STOCKHOLM SCHOOL OF ECONOMICS

BACHELOR THESIS IN FINANCE

Predicting the smile

A study on the properties of the volatility surface of S&P 500 Composite Index options

---

Viktor Berggren[1] Johan Blomkvist[2]

Abstract
The classic Black-Scholes model on option pricing from 1973 has been a widely debated and scrutinized theory over the past decades. Despite its proved limitations and simplifications, it remains the most used pricing model of publicly traded options today. One implication of the model is the implied volatility surface, an empirical anomaly which has emerged over the past 25 years from the volatility of the underlying instrument. In this thesis, our goal is to create a parameterization of the daily volatility smiles of the S&P 500 Index option, and study the time-series properties of these parameters. We do this by testing the explanatory power of exogenous variables, and by using different lag models to predict the shape of the smile. Our results indicate strong correlations with our external factors, but a non-conclusive predictive power of the inherent parameters. The forecasting properties of the smile seem to remain a mystery.

Keywords: Implied volatility, Volatility smiles, Volatility surface, Options

Tutor: Christian Huse

Date: June 11th 2010

Discussants: Richard du Rietz, Philip Lindqvist

Acknowledgements
We would like to give our sincere thanks to our tutor Christian Huse, Assistant Professor of the Department of Finance at Stockholm School of Economics, for his guidance and valuable inputs throughout the writing of this thesis.

---

[1] 20703@student.hhs.se
[2] 21251@student.hhs.se

1. Introduction

Is it possible to predict the shape of the volatility smile, given its shape today? The implications of the classic Black-Scholes formula have been a widely debated field of study since its inception in 1973. According to their option pricing formula, or rather assumed in their model derivation, the implied volatility of options with equal expiration dates but different strike prices should be the same. However, as observed on actual traded option prices, the implied volatilities show a skew when plotted, hence creating a "smiling" surface over time. This surface holds a large amount of information for investors, and is often used as a valuation tool rather than using the price of the option. In this thesis we will first create a parameterization of the smile on a daily basis, and secondly study which inherent and external factors can aid in predicting the shape of the smile using exogenous uni- and multivariate regressions, and also using AR- and VAR-lag models.

1.1 Purpose of study

Our primary goal with this thesis is to understand what actually drives the volatility surface, and to explain the intrinsic properties of our chosen parameters in predicting the shape of the smile. Conventional studies in the area have almost exclusively chosen one of two approaches; either modelling the cross-section of volatilities, i.e. looking at all option maturities and strike prices on a given day and hence neglecting the time-series dimension (Alentorn, 2004), or deciding on a certain moneyness and/or time-to-expiration and looking at the time-series progression of that single option (Peña, Rubio, & Serna, 1998). To our knowledge only one previous study has been done on the two-step regression approach utilized in this thesis (Gonçalves & Guidolin, 2005), and they used a different method in their first-step parameterization.

Rebonato (1999) states that the implied volatility of options is the "*wrong number to plug into the wrong equation to get the right price"*, which summarizes what makes the research regarding the matter so interesting. As the volatility surface is a de facto trading tool, on which investors and traders base multi-billion dollar decisions (CBOE, 2009), it is somewhat startling that the specifics regarding its progression over time is still under scrutiny. With this thesis we hope to bring some light on the matter, while also looking at the extent at which external factors, closely linked to option pricing, can affect the shape of the volatility surface.

1.2 Contribution

Our results will be able to contribute to a more detailed understanding of the structural change of the implied volatility surface and its evolution over time. A deeper comprehension regarding the exogenous factors which affect the smile can i) help in creating more precise pricing models for options, and also ii) come to use as proxies for investors and traders to judge in which ways risk sentiments and prices are moving. Our hope is that these results can add to current research regarding option markets in general, and highly liquid markets such as the S&P500 in particular. With the addition of some further research on lag power of our exogenous variables, we believe that it would be possible to create a parsimonious model to forecast the progression of the volatility surface.

1.3 Outline

An introduction to our study has been given above and the remainder of our thesis will be organised as follows. In Section 2 we will give a more thorough background on the Black-Scholes model and implied volatility, as well as a summary of previous research on volatility surfaces. Section 3 details the data, computed measures and limitations which we have used throughout this thesis, as well as our choice of moneyness. In Section 4 we describe our parameterization choice for the first step of our regression, the manipulations done to our data, and the final regressions testing our parameters. In Section 5 the regression results of our study will be reviewed and analysed. The main conclusions of our thesis will then be summarized in Section 6 and lastly Section 7 will go further into additional research that can be done related to the findings of our thesis. All tables and graphs will be available in the appendix at the end of the thesis, following the bibliography.

2. Previous literature

A review of previous literature and a brief theoretical overview will be presented below. In section 2.1 we begin with a review of the Black-Scholes formula. This section leads us into the discussion about implied volatility and its informational content, which is presented in section 2.2. In section 2.3 we go further into earlier work involving the characteristics of the implied volatility surface. Finally we will discuss the implications of our parameterization choice.

2.1 Black- Scholes

Fischer Black and Myron Scholes released their by-now classic paper on option pricing in 1973, and revolutionized the pricing and trade of derivatives through their intuitive pricing formula (Black & Scholes, 1973). Despite its initially theoretical application and its proven limitations when using it on live options, it remains one of the most used option pricing techniques today, and its inventors were awarded the 1993 Nobel Price in Economics as a homage to their work.

The Black-Scholes formula for a European call option is:

$$C(S_t, K, \tau, \sigma) = S_t e^{-q\tau} N(d_1) - Ke^{-r\tau} N(d_2)$$

where d_1 and d_2 are defined as:

$$d_1 = \frac{\ln(\mathrm{S_t}/\mathrm{K}) + \tau(r - q + \frac{\sigma^2}{2})}{\sigma\sqrt{\tau}} \qquad d_2 = d_1 - \sigma\sqrt{\tau}$$

S_t is the spot price of the underlying[3], K is the strike price, τ is the time-to-maturity, r is the risk-free interest rate, q is the dividend yield, and N(d) is the cumulative normal density function[4] where the upper integral is denoted d. The intuition behind d_1 and d_2 is that they provide the risk-neutral probability measures[5] of an option's expiry, much like a moneyness term.

The model assumes that the underlying follows a geometric Brownian motion[6] and that all options on the same underlying regardless of time-to-maturity[7] and strike price have the same implied volatility. Empirical evidence has however shown that this assumption does not correspond with reality (Hull, 2008). As noted by earlier authors, it is the monotonicity[8] of the volatility parameter in the Black-Scholes formula that makes it possible to calculate the implied volatility. As the inverse of the formula has no closed-form solution, one cannot simply re-solve the formula to obtain the implied volatility, given an option price. However, as the formula predicts a flat implied volatility surface, one can use the Newton-Raphson algorithm to invert the Black-Scholes formula. This method has been widely accepted by researchers over the years, and uses the market price of the call option to compute the implied volatility.

---

[3] The S&P500 index option in our study

[4] $N(u) = (2\pi)^{-1/2} \int_{-\infty}^{u} \exp\left(-\frac{z^2}{2}\right) dz$

[5] Under the martingale probability distribution

[6] A continuous stochastic process, such as a Wiener process

[7] Henceforth referred to as TTM

[8] A function which preserves its given order

2.2 Implied volatility

Implied volatility can be described as the market's expected future volatility, and is linked to an option's market price. The measure should not be mistaken for historical volatility, given that historical volatility is linked to the earlier performance of the underlying, rather than the option price. Both measures have been widely studied over the years, and both have received credit for being the most accurate forecast of future volatility.

Whether or not implied volatility is an adequate measure of future volatility has been subject to much research. Canina & Figlewski (1993) found that implied volatility practically lacked any correlation with future volatility, when studying 17000 OEX call options on the S&P100 index. More recent work by Christensen & Prabhala (1998) and Fleming (1998) however find implied volatility to be a superior measure compared to historical volatility when it comes to forecasting future volatility, in their studies of S&P100 index options. Christensen & Prabhala use longer time series and non-overlapping data, and is probably the most comprehensive study on the subject to date.

2.3 The implied volatility surface

It is stated that the implied volatility of a call option depends on both its time-to-maturity T and its strike price K (Derman & Kani, 1998). The function for the implied volatility surface at date *t* visualizes this relationship by the following function:

$$\sigma_t^{BS}: (K,T) \rightarrow \sigma_t^{BS}(K,T)$$

Previous work by Rubinstein (1994) and Derman & Kani (1998) have shown that the volatility surface is a fluctuating two-dimensional surface forming a smile, a skew or a term structure, hence contradicting the Black-Scholes model. The reasoning behind these patterns is that the implied volatility of deep in-the-money and out-of-the-money options is higher compared to the implied volatility of at-the-money options. When it comes to studying the dynamics of the implied volatility surface, notable work has been done by Cont & Da Fonseca (2002), where they study the dynamics of implied volatility surfaces for S&P500 and FTSE index options. They conclude that the volatility surface is a fluctuating surface and that the shape depends only on a few specific factors which correlate with the underlying. They state that the implied volatility is positively autocorrelated, meaning that there is a positive correlation between the values of implied volatility over time. They also find that the surface shows signs of mean reversion, indicating that the implied volatility will return to its mean over time. No perfect correlation is found between movements of the implied volatility and the underlying.

Another acknowledged study on the dynamics of the volatility surface is Skiadopoulos, Hodges & Clewlow (1999), where they perform a Principal Components Analysis (PCA)[9] on S&P500 index options to observe how the implied volatility surface moves over time. They find two significant features that affect the implied volatility surface; a parallel shift and a Z-shaped slope structure. They believe that their findings could give an enhanced understanding of the dynamics of option prices, and that more detailed models of the implied volatility surface could be constructed as a result of their studies.

Further, a similar study which also examines the movements of the implied volatility surface over time is done by Fengler, Härdle & Villa (2003). They state that too much information is lost when applying the PCA, and thus they present their extended model called the Common Principal Component Analysis (CPCA), which they apply on DAX index options. With this structure they were able to model the implied volatility surface more accurately and predict the implied volatility surface simultaneously for different maturities and moneyness. They also identify a similar parallel shift and Z-shaped slope as the above mentioned study did. Cont, Da Fonseca & Durrleman (2002) form a model with the implied volatility surface as a stochastic variable, and use this model to increase the precision of the "sticky moneyness"[10] rule. The most outstanding characteristic of their model was that smiles and term structures of implied volatility could be correctly utilized, and hence that they were able to arrive at a good prediction of the progression of the implied volatility over time.

The insufficiencies of the Black-Scholes model initiated the development of deterministic volatility function (DVF) option pricing models by Derman & Kani (1998), Dupire (1994) and Rubinstein (1994). The framework of these models relies on the assumption that the *"...asset return volatility is a deterministic function of asset price and time...*" (Dumas, Fleming, & Whaley, 1998). The previously mentioned authors use S&P500 index options to examine the DVF models, and they are actually unable to show that the DVF models are better than the Black-Scholes model.

[9] Principal Components Analysis is a statistical method that identifies systematic behavior in the data by reducing the original dataset to a smaller set of uncorrelated variables, which decreases the dimensionality of the dataset (Dunteman, 1989).
[10] Sticky moneyness means that the volatility of an option solely depends on its moneyness

A study in line with ours is the recent work by Gonçalves & Guidolin (2005). In a two-stage approach they first model the implied volatility surface along cross-sections of TTM and moneyness, and in a second stage they use a vector autoregressive model (VAR) to study the dynamics of the implied volatility surface from stage one. They are able to effectively model the implied volatility surface and find that there is a statistical predictability of the implied volatility surface's movements over time.

2.4 Parameterization

As the volatility surface is a 3-dimensional entity, one needs to proceed in several steps when modeling its movement. An accurate but tedious method is to follow each option strike price throughout its life-time and create a time-series containing all options in a given option series. This method is limited due to the very large data sample one must use, hence making it difficult to find any significance when using exogenous variables in the regression. The solution to this problem is to use a parameterization technique, where one approximates the cross-section of the volatility smile, i.e. a time slice, by a pre-determined equation. A statistics program is used to fit the parameters to the cross-section through a conventional least squares regression technique, and the result is a much smaller sample of parameters on which one can conduct further tests. There are many different parameterization methods used in prior research, but we will look at three separate methods here; Nelson-Siegel, Dumas et al, and Gatheral's SVI technique.

Nelson & Siegel (1987), and later Diebold & Li (2005), devised a parsimonious parameterization model for the yield curve, i.e. the interest rate over time seen on bonds. The model took the form:

$$y(m) = \beta_0 + \beta_1 \left(\frac{1 - e^{-m/\tau}}{m/\tau} \right) + \beta_2 \left(\frac{1 - e^{-m/\tau}}{m/\tau} - e^{-m/\tau} \right) + \varepsilon$$

where m was the moneyness term (or maturity in the case of yield curves), the β-terms are the model parameters, and τ is a time decay factor, which dictates the time-varying shape of the parameterization. The strength of this method is its convergence criteria: as τ is a pre-determined variable which is not estimated by the regression, the model takes on a quite simple form which simplifies the estimation procedure. However, as the model is optimized for the yield curve with its humped and S-shaped surface, it is not optimal for the time-varying and 3-dimensional shape of the volatility surface.

Dumas et al (1998) created a model specifically for the volatility smile, which is an evolution of the regular 2-order polynomial regression model. The specification is as follows:

$$\sigma(M,T) = \beta_0 + \beta_1 M + \beta_2 M^2 + \beta_3 T + \beta_4 TM + \varepsilon$$

where M and T are moneyness and TTM and the β-terms are the model parameters. In this regression, M will be varying over the volatility cross-section, while TTM is a scalar for each time slice. The strength of this model is that it explains the 3-dimensional properties of the volatility surface through its inclusion of a second dependant variable. Several studies have used the Dumas parameterization, in both an un-modified original version (Alentorn, 2004) and a customized version (Gonçalves & Guidolin, 2005).

A third interesting parameterization technique is devised by Jim Gatheral in his 2004 presentation on implied volatility surfaces for volatility derivatives (Gatheral, A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives, 2004). The parameterization that he proposes, which he calls stochastic volatility inspired (SVI), is written as:

$$var(k) = \alpha + \beta\left(\rho(k-m) + \sqrt{(k-m)^2 + c^2}\right) + \varepsilon$$

The description and intuition behind the parameters are the following:

> α is the overall level of volatility, i.e. the volatility of the ATM-option for that given cross-section
>
> β gives the spread between the asymptotes, i.e. gradient of the change in implied volatility
>
> ρ shows the orientation of the graph
>
> m translates the graph to match the α-term
>
> c (b4) determines the vertex of the graph, i.e. the actual shape and smoothness of the smile.

The strength of the SVI model is that it is specifically devised for the volatility surface, and that its parameters are parsimonious and quite intuitive. Also, it has a form which is simple enough to make the estimation procedure smooth (Gatheral, 2006).

3. Data

In this section we describe the data which we use throughout this thesis. Section 3.1 details the data sources for the option data, and which definitions are used for this dataset. Section 3.2 describes the definitions which we have used for moneyness, and in what context the different definitions have been utilized. Section 3.3 explains the manipulation and adjustments made to the data before the first-stage regression. Finally, Section 3.4 details the exogenous data choice which we use in our third-stage regression. All data used throughout this thesis is extracted from Thomson DataStream Advance 4,0.

3.1 Option and market characteristics

We have decided to use SPX Index options in this essay. The option series is based on the S&P Composite 500 Index as the underlying, and is noted on the Chicago Board of Exchange[11]. The S&P 500 Index[12] is the most benchmarked index in the world (Dash & Liu, 2008), which translate into an option which is very liquid (CBOE, 2009). The liquidity of the option is very important when modeling the volatility surface, as the price and implied volatility of un-traded options tend to become stochastic and speculative. The SPX options are released with maturities ending the third Friday of each month, and exist in increments of 5-50 points, depending on the moneyness of the option. The options can be released with widely different TTM; from a few weeks to several years.

For this thesis, we have examined two different option series; 1004C[13][14] and 0804C[15]. The focus of the thesis lies on the 1004C option series, and we use the 0804C series mostly as a control when modeling. When using highly liquid options like the SPX we can assume that the Put-Call Parity generally holds, which means that the implied volatility of a Call and a Put with equal maturities must have the same ATM implied volatility (Hull, 2008). As long as this equality holds, it is sufficient to use either the Call or the Put to model the smile, and we have chosen the Call. Taking into account the exogenous variables used to control for external data, and looking at other studies in the area, we feel it is safe to base our study on only two option series.

[11] Henceforth named CBOE
[12] Henceforth named S&PCOMP
[13] This translates into the Call option series maturing in April 2010
[14] 090324-100317, 152 options
[15] 071224-080416, 113 options

3.2 Option data definitions

For each option series, date and strike price, we have extracted current option price, daily volume, and daily implied volatility. The following definitions have been used (Thomson Reuters, 2010):

Price (Mnemonic MP): Last traded price provided last trade is within bid/ask range at the end of the day. As the CBOE does not display settlement prices for each given trade, this is the industry standard in reporting option prices. We will use the pricing of the option as an exogenous variable in our third-step regression.

Implied Volatility (Mnemonic IV): As is described in Section 2.2, implied volatility cannot be found using a closed-root solution of the Black-Scholes formula. The data taken from Datastream is in turn calculated using MB Risk Management's Univopt tool, which is an industrial and acknowledged pricing and risk system (MB Risk Management, 2010). The system uses the Newton-Raphson algorithm mentioned in Section 2.2 to find the implied volatility given an option price.

Volume (Mnemonic VM): The volume is calculated as the total cumulative trading volume for all individual option series. The volume is used both as an exogenous variable in our third-step regression, and as a controlling factor when removing options which are deemed to be too speculative.

These three variables are the only option-related data used in our methods. They are however in some cases manipulated, sometimes in conjunction with other data mentioned below.

3.3 Choice of moneyness

The volatility surface has three dimensions; implied volatility, TTM and moneyness. Implied volatility is noted above, whereas TTM and moneyness are subject to calculations.

TTM is the time until the option matures, measured in years. For this thesis we only look at options with less than one year until maturity, and hence the value will be less than one. Algebraically, we use:

$$TTM_t = \frac{(T - \tau)}{365}$$

where τ is the current date and T is the date of maturity. The difference in days is then divided by 365, and the TTM is received for that particular day.

Moneyness is a little more sophisticated, as it can be described using several different methods. What we are trying to depict is the amount a certain option is in or out of the money, which is a concept which one can label in several different ways. In its simplest form, moneyness describes the relation between an option's strike price and the spot price of the underlying, in absolute terms. However, as noted in a paper by Dumas et al (1998), having a strictly linear moneyness term creates discrepancies in the model when option maturities grow shorter. Hence, other types of models for the moneyness have been created.

Below we describe the intuition between the main methods, and which ones we use in this essay.

Black-Scholes standard deviation: As described in Section 2.1, the Black-Scholes model is the basis for the volatility surface as depicted in this essay. Defining the moneyness as the average of the two d-terms from the Black-Scholes formula, we get the following expression for the moneyness (McMillan, 1986):

$$m = \frac{\ln\left(\frac{S}{K}\right) + rT}{\sigma\sqrt{T}} = \frac{d_1 + d_2}{2}$$

In the equation, *S* is the spot price of the underlying, *K* is the strike price of the option, *r* is the risk-free rate, *T* is the time-to-market, and σ is the implied volatility. This measure calculates the moneyness in normal standard deviations, i.e. a moneyness of 0 means that the option is at-the-money and has a 50% chance of ending up in-the-money, and a moneyness of 1 equals an 84% chance of ending up in-the-money. In a sense, this is the most accurate measure from a Black-Scholes point of view, as it is in fact derived from the equation. However, as it subsequently uses the implied volatility as a factor in the calculations, and we are using the implied volatility rather than price as the topic in our essay, this method is not optimal.

Gross & Waltners: In their paper *Put Volatility Smile and Risk Aversion*, Gross and Waltners (1995) define moneyness according to the following:

$$m = \frac{\log\left(\frac{F}{K}\right)}{\sqrt{T}}$$

Here, F equals the forward price of the underlying, K is the strike price of the option, and T is the time-to-maturity. F is calculated according to the following[16]:

$$F = S_t e^{T(r-q)}$$

The variable S is the spot price of the underlying, T is the time-to-maturity, r is the yearly risk-free rate, and q is the annual dividend yield on the underlying. The strength of this method is that it gives a very intuitive depiction of moneyness, showing the relation between forward price and spot price, and accounting for time decay. We use this method when looking at which maturities to use in our first-step regression, further specified in Section 3.4.

Gatheral's k: In Jim Gatheral's SVI parameterization, which was touched on in Section 2.4, the following moneyness term is used:

$$k = \log\left(\frac{K}{F}\right)$$

In this case, k is the moneyness term, K is the strike price of the option, and F is the forward price of the underlying. Compared to the method proposed by Gross & Waltners, this will in fact show the inverse of moneyness as K is the numerator, which will give a mirrored version of the smile when modeled. We use the k when using the SVI parameterization to model the smile in our first-step regression.

3.4 Exogenous data choice

Regarding our choice of exogenous data, we have extracted data partly to calculate variables for our regressions, and partly to use as control variables in our third-step regression. Again, all the data is extracted from Datastream, and we look at S&PCOMP for the following values:

Price (Mnemonic PI): Datastream calculates its own price level for the index, which is a weighted average of the different stock in the different indices.

[16] The equation for the future price is used for all future price calculations throughout this thesis

Dividend Yield (Mnemonic DY): The dividend yield for an index is the total dividend amount for that specific index, expressed as a percentage of the total market value of the index.

Apart from the S&P values, we also look at the risk-free rate and the VIX. As a proxy for the risk-free rate, we have used the US Federal Target Rate (Mnemonic USFDTRG), which for our purposes is a very close approximation of the theoretical risk-free rate. One issue with using the target rate as a proxy is that it changes with quite large increments, as it is the actual rate which the Fed chooses. However, for the scope of this essay, the rate has been constant for both datasets used, and as such the choice fits our purposes.

CBOE issues a volatility index called the VIX (Mnemonic CVX), which is a measure of the market's 30-day volatility expectations, as conveyed by the SPX option series. We will use the index as a proxy for the market's general level of volatility in our third-step regression.

4. Methodology

In this section we introduce the method used in this essay. We will describe both the practical application of our models, and the econometric intuition behind. Section 4.1 describes the manipulation to our data, which was done before any actual parameterization was conducted on the values. Section 4.2 describes the first step of our regression, which is the parameterization of the volatility surface. Section 4.3 describes the descriptive summaries and visualizations done on the result of the first-step regression. Section 4.4 details the regressions testing the explanatory power of our exogenous values, and the correlation between the parameters.

4.1 Option data manipulation

For the two sets of options series mentioned above, TTM for the first observation varies between approximately three months to one year. As the range of options for each given day and maturity is very large, with over 150 strike prices present, the volumes traded in a substantial part of the data is close or equal to zero for many option strikes. As mentioned in Section 3.1, option prices and implied volatilities for untraded data becomes very speculative, and is characterized rather by stochastic features than values reach by market equilibrium.

When modeling the smile surface, we have removed options where the moneyness is too far away from the spot price. Cont & DaFonseca (2002) use a moneyness limit of +- 0.5 in their paper on the dynamics of volatility surfaces, and remove options outside of this boundary. However, as we use another measure of moneyness, we have in part tried to convert their limit to our moneyness measure, but also looked at where the volume traded starts to decline. From this, we decided on a boundary of ±2, according to the *Gross & Waltners* definition of moneyness.

Another consideration we have had when preparing the data for the first-step regression is the time component. Compared to several other studies in which all traded options on a given day are studied, regardless of their expiry date, we only look at options with equal TTM. For a given maturity, CBOE only releases a few options strike prices when the option maturity is first introduced on the market. When the TTM comes closer to date of expiry, more options are added to the option series, which makes the incremental gaps smaller. While testing our first-step regression, we realized that it was very difficult to reach convergence when TTM was large, as there were too few options data points to interpolate. Through testing, we found that approximately 70 points were needed to reach convergence, i.e. 70 options with different strike prices on a given day.

4.2 First step – Parameterization

As our ultimate goal is to find a way to predict the progression of the volatility surface, the first step is to find a model to fit the smile on a given day, henceforth denoted as a time slice. As described in Section 2.4, there are several parameterization techniques used in previous research. We have decided to use Gatheral's SVI technique, as it i) gives a parsimonious depiction of the smile with intuitive parameters, ii) uses only moneyness as a dependant variable[17], and iii) has a simple for which enables simpler estimations (Gatheral, 2006). The model to be fitted is the following:

$$\sigma = \alpha + \beta\left(\rho(k - m) + \sqrt{(k - m)^2 + c^2}\right) + \varepsilon$$

For each day in our sample, a summarized file was created containing the implied volatility for each day and strike price, and the corresponding moneyness k for each term, given the manipulation stated in Section 4.1. The file was imported into STATA and each day was

[17] Dumas et al use a model where both moneyness and TTM are used as dependant variables. This gives a very accurate result of the smile, but since we use TTM as an exogenous control variable, this approach would not fit our method

estimated using a non-linear least squares technique, according to the following (Alentorn, 2004):

$$\min SSE(\beta) = \sum_{i=1}^{m} (y_i - prd_i(w))^2$$

Here, m is the amount equations in the series; y is the observed implied volatility; prd is the actual specified equation; w depicts the moneyness term; and β is the given parameter vector. The algorithm for the function minimizes the SSE[18] between observation y_i and the model's prediction for the observation using prd_i. STATA performs iterations until the point where the convergence criterion is either satisfied, or the maximum numbers of iterations specified are exceeded and no convergence is met. From this we obtain an R^2-value to examine the goodness-of-fit, which measures how successful the model is at explaining y. The DO-file created for this loop is attached in Table 1.

4.3 Second step – Descriptive summary and visualization

For each day in our sample, we collect our $\beta_t = (\alpha_t, \beta_t, \rho_t, m_t, c_t)$ and summarize and sort the vectors as to create a time-series for our third-step regression. Before proceeding to our final regressions, we perform visualizations of the results in order to receive some understanding regarding the shape of the parameters over time, and if any qualitative conclusions can be drawn at this stage.

4.4 Third step – Endogenous and exogenous AR and VAR modeling

From the parameters obtained in the first-step regression, we merge the values with the exogenous variables corresponding to the same dates to create a time-series. For each option series, we then have the following variables: α, β, ρ, m, c, date, TTM, spot price of S&PCOMP, future price of S&PCOMP, VIX, and volume traded. What we are trying to do in this step is to understand what actually influences the progression of the respective parameters. We will do this by looking at their correlation with the exogenous factors, the lags on each respective parameter, and finally the lags of all parameters.

The first regressions conducted are regular univariate OLS regressions according to the following:

[18] Sum of Square Errors

$$P_t = \alpha + \beta TTM_t + \varepsilon$$

$$P_t = \alpha + \beta \log\left(Fprice_t\right) + \varepsilon$$

$$P_t = \alpha + \beta \log\left(Volume_t\right) + \varepsilon$$

$$P_t = \alpha + \beta \log\left(Vix_t\right) + \varepsilon$$

Here, P_t is a vector containing all parameters, i.e. $P_t: (\alpha_t, \beta_t, \rho_t, m_t, c_t)$, as all of the regressions are run separately for each parameter. These regressions are conducted to isolate each exogenous factor, to see if there is a correlation between the two in its simplest form. The logarithmic form is used to give the equations as simple form as possible, and as we're only interested in the actual relation between the two variables it is a valid manipulation. We apply robust standard errors to correct for heteroskedasticity throughout all these regressions. Worth to notice here is that the α and β received in these regressions are not the same as the parameters in the regression, but simple results from the regressions.

The next regression we conduct is a multivariate version of the above, where we try to see whether the exogenous variables are successful together in explaining the parameters movement over time. The formula is as follows:

$$P_t = \alpha + \beta TTM_t + \beta \log(Fprice_t) + \beta \log(Volume_t) + \beta \log\left(Vix_t\right) + \varepsilon$$

Until this point, our tests have aimed at describing the fit of our model and which external factors can help explain the movement of the parameters. What is more interesting for our study however is the ability to forecast the movements of the volatility surface over time. We would hence like to test the parameters intrinsic ability to influence its own movement, using an autoregressive model. The first step is to run a #-lag test on each parameter, where # is a number of lags determined by the AIC and BIC criterion. They are calculated as follows (Plasmans, 2006):

AIC (Akaike Information Criterion)

$$AIC(p,q) = \ln(\sigma_\varepsilon^2) + \frac{2(p+q)}{T}$$

BIC (Schwartz (Bayesian) Information Criterion)

$$BIC(p,q) = \ln(\sigma_\varepsilon^2) + \frac{(p+q)}{T}\ln(T)$$

where σ_ε^2 is the estimated error variance, T is the sample size, and $(p+q)$ is the total number of estimated parameters. The difference between the two criteria is that AIC tends to suggest models which are over-parameterized, while BIC penalizes this to a greater extent. In STATA we conduct a lag-order selection statistics test to see which of the criteria returns the lowest value. With the given number of suggested lags returned by the test, we then run a AR(#)-regression on each of the parameters, according to the following equation:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

where X_t is the tested parameter, φ_i are the parameters of the model, c is a constant, p is the number of lags, and ε_t is the error term. Hence, according to the criterion given above, we test each parameter on a certain number of lags and look at the validity of these betas.

The last regression we run is a vectorized autoregressive model[19]. This test is done to capture the evolution and interdependencies between our several time series parameters, hence combining the AR-model with the regular OLS-regression (Plasmans, 2006). A VAR-model of order p can be expressed as:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon$$

where c is a *k*1 vector,* the A_p's are *k*k* matrices (for every *i =1,…,p*) , and ε is a *k*1* vector of error terms[20]. For each of our five parameters, we receive a two-lag result for all other parameters, showing the forecasting properties of the parameters between each other. This is the last regression performed on our parameters.

[19] Henceforth called a VAR-model

[20] For a general matrix notation of a *p*-order VAR matrix, see http://en.wikipedia.org/wiki/General_Matrix_notation_of_a_VAR(p)

5. Results

In this section we present the findings of our research and some brief comments on our results. Section 5.1 shortly describes the results from the parameterization, and touches on the issues found in the regressions. In section 5.2 we give interpretations and visual depictions of our parameter results. Finally, section 5.3 details the findings of our exogenous and AR-regressions. We have attached a 3-dimensional graph showing the actual progression of the 1004C volatility surface in Table 18.

5.1 First step results

The results of the parameterizations are found in Table 2. For the option series on which we focus, 1003C, we have 68 continuous observations, spanning between 2009-12-11 and 2010-03-16. As mentioned above, we have removed the last days of option trading, as the parameters show very random values which distorts our model. For our controlling option series, 0804C, we have a non-continuous sample of 39 observations spanning between 2008-02-04 and 2008-04-16. We experienced difficulties when conducting the parameterization on this series, probably explained by a deep in-the-money negative skew on the implied volatilities which distorted our model. Despite our data manipulation and moneyness restriction of ±2, this issue was very present for the 0804C option series.

When deciding on the validity of the parameters for each time slice, we have focused on the R^2-value. The key asset of the parameterization is to create such an accurate model that the parameters themselves can act as proxies for the volatility smile in the time series. If that property is lost, we will have great difficulties in getting any validity in the third-step regressions. The average R^2 and standard deviation of the R^2 for the two series are found in Table 3. Both series have an average R^2 of above 0.9, which must be considered as a good fit for our parameterization.

As will be detailed below, the parameterization gives some interesting time-series properties of our smile components. The cross-sections which we have parameterized are accurate according to the R^2-statistic, and at least for the 1004C series we have a continuous and sufficiently extensive time span on our observations. The evolution of the volatility surface for options which are far from expiry tends to be static and contain little information (Rubinstein, 1994), and hence we think that the time span used in our regressions is satisfactory for our purposes.

5.2 Second-step results

In the second step, we have examined the parameters from a qualitative perspective, to get a feel of their inherent properties and which tests would be appropriate to further understand the progression of the smile. From Table 2 one can get a generalized feel of the time series progression of the parameters, but naturally the easiest way to investigate their patterns is through graphical representations. We will describe the parameters one by one, discuss the way they move over time, and summarize our findings. Worth to note here is that we have graphically shown the parameter's progression over elapsed time rather than TTM, as it gives a better visual representation of their movement. The regressions in Section 5.3 are however done on TTM. We see no correlation between TTM and the R^2-values received from the regression. The DO-file for all these regressions is attached in Table 17.

5.2.1 The results of α, β and ρ

$\boldsymbol{\alpha}$ represents the approximated volatility of the ATM-option at a given point in time, and is presented in Table 5. For both option series, the graphs show a slightly downward but stochastic trend for the parameter. This would then indicate that the implied volatility of the ATM-option decreases when *i)* the option approaches expiry, *ii)* the value of the underlying displays volatility during the examined period (as it does in both these cases), *iii)* the exogenous variables described in Section 5.3 influences the volatility, or most probably *iv)* a combination of the above.

In Table 4 we have shown the implied volatility for the closest ATM-option for each day, and the corresponding b0. One can directly see that there is a discrepancy between the two series. Table 4 also shows the difference for each day, and the results look quite stochastic. Clearly there are other explanatory factors to describe the evolution of α, which we will further look into in Section 5.3. One possible explanation for the jumpy series could be the effect of volume, or rather the lack thereof. When the implied volatility of the observations for each cross-section is regressed, the options which are actually traded can vary heavily from day to day, hence creating a natural error source in the regression. This would imply that the actual trend of b0 is of clear interest, but that individual parameter observations can and will be random in nature.

$\boldsymbol{\beta}$ shows the spread between the left and right asymptotes of the smile, which translates into the gradient of change in implied volatility. The evolution over time of the parameter is shown in Table 6, and it has a logical but interesting shape for both option series. The parameter is clearly exponentially correlated with the time value, which is natural considering the time-varying shape of the volatility surface. This parameter is of special interest when the TTM becomes small, as the smile tends to be quite steady until this point is reached.

As can be seen clearly in Table 18, and in accordance with current empirical studies on the volatility smile, the surface sees an increased "smile" when the TTM decreases and approaches expiry. The β-parameter hence shows a very logical progression over time, as it exponentially increases until expiry. The time-series properties of this parameter will be very dependant on the choice of time span for the study, as the parameter will be static until – as we can see for our two option series – there is approximately one month of trading left.

$\boldsymbol{\rho}$ shows the orientation of the graph, i.e. the "tilt". We have shown its progression in Table 7 for both options, and the results are somewhat puzzling. Both graphs are stochastic in nature, but still show shape tendencies which seem quite unrelated. The graph for 1004C seems to have heteroskedastic properties, as the estimated parameters have an increasing volatility with time. This is pointed out with supporting lines for the 1004C graph. 0804C, on the other hand, looks like it has a negative parabolic shape, which is also shown with a trend line.

Regarding the properties of the ρ-parameter, we can conclude that it *i)* varies between our two option series, and *ii)* has a role in the model that primarily controls for errors in the β and m-parameter. As such, the information it carries regarding the future shape of the smile might rather be as a controlling factor for the other parameters, rather than being a key determinant in the shape. This will be further accentuated in Section 5.3, when looking at the regression results.

5.2.2 The results of m, c and our conclusion

$\boldsymbol{m}$ translates the graph, meaning that it shifts it horizontally. This parameter is linked to α as it aims to fit the implied volatility with the corresponding strike price or moneyness. We show the parameter regressed on time in Table 8, and both option series show a significantly negative linear relation between the parameter and the expired time. As the time series aims to describe the shape of the volatility surface over time, and our time dimension is "from left to right", this is a logical shape for the parameter. We would expect the parameter to have a similar linear shape regardless of the time period we test for, as the time decrease (TTM) is a linearly decreasing function.

$\boldsymbol{c}$ is the last parameter, and controls the degree of smoothness of the vertex. In a sense, this parameter determines the "level of smile" on the smile. In Table 9, we have visualized the parameter over the course of time, and the shape we see is rather scattered, however with the negative extreme points becoming smaller as time progresses. The value of this parameter will very much be a result of the other four parameters, as it only smoothens the shape determined. Both option series show similar graphics.

What is interesting with the c-parameter is that it controls the gradient of the change close to the ATM-option, i.e. it determines the amount of change in moneyness required for an option to reach the linear increase obtained in deep OTM/ITM-options. This is an interesting concept, as its prediction could help an investor choose which moneyness is appropriate for ex. a hedging strategy.

Conclusively, the result from the parameterizations shows significant and interesting results regarding the parameter's behavior over time. Generally the option series are analogous in their shapes, with the exception of ρ. This could however be an effect of the small sample for the 0804C series, or perhaps have further explanations in the exogenous variables described in Section 5.3.

5.3 Third step results

In our final third step, we have examined the effects which the parameters have on each other, the way the parameters evolve through lags, and the correlation with exogenous factors. Combined with the time-series properties which we observed in Section 5.2, the results in this section aim to help explain the variation and progression of the volatility surface. All regressions apply robust standard errors to correct for possible heteroskedasticity.

5.3.1 Univariate regressions

TTM: We initiated our tests with the univariate regressions between each parameter and the exogenous factors. The first variable which we test is TTM, i.e. Time-to-Market. This regression is in effect an algebraic interpretation of the graphs obtained in Section 5.2, except that we change the time measure from elapsed time to TTM, with the effect being that the regressions become "mirrored" compared to the graphs. The results for the five parameters are tabulated in Table 10, and as indicated in the section above, all values are significant except ρ . Looking at the R^2-values, α and c have low explanatory powers of 0.074 and 0.196 respectively, while β has a value of 0.507 and m a very high coefficient of 0.820, which is explained by its function to account for time.

All in all, TTM is a very significant measure in predicting the shape of the smile, which is expected. The evolution of the volatility surface over time is a well-studied and empirically proven phenomenon, and our results show that we have a parametric model which works well with the time-varying nature of the smile. Our conclusion from Section 5.2 regarding the ρ-parameter seems true in this regression as well, and the results further point towards the parameter's function as a fit control residual rather than a source of information. It is however worth noting that we do not equal its function with the ε-term; the error term accounts for the unexplained error in the model, whereas the ρ-parameter corrects the β- and m-parameters.

VIX: Our next univariate regression is testing the VIX index against our values. As the VIX is a proxy for volatility expectations we would expect to see some correlation in this regression, which also is the case. As can be seen in Table 11, α, β and m are all significant on the 5% level, although the regression indicates somewhat boosted betas due to the logarithmic proxy. The three significant parameters have R^2-values ranging between 0.180 and 0.304, so the explanatory power is moderate at best. Interesting to note is that m is significant, which is a parameter which we would expect to be affected by TTM only.

The VIX seems to explain the key parameters in fitting the smile, which is in accordance with its purpose as a volatility indicator. Our hypothesis regarding the insignificance of ρ is stated above, but the result of c is of further interest. As it controls for the gradient of the smile, one would assume that the market's perception of future volatility would influence the rate at which the implied volatility would rise. Since β is significant this is true for options further from the ATM-option, but for options with small values of k – i.e. options close in or out of the money – there is no significance.

Log(Volume): The third univariate regression tests the correlation with the logarithm of volume traded. The difference between completely untraded and traded options is significant, as we have discussed earlier, but the level of volume is another matter as the pricing mechanisms of the market should work regardless of the amount traded, as long as the option is actually traded. The results are shown in Table 12, and we do actually have significance on three of our five parameters. β, m and c all show p-values below 0.05, but all three have very small betas, and the

R^2-value does not exceed 10% for any of them. This small, but significant, result could be an effect of the volume working as a proxy for TTM, as the volume traded generally increases as the expiry date approaches.

Log(Future): The last univariate regression carried out aims to test the significance of the logarithmic future S&P 500 price level on the five parameters. As the volatility surface in our case is a predictive volatility measure on the option volatility based on the S&P 500, it is an interesting factor to study. The results are presented in Table 13, and we have significant results for α, β and m. All three parameters have R^2-values below 0.2, but the effect could become interesting in combination with the other above mentioned exogenous variables. Again, the large betas are most likely inflated by the logarithmic form of the S&P variable.

5.3.2 Multivariate regressions, AR- and VAR-models

Having conducted the four univariate regressions, we are interested in looking at the combined correlation of the four exogenous variables on the parameters. The results from this multivariate regression are found in Table 14, and show a large range in correlation. The correlation between ρ and the combined exogenous variables is non-existent, with a result showing no significant betas and an R^2-value of merely 0.058. α gives an R^2 of 0.285, with significant correlations with TTM and the future S&P price. β is quite well explained with an R^2-value of 0.656, and significant correlations with TTM, S&P and a weak significance with the VIX. m is very well explained by the exogenous variables, resulting in an R^2 of 0.918, although only TTM and the S&P are significant. Last, c gives an R^2 of 0.206, with a significant correlation with TTM and a weak significance with the volume. Worth noting when looking at the R^2 of multivariate regressions is that the value increases weakly with the number of regressors in the model, which could possibly somewhat distort the value in these regressions. However, as a combined goodness-of-fit measure with the p-values which we present, we believe that there is significance in looking at the value.

AR-model: Having looked at the exogenous variables' ability to fit the time-series of our parameters, we are now interested in the properties of the parameters themselves. If we are to find a model to forecast the shape of the smile, we need to find the correlation between past and current time slices. The method we have chosen is to first look at the lag properties of each isolated parameter using AR-models, and finally test the intra-dependencies of the parameters using a VAR-model.

Having conducted a lag-order significance test, we find the optimal number of lags for which we test the parameters. Conducting the AR-regression, the α-parameter indicated the largest explanatory lag-power, for which the first, third and fourth lag is significant. As presented in Table 15, the parameter receives an R^2 of 0.807, which implies a high level of predictive power through its lags. What makes this result extra interesting is that the α-parameter, given its intrinsic quality of predicting the ATM implied volatility, is a predictable value. β shows a 1-lag significance, while ρ shows no significance whatsoever. The m-parameter shows a 1-lag significance with a high explanatory power of 0.852, while the final c-parameter exhibits a low R^2 of 0.106, despite being significant in the first lag.

VAR-model: The results of the VAR regression, tabulated in Table 16, are somewhat puzzling. Despite the α-parameter showing high correlations with the exogenous variables and its own lags, there is no correlation at all with the two earlier lags of the other four parameters. The same holds for c, ρ and β, who all are completely uncorrelated with the lags of the other parameters. The only parameter which shows any correlation is m, which has a significant beta with the first lag of the ρ-parameter.

Overall, we find some interesting results from our regressions. Previous work by the likes of Gonçalves & Guidolin (2005) and Dumas, Fleming & Whaley (1998) have found significance in their respective models when predicting the smile, but we believe that entering external factors in the equation further increases the explanatory power of the parameters. In contrast to the above, however, we do not find any significant correlation between the lags of the parameters when conducting a VAR-regression. A replicating study of Gonçalves & Guidolin (2005), with the introduction of external variables in the univariate parameter functions, could hence be an interesting attempt at combining the two.

6. Conclusion

In this study we have examined the properties of the implied volatility surface, with a focus on the ability to forecast its evolution with external factors and lag-models . We have done this by conducting a parameterization on the cross-section of the volatility surface, i.e. the time slices, and conducted time-series regressions on the individual parameters using exogenous values and AR/VAR-models. To our knowledge, combining the lag-models used with a control for external variables has not been done before, and should help us explain some of the time-varying features of the volatility smile.

We first study the univariate regression results for each of the five estimated parameters in our model. Only one parameter lacks significance when regressing on TTM, which is expected as the volatility surface is a time-varying entity. More interesting is that we find correlation with our other three exogenous variables as well. When regressing on the VIX, we get significant betas on three of our five parameters, among those the parameter controlling for ATM implied volatility and interestingly the parameter controlling for shift in the smile. The correlation between traded volume and the parameters is also significant in three cases, which also holds true for our regression with the future price of the S&PCOMP.

Our next area of study is the multivariate regression on the parameters, controlling for all factors mentioned above. Here we wanted to study the success at which the external factors could correctly describe the movement of the parameters. The R^2-values ranged from 0.058 to 0.918, and in general TTM and the logarithm of the future S&PCOMP were the only significant variables in the multivariate setting, whereas the VIX and Volume were completely insignificant in this setting.

Apart from trying to understand the variations of the volatility surface with the help of external factors, we have also looked at the predictive power of the parameters which we have estimated. We carried out lag-factor estimate tests and used AR-models to look at the #-lag power of each parameter. One parameter showed a significant 1-,3- and 4-lag beta, indicating a strong past predictive power. Three parameters showed significant 1-lag betas, whereas one parameter showed no prior explanatory power.

Our last test was looking at a vectorized AR-model to study the level of correlation between the lags of the parameters, with the ultimate goal of finding a predictive power in their combined movements. This test showed very little significance however, with only one parameter showing a 1-lag correlation with another parameter at the 5% level.

7. Further research

One problem with conducting this type of research is that the parameterization technique determines many of the results in the later stages of the study. One interesting approach would be to use several different techniques in the first stage and use the same methods in the later stages, as to see how the parameterization influences later results.

Another point of further interest would be to model a complete volatility surface for each day, i.e. using all options traded on a given day, regardless of their TTM. Instead of extrapolating the smiles for each day, one would look at the evolution of the entire smile. This study would most likely be a more practically oriented approach, as traders do in fact look at all maturities on a given day when issuing their orders. When doing a study like this, it might be interesting to account for certain day-of-the-week effects as concluded by Berument & Kiymaz (2001).

A third point of further research is the combination of VAR-regressions and the inclusion of external variables. We believe that the external variables do increase the explanatory power of the parameters, and a successful combination of the two could definitely bring some new light onto the evolution of the smile.

8. Bibliography

Alentorn, A. (2004). *Modelling the implied volatility surface: an empirical study for FTSE options.* University of Essex, Centre for Computational Finance and Economic Agents, Essex.

Berument, H., & Kiymaz, H. (2001). The Day of the Week Effect on Stock Market Volatility. *Journal of Economics and Finance , 25* (2).

Black, F., & Scholes, M. (1973, june). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* , 637-659.

Canina, L., & Figlewski, S. (1993). The Informational Content of Implied Volatility. *The Review of Financial Studies , VI*, 659-681.

CBOE. (2009, September). *S&P 500 (SPX) Index Options.* Retrieved May 18, 2010, from Chicago Board of Exchange: http://www.cboe.com/micro/spx/pdf/SPX_QRG.pdf

Christensen, B. J., & Prabhala, N. R. (1998). The Relation Between Implied and Realized Volatility. *Journal of Financial Economics , 50* (2), 125-150.

Cont, R., & Da Fonseca, J. (2002). Dynamics of Implied Volatility Surfaces. *Quantitative Finance , II*, 45-60.

Cont, R., Da Fonseca, J., & Durrleman, V. (2002). Stochastic Models of Implied Volatility Surfaces. *Economic Notes , 31* (2), 361-377.

Dash, S., & Liu, B. (2008, June). *Capturing the Index Effects via Options.* Retrieved June 18, 2010, from Chicago Board of Exchange: http://www.cboe.com/Institutional/pdf/500_Index_Effect_Options.pdf

Derman, E., & Kani, I. (1998). Stochastic Implied Trees: Arbitrage Pricing with Stochastic Term and Strike Structure of Volatility. *International Journal of Theoretical and Applied Finance , I*, 61-110.

Diebold, F. X., & Li, C. (2005). Forecasting the term structure of government. *Journal of Econometrics , 130*, 337-364.

Dumas, B., Fleming, J., & Whaley, R. E. (1998). Implied Volatility Functions: Empirical Tests. *The Journal of Finance , 53* (6), 2059-2106.

Dunteman, G. H. (1989). *Principal Components Analysis.* Newbury Park, California: SAGE Publications, Inc.

Dupire, B. (1994). Pricing with a smile. *RISK , 7*, 18-20.

Fengler, M., Härdle, W., & Villa, C. (2003). A Common Principal Component Approach. *Review of Derivatives Research , 6* (3), 179-202.

Fleming, J. (1998). The Quality of Market Volatility Forecasts Implied by S&P100 Index Option Prices. *Journal of Empirical Finance , 5* (4), 317-345.

Gatheral, J. (2004). *A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives.* Global Derivatives & Risk Management 2004. Madrid: Merril Lynch.

Gatheral, J. (2006). *The Volatility Surface - A Practitioners Guide.* Hoboken: Johan Wiliey & Sons.

Gonçalves, S., & Guidolin, M. (2005). *Predictable Dynamics in the S&P 500 Index Option series.* Federal Reserve Bank of St. Louis, Research Division. St. Louis: Federal Reserve Bank of St. Louis Working Papers.

Gross, L., & Waltner, N. (1995). S&P 500 Options: Put Volatility Smile and Risk Aversion. *Salomon Brothers memo* .

Hull, J. C. (2008). *Fundamentals of Futures and Options Markets* (Vol. 6th). Upper Saddle River: Pearson Prentice Hall.

MB Risk Management. (2010). *UNIVOPT- Universal Options Add-in*. Retrieved May 18, 2010, from MB Risk Management: http://www.mbrm.com/product_uoa.shtml

McMillan, L. G. (1986). *Options as a strategic investment : a comprehensive analysis of listed option strategies* (2 ed.). New York: New York Institute of Finance.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business , 60* (4), 473-489.

Peña, I., Rubio, G., & Serna, G. (1998, September). Why do we smile? On the determinants of the implied volatility function. *Journal of Banking and Finance* , 1151-1179.

Plasmans, J. (2006). *Modern Linear and Nonlinear Econometrics* (Vol. 9). Antwerpen: Springer.

Rebonato, R. (1999). *Volatility and Correlation.* New York: John Wiley & Sons.

Rubinstein, M. (1994). Implied Binomial Trees. *Journal of Finance , 49*, 771-818.

Skiadopoulos, G., Hodges, S., & Clewlow, L. (1999). The Dynamics of the S&P 500 Implied Volatility Surface. *Review of Derivatives Research , 3*, 263-282.

Thomson Reuters. (2010, 05 18). Option: Market Price. *Datastream* (4.0).

9. Appendix

Table 1

First-step regression DO-file for option series 1004C

Programming code from STATA, for the first-step regression of our thesis. The common directory must be changed according to data source, and for the estout command to work properly one must install the estout feature. The forval-loop can be altered according to the range of options one wishes to loop.

STATA DO-file

```
cd "[DIRECTORY OF CHOICE]"

clear

set mem 500m
set more off

**ssc install estout

insheet using "1004C-FSTP.csv", delimiter (";")

program nlfstp, rclass

   version 9
   syntax varlist(min=2 max=2) if
   local y : word 1 of `varlist'
   local x : word 2 of `varlist'

   return local eq "`y' = {b0=0.1} + {b1=3}*({b2=0.1}*(`x'-{b3=0.02})+((((`x'-
{b3=0.02})^2)+{b4=0.05}^2)^0.5))"
   return local title "First-step reg"

end

forval i=1(1)257{
nl fstp : iv`i' k`i',nolog iterate(100) eps(.001)
estout using 1004C-FSTP-Raw.csv, append cells("b se t p ") stats(N r2 converge) title(iv`i') delimiter(";")
}

save 1004C-fstp, replace
```

Table 2

Parameterization results from first-step SVI regression

The first-step parameterization results from the SVI regression. The model regressed is $\sigma = \alpha + \beta\left(\rho(k-m) + \sqrt{(k-m)^2 + c^2}\right) + \varepsilon$, and the b0-b4 parameters are listed in order of appearance, where k is the dependant variable. First the 1004C option series is listed, and second the 0804C series.

1004C Option Series

Time	**2009-12-11**	**2009-12-14**	**2009-12-15**	**2009-12-16**	**2009-12-17**	**2009-12-18**	**2009-12-21**
b0	0.121	0.149	0.125	0.131	0.146	0.150	0.135
b1	1.350	1.070	1.353	1.271	1.273	1.088	1.261
b2	0.045	0.024	0.087	0.035	-0.013	-0.034	0.237
b3	0.070	0.062	0.071	0.064	0.060	0.057	0.069
b4	0.040	0.021	0.035	0.030	0.020	0.015	0.019
N	68	68	68	68	83	83	83
R2	0.992	0.992	0.994	0.995	0.989	0.992	0.995

Time	**2009-12-22**	**2009-12-23**	**2009-12-24**	**2009-12-25**	**2009-12-28**	**2009-12-29**	**2009-12-30**
b0	0.099	0.136	0.113	0.113	0.124	0.128	0.121
b1	1.711	1.015	1.456	1.456	1.245	1.128	1.300
b2	0.255	-0.064	0.168	0.168	0.088	-0.033	0.103
b3	0.070	0.055	0.061	0.061	0.057	0.054	0.058
b4	0.032	0.017	0.028	0.028	0.021	0.022	0.025
N	83	83	83	83	83	83	83
R2	0.999	0.998	0.999	0.999	0.995	0.998	0.998

Time	**2009-12-31**	**2010-01-01**	**2010-01-04**	**2010-01-05**	**2010-01-06**	**2010-01-07**	**2010-01-08**
b0	0.124	0.124	0.131	0.127	0.115	0.143	0.092
b1	1.243	1.243	1.167	1.182	1.298	2.680	1.567
b2	-0.015	-0.015	-0.048	-0.025	-0.006	0.147	0.077
b3	0.060	0.060	0.049	0.048	0.047	0.060	0.046
b4	0.029	0.029	0.018	0.017	0.026	0.110	0.035
N	83	83	83	83	83	83	83
R2	1.000	1.000	0.998	0.998	0.997	0.901	0.999

Time	**2010-01-11**	**2010-01-12**	**2010-01-13**	**2010-01-14**	**2010-01-15**	**2010-01-18**	**2010-01-19**
b0	0.131	0.083	0.115	0.099	0.116	0.116	0.120
b1	1.204	1.914	1.510	1.683	1.463	1.463	1.436
b2	0.092	0.092	0.158	0.109	0.135	0.135	0.175
b3	0.042	0.045	0.043	0.042	0.042	0.042	0.039
b4	-0.013	0.036	0.019	0.025	0.019	0.019	-0.011
N	83	90	90	90	90	90	90
R2	0.954	0.996	0.990	0.991	0.985	0.985	0.991

Time	**2010-01-20**	**2010-01-21**	**2010-01-22**	**2010-01-25**	**2010-01-26**	**2010-01-27**	**2010-01-28**
b0	0.114	0.150	0.159	0.149	0.163	0.153	0.155
b1	1.549	1.236	1.698	1.643	1.349	1.473	1.394
b2	0.094	0.125	-0.108	0.124	0.205	0.122	0.225
b3	0.040	0.040	0.039	0.045	0.046	0.041	0.049
b4	0.017	0.001	0.009	0.004	0.000	-0.003	-0.004
N	90	90	90	90	90	90	90
R2	0.994	0.950	0.952	0.972	0.924	0.943	0.930

Time	2010-01-29	2010-02-01	2010-02-02	2010-02-03	2010-02-04	2010-02-05	2010-02-08
b0	0.132	0.131	0.154	0.104	0.157	0.167	0.158
b1	1.987	1.721	1.177	1.919	1.729	1.399	1.835
b2	0.291	0.171	0.606	0.041	0.045	0.314	0.074
b3	0.059	0.045	0.053	0.039	0.047	0.052	0.046
b4	0.016	0.013	0.000	0.029	0.017	-0.005	0.016
N	90	90	90	90	90	90	90
R2	0.988	0.974	0.729	0.975	0.965	0.910	0.959

Time	2010-02-09	2010-02-10	2010-02-11	2010-02-12	2010-02-15	2010-02-16	2010-02-17
b0	0.150	0.162	0.137	0.115	0.115	0.140	0.093
b1	1.820	1.650	1.863	2.498	2.498	1.855	2.622
b2	0.227	0.254	0.167	-0.145	-0.145	0.142	-0.075
b3	0.050	0.047	0.046	0.035	0.035	0.035	0.026
b4	0.014	0.008	0.015	0.023	0.023	0.009	0.026
N	90	90	90	90	90	90	90
R2	0.952	0.965	0.980	0.975	0.975	0.981	0.996

Time	2010-02-18	2010-02-19	2010-02-22	2010-02-23	2010-02-24	2010-02-25	2010-02-26
b0	0.093	0.146	0.061	0.038	0.151	0.075	0.111
b1	2.530	1.655	3.185	3.911	2.293	3.074	2.703
b2	0.039	0.496	-0.041	-0.167	0.703	0.014	0.146
b3	0.024	0.035	0.021	0.020	0.038	0.023	0.024
b4	0.023	-0.004	0.028	0.033	0.005	0.025	0.011
N	90	90	90	90	90	90	90
R2	0.982	0.908	0.993	0.997	0.934	0.997	0.980

Time	2010-03-01	2010-03-02	2010-03-03	2010-03-04	2010-03-05	2010-03-08	2010-03-09
b0	0.091	0.099	0.041	0.051	0.120	0.127	0.019
b1	2.976	3.020	4.118	3.874	2.372	2.944	5.386
b2	0.121	0.094	-0.100	-0.005	0.550	0.421	-0.104
b3	0.020	0.018	0.011	0.012	0.021	0.018	0.003
b4	0.017	0.015	0.026	-0.024	-0.010	-0.008	0.023
N	90	90	90	90	90	90	90
R2	0.975	0.988	0.997	0.993	0.962	0.936	0.997

Time	2010-03-10	2010-03-11	2010-03-12	2010-03-15	2010-03-16
b0	0.053	0.039	0.051	0.063	0.118
b1	5.515	6.292	7.401	8.580	9.657
b2	-0.102	-0.140	-0.219	-0.033	0.000
b3	0.006	0.004	0.004	0.004	0.004
b4	0.018	-0.016	-0.012	-0.011	-0.006
N	90	90	90	90	90
R2	0.994	0.997	0.992	0.997	0.970

0804C Option Series

Time	2008-02-04	2008-02-07	2008-02-08	2008-02-11	2008-02-12	2008-02-19	2008-02-20
b0	0.140	0.156	0.176	0.165	0.160	0.136	0.158
b1	2.466	1.613	1.122	1.399	1.298	2.738	1.710
b2	0.524	0.134	-0.020	0.235	0.037	0.381	0.403
b3	0.071	0.073	0.070	0.073	0.068	0.069	0.061
b4	0.002	0.013	0.000	0.003	0.006	0.014	0.000
N	26	26	26	27	30	33	77
R2	0.992	0.989	0.993	0.964	0.982	0.974	0.932

Time	2008-02-22	2008-02-25	2008-02-26	2008-02-29	2008-03-03	2008-03-04	2008-03-05
b0	0.108	0.150	0.140	0.177	0.172	0.165	0.155
b1	2.968	1.947	2.127	1.457	1.740	2.051	2.185
b2	0.430	0.439	0.380	0.136	0.232	0.427	0.290
b3	0.062	0.057	0.054	0.062	0.057	0.059	0.052
b4	0.019	0.000	-0.004	-0.004	0.004	-0.005	0.007
N	77	82	82	84	87	87	87
R2	0.919	0.928	0.958	0.972	0.977	0.912	0.978

Time	2008-03-06	2008-03-07	2008-03-10	2008-03-11	2008-03-12	2008-03-13	2008-03-14
b0	0.177	0.190	0.195	0.056	0.189	0.196	0.201
b1	2.298	1.702	2.277	3.148	2.325	1.891	2.824
b2	0.298	0.303	0.339	0.183	0.414	0.609	0.361
b3	0.062	0.062	0.066	0.050	0.063	0.060	0.057
b4	-0.008	-0.002	-0.006	0.043	-0.002	0.000	0.013
N	87	87	89	89	89	89	89
R2	0.983	0.946	0.970	0.921	0.898	0.624	0.877

Time	2008-03-17	2008-03-18	2008-03-19	2008-03-24	2008-03-27	2008-03-28	2008-03-31
b0	0.149	-0.005	0.193	0.147	0.135	0.150	0.157
b1	3.188	3.697	2.057	2.895	3.513	3.089	2.751
b2	0.382	0.125	0.248	0.245	0.162	0.146	0.248
b3	0.068	0.040	0.069	0.043	0.042	0.039	0.042
b4	0.030	0.054	-0.009	-0.011	-0.016	0.014	0.012
N	89	89	89	90	90	90	91
R2	0.892	0.865	0.737	0.951	0.970	0.947	0.973

Time	2008-04-02	2008-04-03	2008-04-04	2008-04-07	2008-04-08	2008-04-09	2008-04-10
b0	0.037	-0.053	0.128	0.111	-0.027	-0.042	0.044
b1	4.498	5.896	3.900	5.543	7.487	8.057	6.789
b2	0.053	0.116	0.240	-0.065	-0.136	-0.156	-0.021
b3	0.027	0.029	0.027	0.019	0.012	0.010	0.011
b4	0.030	0.041	0.010	-0.009	-0.028	0.029	-0.019
N	92	92	92	92	92	92	92
R2	0.891	0.974	0.944	0.902	0.992	0.980	0.986

Time	2008-04-11	2008-04-14	2008-04-15	2008-04-16
b0	-0.047	0.042	-0.043	0.158
b1	8.241	11.152	12.229	21.769
b2	-0.017	0.000	0.068	-0.194
b3	0.012	0.011	0.006	0.005
b4	0.031	-0.019	-0.024	-0.006
N	92	92	92	92
R2	0.967	0.967	0.979	0.988

Table 3

R^2 statistics for parameterization of option series

R^2 statistics of the parameterization for each cross-section of the option series. Reported is the arithmetic average of the R^2, the standard deviation, and the minimum and maximum reported R^2 of the parameterization.

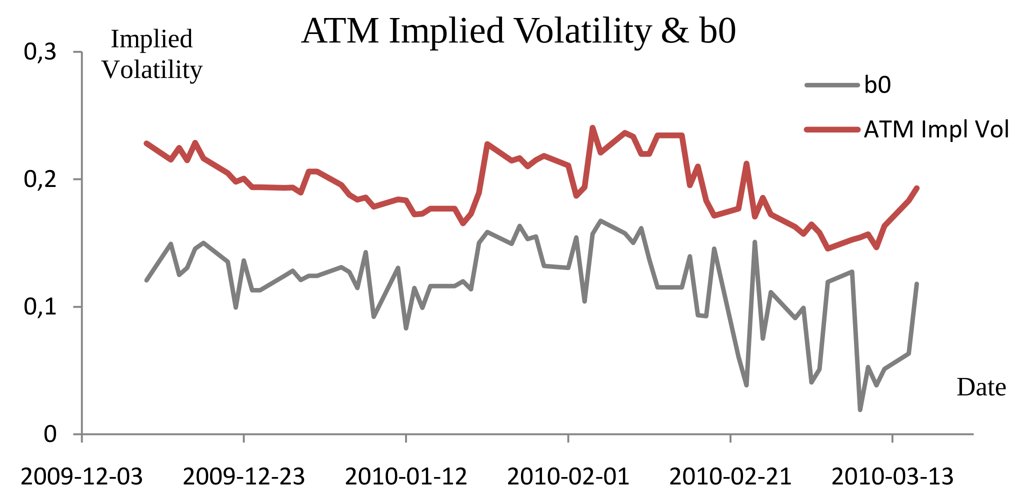
R2 statistics for option series

Series	Average	Std. Dev.	Min	Max
1004C	0.975	0.039	0.729	1.000
0804C	0.938	0.071	0.624	0.993

Table 4

Implied volatility of the ATM option for each day modeled against the b0, and the difference between the two

The first graph shows the ATM Implied volatility for each day, and the corresponding b0. The second graph shows the difference between the two per day.



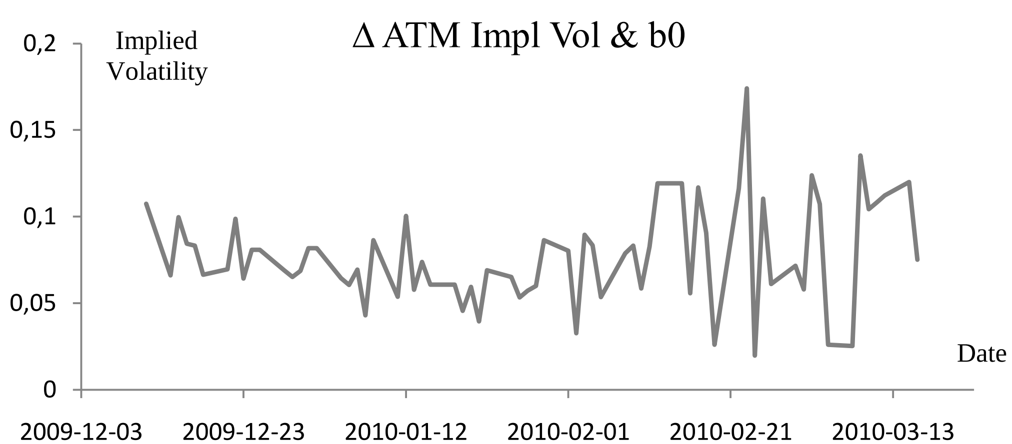
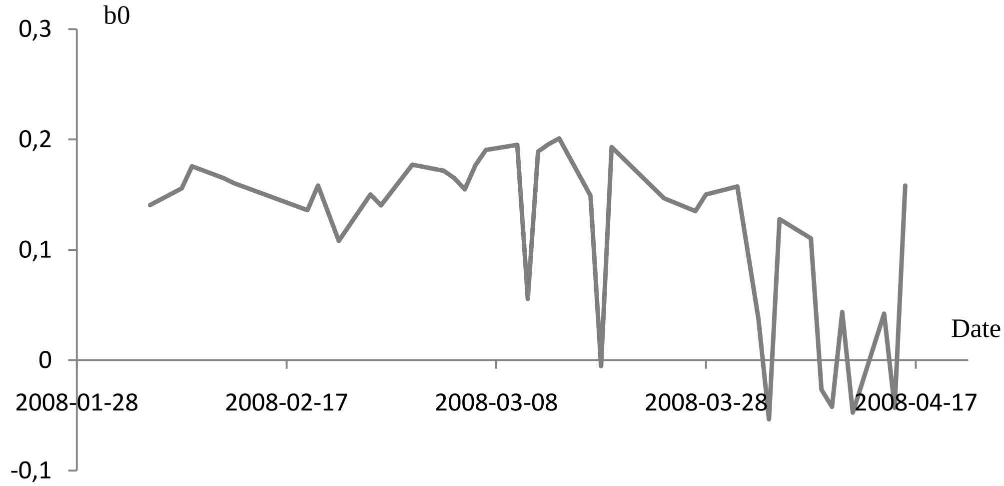


Table 5

Graphical interpretation of α (b0) over time for both option series

The graph shows the evolution of the α-paramter (b0) over the regressed period for each option series. The elapsed time is on the x-axis, while the parameter itself is on the y-axis.

0804C



1004C

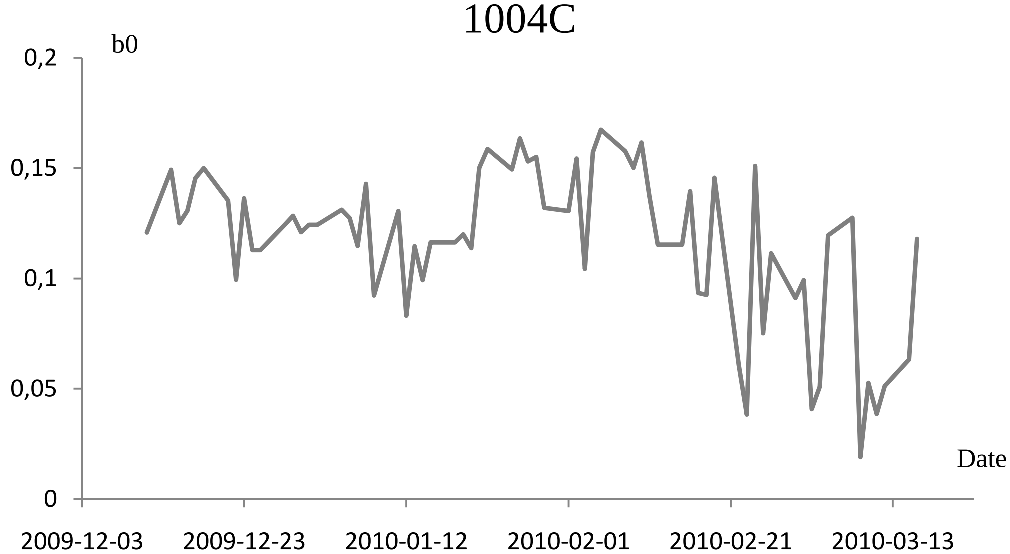
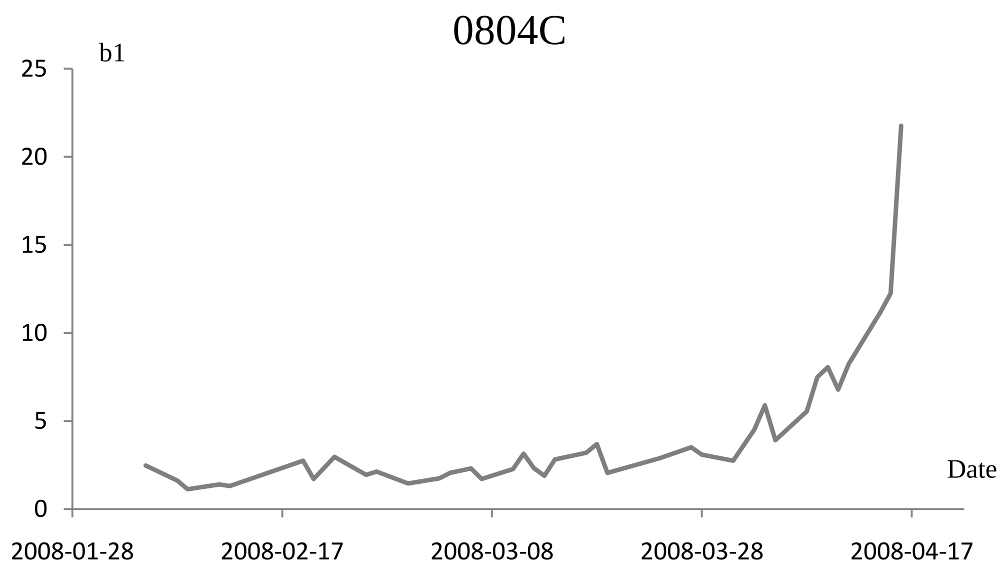


Table 6

Graphical interpretation of β (b1) over time for both option series

The graph shows the evolution of the β-paramter (b1) over the regressed period for each option series. The elapsed time is on the x-axis, while the parameter itself is on the y-axis.



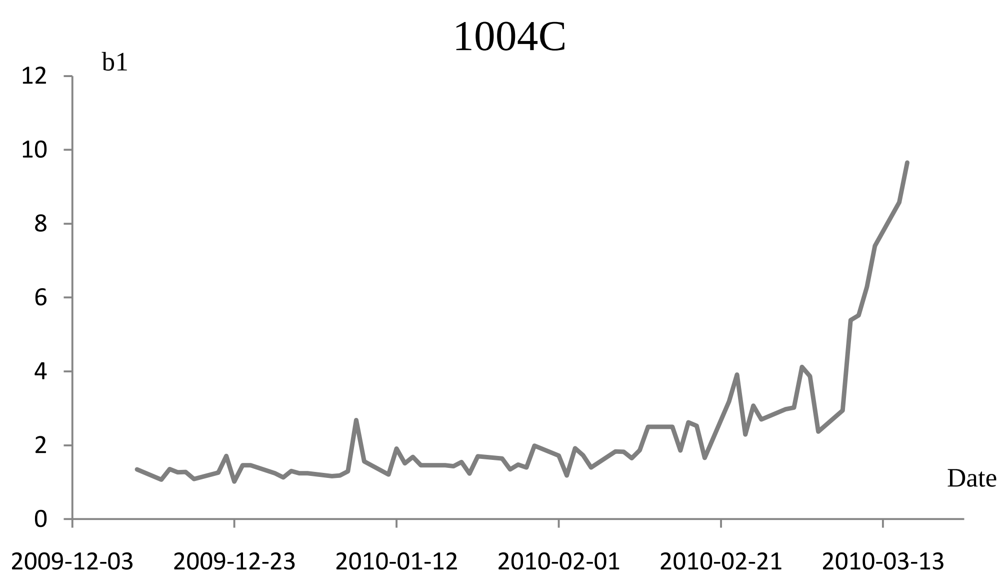
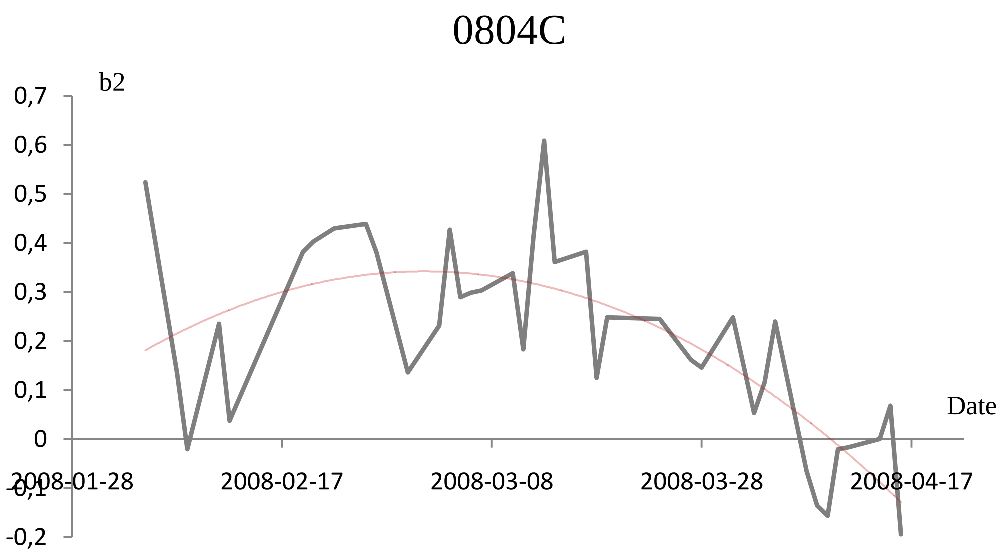


Table 7

Graphical interpretation of ρ (b2) over time for both option series

The graph shows the evolution of the ρ-paramter (b2) over the regressed period for each option series. The elapsed time is on the x-axis, while the parameter itself is on the y-axis.



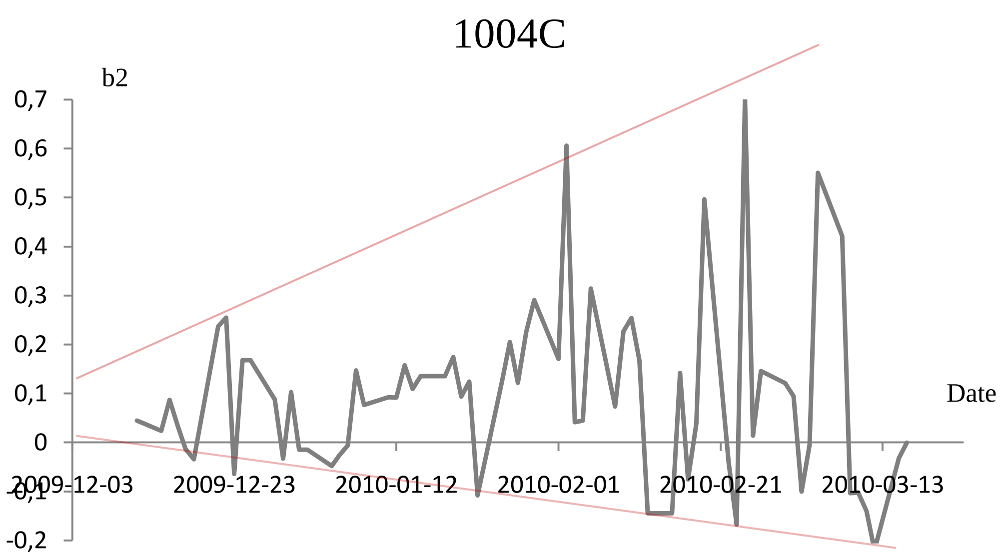
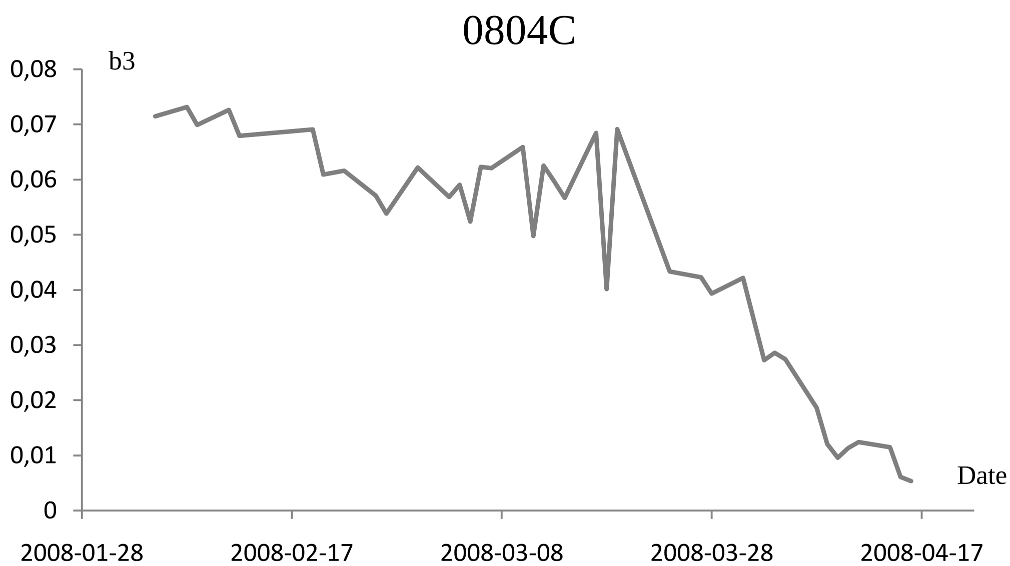


Table 8

Graphical interpretation of m (b3) over time for both option series

The graph shows the evolution of the m-paramter (b3) over the regressed period for each option series. The elapsed time is on the x-axis, while the parameter itself is on the y-axis.



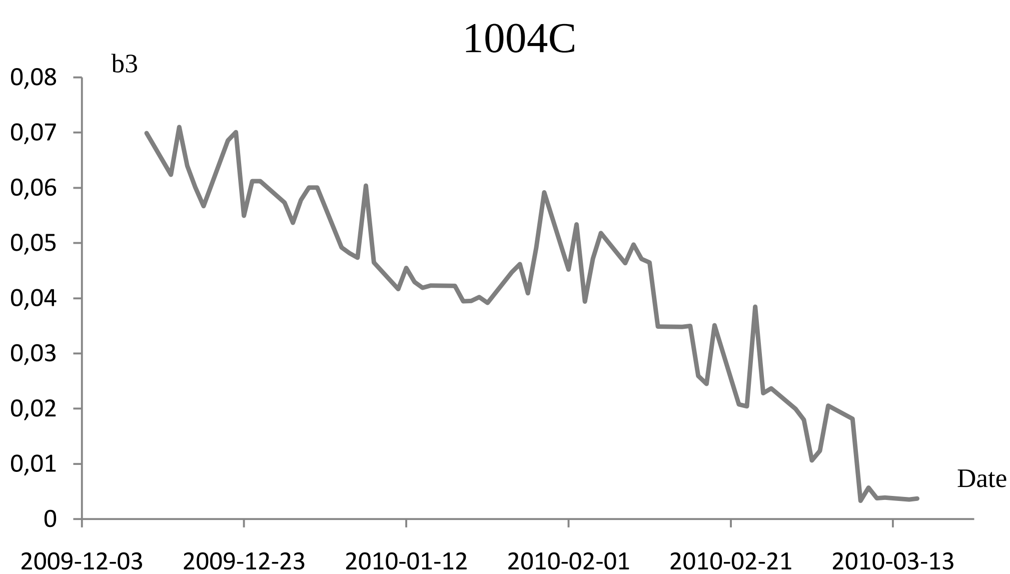
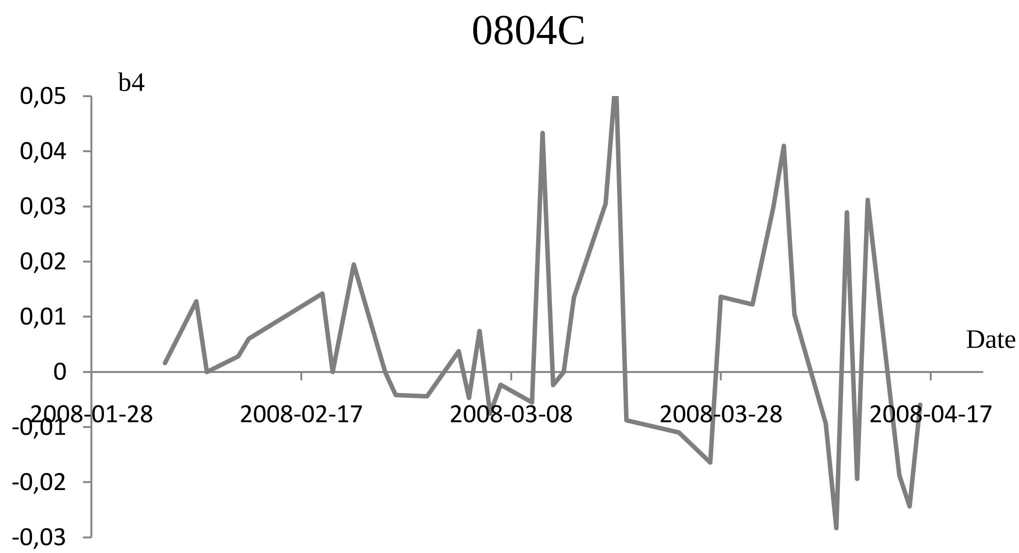


Table 9

Graphical interpretation of *c* (b4) over time for both option series

The graph shows the evolution of the *c*-paramter (b4) over the regressed period for each option series. The elapsed time is on the x-axis, while the parameter itself is on the y-axis.



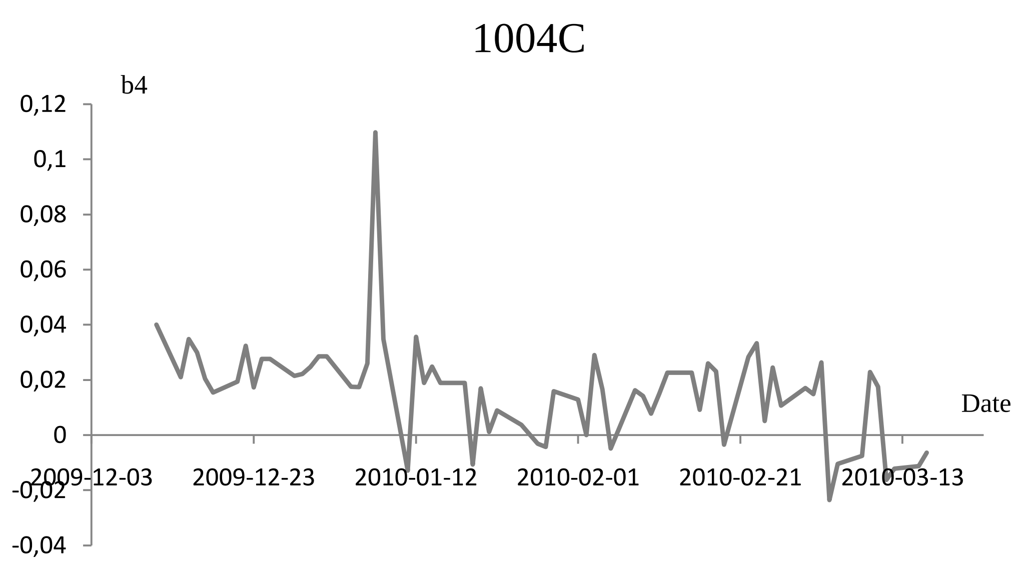


Table 10

Regression results from correlation between TTM and parameters

The table shows the results from the regression $P_t = \alpha + \beta TTM_t + \varepsilon$, for all five of the parameters.

Correlation between TTM and parameters, 1004C Option series

	b0		b1		b2		b3		b4	
	b	**p**	**b**	**p**	**b**	**p**	**b**	**p**	**b**	**p**
TTM	0.168	0.010	-16.250	0.000	-0.032	0.914	0.221	0.000	0.109	0.000
_cons	0.090	0.000	4.510	0.000	0.100	0.082	0.011	0.000	0.001	0.768
N	68		68		68		68		68	
r2	0.074		0.504		0.000		0.820		0.196	

Table 11

Regression results from correlation between VIX and parameters

The table shows the results from the regression $P_t = \alpha + \beta \log{(VIX_t)} + \varepsilon$, for all five of the parameters.

Correlation between VIX and parameters, 1004C Option series

	b0		b1		b2		b3		b4	
	b	**p**	**b**	**p**	**b**	**p**	**b**	**p**	**b**	**p**
ln(VIX)	0.305	0.000	-12.192	0.001	-0.083	0.820	0.157	0.000	0.041	0.131
_cons	-0.845	0.000	40.566	0.000	0.355	0.758	-0.452	0.000	-0.112	0.187
N	68		68		68		68		68	
r2	0.180		0.208		0.001		0.304		0.020	

Table 12

Regression results from correlation between volume and parameters

The table shows the results from the regression $P_t = \alpha + \beta \log{(Volume_t)} + \varepsilon$ for all five of the parameters.

Correlation between volume and parameters, 1004C Option series

	bo		b1		b2		b3		b4	
	b	**p**	**b**	**p**	**b**	**p**	**b**	**p**	**b**	**p**
ln(Vol)	-0.001	0.214	0.123	0.013	0.005	0.412	-0.002	0.012	-0.001	0.001
_cons	0.121	0.000	1.002	0.027	0.042	0.507	0.061	0.000	0.030	0.000
N	68		68		68		68		68	
r2	0.002		0.038		0.006		0.080		0.040	

Table 13

Regression results from correlation between future S&PCOMP and parameters

The table shows the results from the regression $P_t = \alpha + \beta \log(Fprice_t) + \varepsilon$, for all five of the parameters.

Correlation between future S&PCOMP and parameters, 1004C Option series

	bo		b1		b2		b3		b4	
	b	p	b	p	b	p	b	p	b	p
ln(S&PFUT)	-0.954	0.000	30.181	0.013	-0.986	0.258	-0.276	0.001	-0.031	0.768
_cons	6.802	0.000	-209.353	0.014	7.009	0.251	1.980	0.001	0.236	0.752
N	68		68		68		68		68	
r2	0.218		0.158		0.016		0.117		0.001	

Table 14

Regression results from correlation between TTM, VIX, volume, future S&PCOMP and parameters

The table shows the results from the regression

$$P_t = \alpha + \beta TTM_t + \beta \log(Fprice_t) + \beta \log(Volume_t) + \beta \log(Vix_t) + \varepsilon$$

for all five of the parameters.

Correlation between TTM, VIX, volume, future S&PCOMP and parameters, 1004C Option series

	b0		b1		b2		b3		b4	
	b	p	b	p	b	p	b	p	b	p
TTM	0.160	0.003	-17.521	0.000	0.274	0.384	0.217	0.000	0.111	0.000
ln(S&PFUT)	-0.912	0.031	37.573	0.001	-2.582	0.105	-0.257	0.000	-0.072	0.646
ln(Vol)	0.001	0.523	-0.012	0.715	0.006	0.338	0.000	0.321	-0.001	0.090
ln(VIX)	0.013	0.915	4.689	0.082	-0.820	0.185	-0.002	0.912	-0.027	0.538
_cons	6.443	0.052	-273.414	0.002	20.669	0.107	1.818	0.000	0.599	0.625
N	68		68		68		68		68	
r2	0.285		0.656		0.058		0.918		0.206	

Table 15

Results from lag-order selection tests and correspond AR-regressions on all five parameters

Lag-order selection tests to determine the lower of the AIC and the SBIC criteria. After determining the lower criteria, a AR(p)-regression with p lags according to

$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$ is run, with the following results.

Lag-order selection tests and corresponding AR-regressions for all parameters

Selection test (b0)				AR-regression (b0)		
Lag	**p**	**AIC**	**SBIC**	**Lag**	**b**	**p**
0		-4.081	-4.037	**L.b0**	0.183	0.003
1	0.014	-4.388	-4.301	**L2.b0**	-0.038	0.724
2	0.717	-4.244	-4.114	**L3.b0**	0.572	0.001
3	0.005	-4.684	-4.510	**L4.b0**	-0.558	0.014
4	0.006	**-5.111**	**-4.894**	**_cons**	0.116	0.001
				N	13	
				r2	0.8072	

Selection test (b1)				AR-regression (b1)		
Lag	**p**	**AIC**	**SBIC**	**Lag**	**b**	**p**
0		3.902	3.946	**L.b1**	1.057	0.000
1	0.000	**2.457**	**2.544**	**_cons**	-0.025	0.850
2	0.837	2.608	2.738	**N**	53	
3	0.118	2.573	2.747	**r2**	0.840	
4	0.026	2.344	2.561			

Selection test (b2)				AR-regression (b2)		
Lag	**p**	**AIC**	**SBIC**	**Lag**	**b**	**p**
0		**0.037**	**0.080**			
1	0.652	0.175	0.262			
2	0.481	0.291	0.421			
3	0.391	0.388	0.562			
4	0.045	0.231	0.449			

Selection test (b3)				AR-regression (b3)		
Lag	**p**	**AIC**	**SBIC**	**Lag**	**b**	**p**
0		-5.181	-5.137	**L.b3**	0.939	0.000
1	0.000	**-6.966**	**-6.879**	**_cons**	0.002	0.355
2	0.640	-6.829	-6.699	**N**	53	
3	0.104	-6.879	-6.705	**r2**	0.852	
4	0.652	-6.741	-6.523			

Selection test (b4)				AR-regression (b4)		
Lag	**p**	**AIC**	**SBIC**	**Lag**	**b**	**p**
0		-5.432	-5.389	**L.b4**	0.311	0.007
1	0.004	-5.926	-5.839	**L2.b4**	0.110	0.416
2	0.059	**-6.045**	**-5.915**	**_cons**	0.009	0.018
3	0.322	-5.967	-5.793	**N**	39	
4	0.666	-5.827	-5.610	**r2**	0.1016	

Table 16

Results from VAR-regression on all five parameters

The results from the VAR-regression according to the following formula: $y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon$. The results are as follows:

Results from VAR-regression on all five parameters

		b0		b1		b2		b3		b4	
	Lags	**Coef.**	**P>|z|**	**Coef.**	**P>|z|**	**Coef.**	**P>|z|**	**Coef.**	**P>|z|**	**Coef.**	**P>|z|**
b0	**L1.**	*-0.031*	*0.930*	6.287	0.113	-1.308	0.294	0.008	0.842	0.010	0.940
	L2.	*0.197*	*0.747*	3.750	0.578	0.023	0.991	-0.068	0.292	-0.068	0.756
b1	**L1.**	-0.012	0.702	*0.955*	*0.005*	-0.135	0.209	-0.001	0.686	0.001	0.949
	L2.	0.008	0.836	*0.497*	*0.251*	-0.032	0.812	-0.005	0.219	-0.010	0.482
b2	**L1.**	-0.101	0.247	1.190	0.217	*-0.510*	*0.092*	-0.035	0.000	0.039	0.218
	L2.	0.162	0.071	-1.460	0.142	*0.386*	*0.216*	0.010	0.297	-0.040	0.212
b3	**L1.**	4.103	0.086	-33.999	0.199	9.480	0.255	*1.116*	*0.000*	-1.012	0.241
	L2.	-3.570	0.181	43.539	0.140	-18.459	0.047	*-0.395*	*0.163*	1.061	0.270
b4	**L1.**	-0.711	0.356	9.076	0.287	-2.857	0.286	-0.125	0.125	*0.267*	*0.337*
	L2.	0.680	0.559	-15.860	0.219	4.393	0.278	0.047	0.706	*0.005*	*0.990*
	_cons	0.074	0.488	-2.257	0.057	0.956	0.010	0.035	0.002	0.035	0.367

Table 17

Third-step regression DO-file for option series 1004C

Programming code from STATA, for the first-step regression of our thesis. The common directory must be changed according to data source, and for the estout command to work properly one must install the estout feature. The forval-loop can be altered according to the range of options one wishes to loop.

STATA DO-file

```
cd “[DIRECTORY OF CHOICE]"
clear
set mem 500m
set more off

**ssc install estout

insheet using 1003C-SNDSTP.csv, delimiter (";")

gen datenr=date(time,"ymd")
format datenr %d
drop time
order datenr

gen lnvol=ln(volume)
replace lnvol=0 if lnvol==.
gen lnspcomp=ln(spcomp)
replace lnspcomp=0 if lnspcomp==.
gen lnspfuture=ln(spfuture)
replace lnspfuture=0 if lnspfuture==.
gen lnvix=ln(vix)
replace lnvix=0 if lnvix==.

global b "b0 b1 b2 b3 b4"

forval i=0(1)4{
reg b`i' ttm,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b`i' ttm) delimiter(";")
}

forval i=0(1)4{
reg b`i' lnvix,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b`i' vix) delimiter(";")
}

forval i=0(1)4{
reg b`i' lnvol,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b`i' lnvol)
delimiter(";")
}
```

```
forval i=0(1)4{
reg b`i' lnspfuture,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b`i' lnspfuture)
delimiter(";")
}

forval i=0(1)4{
reg b`i' ttm lnspfuture lnvol lnvix,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b`i' ttm lnspfuture
lnvol lnvix) delimiter(";")
}

pwcorr $b,sig

tsset datenr

varsoc b0
reg b0 L(1/4).b0,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b0 lags) delimiter(";")

varsoc b1
reg b1 L(1/1).b1,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b1 lags) delimiter(";")

varsoc b2

varsoc b3
reg b3 L(1/1).b3,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b3 lags) delimiter(";")

varsoc b4
reg b4 L(1/2).b4,r
estout using 1003C-SNDSTP-Raw.csv, append cells("b p") stats(N r2) title(b4 lags) delimiter(";")

var b0 b1 b2 b3 b4

save 1004C-sndstp, replace
```

Table 18

Implied Volatility Surface for option series 1004C

The implied volatility surface for the 1004C option series, modeled during the time series on which we conducted our regressions.