investors to either pull out of the market or borrow more cash on a mass scale; both of which may lead to procyclicality. This *asymmetry* is also reflected in EMIR's Article 28:

> *1. A CCP shall ensure that its policy for selecting and revising the confidence interval, the liquidation period and the lookback period deliver forward looking, stable and prudent margin requirements that limit procyclicality to the extent that the soundness and financial security of the CCP is not negatively affected. This shall include avoiding when possible disruptive or big step changes in margin requirements and establishing transparent and predictable procedures for adjusting margin requirements in re-*

way, our model can enhance the factualness of such works.

[5] Also discussed in chapter 3

> *sponse to changing market conditions. In doing so, the CCP shall employ at least one of the following options:*
>
> *(a) applying a margin buffer at least equal to 25 % of the calculated margins which it allows to be temporarily exhausted in periods where calculated margin requirements are rising significantly;*
>
> *(b) assigning at least 25 % weight to stressed observations in the lookback period calculated in accordance with Article 26;*
>
> *(c) ensuring that its margin requirements are not lower than those that would be calculated using volatility estimated over a 10 year historical lookback period.*

Options *b*) and *c*) are not being considered in this study although this provides an interesting case for further investigation. Article 28 entails *at least* 1.25 times the margin level which can be exhausted to avoid significant margin rises, *not falls.* Contrary to this, we argue that margin falls should also be controlled. Our intuition is as follows:

Referring to Figure 5.2, we observe that $VaR_{t,T}^{0.99,1}$ rises and falls in cycles, which may be due to a number of reasons: Brent crude oil volatility is subjected to economic cycles, seasonal demand patterns and the Samuelson effect; crude oil prices are sensitive to political unrest throughout the world, hence P&L spikes are common. Should margins decrease too quickly in less volatile periods, a possible proceeding jump in the volatility can cause a large step in the margin to follow. In other words, *uncontrolled margin falls may lead to larger, more frequent margin rises.* This can also help avoid sudden increases in trading volume which can provoke procyclicality, see Kupiec and Sharpe (1991) and Coen-Pirani (2005) who also find decreases in stock margins leads to increases in trading volume. Hence, we build our criteria to penalise both directions of margin movements.

For a margin $M_{t,T}$ corresponding to the futures contract $F_{t,T}$, the most desirable margin movement should comprise of the smallest changes possible whose occurrences are as infrequent as possible. The corresponding criteria are hence 1) to minimise the the average margin step size throughout the term structure

$$MS_t = \frac{1}{n_T n_s} \sum_{k=1}^{n_T} \sum_{i=t-n+1}^{t} |\Delta M_{i,T_k}| \, , \tag{5.1}$$

where n_s is the number of margin changes over the time period $t-n+1$ to t , n_T is the number of futures series in the term structure and 2) to maximise the average time between margin changes throughout the term structure

$$MT_t = \frac{1}{n_T(n_s-1)} \sum_{k=1}^{n_T} \sum_{j=2}^{n_s} t^*_{j,T_k} - t^*_{j-1,T_k} \ , \tag{5.2}$$

where

$$t^*_{j,T} = \begin{cases} t, & \text{for } |\Delta M_{t,T}| > 0 \\ 0 & \text{otherwise .} \end{cases} \tag{5.3}$$

Purely controlling for the above criteria however would generate margins which are as high as possible at all times. Although this may be prudent, this is not implementable as exchanges compete with each other and prefer low margin levels to reduce investment costs for investors.[6] Hence, we introduce one further criterion to keep the margin level at bay.

Assuming that one adds on the 25% buffer and hence the margin is now set equal to $1.25VaR^{0.99,1}_{t,T}$. Margins above/below this value would indicate that the buffer is being added/exhausted. The optimal margin level, ignoring stability is when the buffer is not utilised at all, i.e. equal to $1.25VaR^{0.99,1}_{t,T}$. Hence, our final criterion is to minimise absolute margin deviations from this level,

$$MD_t = \frac{1}{n_T n} \sum_{k=1}^{n_T} \sum_{i=t-n+1}^{t} |M_{i,T} - 1.25VaR^{0.99,1}_{i,T}| \ , \tag{5.4}$$

where n is the number of observations in the estimation window. This should allow margins to fluctuate within the vicinity of $1.25VaR^{0.99,1}_{t,T}$. We also find that the first-to-mature SPAN series is roughly on par with this level with $MD_t = 0.840$ USD/barrel.

5.2.3 Margin Rules

We formulate the exchange's decisions when changing margin requirements using margin rules. From hereafter, we present the methods for computing margin requirements based on the long position VaR only; all methods are also applicable to the short position margin which is based on the short position VaR. As VaR is always expressed as a positive number, we express the α% VaR as the absolute value of the α% quantile of the P&L distribution. The 99% short position margin is hence analogous to the long position $|VaR^{0.01,1}_{t,T}|$ (see more details in chapter 3). We consider 4 different rules

[6] We previously discussed possible extension from previous literature in chapter 2, namely Lam et al. (2004) who addresses the same problem. We find that their solution is no longer applicable as this will produce margin levels that are too low comparative to $VaR^{0.99,1}_{t,T}$.

in total, presented here in increasing order of flexibility.

The first rule is where the time t margin requirement for a futures contract with maturity T is set equal to 1.25 times $VaR_{t,T}^{0.99,1}$ (hereafter denoted $1.25VaR_{t,T}^{0.99,1}$) following EMIR's Article 28, option a). Note that, this model produces daily margin changes, i.e. the maximum frequency and the 25% margin buffer is never exhausted.

Next, we alleviate the restriction on the margin buffer so that it can be added/removed to maintain stability. Of course, we restrict the margin to remain above $VaR_{t,T}^{0.99,1}$ following our suggestions for amendments to EMIR's Article 25. Here, the margin remains at the same level, until it breaches $\pm 15\%$ of $1.25VaR_{t,T}^{0.99,1}$, where it re-adjusts to $1.25VaR_{t,T}^{0.99,1}$. The $\pm 15\%$ band is also found in Chiu et al. (2006), and the corresponding margin $M_{t,T}^{C}$ can be described by the relationship

$$M_{t,T}^{C} = \begin{cases} M_{t-1,T}^{C}, & \text{for } 1.1VaR_{t,T}^{0.99,1} < M_{t-1,T}^{C} < 1.4VaR_{t,T}^{0.99,1} \\ 1.25VaR_{t,T}^{0.99,1} & \text{otherwise} . \end{cases} \tag{5.5}$$

We include this model as *ad hoc* parameters may avoid parameter uncertainty and generate more stable margin requirements in an out-of-sample test. In reality the margin may start at any point between $1.1VaR_{t,T}^{0.99,1}$ and $1.4VaR_{t,T}^{0.99,1}$. In this study, we apply $M_{t,T}^{C} = 1.25VaR_{t,T}^{0.99,1}$ as a starting point for all in-sample period calibrations.

For the third rule, we further alleviate the above by allowing the margin thresholds to be calibrated according to the criteria introduced in section 5.2.2. The idea of calibrating margin thresholds is also seen in Lam et al. (2010). The margin under this rule, $M_{t,T}^{L}$ follows:

$$M_{t,T}^{L} = \begin{cases} M_{t-1,T}^{L}, & \text{for } (1.25-\beta_1^L)VaR_{t,T}^{0.99,1} < M_{t-1,T}^{L} < VaR_{t,T}^{0.99,1}(1.25+\beta_2^L) \\ 1.25VaR_{t,T}^{0.99,1} & \text{otherwise} , \end{cases} \tag{5.6}$$

where β_1^L,β_2^L are parameters to be estimated. Unlike Lam et al. (2010), we also allow the thresholds above and below the margin to take different values.

We introduce one further rule, which does not build upon any previous models. Our aim is to generate margins which can reduce cautiously so it is unlikely to be affected by $VaR_{t,T}^{0.99,1}$ jumps. Here, the margin $M_{t,T}^{*}$ for a futures contract expiring at time T can be described by the relationship

$$M^*_{t,T} = \begin{cases} M^*_{t-1,T}, & \text{for } VaR^{0.99,1}_{t,T} < M^*_{t-1,T} < VaR^{0.99,1}_{t,T}\beta_1 \ , \\ VaR^{0.99,1}_{t,T}(1+\beta_2) & \text{for } M^*_{t-1,T} < VaR^{0.99,1}_{t,T} \\ VaR^{0.99,1}_{t,T}(\beta_1-\beta_2) & \text{otherwise} \ . \end{cases} \tag{5.7}$$

The second line of the rule represents downward movements while the third line represents the upward movements. The intuition behind this rule is as follows

- For an upward movement:
    1. The margin increases when it is below $VaR^{0.99,1}_{t,T}$
    2. β_2 governs how much further above the $VaR^{0.99,1}_{t,T}$ the margin increases to
- For a downward movement
    1. The margin decreases when it is above $\beta_1 VaR^{0.99,1}_{t,T}$
    2. β_2 governs how much further below $\beta_1 VaR^{0.99,1}_{t,T}$ the margin decreases to

In other words, low β_2 allows for the margin to increase and decrease cautiously while β_1 governs the upper bound to which the margin is allowed to rise above $VaR^{0.99,1}_{t,T}$. It is not surprising to find that historically, $M^S_{t,T}$ also operate in a similar manner. In fact, $M^S_{t,T}$ have been decreasing in small steps from June 2011 to May 2013.

5.2.4 Margin Term Structures

Here we present detailed analysis of SPAN, paying particular attention to its movements along the term structure. This ultimately leads to the rationale on how margins should decrease along the term structure. First, we observe that as of July 2011, $M^S_{t,T}$ decreases every two contracts. By utilising the OEWMAt-6 method on the P&L's (also defined in the previous chapter), we achieve a smooth-decaying term structure of $VaR^{0.99,1}_{t,T}$. In this case, $VaR^{0.99,1}_{t,T_1} > VaR^{0.99,1}_{t,T_2}$ for $T_1 < T_2$, hence setting margins of a futures contract M_{t,T_2} as high as M_{t,T_1} will always allow for adequate coverage of the 99% movement interval (see Figure 5.3 for snapshots of the SPAN margin and $VaR^{0.99,1}_{t,T_1}$ term structures). Hereafter, we denote T_i as the time to maturity of the i^{th}-to-mature futures.

We observe that M^S_{t,T_i}, for $i = 2$ displays step patterns from June 2011 onwards (see Figure 5.5). This is a result of SPAN margins increasing at roughly 45-50 days to expiry (see Figure 5.4 for some examples). As Brent crude oil futures contracts expire monthly, i.e. the rollover occurs roughly every 30 days, the first-to-mature series's time to expiry is always between 30 and 0 days and is therefore not affected. The

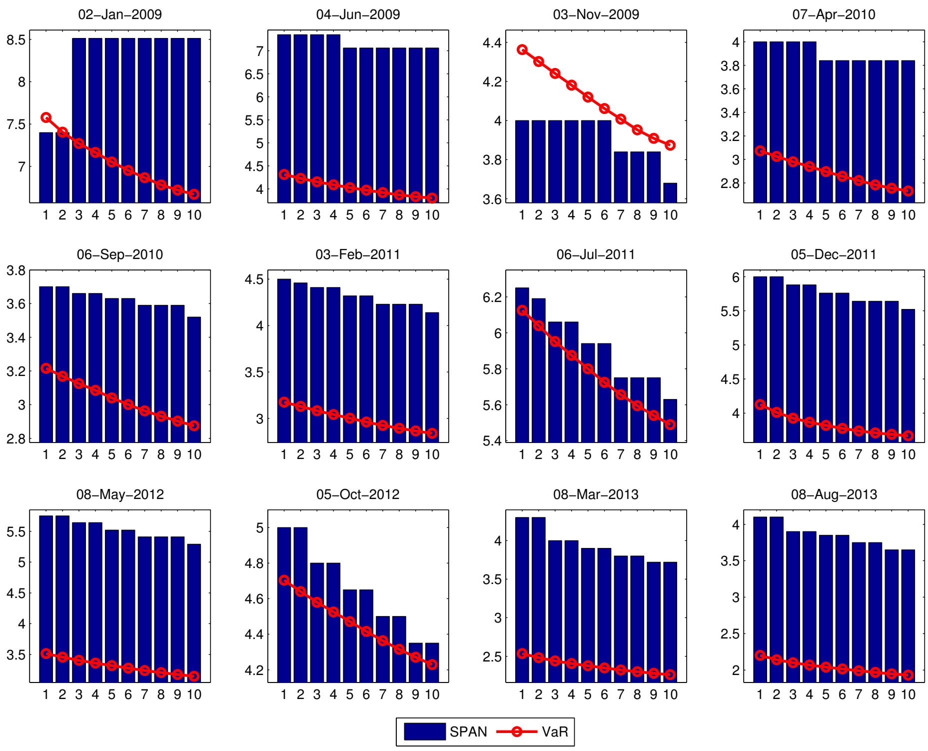

Figure 5.3.: Snapshots of the margin requirements and $VaR_{t,T}^{0.99,1}$ of Brent crude oil futures term structure. Period 28^{th} January 2009 to 26^{th} December 2013. VaR calculation method: Orthogonal EWMA, using student-t distribution innovations with 6 degrees of freedom.

second-to-mature series on the other hand has maturities between 60 and 30 days. As the contract's time to maturity decreases, it experiences a rise in the margin at roughly 45-50 days to expiry. Once the contract reaches 30 days to expiry, the series rolls over to the next contract and the margin decreases down the step again. This pattern produces increasing and decreasing patterns, recurring at every 45-50 days to maturity up to 300 days to expiry. Consequently, we observe similar behaviour for M_{t,T_i}^S, for $i = 4, 6, 8$.

It is unlikely that investors would take positions in such a contract given the fluctuation in the margin can increase investment costs and margin call risk. We do not encourage the use of constant-maturity futures to resolve this issue as one would need to assume that a synthetic a contract with maturity T, will have the same margin as its neighbouring observable contract, with maturity $T_1 < T$ and $T_2 > T$. Moreover, as shown in the previous chapter, VaR extrapolation from constant-maturity futures using a Hermite spline can undermine the VaR estimation process. Alternatively, one could select these margin steps in the term structure as rollover points for the series, which would eliminate them. However, the steps occur at varying time to maturities and thus cannot be replicated easily before 1^{st} January 2009 where the SPAN data does not stretch to.

In this study, we analyse M_{t,T_i} for $i = 1, 3, 5, 7, 9$ only. Although the amount of data used in this study is halved from the previous chapter, the length of the out-of-sample period is not affected. There remains an abundant number of data points which does not undermine the stability of MS_t, MT_t and MD_t. Furthermore, as we planned for margins to decrease in steps of 2 contracts, removing M_{t,T_i} for $i = 2, 4, 6, 8, 10$ has no effect on MS_t and MT_t.

Unlike Lam et al. (2010), we consider margin movements throughout the term structure and not only on the first-to-mature series. According to the Samuelson effect, the volatility increases with decreasing time to maturity, so the margin requirements should also follow the same movement.

We also find that, the VaR term structure may tilt which can sometimes push margins with greater time to maturity above those with lesser time to maturity. To avoid this, we allow the whole term structure to change only when the first-to-mature series changes, that is

$$M_{t,T_i} = \begin{cases} M_{t-1,T_i}, & \text{for } M_{t,T_1} = M_{t-1,T_1} \ , \\ VaR^{0.99,1}_{t,T_i} \frac{M_{t,T_1}}{VaR^{0.99,1}_{t,T_1}} & \text{otherwise} \ , \end{cases} \tag{5.8}$$

where $i = 3, 5, 7, 9$.

5.2.5 Evaluating Margin Requirements

Although we are primarily concerned with stability, for an implementable margin model, we assume that the exchange is interested in keeping low margin requirements to attract new clients. Here, the criterion of minimising MD_t has been employed to keep the margin suitably low. We also observe that M^S_{t,T_i} moves very much on par with $1.25VaR^{0.99,1}_{t,T}$. Hence we deem that the margin is at an adequate level should its corresponding MD_t be less than SPAN's.

To our knowledge, our work is the first to tackle the margin stability issue, thus we introduce some new measures for evaluating the margin in this respect. First, MS_t and MT_t are applicable as measures themselves where a stable low MS_t and high MT_t indicates stable margins. However, consider a scenario where we compare two margin models, should one model generate both lower MS_t and lower MT_t than the

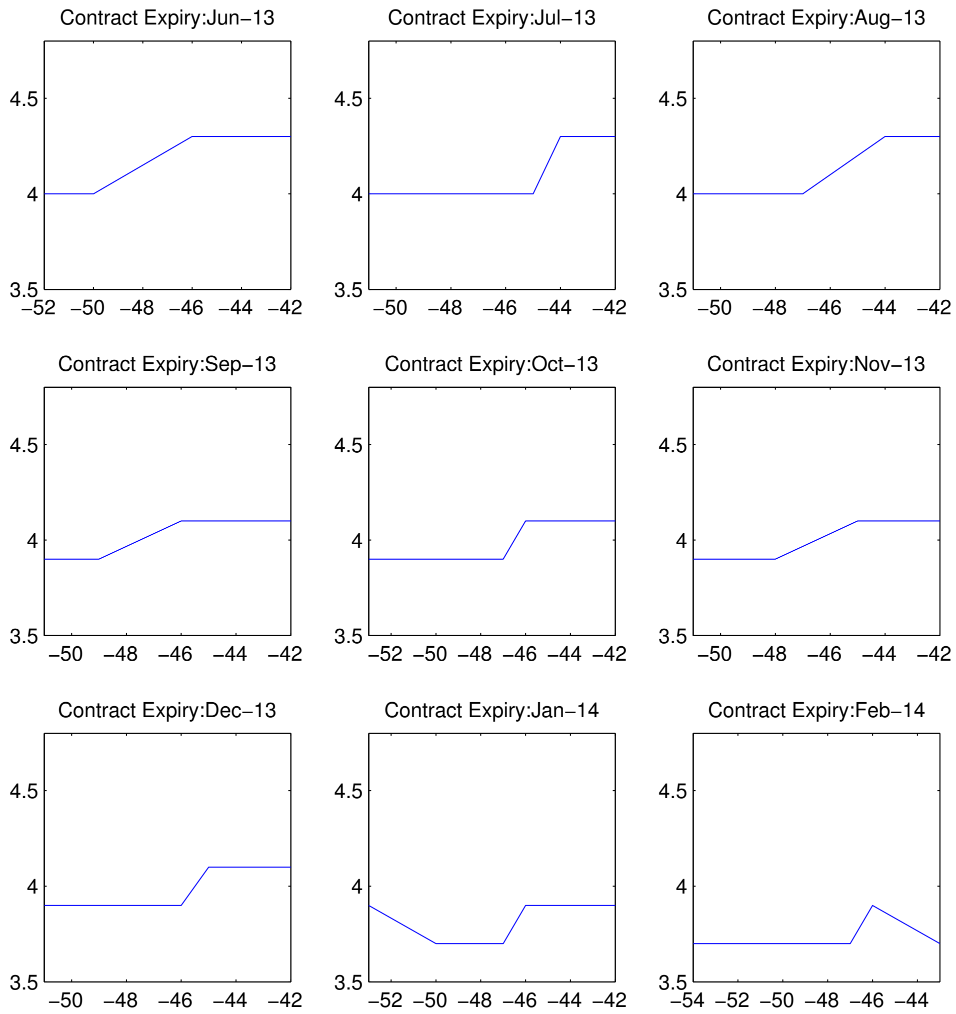

Figure 5.4.: Snapshots of the margin requirements of individual Brent crude oil contracts. Contracts expiry months: June 2013 - February 2014. Range of time to maturity: 54 to 40 days to maturity.

other, we would not be able to determine which is more stable.

To understand the behaviour of MS_t and MT_t, we examine the relationship between them. Here, for the in-sample period 29^{th} January 2009 to 30^{th} December 2013, we utilise relationship 5.6, using several combinations of $0 < \beta^L_{1,t} < 0.25$ and $0 < \beta^L_{1,t} < 0.75$, to calculate $M^L_{t,T}$. We then compute the corresponding MS_t and MT_t for each combination and construct a scatter plot between them with MS_t on the x-axis and MT_t on the y-axis (see Figure 5.6).

The most desirable place to be on this plot is at the top-left corner of the graph,

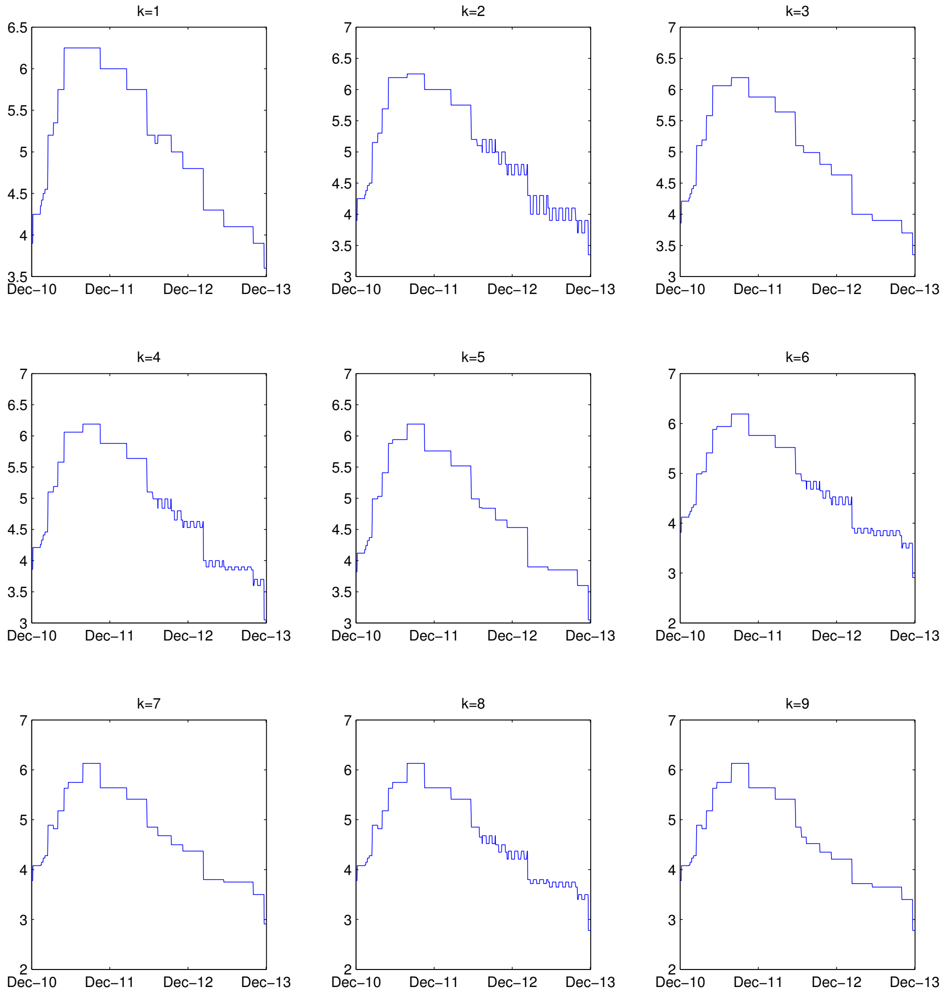

Figure 5.5.: k^{th}-to-mature Brent crude oil SPAN margin series for $k = 1$ to 9. Period 23^{rd} December 2011 to 26^{th} December 2013.

i.e. to maximise MT_t and minimise MS_t. The problem presented here is similar to Markowitz's efficient frontier, where the area covered by the blue points is an opportunity set of stability. Like the efficient frontier, only the points along the top-left edge of this area are optimally stable.

Now, consider a line drawn from the point $MS_t = 0$ and $MT_t = 0$ to any point in the opportunity set. The most stable margin level would create an infinitely high (MT_t/MS_t), while the least stable margin level would generate $MT_t/MS_t = 0$. Hence, similar to the Sharpe ratio, we can use this gradient as a measure of stability.

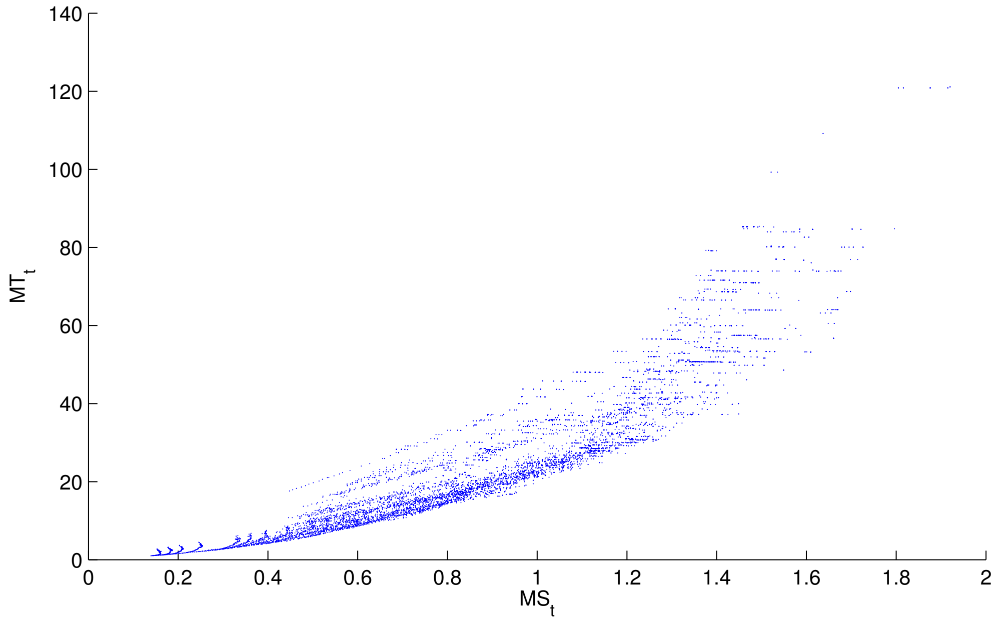

Figure 5.6.: Scatter plot of MS_t versus MT_t corresponding to $M^L_{t,T}$, parameter range $0 < \beta^L_{1,t} < 0.25$ and $0 < \beta^L_{1,t} < 0.75$. Period 29^{th} January 2009 to 30^{th} December 2013.

5.2.6 Calibration Method: Fuzzy Goal Programming for Stable Margin Estimates

A number of issues arise when calibrating β^L_1, β^L_2, β^*_1, β^*_2. First, MS_t, MT_t and MD_t do not share the same units hence they are not directly comparable. Second, we wish for stable parameters as the calibration window is rolled over since this would also generate stable margin bands and ultimately, stable margins. Third, the exchange may wish to place different emphasis on the different criteria, concentrating on the margin level as opposed to stability for example. We employ Goal Programming (GP) to resolve these issues.

Goal Programming (GP) practices include parameters which are adjusted to satisfy, or get as close as possible to a known objective (aka goals). The difference between the final outcome and the goal is known as the achievement function. Hereafter, we refer to MS_t, MT_t, MD_t as the achievement functions. The four main types of GP described in Romero (2004) are

1. Weighted GP (aka Archimedean GP) sets weights on the objectives which both normalises and rank them according to preferences
2. Lexicographic GP (aka non-Archimedean or pre-emptive GP) is similar to Weighted

GP but here objectives can be sorted into groups with different priorities, the group with the highest priority takes an infinite amount of importance and must be satisfied first before moving to the next set of objectives

3. MINMAX GP (aka Chebyshev or Fuzzy GP) is again similar to the Weighted GP, but here minimising the deviation from the goal is also a goal

Minimum-variance hedging as studied in chapter 3 for example can be described as a variant of the Lexicographic GP, where the decision to minimise the variance takes priority over all other criteria such as minimising transaction costs - which is then considered last.

GP and other programming practices are already implemented widely in finance (intentionally or otherwise), mostly in the portfolio management literature, see Board et al. (2003) for an overview of general Operational Research techniques in the financial markets. Note that like most works mentioned in Board et al. (2003), we do not adhere to the same notations as those which may be found in the GP literature. However, given our proposed decision making processes follow the GP formulation, the result is essentially the same.

The nature of margin requirements does not allow for a similar problem, as the goals are criteria to be minimised or maximised (such as margin step size and margin change intervals). Our work is hence more related to Parra et al. (2001) who uses GP to find optimal portfolio weights to maximise both investors utility and liquidity. Parra et al. (2001) describes the objectives being Fuzzy, that is, the approximate location of the objective is known hence one optimises the parameters within the vicinity of the location. For margin requirements, it is clear that stable parameters leads to stable margin bands and subsequently stable margins. As above, to tame the volatility of the optimal parameters, one can also set the algorithm to search for a local optimum around the previous parameter estimates. In this study we implement two types of GP, Archimedean and Lexicographic.

Archimedean GP

Here, for the margin model $i = L$ or $*$, we calibrate the parameters $\boldsymbol{\beta}_t^i = \{\beta_{1,t}^i, \beta_{2,t}^i\}$ by

$$\boldsymbol{\beta}_t^i = \underset{\beta_{1,t}^i, \beta_{2,t}^i}{\operatorname{argmin}} \left\{ w_{MD} \frac{MD_t - \mu_t^{MD}}{\sigma_t^{MD}} + w_{MS} \frac{MS_t - \mu_t^{MS}}{\sigma_t^{MS}} - w_{MT} \frac{MT_t - \mu_t^{MT}}{\sigma_t^{MT}} \right\} , \quad (5.9)$$

where w_i, μ_t^i and σ_t^i are weights, means and standard deviations of achievement function i respectively. Note that $\boldsymbol{\beta}_t^i$ is optimal for the window $t-n+1$ to t only. As the estimation window is rolled over, the optimal parameters may change, hence the addition of the subscript t in this relationship. The negative sign on MT_t indicates maximisation as opposed to minimisation.

We normalise the achievement functions to prevent *incommensurability* given they do not share the same units. For example, MT_t is measured in days, while the others are measured in $USD/barrel$. Without normalising, minimising MT would dominate the other criteria in the optimisation scheme. A popular normalising method in GP is to divide the achievement functions by the goal itself and multiplying by 100, that is, converting each achievement function into a percentage. We cannot apply the same method as this would require dividing MS_t by 0. Alternatively, dividing by the achievement functions by the standard deviation would also solve the incommensurability issue. From preliminary analysis however, we find that the mean of MT_t is much higher relative to MS_t and once divided by the standard deviation, MT_t still significantly overrides MS_t. Hence we normalise by both taking away the mean and dividing by the standard deviation. Note that should we calculate μ_t^i and σ_t^i in a conventional sense, e.g. $\mu_t^{MS} = \frac{1}{n_T n_s}\sum_{k=1}^{n_T}\sum_{i=t-n+1}^{t}|\Delta M_{i,T_k}| = MS_t$, $\sigma_t^{MS} = \sqrt{\frac{1}{n_T n_s}\sum_{k=1}^{n_T}\sum_{i=t-n+1}^{t}(|\Delta M_{i,T_k}| - MS_t)^2}$, we would yield $\mu_t^{MS} = MS_t$, similarly we would find that $\mu_t^{MD} = MD_t$ and $\mu_t^{MT} = MT_t$ and relationship 5.9 would always yield 0. Hence μ_t^i and σ_t^i of the achievement function i are found by using a different method. For margin models $j = \{L, *\}$ with parameters $\beta_{1,t}^j$ and $\beta_{2,t}^j$: 1) we construct $M_{t,T}^j$, 2) we calculate the achievement functions for $M_{t,T}^j$ 3) we repeat this process using all combinations of $\beta_{1,t}^j$ and $\beta_{2,t}^j$ in steps of 0.01 over the calibration range to produce collections of MS_t, MT_t and MD_t. Lastly, we use the mean and standard deviation of each collection as estimates of μ_t^i and σ_t^i for $i = MS,\ MT,\ MD$. Furthermore, we utilise w_i to adjust the achievement functions when the calibration may be over defined (discussed in more detail in section 5.3.1). We perform in-sample calibration over the period 28^{th} January 2009 to 30^{th} December 2013. We first examine whether or not the margin's MD_t is less than or equal to SPAN's (0.840 USD/barrel), should this be the case then we consider the margin to be at a suitable level. We then examine the stability of each model by comparing MT_t/MS_t ($MS_t = 0.367$ USD/barrel, $MT_t =$ 43.381 days, $MT_t/MS_t = 118.202$ days/(USD/barrel) for SPAN margins).

We observe that $M_{t,T}^S$ fulfils the our criteria very well, especially on the stability front. This is because the SPAN system was never designed to interact with the VaR -which is very volatile- in the first place. This comes at a cost of breaching the one-day 99% VaR at several points in time, so even though EMIR were to reduce the two-day VaR restriction down to one-day as we suggest, the SPAN method would still be inadequate.

We carry out the out-of-sample calibration using a 2500-day (10 years) rolling window, starting from the window 29^{th} March 1999 - 28^{th} January 2009, rolling over one business day at a time. Here, at time t, we assume the exchange has knowledge of the information in the time window $t-n+1$ to t. We select the optimal parameters β_1^L, β_2^L, β_1^*, β_2^* which generate optimal levels of MS_t, MT_t and MD_t via the following procedure:

1. At time t, generate a series of $VaR_{t,T}^{0.99,1}$ using the information set $t-n+1$ to t
2. As we are using the OEWMAt-6 method for calculating the VaR, we estimate the time $t+1$ forecast via the relationship $VaR_{t+1,T}^{0.99,1} = VaR_{t,T}^{0.99,1}$
3. Calibrate β_1^L, β_2^L, β_1^*, β_2^* on the estimation window $t-n$ to $t+1$ using relationship 5.9
4. Calculate $M_{t+1,T}$ using the optimal parameters obtained
5. Rollover the calculations for the rest of the out-of-sample period

We assume that $M_{t+1,T}$ is calculated and announced at time t, i.e. one business day prior to the change in margin at time $t+1$. Note that on this period, we cannot combine the achievement functions to examine the overall performance of the margin model as seen in relationship 5.9. The population of each margin level is limited to only one set of movement and given the objective functions are averages themselves, the estimate of μ_t^i for achievement function i is simply equal to itself; we compare each criteria one-by-one instead.

Lexicographic GP

There are many different ways to which the exchange can prioritise the achievement functions. Here, we first identify a pool of parameters suitable for generating margins which meets a particular stability point and from this, determine which parameters give the lowest margin level for competitive reasons. The target stability point in this case is SPAN's. Our calibration should ensure margin stability at roughly the same level whilst keeping the margin suitably low so the exchange maintains competitive advantage. The process for a window of VaR estimates hence follows:

1. We generate an MT_t-MS_t scatter plot for each margin rule
2. We identify a target margin stability point on this map and bind an area for all the margins with higher MT_t and lower MS_t from this point

3. Of this pool of parameters, select the set which produces the lowest MD_t

First, we apply this procedure using the same in-sample range as the Archimedean GP approach. The required area for margins more stable than SPAN is hence $MS_t < 0.367$ USD/barrel, $MT_t > 43.381$ days. We also apply this method out-of-sample using the same periods as the previous section. Here, we roll the estimation window over and repeat all of the above. If no parameters can produce estimates more stable than SPAN, then take the previous parameters instead. However, we have no data for $M^S_{t,T}$ covering the initial calibration period February 1999 - January 2009. More data on this front would be of great benefit to our results, for now, we have to impose the assumption that ICE have operated in the same manner as they have done, that is our stability targets of $MS_t = 0.367$ USD/barrel and $MT_t = 43.381$ days remain constant throughout. Further, VaR movements between February 1999 and January 2009 contain much more extreme swings than the period January 2009 - December 2013. Although unobservable, we assume that the margin in the initial calibration period is less stable than the out-of-sample period. Margins with characteristics $MS_t < 0.367$ USD/barrel, $MT_t > 43.381$ days would be more stable than those which were actually set in practice. Again, to stop out-of-sample parameters from jumping, we impose the vicinity ± 0.05 for the margin movements. Note that as each objective function is being considered one-by-one, there is no need to impose normalising constants as carried out when combining the criteria in an Archimedean GP arrangement. Hereafter, we denote the margin generated using the Lexicographic GP method as $M^{**}_{t,T}$.

5.3 Results and Discussion

5.3.1 In-sample Calibration: $\beta^L_{1,t}$, $\beta^L_{2,t}$

In this analysis, we limit the parameters to the range $0 < \beta^L_1 < 0.25$ and $0 < \beta^L_2 < 0.75$ which should allow for the corresponding margin level to move within the vicinity of $1.25VaR^{0.99,1}_{t,T}$. Note that ranges β^L_1, $\beta^L_2 < 0$ is not applicable according to relationship 5.6, as $M^L_{t,T}$ would always equal $1.25VaR^{0.99,t}_{t,T}$. The restriction $0 < \beta^L_2 < 0.75$ ensures that the margin does not move beyond $2VaR^{0.99,1}_{t,T}$.

Intuitively, relationship 5.6 entails that as β^L_1 and β^L_2 increase, the margin changes should increase in size and decrease in frequency. Maps of the achievement functions are shown in Figure 5.7. The corresponding margin achievement functions behave as expected where $MS_{t,T}$ and $MT_{t,T}$ increase with decreasing $(1.25 - \beta^L_1)VaR^{0.99,1}_{t,T}$ and increasing $(1.25 + \beta^L_2)VaR^{0.99,1}_{t,T}$. The initial calibration, with w^{MD}, w^{MS}, $w^{MT} = 1$ resulted in optimal parameters $\beta^L_1 = 0$, $\beta^L_2 = 0$, i.e. to set $M^L_{t,T} = VaR^{0.99,1}_{t,T}$, which counters our expectations. Upon further investigation, we find that because of the way relationship 5.6 is specified, having both parameters equal to 0 fulfils two of the criteria -minimising both MS_t and MD_t simultaneously. To avoid this issue, we set $w^{MD} = 0$, our intuition follows that the margin level must always return to $1.25VaR^{0.99,1}_{t,T}$ and the criteria to minimise MD_t is already satisfied as long as $\beta^L_{2,t}$ is not too high; even its maximum $\beta^L_{2,t} = 0.75$, MD_t according to this rule is below SPAN's (at 0.78 USD/barrel respectively). Here, the optimal parameters are $\beta^L_{1,t} = 0.22$ and $\beta^L_{2,t} = 0.61$ with corresponding $MS_t = 1.024$ USD/barrel and $MT_t = 87.482$ days. These estimates are far off our targets, upon assessing the stability map for this rule (see Figure 5.8), we observe that none of the parameters in this range can produce margins more stable than SPAN. it is clear that the underlying problem lies in the rule itself and hence it would not be possible to generate favourable margin requirements out-of-sample. We calibrate this rule out of sample to compare with other rules.

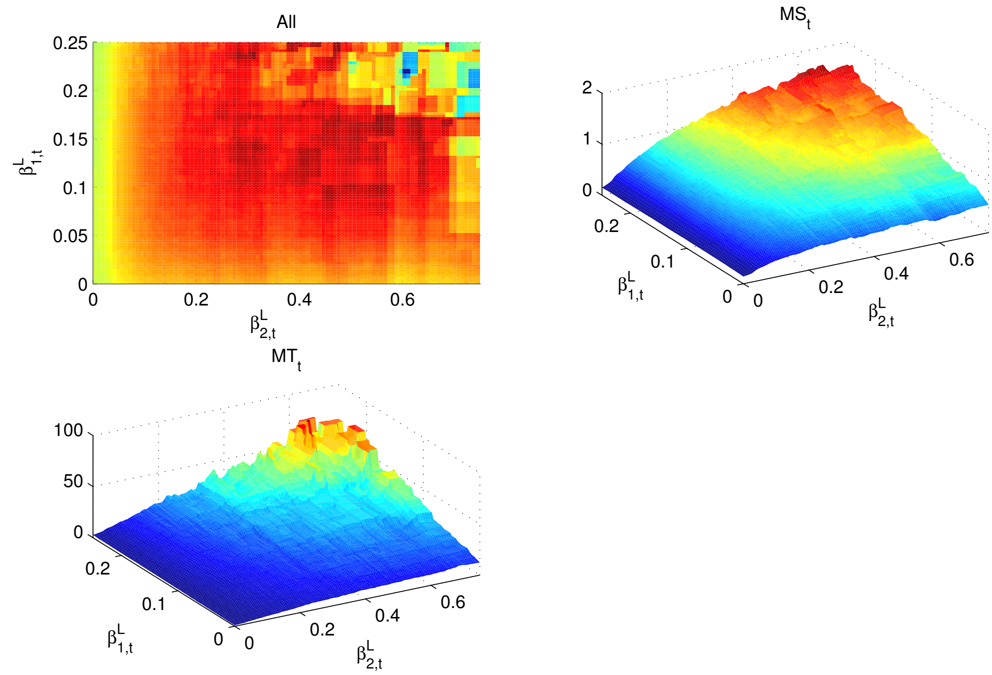

Figure 5.7.: Map of each achievement function for margin $M^L_{t,T}$ with parameters $\beta^L_{t,1}$ from 0 to 0.25 and $\beta^L_{t,2}$ from 0 to 0.75. Period: 28^{th} January 2009 to 30^{th} December 2013. Optimal parameters: $\beta^L_1 = 0.25$ and $\beta^L_2 = 0.70$, producing $MS_t = 1.024$ USD/barrel and $MT_t = 87.482$ days. General statistics: $\mu^{MS}_t = 0.546$, $\sigma^{MS}_t = 0.246$ USD/barrel, $\mu^{MT}_t = 24.97$ and $\sigma^{MT}_t = 17.61$ days.

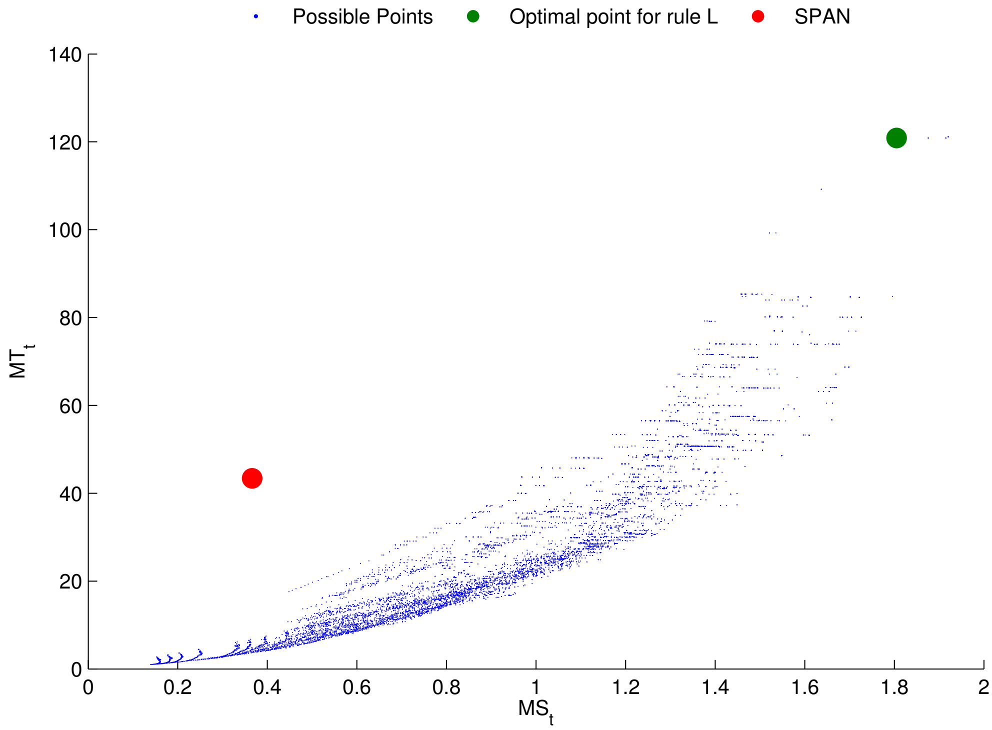

Figure 5.8.: Scatter plot of MS_t versus MT_t corresponding to $M^L_{t,T}$, parameter range $0 < \beta^L_{1,t} < 0.25$ and $0 < \beta^L_{1,t} < 0.75$. Period: 28^{th} January 2009 to 30^{th} December 2013.

5.3.2 In-sample Calibration II: $\beta^*_{1,t}$, $\beta^*_{2,t}$, $\beta^{**}_{1,t}$, $\beta^{**}_{2,t}$

The maps of the in-sample achievement functions are shown in Figure 5.9. As expected, the corresponding margin level increases with $\beta^*_{1,t}$ hence we need to minimise MD_t to allow for suitably low margins. Second, MT_t tend to increase with $\beta^*_{1,t}$, higher upper margin threshold indicates less frequent changes. Third, MS_t decreases with increasing $\beta^*_{1,t}$, this is rather surprising, given margin step sizes should be governed by $\beta^*_{2,t}$ alone. We find that as the margin band narrows (with decreasing $\beta^*_{1,t}$), the margin tend to follow $VaR^{0.99,1}_{t,T}$ more. Given the VaR is quite volatile, lower $\beta^*_{1,t}$ also equates to more volatile margins.

Since $\beta^*_{2,t}$ governs the step sizes, greater $\beta^*_{2,t}$ equates to higher MS_t. Large step sizes also moves the margin further away from the thresholds, hence the margin also changes less frequently, that is higher $\beta^*_{2,t}$ indicates higher MT_t. There is however no relationship with respect to MD_t, which again is not surprising given the margin level is mainly govern by the upper threshold, i.e. $\beta^*_{1,t}VaR^{0.99,1}_{t,T}$.

The initial optimal parameters are: $\beta_1^* = 2.24$ and $\beta_2^* = 0.06$ and $MS_t = 0.388$ USD/barrel, $MT_t = 61.111$ days, $MD_t = 1.262$ USD/barrel. The calibration is successful in producing lower MS_t and higher MT_t than those from the preliminary analysis. This shows that the rule introduced from scratch is more likely to outperform those from the previous literature in an out-of-sample analysis; although the margin level is still too high. From this figure, we also observe that MT_t is highly nonlinear with respect to the parameters which results in patches of minima once the achievement functions are combined. More notably, around $2.27 < \beta_{1,t}^* < 2.30$, $0.05 < \beta_{2,t}^* < 0.09$ indicated in Figure 5.9 by a dark blue patch, and a relatively smaller area at $1.8 < \beta_{1,t}^* < 2.1$, $0.18 < \beta_{2,t}^* < 0.2$. There could hence be several solutions to the optimal parameter estimates. For this procedure, $w^{MS} = w^{MT} = w^{MD} = 1$, an interesting further study may be to explore different sets of weights which may ensure greater performance by the margin models.

With this rule, we find that there are several parameters whose margins are more stable than SPAN ($MS_t < 0.367$ USD/barrel, $MT_t > 43.38$ days, see Figure 5.10). At this point however, the margin level is too high, with $MD_t = 1.512$ USD/barrel and hence may not be to the likings of the exchange. Nonetheless, it is clear that the nature of this rule is more flexible in generating stable margins than relationship 5.6 and we expect our rule to produce the best out-of-sample results among all rules derived from the literature.

Now examining Lexicographic GP, we assume that in this period, the exchange aims to produce margins at least as stable as SPAN. Hence, on the MS_t versus MT_t scatter plot, we enclose an area with $MT_t > 43.48$ days and $MS_t < 0.367$ days, shown by the green positions in Figure 5.10. Out of this set, we select the margin level which has the lowest MD_t, as this would produce the lowest margins possible. We obtain the results with $MS_t = 0.355$ USD/barrel, $MT_t = 45.381$ days and $MD_t = 0.705$ USD/barrel, $MT_t/MS_t = 127.890$ days/(USD/barrel). Lexicographic GP is able to surpass SPAN in all accounts, hence we expect $M_{t,T}^{**}$ to generate the best results out-of-sample.

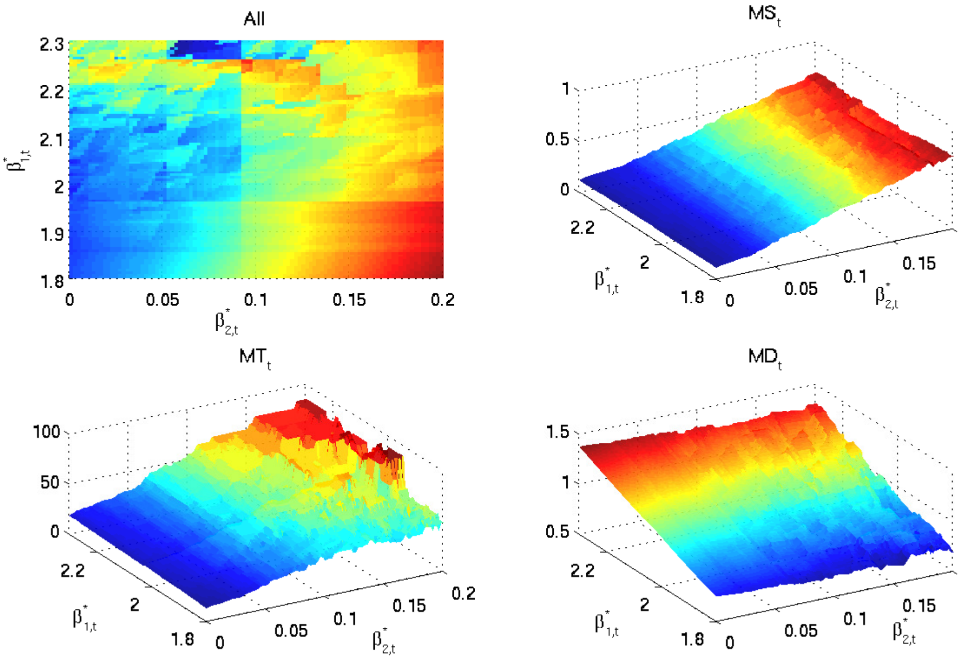

Figure 5.9.: Map of each achievement function for margin $M^*_{t,T}$ with parameters $\beta^*_{t,1}$ from 1.8 to 2.3 and $\beta^L_{t,2}$ from 0 to 0.2. Period: 28^{th} January 2009 to 30^{th} December 2013. Optimal parameters: $\beta^*_1 = 1.92$ and $\beta^*_2 = 0.11$, producing $MS_t = 0.223$ USD/barrel and $MT_t = 32.483$ days, $MD_t = 1.299$ USD/barrel, $MT_t/MS_t = 145.664$ days/(USD/barrel). Other statistics: $\mu^{MS}_t = 0.390$, $\sigma^{MS}_t = 0.181$, $\mu^{MT}_t = 45.017$ days, $\sigma^{MT}_t = 18.989$ days, $\mu^{MD}_t = 1.007$, $\sigma^{MD}_t = 0.177$ USD/barrel.

5.3.3 Out-of-Sample Calibration

Following the in-sample results, the initial calibration is carried out with the parameter range $1.8 < \beta^*_{1,t} < 2.3$, $0 < \beta^*_{2,t} < 0.2$ and $0 < \beta^L_{1,t} < 0.25$, $0 < \beta^L_{2,t} < 0.75$. The range $1.8 < \beta^*_{1,t} < 2.3$ may seem ad hoc at this stage so we carried out a secondary study for $1.25 < \beta^*_{1,t} < 2$, we find that the calibration follows a different minimum patch at roughly $\beta^*_{1,t} = 1.5$ with a much lower MD_t. The margin according to this area however is much less stable and results will not be shown here. We use all margin estimates at the end of the initial in-sample calibration as the starting point of the out-of-sample window.

Rolling over from the initial period, we fix the time t parameter range to ± 0.05 of the time $t-1$ optimal parameters to ensure that optimal parameters do not jump between minimum patches. The ± 0.05 range covers a suitable amount of room for the parameters to move within and also maintain parameter stability. This also helps

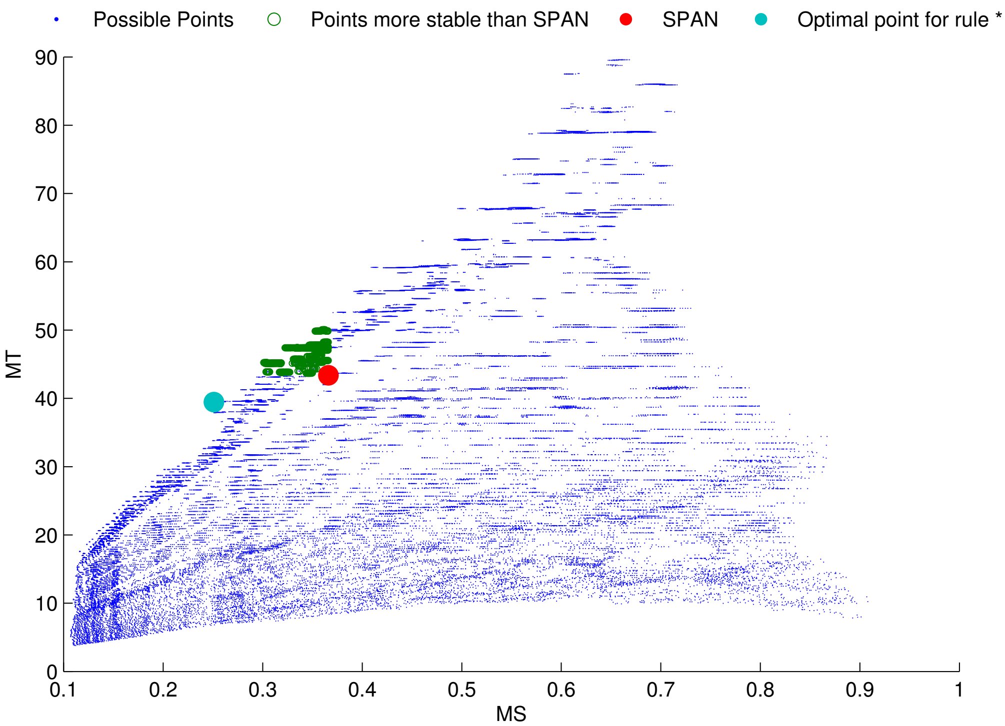

Figure 5.10.: Scatter plot of MS_t versus MT_t corresponding to $M^{**}_{t,T}$, parameter range $1.8 < \beta^{**}_{1,t} < 2.3$ and $0 < \beta^{**}_{2,t} < 0.2$. Period: 28^{th} January 2009 to 30^{th} December 2013.

to reduce the computational speed by over 10 times.

Archimedean GP is successful in producing stable out-of-sample parameters. Figure 5.11 shows the optimal parameter estimates $\beta^L_{t,1}$, $\beta^L_{t,2}$, $\beta^*_{t,1}$, $\beta^*_{t,2}$, $\beta^{**}_{t,1}$, $\beta^{**}_{t,2}$. We observe that $\beta_{1,t}$ bounces between 0.18 and 0.19 roughly around the period January 2013 and December 2013. Note that this is not a result of the parameter confinement range since the fluctuation occurs at $0 < \beta^L_{1,t} < 0.25$. While the cause of this fluctuation is unknown, its only effect is to generate similar fluctuations in the lower margin threshold. Providing such movements are small and the margin does not breach this threshold often, we should see no (or negligible) difference in the margin movements. This is also shown in Figure 5.13, where M^L_{t,T_1} does not contain fluctuations in the same period.

We calculate MS_t, MT_t, MD_t and MT_t/MS_t and compare the margin models.[7] From Table 5.1, we find that keeping the margin equal to $1.25VaR^{0.99,1}_{t,T}$ as

[7] Here we cannot calculate the mean and standard deviation of the margins as we have done in the in-sample estimation given only one margin route is present per model.

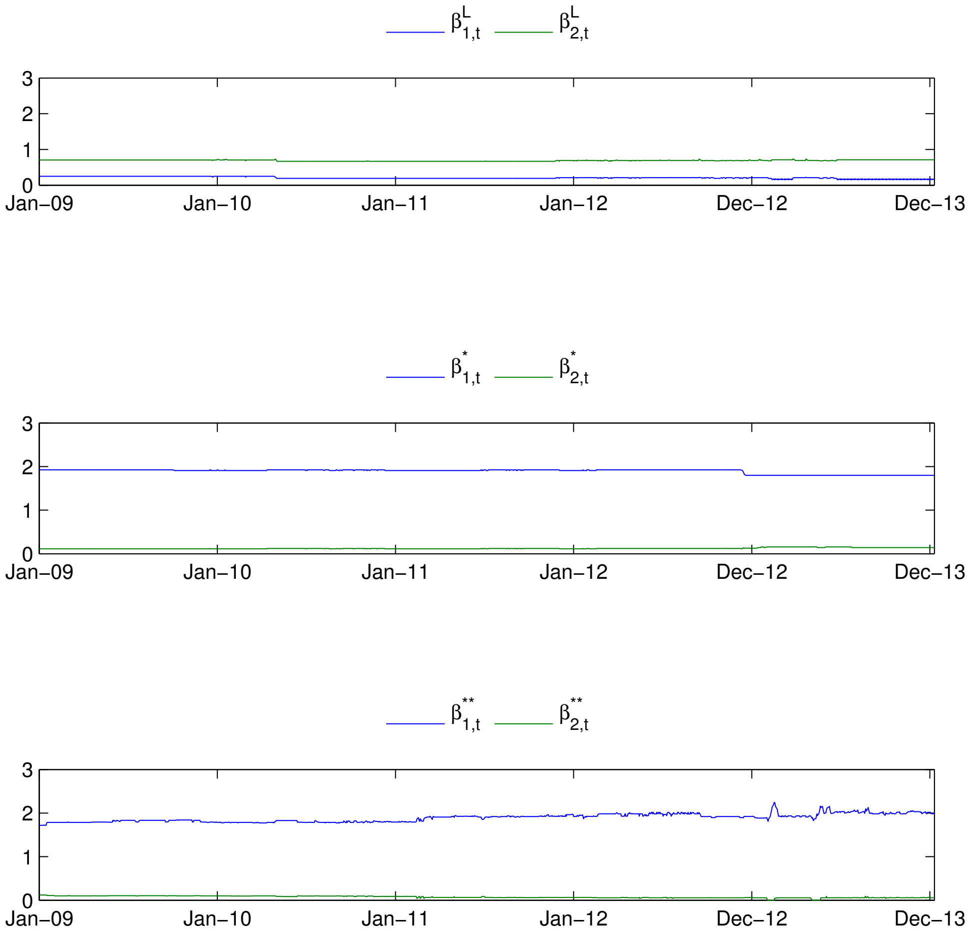

Figure 5.11.: Optimal parameter estimates for the period 28^{th} January 2009 to 30^{th} December 2013. Parameters calibrated using a 2500-day rolling window. $\beta^L_{t,1}$ and $\beta^L_{t,2}$ confined to 0-0.25 and 0-0.75 respectively. $\beta^*_{t,1}$ and $\beta^*_{t,2}$ confined to 0-0.2 and 1.8-2.3 respectively. $\beta^{**}_{t,1}$ and $\beta^{**}_{t,2}$ also confined to 0-0.2 and 1.8-2.3 respectively. Each parameter estimate at time t is restricted to move within the vicinity of ± 0.05 of the estimates at time $t-1$. The first estimation window is calibrated using the entire confinement range. Calibration is carried out using the first 10-to-mature futures series.

assumed by the majority of the literature provides the most unstable margin level with $MT_t/MS_t = 7.189$ days/(USD/barrel). The technique of only changing the margin when it breaches a threshold holds superior on this account where simply imposing a

± 0.15 band on the margin ($M^C_{t,T}$) is more than twice as stable with $MT_t/MS_t = 18.097$ days/(USD/barrel). Further alleviating the flexibility by optimally calibrating the margin band produces margins ($M^L_{t,T}$) which are again two times more stable with $MT_t/MS_t = 45.804$ days/(USD/barrel). This shows that, calibration strongly aids the production of stable margins. Of all the automated systems, our rule provides the most stable margins ($M^*_{t,T}$) with $MT_t/MS_t = 81.454$ days/(USD/barrel), over 2.5 times that of $M^L_{t,T}$. In fact, $M^*_{t,T}$ is superior to $M^C_{t,T}$ on both accounts with greater MT_t and lower MS_t. The most successful margin model here is $M^{**}_{t,T}$ using Lexicographic GP with corresponding $MT_t/MS_t = 101.724$ days/(USD/barrel).

Criteria	$1.25VaR^{0.99,1}_{t,T}$	$M^C_{t,T}$	$M^L_{t,T}$	$M^*_{t,T}$	$M^{**}_{t,T}$	$M^S_{t,T}$
			28^{th} January 2009 - 30^{th} December 2013			
MS_t	**0.139**	0.673	1.345	0.486	0.335	0.367
MT_t	1.000	12.183	**61.611**	39.600	34.029	43.381
MD_t	**0.000**	0.198	0.731	0.903	0.799	0.840
MT_t/MS_t	7.189	18.097	45.804	81.454	101.724	**118.334**
			28^{th} January 2011 - 14^{th} July 2011			
MS_t	**0.145**	0.702	1.354	0.538	0.459	0.417
MT_t	1.000	13.043	**61.444**	32.833	31.105	32.855
MD_t	**0.000**	0.205	0.723	1.161	0.898	0.792
MT_t/MS_t	6.782	18.576	45.375	61.051	61.764	**78.852**
			14^{th} July 2011 - 30^{th} December 2013			
MS_t	**0.131**	0.649	1.337	0.415	0.195	0.477
MT_t	1.000	11.232	**59.667**	35.000	26.250	57.698
MD_t	**0.000**	0.192	0.739	0.643	0.700	0.755
MT_t/MS_t	7.658	17.312	44.629	84.406	134.365	**189.839**

Table 5.1.: Out-of-sample MS_t, MT_t, MD_t and MT_t/MS_t for all models denoted in USD/barrel, days, USD/barrel and days/(USD/barrel) respectively. Estimation periods 28^{th} January 2009 to 14^{th} July 2011, 26^{th} December 2013 and 14^{th} July 2011 - 30^{th} December 2013. Bold fonts indicate the best margin models for each achievement function.

There are some visible differences in the relative margin levels from the first half of the out-of-sample period to the second. More notably, from January 2012 onwards, $M^*_{t,T}$ moves very much on par with $M^S_{t,T}$ with cautious decreases which indicates that our rule is successful in mimicking this recent behaviour of the exchange. Prior to this period however, between January 2009 and June 2010, all other rules generate margins well above $M^S_{t,T}$. This is due to the sudden decrease in margin levels by ICE in April 2009, which is uncharacteristic. In fact, referring to Figure 5.2, M^S_{t,T_1}

is often-times below $VaR^{0.99,1}_{t,T}$, which is inadequate. Therefore, it is not surprising that our margin rules would generate higher margin levels in this period. While $M^{*}_{t,T}$ and $M^{**}_{t,T}$ remain high up until November 2009 and slowly decreases, $M^{L}_{t,T}$ however displays the same sharp decreasing pattern as $M^{S}_{t,T}$. This however rises again in July 2009 following a slight increase in $VaR^{0.99,1}_{t,T}$, in conjunction with our hypothesis that sharp margin falls may lead to more frequent margin rises.

To examine the impact of these movements on the achievement functions, we split the analysis period in two halves: 28^{th} January 2009 to 14^{th} July 2011 and 14^{th} July 2011 to 26^{th} December 2013. The former sub period is slightly more volatile with mean $VaR^{0.99,1}_{t,T_1} = 4.033$ USD/barrel compared to the latter's 3.232 USD/barrel. We do not split the period into smaller sub periods to assess the upward volatility movement from December 2011 to February 2012; we consider this too short to generate statistically significant estimates of the achievement functions. We observe that $1.25VaR^{0.99,1}_{t,T}$ and $M^{C}_{t,T}$ perform worse in the latter sub period, with lower MT_t/MS_t estimates. This confirms that although the market volatility is generally decreasing, the movement of $1.25VaR^{0.99,1}_{t,T}$ is more erratic. The calibrated rules on the other hand perform better in this latter period, demonstrating that applications of rules and proper calibration techniques are essential in generating more stable margin requirements.

The whole sample results show that $M^{**}_{t,T}$ is the best model for the period 28^{th} January 2009 to 14^{th} July 2011 and the overall out-of-sample period. To further examine the robustness of these results, we examine 500-days rolling-window estimates of MT_t/MS_t and MD_t starting on the period 28^{th} January 2009 to 4^{th} January 2011 (see illustrations in Figure 5.12). Here, MD_t of each calibrated model tend to overlap each other but remain roughly at the same level, approximately 0.8 USD/barrel. Although the MD_t for the models are not always consistently below SPAN's, hence there are some estimation windows where margins may be slightly too high. Estimates of MT_t/MS_t tend to display step functions which is not surprising given that these only move as margin changes leaves and enters the estimation window. We find that in fact, $M^{S}_{t,T}$ is the most stable margin for the majority of the instances as the window is rolled over. Of the automated systems however, $M^{**}_{t,T}$ remains superior throughout. Lexicographic GP using our rule is hence the best model in this case. Although, further development is required to generate a suitable margin model. In the same Figure, we observe that SPAN's MT_t/MS_t is increasing with time, this is perhaps due to ICE's own development to promote stable margin requirements. Our rules however, tend to generate more consistent MT_t/MS_t.

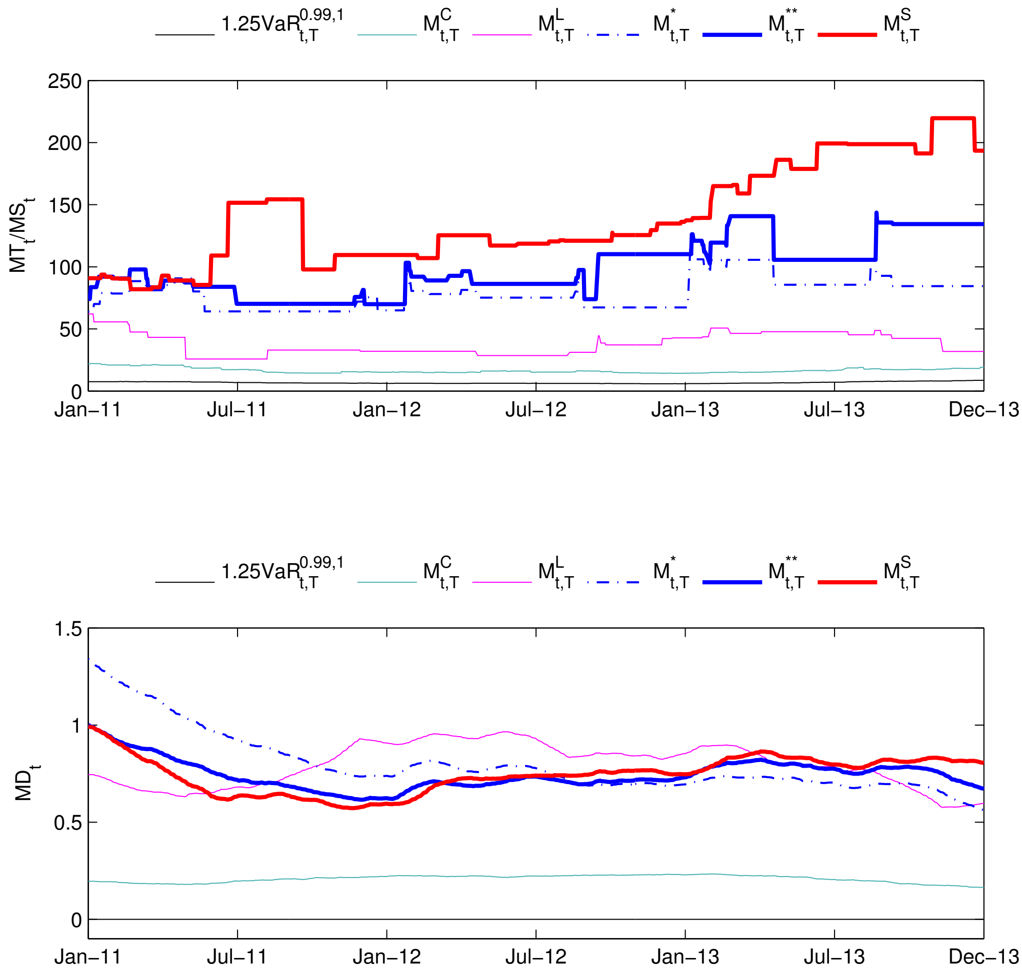

Figure 5.12.: Rolling MT_t/MS_t and MD_t estimates for all margin rules denoted in days/(USD/barrel) and USD/barrel respectively. Rolling window length: 500 days. Estimation period: 4^{th} January 2011 - 30^{th} December 2013.

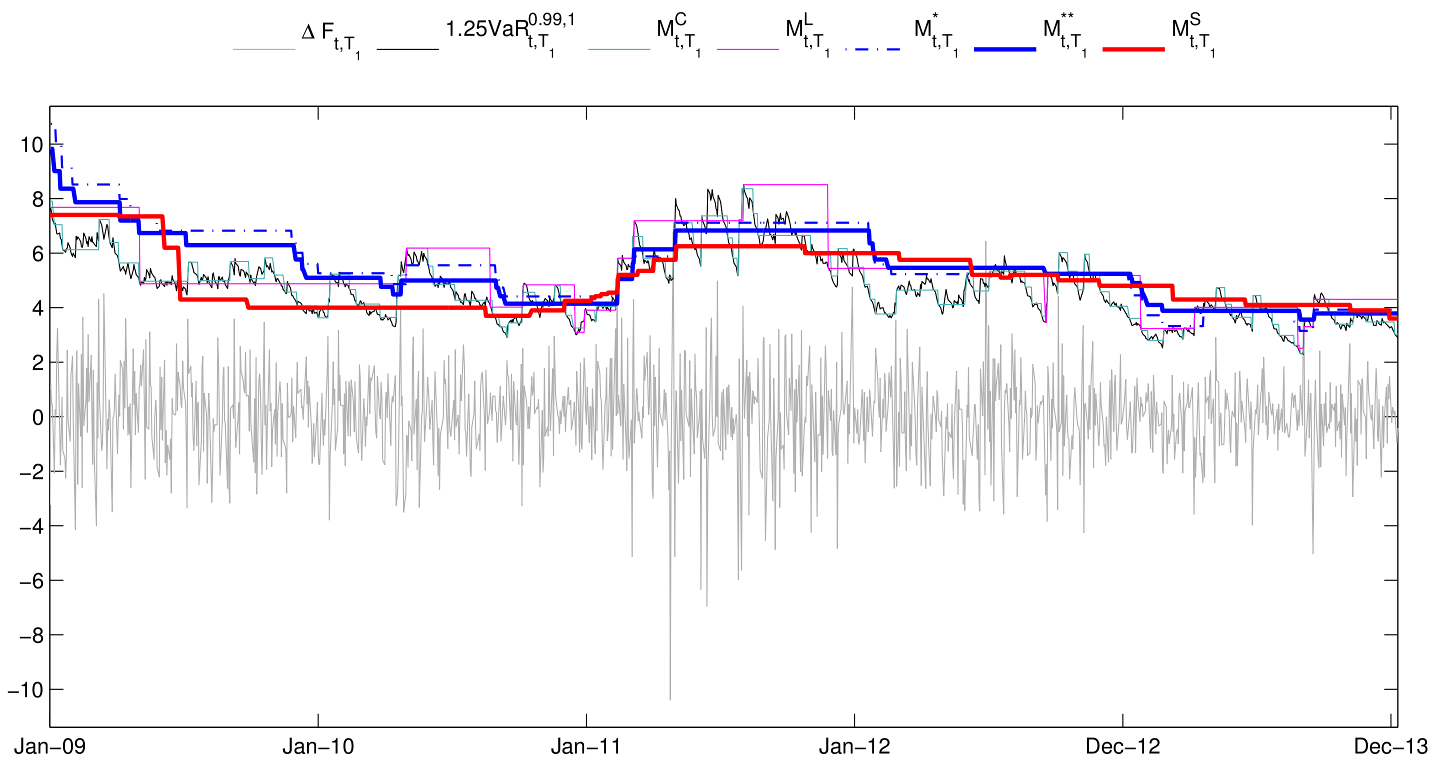

Figure 5.13.: First-to-mature Brent crude futures P&L's and short position margin requirements each model: VaR_{t,T_1}, M^C_{t,T_1}, M^L_{t,T_1}, M^*_{t,T_1}, M^{**}_{t,T_1} and M^S_{t,T_1}. Period: 28th January 2009 to 26th December 2013. VaR estimation method: Orthogonal EWMA using student-t distribution innovations with 6 degrees of freedom. VaR and margin estimations does not Brent futures P&L spikes on the 5th May 2011 and 13th June 2012 as these were caused by rare events which only effected the market temporarily. Long position margins are equivalent the short position margin. Margins and P&L's denoted in USD per barrel.

5.4 Conclusions and Further Work

In this chapter we introduce methods for evaluating margin requirements' stability and demonstrated methods for calibrating margin models. We compare 5 different margin models' ability to produce stable margin requirements for Brent crude oil futures term structure. Although the margin models tested here cannot quite outperform SPAN in producing more stable margin requirements, our methods are fully automated and much simpler to implement. We also suggest a change to current regulations, which require margins to cover the 99%, two-day VaR, to adjust their levels to the 99%, one-day VaR to allow US and European exchanges to compete. Most importantly, our methods ensure that the margin consistently remains above this coverage level. Our measure of margin stability, the fraction between average the time and size of the margin changes (MT_t/MS_t), is an easily understandable measure and also adaptable on a regulatory basis. Together with our works in chapter 4, we generate out-of-sample results on a rolling-window basis, following EMIR requirements.

The most successful rule in producing stable margins, is one which mimics the movements of SPAN, where margins are allowed to decrease slowly to avoid sudden increases in conjunction with the VaR. Prioritising some criteria results in more stable margins as opposed to combining them in a linear manner. Lexicographic GP is hence superior to its Archimedean counterpart for generating stable margin requirements. The rules provided here highlight the how margins generated using different methods result in varying levels of stability, our research provides a platform numerous studies to come. This may include generating more flexible rules with more parameters, or different prioritisations of criteria in the Lexicographic GP formulation.

The statistical methods presented in this work include estimating averages of time and step sizes of the margin movements. The overall distribution of these terms are however undefined. We note that the margin follows a Compound Poisson process, calibrating the margin to such a process would provide better understanding in terms of the statistical behaviour of the margin process. We can then perhaps compare the poisson processes in a statistically meaningful way as was carried out in chapters 3 and 4.

6. Concluding Remarks

Following the substantial growth in the energy market, our studies investigate risk management methods for its players. The numerous players: speculators, hedgers, brokers, central banks, regulators, exchanges, interact in a complex manner, to which we concentrate on two: the hedger, more specifically the refinery and the exchange.

The refinery buys crude oil and sells refined products, earning the spread (aka crack spread) between them. While their aim is to generate as much sales as possible, the crack spread may fluctuate and the refinery is exposed highly exposed to market risks. We examine the short-term delta hedging problem, prevalent in the literature where weekly spot prices are hedged using futures to ensure low volatility in the profits and losses. This is a complex multi-asset hedging problem, which requires attention to volatility estimation of the commodity prices and the correlation between them. Contrary to the the majority of previous works, we find that advanced minimum-variance hedge ratios not only provide negligible differences in variance reduction but also generate excessive transaction costs. Hence we consider the naive hedge the best strategy for hedging oil spreads in this instance. What this study highlight is the importance of transaction costs in investments, of which the margin on the hedging instrument -the futures contract- plays a central role. Upon carrying out this study, we find that the margin data is not only hard to obtain but also impossible to replicate, given the technical documents on the current methods are vague. We further find that the nature of margin requirements are changing rapidly as new regulations are introduced in the aftermath of the world recession in 2008. To this end, we turn our attention to risk management of the exchange, and suggest methods for setting margin requirements optimally in accordance to the regulatory changes.

Margin requirements is heavily based on the VaR, hence our first investigation is to find the most accurate VaR forecasting method. We focus on Brent crude oil futures as this is currently the most highly-traded energy derivative contract. We found a number of issues surrounding the literature where comparisons of log-likelihood ratios may lead to biased outcomes, especially when these ratios are of similar magnitude. Our methods is amongst the first to examine VaR on a term structural basis. By producing rollover log-likelihood ratios, we identify jumps in the independence statistics when exceedances are at the beginning and/or the end of the estimation window. To our knowledge, this behaviour has not been recorded before in the literature and can significantly bias VaR backtesting results. We conclude that the Orthogonal EWMA using t-distribution innovations with 6 degrees of freedom is the best method for forecasting Brent crude oil VaR.

Finally, we apply our VaR model to setting margin requirements for Brent crude oil futures, which has as of late become the bench mark crude oil in the world trading platform. New propositions from EMIR and Dodd Frank restrict ways in which exchanges can set margin requirements. In our study, we formulate methods for clearing houses to set margin requirements optimally, using automated methods which is easy to replicate. Our methods concentrate on creating margin requirements with small, infrequent changes in conjunction to regulations introduced by EMIR. For this, we introduce three criteria for margin calibration, namely: maximising the average time between margin changes and minimising the average step size of the margin change to generate stable margin requirements; and minimising margin deviation from 1.25 times the VaR which stops the margin from rising too high.

We also introduce measures to assess margin stability, the likes of which are easily understandable and can be used in a regulatory context. By generating a scatter plot of the average margin step size versus the average time between margin changes, we create a graph with similar criteria to Markowitz's efficient frontier. The gradient of the line [0,0] to any point on the plot can hence be used as a measure for margin stability where large gradients indicate stable margins.

Our analysis ends in comparing different margin models, three of which developed from the literature to adhere to coming EMIR regulations, one newly introduced and the historical margins themselves. We find that of the automated models, the newly introduced model performs best. While this may not quite outperform SPAN's historical margin levels, our methods are simpler to implement. What we provide here is merely a demonstration of what can be achieved, using detailed calibration methods and measurement systems. The possibilities of margin rules is by no means exhausted. We urge that the academic community continues to invent further rules, on the grounds of our measures and perhaps this can ultimately lead to a truly stable margin method, which can be implemented in the near future.

A. Appendix

Model	Number of points	Proportion (%)
EWMAt-6	1638	12.85
EWMAt-12	2451	19.22
EWMA	1107	86.88
GJRt	2793	21.91
GJR	10586	83.03
K-60	5770	45.25
K-120	5907	46.33
OEWMAt-6	771	6.05
OEWMAt-12	2520	19.78
OEWMA	10325	81.00
OGJR	12563	98.53

Table A.1.: Number of points and proportion above the 95%, χ^2_6 critical statistic for the rolling $LR^{cc}_{t,T}$ statistics for each model. Total number of instances: 12750. Period: 28^{th} January 2009 - 30^{th} December 2013.

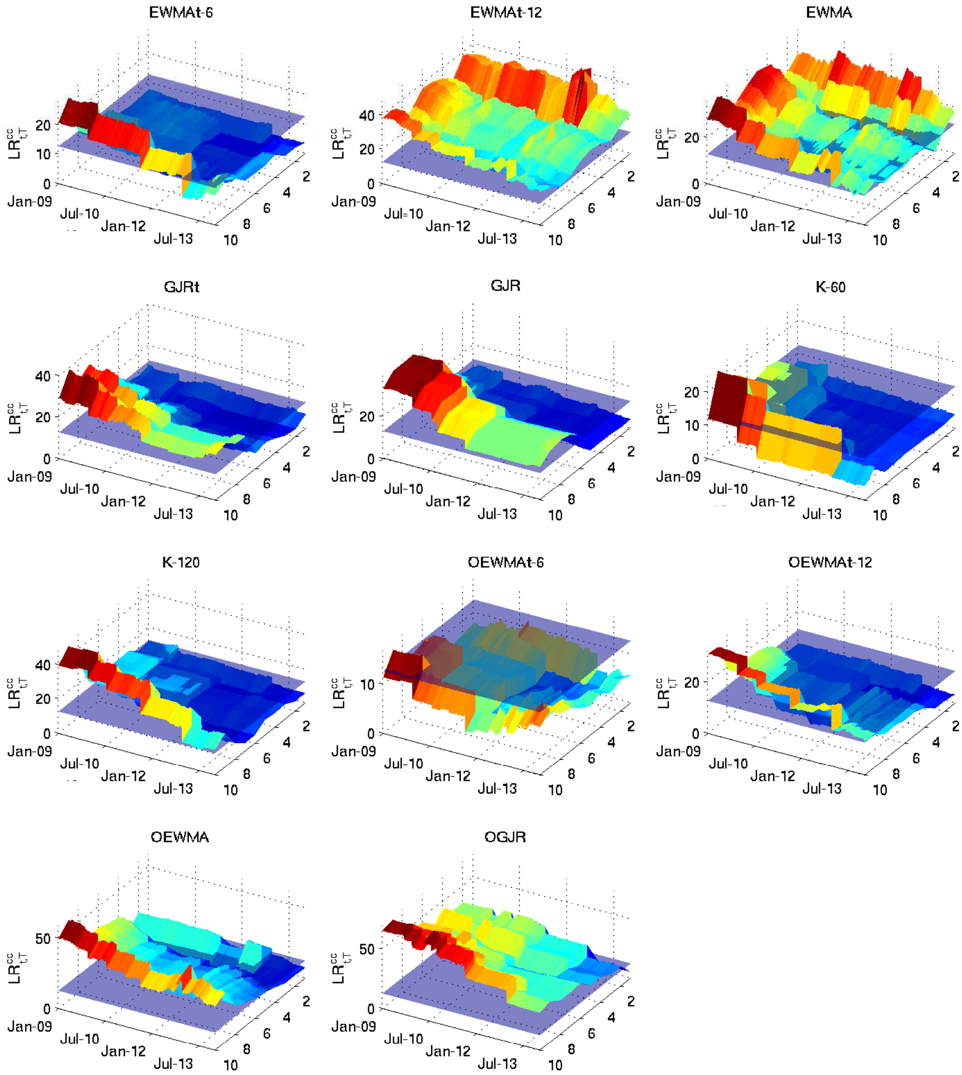

Figure A.1.: Rolling $LR^{cc}_{t,T}$ statistics for all models for the first 10-to-mature series. Period: 28^{th} January 2009 - 30^{th} December 2013. The transparent surface represent the 95% critical statistic for χ^2_6 distribution for the conditional coverage.

Months to Maturity	EWMAt-6	EWMAt-12	EWMA	GJRt	GJR
			$LR^{uc}_{t,T}$		
1	1.94**	4.64*	17.10	1.80**	1.76**
2	0.34**	4.12**	20.91	2.23**	2.18**
3	1.00**	6.22	25.59	3.30**	1.59**
4	1.57**	7.66	25.59	2.11**	2.45**
5	2.41**	9.58	23.19	1.51**	3.18**
6	1.98**	10.43	22.03	1.70**	4.42**
7	2.06**	11.32	22.71	1.87**	3.96**
8	2.43**	10.43	18.76	0.51**	5.38*
9	3.01**	9.58	17.73	0.78**	6.51
10	3.44**	6.92	18.59	1.58**	6.51
			$LR^{cc}_{t,T}$		
1	6.56**	14.12	25.32	17.94	10.53**
2	5.68**	8.19**	30.22	15.02	12.16*
3	5.84**	11.39*	39.25	16.57	15.71
4	8.66**	14.31	39.25	23.34	15.54
5	9.27**	15.62	31.19	21.29	15.30
6	8.50**	16.24	27.83	25.38	21.87
7	8.39**	16.91	26.19	19.72	21.88
8	11.85*	16.24	22.83	27.34	30.89
9	11.93*	15.62	21.99	28.07	30.58
10	12.44*	13.82	22.65	16.34	25.97
			$LR^{in}_{t,T}$		
1	4.62**	9.48*	8.22*	16.14	8.77*
2	5.34**	4.07**	9.31*	12.79	9.98
3	4.84**	5.17**	13.66	13.27	14.12
4	7.09**	6.65**	13.66	21.23	13.09
5	6.86**	6.04**	8.00*	19.78	12.12
6	6.52**	5.81**	5.80**	23.68	17.45
7	6.33**	5.59**	3.48**	17.85	17.92
8	9.42*	5.81**	4.07**	26.83	25.51
9	8.92*	6.04**	4.26**	27.29	24.07
10	9.00*	6.90**	4.06**	14.76	19.46

Table A.2.: Whole-sample, two-tailed, 1% $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998) coverage tests on the Brent futures return series for the first 10 months-to-mature constant-maturity series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29th March 1999 - 30th December 2013. Starting and final return values are not exceedances of the VaR. Models considered: EWMAt-6, EWMAt-12, EWMA, GJRt, GJR.

Months to Maturity	K-60	K-120	OEWMAt-6	OEWMAt-12	OEWMA	OGJR
				$LR^{uc}_{t,T}$		
1	1.57**	0.38**	1.52**	1.82**	12.24	1.95**
2	2.64**	0.42**	0.30**	1.98**	17.06	1.76**
3	3.04**	0.40**	0.73**	3.67**	22.55	3.13**
4	2.86**	0.60**	1.19**	8.77	26.84	6.05
5	2.08**	0.70**	1.98**	9.14	30.10	6.51
6	2.81**	0.40**	3.01**	12.87	28.98	7.63
7	3.04**	0.47**	3.47**	12.61	29.35	9.39
8	3.01**	0.64**	4.64*	10.10	30.22	9.00
9	2.51**	0.42**	3.44**	10.37	32.37	12.19
10	1.65**	0.51**	4.02**	9.95	33.02	12.19
				$LR^{cc}_{t,T}$		
1	8.30**	17.02	7.17**	13.57	20.68	26.09
2	12.19*	22.65	6.00**	6.49**	23.42	31.57
3	13.08	21.49	5.73**	10.10**	28.4	30.45
4	12.56*	20.63	8.53**	15.04	37.91	38.00
5	9.98**	16.32	8.88**	15.28	39.43	43.09
6	9.98**	16.62	11.93*	18.17	38.68	45.24
7	9.99**	16.27	12.27*	17.87	39.05	47.50
8	10.44**	22.41	13.06	16.05	43.86	48.75
9	10.12**	25.400	9.98**	16.16	45.48	62.18
10	16.34	26.16	10.32**	17.59	46.09	64.77
				$LR^{in}_{t,T}$		
1	6.73**	16.64	5.65**	11.75	8.44*	24.14
2	9.55	22.23	5.70**	4.51**	6.36**	29.81
3	10.04	21.09	5.00**	6.43**	5.85**	27.32
4	9.70	20.03	7.34**	6.27**	11.07	31.95
5	7.90*	15.62	6.90**	6.14**	9.33*	36.58
6	7.17**	16.22	8.92*	5.30**	9.70	37.61
7	6.95**	15.80	8.80*	5.26**	9.70	38.11
8	7.43**	21.77	8.42*	5.95**	13.64	39.75
9	7.61**	24.98	6.54**	5.79**	13.11	49.99
10	14.69	25.65	6.30**	7.64**	13.07	52.58

Table A.3.: Whole-sample, two-tailed, 1% $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998) coverage tests on the Brent futures return series for the first 10 months-to-mature constant-maturity series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29^{th} March 1999 - 30^{th} December 2013. Starting and final return values are not exceedances of the VaR. Models considered: K-60, K-120, OEWMAt-6, OEWMAt-12, OEWMA, OGJR.

Months to Mature	EWMAt-6	EWMAt-12	EWMA	GJRt	GJR
			$LR^{uc}_{t,T}$		
1	0.87**	5.33*	20.49	1.47**	13.78
2	1.14**	4.05**	22.03	1.18**	16.51
3	1.09**	7.15	23.74	1.69**	18.40
4	1.46**	8.16	26.84	1.15**	29.32
5	1.00**	9.25	28.79	2.44**	28.32
6	0.87**	9.25	22.71	2.23**	21.47
7	1.57**	7.66	22.71	1.94**	21.95
8	1.68**	6.92	24.17	2.43**	22.74
9	0.73**	7.93	23.42	0.87**	20.48
10	1.15**	7.15	18.76	1.99**	19.44
			$LR^{cc}_{t,T}$		
1	6.00**	8.43**	30.03	10.21**	19.47
2	6.88**	10.74*	29.18	10.50**	22.44
3	13.13	16.10	30.49	10.91*	28.02
4	12.70	14.71	35.12	9.22**	37.01
5	8.91**	13.07	39.39	14.04	37.31
6	6.17**	13.07	29.74	9.71**	30.28
7	6.14**	11.80*	27.27	10.15**	28.86
8	7.06**	11.23*	32.75	9.16**	28.50
9	6.43**	12.19*	35.10	8.49**	23.76
10	6.89**	11.58*	29.51	9.17**	23.90
			$LR^{in}_{t,T}$		
1	5.13**	3.10**	9.54	8.74*	5.69**
2	5.74**	6.69**	7.15**	9.32*	5.93**
3	12.04	8.95*	6.75**	9.22*	9.62
4	11.24	6.55**	8.28*	8.07*	7.69**
5	7.91*	3.82**	10.60	11.60	8.99*
6	5.30**	3.82**	7.03**	7.48**	8.81*
7	4.57**	4.14**	4.56**	8.21*	6.91**
8	5.38**	4.31**	8.58*	6.73**	5.76**
9	5.70**	4.26**	11.68	7.62**	3.28**
10	5.74**	4.43**	10.75	7.18**	4.46**

Table A.4.: Whole-sample, two-tailed, 1% $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998) coverage tests on the Brent futures P&L series for the first 10 months-to-mature constant-maturity series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29^{th} March 1999 - 30^{th} December 2013. Starting and final P&L values are not exceedances of the VaR. Models considered: EWMAt-6, EWMAt-12, EWMA, GJRt, GJR.

Months to Mature	K-60	K-120	OEWMAt-6	OEWMAt-12	OEWMA	OGJR
				$LR^{uc}_{t,T}$		
1	2.08**	1.76**	1.36**	5.71*	13.85	19.51
2	1.77**	3.32**	0.87**	9.41	25.59	24.66
3	1.22**	2.08**	0.87**	8.97	25.05	29.64
4	2.08**	3.55**	0.87**	13.85	30.10	32.37
5	3.54**	3.54**	0.64**	9.82	32.15	35.06
6	4.08**	2.06**	1.00**	10.10	25.76	36.13
7	3.85**	3.32**	1.59**	7.38	28.23	33.71
8	1.65**	2.86**	1.94**	8.74	29.35	31.80
9	0.97**	3.18**	1.94**	6.41	29.35	33.84
10	1.09**	2.81**	1.69**	6.22	24.17	32.55
				$LR^{cc}_{t,T}$		
1	11.70*	9.81**	6.3**	11.44*	23.86	36.74
2	14.16	12.10*	6.00**	19.03	36.03	44.67
3	16.70	15.37	11.96*	20.31	31.56	47.98
4	16.49	17.70	11.96*	23.35	35.66	58.62
5	12.26*	24.33	8.61**	13.63	39.97	60.74
6	13.22	22.14	6.47**	13.77	29.83	58.02
7	10.68*	21.63	8.72**	11.66*	33.97	58.39
8	16.97	25.05	10.15**	12.85	41.12	58.46
9	16.68	25.02	10.15**	16.17	41.12	59.85
10	16.46	25.22	5.00**	19.60	35.13	61.60
				$LR^{in}_{t,T}$		
1	9.62	8.05*	4.94**	5.73**	10.01	17.23
2	12.39	8.78*	5.13**	9.62	10.44	20.01
3	15.48	13.29	11.09	11.34	6.51**	18.34
4	14.41	14.15	11.09	9.50	5.56**	26.25
5	8.72*	20.79	7.97*	3.81**	7.82*	25.68
6	9.14*	20.08	5.47**	3.67**	4.07**	21.89
7	6.83**	18.31	7.13**	4.28**	5.74**	24.68
8	15.32	22.19	8.21*	4.11**	11.77	26.66
9	15.71	21.84	8.21*	9.76	11.77	26.01
10	15.37	22.41	3.31**	13.38	10.96	29.05

Table A.5.: Whole-sample, two-tailed, 1% $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998) coverage tests on the Brent futures P&L series for the first 10 months-to-mature constant-maturity seriess. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29^{th} March 1999 - 30^{th} December 2013. Starting and final P&L values are not exceedances of the VaR. Models considered: K-60, K-120, OEWMAt-6, OEWMAt-12, OEWMA, OGJR.

Rollover series	EWMAt-6	EWMAt-12	EWMA	GJRt	GJR
			$LR^{uc}_{t,T}$		
1^{st}	2.50**	7.52	28.76	2.23**	1.83**
2^{nd}	1.09**	8.66	20.27	0.42**	3.97**
3^{rd}	2.08**	13.39	21.95	2.67**	4.31**
4^{th}	2.74**	15.31	32.26	3.45**	5.04*
5^{th}	5.07*	21.59	34.09	4.13**	5.64*
6^{th}	5.73*	20.17	35.11	6.04	6.31
7^{th}	2.79**	15.63	32.65	2.67**	6.06
8^{th}	3.20**	10.41	28.05	2.34**	7.79
9^{th}	4.44**	16.56	30.87	2.17**	6.42
10^{th}	4.61*	14.78	26.35	3.54**	7.16
			$LR^{cc}_{t,T}$		
1^{st}	6.78**	11.28*	35.53	10.25**	6.38**
2^{nd}	5.95**	12.10*	30.37	6.73**	8.23**
3^{rd}	6.50**	15.15	23.86	9.54**	10.35**
4^{th}	7.52**	18.89	36.71	13.97	10.82*
5^{th}	11.48*	24.06	38.20	16.03	14.23
6^{th}	12.01*	22.87	38.26	16.86	22.84
7^{th}	9.58**	18.75	36.14	29.36	30.00
8^{th}	10.43**	14.48	31.96	26.11	31.43
9^{th}	21.12	26.98	39.90	32.97	30.44
10^{th}	17.35	22.36	33.67	26.83	26.71
			$LR^{in}_{t,T}$		
1^{st}	4.28**	3.76**	6.77**	8.02*	4.55**
2^{nd}	4.86**	3.44**	10.1	6.31**	4.26**
3^{rd}	4.42**	1.76**	1.91**	6.87**	6.04**
4^{th}	4.78**	3.58**	4.45**	10.52	5.78**
5^{th}	6.41**	2.47**	4.11**	11.90	8.59*
6^{th}	6.28**	2.70**	3.15**	10.82	16.53
7^{th}	6.79**	3.12**	3.49**	26.69	23.94
8^{th}	7.23**	4.07**	3.91**	23.77	23.64
9^{th}	16.68	10.42	9.03*	30.80	24.02
10^{th}	12.74	7.58**	7.32**	23.29	19.55

Table A.6.: Whole-sample, two-tailed, 1% $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998) coverage tests on the Brent futures return series for the first 10-to-mature series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29^{th} March 1999 - 30^{th} December 2013. Starting and final return values are not exceedances of the VaR. Models considered: EWMAt-6, EWMAt-12, EWMA, GJRt, GJR.

Rollover series	K-60	K-120	OEWMAt-6	OEWMAt-12	OEWMA	OGJR
			$LR^{uc}_{t,T}$			
1^{st}	0.09**	0.66**	1.75**	1.41**	6.95	2.24**
2^{nd}	1.01**	0.08**	1.19**	3.18**	11.80	2.97**
3^{rd}	1.48**	0.18**	2.24**	9.47	22.16	5.62*
4^{th}	2.19**	0.02**	6.12	16.72	34.58	8.77
5^{th}	1.39**	0.10**	6.33	19.20	37.64	6.95
6^{th}	0.73**	0.10**	7.96	20.38	45.25	7.66
7^{th}	1.00**	0.40**	7.06	20.38	43.62	10.88
8^{th}	0.52**	0.22**	4.95*	17.18	40.78	9.59
9^{th}	2.61**	0.05**	7.01	17.18	52.74	8.24
10^{th}	1.01**	0.00**	5.91*	19.76	45.53	9.48
			$LR^{cc}_{t,T}$			
1^{st}	5.76**	10.39**	7.26**	6.02**	13.59	13.67
2^{nd}	6.16**	9.10**	6.35**	7.27**	27.12	35.11
3^{rd}	9.98**	12.10*	6.75**	12.78	23.08	34.13
4^{th}	10.42**	12.93	9.72**	19.73	38.88	40.85
5^{th}	9.89**	13.78	12.32*	21.86	41.37	30.76
6^{th}	9.92**	13.78	13.23	22.95	47.63	41.54
7^{th}	13.84	14.79	11.75*	22.95	49.19	42.15
8^{th}	9.49**	30.45	10.72*	22.54	48.74	51.68
9^{th}	19.67	41.79	14.34	29.90	70.01	61.08
10^{th}	9.88**	35.85	11.90*	33.10	61.92	58.30
			$LR^{in}_{t,T}$			
1^{st}	5.67**	9.73	5.51**	4.61**	6.64**	11.43
2^{nd}	5.15**	9.02*	5.16**	4.09**	15.32	32.14
3^{rd}	8.5*	11.92	4.51**	3.31**	0.92**	28.51
4^{th}	8.23*	12.91	3.60**	3.01**	4.30**	32.08
5^{th}	8.5*	13.68	5.99**	2.66**	3.73**	23.81
6^{th}	9.19*	13.68	5.27**	2.57**	2.38**	33.88
7^{th}	12.84	14.39	4.69**	2.57**	5.57**	31.27
8^{th}	8.97*	30.23	5.77**	5.36**	7.96*	42.09
9^{th}	17.06	41.74	7.33**	12.72	17.27	52.84
10^{th}	8.87*	35.85	5.99**	13.34	16.39	48.82

Table A.7.: Whole-sample, two-tailed, 1% $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998) coverage tests on the Brent futures return series for the first 10-to-mature series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29^{th} March 1999 - 30^{th} December 2013. Starting and final return values are not exceedances of the VaR. Models considered: K-60, K-120, OEWMAt-6, OEWMAt-12, OEWMA, OGJR.

Rollover series	EWMAt-6	EWMAt-12	EWMA	GJRt	GJR
			$LR^{uc}_{t,T}$		
1^{st}	2.74**	5.30*	21.25	1.32**	21.48
2^{nd}	1.56**	6.27	27.44	1.38**	16.09
3^{rd}	1.98**	7.80	28.74	1.09**	14.00
4^{th}	3.89**	14.25	31.73	3.82**	22.51
5^{th}	4.31**	13.39	34.85	5.33*	22.83
6^{th}	5.27*	15.47	34.78	4.28**	27.16
7^{th}	3.20**	10.87	28.74	3.52**	26.87
8^{th}	1.91**	7.79	26.17	3.40**	18.66
9^{th}	3.71**	12.29	31.73	4.78*	24.59
10^{th}	4.91*	13.23	26.87	4.28**	23.34
			$LR^{cc}_{t,T}$		
1^{st}	5.06**	7.96**	26.69	6.16**	25.42
2^{nd}	4.19**	12.33*	31.33	9.93**	20.49
3^{rd}	5.27**	12.67	34.50	6.62**	18.36
4^{th}	12.39*	17.96	37.27	11.92*	31.69
5^{th}	12.62	16.90	37.85	16.93	30.99
6^{th}	13.23	18.74	37.65	16.31	33.98
7^{th}	8.50**	14.49	32.22	12.22*	33.17
8^{th}	8.01**	12.49*	27.74	15.90	26.55
9^{th}	21.85	24.20	38.47	23.48	38.15
10^{th}	10.58**	21.69	31.87	12.20*	34.40
			$LR^{in}_{t,T}$		
1^{st}	2.32**	2.66**	5.44**	4.84**	3.94**
2^{nd}	2.63**	6.06**	3.89**	8.55*	4.40**
3^{rd}	3.29**	4.87**	5.76**	5.53**	4.36**
4^{th}	8.50*	3.71**	5.54**	8.10*	9.18*
5^{th}	8.31*	3.51**	3.00**	11.60	8.16*
6^{th}	7.96*	3.27**	2.87**	12.03	6.82**
7^{th}	5.30**	3.62**	3.48**	8.70*	6.30**
8^{th}	6.10**	4.70**	1.57**	12.50	7.89*
9^{th}	18.14	11.91	6.74**	18.7	13.56
10^{th}	5.67**	8.46*	5.00**	7.92*	11.06

Table A.8.: Whole-sample, two-tailed, 1% $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998) coverage tests on the Brent futures P&L series for the first 10-to-mature series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29^{th} March 1999 - 30^{th} December 2013. Starting and final P&L values are not exceedances of the VaR. Models considered: EWMAt-6, EWMAt-12, EWMA, GJRt, GJR.

Rollover series	K-60	K-120	OEWMAt-6	OEWMAt-12	OEWMA	OGJR
			$LR^{uc}_{t,T}$			
1^{st}	1.34**	0.18**	2.81**	1.35**	14.05	13.01
2^{nd}	2.33**	1.49**	2.60**	4.95*	22.62	19.15
3^{rd}	3.12**	2.45**	3.82**	10.36	29.51	23.12
4^{th}	2.01**	2.03**	4.91*	16.39	39.59	28.18
5^{th}	2.45**	2.70**	4.91*	17.18	43.62	31.34
6^{th}	3.61**	1.15**	6.12	15.83	44.43	34.78
7^{th}	3.51**	1.01**	4.91*	11.80	41.41	32.05
8^{th}	2.75**	1.00**	3.52**	9.47	37.07	30.69
9^{th}	4.41**	2.36**	4.91*	11.33	37.91	27.44
10^{th}	2.19**	3.64**	4.78*	14.57	36.52	30.08
			$LR^{cc}_{t,T}$			
1^{st}	5.95**	5.62**	5.74**	5.99**	16.25	24.98
2^{nd}	11.07*	6.14**	7.10**	8.71**	25.44	43.87
3^{rd}	10.74*	10.16**	6.08**	14.59	33.14	55.23
4^{th}	12.12*	14.46	12.30*	21.77	43.37	57.01
5^{th}	9.49**	18.00	12.30*	19.92	47.34	51.99
6^{th}	12.46*	19.55	13.19	18.82	47.15	54.38
7^{th}	18.37	20.00	12.19*	15.28	46.69	52.66
8^{th}	19.11	26.17	12.22*	13.85	45.79	55.22
9^{th}	24.89	29.43	15.15	17.35	52.11	57.49
10^{th}	20.76	25.23	12.53*	25.83	53.83	58.97
			$LR^{in}_{t,T}$			
1^{st}	4.61**	5.44**	2.93**	4.64**	2.20**	11.97
2^{nd}	8.74*	4.65**	4.50**	3.76**	2.82**	24.72
3^{rd}	7.62**	7.71**	2.26**	4.23**	3.63**	32.11
4^{th}	10.11	12.43	7.39**	5.38**	3.78**	28.83
5^{th}	7.04**	15.30	7.39**	2.74**	3.72**	20.65
6^{th}	8.85*	18.40	7.07**	2.99**	2.72**	19.60
7^{th}	14.86	18.99	7.28**	3.48**	5.28**	20.61
8^{th}	16.36	25.17	8.70*	4.38**	8.72*	24.53
9^{th}	20.48	27.07	10.24	6.02**	14.20	30.05
10^{th}	18.57	21.59	7.75**	11.26	17.31	28.89

Table A.9.: Whole-sample, two-tailed, 1% $LR^{uc}_{t,T}$, $LR^{cc}_{t,T}$, $LR^{in}_{t,T}$ based on Christoffersen (1998) coverage tests on the Brent futures P&L series for the first 10-to-mature series. ** and * denotes failure to reject the null hypothesis that the VaR model is accurate at 90% and 95% respectively. Out-of-sample period: 29^{th} March 1999 - 30^{th} December 2013. Starting and final P&L values are not exceedances of the VaR. Models considered: K-60, K-120, OEWMAt-6, OEWMAt-12, OEWMA, OGJR.

Bibliography

Abruzzo, N. and Park, Y.-H. (2013). An empirical analysis of futures margin changes: Determinants and policy implications. `http://ssrn.com/abstract=2326687orhttp://dx.doi.org/10.2139/ssrn.2326687`.

Alexander, C. (1999). Optimal hedging using cointegration. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 357(1758):2039–2058.

Alexander, C. (2001). Orthogonal GARCH. *Mastering Risk*, 2:21–38. Financial Times-Pretince Hall: London;.

Alexander, C. and Barbosa, A. (2007). Effectiveness of minimum variance hedging. *Journal of Portfolio Management*, 33(2):46–59.

Alexander, C. and Sheedy, E. (2006). The professional risk managers' guide to financial markets.

Aloui, C. and Mabrouk, S. (2010). Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models. *Energy Policy*, 38(5):2326–2339.

Anderson, R. W. and Danthine, J.-P. (1981). Cross hedging. *Journal of Political Economy*, 89(6):1182–1196.

Back, J., Prokopczuk, M., and Rudolf, M. (2013). Seasonality and the valuation of commodity options. *Journal of Banking and Finance*, 37(2):273–290.

Baillie, R. T. and Myers, R. J. (1991). Bivariate GARCH estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics*, 6(2):109–124.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (2002). Backtesting derivative portfolios with filtered historical simulation (fhs). *European Financial Management*, 8(1):31–58.

Bauwens, L., Hafner, C. M., and Pierret, D. (2013). Multivariate volatility modeling of electricity futures. *Journal of Applied Econometrics*, 28(5):743–761.

Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 35:1–11.

Bennet, B. A. (1990). Commodity futures cross hedging of foreign exchange exposure. *Journal of Futures Markets*, 10:287–306.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Bliss, R. (1997). Movements in the term structure of interest rates. *Economic Review*, 4:16–33.

Board, J., Sutcliffe, C., and Ziemba, W. T. (2003). Applying operations research techniques to financial markets. *Interfaces*, 33(2):12–24.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedacity. *Journal of Econometrics*, 31:301–327.

Booth, G. G., Broussard, J. P., Martikainen, T., and Puttonen, V. (1997). Prudent margin levels in the Finnish stock index futures market. *Management Science*, 43:1177–1188.

Brennan, D. T. (1958). The supply of storage. *American Economic Review*, 48:50–72.

Brennan, M. J. (1986). A theory of price limits in futures markets. *Journal of Financial Economics*, 16(2):213–233.

Brooks, C., Henry, O. T., and Persand, G. (2002). The effect of asymmetries on optimal hedge ratios. *Journal of Business*, 75(2):333–352.

Broussard, J. P. (2001). Extreme-value and margin setting with and without price limits. *The Quarterly Review of Economics and Finance*, 41(3):365–385.

Broussard, J. P. and Booth, G. (1998). The behavior of extreme values in Germany's stock index futures: An application to intradaily margin setting. *European Journal of Operational Research*, 104(3):393–402.

Brown, S. L. (1985). A reformulation of the portfolio model of hedging. *American Journal of Agricultural Economics*, 67:508–512.

Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2238.

Chan, W. H. and Young, D. (2006). Jumping hedges: An examination of movements in copper spot and futures markets. *Journal of Futures Market*, 26(2):169–188.

Chang, C.-L., McAleer, M., and Tansuchat, R. (2011). Crude oil hedging strategies using dynamic multivariate GARCH. *Energy Economics*, 33:912–923.

Chen, Q., Gerlach, R., and Lu, Z. (2012). Bayesian value-at-risk and expected shortfall forecasting via the asymmetric laplace distribution. *Computational Statistics & Data Analysis*, 56(11):3498 – 3516. 1st issue of the Annals of Computational and

Financial Econometrics Sixth Special Issue on Computational Econometrics.

Chiu, C.-L., Chiang, S.-M., Hung, J.-C., and Chen, Y.-L. (2006). Clearing margin system in the futures markets–applying the value-at-risk model to taiwanese data. *Physica A: Statistical Mechanics and its Applications*, 367(0):353 –374.

Chou, P.-H., Lin, M.-C., and Yu, M.-T. (2006). Margins and price limits in Taiwan's stock index futures market. *Emerging Markets Finance & Trade*, 42(1):62–88.

Christoffersen, P., Hahn, J., and Inoue, A. (2001). Testing and comparing value-at-risk measures. *Journal of Empirical Finance*, 8(3):325 – 342.

Christoffersen, P. and Pelletier, D. (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2(1):84–108.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841 – 862.

Coen-Pirani, D. (2005). Margin requirements and equilibrium asset prices. *Journal of Monetary Economics*, 52(2):449–475.

Cotter, J. (2001). Margin exceedences for european stock index futures using extreme value theory. *Journal of Banking & Finance*, 25(8):1475 – 1502.

Cotter, J. and Dowd, K. (2006). Extreme spectral risk measures: An application to futures clearinghouse margin requirements. *Journal of Banking & Finance*, 30(12):3469 – 3485.

Cousin, A. and Bernardino, E. D. (2013). On multivariate extensions of value-at-risk. *Journal of Multivariate Analysis*, 119(0):32 – 46.

Dale, C. (1981). The hedging effectiveness of currency futures markets. *Journal of Futures Markets*, 1(1):77–88.

Day, T. E. and Lewis, C. M. (1997). Initial margin policy and stochastic volatility in the crude oil futures market. *The Review of Financial Studies*, 10(2):303–332.

Day, T. E. and Lewis, C. M. (2004). Margin adequacy and standards: An analysis of the crude oil futures market. *The Journal of Business*, 77(1):101–135.

Dempster, M. A. H., Medova, E., and Tang, K. (2008). Long term spread option valuation and hedging. *Journal of Banking & Finance*, 32:2530–2540.

Diamandis, P. F., Drakos, A. A., Kouretas, G. P., and Zarangas, L. (2011). Value-at-risk for long and short trading positions: Evidence from developed and emerging equity markets. *International Review of Financial Analysis*, 20(3):165–176.

Driessen, J., Klaassen, P., and Melenberg, B. (2002). The performance of multi-factor term structure models for pricing and hedging caps and swaptions. *Journal of*

Financial and Quantitative Analysis, 38:635–672.

Duffie, D. and Zhu, H. (2011). Does a central clearing counterparty reduce counterparty risk? *Review of Asset Pricing Studies*, 1(1):74–95.

Dunis, C., Laws, J., and Evans, B. (2008). Trading futures spread portfolios: applications of higher order recurrent networks. *European Journal of Finance*, 14:503–521.

Eckardt, Walter L., J. and Rogoff, D. L. (1976). 100% margins revisited. *The Journal of Finance*, 31(3):995–1000.

Ederington, L. H. (1979). The hedging performance of the new futures markets. *Journal of Finance*, 34(1).

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, 97(2):526 – 547.

Engle, R. (2004). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008.

Engle, R. F. and Kroner, K. K. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150.

Engle, R. F. and Manganelli, S. (2004). CAViaR. *Journal of Business & Economic Statistics*, 22(4):367–381.

European Union (2013). Commission delegated regulation (EU) no 153/2013. *Official Journal of the European Union*.

Fama, E. F. (1965). The behaviour of stock market prices. *Journal of Business*, 38:34–105.

Fama, E. F. and French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of Futures Market*, 60(1):55–73.

Fama, E. F. and French, K. R. (1988). Business cycles and the behavior of metals prices. *Journal of Finance*, 43:1075–1093.

Fan, Y., Zhang, Y.-J., Tsai, H.-T., and Wei, Y.-M. (2008). Estimating value at risk of crude oil price and its spillover effect using the GED-GARCH approach. *Energy Economics*, 30(6):3156 – 3171.

Fenn, G. W. and Kupiec, P. (1993). Prudential margin policy in a futures-style settlement system. *Journal of Futures Markets*, 13(4):389–408.

Figlewski, S. (1984a). Hedging performance and basis risk in stock index futures. *Journal of Finance*, 39(3):657–669.

Figlewski, S. (1984b). Margins and market integrity: Margin setting for stock index futures and options. *Journal of Futures Markets*, 4(3):385–416.

Gagnon, L., Lupny, G. L., and McCurdy, T. H. (1998). Hedging foreign currency portfolios. *Journal of Empirical Finance*, 5:197–220.

Galai, N. (1979). A proposal for indexes for traded call options. *Journal of Finance*, 34:1157–1172.

Garleanu, N. and Pedersen, L. H. (2011). Margin-based asset pricing and deviations from the law of one price. *Review of Financial Studies*, 24(6):1980–2022.

Gay, G. D., Hunter, W. C., and Kolb, R. W. (1986). A comparative analysis of futures contract margins. *Journal of Futures Markets*, 6(2):307–324.

Geman, H. and Kharoubi, C. (2008). WTI crude oil futures in portfolio diversification: The time-to-maturity effect. *Journal of Banking & Finance*, 32:2553–2559.

Gibson, R. and Schwartz, E. S. (1990). Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, 45(3):959–976.

Giot, P. and Laurent, S. (2003). Market risk in commodity markets: a VaR approach. *Energy Economics*, 25(5):435 – 457.

Girma, P. B. and Paulson, A. S. (1999). Risk arbitrage opportunities in petroleum futures spreads. *Journal of Futures Markets*, 19:931–955.

Grier, K. B., Henry, O. T., Olekalns, N., and Shields, K. (2004). The asymmetric effects of uncertainty on inflation and output growth. *Journal of Applied Econometrics*, 19:551–565.

Haigh, M. S. and Holt, M. T. (2000). Hedging multiple price uncertainty in international grain trade. *American Journal of Agricultural Economics*, 82(4):881–889.

Haigh, M. S. and Holt, M. T. (2002). Crack spread hedging: Accounting for time varying volatility spillovers in the energy futures markets. *Journal of Applied Econometrics*, 17:269–289.

Hammoudeh, S., Malik, F., and McAleer, M. (2011). Risk management of precious metals. *The Quarterly Review of Economics and Finance*, 51(4):435 – 441.

Hardouvelis, G. A. (1990). Margin requirements, volatility, and the transitory component of stock prices. *The American Economic Review*, 80(4):736–762.

Hardouvelis, G. A. and Theodossiou, P. (2002). The asymmetric relation between initial margin requirements and stock market volatility across bull and bear markets.

The Review of Financial Studies, 15(5):1525–1559.

Herbst, A. F., Kare, D. D., and Caples, S. C. (1989). Hedging effectiveness and minimum risk hedge ratios in the presence of autocorrelation: Foreign currency futures. *Journal of Financial Engineering*, 9(3):185–197.

Herbst, A. F. and Marshall, J. F. (1994). Convergence-adjusted composite hedging. *Journal of Fnancial Engineering*, 3(2):109–135.

Hill, J. and Schneeweis, T. (1982). The hedging effectiveness of foreign currency futures. *Journal of Financial Research*, 5:95–104.

Hsieh, D. A. and Miller, M. H. (1990). Margin regulation and stock market volatility. *The Journal of Finance*, 45(1):3–29.

Huang, A. Y. (2010). An optimization process in value-at-risk estimation. *Review of Financial Economics*, 19(3):109 – 116.

Huang, D., Yu, B., Fabozzi, F. J., and Fukushima, M. (2009). CAViaR-based forecast for oil price risk. *Energy Economics*, 31(4):511 – 518.

Huang, G., Wan, J., and Chen, C. (2011). An active margin system and its application in Chinese margin lending market. available at `http://arxiv.org/abs/1101.3974`.

Ji, Q. and Fan, Y. (2011). A dynamic hedging approach for refineries in multiproduct oil markets. *Energy*, 36:881–887.

Johnson, L. L. (1960). The theory of hedging and speculation in commodity futures. *Review of Economic Studies*, 27(3):139–151.

Kahl, H. (1986). A reformulation of the portfolio model of hedging: Comment. *American Journal of Agricultural Economics*, 68(4):1007–1009.

Kaldor, N. (1939). Speculation and economic stability. *Review of Economic Studies*, 7:1–27.

Kawakatsu, H. (2006). Matrix exponential GARCH. *Journal of Econometrics*, 134(1):95 – 128.

Kroner, K. F. and Sultan, J. (1993). Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis*, 28(4):535–551.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Kumar, R., Ferris, S. P., and Chance, D. M. (1991). The differential impact of federal reserve margin requirements on stock return volatility. *Financial Review*, 26(3):343–366.

Kupiec, P. H. (1993). Futures margins and stock price volatility: Is there any link? *Journal of Futures Markets*, 13(6):677–691.

Kupiec, P. H. (1994). The performance of S&P 500 futures product margins under the span margining system. *Journal of Futures Markets*, 14(7):789–811.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2).

Kupiec, P. H. and Sharpe, S. A. (1991). Animal spirits, margin requirements, and stock price volatility. *The Journal of Finance*, 46(2):717–731.

Kupiec, P. H. and White, A. P. (1996). Regulatory competition and the efficiency of alternative derivative product margining systems. *Journal of Futures Markets*, 16(8):943–968.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Lam, K., Sin, C.-Y., and Leung, R. (2004). A theoretical framework to evaluate different margin-setting methodologies. *Journal of Futures Markets*, 24(2):117–145.

Lam, K., Yu, P., and Lee, P. (2010). A margin scheme that advises on when to change required margin. *European Journal of Operational Research*, 207(1):524–530.

Largay, J. A. and West, R. R. (1973). Margin changes and stock price behavior. *Journal of Political Economy*, 81(2):328–339.

Lee, H.-T. (2009). Optimal futures hedging under jump switching dynamics. *Journal of Empirical Finance*, 16:446–456.

Lee, H.-T. (2010). Regime switching correlation hedging. *Journal of Empirical Finance*, 34:2728–2741.

Lee, H.-T. and Yorder, J. K. (2007). A bivariate markov regime switching GARCH approach to estimate time varying minimum variance hedge ratios. *Applied Economics*, 39:1253–1265.

Leland, H. E. (1985). Option pricing and replication with transactions costs. *The Journal of Finance*, 40(5):1283–1301.

Lence, S. H., Kimle, K. L., and L., H. M. (1993). A dynamic minimum variance hedge. *American Journal of Agricultural Economics*, 75(4):1063–1071.

Lien, D. (2005). The use and abuse of the hedging effectiveness measure. *International Review of Financial Analysis*, 14(2):277–282.

Lien, D. (2008). Asymmetric effect of basis on dynamic futures hedging: Empirical evidence from commodity markets. *Journal of Banking & Finance*, 32:187–198.

Lien, D. (2009). A note on the effectiveness of GARCH models. *International Review of Economics and Finance*, 18:110–112.

Longin, F. M. (1999). Optimal margin level in futures markets: Extreme price movements. *Journal of Futures Markets*, 19(2):127–152.

Ma, C. K., Kao, G. W., and Frohlich, C. (1993). Margin requirements and the behavior of silver futures prices. *Journal of Business Finance & Accounting*, 20(1):41–60.

Marshall, B. R., Nguyen, N. H., and Visaltanachoti, N. (2012). Commodity liquidity measurement and transaction costs. *Review of Financial Studies*, 25(2):599–638.

Miller, S. E. (1985). Simple and multiple cross-hedging of millfeeds. *The Journal of Futures Markets,*, 5(1):21–28.

Milonas, N. T. (1991). Measuring seasonalities in commodity markets and the half-month effect. *Journal of Futures Market*, 11(3):331–345.

Moschini, G. and Myers, R. J. (2002). Testing for constant hedge ratios in commodity markets: a multivariate GARCH approach. *Journal of Empirical Finance*, 9(5):589–603.

Myers, R. J. and Thompson, S. R. (1989). Generalized optimal hedge ratio estimation. *American Journal of Agricultural Economics*, 71(4):858–868.

Nguyen, V. T. T., Carbonez, K. A. E., and Sercu, P. (2011). Hedging with two futures contracts: Simplicity pays. *European Financial Management*, 17:806–834.

Nomikos, N. K. and Pouliasis, P. K. (2014). Petroleum term structure dynamics and the role of regimes. *Journal of Futures Markets*, pages n/a–n/a.

Parra, M. A., Terol, A. B., and Uria, M. R. (2001). A fuzzy goal programming approach to portfolio selection. *European Journal of Operational Research*, 133(2):287–297.

Paschke, R. and Prokopczuk, M. (2009). Integrating multiple commodities in a model of stochastic price dynamics. *Journal of Energy Markets*, 2(3).

Pennanen, T. and Penner, I. (2010). Hedging of claims with physical delivery under convex transaction costs. *SIAM Journal on Financial Mathematics*, 1(1):158–178.

Phylaktis, K. and Aristidou, A. (2013). Margin changes and futures trading activity: a new approach. *European Financial Management*, 19(1):45–71.

Rahman, S. M., Turner, S. C., and Cost, E. F. (2001). Cross-hedging cottonseed meal. *Journal of Agribusiness*, 19(2):163–171.

Romero, C. (2004). A general structure of achievement function for a goal programming model. *European Journal of Operational Research*, 153(3):675 – 686.

Routledge, B. R. and Seppi, D. J. (2000). Equilibrium forward curves for commodities. *Journal of Finance*, 55(3).

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Finance*, 6(2):41–49.

Santos, T. and Scheinkman, J. A. (2001). Competition among exchanges. *The Quarterly Journal of Economics*, 116(3):1027–1061.

Schwartz, E. S. (1997). The stochastic behaviour of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52:923–973.

Shanker, L. and Balakrishnan, N. (2005). Optimal clearing margin, capital and price limits for futures clearinghouses. *Journal of Banking & Finance*, 29(7):1611 – 1630.

Shanker, L. and Balakrishnan, N. (2006). Price limits and capital requirements of futures clearinghouses. *European Journal of Operational Research*, 168(2):281 – 290.

Sorensen, C. (2002). Modelling seasonality in agricultural commodity futures. *Journal of Futures Markets*, 22(5):271–289.

Stein, J. L. (1961). The simultaneous determination of spot and futures prices. *American Economic Review*, 51:1012–1025.

Stulz, R. M. (1984). Optimal hedging policies. *Journal of Financial and Quantitative Analysis*, 19(2):127–140.

Su, Y. C., Huang, H. C., and Lin, Y. J. (2011). GJR-GARCH model in value-at-risk of financial holdings. *Applied Financial Economics*, 21(24):1819–1829.

Tanlapco, E., Lawarree, J., and Liu, C.-C. (2002). Hedging with futures contracts in a deregulated electricity industry. *Power Systems, IEEE Transactions on*, 17(3):577–582.

Telser, L. G. (1981). Margins and futures contracts. *Journal of Futures Markets*, 1(2):225–253.

Tolmasky, C. and Hindanov, D. (2002). Principal components analysis for correlated curves and seasonal commodities: The case of the petroleum market. *Journal of Futures Markets*, 22(11):1019–1035.

Van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17(5):549–564.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68:97–126.

Witt, H. J., Schroeder, T. C., and Hayenga, L. M. (1987). Comparison of analytical approaches for estimating hedge ratios for agricultural commodities. *Journal of Futures Market*, 7:135–146.

Working, H. (1953). Futures trading and hedging. *American Economic Review*, 43(3):314–343.

Yang, H.-Y. and Yan, H.-L. (2008). Comparative empirical study on the margin setting of stock index futures calendar spread trading. *Systems Engineering - Theory & Practice*, 28(8):132–138.