

SEMI-PARAMETRIC VOLATILITY SURFACE ESTIMATION

by

Amin Afshinmanesh
BSc. in Accounting, School of Economic Sciences, 2009

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN FINANCE

In the Master of Science in Finance Program
of the
Faculty
of
Business Administration

© Amin Afshinmanesh, 2015
SIMON FRASER UNIVERSITY
Spring 2015

All rights reserved. However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for *Fair Dealing*. Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: **Amin Afshinmanesh**

Degree: **Master of Science in Finance**

Title of Project: **Semi-parametric volatility surface estimation**

Supervisory Committee:

Dr Andrey Pavlov
Senior Supervisor
Professor

Dr Phil Goddard
Second Reader
Adjunct Professor

Date Approved: _____

Abstract

In this paper, we have investigated the performance of a semi-parametric model of volatility surfaces by applying it to the prices of exchange-traded options on crude oil futures in a highly volatile environment. We have also reviewed the performance of a parametric and a non-parametric model in such market conditions. To confirm our findings, we have provided some performance metrics as well as further analysis on the dynamics of the underlying futures. Based on our work, we have shown that the above models do not perform well in highly volatile conditions.

The overall contribution of this paper is to determine whether local volatility models in general are applicable to options on crude oil futures in a highly volatile market or we should move towards stochastic volatility approaches.

Keywords: Semi-parametric; Parametric; Non-parametric; Volatility surface; Local volatility; Stochastic volatility

Dedication

To my parents who are my dearest...

Acknowledgements

First, I would like to thank Ashkan Ziabakhshdeylami who has been an inspiration to me since the very beginning of the program. I would also like to extend my gratitude to Dr Andrey Pavlov and Dr Phil Goddard from whom I have learnt a lot and who have supported me during this research.

Many thanks to Dr Avi Bick and Dr Gary Parker for the valuable knowledge I have gained during their amazing courses.

And last but not least, I would like to thank my friend and my classmate Laleh Samii for the wonderful hours we have spent working together.

Table of Contents

Approval	ii
Abstract	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Figures	vii
List of Tables	viii
1: Introduction	1
1.1 Background	2
1.2 Literature review	3
2: Models	5
2.1 Parametric Model	5
2.2 Non-parametric Model	6
2.3 Semi-parametric Model	7
3: Methodology	10
3.1 Data collection	10
3.2 Risk free rate	12
3.3 Implied volatilities	13
3.4 Fitting the parametric model	13
3.5 Fitting the Non-parametric model	15
3.6 Fitting the Semi-parametric model	17
3.7 Further analysis	21
4: Conclusion and future work	25
Appendices	26
Appendix A: MATLAB Codes	27
Appendix B: Copyright License	36
References	40

List of Figures

Figure 3-1: USD Libor yield curve as of 8 th Oct 2014	13
Figure 3-2: Comparison of volatility surfaces_Parametric model	14
Figure 3-3: Comparison of volatility surfaces_Non-parametric model.....	16
Figure 3-4: Comparison of volatility slices_Semi-parametric model.....	19
Figure 3-5: Comparison of volatility surfaces_Semi-parametric model	20
Figure 3-6: Daily log-return of Crude oil futures	23
Figure 3-7: Empirical distribution of futures' daily log-returns vs normal distribution	23
Figure 3-8: QQ Plot of futures' daily log-returns vs standard normal distribution.....	24

List of Tables

Table 3-1: Crude oil active future contracts on NYMEX.....	10
Table 3-2: Structure of the tickers used (1 st part)	10
Table 3-3: Structure of the tickers used (2nd part).....	11
Table 3-4: Tickers of the underlyings of the 3 datasets.....	11
Table 3-5: Datasets	12
Table 3-6: USD Libor rate as of 8 th Oct 2014	12
Table 3-7: Parameters of the parametric model.....	14
Table 3-8: Parameters of the semi-parametric model (Dataset 1)	18
Table 3-9: Parameters of the semi-parametric model (Dataset 2)	18
Table 3-10: Parameters of the semi-parametric model (Dataset 3)	18
Table 3-11: Mean and standard deviation of residuals, from Borovkova and Permana (2009)	22
Table 3-12: Mean and standard deviation of residuals of three datasets.....	22
Table A-1: MATLAB_Yield curve using LIBOR rates	27
Table A-2: MATLAB_Estimation of implied volatilities.....	27
Table A-3: MATLAB_Calculation of moneyness	27
Table A-4: MATLAB_Calculation of weights for the semi-parametric model.....	27
Table A-5: MATLAB_Parametric model	29
Table A-6: MATLAB_Non-parametric model.....	30
Table A-7: MATLAB_Semi-parametric model	32
Table A-8: MATLAB_Semi-parametric model (2 nd approach)	35

1: Introduction

In this paper, we deal with three models of volatility surfaces. A semi-parametric model that was introduced in Borovkova and Permana (2009), a parametric model (Dumas et al., 1998) and a non-parametric model (Bourke, 1998). We have tested the performance of these models by applying them to options on crude oil futures during highly volatile market conditions.

Our focus in this project is mainly on the performance of the semi-parametric approach. The other two methods were already investigated in Borovkova and Permana (2009), but for comparison purposes and in order to test their feasibility in highly volatile market conditions, we have presented them here.

The crude oil futures market started a heavy downward trend towards the end of June 2014 and ever since and up to the date we have prepared this document, it has not rebounded to the previous levels. This downward trend was accompanied by high volatilities as well as more uncertainties.

As we will discuss in the literature review section, modeling implied volatilities in commodity markets, especially crude oil is less studied than other markets such as equity and foreign exchange. The semi-parametric model that we have tested here is one of the very few attempts to address this shortage.

In the semi-parametric approach, a quadratic function is fitted to implied volatilities at each maturity. We call each of these fitted quadratic functions a volatility slice. The coefficients of these quadratic functions are calculated through either the ordinary least-squares method or a weighted least-squares approach. After this stage, the volatility slices are connected to each other through bi-cubic spline interpolation to build the whole volatility surface. The advantage of a semi-parametric approach is that we can have different shapes of volatility slices at different maturities, such as a smile, a smirk or a skew and it provides a relatively good fit on at-the-money implied volatilities.

Based on our analysis we have shown that when volatilities reach record high levels in the market this method fails to provide a good fit to implied volatilities. This fact becomes more apparent when the number of liquid options increases and when we move away from at-the-money strike/moneyness. In addition, in this model it is implicitly assumed that market only behaves in

certain ways and a quadratic function is able to capture and depict those shapes. This is rather simplistic and restrictive, as we cannot force the market to behave in specific ways.

In order to verify our observations we have estimated some metrics and have compared them with the ones in Borovkova and Permana (2009). Our comparison verifies our findings and the fact that the performance of these models exacerbates when we are dealing with high volatility.

We have also conducted some work to see whether local volatility models in general are useful in modeling implied volatilities in the crude oil market. We have noted volatility clustering in the futures daily returns as well as high peak and fat tails in the distribution. These features along with the fact that daily returns on futures do not follow normal distribution suggest that the volatility itself should be considered as a stochastic process.

Traders usually use volatility surfaces as a tool to price options that do not exist on exchanges or to price OTC options with more complex structures. So in the modeling of the volatility surfaces the choice of model becomes an essential part that we suggest should be decided upon by considering the dynamics of the underlying asset. In the case of options on crude oil futures we believe stochastic volatility models would lead to better results.

This paper is structured as follows: For the remainder of this section we provide background knowledge and review some of the literature in volatility modeling. In section (2), we introduce the models that we have applied in our analysis. Section (3) covers our research methodology, data analysis and results. In section (4) we give our conclusion.

1.1 Background

Volatility is a parameter that cannot be directly observed the same way as interest rates and stock prices. But the best thing we can do is to measure it using statistical methods. But such measure would be backward looking while we really want to take a forward-looking approach. In pricing options, we care about the actual volatility for the remaining life of the option. To deal with this situation, the best thing we can do is to use implied volatilities. Implied volatility is the market's best estimate of what volatility will do in the future. In this sense, the implied volatility of a 6-month option may contain information about how the actual volatility will behave during the next six months (Wilmott, 2007).

In a theoretical sense, implied volatility is the volatility of an underlying asset, which when substituted into an option pricing model like the Black-Scholes, gives a theoretical price equal to the market price of the underlying. The Black-Scholes (1973) formula for an option takes the time to maturity, interest rate, strike price, spot price of underlying together with volatility as input and generates the price as an output.

It has been extensively argued that assumptions underlying the Black-Scholes model, such as constant volatility and lognormal distribution of the underlying asset are not realistic. The Black-Scholes (1973) model, for example, assumes the asset price follows geometric Brownian motion with constant volatility. Consequently, all options on the same asset should provide the same implied volatility. In practice, however, Black-Scholes implied volatilities tend to differ across exercise prices and times to expiration (Dumas et al., 1998).

Traders still use the Black-Scholes model but not in the way it was originally introduced. They allow the volatility parameter to depend on the strike price and time to maturity. In this case, the most straightforward approach for modeling the implied volatility would be to use local volatility models that define volatility as a deterministic function of time and underlying's price (Wilmott, 2007) and (Hull, 2012).

Although local volatility models have some attractive features, but without any doubt we still cannot forecast the future actual volatility based on these models accurately. To further address these issues stochastic volatility models have been developed. Stochastic volatility is the notion that volatility varies in a random fashion (Gatheral, 2006). So in addition to modeling the underlying as random, in stochastic volatility models, the volatility parameter itself is also modeled as a random process. To link stochastic volatilities to local volatilities, Gatheral (2006) shows that local variance is a conditional expectation of the instantaneous variance.

A plot of the implied volatility of an option with a certain life as a function of its strike price is known as a volatility smile. And when the implied volatility at a specific strike price is plotted as a function of time to maturity we have volatility term structure. Volatility surfaces combine volatility smiles with the volatility term structures to tabulate the volatilities appropriate for pricing an option with any strike price and any maturity (Hull, 2012). Volatility surfaces are important in the sense that option traders use them as a tool in order to price options.

In this paper we deal with three types of modeling volatility surfaces: parametric, non-parametric and semi-parametric. In a parametric model, we build a volatility surface by just estimating a finite set of parameters. Using a non-parametric approach, the parameters would be in infinite-dimensional parameter spaces and finally a semi-parametric method is comprised of both finite-dimensional and infinite-dimensional unknown parameters of interest.

1.2 Literature review

Modeling the volatility surfaces is subject of extensive research in the financial markets and more specifically the equity markets. But implied volatility is much less studied in the commodities

market due to limited liquidity and complicated structure of commodity options (Borovkova and Permana, 2009).

Derman and Kani (1994), Dupire (1994), and Rubinstein (1994) hypothesized that asset return volatility is a deterministic function of asset price and time, and developed a deterministic volatility function (DVF) option valuation model that has the potential of fitting the observed cross section of option prices exactly. To verify this, Dumas et al. (1998) examined the predictive and hedging performance of the DVF option valuation model by using S&P 500 options from June 1988 through December 1993 and demonstrated that the DVF model is no better than an ad hoc procedure that merely smooths Black–Scholes (1973) implied volatilities across exercise prices and times to expiration.

In their article, Gatheral and Jacquier (2014) show how to calibrate the widely used SVI parameterization of the implied volatility smile in such a way as to guarantee the absence of static arbitrage. In particular, they exhibit a large class of arbitrage-free SVI volatility surfaces with a simple closed-form representation and demonstrate the high quality of typical SVI fits with a numerical example using SPX options data.

Murphy and Ronn (2014) calibrated the volatility skew by the Merton (1976) jump-diffusion option-pricing model using market prices for crude-oil futures options and the prices of their underlying futures contracts. They demonstrated that the jump-diffusion parameters bear a close relationship to concurrent economic, financial and geopolitical events.

Boudon (2007) investigated implied volatility surface patterns for call options on crude oil futures. They tested the capacity of quadratic Implied Volatility Functions (IVFs) to provide a good forecast of option prices and found that the profile of crude oil implied volatility is too complex to be fully explained by IVFs.

It is generally accepted that the implied volatilities from liquidly traded options, specifically at-the-money options, provide the best forecast of future volatility. In their paper Wang et al. (2009) showed that the degree of the bias of the implied volatility estimation is directly related to the option's moneyness and maturity and such bias can be minimized by choosing option with the highest Vega to estimate volatility. They used the S&P 500 index option to test the informational content of the implied volatility from options with highest Vega in comparison to that of at-the-money implied volatility and demonstrated that the implied volatility from options with highest Vega outperforms at-the-money implied volatility, especially for longer forecasting horizons.

2: Models

2.1 Parametric Model

The parametric model is a quadratic function of either option's strike (E) or option's moneyness (MN), and a linear function of time to maturity (τ). In addition, it captures the combined effect of strike/moneyness and time to maturity.

Moneyness:

Moneyness is a unit-less quantity defined as follows:

$$(2-1) \quad MN = \frac{\log\left(\frac{F(t, T)}{E}\right)}{\sigma\sqrt{\tau}}$$

In which the symbols used are:

$$\tau = T - t$$

$F(t, T)$: Futures price (Here, Crude Oil Futures)

E : Option's strike

σ : Option's implied volatility

τ : Time to maturity of the option

T : Maturity of the option

t : Current time

For a call option, positive moneyness corresponds to In-The-Money options (ITM), negative moneyness corresponds to Out-of-The-Money options (OTM) and zero moneyness corresponds to At-The-Money options (ATM). And for a put option positive moneyness corresponds to OTM options, negative moneyness corresponds to ITM options and zero moneyness to ATM options.

Depending on whether moneyness or strike is used, the parametric approach models the implied volatility surface through either of the following equations:

For moneyness:

$$(2-2) \quad \sigma(MN, \tau) = b_1 + b_2(MN) + b_3(MN)^2 + b_4\tau + b_5(MN)\tau + \varepsilon_{MN,\tau}$$

For strike:

$$(2-3) \quad \sigma(E, \tau) = b_1 + b_2(E) + b_3(E)^2 + b_4\tau + b_5(E)\tau + \varepsilon_{E,\tau}$$

In these equations, σ is the implied volatility of an option with strike (E)/moneyness (MN) and time to maturity τ . b_i 's are unknown model coefficients, and ε 's are model errors.

The coefficients of this model are estimated by ordinary least-squares (OLS) method and using the observed implied volatilities. b_1 determines the displacement of the parabola with respect to at-the-money options, b_3 determines the wideness of the smile, b_4 determines the maturity effect and b_5 captures the combined effect of maturity and moneyness/strike.

The serious drawback of this model is that if we slice it at different maturities, the resulting slices have always the same shape. On the other hand the linear relationship with time to maturity is just a simplistic way of capturing the maturity effect.

2.2 Non-parametric Model

In the non-parametric model, implied volatilities for unobserved options are interpolated using a nearest neighbour weighted interpolation method.

Assume that we have observed N option prices with strikes and times to maturity of (E_i, τ_i) , where $i=1, \dots, N$. Then for an unobserved pair of (E, τ) , the implied volatility is estimated using the following equation:

$$(2-4) \quad \sigma(E, \tau) = \sum_{k=1}^N \left[\frac{\sigma_k}{[(\tau_k - \tau)^2 + (E_k - E)^2]^{p/2}} \right] \Bigg/ \sum_{k=1}^N \left[\frac{1}{[(\tau_k - \tau)^2 + (E_k - E)^2]^{p/2}} \right]$$

The parameter p in the above equation determines the relative importance of distant samples. The estimation is generally better when higher values of p are used. For the purpose of this paper, we have used p=8 which was utilized in Borovkova and Permana (2009).

This method works fine if just a few numbers of liquid options are available, but if there are many liquid options in the market, it will have a rather complex shape. Moreover, if we slice the volatility

surface estimated using this approach at different maturities, each slice may have several minima and maxima, and the whole surface may not address the different arbitrage issues properly (More on this in section 3).

2.3 Semi-parametric Model

This method combines the attractive features of both parametric and semi-parametric approaches. In this method, a quadratic function is fitted to available option prices at different maturities:

$$(2-5) \quad \sigma_\tau = b_{1,\tau} + b_{2,\tau}(X) + b_{3,\tau}(X)^2 + \varepsilon_\tau$$

First we fix the time to maturity and then apply the above equation to the existing options at that maturity. A semi-parametric method can be fitted either based on strike or moneyness:

For strike:

$$(2-6) \quad \hat{\sigma}_i = b_1 + b_2(E_i) + b_3(E_i)^2 + \varepsilon_i$$

For moneyness:

$$(2-7) \quad \hat{\sigma}_i = b_1 + b_2(MN_i) + b_3(MN_i)^2 + \varepsilon_i$$

One of the drawbacks of this method is that the number of options at each maturity should be at least three, otherwise we cannot apply equations (2-6) or (2-7). This issue is troublesome especially during the implementation phase, as we have to check whether three options exist at a maturity or not.

After we have built each slice, then we use the bi-cubic spline method to interpolate between different slices and thus estimate the volatility surface.

The parameters of the quadratic functions above can be estimated either by ordinary least-squares method or a weighted least-squares approach. For doing so, first we minimize the sum of squared residuals as in equation (2-8):

$$(2-8) \quad S = \sum_{i=1}^n \omega_i (\hat{\sigma}_i - \sigma_i)^2$$

In case the volatility surface is estimated using moneyness (MN) we have:

$$(2-9) \quad MN^T = \begin{bmatrix} 1 & MN_1 & MN_1^2 \\ \vdots & \vdots & \vdots \\ 1 & MN_n & MN_n^2 \end{bmatrix}$$

$$(2-10) \quad \hat{b} = ((MN)W(MN)^T)^{-1}(MN)W\sigma$$

And if the surface is based on strike prices (E):

$$(2-11) \quad E^T = \begin{bmatrix} 1 & E_1 & E_1^2 \\ \vdots & \vdots & \vdots \\ 1 & E_n & E_n^2 \end{bmatrix}$$

$$(2-12) \quad \hat{b} = (EWE^T)^{-1}EW\sigma$$

In the above equations, we have

$$(2-13) \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$(2-14) \quad \sigma = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix}$$

$$(2-15) \quad W = \text{diag}\{\omega_1, \dots, \omega_n\}$$

When we use the ordinary least-squares method to estimate the coefficients of equations (2-6) or (2-7), the elements of equation (2-15) become:

$$W = \text{diag}\{1, \dots, 1\}$$

And in case we use weighted least-squares approach we may use Liquidity-based or Gaussian methods. Depending on the availability of data, we can use Volume of trades or Open interest as proxies for liquidity of the options. If such data is not available we can give the weights based on how far an option is from being at the money. The calculation and implementation of this approach is further elaborated on in the section 3.

After we have created the volatility slices at each time to maturity, we utilize bi-cubic spline approach to interpolate between different slices:

$$(2-16) \quad p(x, y) = \sum_{j=0}^3 \sum_{i=0}^3 c_{i,j} t^i u^j$$

$$(2-17) \quad p_x(x, y) = \sum_{j=0}^3 \sum_{i=0}^3 i c_{i,j} t^{i-1} u^j$$

$$(2-18) \quad p_y(x, y) = \sum_{j=0}^3 \sum_{i=0}^3 j c_{i,j} t^i u^{j-1}$$

$$(2-19) \quad p_{xy}(x, y) = \sum_{j=0}^3 \sum_{i=0}^3 i j c_{i,j} t^{i-1} u^{j-1}$$

In these equations, $c_{i,j}$'s are constants and u and t are parameters ranging from 0 to 1. x and y represent strike/moneyness and time to maturity respectively.

3: Methodology

3.1 Data collection

As the process of data collection for this sort of research is lengthy and confusing at times, we have provided a detailed description of our efforts so that it would be a contribution to future work. We have used Bloomberg terminal to obtain option prices of crude oil futures traded on New York Mercantile Exchange (NYMEX). In the Bloomberg terminal, function “MFP<GO>” creates access to NYMEX’s Energy Futures list. This list provides real time data regarding the futures contracts traded on this market. To get more specific, we can use “CTM<GO>” to see all the categories of exchange traded contracts in Commodity markets. From this screen, when we choose “Crude Oil” a new screen appears which lists all the exchange traded contracts for crude oil. By setting the proper searching criteria, we will see the list of Crude oil (WTI)’s futures that are currently active on NYMEX and which are suitable for our work.

Table 3-1: Crude oil active future contracts on NYMEX

Ticker	Description	Exchange
CLA Comdty	Crude Oil, WTI	NYMEX Exchange
HBA Comdty	Crude Oil, European	NYMEX Clearport
G9A Comdty	Crude Oil, WTI Asian options	NYMEX Clearport

In Table (3-1), CLA is the active futures contract on light sweet crude oil, HBA is just set up to mirror the CLA contract, but for the sole purpose of pricing European style options and G9A is set up to mirror WTI financial futures to support the average price options.

The next step is to substitute the letter “A” in the above tickers with an appropriate letter in the following table to get the ticker for each month of the year.

Table 3-2: Structure of the tickers used (1st part)

	January	February	March	April	May	June	July	August	September	October	November	December
	F	G	H	J	K	M	N	Q	U	V	X	Z
CL	CLF	CLG	CLH	CLJ	CLK	CLM	CLN	CLQ	CLU	CLV	CLX	CLZ
HB	HBF	HBG	HBH	HBJ	HBK	HBM	HBN	HBQ	HBU	HBV	HBX	HBZ
G9	G9F	G9G	G9H	G9J	G9K	G9M	G9N	G9Q	G9U	G9V	G9X	G9Z

In the last step and to create the final ticker for a month of interest, we just have to add the year number based on the table below, followed by “Comdty” function.

Table 3-3: Structure of the tickers used (2nd part)

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Suffix	06	07	08	09	10	11	12	13	14*	5	6	7	8	9	0	1	2	3

Below are the tickers of the underlying futures contracts that we have used in this paper:

Table 3-4: Tickers of the underlyings of the 3 datasets

Month	Year	Underlying Futures Contract		
		Dataset 1	Dataset 2	Dataset 3
		CLA	HBA	G9A
November	2014	CLX4	HBX4	G9X4
December	2014	CLZ4	HBZ4	G9Z4
January	2015	CLF5	HBF5	G9F5
February	2015	CLG5	HBG5	G9G5
March	2015	CLH5	HBH5	G9H5
April	2015	CLJ5	HBJ5	G9J5
May	2015	CLK5	HBK5	G9K5
June	2015	CLM5	HBM5	G9M5

One thing to mention here is that the expiry date of the options on above futures is a few days earlier than the maturity of the underlying futures. This issue has been accounted for during the implementation process.

The most complex part of the data collection process was to see what options at which maturities and strikes are linked to the above contracts. Using Bloomberg function “OPT_CHAIN” in Excel, one can obtain the tickers of all of the options linked to an underlying security. But unfortunately, this function just works real time and we cannot observe what options were linked to an underlying security on a specific date in the past. For doing so we have to build the tickers manually.

If we are looking for call options, we add “C”, and for put options we add “P” to any of the tickers in Table (3-4) followed by the strike price and “Comdty”. As an example, the ticker for an option on a

* In 2014, for futures from January to May “14” is used, and for the ones afterwards, “4” is used.

June 2015 futures with a strike of 60\$ becomes CLM5C 60.00 Comdty for call option and CLM5P 60.00 Comdty for put option.

The final step in creating the tickers for the data under consideration is to have a range of strikes to make sure no option contract is ignored in the process. For this purpose, we considered a range from 10 \$ to 200\$, with increments of 0.5 \$. This way we will have a long list of options, but at the same time, we are making sure that everything is included. The advantage of following this procedure is that we somehow replicate the strike prices in the market. As an example, if the starting strike is 40 \$, the next one would be 40.5 \$, then 41 \$ and so on.

So based on our elaborations in the previous paragraphs, we have obtained the following three datasets of options on crude oil futures traded on New York Mercantile Exchange (NYMEX) as of 8th October 2014:

Table 3-5: Datasets

Dataset	Description	Maturity
No. 1	Prices of American options on crude oil futures	Oct 2014 – May 2015
No. 2	Prices of European options on crude oil futures	Oct 2014 – May 2015
No. 3	Crude oil average price (Asian) option prices	Oct 2014 – May 2015

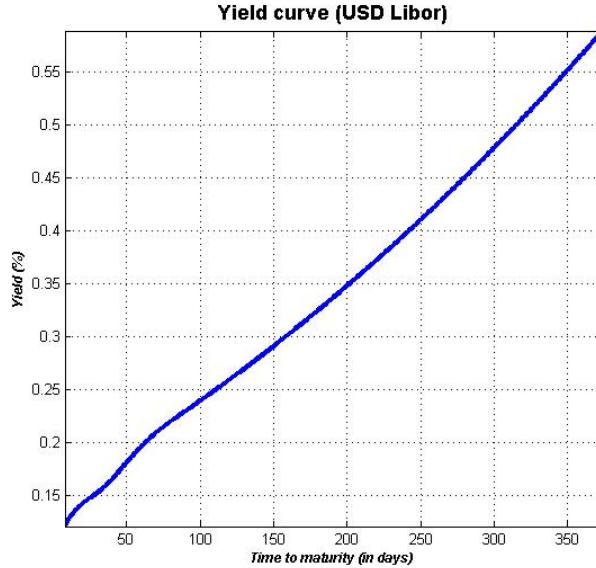
3.2 Risk free rate

Risk-free interest rates are required in order to calculate the implied volatilities of our observed options. We have used USD LIBOR rates as proxy for risk-free interest rates to create the yield curve. Cubic spline method is used to fit a smooth curve to the interest rates as of 8th October 2014.

Table 3-6: USD Libor rate as of 8th Oct 2014

Period	Rate
Intraday	0.09000
1 week	0.11730
1 month	0.15180
2 months	0.19600
3 months	0.22910
6 months	0.32415
1 year	0.56730

Figure 3-1: USD Libor yield curve as of 8th Oct 2014



3.3 Implied volatilities

Implied volatilities for options on futures are calculated by inverting the Black model (1976) and using numerical methods. The Black model is a variant of the famous Black-Scholes option pricing model but its main domains of application include pricing options on futures, bond options, interest rate caps and swaptions. In this paper we have used MATLAB's built-in function "blkimpv" to estimate the implied volatilities for the observed option prices. After implementation of what has been described so far, we are able to apply our models of interest to create the volatility surfaces.

3.4 Fitting the parametric model

As stated in the previous section, the parametric implied volatility surface can be created as a function of either strike or moneyness, and the time to maturity. For the purpose of this paper, the results are illustrated based on moneyness.

Implementation of the parametric model is much easier than other approaches since we do not require many data points, and additionally by just estimating a few parameters, we are able to create the whole surface. Another factor, which contributes to the simple application of these models, is the tools that some programming languages like MATLAB provide at our disposal. In MATLAB we can fit a parametric model by using the built-in function "fit" and the library model

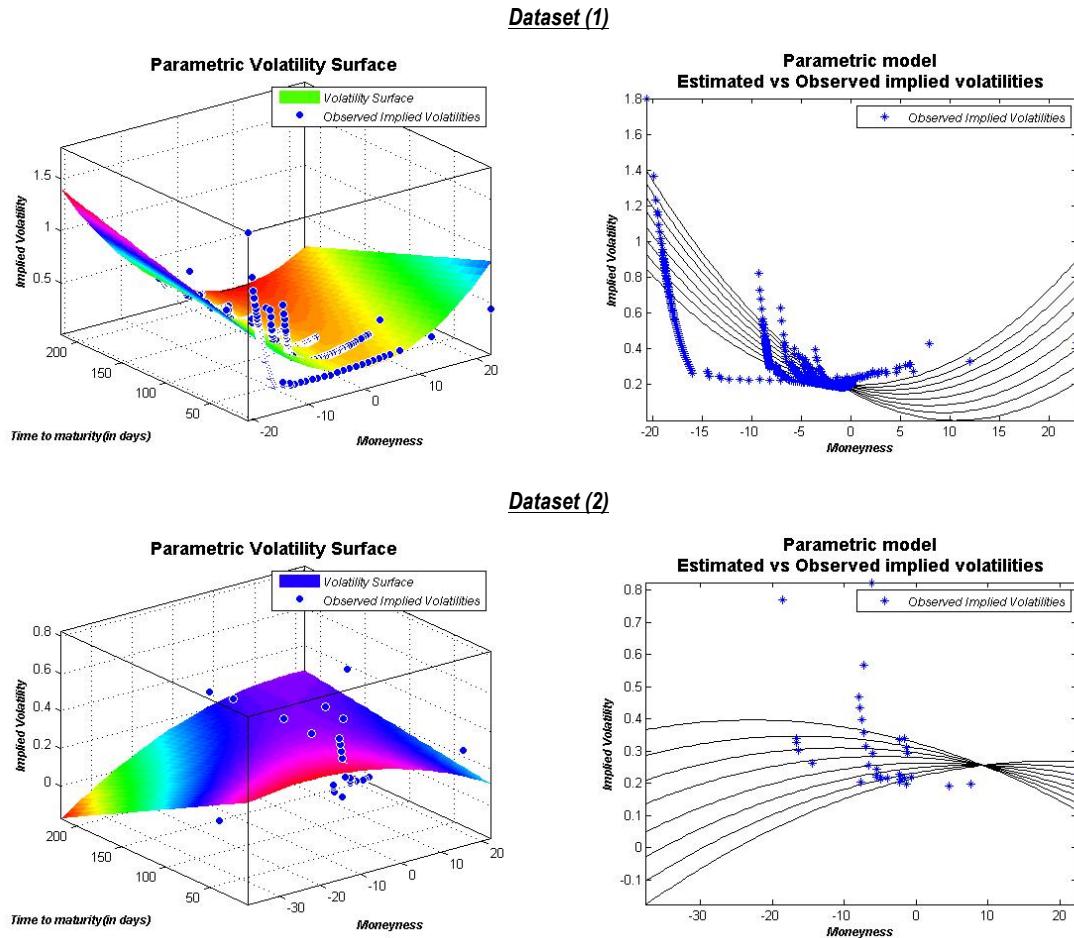
“poly21”. Table (3-7) lists the set of parameters estimated from fitting this model to each of our datasets:

Table 3-7: Parameters of the parametric model

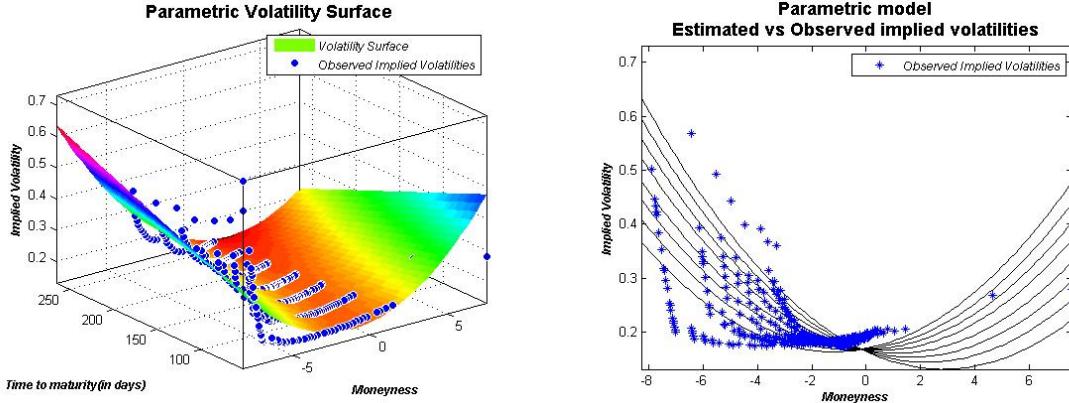
	Dataset 1	Dataset 2	Dataset 3
b1	0.200500	0.330200	0.170300
b2	-0.000747	-0.006709	0.017750
b3	0.001434	-0.000138	0.004082
b4	-0.000183	-0.000513	-0.000026
b5	-0.000135	0.000055	-0.000155

In Figure (3-2), we can see different volatility surfaces and volatility slices that we have fitted to our three datasets under the parametric model:

Figure 3-2: Comparison of volatility surfaces _Parametric model



Dataset (3)



As can be observed in Figure (3-2), the parametric model does not seem to fit the observed implied volatilities correctly. There are many outliers and as stated earlier, in this model the volatility slices maintain the same shape across all maturities. The quadratic function of moneyness/strike is assumed to be the same across all maturities and the linear relationship with time to maturity is rather simplistic. As was the case in Borovkova and Permana (2009), the parametric model still seems not to be a suitable model and it is incapable of generating accurate results during a highly volatile market.

3.5 Fitting the Non-parametric model

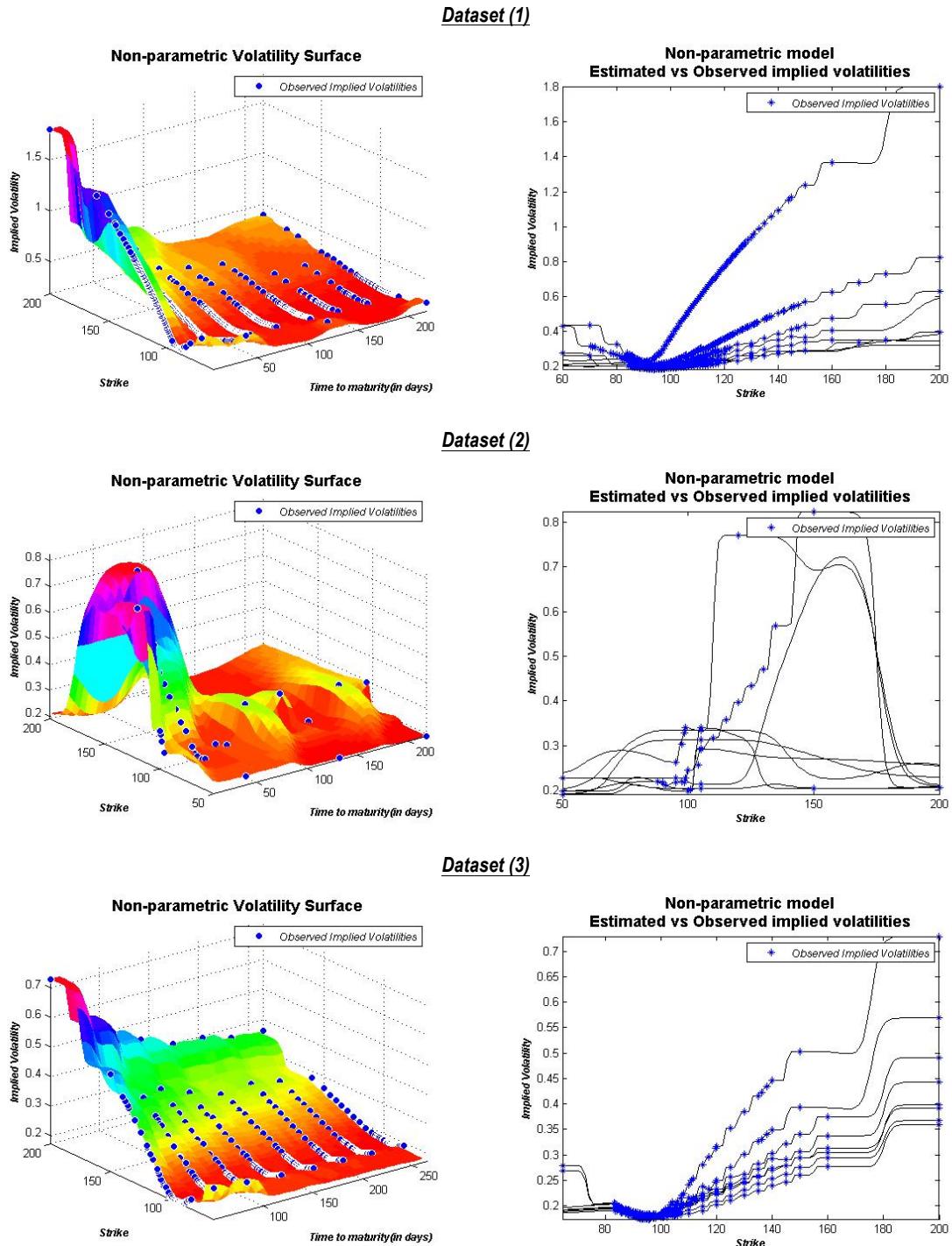
First, we have created a grid of points in two dimensions with the x-axis being the strikes and y-axis being the times to maturity. The strikes start from the minimum strike price of the dataset used and continues to the maximum strike. The steps are set to 0.5. This way we can somehow follow the pattern of the existing strike prices in the market.

The y-axis has the same number of steps as the x-axis, which can be altered. Then we have applied equation (2-4) to each point in the grid, and using the strike, time to maturity and implied volatility vectors of the observed option prices. This way, we have a three-dimension array in which each element is the implied volatility corresponding to the respective strike and time to maturity.

Just note that the non-parametric equation should be applied to data points that are not observed in the market. That is, if the strike and time to maturity of a grid point coincides with an observed point, instead of applying equation (2-4) to compute the respective implied volatility; we should just substitute the observed implied volatility. The reason is that the denominators in the equation become zero and consequently we would get errors.

The plots in Figure (3-3) illustrate how exactly the volatility surfaces under this method fit the observed implied volatilities, but at the expense of having very complex shapes.

Figure 3-3: Comparison of volatility surfaces_Non-parametric model



Our results from application of the non-parametric model in a highly volatile market are consistent with earlier findings in Borovkova and Permana (2009). The non-parametric approach lacks generalization power as it just attempts to fit the observed implied volatilities exactly. Consequently and as can be noted in Figure (3-3), we may have several minima and/or maxima in cross-sections of the surface along the maturity axis. We call each of these cross-sections a volatility slice. As stated in Borovkova and Permana (2009), the multimodality of the volatility slices translates into multimodality of option prices versus strike. Multimodal asset price distributions cannot occur under the risk-neutral probability measure. The multimodality of the volatility slices is inconsistent with risk aversion in the market, hence destroys the correct behaviour of the option's Vega versus strike.

3.6 Fitting the Semi-parametric model

In this paper the semi-parametric volatility surface is presented as a function of time to maturity and strike price. In this method, we initially need to categorize our data points based on their maturities. Then we can fit equation (2-6) to each of these categories in order to get the parameters of the volatility slices. This way we have fitted a quadratic function of different strikes at each maturity. As described in section 2, there are two methods to estimate the quadratic functions' parameters: ordinary least-squares and weighted least-squares.

Under the ordinary least-squares approach, all the elements of the equation (2-15) become "1". But if we take the weighted least-squares approach, we would have two methods to deal with the situation: Liquidity-based weights or Gaussian.

Liquidity-based weights are derived using either daily open interests or volume of trades. Under this method, for each category of options with the same maturity, each data point is weighted based on its liquidity factor relative to the whole group. Obviously, at-the-money options that are more liquid will get a higher weight than out or in-the-money contracts. As a result the quadratic function would become a better fit for at-the-money options.

In case we do not have access to the open interest as well as the volume data, we can utilize the Gaussian method. Under this method, each option is weighted based on how far it is from being at-the-money. To calculate these weights first we determine at-the-money strike (E^*) of the options in each category. Since we are studying the futures, at-the-money strike in each category equals to the futures price in that category. Then we compute the distance of each option's strike from at-the-money strike price ($d_i = E_i - E^*$). Finally, the weight of each option will equal to $\varphi_s(d_i)$, in which (s) is the empirical standard deviation of d_i 's. The sets of quadratic function parameters for our datasets are displayed in the following tables:

Table 3-8: Parameters of the semi-parametric model (Dataset 1)

CLA		Maturity							
		1	2	3	4	5	6	7	8
OLS	<i>b3</i>	-1.212766	0.452226	0.189898	0.477041	0.565004	0.427259	0.492701	0.331865
	<i>b2</i>	0.016929	-0.006367	-0.001641	-0.006630	-0.007839	-0.005582	-0.006716	-0.003172
	<i>b1</i>	-0.000005	0.000046	0.000020	0.000040	0.000042	0.000033	0.000038	0.000018
Gaussian	<i>b3</i>	2.375934	1.741131	1.307842	1.150852	1.000867	0.709714	0.692364	0.758903
	<i>b2</i>	-0.053397	-0.032404	-0.023204	-0.019799	-0.016591	-0.011102	-0.010628	-0.011777
	<i>b1</i>	0.000333	0.000173	0.000123	0.000103	0.000086	0.000060	0.000057	0.000061
Liquidity	<i>b3</i>	-1.783459	-0.077874	0.524895	0.613289	0.795013	0.706657	0.680142	0.493844
	<i>b2</i>	0.024190	0.001668	-0.008061	-0.009207	-0.012625	-0.010434	-0.010170	-0.006555
	<i>b1</i>	-0.000026	0.000016	0.000050	0.000052	0.000067	0.000053	0.000053	0.000035

Table 3-9: Parameters of the semi-parametric model (Dataset 2)

HBA		Maturity			
		1	2	3	4
OLS	<i>b3</i>	-0.824725	-0.483553	-0.019506	1.890681
	<i>b2</i>	0.004347	0.011581	0.005691	-0.051054
	<i>b1</i>	0.000075	-0.000036	-0.000028	0.000341
Gaussian	<i>b3</i>	-0.812376	0.737477	0.025356	1.890681
	<i>b2</i>	0.004116	-0.015125	0.004303	-0.051054
	<i>b1</i>	0.000076	0.000102	-0.000019	0.000341
Liquidity	<i>b3</i>	-0.619237	-0.078948	0.188888	1.890681
	<i>b2</i>	0.000555	0.001873	0.000214	-0.051054
	<i>b1</i>	0.000092	0.000020	-0.000001	0.000341

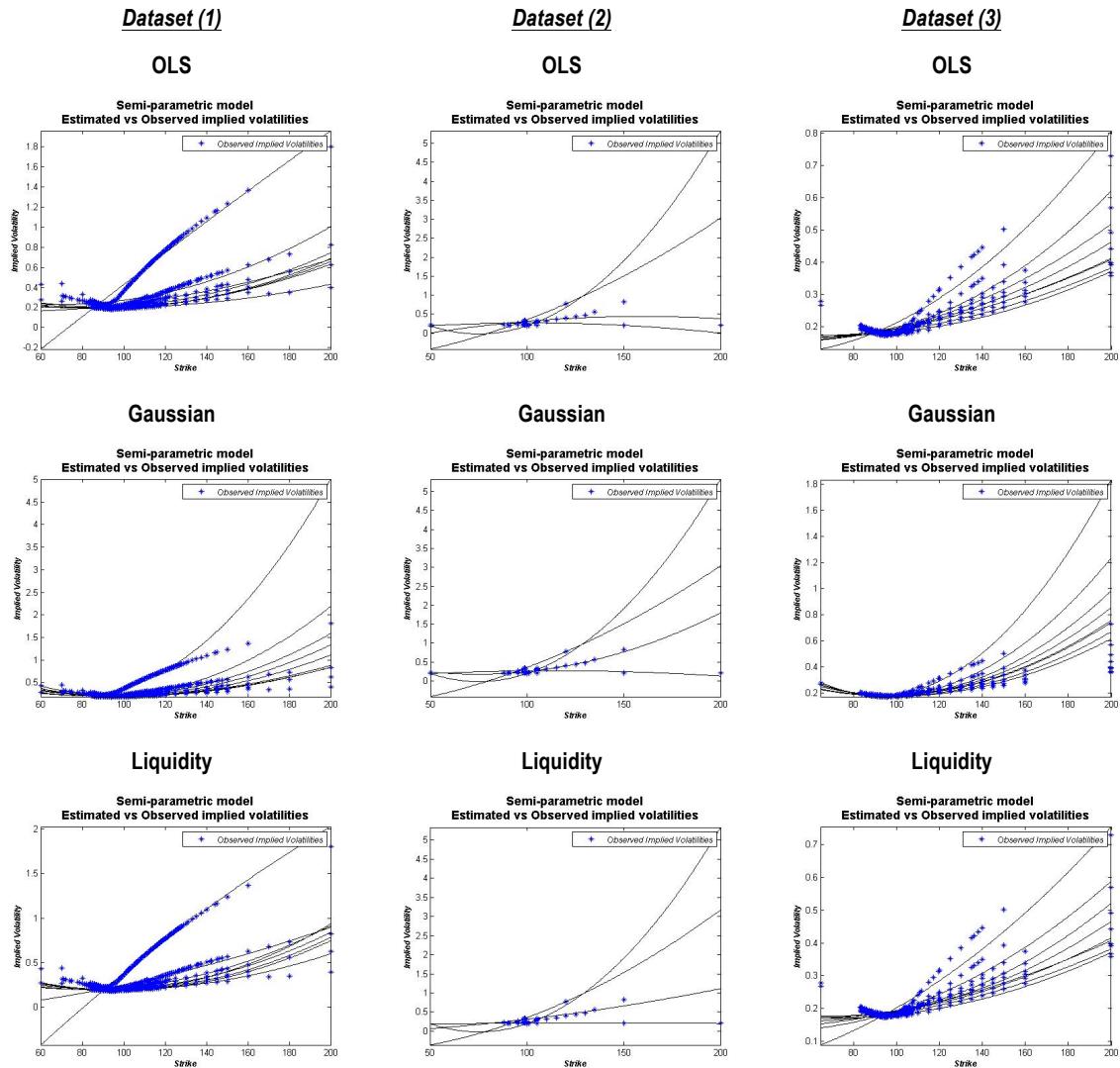
Table 3-10: Parameters of the semi-parametric model (Dataset 3)

G9A		Maturity							
		1	2	3	4	5	6	7	8
OLS	<i>b3</i>	0.154587	0.298086	0.202897	0.207500	0.165080	0.154112	0.193613	0.234800
	<i>b2</i>	-0.002104	-0.003596	-0.001720	-0.001512	-0.000675	-0.000494	-0.001059	-0.001716
	<i>b1</i>	0.000027	0.000026	0.000016	0.000014	0.000010	0.000009	0.000010	0.000012
Gaussian	<i>b3</i>	1.368725	1.041886	0.885555	0.773626	0.650082	0.642999	0.620838	0.597012
	<i>b2</i>	-0.025896	-0.018100	-0.014500	-0.012150	-0.009827	-0.009667	-0.009051	-0.008501
	<i>b1</i>	0.000141	0.000095	0.000075	0.000063	0.000052	0.000051	0.000046	0.000043

G9A		Maturity							
		1	2	3	4	5	6	7	8
Liquidity	b3	-0.005560	0.161870	0.269166	0.254614	0.207084	0.137569	0.180342	0.225908
	b2	0.000353	-0.001493	-0.002771	-0.002266	-0.001352	-0.000283	-0.000896	-0.001613
	b1	0.000017	0.000018	0.000020	0.000017	0.000012	0.000008	0.000010	0.000012

The semi-parametric implied volatility functions (volatility slices) are illustrated in Fig (3-4). Note how the implied volatility functions have distinctly different shapes at different maturities.

Figure 3-4: Comparison of volatility slices_Semi-parametric model



Based on the plots in Figure (3-4), we can see that for:

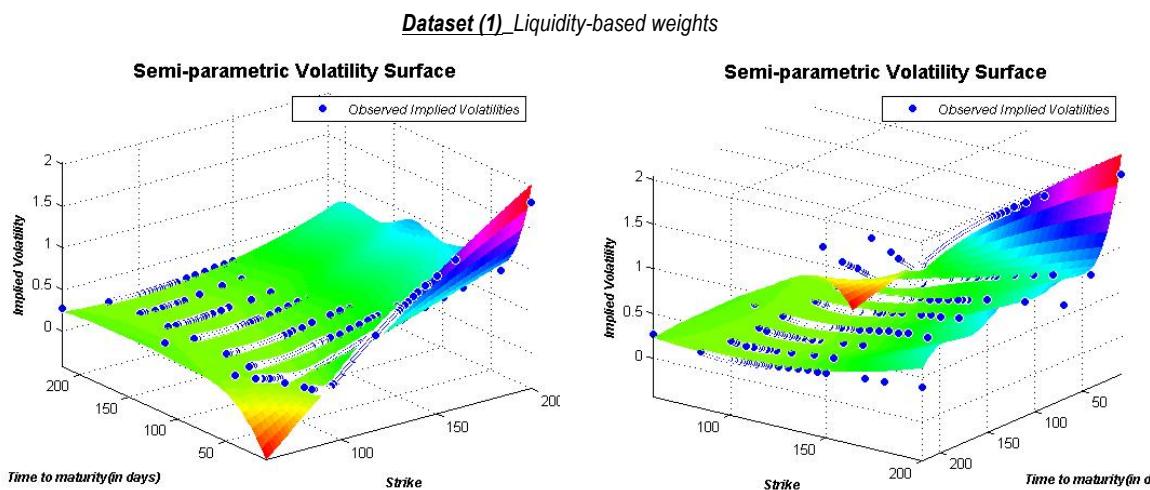
- Dataset (1), the liquidity-based and OLS approaches provide better fits than the Gaussian approach. Options in Dataset (1) are more liquid than the ones in other datasets.
- Dataset (2), the Gaussian approach is the best fit. Options in this category are rarely traded, thus their prices are less reliable.
- Dataset (3), again the Gaussian approach provides a better fit. But in this case, outliers are more noticeable than the other two datasets.
- The one thing that is common in all the cases above is that the semi-parametric model provides a better fit to at-the-money options. But as we move away from the at-the-money strike, we see that this model does not calibrate to the given implied volatilities well.

After we have fitted the quadratic functions to different classes of options and have built the volatility slices at different maturities, we interpolate between these slices using the bi-cubic spline approach to create the volatility surface.

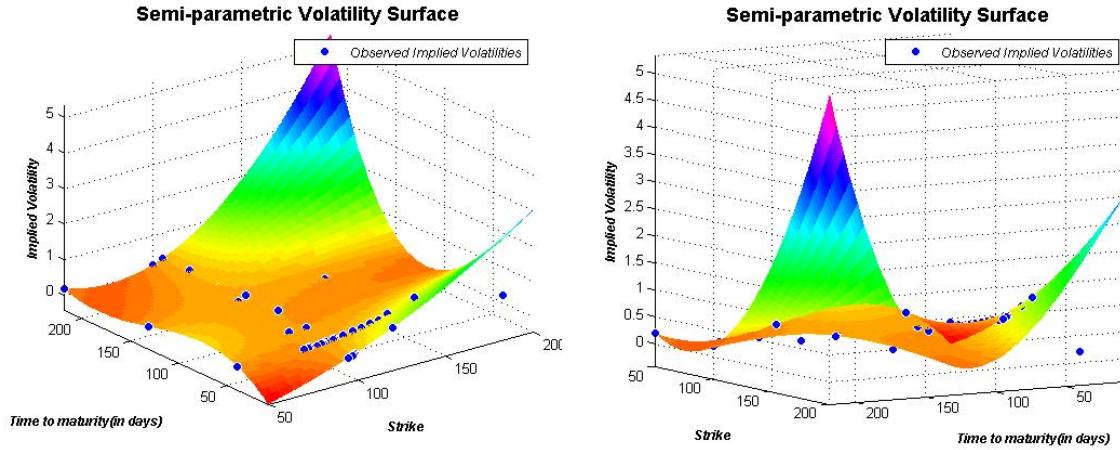
The most important part in fitting the semi-parametric model is when the volatility slices are calibrated at each maturity. It is in this stage that we can figure out whether our model is a good fit for our data or not. The stage of bi-cubic spline interpolation does not make any difference to the slices that are already created. This stage just connects the slices to build a smooth surface.

In Figure (3-5) we can observe different instances of application of the semi-parametric model to our three datasets. Note that the plots on the right hand side are the same as the ones on the left but at a different angle.

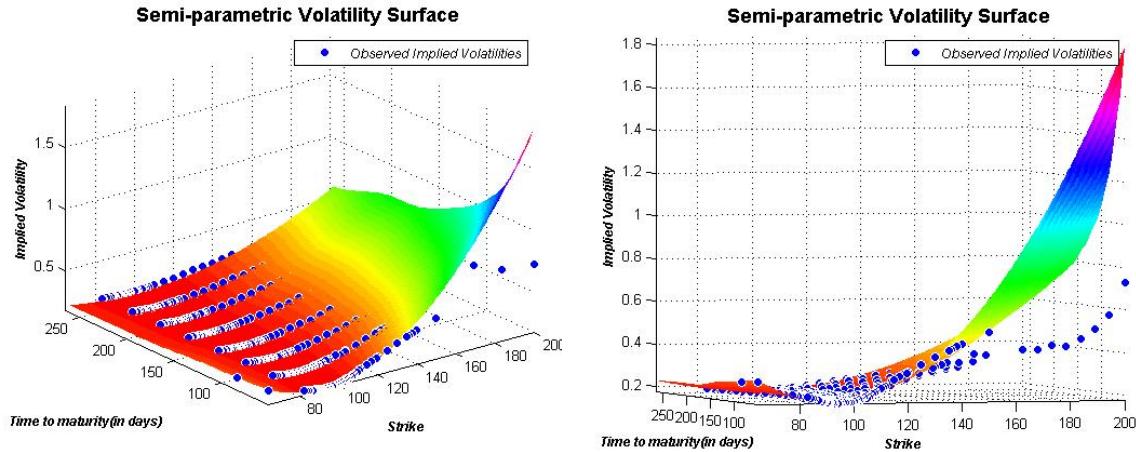
Figure 3-5: Comparison of volatility surfaces_Semi-parametric model



Dataset (2) Gaussian weights



Dataset (3) Gaussian weights



Our visual observation suggests that the semi-parametric model does not perform well in a highly volatile market. In reality we cannot force the market to behave in a specific manner, but the assumptions of this model restrict the volatility slices to certain shapes. One important point that needs to be mentioned here is that in our datasets we have excluded the options that exist in the market but not traded. We believe prices of these options, hence their implied volatilities are not reliable. If we were to include those options in our applications, we would have faced chaos in our volatility surfaces; too many outliers and unrealistic shape of the surfaces. The other drawback of this model is that the arbitrage-free conditions are not set. Even if the model provided a good fit, how could we determine whether the resulting surface was free of arbitrage or not?

3.7 Further analysis

In order to confirm our observations so far, we estimate the mean and standard deviation of our models' residuals and compare them with the metrics that were previously presented in

Borovkova and Permana (2009). Table (3-11) is taken from Borovkova and Permana (2009) and in Table (3-12), we have listed the results of our estimations.

Table 3-11: Mean and standard deviation of residuals, from Borovkova and Permana (2009)

Method	Dataset 1	Dataset 2	Dataset 3
Parametric	0.0096	0.0036	0.0090
	(0.0067)	(0.0034)	(0.0068)
Semi-parametric (<i>ordinary l.s.</i>)	0.0038	0.0016	0.0013
	(0.0061)	(0.0019)	(0.0037)
Semi-parametric (<i>Gaussian weighted l.s.</i>)	0.0044	0.0024	0.0011
	(0.0072)	(0.0070)	(0.0075)
Semi-parametric (<i>liquidity-based weighted l.s.</i>)	0.0048	0.0018	0.0011
	(0.0065)	(0.0021)	(0.0064)

In the table above, the Dataset (1) comprises European options on Crude oil futures and Dataset (2) contains the Average price options. These metrics are estimated based on the application of the three models that we have reviewed in this paper, but the timing of the datasets are different. As stated in Borovkova and Permana (2009), the figures in parentheses are the standard deviations of residuals and the other figures are the mean of residuals. For comparison purposes, the Dataset (1) in the above table corresponds to the Dataset (2) in our paper and the Dataset (2) corresponds to Dataset (3) in this project.

Table 3-12: Mean and standard deviation of residuals of three datasets

Method	Dataset (1)		Dataset (2)		Dataset (3)	
	Mean	STD	Mean	STD	Mean	STD
Parametric	1.65E-16	0.107001	-1.54E-18	0.136917	5.96E-17	0.041923
Semi-parametric (OLS)	3.24E-17	0.097058	-1.71E-15	0.121130	4.91E-17	0.036738
Semi-parametric (Gaussian)	0.017831	0.114362	-0.091162	0.515141	0.008098	0.045297
Semi-parametric (Liquidity-based)	-0.004638	0.244390	-0.067187	0.139537	-0.000330	0.041050

Comparison of Table (3-11) with Table (3-12) suggests that the standard deviations of the residuals have significantly increased when our data is taken from a highly volatile market. Furthermore both our visual observations and the metrics raise serious questions regarding applicability of such local volatility models to option data on crude oil futures.

Figure 3-6: Daily log-return of Crude oil futures

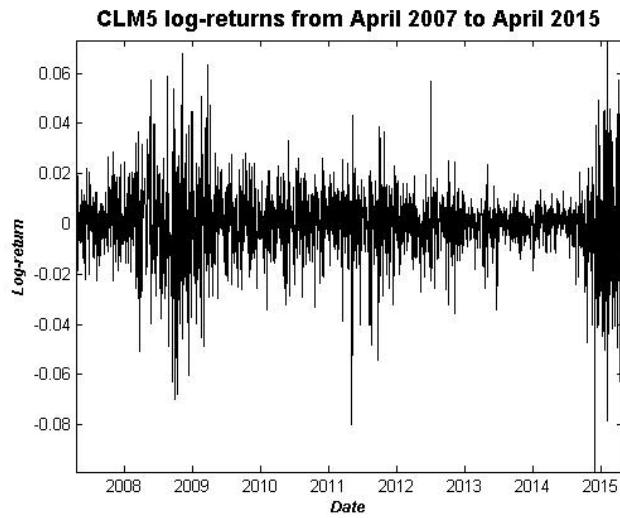
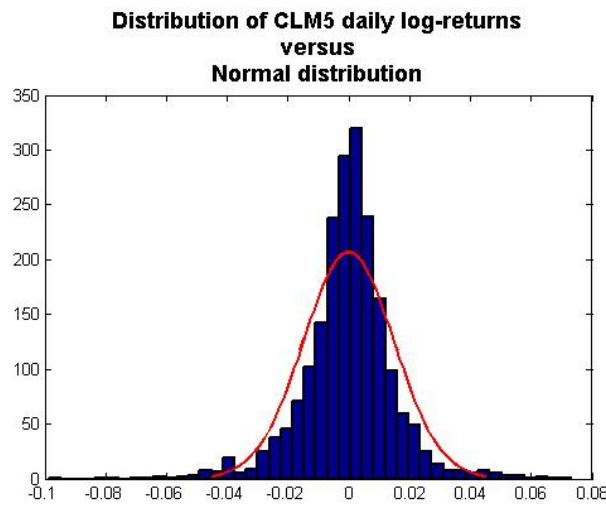


Figure (3-6) displays the daily log-returns of the June 2015 maturity crude oil futures since the first trade in April 2007. As we can see, large moves follow large moves and small moves follow small moves. In other words, we observe volatility clustering in this data.

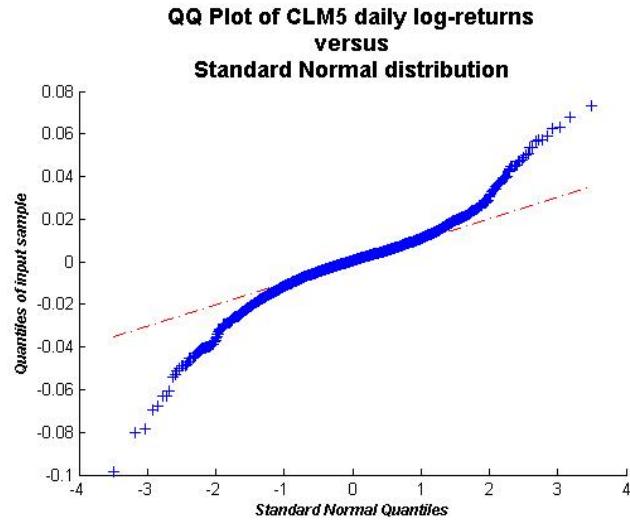
In Figure (3-7), we have plotted the distribution of the above data against normal density function. We observe that the empirical distribution has higher peak and fatter tails compared to the normal distribution.

Figure 3-7: Empirical distribution of futures' daily log-returns vs normal distribution



In Figure (3-8), the Q-Q plot compares the quantiles of the above dataset to the quantiles of the standard normal distribution. In this plot we can observe how the tails of the distribution of log-returns are extremely different from the standard normal distribution. The plot would have been a straight line if the data followed a normal distribution.

Figure 3-8: QQ Plot of futures' daily log-returns vs standard normal distribution



Fat tails and the high central peak are characteristics of mixtures of distributions with different variances. The volatility clustering feature implies that volatility (or variance) is auto-correlated (Gatheral, 2006). Our previous findings and the results presented here motivate us to consider volatility as a random variable. Consequently it would be valuable to consider whether stochastic volatility models such as the SVI parameterization of volatility surfaces that was first introduced at Merrill Lynch in 1999 would be a better fit for the options on crude oil futures?

4: Conclusion and future work

In this paper, we tested three local volatility models using three sets of option prices on crude oil futures during a high volatility period. We have demonstrated that such models are no longer capable of adequately fitting the observed data in the crude oil market. We have performed our tests using crude oil option data in 2014 and beginning of 2015. Based on our investigation, we have concluded that a parametric, a non-parametric and a semi-parametric method of implied volatilities do not give valid results when we are dealing with high volatility in the underlying futures.

Moreover, we have shown that the daily log-returns of crude oil futures display volatility clustering and their empirical distribution displays higher peak and fatter tails compared to the normal density function which leads to considering the volatility itself as a stochastic process. This fact raises the question of whether local volatility models in general are still appropriate for crude oil options or should we move towards using stochastic volatility models. We leave further investigation of this issue for future research.

Appendices

Appendix A: MATLAB Codes

Table 4-1: MATLAB_Yield curve using LIBOR rates

```
%% Term Structure:
CurrentDate='08-Oct-2014';
Data=[0.09000,0.11730,0.15180,0.19600,0.22910,0.32415,0.56730]/100;
Dates = daysadd(CurrentDate,...,
[0 7 (1/12)*360 (2/12)*360 (1/4)*360 (1/2)*360 360],2);
irdc=IRDataCurve('Zero',CurrentDate,Dates,Data,'InterpMethod','spline');

%-----
% Zero Rate Vector:
ZeroRates=irdc.getZeroRates(Exp_Date);
```

Table A-2: MATLAB_Estimation of implied volatilities

```
%% Implied Volatility:
IV=blkimpv(FutPrice, Strike, ZeroRates, Tenor_y,OptPrice,[],[],{'call'});
```

Table A-3: MATLAB_Calculation of moneyness

```
%% Moneyness:
Moneyness=log(FutPrice./Strike)./(IV.*Tenor_y);
```

Table A-4: MATLAB_Calculation of weights for the semi-parametric model

```
function W=Weightings(OpenInt,Vol,Tenor,Strike,FutPrice,method,InputType)
% This function calculates the weights to be used in the Semi-parametric
% model to estimate the coefficients of the quadratic functions.

% Inputs:
% OpenInt: Vector of Open interests of options on a specific date.
% Vol: Vector of trading volume of options on a specific date.
% Tenor: Vector of the time to maturities of the option data corresponding
% to the Strike vector.
% K: Vector of either strike prices or moneynesses.
% FutPrice: Vector of underlying futures prices.
% method: A string indicating the method to be used to estimate the
% quadratic functions' coefficients. It can be any of the followings:
% 'Liquidity_OpenInterest', 'Liquidity_Volume', 'OLS' and 'Gaussian'.
% InputType: A string, either 'Strike' or 'Moneyness'. It determines
```

```

% whether the elements of vector 'K' are Strike prices or Moneyness of
% options.
%
% Outputs:
% A cell containing square matrices. Each square matrix correspond to a
% specific maturity. These matrices are diagonal with their elements being
% the weights to be used in the Semi-parametric model.

%-----%
% Determine the unique maturities and number of them:
TTM=unique(Tenor);NofMat=length(TTM);

%-----%
% Preallocate the required cell:
W=cell(1,1,length(TTM));

%-----%
% Calculate the weights based on each method:
switch method
    case 'Liquidity_OpenInterest'
        for i=1:NofMat;
            W{i}=diag(OpenInt((Tenor==TTM(i)))/sum(OpenInt((Tenor==TTM(i)))));
        end

    case 'Liquidity_Volume'
        for i=1:NofMat;
            W{i}=diag(Vol((Tenor==TTM(i)))/sum(Vol((Tenor==TTM(i)))));
        end

    case 'OLS'
        for i=1:NofMat;
            W{i}=eye(sum(Tenor==TTM(i)));
        end

    case 'Gaussian'
        switch InputType
            case 'Strike'
                d=(Strike-FutPrice);
            case 'Moneyness'
                d=(Strike);
            end
        s=std(d);
        NormWeights=normpdf(d,0,s);

        for i=1:NofMat;
            W{i}=diag(NormWeights(Tenor==TTM(i)));
        end
    end
end
end

```

Table 4-5: MATLAB Parametric model

```

function [IVParam_MN_me,IVParam_MN_stde,IVParam_K_me, ...
    IVParam_K_stde]=IVParametric(Strike,Moneyness,IV,Tenor)
% This function fits a volatility surface to option data using the
% Parametric model.

% Inputs:
% Strike: Vector of strike prices of options.
% Moneyness: Vector of moneyness of each option.
% IV: Vector of implied volatilities corresponding to Strikes and Tenors.
% Tenor: Vector of the time to maturities of the option data corresponding
% to the Strike vector.
%
% Outputs:
% Two plots of the volatility surfaces fitted to the input data.
% Mean and Standard deviation of residuals.

%-----
% Fit the Parametric model based on either Moneyness or Strike prices:
[Parametric_MN,~,output_MN]=fit([Moneyness,Tenor],IV,'poly21');
[Parametric_K,~,output_K]=fit([Strike,Tenor],IV,'poly21');

%-----
% Plot the volatility surface and the observed implied volatilities based
% on Moneyness:
figure
plot(Parametric_MN, [Moneyness,Tenor], IV);
colormap(hsv);
shading flat
title('Parametric Volatility Surface','FontWeight','bold','FontSize',14);
xlabel('Moneyness','FontWeight','bold','FontSize',9,'FontAngle','italic');
ylabel('Time to maturity(in days)','FontWeight','bold','FontSize',9, ...
    'FontAngle','italic');
zlabel('Implied Volatility','FontWeight','bold','FontSize',9, ...
    'FontAngle','italic');
legend({'Volatility Surface';'Observed Implied Volatilities'},...
    'Location','Northeast','FontAngle','italic','FontSize',9);
axis tight

%-----
% Plot the volatility surface and the observed implied volatilities based
% on Strike prices:
figure
plot(Parametric_K, [Strike,Tenor], IV);
colormap(hsv);
shading flat
title('Parametric Volatility Surface','FontWeight','bold','FontSize',14);
xlabel('Strike','FontWeight','bold','FontSize',9,'FontAngle','italic');

```

```

ylabel('Time to maturity(in days)', 'FontWeight', 'bold', 'FontSize', 9, ...
    'FontAngle', 'italic');
zlabel('Implied Volatility', 'FontWeight', 'bold', 'FontSize', 9, ...
    'FontAngle', 'italic');
legend({'Volatility Surface'; 'Observed Implied Volatilities'}, ...
    'Location', 'Northeast', 'FontAngle', 'italic', 'FontSize', 9);
axis tight

%-----
% Calculate the mean and standard deviation of residuals:
IVParam_MN_me=mean(output_MN.residuals);
IVParam_MN_stde=std(output_MN.residuals);
IVParam_K_me=mean(output_K.residuals);
IVParam_K_stde=std(output_K.residuals);

end

```

Table 4-6: MATLAB_Non-parametric model

```

function IVNonParametric(Strike,TTM,IV,p,MaturityType)
% This function fits a volatility surface to option data
% using the Non-parametric model (nearest-neighbor weighted interpolation
% method.)

% Inputs:
% Strike: Vector of strike prices of option data
% Tenor: Vector of the time to maturities of the option data corresponding
% to the Strike vector.
% IV: Vector of implied volatilities corresponding to Strikes and Tenors.
% P: P is just a quantity. It determines the relative importance of distant
% samples. The approximation is better with higher values of "p".
% MaturityType: A string, either 'd' or 'y'. It determines whether the
% input Tenor vector is in years or in days.
%
% Output:
% A plot of the volatility surface fitted to the input data.

%-----
% Set the convention to be used(360 or 365 days):
year=365;

%-----
% This part determines the increments required for the calculations below:
switch MaturityType
    case 'd'
        increment=1;
    case 'y'

```

```

increment=1/year;
end

%-----
% Create the required grid for interpolation in the following sections:
Tenor=min(TTM):increment:max(TTM);
K=linspace(min(Strike),max(Strike),length(Tenor));
[X,Y] = meshgrid(Tenor,K);
ImpliedVol=nan(size(X));
[idx,jdx]=size(X);

%-----
% Interpolate and estimate the implied volatilities at each grid point:
for i=1:idx;
    for j=1:jdx;
        if sum((TTM==X(i,j)).*(Strike==Y(i,j)))==0;
            d=((TTM-X(i,j)).^2+(Strike-Y(i,j)).^2).^(p/2);
            ImpliedVol(i,j)=sum(IV./d)/sum(1./d);
        else
            ImpliedVol(i,j)=sum(IV.*((TTM==X(i,j)).*(Strike==Y(i,j))));
        end
    end
end

%-----
% Plot the volatility surface along with the observed implied volatilities:
figure
surf(X,Y,ImpliedVol);
colormap(hsv);
shading flat
hold on
h=plot3(TTM,Strike,IV,'o','MarkerEdgeColor','w','MarkerFaceColor','b',...
    'MarkerSize',7);
title('Non-parametric Volatility Surface','FontWeight','bold',...
    'FontSize',14);
xlabel('Time to maturity(in days)','FontWeight','bold','FontSize',9,...
    'FontAngle','italic');
ylabel('Strike','FontWeight','bold','FontSize',9,'FontAngle','italic');
zlabel('Implied Volatility','FontWeight','bold','FontSize',9,...
    'FontAngle','italic');
hleg=legend(h,'Observed Implied Volatilities');
set(hleg,'Location','Northeast','FontAngle','italic','FontSize',9);
axis tight
hold off

end

```

Table 4-7: MATLAB_Semi-parametric model

```

function [method,IVSmParam_K_me,IVSmParam_K_stde]=IVSemiParametric_Strike(...
    OpenInt,Vol,Tenor,Strike,FutPrice,IV,method,MaturityType)
% This function fits a volatility surface to option data using the
% Semi-parametric model. In this model, a quadratic function is fitted to
% implied volatilities at each maturity to create the volatility slices.
% Then these slices are connected to each other by bi-cubic spline
% interpolation.

% Inputs:
% OpenInt: Vector of Open interests of options on a specific date.
% Vol: Vector of trading volume of options on a specific date.
% Tenor: Vector of the time to maturities of the option data corresponding
% to the Strike vector.
% Strike: Vector of strike prices of options.
% FutPrice: Vector of underlying futures prices.
% IV: Vector of implied volatilities corresponding to Strikes and Tenors.
% method: A string indicating the method to be used to estimate the
% quadratic functions' coefficients. It can be any of the followings:
% 'Liquidity_OpenInterest', 'Liquidity_Volume', 'OLS' and 'Gaussian'.
% MaturityType: A string, either 'd' or 'y'. It determines whether the
% input Tenor vector is in years or in days.
%
% Outputs:
% A plot of the volatility surface fitted to the input data.
% Mean and Standard deviation of residuals.

%-----
% Determine the unique maturities and number of them:
TTM=unique(Tenor);NofMat=length(TTM);

%-----
% Preallocate the required cells and arrays:
E=cell(1,1,NofMat);Sig=cell(1,1,NofMat);InterpIV=cell(1,1,NofMat);
Residuals=nan(length(Tenor),1);b=nan(3,NofMat);

%-----
% Set the step sizes for the grid of data in the following sections:
switch MaturityType
    case 'd'
        increment=5;
    case 'y'
        increment=5/year;
end

%-----
% Weights:
W=Weightings(OpenInt,Vol,Tenor,Strike,FutPrice,method,'Strike');

```

```

%-----
% Estimate the coefficients of the quadratic functions (for each volatility
% slice):
for i=1:NofMat;
    SelectedOptions=(Tenor==TTM(i));
    CountOptions=sum(SelectedOptions);
    if CountOptions<3;
        continue
    end
    E{i}=nan(3,CountOptions);
    E{i}(1,:)=1;
    E{i}(2,:)=Strike(SelectedOptions);
    E{i}(3,:)=E{i}(2,:).^2;
    Sig{i}=nan(CountOptions,1);
    Sig{i}=IV(SelectedOptions);
    b(:,i)=E{i}*W{i}*E{i}'\E{i}*W{i}*Sig{i};
    % Residuals:
    InterpIV{i}=b(:,i) *[E{i}(1,:);E{i}(2,:);E{i}(3,:)];
    Residuals(SelectedOptions)=Sig{i}-InterpIV{i}';
end

%-----
% Remove the NaNs from residuals and estimate the mean and standard
% deviation of residuals:
Residuals=Residuals(~isnan(Residuals));
IVSmParam_K_me=mean(Residuals);
IVSmParam_K_stde=std(Residuals);

%-----
% Remove the NaNs from the results. NaNs appear when there are less than 3
% options at each maturity:
TTM=TTM(~isnan(b(1,:)));
b=b(:,~isnan(b(1,:)));

%-----
% Build the volatility slices with the coefficients estimated in the
% previous sections:
K=min(Strike):0.5:max(Strike);
X=[ones(1,length(K));K;K.^2];

SigSlice=b'*X;

%-----
% Plot the volatility slices and the observed implied volatilities:
figure('Name',method)

plot(K,SigSlice,'k','LineWidth',1.2);
hold on

```

```

h=plot(Strike,IV,'*');
xlabel('Strike','FontWeight','bold','FontSize',9,'FontAngle','italic');
ylabel('Implied Volatility','FontWeight','bold','FontSize',9,...  

    'FontAngle','italic');
title({'Semi-parametric model';...  

    'Estimated vs Observed implied volatilities'},'FontWeight',...
    'bold','FontSize',14);
hleg=legend(h,'Observed Implied Volatilities');
set(hleg,'Location','Northeast','FontAngle','italic','FontSize',9);
axis tight
hold off

%-----  

% Create the grid required in order to interpolate between volatility  

% slices:  

[Xa,Ya]=meshgrid(K,TTM);
[Xq,Yq]=meshgrid(K,min(TTM):increment:max(TTM));
Vfine=interp2(Xa,Ya,SigSlice,Xq,Yq,'spline');

%-----  

% Plot the volatility surface and the observed implied volatilities:  

figure('Name',method)

surf(Xq,Yq,Vfine);
colormap(hsv);
shading flat
hold on
h=plot3(Strike,Tenor,IV,'o','MarkerEdgeColor','w','MarkerFaceColor',...
    'b','MarkerSize',7);
title('Semi-parametric Volatility Surface','FontWeight','bold',...
    'FontSize',14);
xlabel('Strike','FontWeight','bold','FontSize',9,'FontAngle','italic');
ylabel('Time to maturity(in days)','FontWeight','bold','FontSize',9,...  

    'FontAngle','italic');
zlabel('Implied Volatility','FontWeight','bold','FontSize',9,...  

    'FontAngle','italic');
hleg=legend(h,'Observed Implied Volatilities');
set(hleg,'Location','Northeast','FontAngle','italic','FontSize',9);
axis tight
grid on
hold off

end

```

Table A-8: MATLAB_Semi-parametric model (2nd approach)

```
% Preallocate the required cell:
fitobject1=cell(1,1,NofMat);

%-----
% Fit the quadratic function at each maturity:
for i=1:NofMat;
    SelectedOptions=(Tenor_y==TTM(i));
    if sum(SelectedOptions)<3;
        TTM(i)=nan;
        continue
    end
    fitobject1{i}=fit(Strike(SelectedOptions),IV(SelectedOptions),...
        'poly2','Weights',diag(W{i}));
end

%-----
% Remove non-empty elements from the results of the above loop:
TTM=TTM(~isnan(TTM));
fitobject1=fitobject1(~cellfun(@isempty,fitobject1));

%-----
% Build the volatility slices:
K=min(Strike):0.5:max(Strike);
NofMat=length(TTM);
IVSlice=nan(NofMat,length(K));

for i=1:NofMat;
    IVSlice(i,:)=feval(fitobject1{i},K);
end

%-----
% Prepare the data points from volatility slices to create the volatility
% surface:
[K,TTM,IVSlice]=prepareSurfaceData(K,TTM,IVSlice);

% Fit the volatility surface to the slices using the bicubic spline
% interpolation method:
[fitobject2,gof2,output2]=fit([K,TTM],IVSlice,'cubicinterp');
```

Appendix B: Copyright License

ELSEVIER LICENSE TERMS AND CONDITIONS

This is a License Agreement between Amin Afshinmanesh ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Amin Afshinmanesh
Customer address	****
License number	3617241084448
License date	Apr 27, 2015
Licensed content publisher	Elsevier
Licensed content publication	Computational Statistics & Data Analysis
Licensed content title	Implied volatility in oil markets
Licensed content author	Svetlana Borovkova, Ferry J. Permana
Licensed content date	15 April 2009
Licensed content volume number	53
Licensed content issue number	6
Number of pages	18
Start Page	2022
End Page	2039
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Table 4
Title of your thesis/dissertation	SEMI-PARAMETRIC VOLATILITY SURFACE ESTIMATION: CRUDE OIL FUTURES OPTIONS
Expected completion date	Apr 2015
Estimated size (number of pages)	50
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 CAD
VAT/Local Sales Tax	0.00 CAD / 0.00 GBP
Total	0.00 CAD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. Translation: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word

preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. Posting licensed content on any Website: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at <http://www.elsevier.com>. All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. For journal authors: the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- - immediately
 - via their non-commercial person homepage or blog
 - by updating a preprint in arXiv or RePEc with the accepted manuscript
 - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
 - directly by providing copies to their students or to research collaborators for their personal use
 - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- - after the embargo period
 - via non-commercial hosting platforms such as their institutional repository
 - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- - link to the formal publication via its DOI
- - bear a CC-BY-NC-ND license - this is easy to do
- - if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (PJA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJsAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by/4.0>.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-sa/4.0>.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-nd/4.0>. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- - Associating advertising with the full text of the Article
- - Charging fees for document delivery or access
- - Article aggregation
- - Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

References

- Andersen, L. & Andreasen, J. 2000, "Jump-diffusion processes: volatility smile fitting and numerical methods for option pricing", *Review of derivatives research*, vol. 4, no. 3, pp. 231-262.
- Black, F. & Scholes, M. 1973, "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, vol. 81, no. 3, pp. 637-654.
- Black, F. 1976, "The pricing of commodity contracts", *Journal of Financial Economics*, vol. 3, no. 1, pp. 167-179.
- Borovkova, S. & Permana, F.J. 2009, "Implied volatility in oil markets", *Computational Statistics and Data Analysis*, vol. 53, no. 6, pp. 2022-2039.
- Boudon, A. (2007). The Behaviour of Implied Volatility Surface: Evidence from Crude Oil Futures Options. Available at SSRN 930726.
- Bourke, P., 1998. Nearest neighbour weighted interpolation.
<http://astronomy.swin.edu.au/bourke/modeling/weightinterp>.
- Derman, Emanuel, and Iraj Kani, 1994a, The volatility smile and its implied tree, *Quantitative Strategies Research Notes* ~Goldman Sachs, New York.
- Derman, Emanuel, and Iraj Kani, 1994b, Riding on the smile, *Risk* 7, 32–39.
- Dumas, B., Fleming, J. & Whaley, R.E. 1998, "Implied volatility functions: empirical tests", *The Journal of Finance*, vol. 53, no. 6, pp. 2059-2106.
- Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1), 18-20.
- Gatheral, J. & Jacquier, A. 2014, "Arbitrage-free SVI volatility surfaces", *QUANTITATIVE FINANCE*, vol. 14, no. 1, pp. 59-71.
- Gatheral, J., 2006. *The Volatility Surface: A Practitioner's Guide*. Wiley Finance.

- Heath, D.C. 1992, "Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation", *Econometrica*, vol. 60, no. 1, pp. 77-105.
- Hull, J. 2012, *Options, futures, and other derivatives*, Prentice Hall, Boston.
- Merton, R.C. 1976, "Option pricing when underlying stock returns are discontinuous", *Journal of Financial Economics*, vol. 3, no. 1, pp. 125-144.
- Murphy, F. & Ronn, E.I. 2014, "The valuation and information content of options on crude-oil futures contracts", *Review of Derivatives Research*.
- Rubinstein, M. 1994, "Implied binomial trees", *The Journal of Finance*, vol. 49, no. 3, pp. 771-818.
- Wang, G.J., Yourougou, P. & Wang, Y.D. 2012; 2009, "Which implied volatility provides the best measure of future volatility?", *Journal of economics and finance*, vol. 36, no. 1, pp. 93-105.
- Wilmott, P. 2007, *Paul Wilmott introduces quantitative finance*, John Wiley & Sons Ltd, Hoboken, NJ; Chichester, England.
- Xiang, Y. & Zhang, J.E. 2008, "The implied volatility smirk", *Quantitative Finance*, vol. 8, no. 3, pp. 263-284.