Application of model combination to forecast the implied volatility surface

Master Thesis in Quantitative Finance

Erasmus University Rotterdam

Author: Jerzy Dembski (409534)

Supervisor: Prof. Dr. Dick van Dijk
Co-reader: Dr. Wing Wah Tham

April 2016

Abstract

The aim of this thesis is to explore predictable dynamics in the implied volatility surface of S&P500 index options. I consider three approaches to modelling the surface that can be distinguished in the literature: (i) dynamic factor model where the latent factors drive the dynamics of the surface, (ii) models that assume parametric structure of the surface, (iii) option pricing model consistent with the skew observed for the implied volatilities. I find that the latent factor model provides the best accuracy in most regions of the surface at investigated one-day ahead forecasting horizon. This model combines two-step estimation procedure by means of Principal Component Analysis and VAR model for factor dynamics, with non-parametric Nadaraya-Watson regression that allows to deal with special design of implied volatility data. Forecasting accuracy can be further improved by using combination forecast methods. I implement combining methods based on equal weights, discounted mean square prediction error, and optimal estimated weights. Combining based on the estimated optimal weights yields improvement over the individual models in all regions of the surface. However, neither the individual models nor the combination models are capable of beating random walk forecast, a simple forecast that assumes that tomorrow's value of implied volatility equals its current value.

Keywords: implied volatility surface; forecast combinations; option pricing; factor models

Contents

Chapter 1

Introduction

Understanding volatility behaviour is essential for the purpose of risk management, option pricing, hedging of derivatives and supporting portfolio decisions. Much attention in the past has been paid to the *realized historical volatility*, but recently studies on the *implied volatility* (IV) are gaining popularity. The implied volatility is a measure of volatility that is obtained from the observed market prices combined with a certain option pricing model in a way that the volatility parameter ensures the model price equals the market price. This model is usually Black-Scholes for European or binomial tree model for American options. In contrast to the volatility estimates recovered from the historical data, IV is believed to be a forward-looking measure that reflects the current view on market risk and expected volatility. The seminal model of Black and Scholes (1973) assumes that a volatility of an underlying asset is constant. It implies that all options written on the same underlying should have the same implied volatility, regardless of the strike price and the time-to-maturity. Implied volatilities observed in the market exhibit quite a different pattern. The implied volatility varies systematically with moneyness (most commonly, expressed as the current price of the underlying asset relative to the strike) and time-to-maturity. This dependence gives rise to the implied volatility surface (IVS) which is a collection of implied volatilities across moneyness and time-to-maturity dimensions, and can be formally defined as a function $\sigma_t : (m, \kappa) \rightarrow \sigma_t(m, \kappa)$, where m represents moneyness and κ is the remaining time-to-maturity. There are at least two well-recognized stylized facts regarding the implied volatility surface. The pronounced *volatility smile* or *volatility smirk* (when the smile is skewed) is a U-shape pattern observed across different strikes for options' with the same maturity. The *volatility term structure* is a pattern observed across options' maturity given the strike price or moneyness. These two stylized facts implied that the surface is nonflat, contrary to what one would expect in the Black-Scholes world. Figure 1.1 depicts an actual shape of IVS that can be observed in the real world. The implied volatility surface on 26/Apr/2014 exhibits

a downward sloping shape along the moneyness dimension, which is the volatility smirk commonly observed in the market. The term structure pattern depends on the moneyness region. It slopes downward when the time-to-maturity increases for the options with high moneyness, and slopes slightly upward for options with low values of moneyness.

The implied volatility can be obtained by inverting the Black-Scholes-Merton formula given the option price. The Black-Scholes-Merton formulas for European call and put options are

$$C(t,T,S_t,K,r,\sigma,q) = S_t e^{-q(T-t)}\Phi(d_1) - Ke^{-r_f(T-t)}\Phi(d_2) \tag{1.1a}$$

$$P(t,T,S_t,K,r,\sigma,q) = Ke^{-r_f(T-t)}\Phi(-d_2) - S_t e^{-q(T-t)}\Phi(-d_1) \tag{1.1b}$$

$$d_1 = \frac{\log(\frac{S_t}{K}) + (T-t)(r_f - q + \frac{1}{2}\sigma^2)}{\sigma\sqrt{(T-t)}} \tag{1.1c}$$

$$d_2 = \frac{\log(\frac{S_t}{K}) + (T-t)(r_f - q - \frac{1}{2}\sigma^2)}{\sigma\sqrt{(T-t)}} \tag{1.1d}$$

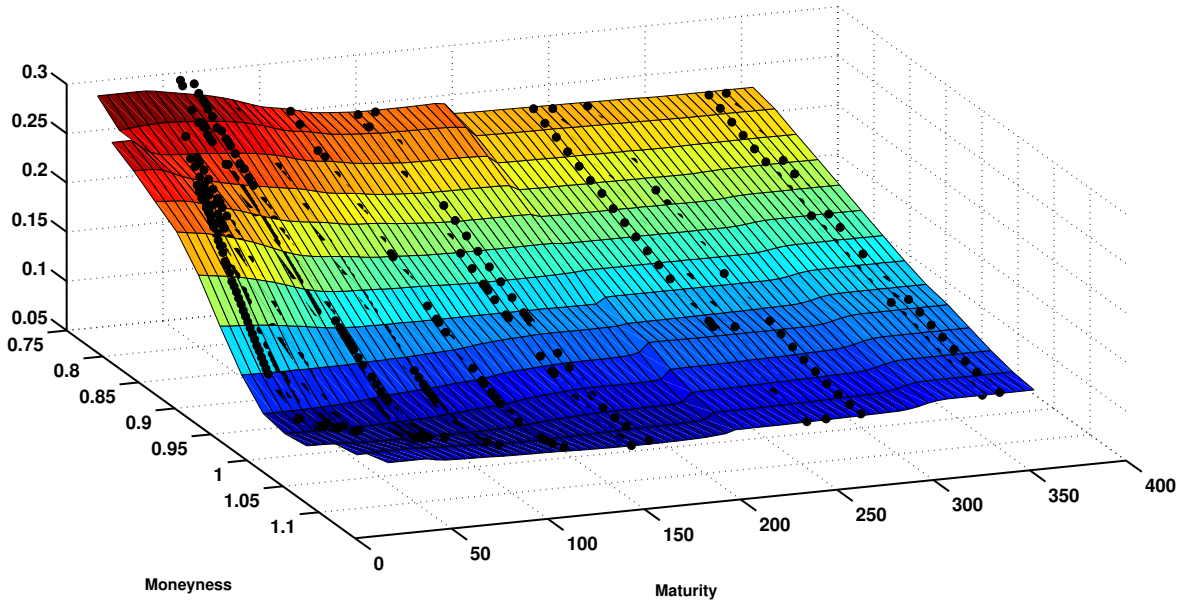
where t is the valuation date, T is the expiry date, $(T-t)$ represents the time-to-maturity, S_t is the price of the underlying asset, r_f is the risk-free rate assumed to be constant, K is the strike price, σ is the underling asset's volatility, q is the continuously compounded dividend rate and $\Phi(x)$ is a standard Normal distribution, that is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy. \tag{1.2}$$

Because the volatility σ is the only unknown parameter when the price is observed, it is easily done via numerical procedure by minimizing the deviation between the theoretical price implied by the Black-Scholes-Merton model and the observed option price. The implied volatility is sometimes referred to as "the wrong number to plug into the wrong formula to get the right price". It is a standard practice to express a value of an option with its implied volatility rather than a price because prices of options with different maturity, moneyness or written on different underlying asset are difficult to compare. Thus, knowing IVs means to known the option prices.

In this thesis, I explore the predictability of IVS using an extensive data set of daily implied volatilities on S&P500 index options. The market of the S&P500 index options is one of the most liquid derivative markets with a wide range of strikes and maturities quoted every day. It makes it very popular in research applications. I examine and asses the out-of-sample forecasting performance of the three approaches to model IVS that can be distinguished in a literature. In general, they are based on: (i) latent factors than span IVS,

Figure 1.1: Implied Volatility Surface



The figure shows the implied volatility surface from options on S&P500 index on 26/Apr/2013. The black dots represent observed implied volatilities, whereas the smooth surface is obtained with a nonparametric kernel regression, see section 2.2 for details. Moneyness (defined as $m = \frac{K}{S}$) is displayed on the left axis and time to maturity κ on the right one (measured in days).

(ii) assumed parametric structure of the surface and (iii) option pricing model.

I estimate the latent factors with Principal Component Analysis (PCA) approach. Here, I contribute to the existing literature in two ways. Firstly, I propose a method to generate out-of-sample IVS forecasts of non-constant time-to-maturity index options that can be evaluated using observed data. The procedure is as follows. I apply Nadaraya-Watson kernel regression to estimate the whole surface each day and then recover the time-series on a given grid of moneyness and maturity. Then, I extract principal components (PCs), model their dynamics and produce the forecasts of PCs with VARX model. Forecasted PCs lead directly to forecasts of selected smoothed points on the surface. Using the forecasted points on the given grid, I again apply kernel regression and obtain the forecast of the whole surface. The second contribution is that I perform PCA on the correlation matrix in contrast to the previous literature, which also used the nonparametric smoothing to estimate the surface and extracted PCs from the covariance matrix. This is motivated by the fact that variance in different regions of the surface (represented by the time-series on the given grid) is not uniform. In the parametric approach I follow Goncalves and Guidolin (2006) and assume the parametric structure of the surface proposed in their paper. The parametric

approach includes two models. The first model simply uses today's estimated factors to produce forecasts of tomorrow's IVs, i.e. this model uses random walk forecasts for estimated factors. It is used by practitioners (it is often refereed to as Practitioners Black-Scholes) and is treated as a benchmarking model for this thesis. I use random walk forecasts as an additional benchmark against all the models. The next parametric model is an extension of the benchmark that tries to capture the factor dynamics with VARX model. For the option pricing model I choose NGARCH(1,1) model of Heston and Nandi (2000). Unlike previously described models, this model is estimated in option prices space. Once the option prices are forecasted, I obtain the forecasts of the implied volatilities by inverting the Black-Scholes-Merton formula. Because the characteristic function of the return in Heston and Nandi model is known analytically, I make use of an efficient numerical method to value options in the line with work of Carr and Madan (1999). This method is known as the fast Fourier transform (FFT) and makes it feasible to price large collection of option contracts relatively fast. This is necessary given the fact that I evaluate over 600,000 forecasts of IVs.

The results I obtain provide several implications. The proposed application of PCA method works especially well for the medium and the long term contracts and PCA model outperforms other approaches largely. Forecasting IVs of the short term options is more challenging. However, PCA model shows its superiority over other investigated approaches with respect to the short-term put options and some call options that are not too deep out-of-the-money. For the short term deep out-of-the-money call options, the parametric models of the surface work best. Inclusion of the factor dynamics in the parametric method sometimes improves the forecasting accuracy for longer maturities. I find that Heston and Nandi model, which is the only one estimated in option prices space instead of implied volatility space, works worst both in terms of in-sample fitting, as well as out-of-sample forecasting of IVS. However, relative deterioration when turning from the in-sample to the out-of-sample case is the smallest for this model. For the evaluation purposes, I partition the surface into 21 segments depending on moneyness-maturity characteristics and find that some segments of the surface are more predictable than others. This leads to the next part of the analysis which examines whether combination of different models can further improve on forecasting IVS. Forecasts combination attracted a lot of attention in empirical studies in numerous areas of economic research such as equity premium forecasting, currency market volatility and various macroeconomic applications. Many researches proved the usefulness of forecasts combination to generate more accurate forecasts, sometimes using even the simple model averaging. I conclude that use of the combining schemes, like the optimal estimated weights or discounted mean square error, can improve on forecasting the surface in all regions. Still, the improvement is not large enough to outperform simple random walk forecasts for implied

volatilities in any region of the surface.

The models of IVS I study in this thesis can be related to three approaches to modelling the implied volatility that can be distinguished in the literature. The factor models of IVS are inspired by the literature on the term-structure of the interests rates. Because there is a strong comovement in the different regions of the surface, latent factors approach has been investigated. There are few studies that examine the dynamics of IVS driven by latent factors estimated with PCA. Skiadopoulos et al. (1999) study the dynamics of the surface of S&P500 index options. Their approach is based on grouping the data in three different maturity buckets. Next, they average IVs which fall into them respectively to options' moneyness and apply PCA to each bucket's covariance matrix. An important disadvantage of this approach is that the common factors can be disturbed both by the within and between group variation, meaning that this approach fails to distinguish between the common and specific latent factors driving IVS. This is because the grouping approach neglects the surface structure of IV data and average the options within the maturity bucket. A more popular approach to PCA modelling of IVS, also undertaken in this thesis, is described in Cont and da Fonseca (2002), Fengler et al. (2003) and Chorro et al. (2014) among others. Here, the whole surface is estimated on a given day with a kernel smoothing procedure, namely Nadaraya-Watson regression. The time-series on a chosen grid of moneyness and maturity can be recovered from the estimated smoothed surfaces. The aforementioned studies focus on extracting and identifying statistical (latent) common factors, while they are not concerned whether the factors are predictable themselves, what would allow to accurately forecast IVS.

Recently, Van der Wel et al. (2015) proposed a likelihood-based general dynamic factor model for the dynamics of the latent factors driving IVS. They estimate the factors and their dynamics in one-step by means of Kalman filter. The other factor approaches studied in the context of IVS dynamics include semi-parametric factor model of Fengler et al. (2007) designed to deal with a special feature of IV data that typically options with only few maturities (5-8) are traded on each day, or the restricted factor model of Christoffersen et al. (2013) where factors represent the level, smile and term structure of IVS. le Roux (2007) proposes the model of S&P500 IVS that combines the parametric approach with the application of principal component analysis, designed to capture the long-term dynamics of IVS.

The other approach popular in the literature on the implied volatility is to assume a linear parametric structure of the surface. Dumas et al. (1998) and Pena et al. (1999) investigate various parametric forms of IVS that link the cross-section of implied volatilities to options' maturity and moneyness. Estimated coefficients in these models are treated as factors. The parametric approach is extended by Goncalves and Guidolin (2006) who introduce VAR

dynamics to the factors.

A number of option pricing models departure from the Black-Scholes assumption that the volatility is constant over time and try to reconcile the stylized facts of IVS such as smiles and the volatility term structure. These models incorporate the stochastic or time-varying volatility (e.g. Hull and White (1987), Heston (1993) Heston and Nandi (2000)), jumps processes (Merton (1976)), or the jump processes combined with the stochastic volatility (Bates (1996), Scott (1997)). However, it may be that even a decent option valuation model in terms of pricing performance can deliver inaccurate predictions of IVS. This is because small errors in option prices can produce large errors in implied volatilities, as investigated by Hentschel (2003). Das and Sundaram (1999) and Skiadopoulos et al. (1999) conclude that none of the aforementioned option pricing models captures the stylized facts of IVS well. This approach to modelling IVS is sometimes referred to as a *no-arbitrage* approach (see Chalamandaris and Tsekrekos (2010)) because option pricing models do not allow for the possibility of arbitrage. An unsatisfactory performance of option pricing models resulted in a development of models that are estimated in implied volatility space, which examples were given above.

Although much has been said on what determines the shape and dynamics of IVS and plausible interpretation of the factors has been proposed, the application to out-of-sample forecasting are relatively rare, especially when it comes to the latent factor models. Goncalves and Guidolin (2006) and Bernales and Guidolin (2014) investigate out-of-sample forecasting ability of the parametric models. An explicit out-of-sample forecasting approach based on the latent factors estimated with PCA is taken by Chalamandaris and Tsekrekos (2010) who study IVS dynamics of over-the-counter (OTC) currency options. In contrast to exchange traded index options, OTC options have constant time-to-maturity. The constant time-to-maturity feature elevates the need for constructing the artificial time-series (by smoothing or grouping procedure) that serve as an input to PCA. Harvey and Whaley (1992), Guo (2000) and Brooks and Oozeer (2002) examine predictable patterns for IVs of the short-term at-the-money (ATM) options rather than the whole implied volatility surface. Konstantinidi et al. (2008) address the question about predictability of the implied volatility from the perspective of European and U.S. implied volatility indices like the well-known CBOE Volatility Index (VIX). All the studies listed above recognize some statistically predictable patterns. However, the common conclusion is that they cannot be easily translated into an economic value, what corroborates the option market efficiency to some extent. The economic value of IVS predictability is assessed by constructing various trading strategies and investigating whether IVS forecasts can efficiently support the portfolio decisions.

The structure of the thesis is as follows. Chapter 2 presents the data and explains the

smoothing procedure that helps to understand and organize the data, and most importantly results in the input time-series to PCA model. Chapter 3 details the estimation procedure for each of the forecasting models. Chapter 4 describes combination models. Chapter 5 presents the in-sample and out-of-sample results. Chapter 6 concludes and outlines possible recommendations for future research.

Chapter 2

The implied volatility surface

2.1 The Data

The data set used in this study contains daily data on the S&P500 index options traded on the Chicago Board Options Exchange (CBOE). The S&P500 options are one of the most frequently traded options in the world. The data set covers period of almost 15 years, from January 4, 1999, until August 30, 2013. Options are European style and are retrieved from OptionMetrics.[1] The data set consists of 16 variables from which 6 are used in this study: (i) current date, (ii) time-to-maturity- κ, (iii) strike price- K, (iv) implied volatility- σ, (v) price of underlying S&P500 index- S and (vi) moneyness which is defined as $m = K/S$. In addition, I use data on LIBOR rates treated as a proxy for the risk-free rates and continuously compounded dividend yields on S&P500 index. Both variables are provided by OptionMetrics and are needed to implement the option pricing model of Heston and Nandi (2000) described in chapter 3.

The data are filtered based on six exclusionary criteria to ensure that the whole surface under consideration is active. Applied criteria follow the literature on implied volatility and options pricing. First, all the options with maturity less than 10 days are dismissed due to noisiness in their prices. Second, options with maturity greater than a year are also excluded. These two steps are similar to Dumas et al. (1998) and Bernales and Guidolin (2014) who argue that such options usually contain little information regarding IVS. Third, to avoid the problem of price discreteness, I omit options with prices less than 3/8$ following Bakshi et al. (1997) and Goncalves and Guidolin (2006). Fourth, similarly to Van der Wel et al. (2015) and Barone-Adesi et al. (2008), observations with missing values for IV and those with IV greater than 0.7 are discarded. Fifth, following Cont and da Fonseca (2002) and Van der Wel et al. (2015), I consider only out-of-the-money (OTM) calls and puts because

[1] I would like to thank prof. dr. Dick van Dijk for making this data set available to me.

they are more frequently traded than in-the-money (ITM) options and as such, contain more information about movements in the implied volatility surface. Focusing on OTM instead of ITM options is motivated by the fact, that ITM are illiquid and therefore their prices contain illiquidity premium. However, the put-call parity implies that taking into account OTM options is equivalent to studying ITM options. Every OTM call (put) can be matched to ITM put (call) where the Δ of the call option is always one plus the Δ of the put option and the corresponding put-call pair should have the same implied volatility. Sixth, I filter out options with moneyness outside the range of $m \in [0.5, 1.5]$ which follows Cont and da Fonseca (2002). All applied filters leaves us with 1,170,893 options, what gives 317 contracts on average per day.

2.2 Smoothing the surface

In order to obtain an arbitrary point on the implied volatility surface, one needs to interpolate or smooth the discrete data. This can be achieved in a non-parametric way. I follow approach by Cont and da Fonseca (2002) and Fengler et al. (2003) who apply non-parametric Nadaraya-Watson estimator to IV data. For a partition of explanatry variables (m,κ), where m is moneyness and κ is time-to-maturity, the two-dimensional Nadaraya-Watson estimator of the implied volatility σ is given by

$$\hat{\sigma}(m,\kappa) = \frac{\sum_{i=1}^{n} K_1\big((m-m_i)/h_1\big) K_2\big((\kappa-\kappa_i)/h_2\big)\sigma_i}{\sum_{i=1}^{n} K_1\big((m-m_i)/h_1\big) K_2\big((\kappa-\kappa_i)/h_2\big)}, \tag{2.1}$$

where σ_i is the observed value of IV, $K_1(u)$, $K_2(u)$ are univariate kernel functions, h_1 and h_2 are bandwidths in moneyness and maturity directions respectively, and n denotes the number of observations. I apply the estimator on the filtered data set. As a kernel function I use a quartic kernel, i.e.,

$$K(u) = \frac{15}{16}(1-u^2)^2 \mathbb{1}[|u| \leq 1], \tag{2.2}$$

where $\mathbb{1}[x]$ is an indicator function that equals 1 when x is true and 0 otherwise. Usually, the choice of the kernel function does not influence empirical results (see Silverman (1986)). However, in contrast to Cont and da Fonseca (2002) and Fengler et al. (2003) I conclude that the use of the quartic kernel is preferred over a Gaussian kernel, especially as the latter is incapable of producing a smoothed surface that fits accurately high values of IV which are observed for far out-of-the money put options. This is especially clear when when there is only a small number of maturities traded. Figure B.1 in Appendix B presents smoothed surface fitted to actual data for quartic and Gaussian kernels.

The bandwidths h_1 and h_2 control the level of smoothness of the estimated surface. Too low values of h_1, h_2 will make IVS bumpy, whereas too high values will lead to oversmoothing. As for the optimal bandwidths selection, I follow Fengler et al. (2003) and for each date in the sample I solve the following optimization problem:

$$\min_{h_1,h_2} \quad n^{-1}\sum_{i=1}^{n}\big(\sigma_i - \hat{\sigma}_{h1,h2}(m_i,\kappa_i)\big)^2 \times \Xi\big(n^{-1}h_1^{-1}h_2^{-1}K_1(0)K_2(0)\big), \tag{2.3}$$

where $\Xi(z) = \exp(2z)$ is the Akaike penalizing function. The Akaike correction factor penalizes bandwidths that are too small (for alternative choices of penalizing functions see Hardle (1990)). Next, I average the penalized bandwidths across observations dates. Given the fact that twice in the whole sample the number of available maturities of S&P500 index options increases dividing the sample into 3 periods, which happens on 21/Feb/2007 and 31/May/2012, I calculate the average optimal bandwidth separately for each of the three sub-periods. The selection procedure yields the average optimal bandwidth $h_1^* = 0.02$ in the moneyness dimension for all the three sub-periods. Average optimal bandwidth h_2^* largely differs between the sub-periods as it equals 22.3, 13.6 and 10.6 days for the observations before 21/Feb/2007, from 21/Feb/2007 until 31/May/2012 and after 31/May/2012 respectively. The sample standard deviation of penalized bandwidths in maturity dimension accounts for 8.1, 4.6 and 0.5 respectively. Low standard deviation of optimized bandwidths indicate that single bandwidth can be used for all estimation dates within abovementioned sub-periods.

However, figure B.2 in Appendix B shows that surfaces obtained with the optimized bandwidths are bumpy and discontinuous. It suggests that the optimal bandwidths with respect to penalizing function are too narrow in both dimensions. This observation is consistent with Fengler et al. (2003), who argue that data points appear like 'pearls in the necklace' in the three dimensional space of IVS. Moreover, they argue that penalizing approaches, as other cross-validation procedures, evaluate the quality of estimator right at the observed data points what results in too small bandwidths when one aims to obtain the estimates on the grid points deviating from actual observations, as it is the case in this thesis. Thus, oversmoothing with respect to the penalizing function cannot be avoided. In finally using bandwidths shown in Table 2.1 I take into account that the number of traded maturities daily differs between abovementioned sub-periods and also that for shorter maturities observations are closer to each other than for longer maturities. Such a choice of bandwidths ensures that the surface can be recovered everywhere on the constant grid. I recover IVS on a fixed grid of moneyness $m_i \in \{0.85, 0.90, 0.95, 0.99, 1.01, 1.05, 1.10, 1.12\}$ and maturity $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$. The smoothing procedure applied day by day results

in 64 time-series of smoothed IV $\hat{\sigma}_t(m_i, \kappa_j)$.

Table 2.1: Bandwidths used to recover the surface

	Maturity κ		
Dates range	**<60**	**60-180**	**>180**
04/Jan/1999 - 20/Feb/2007	h_1=0.08, h_2=40	h_1=0.08, h_2=80	h_1=0.08, h_2=145
21/Feb/2007 - 30/May/2012	h_1=0.08, h_2=35	h_1=0.08, h_2=60	h_1=0.08, h_2=100
31/May/2012 - 30/Aug/2013	h_1=0.08, h_2=30	h_1=0.08, h_2=55	h_1=0.08, h_2=80

The table presents values of the bandwidths that I use to recover the surface on the constant grid of moneyness $m_i \in \{0.85, 0.9, 0.95, 0.99, 1.01, 1.05, 1.1, 1.12\}$ and maturity $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$. h_1 is the bandwidth in the moneyness dimension, h_2 is the bandwidth in the maturity dimension. Differentiating h_2 between maturity groups and three date ranges allows to take into account the number of traded maturities daily and the fact that for shorter maturities observation are closer to each other than for longer term contracts.

2.3 Summary statistics

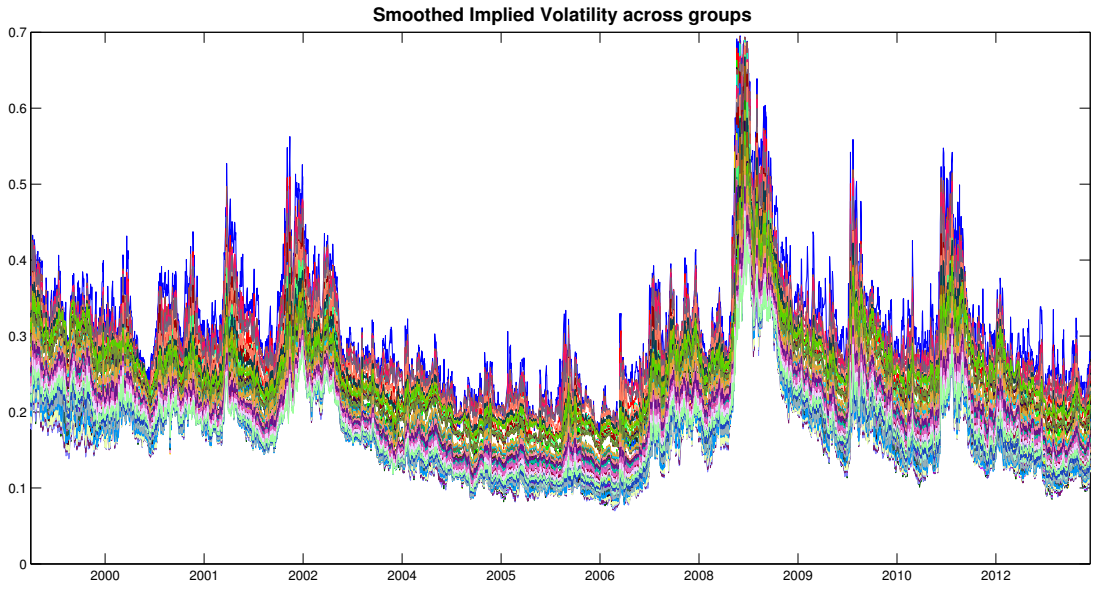
Table 2.2 reports summary statistics for implied volatilities. For the purpose of presenting summary statistics, I divide the data into 3 maturity and 9 moneyness categories. Grouping based on the moneyness level corresponds to 8 grid points in the moneyness dimension selected to recover the surface with the kernel procedure, while maturity categories follow Bakshi et al. (1997), who classify contracts as short term if $\kappa < 60$, medium term if $60 \leq \kappa \leq 180$ and long term if $\kappa > 180$.

The mean values of IV form a pronounced volatility smile, i.e. IV first declines and again inclines when moving from low to high values of m. It is especially visible pattern for the short maturities, while for the medium and the long maturities IV starts to incline again only for very high values of m. The term structure, the pattern for given moneyness across different maturities, is fairly flat, except extreme moneyness groups ($m < 0.85$ or $m > 1.12$) where mean values of IVs are considerably higher for the short maturities than for the medium and the long maturities. The next observed feature is that for a given moneyness group IVs are more volatile for the short term than for the medium and the long term options. Also, contracts that are closer to being at-the-money are less volatile than those with m further from 1. All moneyness-maturity groups exhibit positive skewness, which is a well-known characteristic of implied volatility distribution. It is usually the largest for the short maturities given the moneyness group. The logarithm of IV data exhibits much lower skewness, however after logarithmic transformation the data are still slightly positively skewed. The distributions of the moneyness-maturity groups are leptokurtic, with kurtosis

far above 3[2]. Only short-term put options with $m < 0.85$ and short-term call options with $m > 1.12$ have kurtosis below 3.

Figure 2.1 presents 64 time series obtained from the smoothing procedure. They exhibit very strong comovement. To analyze a degree of this comovement and its persistence in

Figure 2.1: Time series of smoothed IVs on a given grid



IVS is recovered with Nadaraya-Watson estimator (equation (2.1)) on the fixed grid of moneyness $m_i \in \{0.85, 0.90, 0.95, 0.99, 1.01, 1.05, 1.10, 1.12\}$ and maturity $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$. The figure shows 64 time series of $\hat{\sigma}_t(m_i, \kappa_j)$.

the implied volatility surface, tables C.1 and C.2 in Appendix C report cross-correlations and (partial) autocorrelations for the investigated time series. Table C.1 shows the cross-correlations between number of moneyness categories and two maturity categories: the shortest of $\kappa = 20$ and the longest of $\kappa = 320$. Cross-correlations within the maturity category are stronger for the long than for the short maturities. All of them are above 0.9. Even across maturity categories only a few of cross-correlations fall below 0.9. Table C.2 reports (partial) autocorrelations at lags of 1 to 5 days and 1 month, i.e. 22 days, for different moneyness-maturity points. The autocorrelations are higher for the long maturities at every lag. At 22 days lag they stay at around 0.9 for the long maturities and 0.8 for the short maturities. Also partial autocorrelations are higher for the long maturities. However, at the third lag almost all of them are below 0.1. The OTM put options ($m < 1$) exhibit larger drop in partial autocorrelation than OTM calls ($m > 1$).

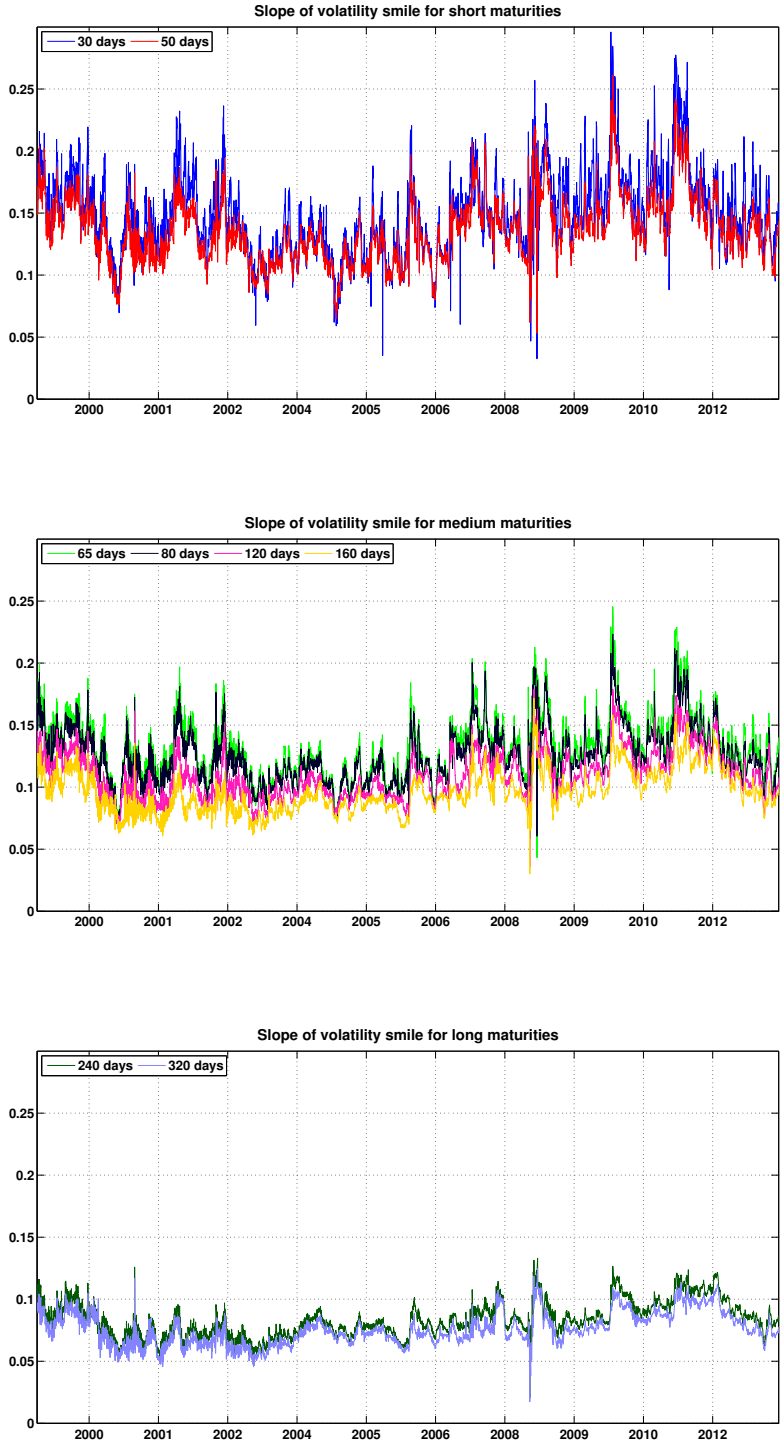
[2]Kurtosis of 3 corresponds to the kurtosis of normal distribution

To further illustrate the dynamics of the surface, I construct measures of the volatility smile and term structure following Van der Wel et al. (2015). Defined measures correspond to the grid points of the smoothed surface. The slope of the volatility smile corresponding to each maturity grid point $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$ is defined as the implied volatility of the furthest OTM put option on the grid for which $m = 0.85$, minus the implied volatility of the furthest OTM call option on the grid for which $m = 1.12$. Similarly, the slope of the volatility term structure corresponding to each moneyness grid point $m_i \in \{0.85, 0.90, 0.95, 0.99, 1.01, 1.05, 1.10, 1.12\}$ is defined as the implied volatility of the longest maturity on the grid $\kappa = 320$, minus the implied volatility of the shortest maturity on the grid $\kappa = 30$.

Figure 2.2 shows that the slope of the smile is positive for all the considered maturity grid points. On average, it achieves higher levels and is more volatile for the short maturities. The smile is larger when the volatility of the financial markets is higher, as measured with VIX index which behaviour is shown in the top panel of Figure 2.3. Figure 2.3 shows time series of the slope of the term structure separately for the put and the call options.

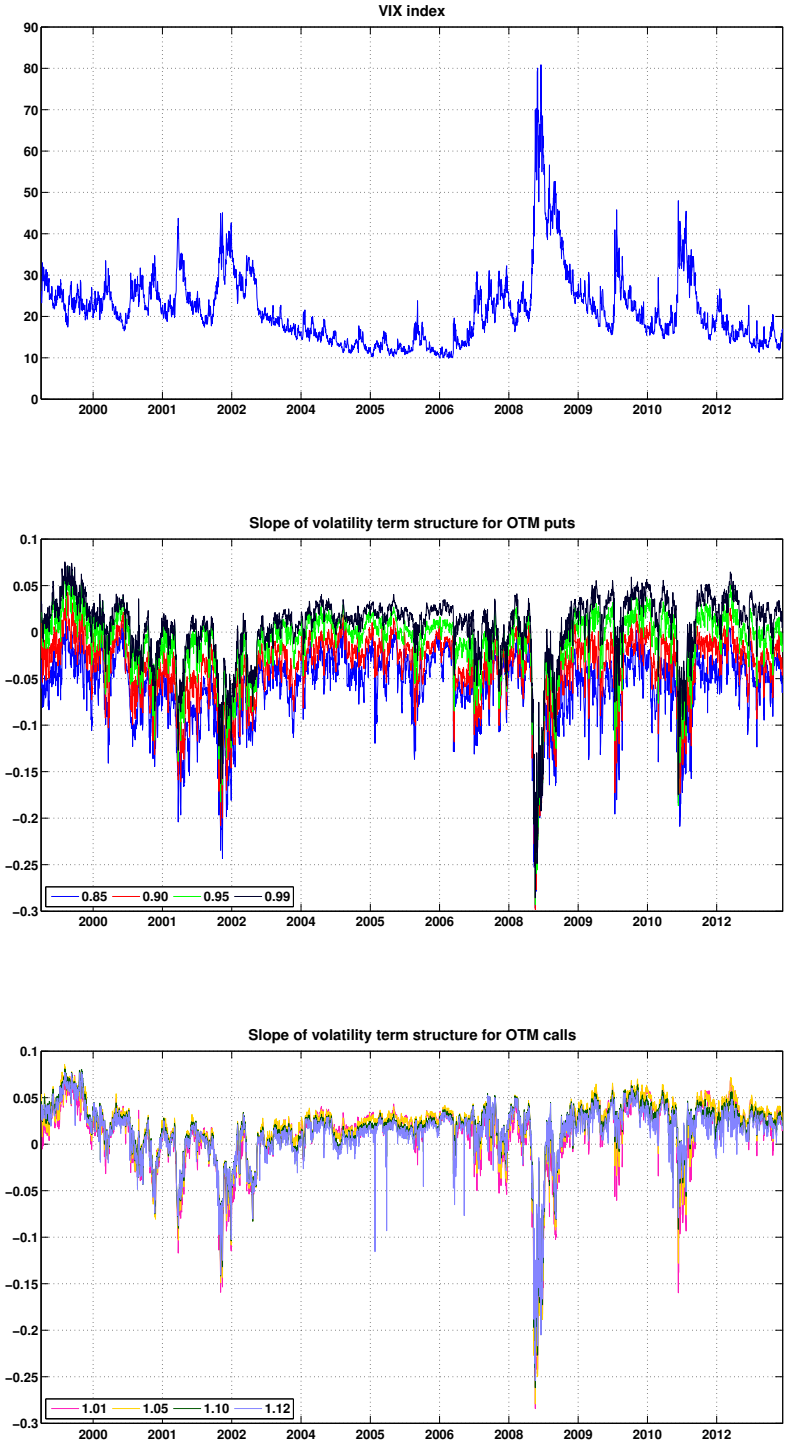
The term structure can be upward- or downward-sloping for all considered moneyness grid points. When the overall level of volatility is high, as measured with VIX index, the term structure tends to be downward-sloping and reach especially low values, hitting the record low in the crisis of 2008. On the other hand, it is upward-sloping when the volatility is low. A popular explanation for this phenomenon in the term structure of IVS is the mean-reverting nature of volatility. Most of the time in the sample the slope of the term structure is negative for the put options with $m \in \{0.85, 0.9, 0.95\}$, while it is positive for the at-the-money put with $m = 0.99$ and all the call options. The slope of the term structure vary across moneyness grid points when $m < 1$, while it is fairly on the same level across all the considered points when $m > 1$. Table C.3 in Appendix C contains (partial) autocorrelations for the slope of the smile and term structure. They exhibit similar pattern as the (partial) autocorrelations of the time series shown in Table C.2. However, the partial autocorrelation of the slope of the volatility smile decays slower, accounting for around 0.2 for the long maturities at the third lag and falling below 0.1 only after the fourth lag.

Figure 2.2: Slope of Volatility Smile



The figure shows VIX index and the slope of the volatility smile. The slope of the smile corresponding to each maturity grid point $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$ is defined as IV of the option with $m = 0.85$ minus IV of the option with $m = 1.12$. Smoothed IV values are used.

Figure 2.3: Slope of Volatility Term Structure



The figure shows VIX index and the slope of the volatility term structure. The slope of the term structure corresponding to each moneyness grid point $m_i \in \{0.85, 0.9, 0.95, 0.99, 1.01, 1.05, 1.1, 1.12\}$ is defined as IV of the option with $\kappa = 320$ minus IV of the option with $\kappa = 30$. Smoothed IV values are used.

Table 2.2: Summary statistics by Moneyness and Maturity

Moneyness		Maturity		
K/S		<60	60-180	>180
<0.85	**Mean**	0.41	0.36	0.31
	Std	0.11	0.09	0.07
	Skewness	0.68	0.75	0.67
	Skewness $\log\sigma$	0.19	0.14	0.03
	Kurtosis	2.72	3.40	3.67
	Observations	75645	143008	106754
0.85-0.90	**Mean**	0.30	0.26	0.25
	Std	0.08	0.07	0.06
	Skewness	1.54	1.52	1.28
	Skewness $\log\sigma$	0.71	0.64	0.45
	Kurtosis	6.25	6.78	6.08
	Observations	50309	39162	24616
0.90-0.95	**Mean**	0.25	0.23	0.23
	Std	0.08	0.07	0.06
	Skewness	1.90	1.59	1.27
	Skewness $\log\sigma$	0.75	0.59	0.39
	Kurtosis	8.64	7.15	6.14
	Observations	66376	44087	25963
0.95-0.99	**Mean**	0.20	0.21	0.21
	Std	0.08	0.07	0.06
	Skewness	2.11	1.57	1.17
	Skewness $\log\sigma$	0.76	0.48	0.31
	Kurtosis	10.02	7.24	5.65
	Observations	60547	39797	21328
0.99-1.01	**Mean**	0.18	0.19	0.20
	Std	0.07	0.07	0.06
	Skewness	2.15	1.59	1.23
	Skewness $\log\sigma$	0.70	0.46	0.34
	Kurtosis	10.35	7.36	5.90
	Observations	31706	21575	10966
1.01-1.05	**Mean**	0.16	0.18	0.19
	Std	0.07	0.06	0.05
	Skewness	2.33	1.57	1.19
	Skewness $\log\sigma$	0.74	0.41	0.32
	Kurtosis	11.51	7.14	5.56
	Observations	59418	41304	21413
1.05-1.10	**Mean**	0.18	0.17	0.18
	Std	0.08	0.06	0.05
	Skewness	2.24	1.75	1.22
	Skewness $\log\sigma$	0.76	0.52	0.25
	Kurtosis	10.12	7.79	5.79
	Observations	39185	43797	25183
1.10-1.12	**Mean**	0.21	0.17	0.18
	Std	0.09	0.06	0.05
	Skewness	2.06	1.96	1.26
	Skewness $\log\sigma$	1.06	0.74	0.27
	Kurtosis	7.93	8.50	5.82
	Observations	7348	14053	9230
>1.12	**Mean**	0.30	0.20	0.18
	Std	0.12	0.07	0.05
	Skewness	0.78	1.27	1.41
	Skewness $\log\sigma$	0.20	0.49	0.48
	Kurtosis	2.87	4.48	5.52
	Observations	17595	56145	74383

The sample covers period from January 4, 1999, until August 30, 2013 for a total of 1,170,893 options after filtering the data. Six summary statistics of IVs are reported: mean, standard deviation (Std), skewness, skewness of log data, kurtosis, and the number of observations within each group.

Chapter 3

Forecasting models

In this chapter I describe 4 different set-ups within 3 main approaches that are employed to forecasts the implied volatility surface. The three approaches are: (i) a model of latent factors that drive IVS, where the factors are estimated with Principal Component Analysis, (ii) two models that assume parametric structure of the surface, (iii) option pricing model for which I choose the model of Heston and Nandi (2000). To preserve consistency, all the first three models estimated in IV space deal with the log transformation on IV data. It gives an advantage that the models always produce non-negative values of IV. At the end of the day, I asses forecasts of IV (without log transformation) such that all four models can be compared. I use the first 1000 days of the data set as an initial estimation period. All the models are estimated using the rolling window framework with the window length of 1000 days. It ensures that the forecasts generated by the different models are conditional on the same information set.

3.1 Principal Component Analysis

The method of Principal Component Analysis (PCA) allows to summarize the main sources of variation and covariation for a large panel of variables into a small number of underlying risk factors. Thus, PCA allows to greatly reduce the dimensionality of the problem under consideration. It may be especially efficient when the investigated time-series are highly correlated, as is the case in this study. I apply PCA technique to extract common factors from the panel of 64 time-series obtained with kernel regression of section 2.2.

I calculate the first differences of the log transformed data $\Delta \log \hat{\sigma}_t(m_i, \tau_j)$ to ensure stationarity of the input variables to PCA model, following works of Cont and da Fonseca (2002), Fengler et al. (2003), Badshah (2009) and Skiadopoulos et al. (1999) who also perform PCA on the first differences of (log) IV data[1]. The smoothed IV time-series illustrated in figure 2.1 suggest that they may contain a unit-root. I test unit-root stationarity of the log time-series with an augmented Dickey-Fuller test. I select a lag order of the test based on the BIC criterion with the maximum value of 5. To illustrate the results of ADF test, I report t-statistics and p-values for 12 time series in total in table C.5 in Appendix C: 6 log and 6 differenced log time series. The test is performed on the time-series having length of 1000 days, what is consistent with the estimation window of PCA model. In each case the null hypothesis that the time series of $\log \hat{\sigma}_t(m, \kappa)$ contain unit root cannot be rejected at any standard significance level, whereas I reject the null for all $\Delta \log \hat{\sigma}_t(m, \kappa)$ being tested with p-value less than 0.001. These results support the decision of modelling the first differences of the smoothed IV time-series. Alternatively, one could work with IV data in levels when the idiosyncratic noise components ϵ_{it} in a factor representation equation are $I(0)$. This is motivated by the fact that PCA estimators of factors and factor loadings are consistent as long as ϵ_{it} are $I(0)$, regardless of stationarity of the factors, as pointed out by Bai and Ng (2004). For more detailed discussion of this approach in the context of IVS see Chalamandaris and Tsekrekos (2010).

I consider static factor representation of the log smoothed implied volatilities $\Delta \log \hat{\sigma}_t(m_i, \tau_j)$:

$$\Delta \log \hat{\sigma}_t(m_i, \tau_j) = \lambda_{ij1} F_{1t} + \cdots + \lambda_{ijr} F_{rt} + \epsilon_{ij,t} = \boldsymbol{\lambda}_{\boldsymbol{ij}}^{'} \boldsymbol{F_t} + \epsilon_{ij,t}, \tag{3.1}$$

where $\boldsymbol{F}_t$ is a vector of common factors, $\boldsymbol{\lambda}_{ij}$ is a vector with factor loadings for the first differences of the log smoothed implied volatility $\hat{\sigma}$ corresponding to i-th moneyness and j-th maturity grid point, and ϵ_{ij} is an idiosyncratic noise. The factors and idiosyncratic

[1] In unreported results, I examined the out-of-sample performance of PCA model estimated in levels. It yielded slightly poorer results. Moreover, the interpretation of the factors estimated on differenced data is clearer.

components are assumed to be uncorrelated, $\mathrm{E}(F_t\epsilon_{ij,s}) = 0$ for all i, j, s, that is they are uncorrelated at all leads and lags. In the "approximate dynamic factor model", as introduced by Chamberlain and Rothschild (1982), weak cross-section correlation between idiosyncratic components is allowed, as long as $\frac{1}{N}\sum_{k=1}^{N}\sum_{l=1}^{N} \mid \mathrm{E}(e_{k,t}e_{l,t}) \mid$ is bounded.

Let $\boldsymbol{X}$ be $N \times T$ matrix of N time-series having length T, $\hat{\Sigma}_{\boldsymbol{XX}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{X}_t\boldsymbol{X}_t'$ be its sample covariance matrix and $\boldsymbol{\Lambda}$ $N \times r$ matrix with r factors' loadings. As described by Stock and Watson (2006) the estimation of $\boldsymbol{\Lambda}$ and $\boldsymbol{F}_t$ can be considered as solving the following non-linear least square problem:

$$\min_{\boldsymbol{F_1},\cdots,\boldsymbol{F_T},\boldsymbol{\Lambda}} T^{-1}\sum_{t=1}^{T}\big(\boldsymbol{X}_t - \boldsymbol{\Lambda}\boldsymbol{F}_t\big)'\big(\boldsymbol{X}_t - \boldsymbol{\Lambda}\boldsymbol{F}_t\big), \tag{3.2}$$

subject to normalisation $\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = I_r$. One possible solution, however not unique, to the above optimization is $\left(\hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{F}}\right)$, where $\hat{\boldsymbol{\Lambda}}$ is set to be the first r eigenvectors corresponding to the r largest eigenvalues of the covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{XX}}$. The estimator of the factors is $\hat{\boldsymbol{F}}_t = \hat{\boldsymbol{\Lambda}}'\boldsymbol{X}_t$, which is the vector consisting of the first r principal components (PCs) of $\boldsymbol{X}_t$.

The standard practice in the literature on the dynamics of the implied volatility surface is to perform PCA on the covariance matrix. However, given that variances of IVs vary across different segments of the surface, as reported in table C.1 in Appendix C, it seems to be reasonable to perform PCA on the correlation matrix instead of the covariance matrix, so that the first PC is not attributed to the few moneyness/maturity groups with the largest variance. When the principal components are extracted from the correlation matrix instead of the covariance matrix, the data are effectively standardized, i.e. z-scores $z_{it} = \frac{x_{it}-\mu_i}{s_i}$ are used, where μ_i and s_i are the mean and the standard deviation of the variable x_i. To find the forecasts $\Delta\log\hat{\sigma}_{t+1|t}(m_i, \tau_j)$ in equation (3.5), I multiply the forecasted z-scores by s_i and add μ_i, that is I 'unstandardize' the data. The rest of the procedure is the same as in the case of non-standardized data. The standardization and 'unstandardization' steps are omitted in the equations' notation for clarity of reading.

Since the primary goal of this study is the out-of-sample forecasting of IVS in contrast to in-sample fitting, I follow common rule of thumb in choosing the number of static factors and set it to be $\hat{r} = 3$. However, there is a vast empirical evidence that 2 to 3 factors drive the dynamics of IVS of index options (see Skiadopoulos et al. (1999), Mixon (2002) and Van der Wel et al. (2015)) or options on futures (Tompkins (2001)). Table 3.1 reports the proportion of explained variance in the full-sample for different kinds of input variables. For the non-stationary panel of $\log\hat{\sigma}$ the first PC explains over 97% of the total variance. The differences between the differenced log-data and its standardized version are only minor. In both cases the first 3 principal components explain over 95% of the variance.

Table 3.1: Principal Component Analysis

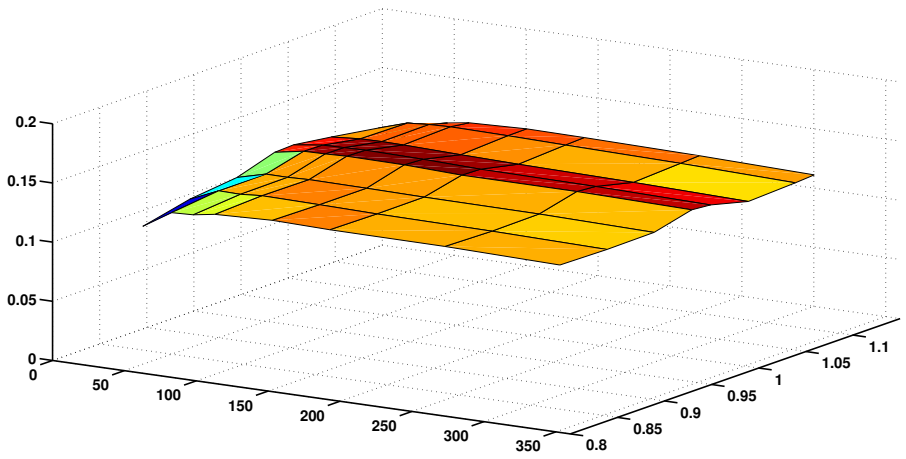
Explained variance						
	$\log \hat{\sigma}$		$\Delta \log \hat{\sigma}$		Z-scores of $\Delta \log \hat{\sigma}$	
	%	Cum. %	%	Cum. %	%	Cum. %
PC 1	97.18	97.18	84.56	84.56	84.51	84.51
PC 2	1.47	98.66	7.57	92.13	8.53	93.03
PC 3	0.86	99.52	3.12	95.25	2.28	95.32
PC 4	0.16	99.67	1.45	96.70	1.23	96.54
PC 5	0.08	99.76	0.80	97.50	0.78	97.32
PC 6	0.08	99.83	0.54	98.03	0.42	97.74
PC 7	0.03	99.86	0.32	98.35	0.29	98.03
PC 8	0.03	99.89	0.23	98.58	0.24	98.27
PC 9	0.02	99.90	0.18	98.77	0.21	98.48
PC 10	0.01	99.92	0.15	98.92	0.19	98.67

The table presents the results of the Principal Component Analysis performed on the full sample of the panel of 64 times-series representing IVS on the fixed grid of moneyness $m_i \in \{0.85, 0.90, 0.95, 0.99, 1.01, 1.05, 1.10, 1.12\}$ and maturity $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$.

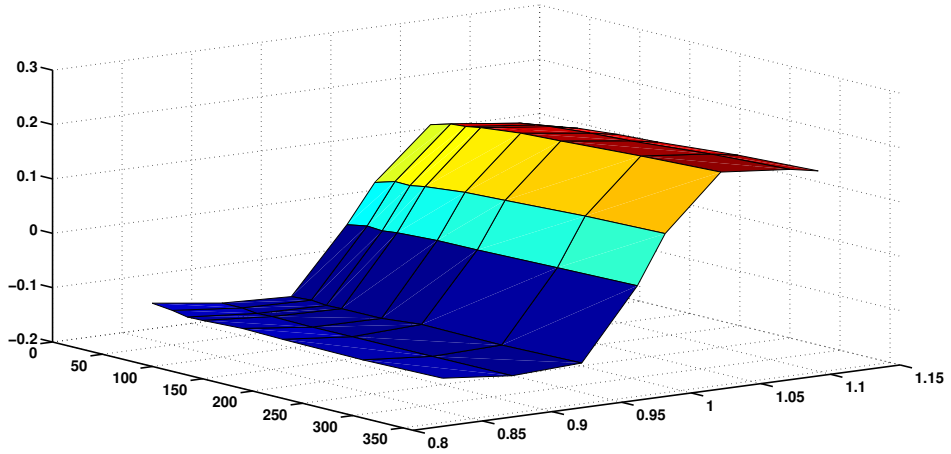
Because further analysis is conducted on the standardized data, I provide possible interpretation of the factors based on the factors' loadings estimated on the panel of Z-scores of $\Delta \log \hat{\sigma}$. The first principal component corresponds with the overall level of IVS, the elements of the first factor loading have approximately the same value for all 64 combinations of moneyness and maturity grid points. The second principal component affects the implied volatility of the call and the put options with a different sign, while the change of the sign takes place between moneyness grid points of 0.99 and 1.01 (ATM put and calls), irrespective of the remaining time-to-maturity. In fact, the effect is almost uniform across the maturity dimension. Thus, the second principal components is a *smile* factor, capturing the volatility smile. Finally, the third principal component corresponds with the volatility term structure, as it changes the sign of how it impacts IV between the maturity of 80 and 120 days and the effect is very similar across the moneyness dimension. Figure 3.1 provides a graphical illustration of the estimated factors' loadings.

Figure 3.1: Graphical illustration of the factors' loadings of the correlation matrix

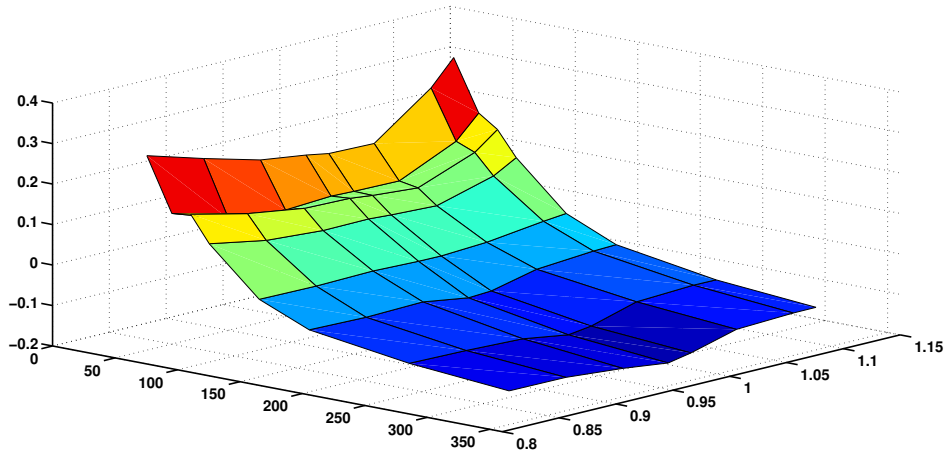
Panel A: First factor's loadings- level factor



Panel B: Second factor's loadings- smile factor



Panel C: Third factor's loadings- term structure factor



The figure shows estimates of the factors' loadings resulted from PCA performed on the full sample of 64 time-series, recovered with Nadaraya-Watson regression on the fixed grid of moneyness $m_i \in \{0.85, 0.90, 0.95, 0.99, 1.01, 1.05, 1.10, 1.12\}$ and maturity $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$.

In order to capture dynamics of the estimated factors $\hat{\boldsymbol{F}}_{\boldsymbol{t}} = (\hat{F}_{1t}, \hat{F}_{2t}, \hat{F}_{3t})'$, I assume they follow p-order Vector AutoRegressive process (VAR) given by equation (3.3). VAR specification can be augmented with exogenous regressors in a form of lagged returns on S&P 500 index, resulting in VARX model for factor dynamics. Chalamandaris and Tsekrekos (2009), who model in levels time-series of implied volatilities of options that have a constant time-to-maturity, show that VAR specification for factor dynamics outperforms both in-sample and out-of-sample univariate AR and VECM specifications for factors driving IVS of OTC currency options.

$$\hat{\boldsymbol{F}}_t = \boldsymbol{c} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j \hat{\boldsymbol{F}}_{t-j} + \sum_{k=1}^{q} \boldsymbol{\Psi}_k r_{SPX,t-j} + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\hat{F}\hat{F}}) \tag{3.3}$$

Equation (3.3) is estimated with MLS estimator. The order of lags p and q is selected based on BIC criterion and repeated in every estimation window. I perform the estimation of VARX process in (3.3) using the moving window of length 1000 of days. Predictability of IVS requires that the factors $\hat{\boldsymbol{F}}_t$ are predictable itself. I use VARX(p,q) specification to produce direct forecasts of $\hat{\boldsymbol{F}}_{t+1}$ which are constructed as

$$\hat{\boldsymbol{F}}_{t+1|t} = \boldsymbol{c} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j \hat{\boldsymbol{F}}_{t-j+1} + \sum_{k=1}^{q} \boldsymbol{\Psi}_k r_{SPX,t-j+1}, \tag{3.4}$$

which allows to calculate forecasts of the differenced smoothed log IVs $\Delta \log \hat{\sigma}_{t+1|t}(m_i, \tau_j)$ as

$$\Delta \log \hat{\sigma}_{t+1|t}(m_i, \tau_j) = \hat{\boldsymbol{\lambda}}_{ij}^{'} \hat{\boldsymbol{F}}_{t+1|t}. \tag{3.5}$$

Next, forecasts of $\log \hat{\sigma}_{t+1|t}(m_i, \tau_j)$ are simply

$$\log \hat{\sigma}_{t+1|t}(m_i, \tau_j) = \log \hat{\sigma}_t(m_i, \tau_j) + \Delta \log \hat{\sigma}_{t+1|t}(m_i, \tau_j) \tag{3.6}$$

Finally, I convert log data to standard implied volatilities. In this step I make use of the property of log-normal distribution which states that when $X \sim LN(a, b^2)$ its expected value equals $\exp(a + \frac{1}{2}b^2)$. Because it was implicitly assumed that conditional distribution of $\log \hat{\sigma}_{t+1|t}(m_i, \tau_j)$ is normal (because of distributional assumptions in equations 3.1 and 3.3) it follows that:

$$\hat{\sigma}_{t+1|t}(m_i, \tau_j) = \exp\left(\log \hat{\sigma}_{t+1|t}(m_i, \tau_j) + \frac{1}{2} \mathrm{Var}_t(\log \hat{\sigma}_{t+1}(m_i, \tau_j)) \right), \tag{3.7}$$

where $\mathrm{Var}_t(\log \hat{\sigma}_{t+1}(m_i, \tau_j))$ corresponds to the appropriate diagonal element of the variance-covariance matrix of the $\log \hat{\boldsymbol{\sigma}}_{t+1|t}$ which arises from the sum the variance of the factors $\boldsymbol{\Sigma}_{\hat{F}\hat{F}}$,

and the variance of the idiosyncratic components $\mathbf{\Sigma}_{\epsilon\epsilon}$:

$$\mathrm{Var}_t(\log \hat{\boldsymbol{\sigma}}_{t+1}) = \mathbf{\Lambda}\mathbf{\Sigma}_{\hat{F}\hat{F}}\mathbf{\Lambda}' + \mathbf{\Sigma}_{\epsilon\epsilon}. \tag{3.8}$$

Now, the whole IVS has to be recovered from the 64 point forecasts of $\hat{\sigma}_{t+1|t}(m_i, \tau_j)$ with moneyness $m_i \in \{0.85, 0.9, 0.95, 0.99, 1.01, 1.05, 1.1, 1.12\}$ and maturity $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$. To obtain forecasts of IVS, I again use the approach presented in section 2.2, that is the non-parametric Nadaraya-Watson regression to recover the surface in an arbitrary number of points. Table 3.2 presents the bandwidth I use to recover the surface on the grid made of options' time-to-maturity and predicted moneyness. To calculate predicted moneyness, I assume that the best forecast of today's index level is its current value. The choice of the bandwidths is now less complicated than it was in section 2.2, as the forecasted data points are regularly spaced in the three dimensional space of the implied volatility surface.

Table 3.2: Bandwidths used to recover the surface

Maturity κ		
<60	**60-180**	**>180**
h_1= 0.05, h_2=30	h_1= 0.05, h_2=40	h_1= 0.05, h_2=60

The table presents the bandwidths used to recover the surface from the forecasted points that lie on the fixed grid of moneyness $m_i \in \{0.85, 0.90, 0.95, 0.99, 1.01, 1.05, 1.10, 1.12\}$ and maturity $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$. The surface is recovered on the grid made of options' time-to-maturity and predicted moneyness.

3.2 Parametric VARX(p, q)

This section describes a deterministic model for IVS, later referred to as parametric VARX model. It consists of two steps that combine cross-sectional fitting of a parametric function to IVS observed on a given day, with the application of Vector AutoRegregressive model with eXogenous regressors (VARX) to the multivariate time-series of estimated in the first step daily coefficients. The exogenous information I include is in the form of lagged returns on S&P500 index. The motivation behind this step is that increased volatility often follows large negative returns in the equity market. The models of this type are commonly called "deterministic", as all the explanatory variables are observable and formed of basic option parameters. I implement the specification successfully applied by Goncalves and Guidolin (2006) and Bernales and Guidolin (2014). They consider and compare in the context of out-of-sample forecasting various parametric IVS specifications presented by Dumas et al. (1998) and Pena et al. (1999). Competing specifications which they compare belong to a general class of polynomials, where the implied volatility is a function of polynomials in moneyness and time to expiration (or functions thereof). The approach presented in this section is also similar to Diebold and Li (2006) who follow two-stage approach in modelling and forecasting the yield curve. First, they impose a parametric structure on the yield curve. Next, they study dynamics of the estimated parameters assuming that they follow VAR(1) process.

In the first step, each day I fit a parametric curve of the following form by ordinary least squares:

$$\log \sigma_i = \beta_0 + \beta_1 M_i + \beta_2 M_i^2 + \beta_3 \frac{\tau_i}{360} + \beta_4 (M_i \times \frac{\tau_i}{360}) + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \delta), \tag{3.9}$$

where ε_i is a random error term assumed to be normally distributed with mean 0 and variance δ, τ_i is time-to-maturity measured in days and divided by 360 in order to be expressed in years, and M_i is a time-adjusted moneyness defined as $M_i \equiv \frac{\log m_i}{\sqrt{\tau_i/360}}$. According to this measure of moneyness, the more time remains to maturity, the larger the difference should be between the strike and the spot price in order for it to have the same normalized maturity as compared to a short term option (see e.g. Tompkins et al. (2001)).

Estimating equation (3.9) for each day in the sample results in the 5-dimensional time series of $\hat{\boldsymbol{\beta}}_t = (\hat{\beta}_{0t}, \hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t}, \hat{\beta}_{4t})'$. Figure 3.2 depicts the time series of daily beta coefficients $\hat{\boldsymbol{\beta}}_{\boldsymbol{t}}$. The daily coefficients are highly unstable over time, implying the instability of IVS in both dimensions. The relative importance in determining the values of IV of each β factor varies over time. For a better understanding what this time-variation implies an interpretation of the factors is needed. If the assumption of the Black-Scholes model of a constant volatility held, the intercept β_0 in equation (3.9) would be equal to the common

for all options log-volatility, while β_j for $j = 1, ..., 4$ would be equal to 0. The intercept β_0 corresponds to the overall level of IVS as it tends to increase when the volatility in the financial markets increases, reaching extraordinary high levels in the financial crisis of 2008. The slope of the volatility smile can be characterized by β_1, β_2 captures the curvature of the smile, β_3 is the factor representing the slope of the implied volatility term structure, and β_4 captures interactions between the two dimensions in which IV is described i.e. time-to-maturity and moneyness. The evolution of the level factor resembles the time-series of 2.3 exhibited in figure 2.3 of chapter 2. The level factor β_0 is also highly positively correlated with β_1 factor which determines the slope of the volatility smile. In fact, the correlation coefficient for these two variables amounts to 0.95. This correlation implies that when the average volatility level is high, the slope of the smile is steeper. On the other hand, the level factor is negatively correlated with β_3 factor that account for the slope of the volatility term structure. Both results are consistent with the observations made in section on the summary statistics. The correlation coefficient between β_0 and β_1 is 0.84, while between β_0 and β_3 stands at -0.83. The next pattern that can be noticed is high positive correlation between the slope and the curvature of the volatility smile.

To capture joint dynamics of the time-series in $\hat{\boldsymbol{\beta}}_t$ I model them with VARX(p, q) model, where p represents the number of lags for the vector of lagged dependent variables, and q is the number of lags of the exogenous regressor. I decide to model the dynamics of $\hat{\boldsymbol{\beta}}_{\boldsymbol{t}}$ in levels, similarly to works of Goncalves and Guidolin (2006), Bernales and Guidolin (2014) and Diebold and Li (2006). This choice is purely pragmatic, as in unreported results I also studied dynamics of $\hat{\boldsymbol{\beta}}_t$ with VARX model, where the non-stationary time-series in $\hat{\boldsymbol{\beta}}_t$ were transformed to stationary time-series by differencing operation. This form of the model yielded poorer out-of-sample performance. I tested the unit-root stationarity of input variables to VARX model with an augmented Dickey-Fuller (ADF) test for a unit root in a univariate time-series, which I perform separately for each time-series of the beta coefficients over estimation windows of 1000 days. The lag order of the test is selected such that it minimizes Bayesian Information Criterion. According to ADF test, a null hypothesis that time-series contain the unit root can be rejected only for $\hat{\beta}_{2t}$, $\hat{\beta}_{3t}$, $\hat{\beta}_{4t}$. Table C.4 in Appendix C reports the results of ADF test for two estimation windows: the first and the last window in which the individual models are estimated.

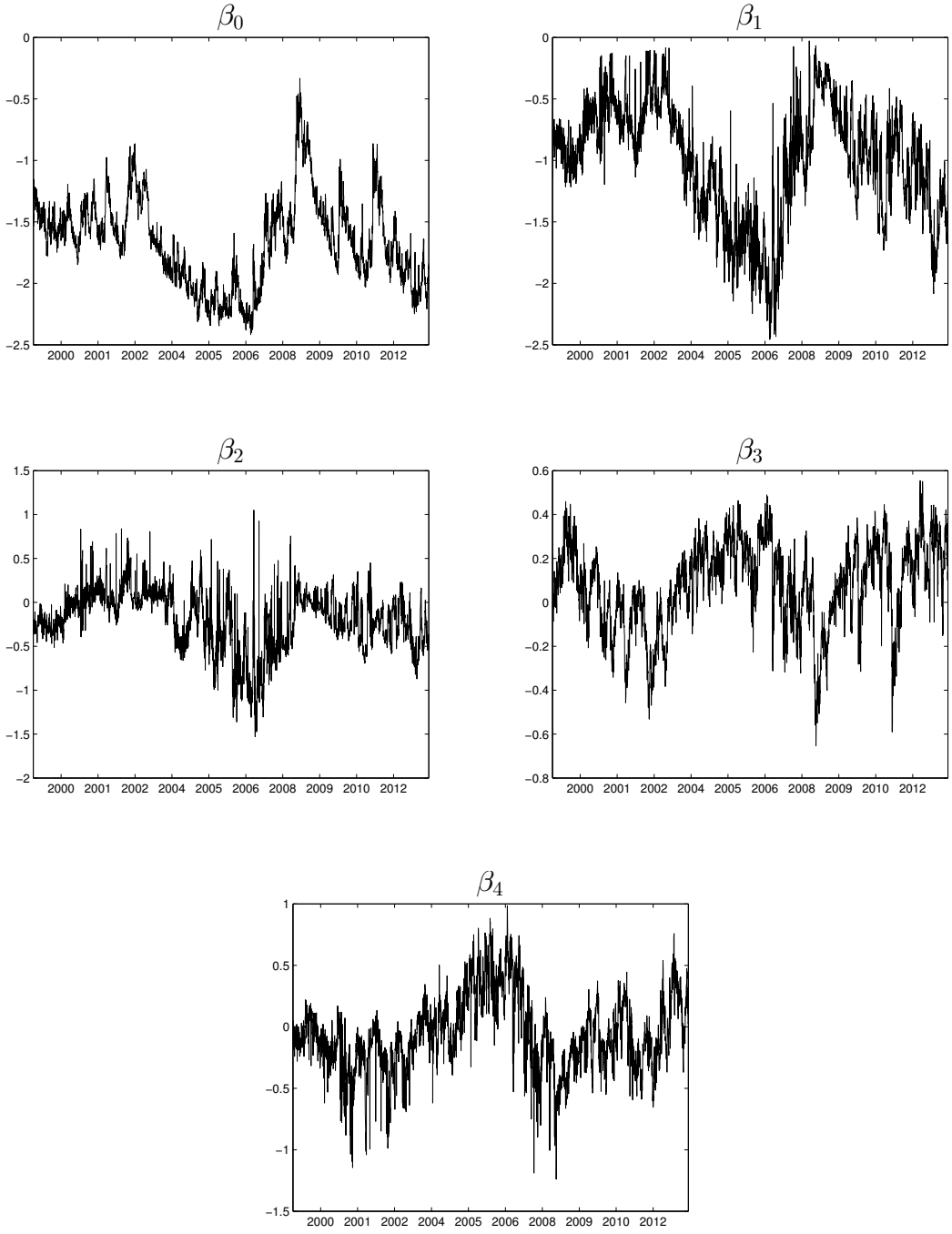
I assume the beta coefficients evolve over time according to the following equation:

$$\hat{\boldsymbol{\beta}}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j \hat{\boldsymbol{\beta}}_{t-j} + \sum_{k=1}^{q} \boldsymbol{\Psi}_k r_{SPX,t-j} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_{\hat{\boldsymbol{\beta}}}), \tag{3.10}$$

where $\hat{\boldsymbol{\beta}}_t$ is 5×1 time series of estimated coefficients in equation (3.9), $\boldsymbol{\mu}$ is 5×1 vector of constant terms, $\boldsymbol{\Phi}_j$ is 5×5 matrix and $\boldsymbol{\Psi}_k$ is 5×1 vector of coefficients for autoregressive terms and lagged values of S&P500 returns respectively, $\boldsymbol{\varepsilon}_t$ is a vector of error terms assumed to follow multivariate white noise process. Equation (3.10) is estimated by multivariate least square (MLS) estimator, which is equivalent to applying OLS estimator equation by equation. I select the lags order p and q by minimizing Baysian Information Criterion (BIC), setting arbitrary maximum values of five lags for p and five lags for q. The restricted version of equation (3.10) without inclusion of exogenous regressors is also allowed in the selection process. The choice of the number of lags is motivated by the analysis of Goncalves and Guidolin (2006) and Bernales and Guidolin (2014) who show that parsimonious models in terms of the number of parameters to be estimated work better for the purpose of the out-of-sample forecasting, when the parametric form of IVS is assumed.

I set up a forecasting exercise for the parametric VARX model as follows. First, I use VARX(p, q) model given in equation (3.10) to produce one-step ahead, out-of-sample forecasts $\hat{\boldsymbol{\beta}}_{t+1|t}$ of $\hat{\boldsymbol{\beta}}_{t+1}$, in the framework of the moving estimation window with the window length of 1000 days. The procedure for selecting the lag orders p, q is repeated for every estimation window. Second, I obtain forecasts of $\log \sigma_{i,t+1|t}$ using again equation (3.9), by plugging in the forecasted values of $\hat{\boldsymbol{\beta}}$. However, parametric VARX(p, q) model does not deliver prediction of the spot price that is necessary to calculate the predicted value of moneyness. Following Goncalves and Guidolin (2006) and Bernales and Guidolin (2014), I assume that the best one-step ahead forecasts of the spot price is its current value. Finally, predictions of $\log \sigma_{i,t+1|t}$ are converted to predictions of $\sigma_{i,t+1|t}$, as the evaluation of the forecasts is done for σ itself. Because conditional distribution of $\hat{\boldsymbol{\beta}}_{t+1|t}$ and distribution of error terms ε are normal, it can be done by calculating expectation of σ_i as $\sigma_{i,t+1|t} = \exp\left(\log \sigma_{i,t+1|t} + \frac{1}{2}\mathrm{Var}_t(\log \sigma_{i,t+1})\right)$, where $\mathrm{Var}_t(\log \sigma_{i,t+1}) = \delta + \boldsymbol{X}_i' \boldsymbol{\Omega}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}} \boldsymbol{X}_i$ and $\boldsymbol{X}_i$ vector of characteristics of $\log \sigma_i$ defined in equation (3.9).

Figure 3.2: Time-variation in $\hat{\boldsymbol{\beta}}$ OLS estimates



Time series of the daily OLS estimates obtained from the cross-sectional fitting of IVS with equation (3.9) over the period 04/Jan/1999-30/Aug/2013.

3.3 Practitioner Black-Scholes

Model of Dumas et al. (1998) known as *ad hoc* Black-Scholes, *ad hoc* Strawman or Practitioner Black-Scholes, henceforth referred to as Practitioner Black-Scholes (PBS), is employed by numerous authors in different applications and proved to be hard to beat benchmark in out-of-sample horse races. It is used by Heston and Nandi (2000) as a benchmark to evaluate the performance of their GARCH-type option valuation model, which usefulness for forecasting movements in IVS is assessed in this paper. It is also used by Goncalves and Guidolin (2006) and Bernales and Guidolin (2014) in the context of forecasting IVS, and Christoffersen and Jacobs (2004) who underlines the importance of the loss function in option pricing applications. I treat this model as a benchmarking approach to model IVS. Moreover, I use random walk forecasts (RW) for IVs as an additional benchmark. Random walk forecasts are obtained by assuming that tomorrow's value of IV equals its today value.

The Practitioner Black-Scholes model can be seen as a restricted case of the parametric VARX model described in section 3.2. The parametric form of IVS is fitted in the same way as in equation (3.9). However, dynamics of $\hat{\boldsymbol{\beta}}_t$ coefficients estimated in the cross-section of the implied volatilities follow a random walk process:

$$\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_{t-1} + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Upsilon}) \tag{3.11}$$

where Υ is a diagonal matrix. Because the error term is assumed to have mean zero, the best n-step ahead forecasts of the set of $\hat{\boldsymbol{\beta}}$ parameters, needed to fit IVS of tomorrow, is today's set of parameters, i.e. $\mathrm{E}_{t-1}(\hat{\boldsymbol{\beta}}_t) = \hat{\boldsymbol{\beta}}_{t-1}$. As in the parametric VARX model, I assume that the best one-step ahead forecast of the spot price is its current value.

3.4 Heston and Nandi GARCH type option valuation model

This section presents the third approach used in this thesis to forecast IVS. GARCH type option valuation model of Heston and Nandi (2000) (HN) is used to forecast option prices. Next, (forecasted) implied volatilities can be backed out from these prices. Heston and Nandi (2000) introduce a model within a class of affine GARCH models for the purpose option pricing. They assume that log-returns follow a particular GARCH(1,1) process of the form:

$$r_t = r_{f,t} + \lambda h_t + \sqrt{h_t} z_t \tag{3.12a}$$

$$h_t = \omega + \beta h_{t-1} + \alpha (z_{t-1} - \gamma \sqrt{h_{t-1}})^2, \tag{3.12b}$$

where $r_{f,t}$ is the daily risk free rate, z_t is a standard normal disturbance, and h_t is the conditional variance of the return which is known at time $t-1$. Such specification of the GARCH(1,1) process, however quite different from the standard GARCH models of Bollerslev (1986) and Duan et al. (1995), resembles models of Engle and Ng (1993) known as NGARCH and VGARCH. The unconditional variance implied by the model is given by

$$\mathrm{E}(h_t) = \frac{\omega + \alpha}{1 - \alpha\gamma^2 - \beta}, \tag{3.13}$$

while the first-order process remains stationary if $\alpha\gamma^2 + \beta < 1$. The α parameter determines the kurtosis of the return distribution. The sign of γ is expected to be positive, as it enables the model to capture a well-known stylized fact of the financial markets that a large negative shock inflates the variance more than a positive shock of the same magnitude. In general, the relationship of the following form can be proved:

$$\mathrm{Cov}_{t-1}\big(h_{t+1}, \log(S_t)\big) = -2\alpha\gamma h_t \tag{3.14}$$

Given positive values of α and γ, equation (3.14) implies the leverage effect described by Black (1976) and documented by Christie (1982). The leverage effect requires negative correlation between the spot price and the variance of the return process. Thus, model of Heston and Nandi (2000) is able to produce the same type of volatility behaviour as the continuous time model of Heston (1993). In fact, Heston and Nandi (2000) show that the model (3.12) contains continuous time stochastic volatility model of Heston (1993) as a special case (that is when the time interval between $t-1$ and t approaches to zero). The parameters of the

model (3.12) can be estimated directly by maximum likelihood estimation (MLE), as there is only a single source of randomness z_t, assumed to follow standard normal distribution. Let $\theta = (\omega, \alpha, \beta, \gamma, \lambda)$ denotes a vector of the model's parameters. The log-likelihood function can be written as

$$l_T(r_1, \ldots, r_T \mid \theta) = \sum_{t=1}^{T} -\frac{1}{2}\left(\log(2\pi) + \log(h_t) + \frac{(r_t - r_{f,t} - \lambda h_t)^2}{h_t} \right), \tag{3.15}$$

where T is the sample size. The log-likelihood function has to be maximized with respect to θ using numerical techniques, as the closed-form solution is not available due to recursive form of the function. However, the log-likelihood function given in (3.15) is of a complex shape with multiple local maxima. Therefore, a careful choice of the starting values for the optimization is of a great importance. The standard practice is to consider various starting values. I choose 4 different starting values for the parameters based on the results obtained by other authors. I initialize h_0 by setting its value equal to the sample variance of the returns.

At this point valuing options is impossible because the risk-neutral distribution of the spot price is still not known. Heston and Nandi (2000) prove that the risk neutral version of the GARCH process given in equations (3.12a) and (3.12b) is obtained by replacing λ by $-1/2$ and γ by $\gamma^* = \gamma + \lambda + 1/2$. Such a replacement has two main implications. First, given that the risk premium parameter $\lambda > -1/2$, we have $\gamma^* > \gamma$. Because of the equation (3.13) the unconditional variance under the risk neutral distribution is higher than in the historical case. Moreover, it can be seen from the equation (3.14) that γ drives the leverage effect. Under the risk-neutral distribution this effect is higher. As noted by Chorro et al. (2014), it is consistent with the skewness premium found in option prices, that the skewness implied in options is stronger than in equity returns (for more detailed discussion see Bates (1997)). Second, in case the risk premium is negative, that is $\lambda < -1/2$, the skewness premium is positive and the unconditional variance is lower under the risk-neutral distribution. This is again consistent with empirical observations: in the bear markets risk premia are negative and the historical volatility is higher than the volatility implied by the option prices.

A European call in the framework of Heston and Nandi (2000) is worth

$$C_t = \frac{1}{2}S_t + \frac{e^{-r(T-t)}}{\pi}\int_0^\infty \Re\left[\frac{K^{-i\phi} f^*(i\phi + 1)}{i\phi}\right] d\phi - Ke^{-r(T-t)}\left(\frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{K^{-i\phi} f^*(i\phi)}{i\phi}\right] d\phi\right), \tag{3.16}$$

where $\Re[\cdot]$ denotes the real part of a complex number, $f^*(i\phi)$ is a risk-neutral version of the

characteristic function $f(i\phi)$ of the logarithm of the spot price. The characteristic function can be obtained by replacing ϕ by $i\phi$ in the conditional generating function of the asset price (which in turn is the moment generating function of the $\log(S_T)$) given by[2]:

$$f(\phi) = \mathrm{E}_t[S_T^{\phi}] = S_t^{\phi} e^{A_t(T,\phi)+B_t(T,\phi)h_{t+1}}, \tag{3.17}$$

where

$$A_t(T,\phi) = A_{t+1}(T,\phi) + \phi r_{f,t} + B_{t+1}(T,\phi)\omega - \frac{1}{2}\log\left(1 - 2\alpha B_{t+1}(T,\phi)\right) \tag{3.18a}$$

$$B_t(T,\phi) = \phi(\lambda+\gamma) - \frac{1}{2}\gamma^2 + \beta B_{t+1}(T,\phi) + \frac{1/2(\phi-\gamma)^2}{1 - 2\alpha B_{t+1}(T,\phi)}. \tag{3.18b}$$

These coefficients can be calculated recursively (but backward) from the terminal conditions: $A_T(T,\phi) = 0, B_T(T,\phi) = 0$. Put values can be obtained from the put-call parity. Because the characteristic function is known analytically, it is possible to apply approach based on the Fast Fourier Transform (FFT) developed by Carr and Madan (1999) to price option. This method allows to at once calculate prices of options for variety of strikes and a given maturity. Thus, numerical computation of the integrals in equation (3.16) is not needed. Even though computational cost of such integration is not heavy when only few prices are to be calculated, it makes calibration of the model infeasible when hundreds of options are quoted every day and the calibration process is to be repeated for over 2600 days. The FFT method is described in greater detail in Appendix A.

Forecasting IVS

Heston and Nandi (2000) finds that their model provides better fit to option prices data than the Practitioner Black-Scholes model of section 3.3, both in-sample and out-of-sample. This subsection explains the procedure how to use the Heston and Nandi model for the purpose of IVS forecasting. In general, the model produce a collection of forecasted option prices from which (forecasted) implied volatility values are backed out. First, I describe how the in-sample fit of IVS is achieved, then I turn to the out-of-sample case. Option valuation formula (3.16) and the moment generating function of the logarithm of the asset price (3.18) depend on the parameters of the GARCH process (3.12). Thus, one could simply plug the MLE parameter estimates of the GARCH process into the equation (3.16) to price the options. However, information used in such approach would be only historical and far from what market practitioners actually do. If one aims to take into account information

[2]for the explicit formula of the characteristic function under the risk-neutral distribution see equations (A.4) and (A.5) in Appendix A

embedded in the cross-section of option prices, which reflects the market expectations on the future evolution of the underlying asset price, then it is more sensible to calibrate the model, that is to look for a set of the parameter estimates that ensure that the difference between theoretical prices implied by the model and market prices is as small as possible, with respect to some penalizing function. For the calibration method I choose non-linear least square (NLS) technique which minimizes the sum of squared pricing errors (SSE). For this purpose I use well-known Matlab optimizer *lsqnonlin* based on the trust-region-reflective algorithm. However, the optimizer easily get stuck in a local minimum. To obtain 'good' (or rather acceptable) minima, I initialize calibration process with 4 different sets of starting points. The objective function at time t is:

$$SSE(t,\theta) = \min_{\theta} \quad \sum_{i=1}^{N} \left(p_{i,t} - \hat{p}_{i,t}(\theta, h_{t+1})\right)^2, \tag{3.19}$$

where p denotes actual option price assumed to be the midpoint between the best closing bid and the best closing offer price of the option, and $\hat{p}$ denotes the price implied by the model. θ is, as before, the vector of model parameters. Even though various starting points are used, the calibrated parameters are highly unstable over time. The mean results of the calibration process are shown in table 3.3. High instability in a short horizon may suggest that better minima are possible. To find those one would have to consider bigger number of starting values for the optimization algorithm. However, it is not really feasible in a research such as this without for example the use of computation power cloud-computing could offer. The option value at time t is not only a function of the GARCH process

Table 3.3: Mean estimates of the calibrated parameters

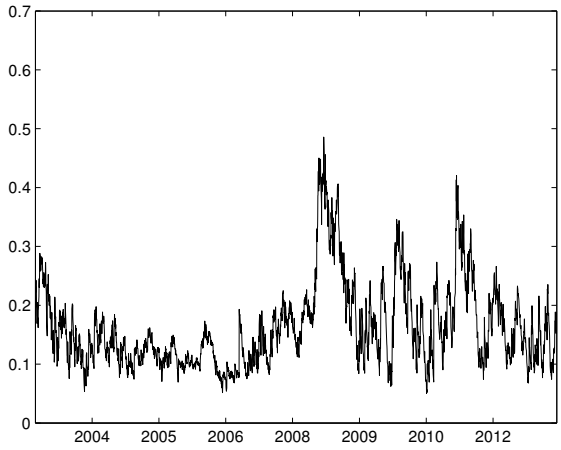
Parameter	Mean	Standard Deviation
ω	-6.08E-06	6.23E-06
α	0.57	0.44
β	8.83E-06	8.44E-06
γ	247.36	152.55

The table reports the mean and standard deviation of the daily parameter estimates from the non-linear least square estimation over the period of 2688 days, from from 24/Dec/2002 to 30/Aug/2013. Carr and Madan (1999) method based on the Fast Fourier Transform is used to calibrate the model.

parameters (the risk neutral versions of $\omega, \alpha, \beta, \gamma$, and λ), but also depends on the conditional variance h_{t+1}. As free parameters in the calibration process I treat only GARCH parameters, whereas h_{t+1} is estimated from the history of the S&P500 returns. This follows Heston and Nandi (2000) who argue that information included in historical prices of the underlying

asset provide information about the future that nature is different from the information embedded in the cross-section of option prices. I estimate GARCH model (3.12) in the rolling-window framework with the window length of 1000 days, that is 4 years of daily data. When estimating the GARCH model and pricing the options, I do not match an option expiration with the corresponding zero-coupon rate of the same maturity. Instead, I use 30 days risk-free rate linearly interpolated from the data on the zero curve. This is motivated by the significant computational savings. Above all, when using the single zero-coupon rate it is enough to calculate the recursion in (3.18) once for all traded maturities on a given day. This saving is especially important in the calibration process. Figure 3.3 illustrates the annualized daily level of the volatility estimated with the GARCH model, which I calculate as $\sqrt{252h_{t+1}}$.

Figure 3.3: Annualized daily level of volatility



The figure shows the annualized daily level of S&P 500 spot volatility from 24/Dec/2002 to 30/Aug/2013, as implied by the asymmetric GARCH model given by equation (3.12). It is estimated by means of moving window with length of 1000 days. Each value of h_{t+1} presented in the figure is obtained using the last 1000 days up to day t.

Once the model is calibrated and in-sample prices are obtained, I numerically 'invert' the prices to obtain implied volatilities in the line with the estimation procedure of OptionMetrics (2011). I back out the Black-Scholes-Merton implied volatilities by inverting equation (1.1). The appropriate interest rate is linearly interpolated to match the maturity of each option, poxing for the risk-free rate. Dividends are assumed to be constant over the remaining period of the option life and paid continuously. This step ends in-sample fitting of IVS.

This framework is easily extendable for the purpose of the out-of-sample forecasting. I use today's calibrated parameters estimates to produce one-day-ahead forecasts of option

prices. Nevertheless, to price options out-of-sample or to back out thier implied volatilities, one-day ahead predictions of the index level, interest rate and dividend yield are necessary. I follow Goncalves and Guidolin (2006) and Bernales and Guidolin (2014) and assume that the best predictions for the index level, interest rates and dividend yield are their today's values, which is in the line with the efficient market hypothesis. Finally, prediction of tomorrow's value of the conditional variance h_{t+1} is also needed. Tomorrow's value of h_{t+1} is simply h_{t+2} conditional on information available today. Having forecasted the option prices, I numerically invert the Black-Scholes-Merton formula and produce one-day ahead predictions of the implied volatilities.

Chapter 4

Combination forecasts

In order to investigate possible benefits that can arise from combining different models in the context of IVS forecasting, I introduce and test some popular and easy to implement combining schemes. As noted by Timmermann (2006) simple combining techniques typically outperform more sophisticated ones. Thus they may serve as a natural starting point for such analysis. I employ three combining techniques, wich are: (i) equal-weights which may serve as a naive (but proved to be hard to beat) benchmark, (ii) Discounted Mean Square Prediction Error (DMSPE) based weights, (iii) estimated optimal weights. DMSPE technique was originally developed in the work of Diebold and Pauly (1987). More recently, Stock and Watson (2004) examine the usefulness of this method in application to forecasting time-series of macroeconomic variables. As IV data is not in the form of time-series but forms a highly unbalanced panel, slight adjustment of this method is needed. I obtain DMSPE-based weights at time t which are used to combine one-step ahead forecasts, by solving:

$$w_{k,t+1|t} = \theta_{k,t}^{-1} / \sum_{k=1}^{K} \theta_{k,t}^{-1} \tag{4.1a}$$

$$\theta_{k,t} = \sum_{s=l}^{t} \varphi^{t-s} \frac{1}{N} \sum_{i=1}^{N} (\sigma_{i,s} - \hat{\sigma}_{i,s|s-1}^{k})^2, \tag{4.1b}$$

where φ is a discount factor, K denotes the number of the models to be combined, $t-l$ is a length of a holdout period, N is the number of IVs in a given moneyness-maturity category on a day s, $\sigma_{i,s}$ is the observed IV value and $\sigma_{i,s|s-1}^{k}$ is the forecasted IV by model k. The average of square errors in a given segment of the surface (the sum over i normalized by the number of N elements) in equation (4.1) is the abovementioned adjustment and allows to preserve the

essential idea of DMSPE method, that the discount factor weights the daily square errors[1]. It allows to preserve consistency with the way I asses the forecasting performance (see how the performance measures are defined in chapter 5). I consider the discount factor φ value of 0.95. In addition, I examine the discount factor values of 0.9 and 1 as a robustness check. The choice of these values follows Stock and Watson (2004).

Finally, I aim to estimate the 'optimal' weights by means of (restricted) regression model. Granger and Ramanathan (1984) propose three kinds of regressions to estimate the optimal weights for the variable of interest y_{t+h}:

$$y_{t+h} = w_{0h} + \boldsymbol{w}_h^{'}\hat{\boldsymbol{y}}_{t+h|t} + \varepsilon_{t+h}, \tag{4.2a}$$

$$y_{t+h} = \boldsymbol{w}_h^{'}\hat{\boldsymbol{y}}_{t+h|t} + \varepsilon_{t+h}, \tag{4.2b}$$

$$y_{t+h} = \boldsymbol{w}_h^{'}\hat{\boldsymbol{y}}_{t+h|t} + \varepsilon_{t+h} \quad s.t. \quad \boldsymbol{w}_h^{'}\iota = 1 \tag{4.2c}$$

The weights' estimates in regressions (4.2a) and (4.2b) can be obtained by means of OLS method. The difference between the two regressions is that an intercept in the former can adjust for the bias if present in the individual forecasts. The third regression can be estimated by constrained least squares. Imposing the constraint that the weights should sum up to one guarantees that the combined forecast is unbiased, in case all of the individual forecasts are also unbiased. The other justification for weights summing up to one is provided by Diebold (1988) who argues that without such restriction the error term in the combined forecast regression will be serially correlated. Furthermore, if the convexity constraint $0 \leq w_{i,h} \leq 1$ is imposed and intercept is omitted, the combined forecast lies in the range of the individual forecasts. There are several reasons why estimated optimal weights perform oftentimes worse than other combining schemes. The one worth to be mentioned, as it can be potentially influential in this study, is multicollinearty of the individual forecasts. Chapter 5 shows that bias is an important contributor to MSPE for all the individual models. In order to adjust the combined forecast for the bias, I estimate optimal weights with regression (4.2a). Thus, the results reported for the optimal estimated weights in chapter 5 are obtained with unconstrained regression (4.2a).

An important issue when implementing DMSPE and regression based weighting is the choice of a track-record period, also known as a holdout period. It determines how many observations are used to estimate the weights by methods given in equations (4.1) and (4.2). So far, no statistical procedure has been developed to select optimal length of the period. Nevertheless, it should balance between two conflicting ideas: on the one hand, a longer

[1] As Stock and Watson (2004) focus on forecasting time-series, one forecasted observation corresponds to a single day and a single square prediction error. Their formula for DMSPE is $\theta_{k,t} = \sum_{s=l}^{t} \varphi^{t-s}(r_s - \hat{r}^k_{s|s-1})^2$.

track record gives more observations in the regression, and thus more accurate estimates of the combination weights. On the other hand, it might be that the optimal weights vary over time, as some models may be superior to others in some periods but not at other times. I set the length of the track-record period equal to 44 days (approximately 2 months of trading days) but consider also 22 and 66 days as a robustness check. Given the fact that the relative performance of the individual models varies in the different regions of the surface (as shown in tables 5.2 and 5.4 in chapter 5), I estimate the optimal weights separately for each of the moneyness-maturity category. I compare this result to an alternative when the weights depend only on the maturity category and are constant across the moneyness dimension. When implementing the combination forecasts, I do not assume that tomorrow's moneyness category for every IV is known, but instead I assign the proper weight based on the best guess for its moneyness, which in turn is based on the today's value of the spot price (recall that $m = K/S$). This manner is the same as across all the individual models and it allows to asses potential benefits of forecasts combination in a fair manner.

Chapter 5

Empirical results

This chapter is arranged in 3 parts that cover number of topics. First, I define performance measures. Second, I provide results regarding the in-sample fit. Third, I present out-of-sample results of the forecasting performance for the individual models. Fourth, I examine bias in individual forecasts. Fifth, I describe the out-of-sample results for the combination forecasts. Sixth, the forecasts are formally compared to the benchmarking Practitioner Black-Scholes model by means of Diebold and Mariano (1995) test.

The out-of-sample forecasts evaluation period covers the time frame from 03/Mar/2002 until 30/Aug/2013 and includes 2644 days[1].

5.1 Performance measures

This section presents statistical evaluation of IVs forecasts. Predictability is assessed in terms of one-step ahead, daily implied volatility forecasts generated by the models presented in chapters 3 and 4. To asses the out-of-sample forecasting performance of the investigated approaches and to enable comparison between the models, I calculate the following three measures for each of them:

1. Root mean square prediction error in implied volatilities (RMSPEV)

$$RMSPEV_{k,t} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_{i,t+1} - \hat{\sigma}^{k}_{i,t+1|t}\right)^2} \tag{5.1}$$

[1]The data set contains 3688 days in total. The first 1000 days serve as an initial estimation period for the individual models and the first out-of-sample forecasts are produced on 26/Dec/2002, the 1001st day. Forecasts generated between 1001st and 1044th serve as a track-record period for the forecasts generated by combination methods. The first out-of-sample combination forecast day is 03/Mar/2003, the 1045th day in the sample. For the comparison to be fair, I begin to evaluate all the models (both individual and combination methods) starting on that day.

i.e. the square root of the average squared deviation between actual implied volatilities provided by OptionMetrics and the one-step ahead IV forecasts conditional on information available at time t. Subscript k denotes one of the four individual or the three combination models used to forecast IVS, N is the number of contracts used to calculate the measure on day t for model k.

2. Mean absolute prediction error in implied volatilities (MAPEV)

$$MAPEV_{k,t} = \frac{1}{N} \sum_{i=1}^{N} \left| \sigma_{i,t+1} - \hat{\sigma}^{k}_{i,t+1|t} \right| \tag{5.2}$$

i.e. the average absolute deviation between the observed implied volatilities and model's forecasted implied volatilities.

3. Mean correct prediction of the direction of change in implied volatilities (MCPV)

$$MCPV_{k,t} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1} \left[\text{sgn} \left(\sigma_{i,t+1} - \sigma_{i,t} \right) = \text{sgn} \left(\hat{\sigma}^{k}_{i,t+1|t} - \sigma_{i,t} \right) \right] \tag{5.3}$$

i.e. the percentage of IV for which the model correctly forecasted the direction of change for the next day. $\mathbb{1}[x]$ is an indicator function that equals 1 when x is true and 0 otherwise, $\text{sgn}(y)$ indicates if y has a positive or a negative sign.

Because the number of quoted contracts per day gradually increases over time as illustrated in figure B.3 in Appendix B, the three abovementioned measures are calculated for each of 2644 days in the out-of-sample period, what corresponds to the subscript t. If the performance measures were not calculated daily and next averaged over the number of quoted contracts per day, much of the weight would be assigned to the more recent observations. Calculating the performance measures daily allows to analyse the time variation in the predictability of IVS and is a frequently used approach in the literature on the out-of-sample forecasting of IVS (see e.g. Goncalves and Guidolin (2006) and Chalamandaris and Tsekrekos (2010)).

To ensure comparability between the models, I restrict the evaluation for all the models to the moneyness interval of $m \in [0.85, 1.15]$, what reduces the number of the options in the out-of-sample period from 1,011,545 to 634,451. The reason for this is that in PCA model the whole surface is recovered from the 64 points which correspond the fixed grid of moneyness $m_i \in \{0.85, 0.90, 0.95, 0.99, 1.01, 1.05, 1.10, 1.12\}$ (see section 2.2). In order to evaluate IV forecasts of the options with moneyness outside the mentioned interval, I would have to use a very large bandwidth to recover the full surface from the forecasted points. This in turn would lead to an extreme oversmoothing and inaccurate forecasts in the moneyness range of

interest $m \in [0.85, 1.15]$ for the PCA model.
I report the performance statistics with respect to different segments of IVS because of two related reasons. First, it is likely that a predictive power of the individual models varies across different regions of the surface, as distinguished by the moneyness-maturity groups. Second, table 2.2 in section 2.1 shows that the number of quoted contracts is irregularly distributed over the surface, with more put than call options. Neglecting this fact, the aggregated result assigns the largest weight to the more numerous put options. The aggregated result is reported under a category *All*.

5.2 In-sample fit

Before analysing the forecasting power of the models under consideration, I evaluate how well they fit to IV data in-sample. In the comparison, I put the factor dynamics of the Parametric VARX(p,q) and PCA models aside, which means I make use of equations (3.9) and (3.1) only when fitting IVS in-sample with these two models respectively (in PCA model the same number of factors is used as in the out-of-sample application which is three). Moreover, Parametric VARX(p,q) model reduces to Practitioner Black-Scholes model in the in-sample case. Hence, when evaluating the in-sample fit of the three approaches studied in this thesis to forecast IVS, I present only three models instead of the four described in chapter 3. I calculate the in-sample version of the measures given in equations 5.1 and 5.2, Root Mean Square Error (RMSEV) and Mean Absolute Error (MAEV). The in-sample evaluation is purely based on the cross-sectional fitting of the models (as explained above), while the idea behind measure 5.3 is to capture and asses their dynamic aspect. Thus, MCPV measure is not considered in the in-sample evaluation.

Table 5.1 compares the in-sample fit of the models in terms of RMSEV metric. PCA model provides the best fit for the medium (60 to 180 days) and the long (over 180 days) maturities across all the moneyness categories. The aggregated RMSEV across all the moneyness categories accounts for 0.0050 and 0.0028 for the medium and the long maturities respectively. The aggregated error for PBS accounts for 0.0080 and 0.0072, while for HN model 0.0173 and 0.0139. PCA also gives the most accurate fit for the short term option with moneyness greater or equal 0.95. It indicates that the nonparametric smoothing of the surface with the kernel regression gives a better fit than assumption of the parametric structure of IVS. The nonparametric regression fits the surface locally, taking into account only these observations that lie within the predefined bandwidths from the fitted point. On the other hand, the parametric approach uses all the available data to estimate the factors that span the surface. This neglects the fact that IVs of the short term options may exhibit quite different features than IVs of the longer contracts. Practitioner Black-Scholes model, which

Table 5.1: In-sample fit as measured by RMSEV

	Maturity								
	<60			60-180			>180		
Moneyness	**PBS**	**PCA**	**HN**	**PBS**	**PCA**	**HN**	**PBS**	**PCA**	**HN**
0.85-0.90	**.0099**	.0180	.0397	.0093	**.0055**	.0209	.0085	**.0019**	.0093
0.90-0.95	**.0090**	.0128	.0371	.0073	**.0046**	.0168	.0074	**.0019**	.0114
0.95-0.99	.0093	**.0072**	.0307	.0054	**.0043**	.0170	.0063	**.0026**	.0159
0.99-1.01	.0110	**.0085**	.0257	.0054	**.0046**	.0167	.0053	**.0030**	.0198
1.01-1.05	.0105	**.0104**	.0216	.0060	**.0041**	.0112	.0052	**.0026**	.0177
1.05-1.10	.0125	**.0086**	.0192	.0075	**.0041**	.0139	.0062	**.0023**	.0112
1.10-1.15	.0211	**.0130**	.0202	.0099	**.0052**	.0169	.0072	**.0037**	.0094
All	**.0113**	.0128	.0314	.0080	**.0050**	.0173	.0072	**.0028**	.0139

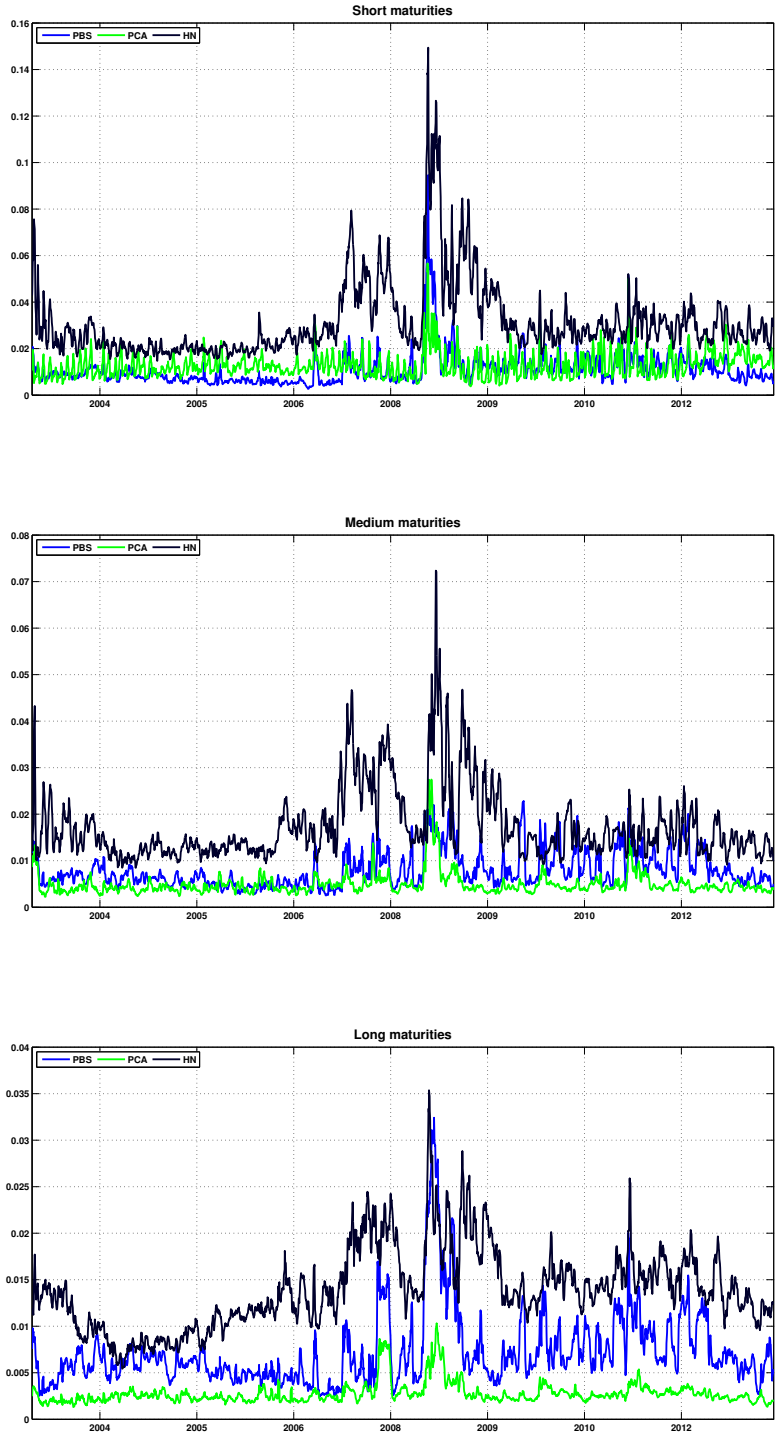
The table contains average daily root mean square error in implied volatilities (RMSEV) for the individual models over the period of 03/Mar/2003 - 30/Aug/2013, which is 2644 days. RMSEV is calculated for different moneyness-maturity categories on the sample restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. PBS is the Practitioner Black-Scholes model, PCA is the PCA model and HN is the Heston and Nandi GARCH type option pricing model. Emboldened values indicate the best performing model within each moneyness-maturity category.

is treated as a benchmark, fits the observed implied volatilities better than other models for the short term OTM put options with $m < 0.95$. It outperforms the other models in this region sufficiently to deliver the smallest RMSEV aggregated across all the moneyness groups. All the three models exhibit better accuracy in fitting IVS as maturity increases. The same conclusions can be drawn based on the table C.6 in Appendix C which shows the in-sample mean absolute errors in the implied volatilities. This result is in the line with what could be expected because when the maturity increases, the surface becomes flatter (see Figure 2.2 in section 2.1 which shows the slope of the volatility smile for different maturities). Figure 5.1 illustrates how the daily RMSEV attributed to the individual models evolve over time in each maturity group. Most of the time, PCA model has the smallest daily RMSEV. The in-sample fit sharply deteriorates during the burst of the 2008 crisis. However, for the long term options the fit of PCA model is hardly affected by the crisis. Figure 2.3 in section 2.1 shows that the slope of the volatility term structure plummeted in the crisis of 2008, meaning that IV of the short term contracts increased relatively more than IV of the long term options. As PCA model fits the surface locally, it explains the fact why the fit of the model deteriorates less in comparison to the other models, which parameters are estimated using all the available data irrespective of the remaining time to maturity.

Model of Heston and Nandi (2000) is the only one estimated in the option prices space in

contrast to PBS, VARX and PCA models which are estimated directly in the implied volatility space. Although the model is of a complex structural type, it gives a poor fit to IVS, with RMSEV of 0.0314, 0.0173 and 0.0139 for the short, medium and the long maturities respectively. Table C.7 in Appendix C contains a size of pricing errors that correspond with the model's poor performance in providing the in-sample fit of the implied volatility surface. The average daily RMSE in option prices, as a percentage of the average daily price, accounts for 29% for the short, 14% for the medium and 12% for the long term contracts. The poor performance in the implied volatility space in not surprising because even small errors in option prices can produce large errors in implied volatilities as noted by Hentschel (2003).

Figure 5.1: In-sample daily RMSEV over time



The figure illustrates the evolution of daily RMSEV for the three models evaluated in terms of the in-sample fit over the period 03/Mar/2003-30/Aug/2013. The plot is based on the 5-day moving averages. RMSEV is calculated for different maturity categories that aggregate all the moneyness groups on the sample restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. PBS- Practitioner Black-Scholes, PCA- Principal Component Analysis model, and HN- Heston and Nandi GARCH model.

5.3 Out-of-sample forecasting performance

Individual models

Table 5.2 reports the average daily RMSPEV of the four individual models. Again, all the models provide better accuracy in forecasting IVS for the medium and the long term options than for the short term contracts. In addition to the reason mentioned in the previous section, this can be explained by the fact that the variability from day to day of IVs is much smaller for the long maturities, as shown in Table C.1 in Appendix C. IV forecasts are less accurate for the put than the call options in the short term category. It can also be explained by the increasing variance along the moneyness dimension. Similarly to the in-sample case, PCA model is the best performer for the medium and the long maturities for all the moneyness-maturity combinations. Its average daily RMSPEV aggregated across all the moneyness groups accounts for 0.0066 for the medium and 0.0040 for the long maturities. PCA model delivers the most accurate forecasts for the short term contracts in 4 out of 7 moneyness categories which are 0.95-0.99, 0.99-1.01, 1.05-1.10 and 1.10-1.15. For the deepest OTM puts considered in the evaluation, i.e. when $m \in [0.85, 0.90)$, Parametric VARX(p,q) model gives the best result with the average daily RMSPEV of 0.0109. Benchmarking Practitioners Black-Scholes yields the smallest forecasting error in two remaining moneyness groups, when $m \in [0.90, 0.95)$ and $m \in [1.01, 1.05)$. However, even the best performing individual model in a given region of the surface is not able to beat simple random walk forecasts for the implied volatilities.

The relative out-of-sample performance of PCA model as compared to the in-sample fit deteriorates the most among all the models. Still, it is capable of producing the most accurate forecasts of IVS in the most segments of the surface. Interestingly, the worst performer which is Heston and Nandi model deteriorates the least of-of-sample what makes it the most stable model. It remains an open question whether with a different (possibly better) calibration approach it could beat other models in the out-of-sample horse race. In general, introducing the factors dynamics to Practitioner Black-Scholes model, that assumes the parametric structure of IVS, improves the forecasting power of the approach only a little as the Parametric VARX model has its average daily RMSPEV very close to PBS. Parametric VARX beats PBS in 12 out of 21 segments of the surface.

Figure 5.2 illustrates how daily RMSPEV of the models evolve over time in each of the maturity groups. Most of the time the blue line of PBS and the red one of the Parametric VARX model lie very close to each other and therefore are hardly distinguishable. For the short maturities, parametric models perform systematically better than PCA model during years 2005-2007 and at the end of the evaluation period. Interestingly, Heston and Nandi

model yields lower RMSPEV at the beginning of the 2008 crisis in the medium and the long maturity categories as compared to parametric models. It can be concluded that the relative models' performance varies over time. I investigate this issue in greater detail in the next subsection. Figure B.4 and table C.9 in Appendices B and C evaluate the models with respect to the mean absolute prediction error in implied volatilities. Even though same conclusion can be drawn based on RMSPEV and MAPEV, I provide the statistics regarding the latter measure to illustrate the average size of the absolute prediction error. Finally, table C.8 in Appendix C provides statistics regarding the size of the pricing errors resulted from the NLS calibration of Heston and Nandi model.

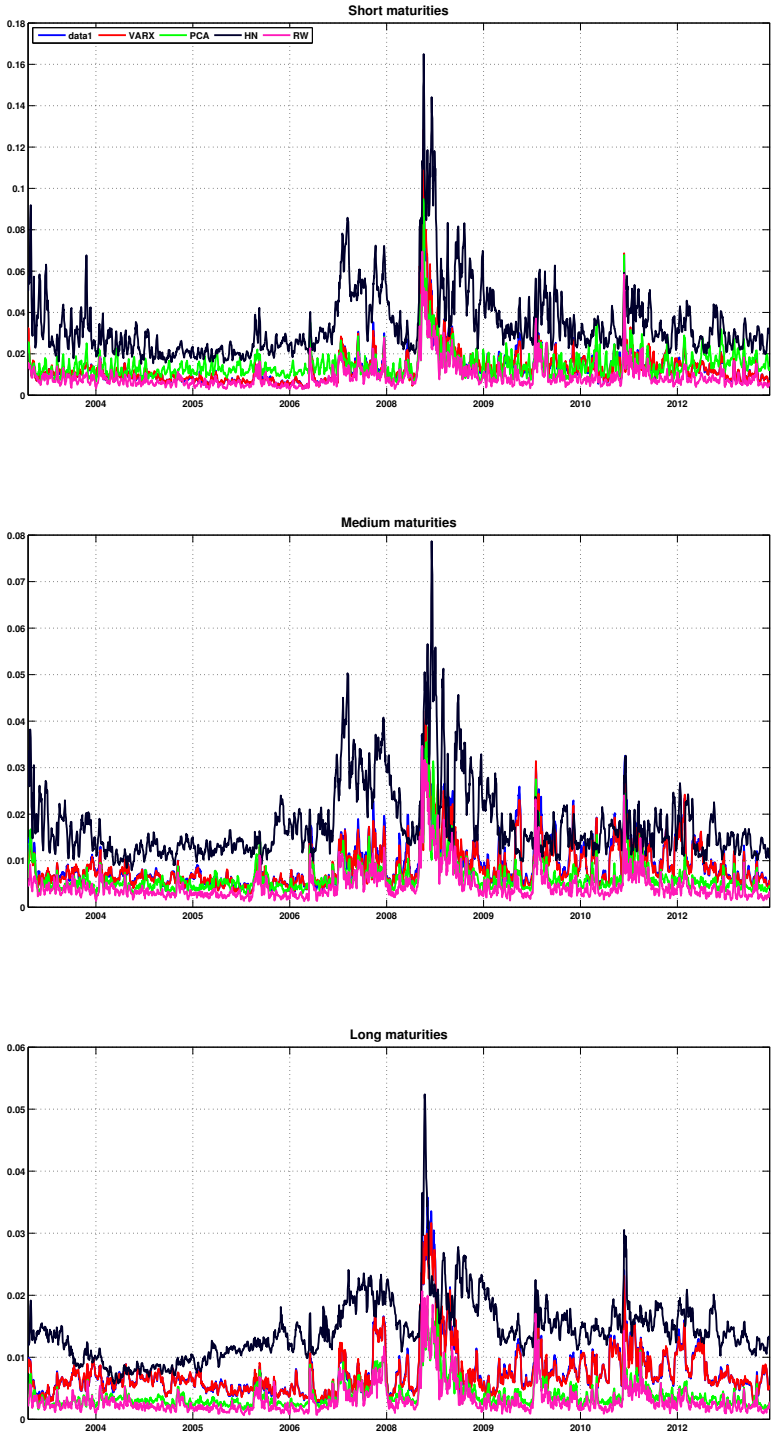
Even though the primary focus of this study is on the forecasting accuracy as the combining schemes aim to reduce MSPEV, it is interesting to compare the forecasting performance from a different perspective. Table 5.3 shows the next performance measure which is the average daily mean correct prediction of the direction of change in implied volatilities, MCPV. When the models are evaluated with respect to this measure, similar conclusion can be drawn when it comes to gains of incorporating the factors' dynamics to the parametric structure of IVS represented by PBS model. Parametric VARX model is superior in predicting the sign change in 19 out of 21 IVS regions, as compared to PBS. Is is also the best performer for the short maturities. PCA model, which was the best performer in terms of RMSPEV, predicts the direction of change most accurate for the medium and the long maturities. When PCA model is investigated in a greater detail, it turns out that when VARX model is employed to capture factor dynamics, it correctly predicts the direction of change in log smoothed implied volatilities $\log \hat{\sigma}_t(m_i, \kappa_j)$ 54% of times on average (average value across 64 smoothed times-series of $\log \hat{\sigma}_t(m_i, \kappa_j)$). This time Heston and Nandi model turns out not to be the worst choice for modelling IVS. It happens to predict the direction of change the most accurate for option with moneyness of 1.05-1.10 in the short and the medium maturity category, and for the deepest OTM long-term contracts. In contrast to the previous performance metric, PBS exhibits the best accuracy in none of the moneyness-maturity groups.

To give an intuition whether the average MCPV values of the models are any good, I compare them to a naive model that assumes parallel shifts of IVS segments. If on a previous day more than half IVs in a given moneyness-maturity group increase (decrease), the naive model assumes that all options in this group will increase (decrease) the next day. The naive model is simplified version of a random walk model that could assume increase (decrease) in specific IV value based on the change from the previous day. However, the naive model is easier to apply, as it does not require that an individual option is in the sample for the three consecutive days (as opposed to the hypothetical random walk model mentioned above). It is a desirable property of the naive model because many options are filtered out from the

OptionMetrics data set (see section 2.1). That is, it is likely that a specific option is present in the data set on day t, excluded from the analysis on day $t+1$, and again returns on day $t+2$. Usually, naive model has MCPV value close to other models. Emboldened figures in table 5.3, which indicate which models are superior to the naive model, show that sign predictability is expected to be higher for the medium and the long maturities that the short term option.

An useful and interesting test of MCPV statistic would be to asses the impact of the sign predictability in terms of trading strategies, as they often entirely rely on the predicted direction of change. However, economic evaluation of the models is out of the scope of this study.

Figure 5.2: Out-of-sample daily RMSPEV over time



The figure illustrates the evolution of daily RMSPEV for the four models evaluated in terms of out-of-sample fit over the period 03/Mar/2003-30/Aug/2013 and the random walk benchmark. The plot is based on the 5-day moving averages. RMSPEV is calculated for different maturity categories that aggregate all the moneyness groups on the sample restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. PBS- Practitioner Black-Scholes model, VARX- Parametric VARX(p,q), PCA-Principal Component Analysis model, HN- Heston and Nandi GARCH model, RW- Random Walk.

Table 5.2: Out-of-sample fit as measured by RMSPEV

	Maturity														
	<60					60-180					>180				
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**
0.85-0.90	.0110	**.0109**	.0186	.0401	.0089	.0094	.0090	**.0068**	.0212	.0042	.0080	.0077	**.0032**	.0096	.0027
0.90-0.95	**.0107**	.0109	.0151	.0394	.0086	.0078	.0075	**.0060**	.0175	.0042	.0072	.0070	**.0032**	.0116	.0028
0.95-0.99	.0110	.0116	**.0096**	.0345	.0079	.0066	.0066	**.0056**	.0172	.0043	.0064	.0063	**.0035**	.0161	.0030
0.99-1.01	.0130	.0135	**.0115**	.0294	.0089	.0072	.0073	**.0063**	.0165	.0051	.0059	.0058	**.0041**	.0197	.0035
1.01-1.05	**.0131**	.0134	.0131	.0252	.0080	.0081	.0081	**.0059**	.0124	.0045	.0061	.0061	**.0037**	.0178	.0030
1.05-1.10	.0144	.0141	**.0113**	.0212	.0082	.0095	.0092	**.0059**	.0144	.0045	.0071	.0072	**.0034**	.0114	.0028
1.10-1.15	.0211	.0210	**.0147**	.0236	.0108	.0108	.0100	**.0067**	.0172	.0045	.0079	.0079	**.0047**	.0094	.0027
All	**.0133**	.0135	.0149	.0344	.0092	.0094	.0091	**.0066**	.0179	.0048	.0076	.0075	**.0040**	.0143	.0032

The table contains the average daily Root Mean Square Prediction Error in implied volatilities (RMSPEV) for the individual models over the period 03/Mar/2003-30/Aug/2013. Forecasting models are estimated on the full sample of the daily S&P500 implied volatilities but RMSPEV is calculated for different moneyness-maturity categories for the forecasts restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. PBS is the Practitioner Black-Scholes model, PCA is PCA model, HN is Heston and Nandi GARCH type option pricing model, and RW is random walk benchmark. Emboldened values indicate the best performing model, excluding RW benchmark, within each moneyness-maturity category. Underlined values indicate the best overall performer.

Table 5.3: Percentage of correctly predicted direction of change in IVs

	Maturity														
	<60					60-180					>180				
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**Naive**	**PBS**	**VARX**	**PCA**	**HN**	**Naive**	**PBS**	**VARX**	**PCA**	**HN**	**Naive**
0.85-0.9	.5516	.5594	.4427	.5268	.6080	**.4834**	**.4882**	**.5295**	**.5179**	.4755	**.5210**	**.5266**	**.5250**	**.5269**	.5123
0.9-0.95	.5668	.5725	.4839	.5121	.5737	**.5061**	**.5118**	**.5372**	**.5136**	.5047	**.5290**	.5345	**.5471**	**.5227**	.5106
0.95-0.99	**.5494**	**.5526**	**.5631**	**.5083**	.4796	**.5356**	**.5305**	**.5503**	**.5235**	.5094	**.5348**	.5393	**.5502**	**.5092**	.4948
0.99-1.01	**.5273**	**.5339**	**.5285**	**.5168**	.4762	**.5328**	**.5317**	**.5418**	**.4873**	.4756	**.5368**	**.5404**	**.5501**	.4460	.4976
1.01-1.05	.5079	.5218	.4894	.5189	.5393	**.4939**	**.5052**	**.5483**	**.5415**	.4648	**.5179**	**.5242**	**.5648**	.4789	.4758
1.05-1.1	**.5033**	**.5230**	**.5291**	**.5316**	.4766	.4859	.4984	**.5455**	**.5562**	.5391	**.4916**	**.4981**	**.6057**	**.5147**	.4805
1.1-1.15	.4965	**.5160**	**.5262**	.4858	.5047	.4891	.4995	.4888	.5294	.5330	.4832	.4884	.4653	**.5242**	.5220
All	**.5399**	**.5498**	**.5110**	**.5138**	.4938	.5066	**.5131**	**.5392**	**.5228**	.5101	.5245	**.5304**	**.5494**	.5119	.5249

The table reports the average values of the daily MCPV. MCPV is a fraction of correctly predicted direction of change in implied volatilities on a given day. Forecast evaluation period for this measure covers dates from 03/Mar/2003 until 30/Aug/2013 Forecasting models are estimated on the full sample of daily S&P500 implied volatilities but MCPV is calculated for different moneyness-maturity categories for the forecasts restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. The naive model is a version of a random walk model that assumes that all IVs in a given IVS segment increase (decrease) if on a previous day in a given moneyness-maturity category more than half IVs increase (decrease). Emboldened values indicate the best performing model within each moneyness-maturity category.

Sub-periods individual models

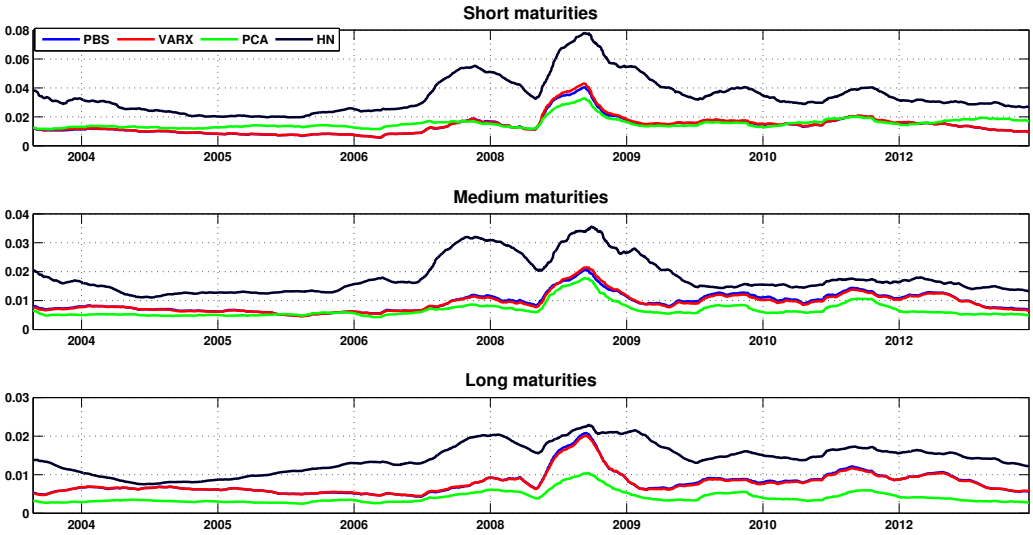
Given that the out-of-sample period covers 11 years and includes such a remarkable event as the financial crisis of 2008, I divide the out-of-sample period in three sub-periods to examine stability of the forecasting performance over time. The sample is divided into three sub-periods based on the NBER US Business Cycle Expansions and Contractions data, such that the first sub-period covers the dates before the recession (03/Mar/2003-31/Dec/2007), the second is the crisis time (01/Jan/2008-30/Jun/2009), and the third represents after the crisis data (01/Jul/2009-30/Aug/2013).

Table 5.4 shows the forecasting performance measured during the three sub-periods. For the medium and the long maturities PCA is unbeatable and yields the smallest RMSPEV at all times. When it comes to the short term options, in the first sub-period which covers years 2003-2007, PCA exhibits the highest forecasting accuracy in four moneyness categories: $m \in [0.95, 0.99)$, $m \in [0.99, 1.01)$, $m \in [1.05, 1.10)$ and $m \in [1.05, 1.15]$. Benchmarking PBS is the most accurate model for the two moneyness categories which $m \in [0.85, 0.90)$ and $m \in [0.90, 0.95)$. Parametric VARX wins the out-of-sample horse race in moneyness category of $m \in [1.01, 1.05)$. During the financial crisis, when IVS shifts to extraordinary levels, as can be observed with VIX behaviour in figure 2.2, PCA captures the out-of-sample dynamics better than the other models for all but the deep OTM put options with $m \in [0.85 - 0.90)$, when PBS and VARX models perform better. The forecasting performance of all the models deteriorates largely in the recession time, for example PCA yields the average RMSPEV for the short maturities of 0.0129 before the crisis and 0.0192 during the crisis. In the post-crisis time, the best performer at the short end of the surface is PCA model winning the out-of-sample horse race in 5 moneyness categories, when $m \geq 0.95$. The short term deep OTM puts are the most numerous option as compared to other moneyness categories of the maturity category, what helps to explain why PBS yields the smallest forecasting error aggregated across the moneyness dimension even though it is the best model only in one out of 7 groups along this dimension. Finally, it is worth to mention that Heston and Nandi model is not always the worst performer, as it beats the parametric approaches in the third sub-period in IVS segment of the deep OTM call options. In the last sub-period RMSPEV of HN model in the mentioned moneyness-maturity category is 0.0172, while for PBS and VARX it accounts for 0.0203 and 0.0201 respectively.

To further explore the time variation in the forecasting ability, I report RMSPEV over the relatively short rolling window. Figure 5.3 shows 126 days (approximately half a year) rolling RMSPEV with respect to options' maturity. For the medium and the long maturities PCA approach outperforms permanently other methods. For the short term options, it begins to

outperform other models since the outburst and for the duration of the financial crisis. In between of the end of the crisis and 2013 parametric models and PCA yield similar forecasting errors. In 2013, which is the last year of out-of-sample period, parametric methods turn out to be the best performers again.

Figure 5.3: 126 days rolling RMSPE over time



The figure illustrates the average daily RMSPEV over the moving window of 126 days. RMSPEV is calculated for the different maturity categories that aggregate all the moneyness groups on the sample restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. PBS- Practitioner Black-Scholes, PCA, and HN- Heston and Nandi GARCH model.

Table 5.4: Average daily RMSPEV in different sub-periods: individual models

Panel A: before the 2008 crisis

	Maturity														
	<60					60-180					>180				
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**
0.85-0.90	**.0081**	.0082	.0173	.0316	.0071	.0079	.0081	**.0056**	.0207	.0034	.0059	.0059	**.0026**	.0075	.0021
0.90-0.95	**.0074**	.0076	.0140	.0331	.0068	.0063	.0064	**.0052**	.0162	.0033	.0049	.0049	**.0026**	.0086	.0022
0.95-0.99	.0080	.0083	**.0076**	.0312	.0064	.0051	.0052	**.0049**	.0161	.0035	.0041	.0041	**.0030**	.0138	.0023
0.99-1.01	.0095	.0098	**.0093**	.0255	.0075	.0055	.0055	**.0052**	.0143	.0043	.0046	.0046	**.0034**	.0173	.0029
1.01-1.05	.0086	**.0086**	.0112	.0208	.0063	.0056	.0056	**.0049**	.0108	.0036	.0055	.0055	**.0030**	.0149	.0025
1.05-1.10	.0106	.0101	**.0092**	.0170	.0062	.0063	.0059	**.0050**	.0130	.0035	.0061	.0061	**.0028**	.0083	.0022
1.10-1.15	.0205	.0201	**.0131**	.0217	.0089	.0082	.0074	**.0053**	.0143	.0035	.0063	.0063	**.0038**	.0074	.0021
All	**.0095**	.0096	.0129	.0290	.0073	.0069	.0068	**.0055**	.0164	.0039	.0057	.0058	**.0032**	.0117	.0025

Panel B: during the 2008 crisis

	Maturity														
	<60					60-180					>180				
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**
0.85-0.90	**.0170**	.0172	.0181	.0661	.0127	.0111	.0111	**.0100**	.0314	.0071	.0117	.0115	**.0053**	.0128	.0046
0.90-0.95	.0196	.0199	**.0167**	.0676	.0135	.0118	.0119	**.0089**	.0317	.0075	.0136	.0134	**.0053**	.0164	.0048
0.95-0.99	.0187	.0192	**.0150**	.0558	.0137	.0115	.0115	**.0087**	.0305	.0079	.0143	.0141	**.0062**	.0198	.0056
0.99-1.01	.0202	.0210	**.0177**	.0443	.0154	.0117	.0118	**.0107**	.0248	.0092	.0114	.0112	**.0078**	.0231	.0074
1.01-1.05	.0210	.0218	**.0184**	.0382	.0149	.0120	.0119	**.0104**	.0178	.0084	.0090	.0087	**.0070**	.0253	.0056
1.05-1.10	.0230	.0238	**.0175**	.0380	.0152	.0138	.0134	**.0089**	.0201	.0082	.0087	.0084	**.0052**	.0206	.0051
1.10-1.15	.0232	.0236	**.0190**	.0381	.0155	.0142	.0134	**.0102**	.0286	.0079	.0085	.0082	**.0064**	.0152	.0046
All	.0225	.0231	**.0192**	.0553	.0159	.0138	.0136	**.0105**	.0284	.0089	.0121	.0118	**.0069**	.0203	.0062

Panel C: after the 2008 crisis

	Maturity														
	<60					60-180					>180				
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**
0.85-0.90	.0124	**.0120**	.0202	.0412	.0098	.0106	.0094	**.0072**	.0184	.0043	.0093	.0087	**.0032**	.0109	.0028
0.90-0.95	**.0112**	.0116	.0158	.0366	.0088	.0081	.0073	**.0058**	.0139	.0041	.0077	.0072	**.0032**	.0134	.0028
0.95-0.99	.0118	.0127	**.0099**	.0306	.0076	.0065	.0063	**.0052**	.0137	.0041	.0063	.0060	**.0032**	.0175	.0028
0.99-1.01	.0145	.0152	**.0117**	.0286	.0081	.0076	.0077	**.0058**	.0161	.0045	.0056	.0055	**.0037**	.0213	.0029
1.01-1.05	.0155	.0159	**.0136**	.0257	.0076	.0097	.0098	**.0055**	.0124	.0041	.0059	.0060	**.0034**	.0186	.0027
1.05-1.10	.0155	.0152	**.0114**	.0198	.0080	.0116	.0114	**.0058**	.0139	.0043	.0078	.0080	**.0034**	.0117	.0027
1.10-1.15	.0203	.0201	**.0133**	.0172	.0094	.0127	.0120	**.0072**	.0167	.0046	.0095	.0097	**.0052**	.0099	.0027
All	**.0144**	.0146	.0156	.0332	.0090	.0105	.0100	**.0065**	.0158	.0045	.0082	.0081	**.0039**	.0152	.0029

The table reports the average daily RMSPEV in the three sub-periods: 03/Mar/2003-31/Dec/2007, 01/Jan/2008-30/Jun/2009, 01/Jul/2009-30/Aug/2013, and for the different moneyness-maturity categories. Forecasting models are estimated on the full sample of the daily S&P500 implied volatilities but RMSPEV is calculated for the forecasts restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. PBS is the Practitioner Black-Scholes model, VARX is Parametric VARX model, PCA is PCA model and HN is Heston and Nandi GARCH type option pricing model. Emboldened values indicate the best performing model within each moneyness-maturity category.

Bias in individual forecasts

This subsection provides assessment of how important the bias contribution to the forecasting errors is. Following Pindyck and Rubinfeld (1998), mean squared prediction error can be decomposed as follows:

$$\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_{i,t+1}-\hat{\sigma}_{i,t+1|t}\right)^2=\left((\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{i,t+1|t})-\bar{\sigma}_{t+1}\right)^2+(s_{\hat{\sigma}_{t+1|t}}-s_{\sigma_{t+1}})^2+2(1-\rho)s_{\hat{\sigma}_{t+1|t}}s_{\sigma_{t+1}},\tag{5.4}$$

where $\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{i,t+1|t}$, $\bar{\sigma}_{t+1}$, $s_{\hat{\sigma}_{t+1|t}}$, $s_{\sigma_{t+1}}$ are the means and biased (obtained with denominator of N) standard deviations of predicted and observed values of IV, denoted by $\hat{\sigma}_{i,t+1|t}$ and σ_{t+1} respectively. This decomposition allows to asses the contribution to MSPE of (squared) bias, variance and covariance, based on the following proportions that sum up to one:

$$\begin{aligned}
&\text{bias contribution:} && \frac{\left((\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{i,t+1|t})-\bar{\sigma}_{t+1}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_{i,t+1}-\hat{\sigma}_{i,t+1|t}\right)^2},\\
&\text{variance contribution:} && \frac{(s_{\hat{\sigma}_{t+1|t}}-s_{\sigma_{t+1}})^2}{\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_{i,t+1}-\hat{\sigma}_{i,t+1|t}\right)^2},\\
&\text{covariance contribution:} && \frac{2(1-\rho)s_{\hat{\sigma}_{t+1|t}}s_{\sigma_{t+1}}}{\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_{i,t+1}-\hat{\sigma}_{i,t+1|t}\right)^2}.
\end{aligned}$$

A desirable property of the forecasts is that much weight should be concentrated on the covariance proportion, which accounts for unsystematic forecasting errors. Investigating how much the bias contributes to the forecasting errors can not only provide better understanding of the roots of the errors, but also can help to decide whether optimal weights obtained by means of regression model should be estimated with or without an intercept. Inclusion of the intercept can adjust for the bias if present in the individual forecasts. Table 5.5 reports average daily bias proportion with respect to different-moneyness maturity, as well as distinguished sub-periods. From the table it can be seen, that bias is an important contributor to the forecasting errors for each of the individual models, especially when assessed in the individual moneyness-maturity bins. On average the bias propotion in the MSPE is around 60-70%. Parametric models, which are PBS and VARX, exhibit on average the largest bias proportion for medium maturities. Contrary to that, PCA model has the largest bias proportion for the options that fall into long maturity category, while the same it true for Heston and Nandi model. Parametric models have smaller bias proportion in the short maturity category for the put options than the call options relatively to other models. Usually, all the models exhibit smaller bias proportion for the call options as compared with the put

options. It is especially the case for the short maturities, where the curvature of the smile is the most significant. In general, patterns across different moneyness-maturity categories are not the same in different sub-periods.

Table 5.5: Bias proportion in forecasting errors

Panel A: before the 2008 crisis

	Maturity											
	<60				60-180				>180			
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**PBS**	**VARX**	**PCA**	**HN**	**PBS**	**VARX**	**PCA**	**HN**
0.85-0.90	63%	64%	69%	**38%**	82%	84%	53%	**43%**	72%	75%	59%	**44%**
0.90-0.95	53%	53%	58%	**33%**	74%	74%	45%	**41%**	62%	65%	**58%**	63%
0.95-0.99	69%	70%	53%	**42%**	65%	66%	67%	**59%**	**55%**	59%	76%	88%
0.99-1.01	63%	63%	57%	**54%**	63%	63%	56%	**55%**	**59%**	62%	75%	84%
1.01-1.05	53%	**54%**	81%	59%	77%	77%	70%	**47%**	**67%**	68%	71%	78%
1.05-1.10	55%	**53%**	66%	67%	64%	**62%**	68%	76%	73%	74%	56%	**43%**
1.10-1.15	62%	**61%**	66%	82%	66%	**63%**	70%	86%	73%	75%	60%	**54%**
All	27%	27%	**16%**	26%	34%	33%	30%	**30%**	20%	21%	28%	**6%**

Panel B: during the 2008 crisis

	Maturity											
	<60				60-180				>180			
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**PBS**	**VARX**	**PCA**	**HN**	**PBS**	**VARX**	**PCA**	**HN**
0.85-0.90	63%	63%	**51%**	65%	78%	78%	**67%**	70%	80%	81%	74%	**73%**
0.90-0.95	65%	66%	**57%**	74%	81%	82%	**67%**	84%	81%	83%	**73%**	85%
0.95-0.99	68%	68%	**67%**	77%	82%	81%	**73%**	86%	82%	84%	**78%**	85%
0.99-1.01	58%	60%	**56%**	72%	64%	64%	**60%**	64%	**71%**	71%	79%	83%
1.01-1.05	73%	**72%**	78%	73%	81%	81%	76%	**57%**	66%	**65%**	82%	84%
1.05-1.10	64%	**63%**	71%	71%	84%	82%	**64%**	65%	64%	**63%**	66%	80%
1.10-1.15	63%	**63%**	70%	78%	85%	85%	**68%**	81%	66%	64%	71%	**57%**
All	41%	42%	**40%**	57%	37%	**35%**	48%	55%	39%	40%	42%	**11%**

Panel C: after the 2008 crisis

	Maturity											
	<60				60-180				>180			
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**PBS**	**VARX**	**PCA**	**HN**	**PBS**	**VARX**	**PCA**	**HN**
0.85-0.90	56%	53%	39%	**30%**	83%	80%	55%	**35%**	71%	71%	**62%**	67%
0.90-0.95	51%	52%	41%	**32%**	72%	67%	47%	**35%**	60%	60%	**59%**	85%
0.95-0.99	63%	67%	51%	**43%**	59%	57%	67%	**47%**	**56%**	57%	70%	91%
0.99-1.01	69%	70%	**58%**	58%	63%	64%	58%	**46%**	**57%**	60%	73%	85%
1.01-1.05	65%	66%	76%	**61%**	72%	74%	64%	**45%**	**60%**	65%	67%	87%
1.05-1.10	46%	**45%**	60%	59%	73%	75%	**60%**	63%	72%	76%	**50%**	63%
1.10-1.15	62%	**61%**	62%	72%	66%	64%	**59%**	76%	78%	81%	62%	**43%**
All	28%	31%	**24%**	27%	29%	28%	37%	**25%**	24%	25%	39%	**9%**

The table reports the average daily proportion of bias contribution to forecasting errors in the three sub-periods: 03/Mar/2003-31/Dec/2007, 01/Jan/2008-30/Jun/2009, 01/Jul/2009-30/Aug/2013, and for the different moneyness-maturity categories. PBS is the Practitioner Black-Scholes model, VARX is Parametric VARX model, PCA is PCA model and HN is Heston and Nandi GARCH type option pricing model. Emboldened values indicate the model with the smallest bias proportion in its forecasting errors within each moneyness-maturity category.

Forecast combination

Table 5.6 shows that using forecasts combination can improve on forecasting the implied volatility surface as compared to the individual models- although the improvement is not large enough to beat RW forecasts. However, combination forecasts based on the equal weights obtained by combining all the four models often produce poor results. It has been documented in the literature on the model combination that trimming of the worst performing models often improves performance. Winkler and Makridakis (1983) and Stock and Watson (2004) find out that combining schemes like the equal weighting perform better when the worst performing model is excluded from the combination. The same is true with respect to forecasting IVS. DMSPE model yields mix results when Heston and Nandi model is excluded from the combination. In general, trimmed version of DMSPE yields smaller errors for the put options, while it does not improve forecasting IVs of the call options. It may be explained by the fact that Heston and Nandi model is relatively more accurate in forecasting IVS for options with larger moneyness. I find that trimming is not effective when the regression method is employed. I report the evaluation results for the trimmed version of the equal weighting. DMSPE and regression based weights are calculated using all the four models.

Table 5.6: Out-of-sample performance of combining methods

	Maturity														
	<60					60-180					>180				
Moneyness	**RW**	**Ind**	**EW**	**Disc**	**Opt**	**RW**	**Ind**	**EW**	**Disc**	**Opt**	**RW**	**Ind**	**EW**	**Disc**	**Opt**
0.85-0.90	.0089	.0109	.0121	.0112	**.0104**	.0042	.0068	**.0066**	**.0066**	**.0052**	.0027	.0032	.0054	.0037	**.0031**
0.90-0.95	.0086	.0107	.0109	**.0105**	**.0100**	.0042	.0060	.0061	.0060	**.0052**	.0028	.0032	.0051	.0039	.0032
0.95-0.99	.0079	.0096	.0098	.0096	**.0091**	.0043	.0056	.0057	.0058	**.0051**	.0030	.0035	.0049	.0043	**.0033**
0.99-1.01	.0089	.0115	.0123	.0118	**.0099**	.0051	.0063	.0064	.0064	**.0057**	.0035	.0041	.0048	.0046	**.0039**
1.01-1.05	.0080	.0131	**.0126**	**.0119**	**.0095**	.0045	.0059	.0070	.0062	**.0053**	.0030	.0037	.0048	.0044	**.0034**
1.05-1.10	.0082	.0113	.0123	.0114	**.0102**	.0045	.0059	.0075	.0064	**.0054**	.0028	.0034	.0051	.0039	**.0032**
1.10-1.15	.0108	.0147	.0181	.0156	.0148	.0045	.0067	.0084	.0070	**.0058**	.0027	.0047	.0063	.0049	**.0036**
All	.0092	.0133	.0128	.0121	**.0107**	.0048	.0066	.0074	.0069	**.0058**	.0032	.0040	.0057	.0046	**.0037**

The table shows the average daily RMSPEV for the different combining schemes. EW denotes trimmed equal weights, where Heston and Nandi model is not included, Disc is Discounted Mean Square Prediction Error (DMSPE) method that follows Stock and Watson (2004) and Opt are the estimated optimal weights with unrestricted regression with inclusion of the intercept (equation (4.2a)). Indv denotes the best performing individual model in a given moneyness-maturity category which implies that different individual models are included in columns under this heading. The hold-out period for DMSPE and regression methods is set to 44 days. Therefore, the evaluation period covers 2642 days, from 03/Mar/2003 until 30/Aug/2013. The emboldened values indicate the combining schemes that improve on the best individual forecast in a given moneyness-maturity category.

Forecasting based on the estimated optimal weights (with unrestricted regression (4.2a)) yields the best results, superior not only to other combining schemes, but also to the best performing individual model in a given IVS segment. On the other hand, forecasts based on the equal weights (but trimmed) usually result in larger RMSPEV as compared to the best individual model. The equal weighting beats the best individual model only for the short term calls with $m \in [1.01 - 1.05)$, and ATM put options with moneyness in the range of 0.99-1.01 for the medium term options. DMSPE-based weighting improves on the best individual model for two moneyness categories within the short maturities: 0.90-0.95 and 1.01-1.15. For the medium maturities, it yields better results for the deep OTM puts with $m \in [0.85 - 0.90)$. For the long maturities, only the regression-based weighting results in smaller RMSPEV in all the moneyness categories but one, yielding the largest improvement for the deep OTM calls with $m \in [1.10 - 1.15]$. For the medium term options, the most significant improvement can be observed for the deep OTM puts and the deep OTM calls. In these two categories RMSPEV reduces from 0.0068 for the best individual model to 0.0052, and from 0.0067 to 0.0058 respectively when regression-based weights are employed. The most notable reduction in forecasting errors with respect to the short maturities is observed for the contracts with $m \in [1.01 - 1.05)$, when RMSPEV drops from 0.0131 to 0.0095. Table 5.2 shows that the aggregated pricing errors in the short term maturity category are the largest, which implies that the forecasting the short end of the surface is the most challenging task. All the combining schemes produce smaller average RMSPEV aggregated across the moneyness dimension than the best performing individual model. In general, implementation of the forecasts combination is partly successful, as the estimated optimal weights are able to generate more accurate forecasts for almost all the moneyness-maturity categories than the individual, yet they do not beat random walk forecasts.

Based on the evolution of RMSPEV of the individual models presented in figure 5.2, it can be presumed that the relative importance each of the individual models plays in the combination forecasts varies over time. Figures 5.4 and 5.5 illustrate what weights are assigned to the individual models by DMPSE and the regression methods respectively. They are aggregated across the moneyness categories and plotted with respect to options' maturity. At all times, the regression method assigns weights with the opposite sign to the parametric models' (PBS and VARX) forecasts, meaning they partly cancel out. This is due to the fact that they lie very close to each other and are highly collinear. Thus, I plot the sum of the weights for PBS and VARX models represented by the red line in Figure 5.4. Figure B.5 in Appendix B illustrate the weights separately. Averaged regression-based weights have a lower variance than the weights obtained with DMSPE method, which is indicated by the fact that they lie close to their 1-month moving average (MA), represented by emboldened

lines in the figures. On the other hand, DMSPE weights are more unstable from day to day, fluctuating around 1-month MA. In figure 5.6, I provide an example how the optimal estimated weights vary in the given moneyness-maturity category.

Because the relative forecasting performance of the individual models vary across different regions of the surface, I find that the weights depended on the moneyness-maturity category are more effective than the weights assigned irrespective to the options' moneyness. Table 5.7 compares RMSPEV resulted from the combination weights estimated with respect to every moneyness-maturity categories with the weights estimated irrespective of the moneyness dimension. In every maturity category the moneyness dependent weighting gives smaller error. The most notable difference in RMSPEV can be seen for the short maturity category. The estimated optimal weights with respect to every moneyness-maturity category yield the average overall square error of 0.0107 in comparison to 0.0116 of the m independent optimal weights.

Table 5.7: RMSPEV comparison of moneyness dependent and independent weights

	Maturity					
	<60		**60-180**		**>180**	
	Disc	**Opt**	**Disc**	**Opt**	**Disc**	**Opt**
m **independent**	.0125	.0116	.0069	.0062	.0047	.0039
m **dependent**	.0121	.0107	.0069	.0058	.0046	.0037

The table compares RMSPEV resulted from the combination weights estimated with respect to every moneyness-maturity category with the weights estimated irrespective of the moneyness category and dependent only on the maturity. Disc- DMSPE method and Opt- optimal estimated weights with unrestricted regression. The performance measure is calculated for the whole evaluation period covering dates from 03/Mar/2003 until 30/Aug/2013.

Sub-periods combination forecasts

IV combination forecasts are also unstable over time. Table 5.8 reports the forecasting accuracy of the combination forecasts measured with RMSPEV for the three sub-periods. In the period before the recession of 2008, combination forecasts are more effective. Simple equal weighting method is able to outperform the most accurate individual model in a given moneyness-maturity category in 1 out of 7 moneyness groups for the short term contracts and 3 for the medium term options. Using this weighting scheme does not result in any improvement for the long term contracts in any of the three sub-periods. During the crisis time it beats the best individual model for the deep OTM puts for the short term as well as the medium term maturities, reducing RMSPEV from 0.0170 to 0.0158, and from 0.0100 to 0.0087 respectively. In the last sub-period, using equal weighting slightly reduces RMSPEV for the medium-term OTM put options. I observe quite similar behaviour for DMSPE

method. It is unable to systematically deliver a reduction in RMSPEV for the long term options in any of the three sub-periods. For the short maturities, it outperforms the best individual model in 4 out of 7 moneyness categories in the first sub-period, just the deep OTM puts during the crisis, and none in the most recent sub-period. For the medium maturities, it starts with smaller RMSPEV than the best individual model in all the moneyness categories, during the crisis it reduces the forecasting error in moneyness categories of $[0.99 - 1.01)$ and $[1.01 - 1.05)$, in the last sub-period it is effective only for the deep OTM puts. The regression based weights yield the most stable forecasting performance in the sense that at almost all occasions they are able to beat the best performing individual model in a given segment of IVS. In the first sub-period the estimated optimal weights do not yield improvement in reducing RMSPEV only for the short term deep OTM calls. In the times of the crisis, there are 3 moneyness categories for which the regression weights are inferior to the best individual model: $m \in [0.90 - 0.95)$, $m \in [0.95 - 0.99)$ and $m \in [1.10 - 1.15]$ for the short maturities. It reduces forecasting error for moneyness categories of $m \in [0.99 - 1.01)$, $m \in [1.01 - 1.05)$ and $m \in [1.10 - 1.15]$ for the long maturities. In the third sub-period, it reduce RMSPEV for all the short and medium term moneyness categories. For the long term options, it delivers forecasting improvement in the same categories as in the crisis time.

Table 5.8: Average daily RMSPEV in different sub-periods: combination forecasts

Panel A: before the 2008 crisis

	Maturity														
	<60					60-180					>180				
Moneyness	**RW**	**Ind**	**EW**	**Disc**	**Opt**	**RW**	**Ind**	**EW**	**Disc**	**Opt**	**RW**	**Ind**	**EW**	**Disc**	**Opt**
0.85-0.90	.0071	.0081	.0098	.0082	**.0075**	.0034	.0056	.0057	**.0055**	**.0041**	.0021	.0026	.0042	.0029	**.0024**
0.90-0.95	.0068	.0074	.0082	.0075	**.0073**	.0033	.0052	**.0050**	**.0051**	**.0040**	.0022	.0026	.0036	.0029	**.0023**
0.95-0.99	.0064	.0076	**.0070**	**.0069**	**.0067**	.0035	.0049	**.0047**	**.0048**	**.0041**	.0023	.0030	.0033	.0032	**.0024**
0.99-1.01	.0075	.0093	.0093	**.0088**	**.0073**	.0043	.0052	**.0051**	**.0050**	**.0046**	.0029	.0034	.0038	.0037	**.0030**
1.01-1.05	.0063	.0086	.0087	**.0082**	**.0069**	.0036	.0049	.0050	**.0046**	**.0040**	.0025	.0030	.0044	.0037	**.0026**
1.05-1.10	.0062	.0092	.0086	**.0083**	**.0079**	.0035	.0050	.0050	**.0047**	**.0041**	.0022	.0028	.0043	.0030	**.0024**
1.10-1.15	.0089	.0131	.0175	.0150	.0143	.0035	.0053	.0060	**.0052**	**.0044**	.0021	.0038	.0050	**.0035**	**.0027**
All	.0073	.0095	.0095	**.0088**	**.0078**	.0039	.0055	.0056	**.0054**	**.0045**	.0025	.0032	.0044	.0035	**.0027**

Panel B: during the 2008 crisis

	Maturity														
	<60					60-180					>180				
Moneyness	**RW**	**Ind**	**EW**	**Disc**	**Opt**	**RW**	**Ind**	**EW**	**Disc**	**Opt**	**RW**	**Ind**	**EW**	**Disc**	**Opt**
0.85-0.90	.0127	.0170	**.0158**	**.0161**	**.0168**	.0071	.0100	**.0087**	**.0094**	**.0079**	.0046	.0053	.0077	.0061	.0054
0.90-0.95	.0135	.0167	.0174	.0177	.0169	.0075	.0089	.0094	.0095	**.0087**	.0048	.0053	.0094	.0069	.0058
0.95-0.99	.0137	.0150	.0166	.0169	.0156	.0079	.0087	.0098	.0099	.0090	.0056	.0062	.0107	.0086	.0063
0.99-1.01	.0154	.0177	.0188	.0187	**.0171**	.0092	.0107	.0107	**.0106**	**.0100**	.0074	.0078	.0092	.0089	**.0077**
1.01-1.05	.0149	.0184	.0194	.0186	**.0165**	.0084	.0104	.0107	**.0099**	**.0093**	.0056	.0070	.0070	.0070	**.0064**
1.05-1.10	.0152	.0175	.0203	.0193	**.0172**	.0082	.0089	.0113	.0100	.0093	.0051	.0052	.0066	.0061	.0055
1.10-1.15	.0155	.0190	.0206	.0199	.0198	.0079	.0102	.0116	.0106	**.0094**	.0046	.0064	.0070	**.0068**	**.0055**
All	.0159	.0192	.0202	.0200	**.0188**	.0089	.0105	.0114	**.0110**	**.0101**	.0062	.0069	.0090	.0079	**.0069**

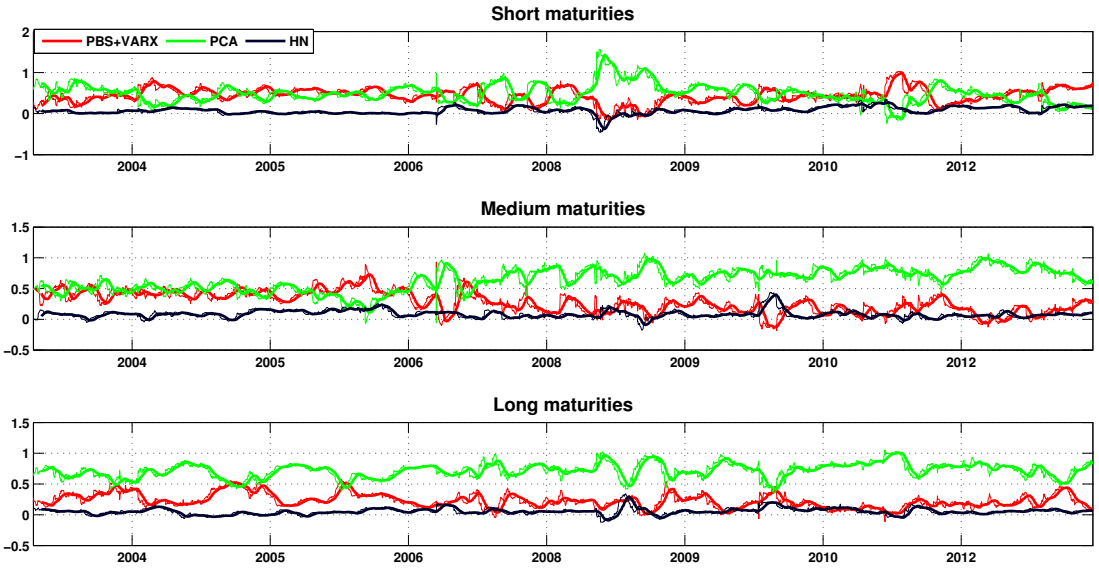
Panel C: after the 2008 crisis

	Maturity														
	<60					60-180					>180				
Moneyness	**RW**	**Ind**	**EW**	**Disc**	**Opt**	**RW**	**Ind**	**EW**	**Disc**	**Opt**	**RW**	**Ind**	**EW**	**Disc**	**Opt**
0.85-0.90	.0098	.0120	.0135	.0129	**.0116**	.0043	.0072	**.0071**	**.0069**	**.0055**	.0028	.0032	.0062	.0040	.0033
0.90-0.95	.0088	.0112	.0117	.0114	**.0108**	.0041	.0058	.0061	.0059	**.0053**	.0028	.0032	.0053	.0039	.0034
0.95-0.99	.0076	.0099	.0108	.0102	**.0096**	.0041	.0052	.0055	.0054	**.0049**	.0028	.0032	.0048	.0041	.0033
0.99-1.01	.0081	.0117	.0135	.0128	**.0103**	.0045	.0058	.0065	.0064	**.0055**	.0029	.0037	.0045	.0043	**.0035**
1.01-1.05	.0076	.0136	.0147	.0139	**.0102**	.0041	.0055	.0079	.0067	**.0053**	.0027	.0034	.0046	.0041	**.0033**
1.05-1.10	.0080	.0114	.0135	.0120	**.0103**	.0043	.0058	.0091	.0070	**.0057**	.0027	.0034	.0056	.0041	.0034
1.10-1.15	.0094	.0133	.0172	.0137	**.0127**	.0046	.0072	.0100	.0080	**.0063**	.0027	.0052	.0076	.0059	**.0040**
All	.0090	.0144	.0139	.0132	**.0113**	.0045	.0065	.0081	**.0071**	**.0058**	.0029	.0039	.0061	.0047	**.0037**

The table reports the average daily RMSPEV in the three sub-periods: 03/Mar/2003-31/Dec/2007, 01/Jan/2008-30/Jun/2009, 01/Jul/2009-30/Aug/2013, and for the different moneyness-maturity categories. RMSPEV is calculated for the forecasts restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. EW denotes trimmed equal weights, where Heston and Nandi model is not included, Disc is Discounted Mean Square Prediction Error (DMSPE) method that follows Stock and Watson (2004) and Opt are the estimated optimal weights with unrestricted regression with inclusion of intercept. Indv denotes the best performing individual model in a given moneyness-maturity category. The hold-out period for DMSPE and regression methods is set to 44 days. The emboldened values indicate the combining schemes that improve on the best individual forecast in a given moneyness-maturity category.

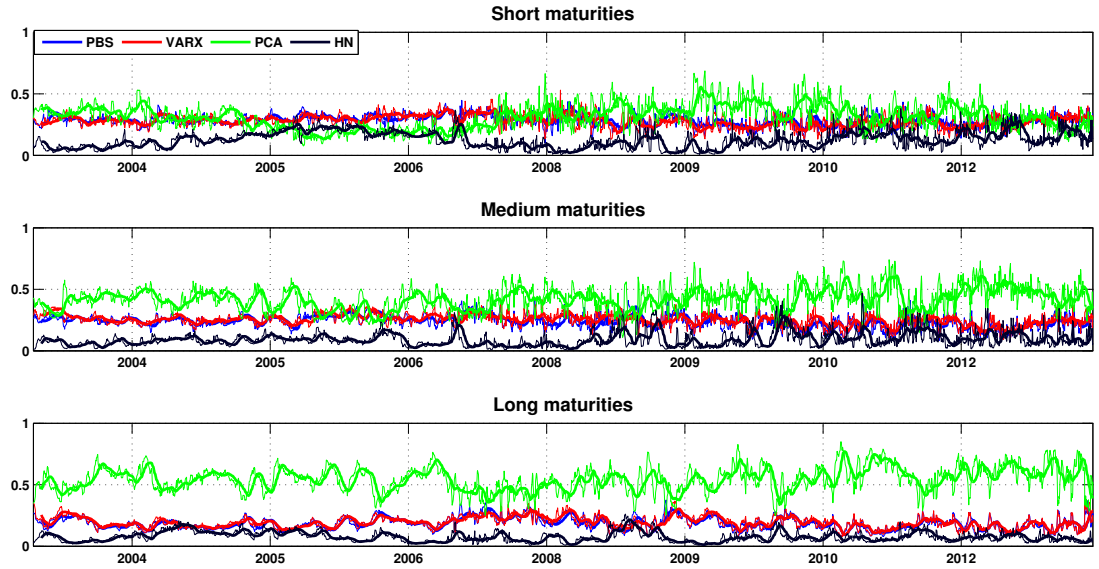
At the end of this subsection I provide two robustness checks regarding the combination forecasts methodology. First, I examine sensitivity of the out-of-sample results to the choice of the track-record period required to implement DMSPE and the regression based weighting. I consider the track-record periods of 22 and 66 days (approximately 1 and 3 month) as compared to the base period of 44 days. Second, I consider the discount factor values for DMSPE method of 0.9 and 1, as compared to the base value of 0.95. Table C.10 in Appendix C indicates that there is a little difference in RMSPEV if the models are estimated using a longer track record period of 66 or a shorter of 22 days. Also the value of the discount factor does not play an important role and the results are almost the same irrespective whether φ equals 0.9, 0.95 or 1.

Figure 5.4: Regression-based weights over time



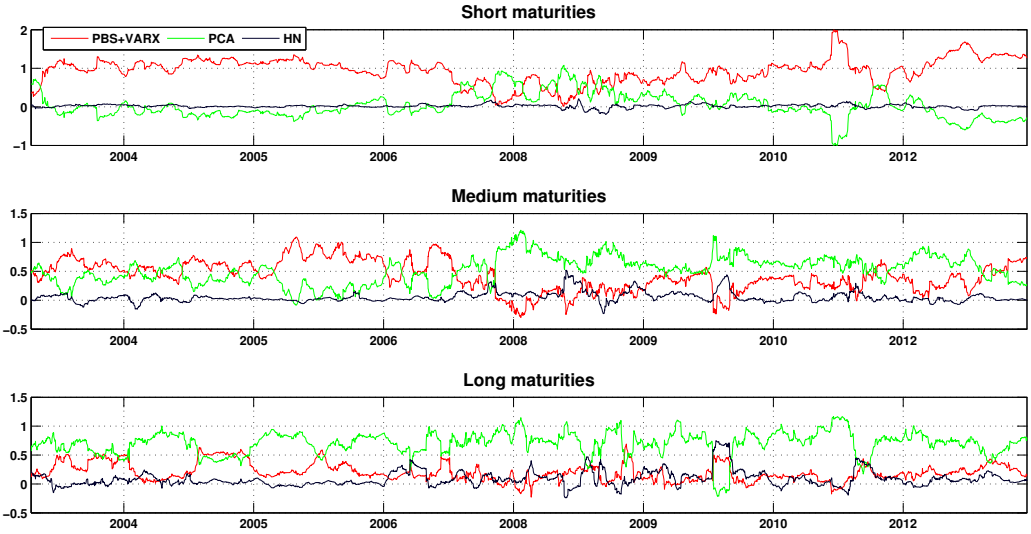
The figure illustrates the time variation of the estimated optimal weights from the unrestricted regression. The length of the track-record of the past forecasts is 44 days. Weights corresponding to PBS and VARX are added, as the models yield similar forecasts and their weights partly cancel out. PBS- Practitioner Black-Scholes model, VARX- Parametric VARX(p,q), PCA, and HN- Heston and Nandi GARCH model. For visual reasons the weights are normalized such that they sum up to one.

Figure 5.5: DMSPE-based weights over time



The figure illustrates the time variation of DMSPE-based weights. The combining method assigns greater weight to a model that exhibit relatively smaller MSPEV in the recent period. The length of the track-record of the past forecasts is 44 days. The discount factor is set to 0.95. PBS- Practitioner Black-Scholes model, VARX- Parametric VARX(p,q), PCA is the PCA model and HN is Heston and Nandi model.

Figure 5.6: Regression based weights for ATM options over time



The figure illustrates the time variation of the estimated optimal weights for ATM option with $m \in [0.99 - 1.01)$. The length of the track-record of the past forecasts is 44 days. Weights corresponding to PBS and VARX are added, as the models yield similar forecasts and their weights partly cancel out. PBS- Practitioner Black-Scholes model, VARX- Parametric VARX(p,q), PCA, and HN- Heston and Nandi GARCH model. For visual reasons the weights are normalized such that they sum up to one.

Statistical significance of forecasts accuracy

To formally compare a statistical significance of the differences in the forecasting performance of the benchmarking PBS model relative to VARX, PCA, HN and combination models, I employ the equal forecasting ability test developed by Diebold and Mariano (1995). The test uses mean squared prediction error in implied volatilities (MSPEV) as a relevant loss function. I test the null hypothesis of equal forecasts accuracy, against the alternative hypothesis that the benchmark model performs worse than a given competitor. DM statistic has an asymptotic standard normal distribution under the null hypothesis, however this result is valid only for nonnested models. Clark and McCracken (2001) and McCracken (2007) show that DM statistic has a nonstandard distribution when forecasting accuracy of nested models is assessed. In order to compare forecasts from PBS and VARX models (PBS is nested in VARX) a solution might be to use adjusted version of DM test introduced by Clark and West (2007) that provides asymptotically valid inferences for nested linear models. The adjusted-DM statistic is conveniently obtained by first defining the adjusted-MSPE statistic

f_{t+1} for one-step ahead forecasts as:

$$f_{t+1} = (y_{t+1} - \hat{y}_{1,t+1|t})^2 - [(y_{t+1} - \hat{y}_{2,t+1|t})^2 - (\hat{y}_{1,t+1|t} - \hat{y}_{2,t+1|t})^2], \tag{5.6}$$

where y_{t+1} is an observed variable of interest, $\hat{y}_{1,t+1|t}$ and $\hat{y}_{2,t+1|t}$ are one-step ahead forecasts from two competing models. Next, $\{f_{s+1}\}_{s=1}^{S}$, where S denotes the length of the evaluation period, is regressed on a constant and a t-statistic corresponding to the coefficient is obtained. P-value for the test is obtained with the standard normal distribution function. This procedure can be seen as testing whether the estimated mean of f_{t+1} differs significantly from zero, because regression on a constant is equivalent to taking the average of the dependent variable. Similarly to DMSPE-based combination forecasts, I adjust f_{t+1} such that it takes into account the fact that performance measures focus on the average daily MSPEV. Hence, the first and the second term are replaced by the MSPEV measure, while the squared forecasts difference is replaced with $\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\sigma}_{i,t+1|t}^{PBS} - \hat{\sigma}_{i,t+1|t}^{VARX}\right)^2$.
The last term in equation 5.6, $(\hat{y}_{1,t+1|t} - \hat{y}_{2,t+1|t})^2$, adjusts MSPE of a larger model for the upward bias in MSPE produced by estimation of parameters that are zero under the nested model. The idea behind the adjusted-DM test is to solve the problem that the DM test statistic can be heavily undersized when comparing forecasts from the nested models, leading to tests with low power. For example Rapach and Wohar (2006) find stronger evidence for the out-of-sample predictability of stock returns when using tests with appropriate size and power. The risk of using adjusted-DM test is that it indicates the VARX is significantly better than the nested benchmark even if the difference is MSPEV is very small. Ultimately, I take more conservative approach in the sense that VARX's statistic may be undersized and report standard DM test results for all the models, testing the null of equal forecasting ability, against the alternative that the PBS model has larger MSPEV.

Table 5.9 shows the results of Diebold Mariano test with respect the three distinguished sub-periods. The negative values of the test statistic indicate than the benchmark PBS model performs better. To start with VARX model, it improves its relative forecasting accuracy for the short maturities over time. In the before the crisis sub-period, it beats PBS significantly only in the moneyness category of $m \in [1.05, 1.10)$ at the $p = 0.001$ significance level. In the after-the-crisis sub-period it significantly outperforms the benchmark additionally for options with $m \in [0.85, 0.90)$. For the medium maturities, there is a noticeable switch, as VARX outperforms PBS for all the OTM calls, ATM options and none puts in the first sub-period, while in the last sub-period it remains to be significantly better for the deep OTM calls and calls with $m \in [1.05, 1.10)$ and turns out to be significantly better for all the OTM put options. For the long maturities VARX exhibits significant improvement over the benchmark

for the deep OTM calls in the first sub-period at 5% level. I reject the null in favour of the alternative that VARX model yields more accurate forecasts than the benchmark for all the OTM puts and ATM options in the last sub-period. During the financial turmoil it can be regarded as a more accurate model for predicting IV of the long term options in all the moneyness categories, as well as irrespectively of the moneyness category. When it comes to the option pricing approach based on Heston and Nandi (2000) the null hypothesis of equal forecasting ability can be rejected only in the last sub-period for the deep OTM calls in the short maturity category, at the 5% significance level.

PCA model outperforms PBS significantly in all regions of the surface for the medium and the long term options usually at the highest significance level, even during the financial crisis. The exception are the deep OTM put options in the medium term category in the crisis period. For the short term contracts, it outperforms the benchmark for two maturity categories when $m \in [1.05 - 1.15]$ in the first sub-period and all the OTM calls together with ATM options and put with $m \in [0.95 - 0.99]$ in the last sub-period. During the crisis it is significantly better for all the moneyness categories excluding the deep OTM puts.

For combination forecasts based on DMSPE and the estimated optimal weights methods the null cannot be rejected only at the rare occasions in the short maturity category. For the regression-based weights it happens for the puts with $m \in [0.85 - 0.90)$ in the first sub-period, $m \in [0.85 - 0.95)$ in the last sub-period. In the crisis time- regression-based weights do not significantly improve over PBS for the three categories of the short term OTM put options. When considering forecasts with DMSPE-based weights, the null cannot be rejected for the short term put options in the moneyness range of [0.85-0.95) for the first sub-period, as well as the last one.

Table 5.9: Diebold-Mariano test of equal forecasting ability- individual and combination forecasts

	Maturity														
	<60					60-180					>180				
Moneyness	**VARX**	**PCA**	**HN**	**Disc**	**Opt**	**VARX**	**PCA**	**HN**	**Disc**	**Opt**	**VARX**	**PCA**	**HN**	**Disc**	**Opt**
0.85-0.90	-2.5	-23.35	-24.04	-3.63	3.02**	-1.8	10.55***	-22.88	20.97***	22.07***	0.01	24.7***	-10.82	28.89***	25.41***
0.90-0.95	-4.51	-19.3	-25.65	-4.69	0.28	-3.59	8.53***	-23.21	17.23***	16.03***	-2.61	21.02***	-24.33	24.74***	17.22***
0.95-0.99	-7.04	-0.76	-24.6	6.63***	6.3***	-1.43	3.4***	-24.48	5.5***	8.96***	-3.16	12.98***	-44.4	15.56***	15.72***
0.99-1.01	-4.8	1.37	-19.95	8.58***	10.04***	2.06*	4.61***	-22.05	8.04***	7.44***	-2	11.43***	-42.58	14.79***	13.43***
1.01-1.05	1.52	-11.14	-15.66	3.96***	7.55***	4.15***	8.5***	-14.31	11.7***	10.65***	-2.92	23.89***	-31.13	25.5***	23.33***
1.05-1.10	4.53***	6.9***	-10.47	10.29***	7.6***	6.56***	8.62***	-15.92	10.74***	10.87***	-0.23	28.54***	-10.46	30.16***	29.58***
1.10-1.15	-0.16	6.78***	-0.73	5.67***	4.14***	7.38***	11.85***	-14.13	14.41***	11.16***	2.17*	21.28***	-7.28	28.83***	29***
All	-1.48	-18.89	-26.74	7.18***	10.36***	4.76***	14.37***	-26.42	19.11***	18.61***	-1.38	35.54***	-35.4	40.17***	35.05***

Panel B: during the 2008 crisis

	Maturity														
	<60					60-180					>180				
Moneyness	**VARX**	**PCA**	**HN**	**Disc**	**Opt**	**VARX**	**PCA**	**HN**	**Disc**	**Opt**	**VARX**	**PCA**	**HN**	**Disc**	**Opt**
0.85-0.90	-1.06	-0.07	-17.46	2.12*	0.24	-0.22	0.05	-12.91	3.54***	3.02**	2.23*	9.65***	-0.79	11.06***	10.39***
0.90-0.95	-1.77	4***	-17.39	2.89**	1.08	-0.04	2.39**	-12.65	5.23***	2.16*	2.69**	11.59***	-1.95	11.82***	10.92***
0.95-0.99	-1.7	4.7***	-14.61	3.39***	1.27	0.21	2.8**	-12.45	4.38***	1.77*	2.61**	10.88***	-4.37	10.41***	10.12***
0.99-1.01	-2.81	4.39***	-11.44	3.54***	1.97*	-0.15	1.68*	-8.89	3.34***	1.78*	1.93*	5.7***	-11.43	5.92***	5.05***
1.01-1.05	-3.28	4.58***	-8.36	4.89***	2.58**	-0.14	2.25*	-4.8	4.9***	3.31***	3.04**	4.6***	-22.44	6.47***	4.5***
1.05-1.10	-2.19	6.94***	-7.94	5.85***	3.87***	0.71	6.39***	-4.74	7.53***	5.08***	2.02*	7.59***	-16.17	7.49***	4.85***
1.10-1.15	-1.79	6.21***	-7.27	3.02**	3.78***	2.37**	7.17***	-11.24	8.01***	7.12***	2.05*	5.28***	-7.48	6.21***	6.3***
All	-2.47	5.33***	-15.81	4.16***	2.04*	0.74	3.66***	-12.61	6.56***	4.44***	2.69**	10.3***	-11.47	10.62***	9.8***

Panel C: after the 2008 crisis

	Maturity														
	<60					60-180					>180				
Moneyness	**VARX**	**PCA**	**HN**	**Disc**	**Opt**	**VARX**	**PCA**	**HN**	**Disc**	**Opt**	**VARX**	**PCA**	**HN**	**Disc**	**Opt**
0.85-0.90	2.26*	-16.44	-29.01	-4.12	1.23	12.44***	10.42***	-19.69	17.47***	14.47***	14.21***	25.6***	-6.87	27.15***	22.47***
0.90-0.95	-4.04	-13.31	-29.75	-3.44	0.01	7.13***	9.86***	-15.46	14.33***	8.61***	12.1***	19.65***	-22.7	20.38***	16.12***
0.95-0.99	-8.83	2.19*	-22.87	8.97***	2.08*	2.28*	5.97***	-21.79	6.48***	4.9***	7.43***	15.84***	-47.2	15.49***	10.03***
0.99-1.01	-6.93	9***	-18.52	9.31***	5.07***	0.83	6.46***	-21.26	6.26***	7.84***	2.2*	11.53***	-59.65	12.65***	8.84***
1.01-1.05	-0.75	8.09***	-14.68	9.28***	7.86***	1.26	12.09***	-4.27	11.62***	14.53***	-3.84	17.42***	-50.15	17.77***	12.17***
1.05-1.10	4.43***	13.03***	-6.6	11.57***	7.02***	4.17***	15.36***	-2.41	15.88***	16.84***	-6.43	25.36***	-14.91	25.75***	17.94***
1.10-1.15	0.8	11.74***	1.99*	8.83***	8.27***	8.21***	16.98***	-7.2	18.15***	17.74***	-4.59	21.27***	-1.52	21.54***	23.07***
All	-1.21	-5.74	-33.39	9.11***	4.37***	6.04***	15.77***	-14.08	16.99***	18.3***	7***	31.06***	-47.31	31.5***	20.61***

The table contains the results of Diebold and Mariano (1995) test for equal predictive ability in the three sub-periods: 03/Mar/2003-31/Dec/2007, 01/Jan/2008-30/Jun/2009, 01/Jul/2009-30/Aug/2013, and for the different moneyness-maturity categories. It reports the DM statistic which is calculated to test the null hypothesis of the equal forecasting against the benchmark Practitioner Black-Scholes model. The loss function is MSPEV. Under the null DM statistic is assumed to be standard normally distributed. *** denotes the significance at 0.1% level, ** 1%, *5%, ' 10%. VARX, PCA, HN are the individual models, Disc is DMSPE combination model, Opt is the combination model with regression based weights.

Chapter 6

Conclusion

In this thesis I investigate predictable dynamics of the implied volatility surface. In assessing predictability, I concentrate on short-horizon, one-step ahead forecasts of IVS. I evaluate forecasting performance of the three approaches, which are as follows. In the first approach, I model dynamics of the factors and produce their forecasts by means of VAR-type model. The second approach tries to identify persistent latent factors that drive dynamics of IVS. I estimate the (latent) factor model with PCA and find that the first three principal components can be interpreted as the level, smile and term-structure factors. To capture the factor dynamics, I again use the parametric VAR-type model. Because IV data has a three-dimensional surface structure (IV can be seen as a function of moneyness and maturity) rather than form of time-series, PCA cannot be directly applied. In the smoothing procedure with the nonparametric kernel regression, I recover time series on a given grid of moneyness and maturity. To enable application of PCA model for the purpose of out-of-sample forecasting, I again apply smoothing procedure to the forecasted points on the grid which delivers forecasts of IVs for all the options. The second approach is treated as a benchmark and can be characterized as a practitioners' framework which relies on a linear regression model that links cross-section of IVs to polynomials in options' moneyness and maturity. Estimated daily regression coefficients are interpreted as factors and exhibit considerable time-variation. The extension of this approach includes factor dynamics, which follows Goncalves and Guidolin (2006). Here, the procedure consists of two steps. First, I estimate regression model on each day and obtain multivariate time-series of factors. The third approach relies on the option pricing model estimated in the prices space, for which I choose NGARCH(1,1) option valuation model of Heston and Nandi (2000). The pricing formula depends on the one-step ahead forecast of the variance and 5 parameters driving NGARCH(1,1) process. The model is estimated such that it exploits the combined information in the history of S&P500 prices and the cross-section of options. The variance parameter

is estimated from the history of the prices, while remaining parameters are obtained in calibration process by means of the non-linear least squares procedure. These parameters are used to predict tomorrow's prices. Once the forecasted prices are obtained, I invert the Black-Scholes-Merton formula and back-out the (forecasted) implied volatilities from the forecasts of the prices.

I evaluate forecasting performance in statistical terms with respect to different regions of IVS by partitioning it in 21 moneyness-maturity categories. I find that PCA model considerably outperforms other approaches for the medium and the long term options irrespective of moneyness bin. The most challenging task is forecasting IVs of the short term contracts. Here, the best performers are PCA and parametric approaches interchangeably. PCA shows the overall improvement in this region of the surface over time and outperforms parametric approaches in 6 out of 7 moneyness categories during years 2009-2013. Forecasting accuracy can be further improved when model combination is used. I find that the combining method based on the estimated optimal weights yields more accurate forecasts in nearly every segment of IVS consistently over time. However, none of the approaches is accurate enough to beat the random walk forecasts, that assume that IVs do not change in the forecasting horizon of one day.

This thesis addressed the predictability of IVS in one-day horizon. Considered models were not able to beat random walk forecasts. However, some statistical patterns were identified. The direct extension of the presented analysis would be to examine IVS predictability with a focus on the forecasting horizon of a different length, for example one week. Moreover, the analysis conducted in this thesis can be further extended in various directions in future research. First, the usefulness of option pricing approach to modelling IVS based on a selected model, not necessarily the GARCH-type option valuation model considered in this research, can be further investigated by using different calibration settings. Perhaps the most interesting question is whether the model could improve its forecasting accuracy of IVS if it was calibrated directly in the IVS space. Christoffersen and Jacobs (2004) argue that when comparing models the estimation loss function should be the same for all of them, otherwise comparison is inappropriate. Moreover, the loss functions used in the estimation and evaluation of a given model should be aligned, otherwise the estimated set of parameters may be suboptimal. This problem was not addressed in this thesis because of numerical reasons. The calibration process repeated on every of 2644 days to price and forecast over 1 million of options in total took fairly long time, while the calibration in the implied volatility space would even lengthen that time, making the model infeasible to apply on such a large data set. Finally, it could be investigated how the model performance depends on the specific loss function used in the calibration process. I used squared deviation as a relevant loss function,

but alternatives include (weighted) relative squared or absolute deviation. In order to obtain better minima more starting values for the numerical optimization could be considered. When it comes to PCA model, there is also room for improvement, especially in the kernel smoothing part of the model as some choices I made were of *ad-hoc* character. First, the formal procedure for bandwidth selection should be proposed that takes into account design of IV data. Second, one could investigate how forecasting accuracy depends on a number of chosen grid points where the surface is recovered. Third, an alternative estimation procedure for latent factors could be used, which is Kalman filter approach as presented in Van der Wel et al. (2015). One may focus on improving the forecasting performance of the parametric models of IVS. Given the fact that slopes of the volatility smile and term structure differ across different regions of the surface, it may be sensible to make the parameters region dependent. To further investigate benefits of model combination, the economic significance of the results could be tested. This can be done by analyzing whether one can design profitable trading strategies based on the combined forecasts.

Bibliography

Badshah, I. (2009), 'Modeling the dynamics of implied volatility surfaces', *Available at SSRN 1347981* .

Bai, J. and Ng, S. (2004), 'A panic attack on unit roots and cointegration', *Econometrica* pp. 1127–1177.

Bakshi, G., Cao, C. and Chen, Z. (1997), 'Empirical performance of alternative option pricing models', *The Journal of Finance* **52**(5), 2003–2049.

Barone-Adesi, G., Engle, R. F. and Mancini, L. (2008), 'A garch option pricing model with filtered historical simulation', *Review of Financial Studies* .

Bates, D. S. (1996), 'Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options', *Review of financial studies* **9**(1), 69–107.

Bates, D. S. (1997), 'The skewness premium: Option pricing under asymmetric processes', *Advances in Futures and Options Research* **9**, 51–82.

Bernales, A. and Guidolin, M. (2014), 'Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests', *Journal of Banking & Finance* **46**, 326–342.

Black, F. (1976), 'Stuedies of stock price volatility changes, proceedings of the 1976 meetings of the american statistical association, business and economic statistic section', pp. 177–181.

Black, F. and Scholes, M. (1973), 'The pricing of options and corporate liabilities', *The journal of political economy* pp. 637–654.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of econometrics* **31**(3), 307–327.

Brooks, C. and Oozeer, M. C. (2002), 'Modelling the implied volatility of options on long gilt futures', *Journal of Business Finance & Accounting* **29**(1-2), 111–137.

Carr, P. and Madan, D. (1999), 'Option valuation using the fast fourier transform', *Journal of computational finance* **2**(4), 61–73.

Chalamandaris, G. and Tsekrekos, A. E. (2009), 'can static models predict implied volatility surfaces? evidence from otc currency options', Technical report, Working Paper, Athens University of Economics and Business, Athens.

Chalamandaris, G. and Tsekrekos, A. E. (2010), 'Predictable dynamics in implied volatility surfaces from otc currency options', *Journal of Banking & Finance* **34**(6), 1175–1188.

Chamberlain, G. and Rothschild, M. (1982), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometica* **51**, 1281–1304.

Chorro, C., Guégan, D. and Ielpo, F. (2014), *A Time Series Approach to Option Pricing: Models, Methods and Empirical Performances*, Springer.

Christie, A. A. (1982), 'The stochastic behavior of common stock variances: Value, leverage and interest rate effects', *Journal of financial Economics* **10**(4), 407–432.

Christoffersen, P., Fournier, M. and Jacobs, K. (2013), 'The factor structure in equity options', *Rotman School of Management Working Paper* (2224270).

Christoffersen, P. and Jacobs, K. (2004), 'The importance of the loss function in option valuation', *Journal of Financial Economics* **72**(2), 291–318.

Clark, T. E. and McCracken, M. W. (2001), 'Tests of equal forecast accuracy and encompassing for nested models', *Journal of econometrics* **105**(1), 85–110.

Clark, T. E. and West, K. D. (2007), 'Approximately normal tests for equal predictive accuracy in nested models', *Journal of econometrics* **138**(1), 291–311.

Cont, R. and da Fonseca, J. (2002), 'Dynamics of implied volatility surfaces', *Quantitative finance* **2**(1), 45–60.

Das, S. R. and Sundaram, R. K. (1999), 'Of smiles and smirks: A term structure perspective', *Journal of Financial and Quantitative Analysis* **34**(02), 211–239.

Diebold, F. X. (1988), 'Serial correlation and the combination of forecasts', *Journal of Business & Economic Statistics* **6**(1), 105–111.

Diebold, F. X. and Li, C. (2006), 'Forecasting the term structure of government bond yields', *Journal of econometrics* **130**(2), 337–364.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business & economic statistics* .

Diebold, F. X. and Pauly, P. (1987), 'Structural change and the combination of forecasts', *Journal of Forecasting* **6**(1), 21–40.

Duan, J.-C. et al. (1995), 'The garch option pricing model', *Mathematical finance* **5**(1), 13–32.

Dumas, B., Fleming, J. and Whaley, R. E. (1998), 'Implied volatility functions: Empirical tests', *The Journal of Finance* **53**(6), 2059–2106.

Engle, R. F. and Ng, V. K. (1993), 'Measuring and testing the impact of news on volatility', *The journal of finance* **48**(5), 1749–1778.

Fengler, M. R., Härdle, W. K. and Mammen, E. (2007), 'A semiparametric factor model for implied volatility surface dynamics', *Journal of Financial Econometrics* **5**(2), 189–218.

Fengler, M. R., Härdle, W. K. and Villa, C. (2003), 'The dynamics of implied volatilities: A common principal components approach', *Review of Derivatives Research* **6**(3), 179–202.

Goncalves, S. and Guidolin, M. (2006), 'Predictable dynamics in the s&p 500 index options implied volatility surface*', *The Journal of Business* **79**(3), 1591–1635.

Granger, C. W. and Ramanathan, R. (1984), 'Improved methods of combining forecasts', *Journal of Forecasting* **3**(2), 197–204.

Guo, D. (2000), 'Dynamic volatility trading strategies in the currency option market', *Review of derivatives research* **4**(2), 133–154.

Hardle, W. (1990), 'Applied nonparametric regression', *Cambridge, UK* .

Harvey, C. R. and Whaley, R. E. (1992), 'Market volatility prediction and the efficiency of the s & p 100 index option market', *Journal of Financial Economics* **31**(1), 43–73.

Hentschel, L. (2003), 'Errors in implied volatility estimation', *Journal of Financial and Quantitative analysis* **38**(04), 779–810.

Heston, S. L. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *Review of financial studies* **6**(2), 327–343.

Heston, S. L. and Nandi, S. (2000), 'A closed-form garch option valuation model', *Review of Financial Studies* **13**(3), 585–625.

Hull, J. and White, A. (1987), 'The pricing of options on assets with stochastic volatilities', *The journal of finance* **42**(2), 281–300.

Konstantinidi, E., Skiadopoulos, G. and Tzagkaraki, E. (2008), 'Can the evolution of implied volatility be forecasted? evidence from european and us implied volatility indices', *Journal of Banking & Finance* **32**(11), 2401–2411.

le Roux, M. (2007), 'A long-term model of the dynamics of the s&p500 implied volatility surface', *North American Actuarial Journal* **11**(4), 61–75.

McCracken, M. W. (2007), 'Asymptotics for out of sample tests of granger causality', *Journal of Econometrics* **140**(2), 719–752.

Merton, R. C. (1976), 'Option pricing when underlying stock returns are discontinuous', *Journal of financial economics* **3**(1), 125–144.

Mikhailov, S. and Nögel, U. (2004), *Heston's stochastic volatility model: Implementation, calibration and some extensions*, John Wiley and Sons.

Mixon, S. (2002), 'Factors explaining movements in the implied volatility surface', *Journal of Futures Markets* **22**(10), 915–937.

Moodley, N. (2005), 'The heston model: A practical approach with matlab code', *Bachelor Thesis, University of the Witwatersrand, Johannesburg, math. nyu. edu* .

OptionMetrics, L. (2011), 'Ivy db file and data reference manual , version 3.0'.

Pena, I., Rubio, G. and Serna, G. (1999), 'Why do we smile? on the determinants of the implied volatility function', *Journal of Banking & Finance* **23**(8), 1151–1179.

Pindyck, Robert, S. and Rubinfeld, D. L. (1998), 'Econometric models and econometric forecasts'.

Rapach, D. E. and Wohar, M. E. (2006), 'In-sample vs. out-of-sample tests of stock return predictability in the context of data mining', *Journal of Empirical Finance* **13**(2), 231–247.

Scott, L. O. (1997), 'Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: Applications of fourier inversion methods', *Mathematical Finance* **7**(4), 413–426.

Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, Vol. 26, CRC press.

Skiadopoulos, G., Hodges, S. and Clewlow, L. (1999), 'The dynamics of the s&p 500 implied volatility surface', *Review of Derivatives Research* **3**(3), 263–282.

Stock, J. H. and Watson, M. W. (2004), 'Combination forecasts of output growth in a seven-country data set', *Journal of Forecasting* **23**(6), 405–430.

Stock, J. H. and Watson, M. W. (2006), 'Forecasting with many predictors', *Handbook of economic forecasting* **1**, 515–554.

Timmermann, A. (2006), 'Forecast combinations', *Handbook of economic forecasting* **1**, 135–196.

Tompkins, R. G. (2001), 'Implied volatility surfaces: uncovering regularities for options on financial futures', *The European Journal of Finance* **7**(3), 198–230.

Tompkins, R. G. et al. (2001), 'Stock index futures markets: stochastic volatility models and smiles', *Journal of Futures Markets* **21**(1), 43–78.

Van der Wel, M., Ozturk, S. R. and Van Dijk, D. J. (2015), 'Dynamic factor models for the volatility surface', *Available at SSRN 2558018* .

Walker, J. S. (1996), *Fast fourier transforms*, Vol. 24, CRC press.

Winkler, R. L. and Makridakis, S. (1983), 'The combination of forecasts', *Journal of the Royal Statistical Society. Series A (General)* pp. 150–157.

Appendix A

Fast Fourier Transform (FFT)

Carr and Madan (1999) introduce a method based on the Fast Fourier Transform (FFT) to price option contracts that can be exploited when the characteristic function of the return is known analytically. Roughly speaking, FFT is an efficient algorithm to calculate the sums $w_1, \ldots, w_N$ where

$$w_k = \sum_{j=1}^{N} \exp\left(-i\frac{2\pi}{N}(j-1)(k-1)\right) x(j). \tag{A.1}$$

For a more detailed description on what FFT is, see Walker (1996). Carr and Madan derive two formulas depending whether an option has an intrinsic value or not, that is they distinguish between in-the-money (including at-the-money) and out-of-the-money contracts. The in-the-money call is worth

$$C_T(k_u) \approx \frac{e^{-\alpha k_u}}{\pi} \sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(u-1)} \underbrace{e^{ibv_j}\psi(v_j)\frac{\eta}{3}(3+(-1)^j-\delta_{j-1})}_{\star}, \tag{A.2}$$

where

$$\begin{aligned} b &= \frac{\pi}{\eta} \\ \eta &= 0.25 \\ N &= 2^{12} \\ \eta &= 0.25 \\ v_j &= \eta(j-1) \\ k_u &= -b + \frac{2b}{N}(u-1), \text{ for } \quad u = 1, 2, \ldots, N+1 \\ \alpha &= 1.5 \end{aligned}$$

and $\psi(\cdot)$ is a characteristic function of the logarithm of the spot price under the risk neutral distribution. In the model of Heston and Nandi (2000) it can be obtained by replacing ϕ by $i\phi$ in the moment generating function of the $\log(S_T)$, which is expressed under the risk-neutral distribution as

$$f^*(\phi) = S_t^{\phi} e^{A_t(T,\phi)+B_t(T,\phi)h_{t+1}}, \tag{A.4}$$

where

$$A_t(T,\phi) = A_{t+1}(T,\phi) + \phi r_{f,t} + \omega B_{t+1}(T,\phi) - \frac{1}{2}\log\left(1 - 2\alpha B_{t+1}(T,\phi)\right) \tag{A.5a}$$

$$B_t(T,\phi) = -\frac{1}{2}\phi + \beta B_{t+1}(T,\phi) + \frac{\frac{\phi^2}{2} - 2\alpha\gamma^* B_{t+1}(T,\phi)\phi + \alpha B_{t+1}(T,\phi)(\gamma^*)^2}{1 - 2\alpha B_{t+1}(T,\phi)}, \tag{A.5b}$$

with terminal conditions $A_T(T,\phi) = B_T(T,\phi) = 0$. In choosing parameters values in equation (A.2) I follow Carr and Madan (1999). The authors argue that such values of the parameters deliver speedup of FFT without compromising the accuracy that other methods can provide. In order to calculate the output of equation (A.2), one needs to plug the part of (A.2) represented by "$\star$" into the Matlab function *fft(X)*.

The out-of-the-money call in the framework of Carr and Madan (1999) is worth

$$C_T(k_u) \approx \frac{1}{\pi \sinh(\alpha k_u)} \sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(u-1)} e^{ibv_j} \gamma_T(v_j) \frac{\eta}{3}(3 + (-1)^j - \delta_{j-1}), \tag{A.6}$$

where

$$\gamma_T(v_j) = \frac{\varsigma_T(v_j - i\alpha) - \varsigma_T(v_j + i\alpha))}{2} \tag{A.7a}$$

$$\varsigma_T(v_j) = e^{-r_f T}\left[\frac{1}{1+iv_j} - \frac{e^{r_f T}}{iv_j} - \frac{\psi(v_j - i)}{v_j^2 - iv_j}\right]. \tag{A.7b}$$

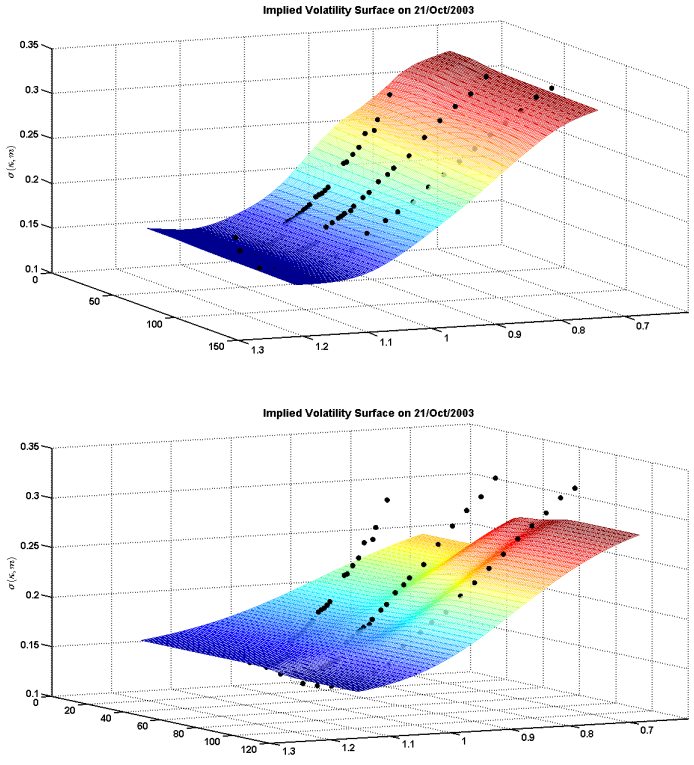
Similarly to Moodley (2005), who applies FFT approach in the context of the Heston (1993) model, I find very little difference between prices obtained with equations (A.2) and (A.6). Thus, following Chorro et al. (2014) and Moodley (2005) I price the options using equation (A.2) regardless of the moneyness of the option. This allows for significant savings in the computation time during the calibration process, as the characteristic function (A.4) has to be calculated only once in each iteration of the calibration, instead of being calculated twice if equation (A.6) was used.

A more elaborate discussion on the implementation of the method of Carr and Madan (1999) can be found in Chorro et al. (2014), Mikhailov and Nögel (2004) and Moodley (2005).

Appendix B

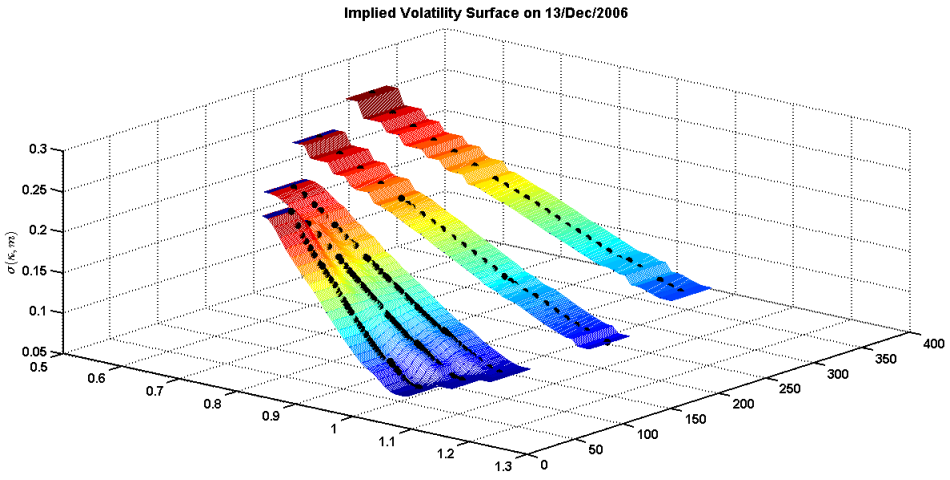
Figures

Figure B.1: Kernel smoothing of IVS



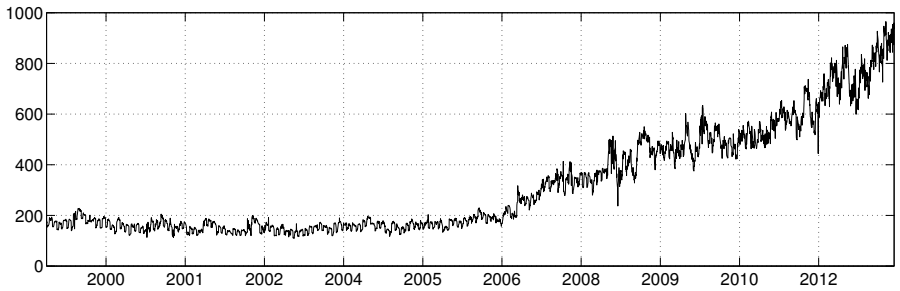
IVS obtained with quartic kernel (top panel) and Gaussian kernel (bottom panel) on 21/Oct/2002. Gaussian kernel fits clearly worse to actual data.

Figure B.2: Smoothed IVS with optimal bandwidths



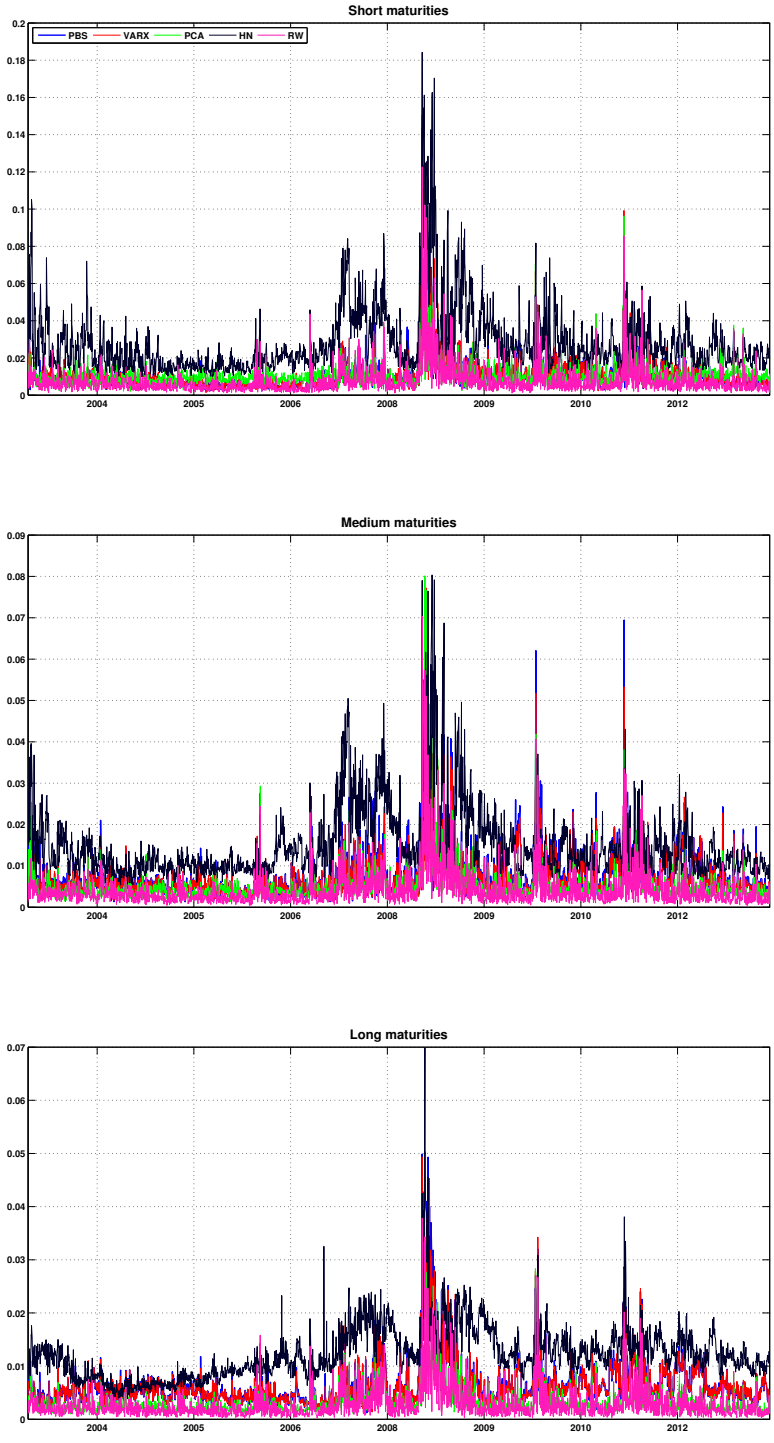
IVS obtained on 13/Dec/2006 with optimal bandwidths with respect to penalizing function given by equation (2.3). Clearly the bandwidths are to narrow in both dimensions, causing the smoothed surface bumpy and discontinuous.

Figure B.3: Number of quoted contracts per day over time



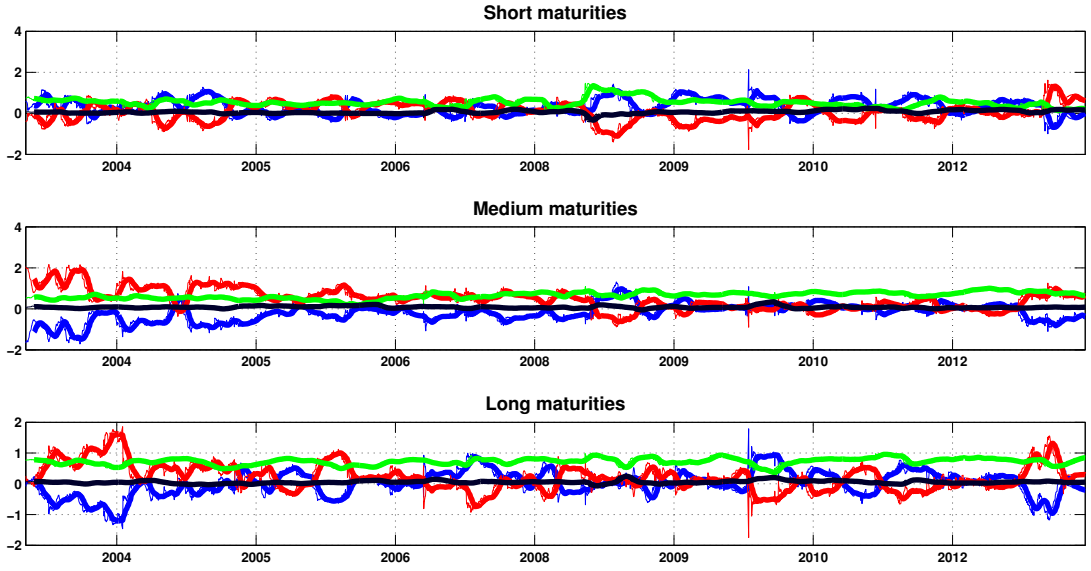
The figure illustrates a number of quoted contracts per day in the data set used in this study. It takes into account only the contracts that are left after the filters of section 2.1 are applied, i.e. those which are effectively used in to estimate or calibrate forecasting models.

Figure B.4: Out-of-sample daily MAPEV over time



The figure illustrates the evolution of daily MAPEV for the four models evaluated in terms of out-of-sample fit: PBS- Practitioner Black-Scholes, VARX- Parametric VARX(p,q), PCA, and HN- Heston and Nandi GARCH model.

Figure B.5: Regression-based weights over time



The figure illustrates the time variation of the estimated optimal weights from the unrestricted regression. The length of the track-record of the past forecasts is 44 days. PBS- Practitioner Black-Scholes model, VARX- Parametric VARX(p,q), PCA, and HN- Heston and Nandi GARCH model.

Appendix C

Tables

Table C.1: Cross-correlations of IV

		20								320							
Maturity	**Moneyness**	**0.85**	**0.90**	**0.95**	**0.99**	**1.01**	**1.05**	**1.10**	**1.12**	**0.85**	**0.90**	**0.95**	**0.99**	**1.01**	**1.05**	**1.10**	**1.12**
20	**0.85**	0.007	0.992	0.977	0.965	0.957	0.942	0.934	0.926	0.909	0.914	0.916	0.915	0.913	0.909	0.909	0.909
	0.90		0.007	0.994	0.983	0.977	0.963	0.955	0.944	0.907	0.913	0.917	0.917	0.915	0.911	0.911	0.911
	0.95			0.007	0.995	0.990	0.978	0.971	0.960	0.904	0.912	0.918	0.918	0.916	0.912	0.912	0.914
	0.99				0.007	0.999	0.992	0.986	0.975	0.900	0.910	0.917	0.920	0.919	0.918	0.920	0.923
	1.01					0.007	0.997	0.992	0.982	0.895	0.905	0.913	0.917	0.917	0.917	0.920	0.924
	1.05						0.006	0.999	0.990	0.880	0.891	0.901	0.906	0.908	0.911	0.915	0.920
	1.10							0.005	0.995	0.871	0.883	0.893	0.900	0.901	0.905	0.910	0.916
	1.12								0.005	0.863	0.874	0.885	0.891	0.893	0.897	0.902	0.909
320	**0.85**									0.003	0.999	0.997	0.993	0.990	0.984	0.981	0.977
	0.90										0.003	0.999	0.996	0.994	0.988	0.985	0.982
	0.95											0.003	0.998	0.996	0.991	0.989	0.987
	0.99												0.003	0.999	0.996	0.995	0.993
	1.01													0.003	0.999	0.997	0.995
	1.05														0.003	1.000	0.998
	1.10															0.003	0.999
	1.12																0.003

The table contains cross-correlations (upper triangle) and variances (diagonal) of the implied volatilities. Time series used to calculate the statistics are obtained with Nadaraya-Watson estimator given by equation (2.1).

Table C.2: (Partial) autocorrelations of IV

		ACF						PACF					
κ	m	**1**	**2**	**3**	**4**	**5**	**22**	**1**	**2**	**3**	**4**	**5**	**22**
20	**0.85**	0.979	0.963	0.949	0.937	0.927	0.803	0.979	0.094	0.080	0.043	0.057	-0.030
	0.90	0.978	0.962	0.949	0.938	0.929	0.800	0.978	0.101	0.101	0.072	0.049	-0.022
	0.95	0.979	0.964	0.952	0.943	0.934	0.805	0.979	0.137	0.098	0.085	0.030	-0.006
	0.99	0.980	0.966	0.956	0.947	0.938	0.815	0.980	0.165	0.102	0.057	0.036	-0.014
	1.01	0.978	0.966	0.956	0.946	0.938	0.817	0.979	0.192	0.109	0.051	0.045	-0.022
	1.05	0.975	0.962	0.952	0.942	0.934	0.817	0.975	0.235	0.119	0.051	0.063	-0.030
	1.10	0.974	0.961	0.951	0.942	0.934	0.818	0.974	0.238	0.122	0.048	0.068	-0.026
	1.15	0.969	0.956	0.947	0.937	0.929	0.811	0.969	0.274	0.146	0.060	0.060	-0.014
320	**0.85**	0.994	0.988	0.983	0.978	0.974	0.898	0.994	0.063	0.042	0.000	0.044	-0.002
	0.90	0.994	0.988	0.983	0.978	0.974	0.899	0.994	0.067	0.049	0.001	0.049	-0.002
	0.95	0.994	0.988	0.983	0.978	0.974	0.900	0.994	0.083	0.043	-0.006	0.043	-0.005
	0.99	0.994	0.988	0.984	0.979	0.974	0.906	0.994	0.094	0.046	-0.010	0.040	-0.010
	1.010	0.993	0.988	0.984	0.979	0.975	0.909	0.993	0.128	0.052	-0.001	0.029	-0.012
	1.05	0.992	0.987	0.983	0.978	0.974	0.912	0.992	0.192	0.071	0.015	0.021	-0.016
	1.10	0.992	0.987	0.983	0.978	0.974	0.913	0.992	0.191	0.072	0.016	0.022	-0.021
	1.15	0.992	0.987	0.983	0.979	0.975	0.915	0.992	0.178	0.071	0.015	0.023	-0.021

The table contains (partial) autocorrelations for each considered in the paper moneyness category across two selected maturities: the shortest 20 days and the longest 320 days. (Partial) autocorrelations are reported for 1,2,3,4,5 and 22 lags. Times series underlying the statistics are obtained with equation (2.1).

Table C.3: (Partial) autocorrelations of the slope of the volatility smile and term structure

Panel A: Volatility smile

	ACF						PACF					
κ	**1**	**2**	**3**	**4**	**5**	**22**	**1**	**2**	**3**	**4**	**5**	**22**
30	0.904	0.857	0.812	0.770	0.739	0.566	0.904	0.221	0.051	0.017	0.055	-0.009
50	0.937	0.904	0.878	0.852	0.831	0.618	0.937	0.207	0.107	0.029	0.048	-0.023
65	0.938	0.906	0.879	0.854	0.834	0.613	0.938	0.213	0.094	0.031	0.057	-0.035
80	0.942	0.913	0.890	0.869	0.851	0.620	0.942	0.231	0.105	0.052	0.052	-0.033
120	0.946	0.929	0.914	0.896	0.884	0.685	0.946	0.330	0.136	0.037	0.055	0.008
160	0.949	0.937	0.926	0.912	0.900	0.704	0.949	0.373	0.175	0.044	0.034	-0.003
240	0.944	0.934	0.927	0.918	0.910	0.768	0.944	0.402	0.226	0.120	0.068	-0.005
320	0.940	0.930	0.920	0.912	0.903	0.744	0.940	0.403	0.207	0.124	0.070	-0.010

Panel B: Volatility Term Scrutcure

	ACF						PACF					
m	**1**	**2**	**3**	**4**	**5**	**22**	**1**	**2**	**3**	**4**	**5**	**22**
0.85	0.946	0.906	0.873	0.843	0.818	0.585	0.946	0.102	0.061	0.034	0.048	-0.043
0.90	0.947	0.906	0.876	0.853	0.832	0.590	0.947	0.093	0.098	0.086	0.034	-0.039
0.95	0.948	0.913	0.889	0.873	0.854	0.610	0.948	0.141	0.110	0.118	0.010	-0.014
0.99	0.952	0.924	0.904	0.887	0.871	0.639	0.952	0.186	0.117	0.083	0.029	-0.021
1.01	0.952	0.925	0.905	0.889	0.873	0.644	0.952	0.206	0.119	0.070	0.040	-0.030
1.05	0.947	0.921	0.900	0.882	0.867	0.639	0.947	0.231	0.115	0.059	0.064	-0.034
1.10	0.944	0.916	0.896	0.876	0.861	0.630	0.944	0.235	0.117	0.054	0.067	-0.022
1.12	0.912	0.882	0.860	0.839	0.821	0.579	0.912	0.297	0.159	0.078	0.049	-0.009

The table contains (partial) autocorrelations for the slope of the volatility smile and the slope of the volatility term structure. The slope of the smile corresponding to each maturity grid point $\kappa_j \in \{30, 50, 65, 80, 120, 160, 240, 320\}$ is defined as IV of the furthest put option on the grid for which $m = 0.85$, minus IV of the furthest call option on the grid for which $m = 1.12$. Similarly, the slope of the volatility term structure corresponding to each moneyness grid point $m_i \in \{0.85, 0.90, 0.95, 0.99, 1.01, 1.05, 1.10, 1.12\}$ is defined as IV of the longest maturity on the grid $\kappa = 320$, minus IV of the shortest maturity on the grid $\kappa = 30$.

Table C.4: Augmented Dickey-Fuller for the input time-series to VARX model

	$\hat{\beta}_{0t}$		$\hat{\beta}_{1t}$		$\hat{\beta}_{2t}$		$\hat{\beta}_{3t}$		$\hat{\beta}_{4t}$	
Period	t.stat	p-value	t.stat	p-value	t.stat	p-value	t.stat	p-value	t.stat	p-value
04/Jan/99-24/Dec/02	-0.61	0.43	-1.42	0.15	-3.15	<0.001	-4.37	0.002	-4.03	<0.001
09/Sep/09-30/Aug/13	-0.39	0.51	-0.86	0.34	-3.74	<0.001	-3.93	<0.001	-5.37	<0.001

The table illustrates the results of an augmented Dickey-Fuller test for a unit root in the logarithm of the input time-series to VARX model. The time-series being tested are of the length 1000 days what corresponds to the moving estimation window of VARX model. The test is performed on 10 time-series in total. I select two time periods- the first and the last estimation window.

Table C.5: Augmented Dickey-Fuller for the input time-series to PCA model

		Maturity κ					
		30 days		120 days		240 days	
	Period	t.stat	p-value	t.stat	p-value	t.stat	p-value
$\log \hat{\sigma}_t(0.99, \kappa_i)$	04/Jan/99-24/Dec/02	-0.65	0.41	-0.36	0.52	-0.26	0.56
	09/Sep/09-30/Aug/13	-0.58	0.44	-0.12	0.61	0.09	0.68
$\Delta \log \hat{\sigma}_t(0.99, \kappa_i)$	04/Jan/99-24/Dec/02	-34.71	<0.001	-31.45	<0.001	-31.86	<0.001
	09/Sep/09-30/Aug/13	-36.25	<0.001	-33.87	<0.001	-33.39	<0.001

The table illustrates the results of an augmented Dickey-Fuller test for a unit root in the logarithm of the input time-series to PCA model. The time-series being tested are of the length 1000 days what corresponds to the moving estimation window of PCA model. The test is performed on 12 time-series in total. I select two time periods- the first and the last estimation window. In each period there are six time-series: $\log \hat{\sigma}_t(m, \kappa)$ and $\Delta \log \hat{\sigma}_t(m, \kappa)$ corresponding to the short, medium and the long term ATM put options with $m = 0.99$ and $\kappa \in \{30, 120, 240\}$.

Table C.6: In-sample fit as measured by MAEV

	Maturity								
	<60			**60-180**			**>180**		
Moneyness	**PBS**	**PCA**	**HN**	**PBS**	**PCA**	**HN**	**PBS**	**PCA**	**HN**
0.85-0.90	**.0086**	.0141	.0340	.0090	**.0046**	.0178	.0080	**.0017**	.0083
0.90-0.95	**.0076**	.0100	.0315	.0069	**.0038**	.0146	.0069	**.0016**	.0106
0.95-0.99	.0082	**.0058**	.0264	.0049	**.0038**	.0156	.0058	**.0023**	.0154
0.99-1.01	.0097	**.0071**	.0228	.0049	**.0040**	.0149	.0049	**.0029**	.0194
1.01-1.05	**.0094**	.0096	.0197	.0056	**.0037**	.0097	.0047	**.0023**	.0167
1.05-1.10	.0106	**.0076**	.0177	.0069	**.0035**	.0126	.0056	**.0019**	.0100
1.10-1.15	.0186	**.0113**	.0188	.0090	**.0045**	.0159	.0065	**.0032**	.0082
All	**.0090**	.0092	.0253	.0068	**.0039**	.0143	.0062	**.0022**	.0119

The table contains average daily mean absolute error in implied volatilities (MAEV) for different models over the period 03/Mar/2003-30/Aug/2013. PBS is Practitioner Black-Scholes model, PCA is PCA model and HN is Heston and Nandi GARCH type option pricing model. MAEV is calculated for different moneyness-maturity categories on the sample restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. Emboldened values indicate the best performing model within each moneyness-maturity category.

Table C.7: In-sample pricing errors of Heston and Nandi model

	Maturity					
	<60		60-180		>180	
Moneyness	**RMSE**	**Avg. price**	**RMSE**	**Avg. price**	**Avg. RMSE**	**Avg. price**
0.85-0.90	1.69	3.40	2.39	13.07	2.67	36.02
0.90-0.95	2.33	6.66	2.74	21.47	3.89	49.71
0.95-0.99	2.90	14.10	3.55	33.77	6.02	65.99
0.99-1.01	3.14	22.07	3.98	43.04	8.31	76.98
1.01-1.05	2.16	10.65	2.44	28.55	7.74	61.54
1.05-1.10	1.00	3.22	1.83	12.53	4.65	38.61
1.10-1.15	0.60	1.78	1.37	5.04	2.80	21.42
All	2.33	8.74	2.81	21.56	5.41	47.35

The table presents in-sample pricing errors of Heston and Nandi model resulted from the NLS estimation over the period 03/Mar/2003-30/Aug/2013. Reported are average daily RMSE for different moneyness-maturity groups, expressed in $. For the sake of comparison column Avg. price reports the average of the mean daily prices in a given group. Emboldened values indicate the best performing model within each moneyness-maturity category.

Table C.8: Out-of-sample pricing errors of Heston and Nandi model

	Maturity					
	<60		60-180		>180	
Moneyness	**RMSPE**	**Avg. Price**	**RMSPE**	**Avg. Price**	**RMSPE**	**Avg. Price**
0.85-0.90	2.01	3.42	3.10	13.11	3.70	36.12
0.90-0.95	3.13	6.66	3.93	21.48	4.88	49.72
0.95-0.99	4.57	14.10	5.04	33.78	6.96	65.99
0.99-1.01	5.47	22.07	6.39	43.05	10.72	76.99
1.01-1.05	3.63	10.65	4.54	28.56	8.33	61.55
1.05-1.10	1.54	3.21	2.93	12.51	5.84	38.58
1.10-1.15	0.87	1.77	1.72	5.03	3.81	21.54
All	3.65	8.70	4.34	21.46	6.75	47.22

The table presents out-of-sample pricing errors of Heston and Nandi model resulted from the NLS estimation over the period 03/Mar/2003-30/Aug/2013. Reported are average daily RMSPE for different moneyness-maturity groups, expressed in $. For the sake of comparison column Avg. price reports the average of the mean daily prices in a given group. Emboldened values indicate the best performing model within each moneyness-maturity category. Emboldened values indicate the best performing model within each moneyness-maturity category.

Table C.9: Out-of-sample fit as measured by MAPEV

	Maturity														
	<60					60-180					>180				
Moneyness	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**	**PBS**	**VARX**	**PCA**	**HN**	**RW**
0.85-0.90	.0098	**.0097**	.0153	.0346	.0079	.0090	.0087	**.0061**	.0182	.0040	.0075	.0073	**.0030**	.0087	.0026
0.90-0.95	**.0094**	.0097	.0124	.0336	.0076	.0074	.0071	**.0053**	.0152	.0040	.0067	.0065	**.0030**	.0108	.0027
0.95-0.99	.0101	.0107	**.0083**	.0297	.0073	.0061	.0061	**.0052**	.0157	.0041	.0059	.0058	**.0033**	.0156	.0029
0.99-1.01	.0118	.0123	**.0101**	.0263	.0078	.0067	.0067	**.0057**	.0150	.0046	.0054	.0054	**.0039**	.0193	.0034
1.01-1.05	**.0120**	.0123	.0124	.0230	.0074	.0078	.0078	**.0055**	.0109	.0043	.0056	.0057	**.0035**	.0169	.0029
1.05-1.10	.0129	.0125	**.0105**	.0195	.0076	.0090	.0086	**.0053**	.0131	.0042	.0066	.0067	**.0030**	.0102	.0027
1.10-1.15	.0193	.0190	**.0135**	.0223	.0102	.0101	.0093	**.0062**	.0162	.0043	.0073	.0074	**.0043**	.0083	.0026
All	**.0111**	.0112	.0116	.0281	.0078	.0080	.0077	**.0056**	.0148	.0043	.0065	.0065	**.0034**	.0122	.0028

The table contains average daily mean absolute prediction error in implied volatilities (MAPEV) for different models over the period 03/Mar/2003-30/Aug/2013. PBS is Practitioner Black-Scholes model, PCA is PCA model and HN is Heston and Nandi GARCH type option pricing model. MAPEV is calculated for different moneyness-maturity categories for the forecasts restricted to contracts in the moneyness range of $m \in [0.85, 1.15]$. Emboldened values indicate the best performing model within each moneyness-maturity category.

Table C.10: Robustness checks of the combination forecasts- RMSPEV comparison

Panel A: Sensitivity to the track record period length for the regression method

	Maturity								
	<60			60-180			>180		
	22	**44**	**66**	**22**	**44**	**66**	**22**	**44**	**66**
0.85-0.9	.0105	.0104	.0104	.0052	.0052	.0052	.0032	.0031	.0031
0.9-0.95	.0101	.0100	.0100	.0052	.0052	.0052	.0033	.0032	.0032
0.95-0.99	.0094	.0091	.0090	.0052	.0051	.0050	.0033	.0033	.0032
0.99-1.01	.0103	.0099	.0098	.0059	.0057	.0056	.0039	.0039	.0038
1.01-1.05	.0100	.0095	.0095	.0055	.0053	.0052	.0035	.0034	.0034
1.05-1.1	.0106	.0102	.0100	.0056	.0054	.0053	.0033	.0032	.0032
1.1-1.15	.0155	.0148	.0146	.0060	.0058	.0058	.0036	.0036	.0036
All	.0111	.0107	.0106	.0059	.0058	.0057	.0037	.0037	.0037

Panel B: Sensitivity to the track record period length for the DMSPE method

	Maturity								
	<60			60-180			>180		
	22	**44**	**66**	**22**	**44**	**66**	**22**	**44**	**66**
0.85-0.9	.0112	.0112	.0110	.0064	.0066	.0064	.0037	.0037	.0037
0.9-0.95	.0106	.0105	.0104	.0060	.0060	.0059	.0038	.0039	.0038
0.95-0.99	.0096	.0096	.0096	.0059	.0058	.0057	.0043	.0043	.0043
0.99-1.01	.0118	.0118	.0118	.0064	.0064	.0063	.0046	.0046	.0046
1.01-1.05	.0119	.0119	.0119	.0062	.0062	.0061	.0043	.0044	.0043
1.05-1.1	.0112	.0114	.0112	.0064	.0064	.0063	.0038	.0039	.0039
1.1-1.15	.0152	.0156	.0157	.0072	.0070	.0070	.0048	.0049	.0049
All	.0121	.0121	.0120	.0069	.0069	.0068	.0046	.0046	.0046

Panel C: Sensitivity to the discount factor value for DMSPE method

	Maturity								
	<60			60-180			>180		
	0.9	**0.95**	**1**	**0.9**	**0.95**	**1**	**0.9**	**0.95**	**1**
0.85-0.9	.0112	.0112	.0111	.0066	.0066	.0065	.0038	.0037	.0037
0.9-0.95	.0106	.0105	.0105	.0061	.0060	.0060	.0039	.0039	.0039
0.95-0.99	.0097	.0096	.0096	.0058	.0058	.0058	.0043	.0043	.0043
0.99-1.01	.0118	.0118	.0118	.0064	.0064	.0063	.0046	.0046	.0046
1.01-1.05	.0120	.0119	.0119	.0062	.0062	.0062	.0044	.0044	.0044
1.05-1.1	.0114	.0114	.0114	.0064	.0064	.0064	.0039	.0039	.0039
1.1-1.15	.0155	.0156	.0156	.0071	.0070	.0070	.0049	.0049	.0049
All	.0122	.0121	.0121	.0069	.0069	.0069	.0046	.0046	.0046

The table reports RMSPEV sensitivity for different setups of the combination methods. Panel A and Panel B report RMSPEV for different choices of the track record period for the regression and DMSPE based weighting respectively. Panel C reports RMSPEV sensitivity to a value of the discount factor in DMSPE method, when the base track record period of 44 days is used. Value of 1 indicate no discounting. The evaluation period for all the three panels is 03/Mar/2003-30/Aug/2013.