



LUND UNIVERSITY

Faculty of Economics

M.Sc in Finance

Master's Thesis

May 29, 2019

Smile! It increases your face value

An empirical analysis of the implied volatility surface on the Swedish stock market index

OMXS30

Authors:

Dennis Ellegren 920413-3475

Anton Evilevitch 970805-5810

Supervisor:

Prof. Frederik Lundtofte, Ph.D

Abstract

This thesis examines some of the multiple variations of the previously established Rules of Thumb; which are used in attempting to explain implied volatility surfaces. Here, these Rules are extensively tested on the Swedish stock market index (OMXS30) using rolling window analysis and linear stepwise regressions with forward selection. Data is collected on options at four different maturities and nine different levels of strike prices for ten years (2512 trading days), resulting in a linearly interpolated volatility surface consisting of 36 points for each trading day.

The results support previous research on other stock indices in that the Square Root of Time Rule is the most accurate according to both in-sample information criteria and out-of-sample measures of fit. According to the results, the most accurate regression specification for one-day-ahead prediction is the Square Root of Time Rule with, in contrast to previous research, an observation window of two days and including explanatory variables up to twelve powers.

Contents

1	Introduction	1
1.1	Previous Research	2
2	Theory	3
2.1	Options	3
2.2	Moneyness	4
2.3	Model Selection Criteria	4
2.3.1	Root Mean Square Error, RMSE	4
2.3.2	Akaike Information Criterion (AIC)	5
2.3.3	Bayesian Information Criterion (BIC)	5
2.4	Volatility	6
2.4.1	Implied Volatility	6
2.4.2	Volatility Surface	7
2.5	Risk Neutral Valuation	7
2.6	No-arbitrage condition	8
2.6.1	Put-Call Parity	8
2.7	The Black-Scholes Formula	9
2.7.1	Assumptions	10
2.7.2	Black-Scholes Pricing Formulas for European Options	11
2.8	The Greeks	11
2.8.1	Delta	12
2.9	Rules of Thumb	13
2.9.1	The Sticky Strike Rule	14
2.9.2	The Sticky Delta Rule	14
2.9.3	The Square Root of Time Rule	15
3	Data and Methodology	17
3.1	Data	17
3.1.1	Data Processing	17
3.2	Stepwise Regression	17
3.2.1	Data-Mining Bias	18
3.2.2	Period Partitioning	18

3.2.3	Forward Selection	19
3.2.4	Walk Forward Analysis	19
3.3	Model Selection	19
3.3.1	All Rules of Thumb Tests	20
3.3.2	Further Analysis: Square Root of Time Rule	22
3.3.3	Volatility Climate	22
4	Results	23
4.1	All Rules of Thumb Tests	23
4.2	Further Analysis: Square Root of Time Rule	25
4.3	Volatility Climate	25
5	Analysis and Interpretation	26
5.1	All Rules of Thumb	26
5.2	Further Analysis: Square Root of Time Rule	27
5.3	Volatility Climate	29
5.4	Model Limitations	29
6	Conclusion	31
6.1	Further Research	32
	Appendices	35

1 Introduction

The volatility of options is an important measure in the financial industry. As true volatilities of assets cannot be directly observed the Black-Scholes formula (see. Section 2.7) is used to estimate the implied volatilities of various assets. These are then often used in place of the true volatility. Some of the assumptions of the original model (see. Section 2.7.1) are criticized for not holding in practice, e.g., the assumption of constant volatility (see. Section 2.4).

What is instead empirically observed is a complex relationship, known as the 'Volatility Smile', because when plotted the shape is reminiscent of a smile (u-shape) or smirk (reverse skew) according to Bates (1997). With the extension of the options' time-to-maturity, it forms a three-dimensional surface called the volatility surface (see. Section 2.4.2). The figures below show the smirk (reverse skew for stock indices) and implied volatility surface for a sample date on the OMXS30:

[INSERT FIGURE 1 ABOUT HERE]

[INSERT FIGURE 2 ABOUT HERE]

There are certain "Rules of Thumb" (see. Section 2.9) that are used in the industry to try and explain this surface. The authors test some of these Rules of Thumb on the Swedish stock index OMX Stockholm 30 (OMXS30) in multiple different periods for European style options. The results are compared using both in-sample and out-of-sample measures of fit.

In addition to previous research, this research will include higher power tests on the Rules of Thumb, a Stepwise, walk-forward analysis on daily data, and an optimal observation window size analysis. Previous research has mainly focused on weekly or monthly data with a seemingly arbitrary observation window size. The approach proposed in this paper is intended to be better suited for one-day ahead predictions. The purpose of this thesis is thus to try and see if analysis using a more extensive methodology supports previous research in the choice of Rule of Thumb and if increased predictive power can be achieved with a Rule of Thumb and regression specification found with the proposed approach.

1.1 Previous Research

Previous research has established three main Rules of Thumb (see. Section 2.9) used in the industry. In a paper called 'Volatility Surfaces: Theory, Rules of Thumb, and Empirical Evidence' by Hull et al. (2006), the authors examine the evolution of a volatility surface using the no-arbitrage condition. They further empirically test some Rules of Thumb used by traders and some extensions to these. Their tests are performed using monthly data on the stock index, S&P 500. The results show that the Sticky Strike model in its basic form, only dependent on the strike price, K , and the time-to-maturity, T , is consistent with the no-arbitrage condition only if the implied volatility surface is observed to be flat, per the original Black-Scholes model assumptions. They further show that skewness in the volatility surface cannot be consistent with either the Sticky Strike Rule or the basic Sticky Delta Rule, but that the Relative Sticky Delta Rule and the Square Root of Time Rule can be approximately consistent with the no-arbitrage condition.

The empirical tests of this previous research left some questions open in regards to the chosen methodology. While their purpose was not to make the best one-day ahead prediction, this is often what is most applicable in the industry. While the choice of observation window size can potentially impact the results, the researchers disclosed no analysis of how it is chosen, and they applied the same observation window for all tested models. This analysis is also made on the S&P 500, which can be considered the prime mover of stock market indices and so the results may not be directly transferable to a smaller stock market index. Finally, previous research limited the size of the models based on monthly data, whereas daily data might yield a better fit using an alternative regression specification.

2 Theory

2.1 Options

Options are a category of financial instruments known as derivatives. These derivatives are contracts based on some underlying asset, commonly, stocks or stock indices such as the OMXS30. A call option gives the buyer the right (but not the obligation) to purchase the underlying asset at an agreed upon price, the strike price, at a predetermined amount of time into the future (time-to-maturity). Similarly, a put option gives the buyer the right (but not the obligation) to instead sell the underlying asset at an agreed upon price, the strike price, at a predetermined date in the future. The payoff of an options contract is, therefore, a contingent claim as it depends on the price outcome of the underlying asset.

The seller of both types of options contracts is mandated to act on the wish of the buyer of the contract. The seller is thus said to be "short" the options contract, and the buyer is said to be "long" the contract. The buyer of an options contract pays a premium for this privilege; this is true whether or not the buyer exercises the option. The buyer of the contract is assumed to only exercise the contract when there is an economic benefit to doing so; when the strike price in relation the current market spot price is of positive value to the holder of the option. Whether an option is in-, out-, or at-the-money is described as the options "moneyness" (see. Section 2.2).

Options are in general considered to be a relatively complex financial instrument. Despite this, when valuing an option, the same principles hold as to value any claim to an uncertain income stream. One must estimate the expected amount and timing of the claim then discount it to the present value at a risk-adjusted discount rate (Constantinides et al. 2003, p. 1148). On a final note, this thesis only considers European style options which, unlike American style options, can be exercised only at the maturity date.

2.2 Moneyness

Moneyness describes the intrinsic value of an option and is in this thesis defined as:

$$MN = K/S \tag{1}$$

Where, MN = moneyness, K = strike price, S = Current price of the underlying asset, which means that when:

- $MN = 1 \iff$ The option is at-the-money for both a call and a put
- $MN < 1 \iff$ Call option: in-the-money; put option: out-of-the-money
- $MN > 1 \iff$ Call option: out-of-the-money; put option: in-the-money

In other words, the holder of the contract will choose to exercise either option when it is in-the-money. If the option is out-of-the-money, the investor gains economic benefit from going directly to the market at current prices instead. If the option is at-the-money, the investor is theoretically indifferent between exercising or not. The common definition of an option involves the transfer of the underlying asset, while in reality, cash-settlement is more prevalent. This means that the side of the contract that is out-of-the-money, if short the contract, will pay the difference between the agreed-upon strike price and the current spot price to the side that is in-the-money. Hence the receiver of the funds can go to the market and purchase the underlying asset with the proceeds from the options contract making up the difference.

An option is also commonly referred to as being "near-the-money" this implies quite intuitively that the option is near the at-the-money point. In practice, the terms at-the-money and near-the-money are often used interchangeably as an option rarely is precisely at-the-money.

2.3 Model Selection Criteria

2.3.1 Root Mean Square Error, RMSE

Root Mean Square Error (RMSE) is a measure of fit. For every observed data point, the vertical distance to the model estimated value (the error term) is measured, squared and divided by the number of data points; this results in the Mean Squared Error, MSE. To arrive at RMSE, the square root is taken from the MSE.

RMSE has a simple interpretation as it is the average distance between the model fitted and observed values, in the same units as that of the vertical axis. A lower RMSE is hence indicative of a better fit. It is important to note that this is a relative measure of fit so, if viewed in isolation, it does not say how well the model fits the data. It is, however, useful when comparing models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (2)$$

where, P_i = model predicted value at observation (or date) i , O_i = actual observed value at observation i , and n = number of observations. An advantage of the RMSE is that it can easily be applied to an out-of-sample prediction when comparing to observed values, this is the form of RMSE that will be focused on in this research and it will be referred to as $RMSE_{T+1}$. Lastly, a drawback of RMSE as an in-sample measure of fit is that it does not include an over-parametrization penalty and will thus tend to reward additional explanatory variables. This motivates the use of additional measures of fit for comparison of models with more parameters.

2.3.2 Akaike Information Criterion (AIC)

Similar to RMSE, Akaike information criterion gives a measure of a statistical model's relative quality for a given data set. The criterion, founded in information theory, describes the model's likelihood to predict values. For each model, AIC balances good fit with parsimony by taking into account the amount of information lost from the reduction of degrees of freedom and factoring this into the overall score of the model. Hence, AIC is a useful tool in selecting the most accurate model, especially when considering additional parameters.

$$AIC = 2k - 2\ln(\widehat{L}) \quad (3)$$

where, k is the number of estimated parameters in the model and \widehat{L} is the maximum value of the likelihood function for the model.

2.3.3 Bayesian Information Criterion (BIC)

The Bayesian Information Criterion is closely related to the AIC as it is also a parameter-penalizing likelihood function criterion. The main difference is that in practice, BIC

penalizes larger models, i.e., model complexity, more heavily than AIC. Thus the only way in which the choice of model between the two criteria can be different is if AIC prefers a larger model. Therefore, it is useful to use the two criteria in conjunction to see if they chose differently sized models.

$$BIC = \ln(n)k - 2\ln(\hat{L}) \quad (4)$$

where k is the number of parameters estimated by, and \hat{L} is the maximized value of the likelihood function of, the model. n is the number of observations.

2.4 Volatility

Volatility is a statistical measure of dispersion; in the case of financial assets, it is the dispersion of returns. The higher the volatility, the riskier the asset is considered as it is associated with larger swings in either direction. Volatility is not directly observable but is estimated for the large market indices as a measure of the market's expectation of volatility. An example of this is VSTOXX, which is an index estimating implied volatility for the EUROSTOXX index. The volatility is of particular interest to options traders as they are not only interested in the direction of the market of the underlying asset, but also the speed of the market. The volatility can be considered a measure of this speed (Natenberg 1994, p. 51). A more volatile market is thus more likely to move through the strike price and benefit the holder of an options contract. If increasing volatility also increases the values of options contracts, it is of high interest to be able to estimate whether a certain market will have relatively high or low volatility in the next period, and take options positions accordingly.

2.4.1 Implied Volatility

The volatility of an asset is the variable of the Black-Scholes formula that cannot be directly observed; what is commonly used in its place is the implied volatilities of the asset (Hull 2015). The implied volatility can be viewed as consensus volatility among the market participants for a given underlying during the remaining life of an option. The historical volatility is known to be at mean reverting at various degrees (Tsay 2010, p. 80) and the implied volatility has been shown to usually have smaller fluctuations around its mean (Natenberg 1994, p. 290). From these empirical findings, it becomes

evident that in the long run, the primary factor that affects the implied volatility is the actual volatility of the underlying as the options traders will adjust their expectations. In the short run, however, there can be significant and sometimes even dominant effects other than the movements of the historical volatility that affects the implied volatility, such as government reports, news and other surprises which can temporarily make the implied volatility higher than the historical.

As expanded on in later sections, the implied volatilities are calculated through the use of the Black-Scholes formula (Constantinides et al. 2003, p. 705) as the volatilities that are implied by options on the asset that is available in the market.

2.4.2 Volatility Surface

The volatility surface is a visual representation of the implied volatility, for a particular asset, and it is commonly shown as a three-dimensional graph of the calculated implied volatilities of options available in the marketplace, as a function of time-to-maturity and moneyness. At any given day a limited number of maturities are observed which means that some points on the surface will be estimated directly while the rest of the surface most commonly is filled out by interpolating between these directly estimated points (Hull et al. 2006).

If the original Black-Scholes assumptions hold (see. Section 2.7.1), we expect to observe a flat volatility surface; this is however not what is generally observed in the empirical analysis of most assets (Hull et al. 2006). The so-called "smile" part of the surface refers to the relationship between the implied volatility and the strike price which for various assets shows an increasing implied volatility as the spot price moves from the at-the-money point in either direction. For stock indices a so-called "reverse skew" is commonly observed which is a situation where the implied volatility increases with decreasing strike prices, Figure 1 shows the reverse skew/smirk of the surface in Figure 2 for OMXS30 on a sample date.

2.5 Risk Neutral Valuation

Risk-neutral valuation states that an options contract can be valued in terms of its expected payoffs, after those have been discounted from the expiration date to the present, under the assumption that they grow at the risk-free rate. This means that the real rate

that the underlying asset grows at does, on average, not affect the value. This is evident since it is possible to completely hedge an option with the underlying asset, thereby removing the exposure to the direction of the aforementioned underlying asset.

Thus the asset and the options have a risk-neutral probability measure in that if the relationship does not hold, there exists an arbitrage opportunity, which is explained further in the next section. This method can then be used to obtain a no-arbitrage price of derivatives.

2.6 No-arbitrage condition

The no-arbitrage condition is one of two central concepts in asset-pricing models, the other being financial market equilibrium (Constantinides et al. 2003, pp. 746). An arbitrage opportunity is a situation where assets can be combined into a portfolio, with zero risk for losses, zero cost, and a positive probability of a non-negative return.

The intuition behind the concept can be seen in the empirical observation that when potential arbitrage opportunities arise, market participants tend to act as to align the prices so that the opportunities are instantly eliminated. This is at least the case if one assumes that the markets are to some degree efficient.

For options contracts, the implication of the no-arbitrage assumption leads to a concept called Put-Call Parity.

2.6.1 Put-Call Parity

The Put-Call Parity is a principle that is derived from creating a risk-free portfolio using a trade in a call, a put, and the underlying asset at the same time (Constantinides et al. 2003, p. 1146).

The parity is given for European style option by:

$$p + S_0e^{-iT} = c + Ke^{-rT} \quad (5)$$

where,

- c is a call option on the underlying asset
- p is a put option on the underlying asset
- S_0 is the spot/current market price of the underlying asset at time zero
- K is the exercise/strike price of both the call and put contract
- r is the interest cost (commonly assumed at risk-free rate)
- i is the non-interest cost/benefit of holding the underlying

This assumption can be used to explain why we should be able to use implied volatilities as:

$$p_{BS} + S_0e^{-iT} = c_{BS} + Ke^{-rT} \quad (6)$$

$$p_{mkt} + S_0e^{-iT} = c_{mkt} + Ke^{-rT} \quad (7)$$

where, BS = Black-Scholes and mkt = Market. If these two equations are subtracted:

$$p_{BS} - p_{mkt} = c_{BS} - c_{mkt} \quad (8)$$

Thus the Put-Call parity indicates that any deviation between the Black-Scholes prices and the market prices should be equivalent for puts and calls.

2.7 The Black-Scholes Formula

The Black-Scholes builds on the assumption of no-arbitrage, and the model allows for an analytical calculation of the rational price of a European option and can be used to create a risk-free portfolio. In financial markets adhering to the assumptions outlined below, synthetic options can be created through the utilization of bonds and stocks; effectively hedging against all risk (i.e., a replicating portfolio). The intuition behind the model is that if the options available in the market are priced correctly, it should not be possible to make certain profits (arbitrage) by using long and short positions in options and their

underlying assets (Black & Scholes 1973, p. 637). If the formula in its original form is the correct way to price options, then there can only be one implied volatility implied by the formula, regardless of the strike price or type of option, i.e., a call or a put. The empirical observations are overwhelmingly against this assertion as the option price most commonly varies with the strike price and maturity. Despite its shortfalls, the BS implied volatilities have over time been shown to be one of the best predictors of volatility (Poon 2005, p. 97) and so the area of interest is to try and explain the links between the implied volatilities from the formula and the actual volatilities.

2.7.1 Assumptions

The Black-Scholes model makes the following assumptions in regards to the pricing of a European option:

1. The stock price follows an Itô-process (Geometric Brownian Motion)
2. Securities are traded continuously
3. There are no arbitrage opportunities
4. It is possible to buy and sell securities at any time to any amounts
5. There are no transaction costs
6. The interest rate of the market is constant
7. Securities do not pay out dividends
8. No short-selling constraints

The intuition behind the BS valuation model in the case of options is that if the underlying asset and the derivative contract share the same source of risk, then it is possible to create a risk-free portfolio (hedge portfolio) by for example buying the underlying asset and selling the derivative contract. As the return of such a portfolio is certain and thus risk-free the return should be equal to that of an investment in the risk-free rate as perfect substitutes must have the same price (Constantinides et al. 2003, p. 1150).

Valuation through replication and the no-arbitrage assumption lays the framework for the Black-Scholes model which recognized that since a risk-free hedge can be created between an option and the underlying asset, it is also possible to replicate the payoffs of for example a call option using a portfolio that consists of the asset and risk-free bonds.

The formula then estimates the composition of the risk-free bond and underlying asset that replicates the payoffs of the option.

While in this thesis the scope is limited to the European style options on a stock index it is worth mentioning that the principles underlying the BS framework for valuation using the no-arbitrage assumption can and has been applied to more complicated assets and derivatives.

2.7.2 Black-Scholes Pricing Formulas for European Options

In this version of the Black-Scholes formula an analytic expression for a European option is derived to the following expressions:

$$C = S_t N(d_1) - K e^{-r(T-t)} N(d_2) \quad (9)$$

$$P = K e^{-r(T-t)} N(-d_2) - S_t N(-d_1) \quad (10)$$

where, C = Call Option, P = Put Option, and

$$d_1 = \frac{1}{\sigma \sqrt{T-t}} [\ln(S_t/K) + (r + \sigma^2/2)(T-t)] \quad (11)$$

$$d_2 = d_1 - \sigma \sqrt{T-t} \quad (12)$$

The BS model is elegant and popular in the industry despite its apparent weaknesses. One problem is that stock prices do not always follow a smooth movement predicted by the Geometric Brownian Motion but instead on occasion jump or fall abruptly. As previously mentioned if all assumptions of the model hold, one expects to observe a flat volatility surface which is not a common empirical finding.

2.8 The Greeks

The Greeks are certain statistical values that are used by traders to evaluate their positions in options. They measure in different ways the risk that is involved in option contracts in relation to its underlying assets. Some of the more popular of these so-called Greeks are Delta, Gamma, Vega, Theta and Rho. For the subject at hand, the most important one is Delta, which is described in the next section.

2.8.1 Delta

Delta is the option's sensitivity to small changes in the underlying price. The specification of "small changes" comes from the fact that it is a first partial derivative, and so is a linear approximation of the price risk of the option. Thus the approximation will get worse the larger the change in the price of the underlying asset.

It is perhaps the most popular of the Greeks for among other things its frequent use in hedging applications in a concept explicitly called delta-hedging. A delta-hedge is motivated from the previous empirical observation that the volatility surface of options is not flat as per the original BS assumptions.

A delta-hedge can be formed as per the previous example by taking a short position on a call option on an asset and a long position in the same underlying asset. The so-called delta-hedge is the amount one needs to invest in the underlying asset to protect the option against the price risk of the underlying. Investment in the underlying asset, by definition, have a delta of one. Thus delta is the most important parameter to consider in hedging and has remained so since the inception of the exchange-traded options market in 1973 (Hull & White 2017, p. 2). Delta can be explained mathematically as:

$$\Delta_C = \frac{\partial C}{\partial S} = N(d_1) > 0 \quad (13)$$

$$\Delta_P = \frac{\partial P}{\partial S} = N(-d_1) < 0 \quad (14)$$

So when taking a position on the options, the delta-hedge will be positive for a call option and negative for a put option. For practical application one would consider dynamic delta-hedging; which means that at certain frequencies in time the investor re-balances the hedge-position according to how the delta changes with current market prices for the underlying and time.

The importance of delta is not only apparent in hedging risk but also for speculators using options replicators. They aim to lock in a potential difference between future realized and current implied volatility, for example, by delta-hedging their options to expiration (Derman & Zou 1999, p. 1). This becomes more complicated if a volatility skew is observed as the entire range of implied volatilities have to be compared to a single historical, realized volatility.

The implications are obvious for portfolio managers that have a risk profile to maintain and so have to take positions and re-balance regularly as the risk for each

included security changes. The manager that is able to predict the risk levels of the securities that are or is to be held in the portfolio can act accordingly and maintain a portfolio with the desired risk characteristics (Andersson 1995, p. 101).

2.9 Rules of Thumb

Rules of Thumb are conditions set by traders to aid in estimation and analysis of volatility surfaces. These apocryphal Rules of Thumb do not always describe the behavior of the volatility surface consistently, which is clearly shown by Derman (1999). They are, however, still widely applied, presumably because they are, to some degree, useful. The various Rules of Thumb that are applied in this thesis can be divided into two main categories.

1. The Rule(s) that are concerned with how the volatility surface changes through time.
2. The Rule(s) that concerns the relationship between the volatility smiles for different option maturities at a given point in time.

The three main Rules of Thumb tested in this thesis is the Sticky Strike, Sticky Delta, and Square Root of Time Rules. Sticky Strike and Sticky Delta belong in the first of the aforementioned categories whilst the Square Root of Time Rule belongs in the second category. The first category of Rules is used to provide a basis to calculate the Greek letters delta and gamma, while the second category, i.e., the Square Root of Time Rule, is used for "filling in the blanks" when producing a complete volatility surface (Hull et al. 2006). In this thesis, the scope is initially limited to these three Rules of Thumb suggested by Hull et al. (2006) but extensions such as higher powers are considered for the purpose of trying to improve the performance. Hull et al. (2006) also mentions that some traders prefer to use the forward value of the asset, F , in place of S , but in this analysis, the scope will be limited to using S . These three main Rules of Thumb are described in the following sections.

2.9.1 The Sticky Strike Rule

The Sticky Strike Rule of Thumb assumes that implied volatility is independent of the asset price. It implies that the sensitivity of the price is equal to:

$$\frac{\partial c}{\partial S} \quad (15)$$

where c is the option price, and so it varies only dependent on the strike price, K , and the time-to-maturity, M . Regression specification according to Hull et al. (2006):

$$\sigma_{KM} = b_0 + \sum_{i=1}^m b_i K^i + \sum_{j=1}^n b_{m+j} M^j + b_{m+n+1} KM + \varepsilon \quad (16)$$

where σ_{KM} is the implied volatility at strike price K and maturity M . In this paper, $m, n \in \{2, 3, 4, 5, 6, 7, 8\}$, b_x are coefficient estimates and ε is a normally distributed error term. This assumption behind this Rule allows delta to be calculated with the Black-Scholes formula with the volatility parameter set equal to the options implied volatility (Hull et al. 2006).

2.9.2 The Sticky Delta Rule

The Sticky Delta Rule of Thumb assumes that implied volatility depends on the moneyness variable K/S (strike price in relation to asset price). The delta, for a European option, is:

$$\Delta = \frac{\partial c}{\partial S} + \frac{\partial c}{\partial \sigma_{KM}} \frac{\partial \sigma_{KM}}{\partial S} \quad (17)$$

Where the first term, $\frac{\partial c}{\partial S}$, is an expression for the delta that has been calculated using the Black-Scholes formula using the implied volatility in place of the volatility, the value of the third term, $\frac{\partial \sigma_{KM}}{\partial S}$, decide whether the implied volatility, σ_{KM} , is increasing or decreasing with the strike. For equity indices which empirically tend to show the reverse skew relationship σ_{KM} should, in general, be a declining function of the strike, K , which is true at least for the chosen sample shown by Figure 1. Thus the Black-Scholes delta will, in this case, be understated compared to the actual delta (Hull et al. 2006).

Regression specification according to Hull et al. (2006):

$$\begin{aligned} \sigma_{KM} = & b_0 + \sum_{i=1}^m b_i \left(\frac{K}{S}\right)^i \\ & + \sum_{j=1}^n b_{m+j} M^j + b_{m+n+1} \left(\frac{K}{S}\right) M + \varepsilon \end{aligned} \quad (18)$$

where, in this paper, $m, n \in \{2, 3, 4, 5, 6, 7, 8\}$, b_x are coefficient estimates and ε is a normally distributed error term. Another version of this Rule presented by Hull et al. (2006) is Relative Sticky Delta. Regression specification for this version:

$$\begin{aligned} \sigma_{KM} - \sigma_{SM} = & b_0 + \sum_{i=1}^m b_i \left[\ln\left(\frac{K}{S}\right) \right]^i \\ & + \sum_{j=1}^n b_{m+j} M^j + b_{m+n+1} \ln\left(\frac{K}{S}\right) M + \varepsilon \end{aligned} \quad (19)$$

where, as before, σ_{KM} is the implied volatility at strike price K and maturity M while σ_{SM} is the implied volatility at spot price S and maturity M . In this paper, $m, n \in \{2, 3, 4, 5, 6, 7, 8\}$, b_x are coefficient estimates and ε is a normally distributed error term. Previous research by Andersson (2003) and later confirmed by Hull et al. (2006) asserts that the Relative Sticky Delta Rule is at least approximately consistent with the no-arbitrage condition even when skewness is observed which could imply more accurate results also on the Swedish equity index.

2.9.3 The Square Root of Time Rule

This Rule was named the Square Root of Time Rule by Hull et al. (2006) but has been suggested at least since Natenberg (1994). It belongs as previously mentioned in the second category of Rules, which means that it concerns the relationship between the implied volatilities for a variety of maturities, at a given point in time. Two versions were suggested by Hull et al. (2006). First is the regression specification that is tested by Hull et al. (2006):

$$\sigma_{KM} - \sigma_{SM} = \sum_{i=1}^k b_i \left(\frac{\ln(K/S)}{\sqrt{M}} \right)^i + \varepsilon \quad (20)$$

and another version that Hull et al. considers but do not test:

$$\frac{\sigma_{KM}}{\sigma_{SM}} = \sum_{i=1}^k b_i \left(\frac{\ln(K/S)}{\sqrt{M}} \right)^i + \varepsilon \quad (21)$$

This paper will also focus on the former, where $k \in \{2, 3, 4, 5, 6, 7, 8\}$, b_x are coefficient estimates and ε is a normally distributed error term. Hull et al. (2006) doesn't go into detail as to why they do not test the second version of the Square Root of Time Rule but it is most likely because the ratio expression can impair intuitive interpretation of the estimation results. They make comparisons between models mainly using the RMSE which may not be easily comparable when applied to a ratio. It would also be more complicated to back out a complete predicted volatility surface which motivates the usage regression specification outlined in Equation 20.

Like with the Relative Sticky Delta Rule, Hull et al. (2006) asserts that it can be approximately consistent with the no-arbitrage condition, even with the skewness observed in equity indices. This again can imply that this Rule possibly has increased predictive power as compared to some of the previously mentioned Rules, i.e., the Sticky Strike and the basic Sticky Delta Rule.

3 Data and Methodology

3.1 Data

Ten years of daily data of the OMXS30 index is gathered through the Bloomberg terminal totaling 2512 daily (trading days) observations. This includes implied volatility for European options on the index at maturities 1, 3, 6 and 18 months and moneyness' 0.8, 0.9, 0.95, 0.975, 1.0, 1.025, 1.05, 1.1, 1.2; totaling 90,432 (36 daily) observations of implied volatilities.

The data for implied volatilities is expressed in units of the decimal representation multiplied by one hundred. The data for the maturities is expressed in actual months as opposed to a percentage of a year, this in combination with the larger units for the implied volatilities limits the numbers of decimals in the estimates and aids in intuitive interpretation.

3.1.1 Data Processing

The data is compiled into a table and imported into Matlab. Here the data points are structured in a key-value map where each key corresponded to a day, and each value is a 9-by-4 matrix of implied volatilities. The various models and specifications are then scripted in Algorithm A and B outlined in Section 3.3.1.

3.2 Stepwise Regression

A stepwise procedure involves starting with a preset number of explanatory variables and then adding or removing variables. Further addition or subtraction of variables depends on measuring the improvement according to some criterion such as p-values or improvements in fit (Brooks 2008, p. 111). If the new version of the model is considered an improvement on the previous one the process repeats itself until the addition or subtraction of a variable no longer improves the model according to the chosen criterion. The method is commonly criticized for its tendency to produce results that suffer from data-mining bias which is described further in a section below. In this paper, the stepwise methodology is applied to linear regressions of various specifications with robust standard errors.

3.2.1 Data-Mining Bias

A common problem with doing a large number of tests on a given data set, and likewise with the stepwise methodology, is that of data-mining bias. Data-mining, in itself, is the methodology of scraping large amounts of data to try and find correlations, trends, and similarities. When using semi- or fully automated testing procedures to identify patterns in a set of data a statistically significant pattern is likely to appear.

The apparent relationship found, such as significant correlation, can be a spurious correlation, i.e., when the variables are statistically related but not causally related. A spurious relationship might be an unlikely coincidence but given a process which systematically tests for various relationships in large volumes this rare coincidence is more likely to appear. The stepwise methodology thus tends to overfit models when relying on some in-sample measure of fit. There are various methods for trying to diagnose spurious relationships, but the most reliable method to avoid it is through the use of out-of-sample tests (Constantinides et al. 2003, p. 1090). The researchers, therefore, rely on out-of-sample RMSE to avoid potential Data-mining bias.

3.2.2 Period Partitioning

In empirical analysis, the size of the observation windows is often chosen seemingly arbitrarily or according to industry praxis. In extension, the same observation window size is commonly used for all tested models and specifications. This, however, can potentially have a large impact on the model accuracy as it is not certain that the same observation window size gives the most accurate predictions over the entire set of data.

If differences are observed in regards to how the models perform with different sizes of the observation windows, both inter- and intra-models, it can be interesting to analyze whether there are some identifiable patterns or trends and in this case, if there are underlying variables that explain these patterns. In practical applications, the point of interest is often to try and estimate one-day ahead. While previous research has primarily focused on monthly or weekly data it seems likely that daily data may, in general, make more accurate predictions for the one-day-ahead implied volatility. The effect of the size of the observation window on each model's performance is thus tested to determine what window size yields the best fit and if any trends in the optimal window size can be identified.

3.2.3 Forward Selection

A stepwise regression can be performed in three main ways with regards to the direction of the explanatory variables that are added or removed; Forward Selection, Backward Elimination, or Bidirectional Elimination. In this thesis, the Forward Selection method is utilized which means that the models start with the lowest amount of explanatory variables and then adds additional ones for evaluation. One difference between the typical Forward Selection approach and the one that is used in this research is that it does not stop once an additional variable does not improve the fit but continues testing all predetermined combinations of regression specifications, window sizes, and Rules of Thumb. This avoids the possibility of not testing a model which may have yielded better fit.

3.2.4 Walk Forward Analysis

The concept of Walk Forward Analysis is simple but efficient in avoiding many of the pitfalls in statistical analysis. By optimizing the model on a selected number of days and then testing it on the period after the training set (test set) biases such as overfitting and data-mining are avoided (Pardo 2008, pp. 237-241). The process then moves forward one period and repeats. This is essentially an advanced version of out-of-sample testing and a specific version of Cross-validation. It results in a sizeable out-of-sample period and truly tests the predictory power of the models considered over the entire data set.

3.3 Model Selection

The model selection in this paper is based on three different criteria: in-sample AIC, in-sample BIC, and out-of-sample, one-day-ahead RMSE. The purpose of AIC and BIC criteria are mainly to compare different numbers of parameters to be included in the regression specification since both of these criteria penalize additional terms. The purpose of the out-of-sample RMSE is to test the models' one-day-ahead predictive power.

First, some definitions. Throughout the following sections, the following terms will be utilized:

- *Rule of Thumb*: An individual Rule of Thumb (e.g., 'Sticky Strike').
- *Power Component*: An individual power term in a regression specification (i.e., Sticky Strike with $m = 3$ and $n = 2$ has five Power Components total)
- *Regression Configuration*: An individual specification of the number of Power Components and the window size for a given regression (e.g., 'three powers components on the first terms of the regression ($m = 3$) and two Power Components on the second ($n = 2$) and window size two days').
- *Model Specification*: A collective term for a combination of Rule of Thumb and Regression Configuration (e.g., 'Sticky Strike with $m = 3$, $n = 2$, and window size two days')
- *Model Criterion*: Collective term for $RMSE_{T+1}$, AIC, and BIC.

As previously stated, the authors conduct a rolling window analysis for in-sample and out-of-sample predictive performance. This is done through backtesting of different observation window size (number of days) between 2 and 40 for each Model Specification and recording the combination that yielded the best fit for the three fit criteria ($RMSE_{T+1}$, AIC, and BIC). Initial tests suggested that 40 days is more than enough for the upper limit of increasing fit and two days is chosen as the lower bound as it results in 72 observations (implied volatilities) which are statistically more sound than just 36 observations which is achieved from a one-day observation window.

3.3.1 All Rules of Thumb Tests

Two main strategies are used to evaluate a best fit model. The first one (Algorithm A) tests the performance of Model Specifications on the entire data set and outputs the best Model Specification in terms of *average* $RMSE_{T+1}$, AIC, and BIC. In addition, the algorithm outputs the optimal Regression Configuration for each isolated Rule of Thumb. Algorithm B tests the performance of Model Specifications on individual days and outputs the best Model Specification for *each* day. In addition, this algorithm also outputs the optimal Regression Configuration for each isolated Rule of Thumb for each day. Preliminary testing showed that the performance of tests with large observation window performs

poorly and therefore both algorithms are limited to test observation windows between 2 and 30 days, inclusive. Due to the exponential nature of the computational requirements needed for each addition Power Component m , n and k are limited between two and eight, inclusive. This is reflected in the regression specifications in Equations 16, 18, & 19 for the Sticky Strike, Sticky Delta, and Relative Sticky Delta Rules. Here there are $7 * 7 = 49$ different combinations of m and n for each of these Rules of Thumb. For the Square Root of Time Rule, k may only take on seven different values, as expressed in the regression specification in Equation 20. Hence each algorithm tests $39 * 3 * 49 + 39 * 7 = 6006$ different Model Specifications.

Algorithm A performs the following operations:

1. Specify Rule of Thumb
2. Specify Regression Configuration
3. Run rolling window regressions on the entire data set
4. Save Regression Configuration and *average* Model Criterion
5. Repeat step 1 through 4 on the next regression configuration
6. Compare the *individual* saved average Model Criteria with the latest run's average Model Criteria.
7. If anyone Model Criterion is lower (read: better) than in the previous run, the previously saved Regression Configuration(s) and Model Criterion/a is replaced to reflect this.
8. Repeat for all possible Model Specifications

In other words, the algorithm runs a rolling window regression on the entire data set for every Model Specification and returns the superior Model Specification for each of the three Model Criteria as well as the superior Regression Configuration for each Rule of Thumb per the Model Criteria. Different Model Specifications can be selected for each of the three Model Criteria.

As the researchers want to test for the best model at individual days, Algorithm B performs the following operations:

1. Specify Rule of Thumb
2. Specify Regression Configuration
3. Run single regression on a single day
4. Save Regression Configuration and Model Criterion in an indexed matrix where the index corresponds to the day
5. Repeat step 1 through 4 on the next Regression Configuration
6. If any one of the Regression Criterion is lower than the previous run, the matrix entry for that day is changed to reflect this
7. Repeat for all possible Model Specifications
8. Move one day forward and repeat steps 1 through 7

In other words, once all Model Specifications are exhausted, the script moves on to the next observation day, and so on until it has progressed through the entire data set. Once complete, Algorithm B yields a 2468-by-76 matrix represented in Figure 2 in the Results section. Both algorithms run a total of approximately $6006 * 2468 \approx 14.8$ million regressions.

3.3.2 Further Analysis: Square Root of Time Rule

As the Square Root of Time Rule is established as the Rule with the most accurate predictions, additional analysis is made to ensure that the most accurate specification of the Rule is utilized. Thus the powers of the Rule is extended in further tests so that $k \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$ in Equation 21.

3.3.3 Volatility Climate

In order to evaluate how the predictive performance of the suggested models vary with the volatility climate they are compared to an implied volatility index using linear regression. The VSTOXX index is the most relevant measure of the current volatility climate in Europe. The 5, 10 and 50-day simple moving averages of this index are also used in the statistical analysis.

4 Results

4.1 All Rules of Thumb Tests

The results from running Algorithm A are presented in the table below:

[INSERT TABLE 1 ABOUT HERE]

The Model Specification which yields the lowest average $RMSE_{T+1}$ in the rolling window analysis of the entire data set, Square Root of Time with $k = 8$ and observation window size of three days, is shown in the first row. The Model Specification which yields the lowest AIC is shown below, and so on. The last three rows show the lowest $RMSE_{T+1}$ obtained by the optimal Regression Configurations for the remaining Rules of Thumb. A small section of the output matrix from Algorithm B is presented in the table below:

[INSERT TABLE 2 ABOUT HERE]

The left group of columns titled 'Optimal Overall' shows the overall optimal Model Specification(s) per the three Model Criteria for each day. The four right groups of columns show the optimal Regression Configurations for each individual Rule of Thumb per the three Model Criteria. Two of these are omitted as they do not fit on the page. The rows show the observation for which the results pertain to. Observation 47 to 2505 are omitted as they do not fit on the page.

A variety of graphs are created to aid in the analysis of the above table. The overall optimal window size results from Algorithm B are plotted in a histogram below:

[INSERT FIGURE 3 ABOUT HERE]

The top, middle, and bottom graphs show the number of times a window size yield the lowest $RMSE_{T+1}$, AIC, and BIC, respectively across all Model Specifications for the entire data set. In other words, the x-axes show the window sizes, and the y-axes show the number of times each is superior.

The following histogram shows the frequency of which a model is superior:

[INSERT FIGURE 4 ABOUT HERE]

Here the x-axes show the Rules of Thumb tested. Similar to the previous plot, the y-axes show the number of times each Rule of Thumb had the lowest $RMSE_{T+1}$, AIC, and BIC. The next graph shows the optimal Power Components of Regression Configuration:

[INSERT FIGURE 5 ABOUT HERE]

The histograms are arranged by columns and rows where the columns pertain to the Rules of Thumb and the rows pertain to the Model Criterion. All histograms show the number of times a Regression Configuration is superior on the y-axis (read: up) in terms of Power Component specification. The left six bivariate histograms show the Power Component specifications on the x and z-axes. The x-axis shows the number of Power Components for the first term in the Regression Configuration (i.e., m), and the z-axis shows the number of Power Components on the second (i.e., n). Similarly, the right histograms show the number of Power Components for the Regression Configuration for Square Root of Time (i.e., k) on the x-axis.

The final comparison of the Rules of Thumb yields the results from running rolling window regressions on the single best Regression Configurations for each Rule of Thumb per the results above on the entire data set and recording $RMSE_{T+1}$ for each observation. The chosen Regression Configuration for each Rule of Thumb can be seen in the table below:

[INSERT TABLE 3 ABOUT HERE]

The resulting 4-by-2508 matrix of $RMSE_{T+1}$ values is plotted in the following bivariate histogram:

[INSERT FIGURE 6 ABOUT HERE]

The z-axis shows the Rules of Thumb and the x-axis shows $RMSE_{T+1}$. The y-axis shows the number of observations between the $RMSE_{T+1}$ interval specified on the x-axis. For example, it is observed that just over 350 observations' regressions generated an out-of-sample RMSE between 0.5 and 0.75 for the Square Root of Time run.

4.2 Further Analysis: Square Root of Time Rule

Upon analyzing the results above more, the tests for Square Root of Time are expanded. The reasoning for this is described in the Analysis section below. The following plots show the results of running the script on Square Root of Time with power terms from 2 to 15 ($k = 2 : 15$) and window size between 2 and 50:

[INSERT FIGURE 7 ABOUT HERE]

The left, middle, and right bivariate histograms pertain to $RMSE_{T+1}$, AIC, and BIC, respectively. In these, the number of Power Components (k) appears on the x-axis, and the window size appears on the z-axis. As per usual, the y-axis shows the number of times a Regression Configuration is superior. Based on these results, a final figure is generated. This is Figure 8 below, which pertains to running a rolling window regression on Square Root of Time with twelve Power Components ($k = 12$) and a window size of two days. For each run, the p-values for each coefficient is saved, and these make up the bivariate histogram:

[INSERT FIGURE 8 ABOUT HERE]

This plot shows each Power Component coefficient index on the z-axis (i.e., i) and the respective p-values for that coefficient for all the runs on the x-axis. The y-axis shows the number of observations generating a significance level *between* the values represented on the x-axis for each coefficient indexed by the z-axis. To read the graph note that each of the twelve 'rows' from 1 to 12 represents 2500 p-values. Looking at row one, it can be seen that the first coefficient is significant for the vast majority of observations, and so on.

4.3 Volatility Climate

Plotting the $RMSE_{T+1}$ for the above run against VSTOXX generates the following graph:

[INSERT FIGURE 10 ABOUT HERE]

The left and right y-axes show the $RMSE_{T+1}$ and VSTOXX Index levels, respectively. The x-axis shows the observation number. The various regressions tested on the relationship between these two did not yield any meaningful relationships.

5 Analysis and Interpretation

As a large number of Model Specifications and observation window sizes are tested in various ways in order to thoroughly examine the data, the results and the distributions of these are sometimes more intuitively interpretable with a visual representation. Therefore a variety of graphs and tables representing the results are supplemented throughout the Results section, which are referred to accordingly.

5.1 All Rules of Thumb

From the results shown in Table 1 and Figure 4, it is evident that the Square Root of Time Rule is preferred in both algorithms by all three Model Criteria; AIC, BIC and most importantly, out-of-sample RMSE. It is interesting that this Rule outperforms the others to such a large degree, and this likely has to do with the skewness observed in equity indices such as the OMXS30. For such indices, the Square Root of Time Rule allows more explanatory flexibility according to Hull et al. (2006). This makes the additional analysis first considered by the authors to try and find which Rule of Thumb that performs best during different periods and volatility climates unnecessary as the best strategy is seemingly to use the Square Root of Time Rule throughout.

Upon analyzing the outcomes of the two algorithms, a conflicting result arises on which window size yields the optimal results. In Table 1, Algorithm A suggests an optimal window size of three days for Square Root of Time while Figure 3 suggests that a two-day observation window size is optimal overall. On closer inspection of the Square Root of Time results from Algorithm B it is confirmed that this algorithm favors a two-day observation window for this Rule. These discrepancies do not arise for the other Rules as both Algorithm A and B selected a two-day observation window for all three. It appears that when the window size is not allowed to vary between days for the Square Root of Time Rule, there are observations which generate an $RMSE_{T+1}$ great enough to negatively affect the average for a two-day observation window. Hence, Algorithm A prefers a three-day observation window even though Algorithm B shows that a two-day observation window produces the highest frequency of most accurate predictions. The best approach is seemingly to use a fixed observation window size of two days with perhaps a sanity check consisting of a prediction with an observation window size of three days.

Some might make the case that a three-day observation window is a preferable model as it can be considered less risky. Both AIC and BIC do, however, consistently prefer an observation window size of two days which means that the \widehat{L} term is higher since the number of parameters, k , is the same regardless of observation window size, this further supports the choice of an observation window size of two days. These results are interesting and the authors consider the analysis merited in multiple areas of the industry where the choice of observation window size is sometimes ignored or chosen according to praxis. The implied volatility is highly variable and notoriously difficult to identify consistent patterns in. It makes intuitive sense then that the smaller the period that the proposed model is fitted onto, the easier it is to get an improved fit. Taking all factors into account, the authors decided to focus on a window size of two days.

Figure 5 also reveals an interesting result which is that both the in-sample information criteria, AIC, and BIC, also prefer the highest power for the Square Root of Time Rule which means that the punishing effect of additional parameters is not enough for the exclusion of them. These results are again echoed by the results from Algorithm A. The tendency of AIC to fit a larger model than BIC is also visible in Figure 5. Figure 6 shows the results of running the specifications of each Rule of Thumb that performed best the most times according to the $RMSE_{T+1}$ in the previous tests. These results speak for how much more accurate the Square Root of Time Rule is when compared to other Rules of Thumb.

5.2 Further Analysis: Square Root of Time Rule

After expanding the analysis on the Square Root of Time Rule it becomes apparent that the model performs even better when k is allowed to take the value twelve as seen in Figure 7. This is consistent amongst all three Model Criterion.

Based on this analysis on the given set of data, the researchers have landed on the following optimal Model Specification for predicting a $T + 1$ implied volatility surface:

$$\sigma_{KM} - \sigma_{SM} = \sum_{i=1}^{12} b_i \left(\frac{\ln(K/S)}{\sqrt{M}} \right)^i + \varepsilon \quad (22)$$

where, observation window size = 2. The inclusion of parameters up to the twelfth power can be considered excessive as the coefficient estimates become rather uninterpretable for the higher powers. These estimates are usually not all statistically significant simultaneously which also makes interpretation of the relationships complicated. This Model Specification is selected on the basis of being the best performing model the most times and by a large margin.

The coefficient estimates are to a large degree statistically significant at the 0.05, 5%, significance level as is shown by the p-values in Figure 8. Also, since the Model Specification is chosen using out-of-sample predictions, the inclusion of these additional parameters are motivated for the improvement of $RMSE_{T+1}$. For powers higher than twelve no improvement in $RMSE_{T+1}$ is observed at any point, this is likely because the term, $\left(\frac{\ln(K/S)}{\sqrt{M}} \right)^i$, closely approaches zero and so Matlab does not produce an estimate. Figure 9 compares a sample date's predicted surfaces using Square Root of Time with four Power Components to one with twelve Power Components. Four Power Components for the Rule is the greatest number included in the analysis of Hull et al. (2006). The improvement is visually noticeable for the sample observation.

The choice of Rule of Thumb is consistent with the previous research on the S&P 500 by Hull et al. (2006) which is reasonable since that analysis also was made on a stock market index also showing the reverse skew relationship. The Swedish stock market is also influenced by the American counterpart which may explain why the same Rule of Thumb yields the most accurate results. As previously stated, the Square Root of Time Rule allows for greater flexibility in the shaping of the surface, especially when higher powers are included in the model. Thus, while the assumptions of the original Black-Scholes model are as expected violated, it seems that by using versions of the apocryphal Rules of Thumb one can, with decent accuracy, predict what the volatility surface on-day-ahead.

The implications can potentially be significant especially for traders involved with delta-hedging or speculation for whom every increase in accuracy means a competitive advantage. The model proposed is still relatively simple in its current form but can be built upon using more advanced methods discussed in the further research section below (see. Section 6.1).

5.3 Volatility Climate

The researchers are interested in exploring whether each Model Specification's $RMSE_{T+1}$ correlates with the VSTOXX implied volatility index. Figure 10 shows $RMSE_{T+1}$ for one Model Specification, as represented by the blue X's, in relation to the VSTOXX index level, represented by the orange line. Multiple regressions are run on these yielding no meaningful relationship. These tests also include smoothing measures such as a 5, 10, and 50-day simple moving averages; these could be interpreted as a measure of the overall current volatility climate in Europe. It is therefore concluded that VSTOXX cannot, at least for the used methodology, explain the performance of the prediction.

Some initial tests are run in an attempt to explain the variability in observation window size in relation to VSTOXX. These tests also included smoothing measures such as simple moving averages. No meaningful relationships are found culminating in a conclusion that a fixed observation window size of two days is, according to current empirical findings and with the limitations of this data set, optimal for the Rules of Thumb.

A general shift in the volatility climate can, however, be observed around the five-year mark in Figure 10; around observation 1250. This downward shift in volatility coincides with a shift to a lower general $RMSE_{T+1}$. This confirms the quite obvious fact that the model has stronger predictive power in more stable volatility climates.

5.4 Model Limitations

The results and the preferred Model Specification suffer some limitations. First of all, the initial data collected features an uneven interval between maturity as well as moneyness. While this reflects the authors approach of collecting the most relevant data (as a majority of options feature a maturity under one year and are near-the-money), this has the effect of making the calculations predisposed to selecting models which performed best in the middle-left area of the volatility surface. As seen in Figure 2 the volatility surface

generated with this data created a surface with higher resolution in certain areas (each intersection of the surface wire-frame represents a data point), and thus the middle-right area of the aforementioned surface is to a large degree interpolated. This results in the $RMSE_{T+1}$ being disproportionately weighted for the high-resolution areas of the surface.

The Model Specification, therefore, generates less accurate predictions on the extremes. In other words, the Model Specification is at its weakest at the longest maturities and lowest/highest levels moneyness'. To remedy this, researchers could evenly collect data across all maturities and moneyness'. An alternative approach could be to give greater weight to the lower resolution areas of the surface or, for example, weight the data points by multiplying the errors, ϵ , for these underrepresented points with some constant C where:

$$C = f(D) \tag{23}$$

where, D is the distance from the highest resolution point. This said, the researchers favor the former approach as the second may cause further complications if function f is not defined properly. One way to implement f is through the use of a so-called Kalman filter which, in simple terms, means that the weighting factor adjusts as the predicted values are compared to the actual outcomes.

Since the implied volatility can be viewed as consensus volatility among all market participants, it will usually adjust when the historical volatility increases or decreases as traders adjust their expectations. Over long samples, they are empirically found to correlate well enough for the implied volatility to be one of the more popular methods for estimating the volatility (Natenberg 1994). These discrepancies can, however, be quite substantial, especially in the short term, which is the case when using daily data as opposed to weekly or monthly. The results indicate that the use of models based on daily data, in general, make stronger predictions for the one-day-ahead implied volatilities. However, it is noteworthy that such a model should be applied with caution as a model shaped only on the data of the previous two days is be more susceptible to make the occasionally unreasonable prediction.

6 Conclusion

In the effort to analyze the implied volatility surface on the Swedish stock market index OMXS30 and how well the already established Rules of Thumb manage to explain and predict it the results are consistent with previous research in that the Square Root of Time Rule is found to be the best performing. The additional analysis of observation window size yielded unsurprising results but is nonetheless interesting and merited in that it further supports the final choice of two days. In the end, the inclusion of powers up to twelve made the most accurate predictions for the Square Root of Time Rule and so that is the best Model Specification among the ones that are tested. This version of the model achieved an average $RMSE_{T+1}$ of 1.4077, which is a substantial improvement over the best specifications of the other Rules of Thumb.

When each specification is held constant over the entire data set the Square Root of Time Rule with power up to twelve and an observation window size of three days delivered a slightly lower $RMSE_{T+1}$ than the same model but with an observation window size of two days. The interpretation of this is that the version with a window of two days predicts more accurately most of the time, but sometimes inaccurate predictions occur which affect the average performance. The suggestion is to either use two days with caution or opt for an approach combining the two different specifications with one version acting as a sanity check.

The analysis builds on previous research that has mainly been performed using less frequent data and fewer variations of the Model Specifications. Higher powers have frequently been ruled out from analysis since the parameters with the lower powers explain most of the surface. The inclusion of these higher powers in the analysis supports the decision to include or exclude them from the final Model Specifications. The authors also aimed to make the material more digestible for a broad group of readers by supplementing the text with a wide variety of graphs and tables that gives some intuition to the vast amount of data.

The question on whether patterns are identifiable for when specific Rules of Thumb performed better in certain volatility climates is ultimately unfruitful since no significant relationships can be found and the results are strongly in favor of the Square Root of Time Rule nevertheless.

The final model specification is one that potentially can be useful in practical applications as it is derived from the goal of making the most accurate one-day-ahead prediction. This is often what industry professionals are interested in as any further predictions tend to get increasingly inaccurate. The high variability of volatility makes it a good candidate for more frequent data use, but as previously mentioned, it should be applied with caution as a few irregular days occasionally can yield unreasonable predictions.

6.1 Further Research

The authors see a couple of areas where the research can be continued and improved. The inclusion of more implied volatility data points results in shorter distances of interpolations in the surface and is a simple way to increase the information density, especially a maturity between the 6 and 18-month points as this is currently quite a sizable interpolation. There are relatively low volumes of options actively traded with even longer maturities, but one can consider including, for example, 24 months as well. The number of moneyness' levels to include is also a consideration that future research might want to alter.

In terms of methodology, the analysis is quite computationally intensive, which perhaps can be performed more efficiently with, for example, maximum likelihood estimation. Instead of the version of stepwise regression with forward selection, one could utilize least absolute shrinkage and selection operator (LASSO) which is an analysis method that potentially can improve the prediction accuracy and the interpretability of the regressions. The common criticism against the stepwise regression method, the risk of data mining, should though have been avoided by using out-of-sample RMSE.

Some researchers such as Bedendo & Hodges (2009) and Andersson (2014) have looked into the possibility of utilizing linear quadratic estimation, also known as Kalman filtering, in order to aid in the estimation of the volatility surface with mixed results. Researchers looking further into the topic at hand likely want to consider these methods as well.

References

- Andersson, G. (1995), *Volatility Forecasting and Efficiency of the Swedish Call Options Market*, Nationalekonomiska Institutionen Handelshögskolan vid Göteborgs Universitet, p. 101.
- Andersson, K. (2003), ‘Stochastic volatility’, *U.U.D.M. Project Report, Department of Mathematics, Uppsala University* .
- Andersson, M. (2014), Models for the dynamics of implied volatility surfaces, Master’s thesis, Chalmers - Göteborg University.
- Bates, D. S. (1997), ‘The skewness premium: Option pricing under asymmetric processes’, *Advances in Futures and Options Research* **9**, pp. 51–82.
- Bedendo, M. & Hodges, S. D. (2009), ‘The dynamics of the volatility skew: A kalman filter approach’, *Journal of Banking Finance* **33**(6).
- Black, F. & Scholes, M. (1973), ‘The pricing of options and corporate liabilities’, *The Journal of Political Economy* **81**(3), p. 637.
- Brooks, C. (2008), *Introductory Econometrics for Finance*, ninth edition edn, Cambridge University Press, p. 111.
- Constantinides, G., Harris, M. & Stulz, R., eds (2003), *Handbook of the Economics of Finance*, Vol. 1, Part B, 1 edn, pp. 705, 1090, 1146–1150.
- Derman, E. (1999), ‘Regimes of volatility: Some observations on the variation of S&P 500 implied volatilities’, *Quantitative Strategies Research Notes, Goldman Sachs* .
- Derman, E. & Zou, J. (1999), ‘Strike-adjusted spread: A new metric for estimating the value of equity options’, *Quantitative Strategies Research Notes, Goldman Sachs* .
- Hull, J. C. (2015), *Options, Futures, and Other Derivatives*, ninth edition edn, Pearson.
- Hull, J. C., Daglish, T. & Suo, W. (2006), ‘Volatility surfaces: Theory, rules of thumb, and empirical evidence’, *Quantitative Finance* **7**(5).
- Hull, J. C. & White, A. (2017), ‘Optimal delta hedging for options’, *The Journal of Banking and Finance* **82**, p. 2.

- Natenberg, S. (1994), *Option Volatility & Pricing: Advanced Trading Strategies and Techniques*, pp. 51, 290.
- Pardo, R. (2008), *The Evaluation Optimization Trading Strategies*, second edition edn, John Wiley & Sons Ltd, pp. 237–241.
- Poon, S.-H. (2005), *A Practical Guide to Forecasting Financial Market Volatility*, John Wiley & Sons Ltd, p. 97.
- Tsay, R. (2010), ‘Analysis of financial time series: Third edition’, *Wiley series in probability and statistics. Financial engineering section* p. 80.

Appendices

Appendix: Figures

Figure 1: Volatility Skew/Smirk

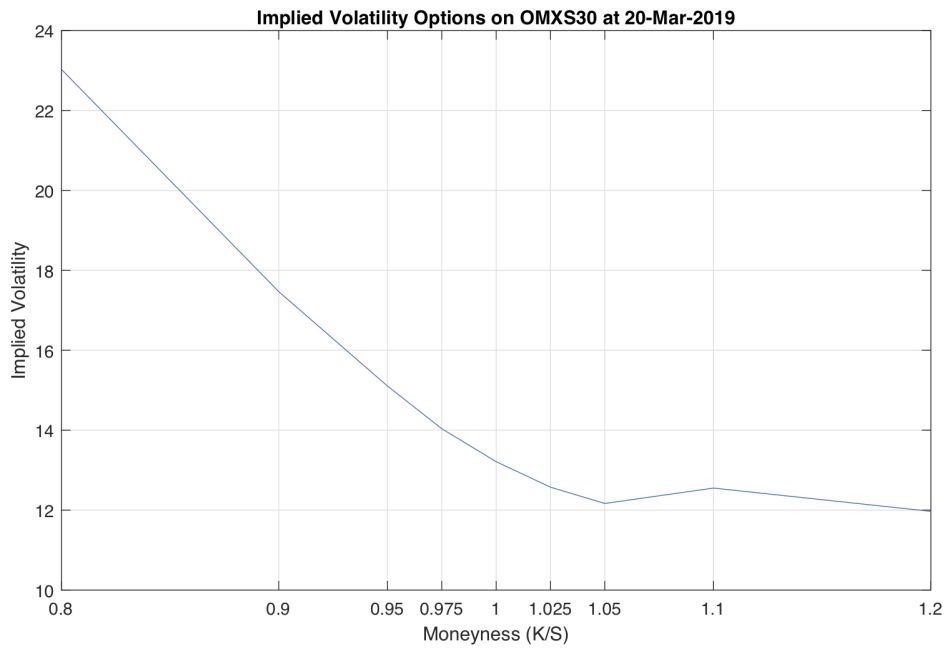


Figure 2: Volatility Surface

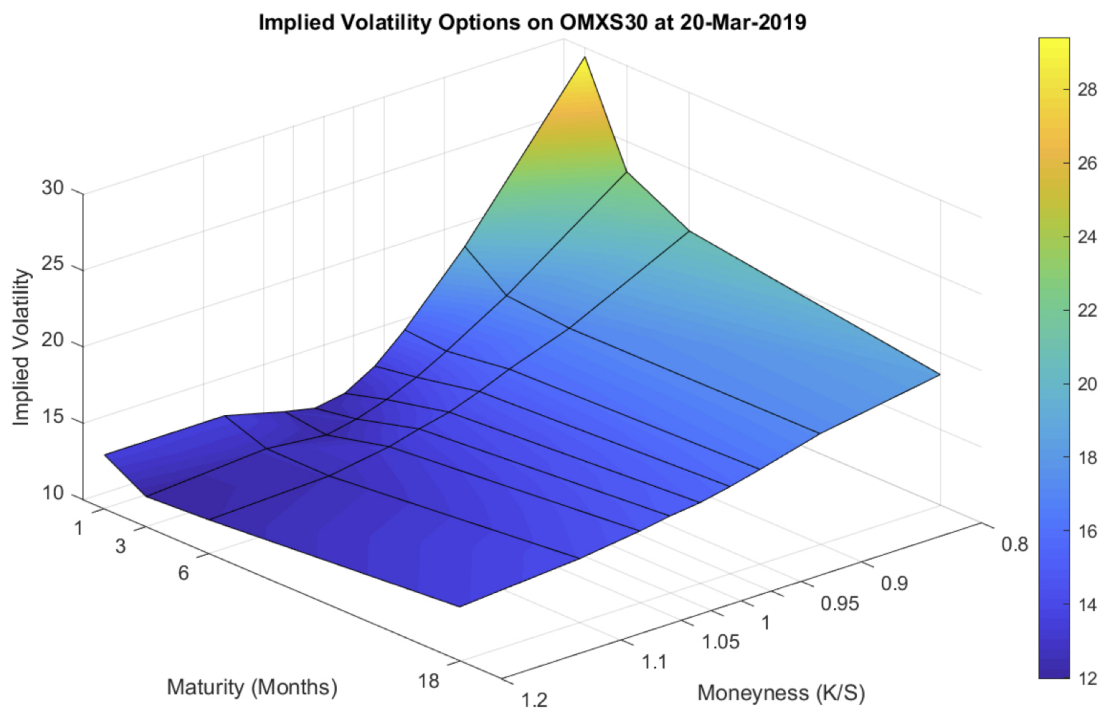


Figure 3: Window Size Analysis

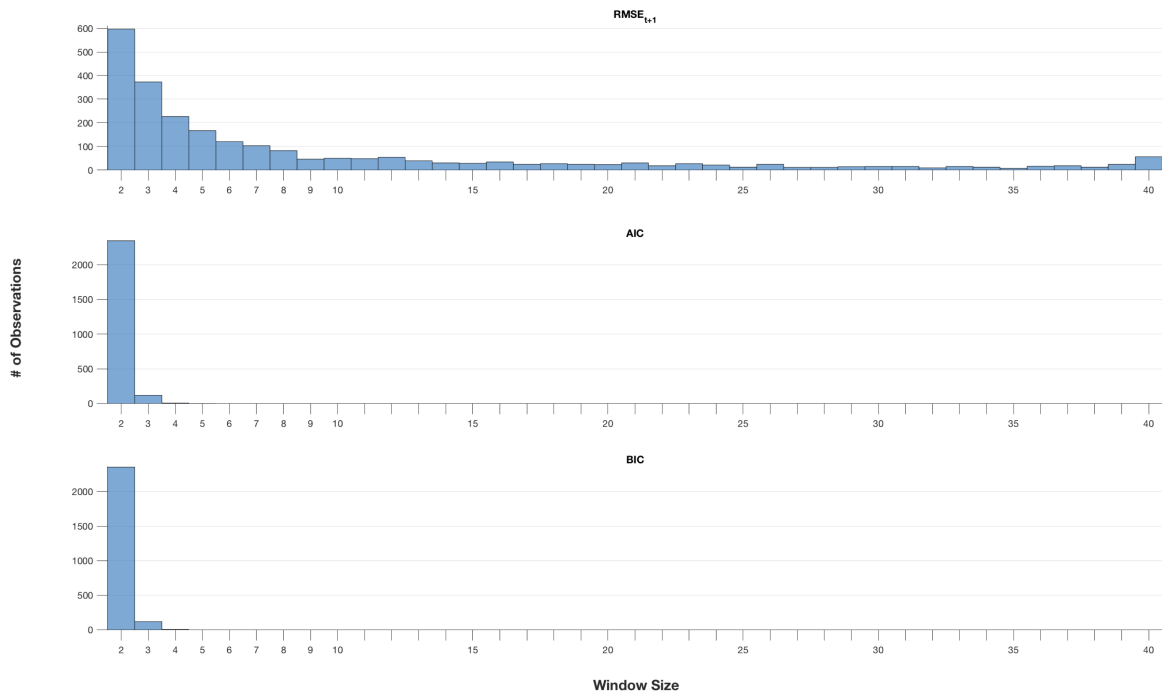


Figure 4: Model Selection Analysis

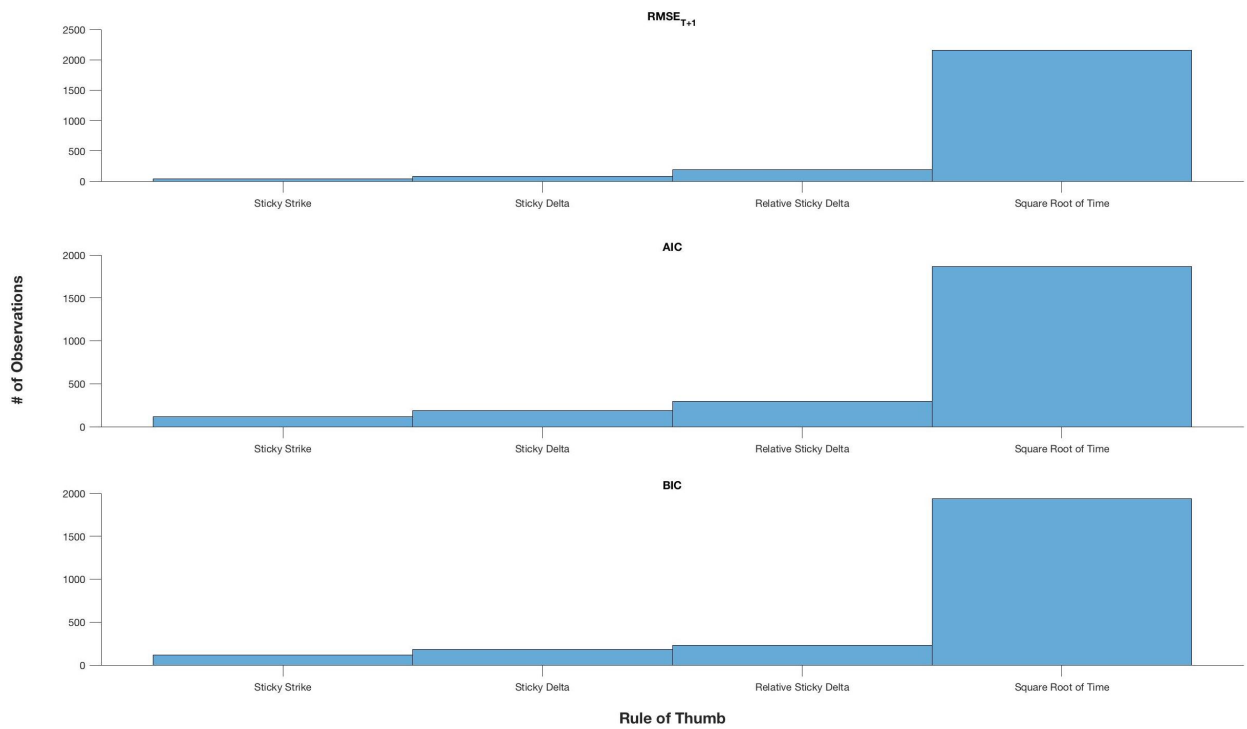


Figure 5: Model Powers Analysis

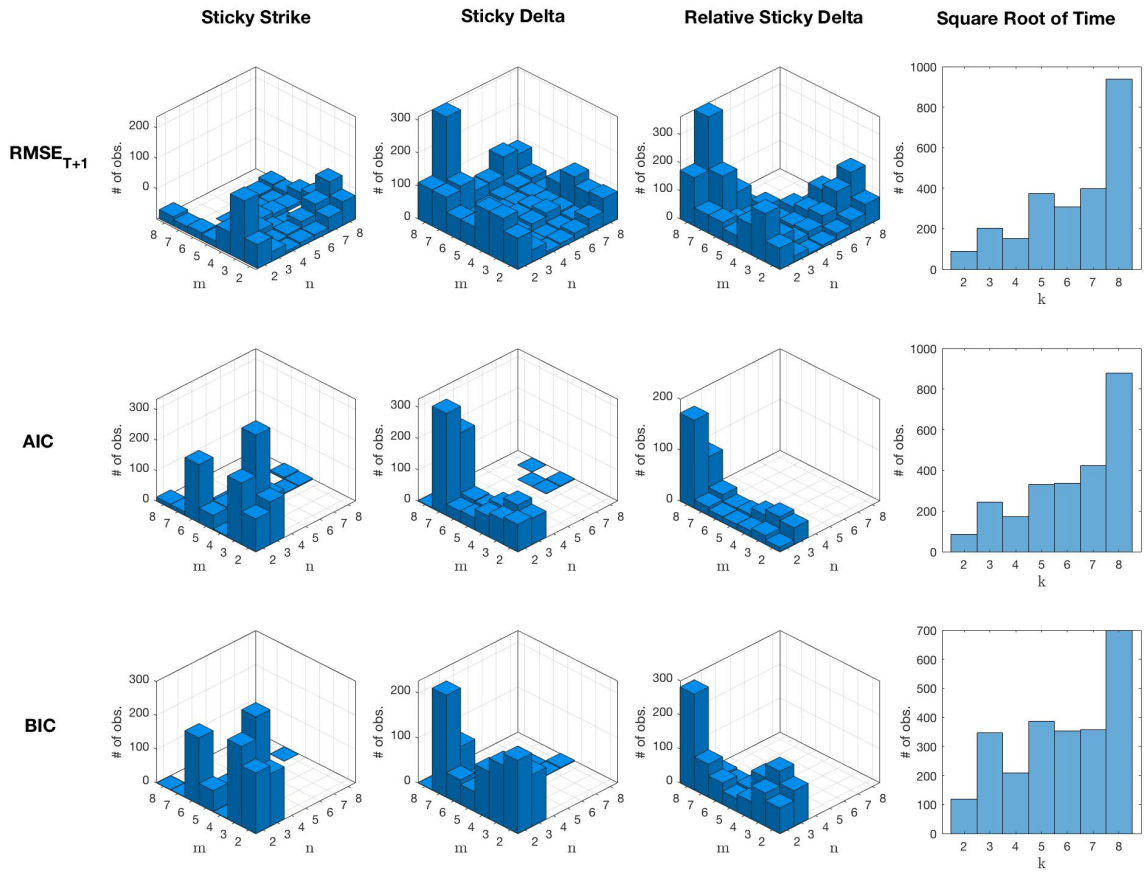


Figure 6: Optimal Model Specification $RMSE_{T+1}$

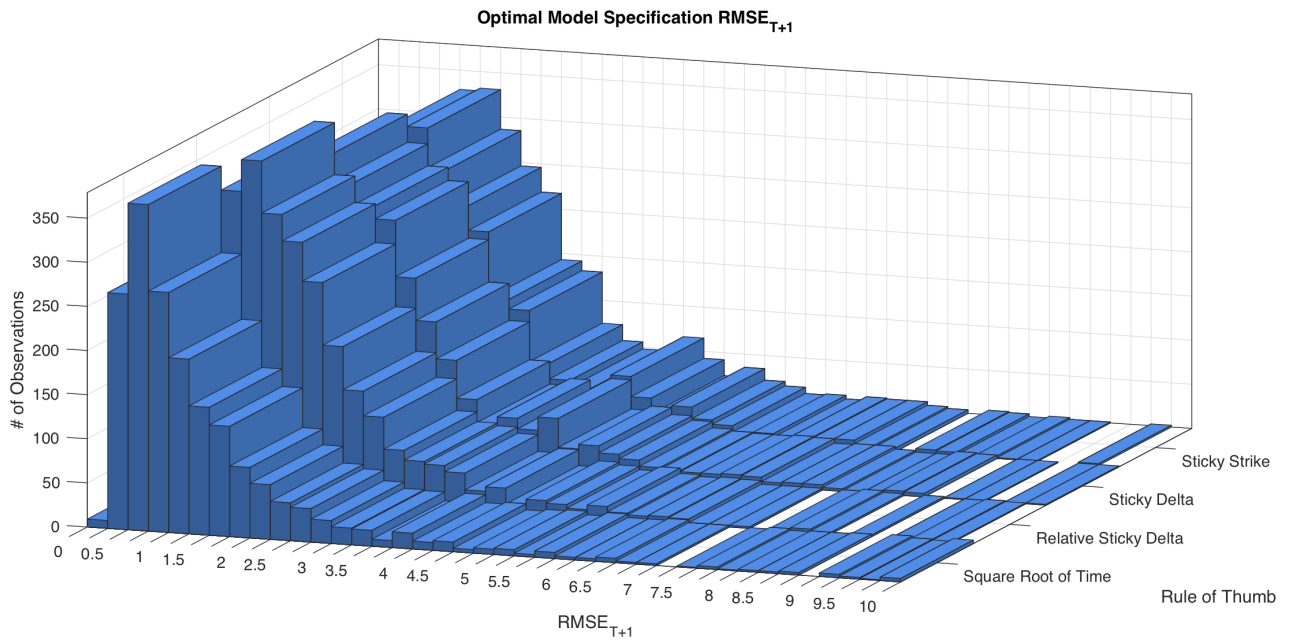


Figure 7: Square Root of Time

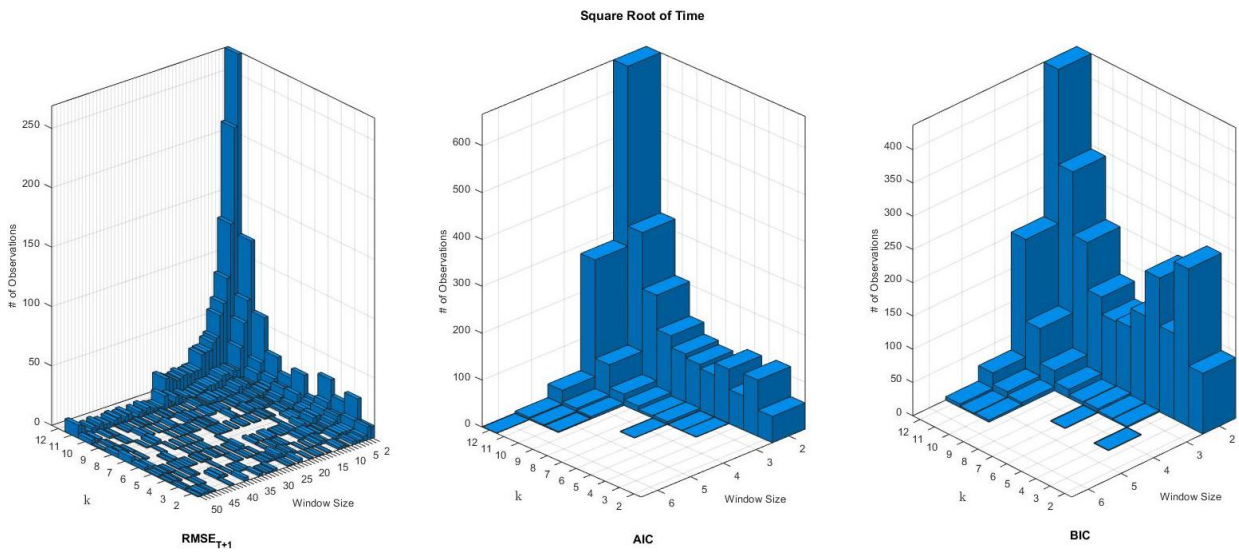


Figure 8: Significance Level for Square Root of Time

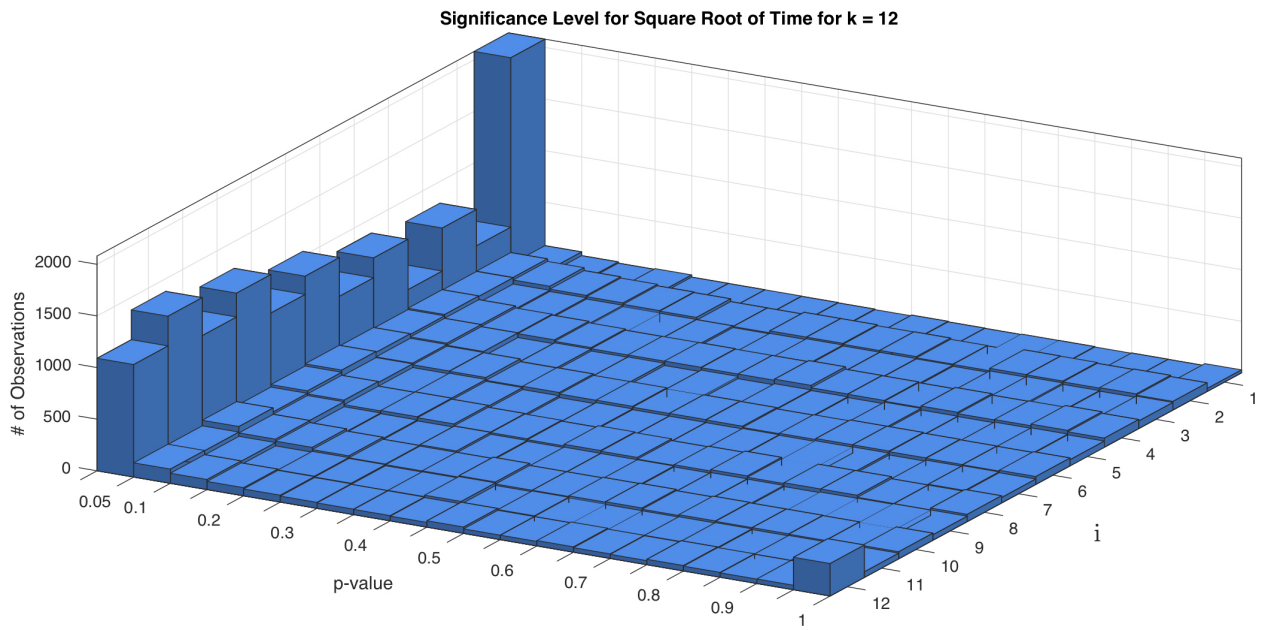


Figure 9: Surface Prediction Comparison

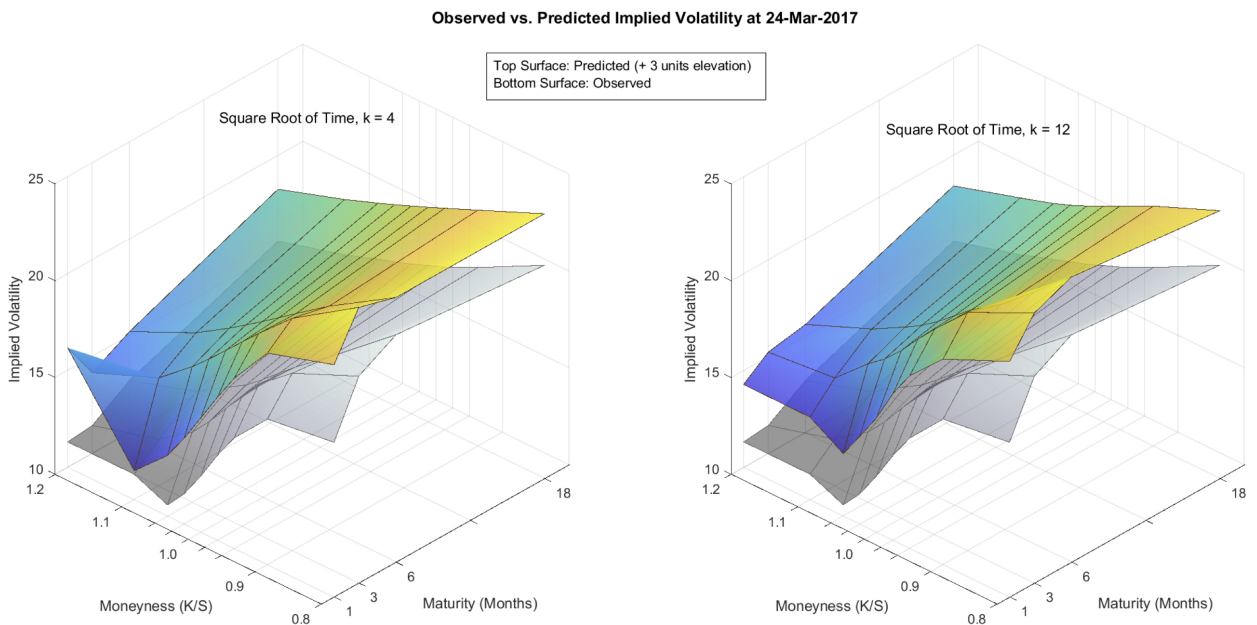
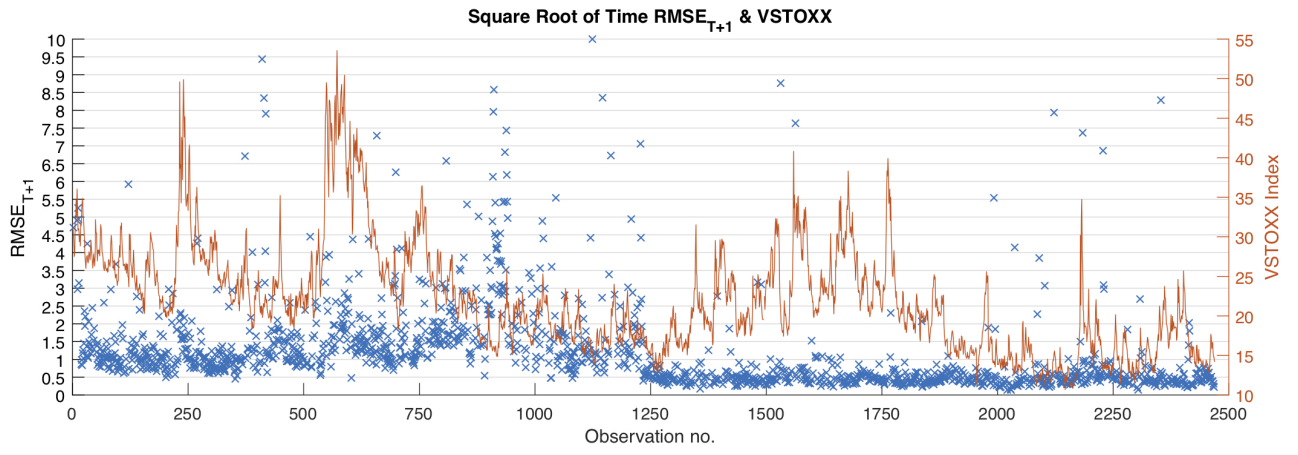


Figure 10: Square Root of Time $RMSE_{T+1}$ and VSTOXX



Appendix: Tables

Table 1: Results Algorithm A

OPTIMAL OVERALL:				
Model Criterion: Avg.	Rule of Thumb	m/k	n	Window Size
RMSE_{T+1}: 1.2874	Square Root of Time	8	-	3
AIC: 234.1125	Square Root of Time	8	-	2
BIC: 251.7673	Square Root of Time	8	-	2
OPTIMAL OTHER RULES OF THUMB:				
RMSE_{T+1}: 2.1376	Sticky Strike	3	8	2
RMSE_{T+1}: 2.1344	Sticky Delta	8	3	2
RMSE_{T+1}: 1.8435	Relative Sticky Delta	8	3	2

Table 2: Results Algorithm B

Obs. #	Optimal Overall				Optimal Sticky Strike				Optimal Square Root of Time					
	RMSE _{T+1}		AIC		BIC		RoT		RMSE _{T+1}		AIC		BIC	
	RoT*	m/k	n	W.S.	RoT	m/k	n	W.S.	m	n	W.S.	m	n	W.S.
42	3.2323	4	-	7	281.5476	5	2	2	3.5490	292.8319	3	2	2	300.1370
	SqrT				SD				3	7	2	3	2	2
43	4.9940	2	-	2	365.9678	8	3	2	5.6057	375.1671	3	3	2	387.6532
	SqrT				RSD				2	2	3	3	3	2
44	4.0663	4	-	2	453.1002	7	2	2	4.5750	467.2612	6	2	2	477.3099
	SqrT				SD				6	2	2	3	3	2
45	3.1629	6	3	2	486.6947	8	2	2	3.2747	499.7333	6	2	2	514.0147
	RSD				RSD				4.7690	493.5985	2	2	5	2
46	2.7295	2	-	4	463.7024	8	3	2	2.2	493.5985	2	2	5	493.2385
	SqrT				RSD				2	2	5	2	3	2
...
2506	0.3499	6	-	2	120.1000	5	-	2	1.1114	271.2845	3	5	8	131.4834
	SqrT				SqrT				1.0527	268.5039	3	2	12	116.4520
2507	0.3895	6	-	3	100.5153	7	-	2	3	2	4	2	2	100.5153
	SqrT				SqrT				1.0882	262.3058	2	2	4	100.5988
2508	0.3094	7	-	2	89.2155	5	-	2	2	2	4	7	2	89.2155
	SqrT				SqrT				1.1927	257.0045	2	2	5	90.1293
2509	0.3461	7	-	3	78.7460	5	-	2	1.0343	259.7048	2	2	6	106.5096
	SqrT				SqrT				1.0343	259.7048	2	2	6	106.5096
2510	0.3221	7	-	3	90.5730	7	-	2	2	2	6	3	3	2
	SqrT				SqrT				2	2	6	3	3	2

*RoT: Rule of Thumb, SD: Sticky Delta, RSD: Relative Sticky Delta, SqrT: Square Root of Time, W.S.: Window Size.

†Not shown in table: Optimal Sticky Delta, Optimal Relative Sticky Delta; Observations 47-2505.

Table 3: Model Specifications for Results in Figure 6

Rule of Thumb	m/k	n	Window Size
Sticky Strike	3	2	2
Sticky Delta	8	3	2
Relative Sticky Delta	8	3	2
Square Root of Time	8	–	2