

ERASMUS SCHOOL OF ECONOMICS

BACHELOR'S THESIS IN ECONOMETRICS AND OPERATIONAL RESEARCH

How to model the implied volatility surface accurately and obtain efficient predictions? Evidence from index options.

HAM, R., VAN DER, 434028 supervisor: Gong, X. second assessor: XIAO, X.

July 8, 2018

Abstract

In this paper we show applicability of the 7-factor implied volatility surface specification, proposed by Chalamandaris and Tsekrekos (2011), on index options. Moreover, we claim that any previously proposed parametric specification contains heteroskedasticity and autocorrelation in the fitting residuals, which results in inefficient volatility estimations. We propose a Residual Correction Model that enables us to reduce the heteroskedasticity and autocorrelation and is superior in both in- and out-of-sample performances over alternative implied volatility surface specifications. We find statistical and economical evidence of the relevance of both the 7-factor specification and Residual Correction Model. Moreover, we show that a combination of the two captures valuable information regarding describing and predicting index implied volatility surfaces, as it produces accurate volatility estimates and efficient surface predictions, and enables us to support profitable trading strategies in the absence of transaction costs. Furthermore, we show that making use of error-correction models and restricted dynamic factor State Space Models help us to obtain more accurate option volatility predictions than alternative time series models.

Contents

1	Introduction													
2	Data													
3	Deterministic IVS modeling 3.1 Five and Seven Factor IVS Specifications 3.2 Implied Volatility Surface Residual Correction Model 3.3 In-sample performances	7 7 10 11												
4	Modeling the time-variation of the IVS 4.1 Time-series models 4.2 Dynamic factor model 4.3 Out-of-Sample Performances	12 12 13 15												
5	Economic Evaluation	19												
6	Conclusion	21												
A	Robustness Check for Lambda	23												
в	Out-of-Sample Performances State Space Model	24												

1 Introduction

The famous Black-Scholes formula that is used to calculate option prices, proposed in Black and Scholes (1973), assumes constant volatility. However, the implicit volatility in option contracts seem to differ for different strike prices and time-to-maturities. These two features that characterize the implied volatility of option contracts attract the interest of researchers and practitioners. The option-volatilities, written on the same underlying asset, with different strike prices and dates to expiration form a shape that is non-flat, also known as the implied volatility surface (IVS). Moreover, the IVS seems to change dynamically over time, because option prices respond to new information that affects investors' expectations. This feature of the IVS is important, as it could prove predictability of the future IVS state.

The derivatives market is expanding to a large extend, but there is still limited information and knowledge about the dynamics of option contracts. Albeit the existence of satisfactory parametric frameworks for describing the IVS and significant prove of predictability in the prospective IVS movements, there are still some deficiencies in estimating and predicting implied volatility surfaces. Existing IVS specifications have difficulties in fitting option contracts that are deep-in -or deep-out-of-the-money. In general, this results in poor in-sample fits and out-of-sample predictions, as large IVS estimation (or prediction) errors occur for option contracts with extreme moneyness values (positive or negative). Moreover, previous studies generally use a vector autoregressive process to prove the existence of predictability of the future IVS states, but there is little known about possible improvements in forecasting the future IVS by addressing different types of time series models.

In this paper we will combine and extend existing literature to model and forecast the implied volatility for both call and put options. We will follow - and use the deterministic 7-factor model proposed in - the paper by Chalamandaris and Tsekrekos (2011). Although they make use of foreign exchange options to proof efficiency of their proposed 7-factor model, we will examine its applicability on index options. Moreover, we introduce an IVS Residual Correction Model for the fitting errors accompanied with the 7-factor model, which reduces heteroskedasticity and autocorrelation of the residuals across different moneyness and maturity levels. This model enables us to capture the implied volatility more sufficient for option contracts that are deep-in -or deep-out-of-the-money - and with very short or long expiration dates - than alternative IVS specifications. Furthermore, we shed light on different types of time series models to come up with the best performing model regarding predictive abilities of future IVS states. We also introduce a restricted dynamic model to obtain IVS forecasts, which is a quite new area and briefly examined by van der Wel et al. (2016). We will use the restricted dynamic model, applied on the 7-factor model proposed by Chalamandaris and Tsekrekos (2011), and is in our knowledge not done before. In order to evaluate the economic relevance of our proposed IVS specifications and time series models, we make use of simulating a simple trading strategy.

We make use of daily option data on three different indices; S&P 500 index, Dow Jones Industrial Average, and the CBOE Mini-NDX Index for the period January 2, 2003 up to December 29, 2017. We find, in accordance with Chalamandaris and Tsekrekos (2011) for foreign exchange options, that the proposed 7-factor model achieves improvements in accurately reproduce the IVS shapes, that are observed in practice, over alternative parametric IVS specifications. Moreover, our proposed IVS Residual Correction Model enables us to substantially reduce the heteroskedasticity and autocorrelation in the errors and represents a more efficient description of volatility surfaces. We show that combining the 7-factor IVS specification with the IVS Residual Correction Model can consistently and considerably improve the ability of describing the implied volatility surface of index options.

If the combination of the two models mentioned above contain valuable information, then they should not only improve in-sample fits, but also produce out-of-sample implied volatility predictions that are more accurate than alternative IVS specifications. We find that making use of the 7-factor model and the IVS Residual Correction Model consistently improve producing out-of-sample IVS forecasts, in means of root mean squared error and mean absolute error, over other hard-to-beat benchmarks. In addition, we will show that error correction models, applied on the 7-factor model, out-perform any other type of time series models for short term horizons. For longer forecasting horizons, the use of a univariate autoregressive process performs approximately equivalent to - or slightly better than - the error-correction model. Moreover, we find that the use of a restricted dynamic factor model contains valuable information regarding predicting the implied volatility of option contracts and could consistently produce more accurate IVS forecasts relatively to any other time series model.

In order to assess the economic relevance of the obtained forecasting improvements, we show that abnormal returns can be achieved, in the absence of transaction costs, and could be increased by making use of the 7-factor model and the IVS Residual Correction Model. However, earnings turn negative by including transaction costs, which is in line with the efficiency of the market.

Previous research mainly focused on modeling the IVS by fitting parametric specifications in a cross sectional regression on options written at a point in time, connecting the implied volatility to (transformations of) maturity and moneyness of these options. A study by Dumas et al. (1998) proved that such parametric frameworks are a realistic and efficient description of the true IVS. However, similar studies to this paper generally neglect option contracts with long expiration dates. These kind of options are available - and actively traded - in over-the-counter (OTC) markets. The existing paper of Goncalves and Guidolin (2006) is closely related to our study and only researches option contracts with dates to expiration less than a year. Furthermore, they propose a widely used 5-factor parametric IVS specification that does not allow for any asymmetric characteristics across different option contracts. Nevertheless, a paper by Peña et al. (1999) show significant asymmetries in the shape of the implied volatility smile.

A study by Chalamandaris and Tsekrekos (2011) find improvements in fitting the IVS more flexible for term structures through a Nelson and Siegel (1987) factorization (proposed by Diebold and Li (2006) for yield curve modeling) and by allowing asymmetries in the IVS shape. Making use of the 7-factor IVS specification achieves the fitting errors to be less heteroskedastic across the surface compared to alternative specifications, but more efficient estimates could be obtained by making use of our proposed IVS Residual Correction Model, which diminish the fitting errors' heteroskedasticity substantially.

Furthermore, studies by Cont and Fonseca (2002), Goncalves and Guidolin (2006), and Fengler et al. (2007) have examined the aspect of modeling and forecasting the IVS dynamically and found significant evidence of predictability in the future IVS movements. Previous literature generally use (simple) vector autoregressive processes to prove the existence of IVS predictability, but do not necessarily seek for the best performing time series models regarding predictive abilities. However, accurate predictions of the future IVS is of big importance in finance, because in many economic fields are dealing with the dynamics in option prices. Accurate forecasts are essential for portfolio allocation, risk management, and speculative purposes. Trading strategies - that make use of option contracts - highly depend on accurate implied volatility forecasts. Hence, more efficient forecasts could lead to higher risk-adjusted returns. Therefore, time series models that consistently improve IVS forecasts could contain valuable information regarding economic affairs.

We contribute to the existing literature by showing applicability of the 7-factor IVS specification for index options and by showing the valuable information it contains. Furthermore, we are proposing a model that enables market practitioners and researchers to describe implied volatility surfaces more accurate and enables them to obtain more efficient predictions for the future IVS states than simpler IVS specifications do, due to reducing the error's heteroskedasticity and autocorrelations across different moneyness and maturity levels. In addition, we show which time series model is best in forecasting the dynamics of option contracts and how optimal IVS predictions could be obtained.

The rest of our paper is organized as follows: We will start with giving a more extensive explanation about the data we will use in this research in Section 2. Section 3 introduces, estimates, and analyzes the IVS deterministic model and the IVS Residual Correction Model. Subsequently, we highlight different time series models and the dynamic factor model to evaluate the forecasting performances in Section 4. Section 5 assesses the economic value of the predictions made in Section 4 with a simple trading strategy simulation. Lastly, Section 6 concludes.

2 Data

To find evidence for our research, we use data from the OptionMetrics database, which provides a large collection of US equity and index options. We use three different indices; S&P 500 index (SPX), Dow Jones Industrial Average (DJX), and CBOE Mini-NDX Index (MNX). We have daily data over the period January 2, 2003 up to December 29, 2017. The sample period covers recent data and contains a fair amount of observations over a large number of years. In this way, we aim at reliable findings that are relevant for today's financial market. For each day in the covering sample period we extract data for call and put options. For each individual option contract we obtain the following characteristics; security ID, date, call or put flag, implied volatility, option delta, expiration date, strike price, closing bid and asking price, and the prices of the underlying index levels.

To filter out observations that contain only a little - or unreliable - information regarding the IVS, we remove option contracts with less than six days to expiration. Option prices are calculated as the midpoints between bid and ask prices, following the literature of Bernales and Guidolin (2014). We remove option contracts with prices below \$0.375 to avoid effects of price discreteness on the shape of the IVS. To measure the moneyness variable, we use the option delta, because the likelihood that an option expires in-the-money does not only depend on the expiration date of the contract, but also on the underlying volatility rate. The option delta is sensitive to both the option's expiration date as well as to the volatility of the underlying asset or index. Moreover, Chalamandaris and Tsekrekos (2011) explain in their paper that the option delta has an additional number of positive features; (i) it has less noisy volatility observations, (ii) it saves the IVS shape, and (iii) it reduces the flattening effect at longer expiration days. We omit variables with an option delta, in absolute value, larger than 0.98 or smaller than 0.02, because these observations are rather noisy, see also Bollen and Whaley (2004). Option contracts with an implied volatility higher than 100% are removed, to extract data errors from our dataset. We also drop option contracts where we miss observations for either the implied volatility or the option delta. For estimation and comparison purposes, we scale the option deltas and days to maturity to the interval [-0.5, 0.5], such that the regression variables have the same order of importance.¹ Note that studying out-of-the-money call (put) option contracts is identical to considering in-the-money put (call) option contracts due to put-call parity. An in-the-money call (put) option can be matched to an out-of-the-money put (call) option, by the fact that the delta of a call option is always 1 plus the delta of a put option (only applies for options with the same characteristics; same expiration date, strike price, and underlying asset).

For each index, we create the daily implied volatility surfaces by grouping the data in moneyness and maturity dimensions. In the moneyness dimension we divide the data in five groups, separated by deep-out-of-the-money (DOTM), out-the-money (OTM), at-the-money (ATM), in-the-money (ITM), and deep-in-the-money (DITM). We denote Δ_c and Δ_p as option delta's for call and put options, respectively. We consider call (put) options with $0.02 < \Delta_c \leq 0.125$ ($-0.125 < \Delta_p \leq -0.02$) as DOTM, with $0.125 < \Delta_c \leq 0.375$ ($-0.375 < \Delta_c \leq -0.125$) as OTM, with $0.375 < \Delta_c \leq 0.625$ ($-0.625 < \Delta_c \leq -0.375$) as ATM, with $0.625 < \Delta_c \leq 0.875$ ($-0.875 < \Delta_c \leq -0.625$) as ITM, and with $0.875 < \Delta_c \leq 0.98$ ($-0.98 < \Delta_c \leq -0.875$) as DITM. In the maturity dimension, we also consider five different categories. Very short-term contracts have expiration dates up to 60 days, short-term contracts have 60-120 days to expiration, medium-term contracts have 120-240 days to expiration, long-term contracts have 240-365 days to expiration, and very long-term contracts have more than 365 days to expiration.

For each of the three indices, Table 1 reports (i) the average implied volatility, (ii) the number of observations in percentage of the total sample, and (iii) the longest maturity. Moreover, the table shows the descriptive statistics (i) and (ii) for all different moneyness-maturity groups and distinguishes call and put options over the different categories. The average implied volatility of call options increases when it gets deeper in-the-money. The contrary is observable for put options, except that the volatility increases for

¹We add 1 to the deltas of put option contracts. Then we subtract 0.5 from all option deltas, which results in a delta interval [-0.5, 0.5]. This means that call (put) option contracts with a delta close to -0.5 are deep-out-of-the-money (deep-in-the-money) and call (put) options with a delta close to 0.5 are deep-in-the-money (deep-out-of-the-money). This also indicates that at-the-money option contracts have a delta close to 0. To construct the same interval for maturities, we divide the expiration date by the longest maturity for the corresponding index and subtract 0.5.

Table 1: Descriptive Statistics for the Index Options by different Moneyness-Maturity groups January 2, 2003 - December 29, 2017

	Very Short-Term		Short-Term		Mediu	Medium-Term		Long-Term		ong-Term		
	Call	Put	Call	Put	Call	Put	Call	Put	Call	Put	Observations (%)	longest maturity
Panel A: Average Vol	atilities											
DOTM:												
S&P 500	0.117	0.235	0.123	0.269	0.137	0.280	0.144	0.287	0.147	0.291	22.59	1095
DJX	0.298	0.291	0.185	0.279	0.171	0.292	0.169	0.293	0.174	0.278	11.60	1095
MNX	0.214	0.343	0.189	0.332	0.184	0.328	0.185	0.328	0.191	0.322	15.42	1123
OTM:												
S&P 500	0.129	0.180	0.142	0.205	0.154	0.219	0.160	0.225	0.164	0.229	18.28	1095
DJX	0.166	0.213	0.159	0.220	0.171	0.238	0.176	0.241	0.182	0.239	21.94	1095
MNX	0.210	0.263	0.204	0.269	0.198	0.267	0.201	0.269	0.205	0.265	22.31	1187
ATM:												
S&P 500	0.150	0.151	0.168	0.169	0.181	0.181	0.186	0.186	0.193	0.191	13.67	1095
DJX	0.178	0.179	0.182	0.182	0.197	0.198	0.203	0.201	0.205	0.201	17.44	1095
MNX	0.228	0.229	0.231	0.232	0.226	0.227	0.228	0.228	0.230	0.229	18.05	1187
ITM:												
S&P 500	0.181	0.133	0.204	0.144	0.220	0.157	0.228	0.162	0.236	0.165	18.88	1095
DJX	0.210	0.165	0.221	0.162	0.239	0.175	0.247	0.180	0.249	0.190	23.80	1095
MNX	0.264	0.216	0.270	0.207	0.267	0.200	0.270	0.202	0.268	0.206	22.77	1187
DITM:												
S&P 500	0.280	0.186	0.280	0.162	0.301	0.185	0.322	0.171	0.329	0.165	26.58	1095
DJX	0.338	0.301	0.312	0.239	0.325	0.234	0.319	0.208	0.299	0.197	25.22	1094
MNX	0.391	0.338	0.352	0.264	0.340	0.222	0.343	0.201	0.329	0.188	21.45	1123
Panel B: Observations	s by Matu	rity(%)										
S&P 500	40).18	22	.44	11	.25	10.66		15.48			
DJX	17	7.27	16	.88	18	.28	19	.21	28	3.37		
MNX	10.50		11.	.51	17	.18	18	.39	42	2.42		

Notes to Table: Panel A reports the average implied volatilities for each of the three different index surfaces and categorized by 25 different moneyness-maturity groups. The results are obtained after applying the exclusionary criteria. We denote Δ_c and Δ_p as option delta's for call and put options, respectively. For the call options, DOTM denotes deep-out-of-the-money ($0.25 < \Delta_c \leq 0.375$); ATM, at-the-money ($0.375 < \Delta_c \leq 0.625$); ITM, in-the-money ($0.625 < \Delta_c \leq 0.875$); and DITM, deep-in-the-money ($0.875 < \Delta_c \leq 0.98$). For put options hold the same, but with negative signs for the Δ_p boundaries, so for example, DOTM is specified as $-0.125 < \Delta_p \leq 0.02.$ Very short-term contracts have option maturities up to 60 days, short-term contracts have 60-120 days to maturity, medium-term contracts have 120-240 days to maturity, long-term contracts have 240-365 days to maturity, and very long-term contracts have more than 365 days to expiration. Panel A also reports the percentage of total observations for each moneyness category.

DITM put options. These features causes a sort of 'U-shape' over the moneyness dimension and is also known as the volatility smile. The accumulated average implied volatilities of DOTM puts and DITM calls, across all maturity groups, is substantially higher than the accumulated average implied volatilities of DITM puts and DOTM calls, which is an indication of possible asymmetries in the volatility smile. The table also shows an indication of an increasing term-structure, as the implied volatility tends to increase for option contracts with longer expiration dates.

Table 1 shows that the observations are quite evenly distributed over the different moneyness categories for all three indices and that there is at least 10% of the total number of option contracts in each maturity category (Panel B), such that each moneyness-maturity group has an adequate number of observations. For each index, the longest date to expiration is about three years and Panel B reports that 15.48% of the total observations for S&P 500 up to 42.42% for the MNX is above one year to maturity. Most previous research omitted these option contracts, but this amount of observations could contain valuable information and indicates possible practicality of including term structures in the IVS specification.

The combination of the different moneyness and maturity categories results in 25 different groups, all containing a fair amount of daily option contracts. For each group, we select the option contract closest to the midpoint of this category.² We use interpolated data, such that options - across the same groups - have similar option characteristics. Grouping the data of a large cross-sectional data set is also done and explained in more detail by Bollen and Whaley (2004).

The cross-sectional and time-series characteristics of the total dataset over the different moneynessmaturity groups can be better understood with the aid of Figure 1. For illustrative purposes, we have plotted the average IVS, the standard deviation of the IVS, and the time evolution of the IVS over the different moneyness and maturity categories for the S&P 500 index options. Panel (a) shows that the average implied

 $^{^{2}}$ Closest to the midpoint is defined as the minimal sum of the squared distance for maturity and delta. Note that this is done after scaling the option delta and maturity to [-0.5, 0.5], such that both variables have the same degree of importance.

volatility smile is far from symmetric and contains a small increasing term structure, at least for moneyness group five. Furthermore, Panel (b) shows that there is a lot of variability in the average IVS structure, especially for large absolute delta values. Panel (c) and (d) both indicate that the IVS is not constant over time, but tends to dynamically change over time. However the average IVS does not seem to be symmetric, Panel (c) shows that its level and shape can vary over time and might be more symmetric for some days than Panel (a) suggests. Although Table 1 suggests that the average term structure is increasing, Panel (d) indicates that the term structure changes from 'up-sloping' to 'down-sloping', to be 's-'or 'u-shaped' over time.





(c) Average Maturity IVS over Time, S&P 500

(b) Standard Deviation IVS S&P 500



(d) Average Moneyness IVS over Time, S&P 500



Figure 1: The figure is divided in four different panels: Panel (a): Average implied volatility surface over the 25 moneyness-maturity groups. Panel (b): Daily standard deviation of the implied volatility over the different groups. Panel (c): Daily implied volatility surface dynamics over the five moneyness categories. We took the average of all maturity values over the different moneyness categories. Panel (d): Daily implied volatility surface dynamics over the five maturity categories. We took the average of all moneyness values over the different maturity categories. All plots are obtained from the S&P 500 index options and cover the full sample period January 2, 2003 - December 29, 2017

3 Deterministic IVS modeling

This section consists of three parts. First we introduce and estimate a 5- and 7-factor model to describe the implied volatility surface. Secondly, we introduce and clarify our proposed IVS Residual Correction Model. Lastly, we evaluate the in-sample fit of all the mentioned models.

3.1 Five and Seven Factor IVS Specifications

Let M be the number of moneyness groups, T the number of maturity groups, and N be the number of option contracts written on day t for one of the three index implied volatility surfaces.³ Then we have for this specific day and index surface a set of (i) deltas $\{\Delta_i\}_{i=1}^M$, as the measure for moneyness, (ii) time to maturities $\{\tau_j\}_{j=1}^T$, and (iii) implied volatilities $\{\sigma_{i,j}\}_{i,j=1}^{M,T}$, which are N group combinations. A common method to describe the IVS is to fit a parametric function with (transformations of) Δ_i and τ_j as exogenous variables in a cross-section regression for every day t and underlying asset or index k. Goncalves and Guidolin (2006) found that the following model describes the daily (on day t) IVS of S&P 500 index options the best:

$$\sigma_{i,j,t} = \beta_{0,t} + \beta_{1,t} \Delta_{i,t} + \beta_{2,t} \tau_{j,t} + \beta_{3,t} \Delta_{i,t}^2 + \beta_{4,t} \Delta_{i,t} \tau_{j,t} + \epsilon_{i,j,t}, \tag{1}$$

where they use the logarithmic transformation on sigma, that is $ln\sigma_{i,j,t}$ as dependent variable, for i = 1, ..., M, j = 1, ..., T, and where $\epsilon_{i,j,t}$ is a random error term. This model is widely used after its creation. This IVS specification is able to fit the volatility smile modestly through its quadratic term of moneyness. However, this model is mostly used by studies where they consider maturities less than one year and it is questionable whether this model performs well for option contracts with longer days to expiration. We will use this parametric IVS specification as a benchmark model with $\sigma_{i,j}$ as dependent variable instead of the logarithmic transformation of the implied volatility as in Goncalves and Guidolin (2006).

We suggests, following Chalamandaris and Tsekrekos (2011), that the term structure contains useful information regarding the IVS. Moreover, Section 2 shows possible asymmetries in the volatility smile for each of the index surfaces. In order to capture those two features, we propose the following cross-sectional model to describe the IVS of index options on any day t and underlying index k:

$$\sigma_{i,j,t} = \beta_{0,t} + \beta_{1,t}I_{1,i,t} + \beta_{2,t}I_{2,i,t} + \beta_{3,t}I_{3,j,t} + \beta_{4,t}I_{4,j,t} + \beta_{5,t}I_{5,i,j,t} + \beta_{6,t}I_{6,i,j,t} + \epsilon_{i,j,t}, \tag{2}$$

where

$$I_{1,i,t} = \mathbf{1}_{\{\Delta_{i,t}>0\}}\Delta_{i,t}^{2}, \qquad Right \quad smile$$

$$I_{2,i,t} = \mathbf{1}_{\{\Delta_{i,t}<0\}}\Delta_{i,t}^{2}, \qquad Left \quad smile$$

$$I_{3,j,t} = \frac{1 - e^{-\lambda\tau_{j,t}}}{\lambda\tau_{j,t}}, \qquad Short - term \quad structure$$

$$I_{4,j,t} = \frac{1 - e^{-\lambda\tau_{j,t}}}{\lambda\tau_{j,t}} - e^{-\lambda\tau_{j,t}}, \qquad Medium - term \quad structure$$

$$I_{5,i,j,t} = \mathbf{1}_{\{\Delta_{i,t}>0\}}\Delta_{i,t}\tau_{j,t}, \qquad Right \quad smile \quad interaction$$

$$I_{6,i,j,t} = \mathbf{1}_{\{\Delta_{i,t}<0\}}\Delta_{i,t}\tau_{j,t}, \qquad Left \quad smile \quad interaction,$$

$$(3)$$

for i = 1, ..., M, j = 1, ..., T and where $\mathbf{1}_{\{x\}}$ is an indicator function that takes the value 1 if condition x is true and 0 otherwise. Note that $\Delta_{i,t}$ and $\tau_{j,t}$ are the option delta and time to expiration after the transformations noted in Section 2, respectively.

If constant volatility is assumed, as in Black and Scholes (1973), then the implied volatility for every option contract is equal to β_0 , which captures the level of the IVS on any day t. The parameters $\beta_{1,t}$ and

³Note that N is equal to $M \times T$, which is 25, except for days where no option contracts were traded. However, we restrict our sample to points in time where at least 10 option contracts were traded on that day to have enough variation in the regression variables.

 $\beta_{2,t}$ describe the curvature in the moneyness dimension and (partly) captures the volatility smile. Dividing the implied volatility smile in a right and left part, enables us to capture any asymmetries in DITM calls (DOTM puts) and DOTM calls (DITM puts) with the parameters $\beta_{1,t}$ and $\beta_{2,t}$, respectively.

The variables $I_{3,j,t}$ and $I_{4,j,t}$ are functions of time-to-maturity and correspond with the IVS termstructures, proposed by Nelson and Siegel (1987) and adopted by Diebold and Li (2006), who found prove of successfully describing and forecasting the yield curve. The parameter λ is the time constant and represents the term-structure's decay rate. Large values of λ corresponds to a rapid decay in the regressors and is able to fit the curvature of low maturities very well, but rather poor for longer maturities. A small value for λ corresponds to a slow decay in the regressors and thus follows in an adequate fit for the curvature of option contracts with long maturities, but is unable to fit option contracts with short maturities. Hence, $\beta_{3,t}$ and $\beta_{4,t}$ captures the short-term and medium-term structure of the IVS on a given day t, respectively.

The parameters $\beta_{5,t}$ and $\beta_{6,t}$ capture potential interaction between moneyness and maturity. The main reason of several other IVS studies to drop option contracts with maturities larger than a year is due to the fact that those contain little information regarding the IVS, as the volatility smile is getting flatter for larger maturities. It is not unreasonable to think that the flatting effect of the IVS is also asymmetric, due to the uncertainty of the IVS direction, and thus $\beta_{5,t}$ and $\beta_{6,t}$ captures the volatility smile flattening effect for the right and left smile, respectively.

Model 2 is a non-linear cross-sectional specification due to the parameter λ in the term structures. However, non-linear estimation causes optimization problems, such as inaccurate end solutions - due to misguided initial starting positions - and non-linear optimization methods are mathematical and time intensive. Usually, researcher that encounter these problems fix the λ coefficient to a specific constant.⁴ However, fixing the lambda on a constant - examined in other studies - gives no assurance of applicability in our research, as we use different (transformations of) data and we do not use restrictions on option contracts with large maturities. Contrary to the use of a predetermined value for lambda, Chalamandaris and Tsekrekos (2011) uses a two-stage approach. They use non-linear regression to all parameters in the first step and re-estimate the β parameters using ordinary least squares (OLS) in the second step, using λ that is fixed to its median for each of the three surfaces that is estimated in the first step. They argue that the location of the IVS structure is much likely to be constant. The Appendix A reports a robustness check for λ . We conclude that we do not take the median for λ in the two-stage approach, but choose for each index surface the λ that minimizes the average RMSE over all daily observations. In this way, we aim at the most accurate, at least on average, in-sample fit and avoid numerical complications.

Table 2 reports summary statistics for the seven beta coefficients of deterministic IVS Model 2 and for the five beta coefficients of deterministic IVS Model 1. The table reports, for the three indices separately, and for each of the beta estimates, (i) the average coefficient estimates across all days in the total sample period, (ii) the average t-test statistic, to test for individual coefficient significance, (iii) the Augmented Dickey Fuller (ADF) test statistic, to test the presence of a unit root in the coefficient series, and (iv) the Ljung-Box (LB) test statistic, to test for significant autocorrelation in the coefficient series. For the ADF test, we use the average across all rolling window samples, with a length of 126 days, as we will use these segments for constructing predictions, which will be further explained in Section 4. The percentage of rejections of the null-hypotheses, using a significance level of 5% for the t-test and 10% for the ADF and LB tests, is given in parentheses.

The three surfaces have, at least on average, similar volatility levels. The positive β_1 and negative β_2 coefficients indicate asymmetry in the volatility smile and shows that the volatility is upward sloping for option contracts with a delta larger than zero and downward sloping for option contracts with a delta smaller than zero. The coefficients β_3 and β_4 indicate that, on average, the term structures are close to zero (see also Figure 1a). However, as Figure 1d indicates, the term structure coefficients might be (large) positive or negative for different trading days. The opposite signs for β_5 and β_6 again signify importance of taking into account asymmetries in IVS modeling.

⁴The paper of Diebold and Li (2006) fix their lambda to 0.0609, which is considered to be the midpoint of the structure.

			IVS specific	cation in eq	IV	IVS specification in equation 1 from						
									Goncalves	and Guide	olin (2006)	
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_0	β_1	β_2	β_3	β_4
S&P 500: Average	0.19	0.62	-0.25	0.00	-0.00	0.18	-0.18	0.19	0.17	0.04	0.19	0.07
t-test	27.12 (100.00)	$11.93 \\ (98.44)$	-4.62 (88.69)	1.19 (32.63)	-0.97 (26.85)	$1.43 \\ (46.05)$	-1.02 (24.18)	33.37 (100.00)	12.11 (99.36)	3.58 (81.33)	4.92 (96.93)	1.63 (51.93)
ADF-test	-2.11 (32.79)	-3.45 (53.70)	-3.58 (63.10)	-4.17 (72.96)	-4.29 (75.89)	-4.25 (68.82)	-5.11 (75.59)	-2.10 (31.12)	-2.96 (46.44)	-2.77 (52.71)	-6.03 (81.89)	-4.06 (63.15)
LB(1)-test	3617.49 (100.00)	2553.74 (100.00)	1775.32 (100.00)	1374.27 (100.00)	1069.56 (100.00)	1384.11 (100.00)	371.41 (100.00)	3675.67 (100.00)	3032.57 (100.00)	3298.56 (100.00)	1125.28 (100.00)	1392.80 (100.00)
LB(3)-test	10774.09 (100.00)	7408.50 (100.00)	4875.63 (100.00)	4083.80 (100.00)	3171.07 (100.00)	3774.49 (100.00)	1027.03 (100.00)	10957.65 (100.00)	8835.33 (100.00)	9478.65 (100.00)	3200.36 (100.00)	3722.40 (100.00)
Dow Jones: Average	0.18	0.60	-0.31	0.00	-0.00	0.14	-0.40	0.18	0.17	0.04	0.19	0.08
t-test	21.11 (99.97)	8.42 (97.22)	-3.71 (84.08)	1.17 (28.36)	-0.99 (24.28)	0.65 (35.28)	$^{-1.31}_{(31.12)}$	26.70 (100.00)	9.08 (98.89)	2.70 (75.08)	3.94 (94.84)	1.29 (49.36)
ADF-test	$^{-2.41}_{(43.53)}$	-3.47 (56.63)	-3.40 (59.73)	-5.55 (89.75)	-5.64 (90.93)	-4.00 (66.93)	-3.98 (68.93)	$^{-2.23}_{(39.92)}$	-2.40 (34.74)	$^{-2.64}_{(49.01)}$	-5.57 (91.37)	-3.46 (63.26)
LB(1)-test	3555.97 (100.00)	2781.58 (100.00)	1375.60 (100.00)	1019.14 (100.00)	850.66 (100.00)	1751.70 (100.00)	559.60 (100.00)	3638.05 (100.00)	3098.64 (100.00)	3346.38 (100.00)	1526.64 (100.00)	1350.60 (100.00)
LB(3)-test	10556.31 (100.00)	7888.28 (100.00)	3950.48 (100.00)	2743.55 (100.00)	2247.74 (100.00)	4773.11 (100.00)	1452.51 (100.00)	10818.52 (100.00)	9015.14 (100.00)	9615.42 (100.00)	3982.40 (100.00)	3696.32 (100.00)
CBOE Mini-NDX: Average	0.22	0.55	-0.36	0.00	-0.00	0.03	-0.50	0.22	0.16	0.03	0.18	0.06
t-test	29.02 (100.00)	9.23 (96.24)	-4.26 (88.93)	1.12 (40.73)	-0.93 (37.34)	0.50 (30.03)	-1.55 (38.69)	35.66 (100.00)	9.28 (98.09)	2.44 (73.78)	3.92 (94.33)	1.06 (39.59)
ADF-test	-2.41 (37.37)	-3.46 (50.63)	-5.12 (69.56)	-5.54 (86.11)	-5.63 (87.42)	$^{-5,42}_{(81.70)}$	-5.94 (82.71)	-2.24 (30.27)	-3.06 (45.67)	-2.88 (55.81)	-6.06 (76.44)	-5.66 (83.32)
LB(1)-test	3528.87 (100.00)	2437.68 (100.00)	1462.89 (100.00)	892.00 (100.00)	725.95 (100.00)	654.69 (100.00)	451.30 (100.00)	3573.07 (100.00)	2812.12 (100.00)	3138.93 (100.00)	590.53 (100.00)	352.16 (100.00)
LB(3)-test	10475.41 (100.00)	7314.07 (100.00)	3371.38 (100.00)	2610.93 (100.00)	2115.85 (100.00)	1793.55 (100.00)	1079.32 (100.00)	10630.19 (100.00)	8413.61 (100.00)	8979.22 (100.00)	1843.77 (100.00)	1044.31 (100.00)

Table 2: Summary Statistics Deterministic IVS Coefficients for Each Index January 2, 2003 - December 29, 2017

Notes to Table: β_l , for l = 0,...,6 for IVS specification 2 and l = 0,...,4 for IVS specification 1 correspond to the 7-and 5-factor deterministic IVS coefficients, respectively. The table reports some summary statistics and statistical tests, where the Augmented Dickey Fuller test, including an intercept in the test equation, is denoted as ADF-test and the Ljung-Box test - with one or three lags included in the test equation - as LB(1)-test and LB(3)-test, respectively. For all test statistics the value in parentheses reports the percentage of significant values at a 5% level for the t-test statistics and 10% for the ADF-and LB-test statistics.

The significance of the estimated beta coefficients suggests that the level and the two asymmetric smile variables are very important in describing the implied volatility surface for all three indices. These three factors are significantly important in at least 84.08 percent of all sample days.

The short-and medium-term Nelson-Siegel structures show significant importance for at least 28.36 and 24.28 percent, respectively. Hence, these indicators are quite important in describing the IVS on daily basis. Moreover, the last two indicators I_5 and I_6 also seem to have a reasonable percentage of significant explanatory power in explaining the IVS for the three different indices. Hence, this suggests that allowing interaction asymmetries - between moneyness and maturity - adds value in describing the IVS.

The ADF test statistics imply that we can not reject the absence of a unit-root for a large percentage of total rolling window samples across all beta estimates. This implies non-stationary processes and must be taken into account during time-variation modeling, elaborated in Section 4. The Ljung-Box test statistics, both with one and three lags, imply significant autocorrelation for all the beta coefficients, suggesting predictability of the three implied volatility surfaces.

3.2 Implied Volatility Surface Residual Correction Model

A common difficulty in modeling the IVS is to accurately fit the volatility of option contracts with extreme moneyness or maturity values. Figure 2 shows the RMSE against the 25 different moneyness-maturity groups for the two deterministic models 2 and 1 in Panel (a) and Panel (b), respectively. The figures show that the residuals are heteroskedactic among the different moneyness-maturity groups. Especially the volatility of option contracts with short or long days to expiration seem to be more challenging to estimate. The deterministic IVS specification from Model 2 seem to perform better, in means of a lower and more constant RMSE over all moneyness-maturity groups, than Model 1.

(a) Average RMSE from Model 2

(b) Average RMSE from Model 1



Figure 2: Panel (a) shows the average RMSE after applying IVS specification 2 and Panel (b) shows the average RMSE after applying IVS specification 1 over the period July 7, 2003 - December 29, 2017.

In addition to heteroskedastic error terms, the residuals also show abundant autocorrelations over time. Table 3 reports the Ljung-Box statistics (one lag included), applied on the grouped residuals $\epsilon_{i,j,t}$ for i, j = 1, ..., 5, and subsequently aggregated over all moneyness-maturity groups. Therefore, we propose an IVS Residual Correction Model that copes with the the error terms of Model 1 and Model 2. We do this by estimating an AR(1) model for the grouped residuals $\epsilon_{i,j,t}$, for all i, j = 1, ..., 5:

$$\epsilon_{i,j,t} = \rho_{i,j}\epsilon_{i,j,t-1} + \eta_{i,j,t} \tag{4}$$

(a) Average RMSE from Model 2 after applying IVS Residual Correction Model





Figure 3: Panel (a) shows the average RMSE after applying IVS specification 2 and Panel (b) shows the average RMSE after applying IVS specification 1, but the the RMSE values are obtained after applying the IVS Residual Correction Model on the fitting errors corresponding to the right deterministic model and covers the period July 7, 2003 - December 29, 2017.

Figure 3 shows the same illustration as Figure 2, namely the RMSE after applying the two deterministic models 2 and 1 in Panel (a) and Panel (b), respectively, but now after applying the IVS Residual Correction Model in Equation 4. The heteroskedasticity across the different moneyness-maturity groups reduces substantially after applying our proposed IVS Residual Correction Model 4, as the figures are more flat than in Figure 2.

Table 3 shows the difference in aggregated Ljung-Box test statistics, with one lag, between the errors $\epsilon_{i,j,t}$ and $\eta_{i,j,t}$. Applying our proposed IVS Residual Correction Model reduces the fitting errors' autocorrelation considerably. Model 4 reduces the LB(1) test with 21.40 to 37.37 percent and 37.41 to 53.57 percent for deterministic IVS Model 2 and 1, respectively. Moreover, the errors corresponding with the 7-factor model are correlated weaker than the fitting errors that correspond to the 5-factor model (the LB(1) test statistics are nearly twice as small). All by all, the combination of our proposed IVS Residual Correction Model and the 7-factor model is able to reduce the residuals' heteroskedasticity and autocorrelation considerably, relatively to the alternative 5-factor specification and might have the ability to obtain more accurate IVS estimates and more efficient predictions of the future IVS states.

Table 3: Ljung-Box test statistics on the residuals before and after applying the IVS Residual Correction Model July 7, 2003 - December 29, 2017

	IVS speci	fication in equat	tions 2-3	IVS specification in equation 1 from					
				Goncaly	(2006) (
	Exc. Model 4	Inc. Model 4	% of change	Exc. Model 4	Inc. Model 4	% of change			
S&P 500:	8790.32	6901.09	(-21.49)	25104.87	12500.77	(-50.21)			
Dow Jones:	14780.97	9873.14	(-33.20)	25820.83	16160.58	(-37.41)			
CBOE Mini-NDX:	4384.84	2746.05	(-37.37)	14401.80	6686.61	(-53.57)			

Notes to Table: The table reports the Ljung-Box test statistic - one lag included - over the residuals corresponding to IVS specification 2-3 and 1, both including and excluding the IVS Residual Correction Model. The residuals are classified per moneyness-maturity group and the LB(1)-tests are accumulated among the different groups. The table also reports the change in percentages before and after including the IVS Residual Correction Model.

3.3 In-sample performances

Table 4 reports the average goodness of fit for the two different deterministic IVS specifications, both including and excluding the IVS Residual Correction Model. Note that the 7-factor model contains two more variables than the model proposed by Goncalves and Guidolin (2006) and thus it is expected to have a better R^2 . To correct for the number of variables in the IVS specification, we use the adjusted $-R^2$ for comparison purposes. First of all, the 7-factor model has a better in-sample fit, relatively to the 5factor model. Enabling the IVS specification to be more flexible towards asymmetries in the moneyness dimension and including term-structures provide consistently higher (average) $adjusted - R^2$ among the three indices. Secondly, over-performing abilities in re-producing the observed implied volatility surfaces can be achieved by making use of the IVS Residual Correction Model. Reducing the error's heteroskedasticity and autocorrelations result in better in-sample fits. Moreover, the combination of the 7-factor model and the IVS Residual Correction Model gives superior in-sample performances. Minimum and maximum goodness of fit values increase, even up to 100 percent for the latter. The increase of the median suggests better in-sample fits for more than half of the total sample period. The average $adjusted - R^2$ increases with 5.02, 6.30, and 5.68 percent for SPX, DJX, and MNX, respectively. Summarized, both the 7-factor model and IVS Residual Correction Model give us the ability to increase descriptive accuracies of the index implied volatility surfaces.

Table 4: Goodness of fit test for the different IVS specifications, including and excluding the IVS Residual Correction Model July 7, 2003 - December 29, 2017

]	IVS specification	n in equations 2	-3	IVS specification in equation 1 from							
					Goncalves and Guidolin (2006)							
	Mean R_{adj}^2	Median R^2_{adj}	Min R^2_{adj}	Max R_{adj}^2	Mean R_{adj}^2	Median R^2_{adj}	Min R^2_{adj}	Max R_{adj}^2				
S&P 500:	93.12(95.42)	96.29(97.94)	4.69(7.30)	99.28(100.00)	90.40(91.99)	93.32(94.71)	-0.65(-2.39)	99.33(100.00)				
	()		((()	()	((
Dow Jones:	89.02(92.00)	93.80(95.13)	13.34(32.40)	99.42(100.00)	85.70(89.53)	90.00(93.01)	-2.98(17.44)	99.05(100.00)				
CBOE Mini-NDX:	90.60(92.53)	95.05(96.72)	-13.01(-8.26)	99.37(100.00)	86.85(88.13)	91.52(92.52)	-18.77(-11.41)	98.81(100.00)				

Notes to Table: For each index, the table reports the (i) mean, (ii) median, (iii) minimum, and (iv) maximum $adjusted - R^2$ - over the sample period July 7, 2003 - December 29, 2017 - for both the 7-and 5-factor IVS specification. The same four statistics - after applying the IVS Residual Correction Model on the fitting errors of the corresponding deterministic IVS model - are given in parentheses.

4 Modeling the time-variation of the IVS

This section consists of three parts. We first introduce and explain five simple time series models that enables us to construct implied volatility forecasts. Thereafter, we propose and clarify the restricted dynamic factor model and explain the Kalman Filter estimation procedure. Lastly, we will analyze the out-of-sample performances of all proposed time series models.

4.1 Time-series models

Section 3.3 shows that the proposed 7-factor model produces a better in-sample fit than the 5-factor model of Goncalves and Guidolin (2006). Moreover, our proposed IVS Residual Correction Model 4 also shows improving in-sample performances, in means of $adjusted - R^2$. However, if the proposed models are informatively more efficient than alternative existing IVS specifications, then the proposed models should not only perform better in in-sample performances, but also in - more importantly - out-of-sample predictive abilities. The Ljung-Box test in Section 3.1 finds significant autocorrelation in the beta estimates, and thus suggests predictability in the beta coefficients and subsequently predictability in the implied volatility of option contracts. We will use the following time series models to make forecasts for the β estimates ($\hat{\beta}_{l,t}$ for l = 1, ..., 7 for Model 2 and l = 1, ..., 5 for Model 1):⁵

• Univariate autoregressive models of order p_{AR} , denoted as $\hat{\beta} - AR(p_{AR})$

$$\hat{\beta}_{l,t} = c_l + \sum_{i=1}^{p_{AR}} \phi_i \hat{\beta}_{l,t-i} + u_{l,t}$$
(5)

• Vector autoregressive models of order p_{VAR} , denoted as $\hat{\beta} - VAR(p_{VAR})$

$$\hat{\beta}_t = \mathbf{c} + \sum_{i=1}^{p_{VAR}} \Phi_i \hat{\beta}_{t-i} + \mathbf{u}_t \tag{6}$$

• Bayesian multivariate autoregressive models of order p_{BVAR} , denoted as $\hat{\beta} - BVAR(p_{BVAR})$

$$\hat{\beta}_t = \mathbf{c} + \sum_{i=1}^{p_{BVAR}} \Phi_{B,i} \hat{\beta}_{t-i} + \mathbf{e}_t, \tag{7}$$

 $^{^{5}}$ As mentioned before, estimating the following time series models will be done recursively, with a rolling window of six trading months - or 126 trading days - to allow for possible non constant parameters over time. The economic environment changes dynamically over time and thus it is not reasonable to assume that the time series model's parameters are constant over time considering daily beta coefficients with a sample period of fifteen years.

where we use the prior of Minnesota by Litterman (1979), and the full VAR estimates will be used to initialize the residual covariance matrix. Furthermore, the optimal lag orders p_i , where $i = \{AR, VAR, BVAR\}$, are obtained by minimizing the Akaike Information Criterion (AIC) for every window sample and is restricted on a maximum of ten. For the $\hat{\beta} - AR(p_{AR})$ model, the AIC is applied on all five or seven individual equations.

The Augmented Dickey Fuller test in Section 3.1 safely rejects the presence of a unit root for the beta factors in only a small percentage of all window samples. Models 5-7 do not take the presence of a possible unit root into account and are likely to give inaccurate estimates and predictions for window samples where a unit root is detected. Therefore, we propose the following two time series models that correct for non-stationary processes, with the first differences defined as $\hat{\gamma}_t \equiv \hat{\beta}_t - \hat{\beta}_{t-1} = (1-L)\hat{\beta}_t$, where L is the lag operator:

• Univariate autoregressive fractionally integrated models, denoted as $\hat{\beta} - AFRI(1, d)$

$$\phi_l (1-L)^{d_l} (\hat{\beta}_{l,t} - c_l) = \epsilon_{l,t}, \tag{8}$$

where d_l is the fractional difference parameter and will be random generated over the uniform distribution [-0.5, 0.5].

• Error-correction models with $r \leq 7$ (or $r \leq 5$ for the 5-factor model) common stochastic trends, denoted as $\hat{\gamma} - ECM$

$$\hat{\gamma}_t = \mathbf{c} + \mathbf{B}\hat{\gamma}_{t-1} + \Theta\beta_{t-1} + \mathbf{v}_t,\tag{9}$$

where Θ is a matrix consisting of the coefficients of the error correction terms. For the latter time series model, we use the VAR-based Johansen cointegration tests in every window sample, by Johansen (1988), to test for the number of cointegration relations.⁶ We make the assumption of a linear trend in the beta coefficients and an intercept - but no trend - in the cointegration relations. However, if this results in a non-singular cointegration matrix, we add a trend in the cointegration relations. Both assumptions are proposed in Johansen (1995). If the number of cointegration relations is estimated to be zero in a window sample, we use $\hat{\beta} - VAR(p_{VAR})$ to model the beta factors dynamically.

4.2 Dynamic factor model

Besides the time series models in Section 4.1, we propose an additional dynamic factor model to study the dynamics of the IVS. In particular, we use a restricted factor model that captures the seven key indicators, noted in Equation 3, of the IVS. We will construct a general State Space Model to obtain predictions of the future IVS. A similar study on S&P 500 index options is done by van der Wel et al. (2016). However, such a dynamic factor model applied on the 7-factor IVS specification, including term structures and asymmetries in option contracts, is in our knowledge not done before.

The observation vector, y_t , consists of the implied volatilities on day t. Vectorizing the volatilities of the 25 moneyness-maturity midpoint option contracts results in:

$$y_t = \begin{pmatrix} IV_{\tau_1,\Delta_1,t} \\ \vdots \\ IV_{\tau_T,\Delta_1,t} \\ \vdots \\ \vdots \\ IV_{\tau_T,\Delta_M,t} \end{pmatrix},$$

⁶These test are elaborated in more detail after its revelation, in particular by Johansen (1991), Johansen (1992), and Johansen (1994)

which is a $(TM \times 1)$ vector and $IV_{\tau_j,\Delta_i,t}$ denotes the implied volatility for an option contract with maturity τ_j and moneyness Δ_i , for j = 1, ..., T and i = 1, ..., M.⁷ Note that option contracts on sequential days are not identical, because it is not possible to have the same maturity and moneyness (delta) values on each day and across the different moneyness-maturity groups. However, we make use of interpolated data and for notational conveniences, we use the time-subscripts for τ and Δ , and use the moneyness (maturity) subscripts for maturity (moneyness). The general State Space Model is given by:

$$y_t = Zf_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, H)$$

$$f_{t+1} = d + Tf_t + \eta_{t+1}, \qquad \eta_{t+1} \sim \mathcal{N}(0, Q),$$
(10)

where y_t is the observation vector defined above, f_t is a vectorization of the beta estimates in Model 2, and ϵ_t is a vector of measurement residuals with variance-covariance matrix H. We use a simple VAR model of order one in the transition equation, as higher order VAR models will increase the computational time substantially, but might be interesting for future research. We add a constant term d in the transition equation to allow for possible trends in the f_t factors. The transition matrix T contains the VAR-coefficients and η_t are the factor innovations, which we assume to be normally distributed with variance-covariance matrix Q. The factors f_t are obtained by running regression Model 2 and collecting the beta coefficients as latent factors, resulting in a restricted dynamic factor model with a restricted factor loading matrix Z, constructed as:

where $I_{l,i}$, $I_{l,j}$, and $I_{l,i,j}$ (for l = 0, ..., 6) are the seven indicators defined in 3. Note that matrix Z contains the actual indicator values corresponding to each moneyness-maturity group on day t and - due to interpolated data - is constant over time.

We will use the Kalman Filter method of Durbin and Koopman (2012) to estimate the State Space Model defined above. The Kalman Filter procedure is a recursive estimation method that estimates latent factors from past observations. We define the mean and variance of the latent factors at time t + 1 as $a_{t+1} = \mathbb{E}[f_{t+1}|I_t]$ and $P_{t+1} = \mathbb{V} ar(f_{t+1}|I_t)$, where I_t is the information available on time t. The initial state distribution of the latent factor is equal to $f_1 \sim \mathcal{N}(a_1, P_1)$. Hence, after specifying the initial states for a_1 and P_1 as a vector of ones and as the identity matrix, respectively, we can obtain the optimal filtered estimates $a_{t|t}$ and $P_{t|t}$ and optimal predictions a_{t+1} and P_{t+1} by the following Kalman Filter iteration procedure:

$$v_{t} = y_{t} - Za_{t}$$

$$F_{t} = ZP_{t}Z' + H$$

$$a_{t|t} = a_{t} + P_{t}Z'F_{t}^{-1}v_{t}$$

$$P_{t|t} = P_{t} - P_{t}Z'F_{t}^{-1}ZP_{t}$$

$$a_{t+1} = Ta_{t|t} + d$$

$$P_{t+1} = TP_{t|t}T' + Q$$

$$(11)$$

The Kalman gain is the optimal weighting matrix for the new evidence and equals:

$$K_t = T_t P_t Z' F_t^{-1} \tag{12}$$

 $^{^{7}}$ T and M are the number of maturity and moneyness categories, respectively. Note that in our application T and M both equal 5.

We again use a moving window of six trading months to estimate the Kalman Filter procedure. After every iteration we obtain the latent variable's mean, variance, and the one day ahead forecasts of those two. We will examine the performances of all discussed time series models in the next section.

4.3 Out-of-Sample Performances

If the 7-factor model and the IVS Residual Correction Model contain valuable information, regarding the dynamics of the implied volatility, then it should perform better than the 5-factor model proposed by Goncalves and Guidolin (2006) in economically context, such as out-of-sample forecasting performances and trading strategies. The future IVS state at time t + h depends on the beta coefficients $\beta_{l,t+h}$ and the indicators on time t + h, defined in 3. For each window sample, we are able to make implied volatility h-step ahead forecasts - for each available index option - by plugging in the obtained beta forecasts $\hat{\beta}_{l,t+h}$ into Equation 2, for $h = 1, 2, 5, 10, 15, 20.^8$ Consequently, we correct the volatility prediction with the IVS Residual Correction Model, where we use the residual predictions $\hat{\epsilon}_{i,j,t+h}$. The estimation and forecasting procedure of the IVS Residual Correction Model is similar to the time series models, namely using a rolling window of six trading months, as explained in Section 4.1. Therefore, our sample period starts six months later (126 trading days), which becomes July 7, 2003.

Forecasting the beta estimates for the time series models 5-9 and the residuals $\epsilon_{i,j,t}$ is straightforward, but obtaining multiple-step ahead forecasts from the State Space Model might require a small clarification. Note that the 1-step ahead forecast is already calculated in Equation 11. Multiple step ahead forecasts are obtained by treating future observations j as missing. Thus, put v_j as unknown and $F_j \to \infty$, for j = t + 1, ..., t + h. This gives $K_j \to 0$ and proceeding the Kalman Filter procedure as normal gives the forecasts $a_{t+h} = Ta_{t+h-1} + d$ and $P_{t+h} = TP_{t+h-1}T' + Q$.

In order to evaluate the efficiency of the implied volatility predictions, we propose the following benchmark models:

• The ad-hoc 'Strawman' model, also used in Dumas et al. (1998). Here, the IVS beta coefficients follow a random walk process:

$$\hat{\beta}_{t+h} = \hat{\beta}_t + \omega_{t+h}, \quad \omega_{t+h} \sim IID(0, \Sigma)$$
(13)

- The 7-factor model excluding the IVS Residual Correction Model.
- All the proposed time series models applied on the 5-factor model of Goncalves and Guidolin (2006), also containing the 'Strawman' random walk model both including and excluding the IVS Residual Correction Model.
- A VAR(1) process for the beta coefficients corresponding to the 5- and 7-factor IVS specifications. In order to assess the performances of the State Space Model that contains a VAR(1) process as transition equation, it is biased to compare those with time series models that are more flexible in the number of lags. Therefore, we will compare the performances of the State Space Model with a VAR process of lag order one, also including and excluding the IVS Residual Correction Model.

We evaluate the out-of-sample performances of the competing (benchmark) models across six different forecast horizons by three statistical measures. We compute the average (across all out-of-sample periods) (i) root mean squared forecast error (RMSE), (ii) mean absolute forecast error (MAE), and the (iii) mean correct predicted direction of change statistic (MCPDC), also used in Bernales and Guidolin (2014). The MCPDC is defined as the percentage of correctly predicting the direction of change in the implied volatility. The direction of change is commonly used in finance to construct trading strategies. Hence, the MCPDC is a

⁸Note that the delta of an option contract on day t + h is unknown on day t. Therefore, we assume that the best prediction of an option delta h days later is simply today's delta. This is a reasonable assumption according to the efficient market hypothesis, as the option delta only depends on the underlying stock price.

good indicator whether the IVS forecasts contain valuable economic information. Panel A in Table 5 reports these three statistics for the 1-day ahead forecasts obtained by the time series models 5-9, for both the 5- and 7-factor IVS specifications. The out-of-sample performances including the IVS Residual Correction Model are given in parentheses. Note that we use the notation p as optimal lag operator for all the time series models instead of p_i , where $i = \{AR, VAR, BVAR\}$, for notational conveniences.

Table 5: Evaluation of 1-day ahead out-of-sample forecasting performances July 7, 2003 - December 29, 2017

		IV	S specification in	equations 2-3			IVS specification in equation 1 from						
								C	Goncalves and Gu	idolin (2006)			
	RW	AR(p)	AFRI(1, d)	VAR(p)	BVAR(p)	ECM	RW	AR(p)	AFRI(1, p)	VAR(p)	BVAR(p)	ECM	
Panel A: Forecasts O	btained from	Time Series Mod	lels 5-9										
S&P 500:													
RMSE	0.0423	0.0418	0.2685	0.0430	0.0419	0.0399	0.0445	0.0429	0.0554	0.0437	0.0433	0.0427	
	(0.0321)	(0.0314)	(0.2671)	(0.0328)	(0.0313)	(0.0288)	(0.0370)	(0.0351)	(0.0494)	(0.0360)	(0.0354)	(0.0348)	
MAE	0.0218	0.0212	0.0605	0.0215	0.0212	0.0197	0.0240	0.0223	0.0293	0.0229	0.0225	0.0223	
	(0.0174)	(0.0164)	(0.0562)	(0.0166)	(0.0166)	(0.0145)	(0.0202)	(0.0181)	(0.0258)	(0.0188)	(0.0184)	(0.0180)	
MCPDC	53.09%	53.58%	53.54%	52.66%	53.85%	52.32%	53.20%	53.36%	52.18%	52.72%	53.47%	52.58%	
	(52.80%)	(53.71%)	(53.83%)	(52.72%)	(54.05%)	(52.24%)	(53.17%)	(53.62%)	(52.09%)	(52.81%)	(53.73%)	(52.63%)	
Dow Jones:													
RMSE	0.0522	0.0531	0.2451	0.0535	0.0521	0.0503	0.0533	0.0518	0.0863	0.0533	0.0524	0.0516	
	(0.0380)	(0.0388)	(0.2419)	(0.0393)	(0.0373)	(0.0349)	(0.0392)	(0.0369)	(0.0786)	(0.0388)	(0.0375)	(0.0366)	
MAE	0.0249	0.0247	0.0548	0.0251	0.0246	0.0232	0.0261	0.0239	0.0383	0.0251	0.0239	0.0243	
	(0.0197)	(0.0190)	(0.0499)	(0.0193)	(0.0191)	(0.0170)	(0.0205)	(0.0178)	(0.0333)	(0.0190)	(0.0179)	(0.0180)	
MCPDC	53.92%	54.18%	52.90%	53.68%	54.03%	53.59%	53.99%	54.34%	53.12%	54.04%	54.60%	53.69%	
	(53.65%)	(54.34%)	(52.82%)	(53.83%)	(54.01%)	(53.77%)	(53.74%)	(54.51%)	(53.08%)	(54.20%)	(54.73%)	(52.63%)	
CBOE Mini-NDX:													
RMSE	0.0370	0.0373	0.1645	0.0384	0.0358	0.0345	0.0398	0.0374	0.1863	0.0394	0.0373	0.0375	
	(0.0341)	(0.0343)	(0.1637)	(0.0355)	(0.0327)	(0.0313)	(0.0358)	(0.0331)	(0.1852)	(0.0352)	(0.0329)	(0.0331)	
MAE	0.0183	0.0181	0.0397	0.0182	0.0177	0.0166	0.0208	0.0186	0.0417	0.0196	0.0185	0.0191	
	(0.0170)	(0.0169)	(0.0386)	(0.0168)	(0.0164)	(0.0151)	(0.0190)	(0.0166)	(0.0340)	(0.0176)	(0.0166)	(0.0169)	
MCPDC	55.08%	55.42%	54.68%	55.73%	55.81%	55.73%	54.58%	55.72%	54.50%	55.53%	56.12%	55.43%	
	(54.82%)	(55.25%)	(54.49%)	(55.69%)	(55.63%)	(55.69%)	(54.38%)	(55.69%)	(54.47%)	(55.49%)	(56.14%)	(55.43%)	
Panel B: Forecasts ()	btained from	the State Space I	Model and VAR(1	DECOGRES									
Taller D. Torceases O	branico nom	VAR(1) - exc	VAR(1) - inc	SSM - exc	SSM - inc			VAR(1) - exc	VAR(1) - inc	SSM - exc	SSM - inc		
S&P 500:		()	(/					()	()				
RMSE		0.0431	0.0329	0.0403	0.0291			0.0442	0.0366	0.0429	0.0350		
MAE		0.0215	0.0167	0.0195	0.0143			0.0235	0.0195	0.0223	0.0179		
MCPDC		52.82%	53.01%	53.57%	53.56%			52.85%	52.97%	53.42%	53.54%		
Dow Jones:													
RMSE		0.0537	0.0370	0.0509	0.0355			0.0536	0.0385	0.0521	0.0370		
MAE		0.0248	0.0184	0.0228	0.0168			0.0248	0.0193	0.0238	0.0176		
MCPDC		53.73%	53.86%	53.97%	53.79%			53.97%	54.03%	54.15%	53.94%		
CBOE Mini-NDX:													
RMSE		0.0386	0.0318	0.0342	0.0310			0.0395	0.0338	0.0373	0.0328		
MAE		0.0183	0.0154	0.0160	0.0145			0.0196	0.0173	0.0184	0.0163		
MCPDC		56.15%	56.12%	55.46%	55.24%			55.99%	56.05%	55.07%	54.95%		

The probability of the statistic for the late statistic for the State Space Model and VAR(1) process. AR(L), the and a space of the statistic for the sta

As evident from the table, we see that the BVAR and ECM models, based on the betas of the 7-factor specification, can improve the 1-day ahead forecasts in means of RMSE and MAE for each index, relatively to the Strawman random walk. For the IVS specification with 5 factors, this also applies for the AR model. The ECM (AFRI) time series model is consistently the best (worst) performing model across all indices, at least statistically, regarding to the RSME and MAE statistics. There is at least one time series model, in both the 5-factor and 7-factor IVS specification, that outperforms the Strawman random walk in means of MCPDC, which could be helpful in construction trading strategies (elaborated further in Section 5).

Table 5 shows two more very important statistical findings. The first finding is that IVS specification 2 performs better for each index, in means of RMSE or MAE, than the IVS specification of Goncalves and Guidolin (2006) across all time series models (except for the AFRI model, which performs substantially lower than any of the other time series models). Secondly, our proposed IVS Residual Correction Model diminishes the RMSE and MAE statistics consistently across all surfaces and time series models. We find reductions in RMSE and MAE of around 25% for the S&P 500 and Dow Jones option contracts and around 8% for the CBOE Mini-NDX options. In means of MCPDC, we do not find consistent improvements for the 7-factor model nor by using our IVS Residual Correction Model, but improvements depend on which time series model is used.

Panel B in Table 5 reports the VAR(1) and State Space Model out-of-sample performances for 1-day ahead forecasting horizons, based on both the 5- and 7-factor specifications. The RMSE, MAE, and MCPDC -



Figure 4: The figure plots in percentages - for each surface and different forecasting horizons - the number of option contracts for which the four best performing time series models statistically outperform the 7factor IVS specification Strawman random walk model in predictive accuracy (significance tested by means of the equal predictive power test of Diebold and Mariano (1995)). Notes to the legends: AR(p) denotes the univariate autoregressive model of order p_{AR} , BVAR(p) denotes the Bayesian vector autoregressive model with lag order p_{BVAR} , ECM denotes the error-correction model, and the SSM denotes the restricted factor dynamic State Space Model. Furthermore, the 5 (7) denotes that the forecasts are obtained from the 5- (7-) factor IVS specification and a * stands for IVS Residual Correction Model included. July 7, 2003 - December 29, 2017.

after applying the IVS Residual Correction Model - are denoted by -inc. The table confirms the statistical superior out-of-sample performances, in means of RMSE and MAE, for the 7-factor model. Furthermore, the State Space Model, estimated by the Kalman Filter procedure, consistently improves statistical out-of-sample performances across all indices. Even more accurate predictions can be achieved by applying the IVS Residual Correction Model. These gains in efficiency become larger for longer forecasting horizons, as shown in Table 7 in Appendix B. The use of our proposed State Space Model can also improve the percentage of MCPDC, but is not consistent over all indices and horizons.

If we compare the 1-step ahead forecast performances with the other time series models in Panel A of Table 5, we see that our proposed State Space Model, with a VAR(1) process as stationary equation, competes pretty well with the more advanced - and more flexible in the lag order - models. Based on the 7-factor IVS specification, it even over-performs alternative time series models in means of RMSE and MAE for the CBOE Mini-NDX surface.

Figure 4 plots the percentage of option contracts for which the four best performing time series models outperform the Strawman random walk model, based on the 7-factor IVS specification, in predictive accuracy, against different forecasting horizons h.⁹ We exclude the AFRI time series processes and the Strawman random walk model based on the 5-factor IVS specification, as those are the lowest performing models across all forecasting horizons. We do not show the VAR process in the figures, as it performs nearly equivalent to, but slightly less than, the BVAR models and would clutter the figures. We also remove the predictive performances of the 5-factor IVS specification including the IVS Residual Correction Model, because it has similar predictive abilities as the 7-factor specification excluding the IVS Residual Correction Model. Including these performances in the figure would confuse the reader and does not add valuable information to - or alter - any conclusions we draw.

Ignoring the State Space Model, the ECM and AR time series processes enables us to obtain the most accurate IVS predictions, based on the 7-factor IVS specification, both for including and excluding the IVS Residual Correction Model. The State Space Model performs best - or slightly less than the best performing time series model - across all forecasting horizons, in means of out-of-sample forecasting performances. Keeping in mind that the transition equation in the State Space Model is just a simple VAR(1) process, its predictive accuracy is remarkable. It performs superior relatively to more advanced (than a simple VAR(1) process) time series models. Enabling the transition equation to be more advanced should be able to increase the predictive accuracy and is a good idea to examine in future research. However, we already find its usefulness in predicting the implied volatility in option contracts.

The figure shows superior predicting abilities for the 7-factor model, relatively to the 5-factor model of Goncalves and Guidolin (2006), across all forecasting horizons and index surfaces. Moreover, our proposed IVS Residual Correction Model improves the implied volatility forecasting accuracy substantially. The predictive ability in outperforming the 7-factor IVS specification Strawman random walk model increases with a range of (approximately) 15% for NDX up to 35% for Dow Jones for short forecasting horizons. However, the 7-factor IVS predictions, including the IVS Residual Correction Model - for longer horizons. Summarized, the combination of the 7-factor model and the IVS Residual Correction Model improves the out-of sample performances with a range of 30%-50% for short horizons and around 10%-15% for long horizons, relatively to the more common 5-factor model, and thus enables us to get superior IVS predictions.

⁹Significance of outperforming the ad-hoc Strawman random walk model is tested by means of the predictive power test of Diebold and Mariano (1995). We used this test to find significance differences in squared forecast errors (at a level of 5%). We implemented the test by considering the obtained forecasting errors - of the different time series models and IVS specifications - for each option contract and computed the test statistic for each option separately. Subsequently, we computed the percentage of total option contracts for which the denoted models have significant more accurate forecasts than the 7-factor IVS specification Strawman random walk.

5 Economic Evaluation

Sections 3.3 and 4.3 show the superior in- and out-of-sample performances of the 7-factor model - combined with the proposed IVS Residual Correction Model - and the gains in predictive improvements by making use of specific time series models (such as the univariate autoregressive model, error-correction model, and the proposed restricted dynamic State Space Model). To examine the economic significance of these prediction improvements, we simulate a simple trading strategy based on the implied volatility forecasts. We use the h-day step ahead forecasts to construct a portfolio, containing option contracts, to examine if we can achieve abnormal returns. We will compare the economic performance of the best performing models, based on the previous sections, to see if our proposed models also outperform economically.

Our simulated investor trades for \$1000 in a delta-hedge portfolio, which is held for only h trading days. Delta hedge portfolios are not exposed to risk that is caused by the price movements of the underlying. The mean idea of the strategy is to purchase (sell) a option contract if it is predicted that the implied volatility on time t + h exceeds (falls below) today's implied volatility. Therefore, we see the importance if the calculated MCPDC statistics in the previous section. We filter out option contracts with an expected change in implied volatility smaller than two percent, to drop noisy forecasts. We also drop option contracts where the option delta $\Delta < 0.3$ or $\Delta > 0.7$, as these could contain differences of such a magnitude that the results - to compare the different models - may be biased. The portfolio will be re-balanced every day t, such that the initial investment of \$1000 stays constant over time.

Suppose that Q_t is the set of option contracts that should be traded at time t. If there are no option contracts traded on day t, that means $Q_t = 0$, then the initial \$1000 will be invested in the risk-free rate $r_{f,t}$, for which we take the daily spot rates of the three months U.S Treasury Bill (T-Bill). Let V_t be the total value of our delta hedge portfolio on day t, then we have:

$$V_{t} = \sum_{m \in Q_{t,+}^{Call}} (C_{m,t} - S_{t} \Delta_{m,t}^{C}) + \sum_{m \in Q_{t,+}^{Put}} (P_{m,t} + S_{t} \Delta_{m,t}^{P}) - \sum_{m \in Q_{t,-}^{Call}} (C_{m,t} - S_{t} \Delta_{m,t}^{C}) - \sum_{m \in Q_{t,-}^{Put}} (P_{m,t} + S_{t} \Delta_{m,t}^{P}),$$
(14)

where $C_{m,t}$ $(P_{m,t})$ is the price of call (put) option m, S_t is the price of the underlying, $Q_{t,+}^{Call}$ $(Q_{t,-}^{Call})$ is the subset of call options that should be purchased (sold) (the same interpretation holds for the put options), and $\Delta_{m,t}^C$ $(\Delta_{m,t}^P)$ is the option's delta of call (put) option m in absolute value. If the total net value of the portfolio is positive, we invest the quantity $X_t = \$1000/V_t$ in our delta hedge portfolio, which produces a total cost of \$1000. Therefore, the h-day net gain (or loss if negative) on time t+h is:

$$G_{t+h} = X_t \left[\sum_{m \in Q_{t,+}^{Call}} \left((C_{m,t+h} - S_{t+h} \Delta_{m,t}^C) - (C_{m,t} - S_t \Delta_{m,t}^C) \right) \right] + X_t \left[\sum_{m \in Q_{t,+}^{Put}} \left((P_{m,t+h} + S_{t+h} \Delta_{m,t}^P) - (P_{m,t} + S_t \Delta_{m,t}^P) \right) \right] + X_t \left[\sum_{m \in Q_{t,-}^{Call}} \left(-(C_{m,t+h} - S_{t+h} \Delta_{m,t}^C) + (C_{m,t} - S_t \Delta_{m,t}^C) \right) \right] + X_t \left[\sum_{m \in Q_{t,-}^{Put}} \left(-(P_{m,t+h} + S_{t+h} \Delta_{m,t}^P) + (P_{m,t} + S_t \Delta_{m,t}^P) \right) \right] \right]$$
(15)

If the net total value of our simulated portfolio is negative, we sell the amount $X_t = \frac{1000}{|V_t|}$ of the portfolio and invest the incoming cashflow of \$1000, plus the additional initial \$1000, in the risk free rate

 $r_{f,t+i}$, for every day between t and h, i = 0, ..., h - 1. Hence, the net gain (or loss if negative) is the amount $G_{t+h} + \sum_{i=0}^{h-1} \$2000 \cdot (exp(r_{f,t+i}/252) - 1).$

If the value of the portfolio V_t is below \$10, we do not invest in this portfolio, but invest the initial \$1000 at the risk free rate, because this could give unfeasible positions in reality. Furthermore, the delta hedge strategy proposed above is self-financing and, as a result, the leverage used in the portfolio is varying over time. Therefore, the risk that the simulated trader takes, is not constant over time. Large outliers could occur and would affect our results. Hence, we use 99%-winsorized returns to get more robust and reliable results for the simulated mean returns.

Table 6: Average returns over the forecasting horizon period h July 7, 2003 - December 29, 2017

	IVS specification in equations 2-3							IVS specification in equation 1 from					
								Go	ncalves and	Guidolin (20	06)		
	RW	AR(p)	VAR(p)	BVAR(p)	ECM	SSM	RW	AR(p)	VAR(p)	BVAR(p)	ECM	SSM	
S&P 500:													
1	0.161	0.173	0.183	0.204	0.202	0.204	0.121	0.138	0.148	0.151	0.105	0.181	
	(0.181)	(0.189)	(0.184)	(0.208)	(0.220)	(0.234)	(0.136)	(0.110)	(0.183)	(0.148)	(0.106)	(0.183)	
2	0.128	0.112	0.161	0.150	0.177	0.145	0.079	0.152	0.113	0.102	0.109	0.120	
	(0.126)	(0.144)	(0.185)	(0.152)	(0.179)	(0.149)	(0.142)	(0.161)	(0.124)	(0.122)	(0.120)	(0.128)	
5	0.209	0.194	0.239	0.159	0.181	0.185	0.194	0.147	0.197	0.106	0.167	0.163	
	(0.194)	(0.208)	(0.289)	(0.199)	(0.178)	(0.197)	(0.182)	(0.178)	(0.199)	(0.116)	(0.181)	(0.165)	
10	0.142	0.171	0.192	0.177	0.192	0.190	0.147	0.198	0.151	0.151	0.176	0.161	
	(0.172)	(0.186)	(0.198)	(0.182)	(0.211)	(0.210)	(0.147)	(0.204)	(0.167)	(0.168)	(0.178)	(0.168)	
15	0.116	0.194	0.201	0.157	0.235	0.175	0.133	0.170	0.190	0.095	0.182	0.165	
	(0.154)	(0.204)	(0.225)	(0.182)	(0.233)	(0.176)	(0.168)	(0.179)	(0.198)	(0.103)	(0.192)	(0.184)	
20	0.128	0.187	0.196	0.170	0.215	0.170	0.140	0.132	0.157	0.076	0.144	0.141	
	(0.139)	(0.189)	(0.226)	(0.187)	(0.212)	(0.190)	(0.156)	(0.132)	(0.170)	(0.085)	(0.145)	(0.153)	
Dow Jones:													
1	0.129	0.097	0.137	0.139	0.143	0.165	0.103	0.090	0.075	0.065	0.038	0.065	
	(0.140)	(0.104)	(0.162)	(0.145)	(0.183)	(0.211)	(0.149)	(0.089)	(0.109)	(0.104)	(0.118)	(0.111)	
2	0.180	0.162	0.208	0.145	0.168	0.190	0.129	0.061	0.128	0.114	0.111	0.096	
	(0.159)	(0.193)	(0.213)	(0.175)	(0.185)	(0.208)	(0.152)	(0.104)	(0.190)	(0.110)	(0.120)	(0.149)	
5	0.115	0.135	0.119	0.133	0.117	0.156	0.068	0.098	0.072	0.088	0.116	0.094	
	(0.155)	(0.168)	(0.114)	(0.154)	(0.132)	(0.174)	(0.075)	(0.098)	(0.077)	(0.089)	(0.121)	(0.102)	
10	0.153	0.170	0.212	0.166	0.200	0.183	0.098	0.104	0.158	0.083	0.099	0.145	
	(0.195)	(0.168)	(0.203)	(0.180)	(0.199)	(0.195)	(0.112)	(0.129)	(0.160)	(0.089)	(0.107)	(0.161)	
15	0.117	0.144	0.157	0.128	0.136	0.125	0.103	0.080	0.060	0.131	0.104	0.107	
	(0.122)	(0.170)	(0.177)	(0.141)	(0.158)	(0.144)	(0.109)	(0.081)	(0.064)	(0.135)	(0.108)	(0.141)	
20	0.161	0.140	0.100	0.161	0.106	0.165	0.087	0.098	0.055	0.073	0.103	0.091	
	(0.168)	(0.144)	(0.108)	(0.176)	(0.113)	(0.183)	(0.097)	(0.102)	(0.053)	(0.073)	(0.107)	(0.105)	
CBOE Mini-NDX:													
1	0.244	0.190	0.397	0.204	0.466	0.367	0.158	0.142	0.223	0.167	0.158	0.197	
	(0.291)	(0.207)	(0.449)	(0.294)	(0.488)	(0.391)	(0.208)	(0.215)	(0.225)	(0.172)	(0.208)	(0.225)	
2	0.055	0.152	0.209	0.195	0.155	0.187	0.035	0.047	0.115	0.122	0.144	0.101	
	(0.051)	(0.168)	(0.288)	(0.283)	(0.298)	(0.243)	(0.079)	(0.057)	(0.189)	(0.132)	(0.164)	(0.139)	
5	0.098	0.153	0.185	0.164	0.183	0.138	0.095	0.021	0.033	0.071	0.163	0.156	
	(0.105)	(0.167)	(0.215)	(0.165)	(0.284)	(0.175)	(0.102)	(0.072)	(0.040)	(0.097)	(0.165)	(0.169)	
10	0.081	0.124	0.151	0.148	0.145	0.195	0.028	0.063	0.051	0.045	0.066	0.169	
	(0.059)	(0.185)	(0.165)	(0.153)	(0.166)	(0.223)	(0.034)	(0.077)	(0.086)	(0.082)	(0.080)	(0.222)	
15	0.077	0.075	0.212	0.198	0.183	0.208	0.108	0.132	0.143	0.138	0.094	0.133	
	(0.082)	(0.084)	(0.269)	(0.216)	(0.191)	(0.255)	(0.112)	(0.148)	(0.149)	(0.156)	(0.103)	(0.144)	
20	0.197	0.127	0.135	0.177	0.220	0.172	0.109	0.063	0.128	0.171	0.166	0.141	
	(0.205)	(0.170)	(0.152)	(0.220)	(0.214)	(0.184)	(0.137)	(0.096)	(0.161)	(0.176)	(0.177)	(0.145)	

Notes to Table: The table reports average h-day(s) returns by making use of volatility forecasts obtained from different models. The models that are used are; the Strawman random walk model (RW), univariate autoregressive model (AR), vector autoregressive model (VAR), Bayesian vector autoregressive model (BVAR), error-correction model (ECM), and the restricted dynamic factor State Space Model (SSM). All models are applied on the 5- and 7-factor IVS specifications. The average h-day(s) returns of the forecasts - obtained after using the IVS Residual Correction Model - are given in parentheses for the corresponding IVS specification. The returns are constructed following a delta-hedge portfolio and do not take transaction costs into account.

Table 6 reports the average h days returns of the best performing time series models across all indices. The table reports the returns for the 5- and 7-factor IVS specifications separately and the returns based on the volatility forecasts obtained from the corresponding time series model and IVS specification - including the IVS Residual Correction Model - are given in parentheses. The 7-factor IVS specification does not only have statistically superior out-of-sample performances, but also economically. Except for the Strawman random

walk model, the 7-factor model achieves higher average returns across all time series models and forecasting horizons h, than the 5-factor model. Moreover, the IVS Residual Correction Model enables our simulated investor to earn even higher average returns. For each forecasting horizon and index, there is a time series model that yields positive trading profits in the absence of transaction costs, but none consistently performs better than all other models. However, the table shows evidence of over-performing VAR, ECM, and SSM processes. Summarized, the 7-factor combined with the IVS Residual Correction Model outperforms hardto-beat benchmark models, both statistically and economically. In economic sense, the vector autoregressive model, error correction model, and the restricted dynamic factor State Space Model yield the highest average returns and therefore economically outperform the remaining time series models. However, including the modest transaction costs turn the trading profits negative, which is line with the efficiency of the market.

6 Conclusion

In this paper we have examined the applicability of the 7-factor model, proposed by Chalamandaris and Tsekrekos (2011), on index options. Moreover, we showed that existing IVS specifications involve heteroskedasticity, across different moneyness and maturity values, and large autocorrelations in the fitting residuals. Therefore, we proposed a model that corrects for large fitting errors, which correspond to option contracts containing extreme maturity and moneyness values. Besides describing the implied volatility of option contracts as accurate as possible, we also focused on efficiently predicting future IVS states. Previous research mainly used a standard and simple vector autoregressive time series model (of lag order one) to prove predictability. However, accurate predictions are important in many financial fields. Accordingly, we have examined many time series models and proposed a restricted dynamic factor State Space Model, applied on the 7-factor model, to study their predictive abilities.

Making use of the 7-factor IVS specification - which allows asymmetries in the moneyness dimension and includes IVS term structures - can accurately describe index option IVS shapes that are observed in reality. We have shown consistent improvements in fitting index surfaces over alternative - and more common - IVS specifications. Moreover, we found significant improvements by combining the 7-factor model with the IVS Residual Correction Model, which enables us to diminish heteroskedasticity and autocorrelations in the fitting residuals. We established that the combination of the 7-factor model and the IVS Residual Correction Model captures information that is valuable regarding describing and predicting the dynamics of option contracts. The inclusion of term structures, allowing asymmetries in the IVS specification, and including IVS Residual Correction Models appear to be important components of concoction for describing and predicting option market dynamics through time, especially for option contracts that hold extreme moneyness values and/or long expiration dates.

From an efficient predictive perspective, our contribution to IVS time-variation modeling could be of great assistance - for market participants and further researches - in constructing efficient IVS predictions. We examined the strong statistical out-of-sample performances by making use of many (advanced) time series models and found three processes that have remarkable performances and enables us to obtain accurate future volatility predictions; (i) the univariate autoregressive time series models with flexible lag orders, (ii) the error-correction models, due to non-stationarity in the time series, and (iii) the restricted dynamic factor State Space Models. Furthermore, we found prove that the latter two are efficient in a market perspective, as they produce high abnormal risk returns. However, exploiting its predictability is difficult as the constructed predictions yield negative profits in the presence of transaction costs. Nonetheless, we revealed the limitations of relevant existing literature. More efficients forecasts can be made and should instigate other researchers to be aware of alternative dynamic models' predictive abilities.

Of great interest are the dynamic factor State Space Models. Even if the State Space Model, estimated by a Kalman Filter procedure, with a vector autoregressive transition equation of lag order one outperforms alternative time series models, it might be of big interest to consider more advanced time series processes as transition equations in future researches. It is not unlikely to think that such dynamic models should improve - and consistently outperform alternative dynamic processes - in means of predictive abilities.

References

- 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(1):326–342.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy, 81(3):637–654.
- Bollen, N. P. B. and Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? The Journal Of Finance, 59(2):711–754.
- Chalamandaris, G. and Tsekrekos, A. E. (2011). How important is the term structure in implied volatility surface modeling? evidence from foreign exchange options. *Journal of International Money and Finance*, 30(4):623–640.
- Cont, R. and Fonseca, J. (2002). Dynamics of implied volatility surfaces. Quantitative finance, 2(1):45–60.
- Diebold, F. 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 and Economic Statistics, 13(3):253-263.
- Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. The Journal of Finance, 53(6):2059–2106.
- Durbin, J. and Koopman, S. (2012). Time series analysis by state space methods. 38(2).
- Fengler, M., Hardle, W., and Mammen, E. (2007). A semiparametric factor model for implied volatility surface dynamics. Journal of Financial econometrics, 5(2):189–218.
- 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.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control, 12(2):231-254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551—1580.
- Johansen, S. (1992). Determination of cointegration rank in the presence of a linear trend. Oxford Bulletin of Economics and Statistics, 54(3):383—397.
- Johansen, S. (1994). The role of the constant and linear terms in cointegration analysis of nonstationary variables. *Econometric Reviews*, 13(2):205–229.
- Johansen, S. (1995). Likelihood-based inference in cointegrated vector autoregressive models. 14(4):517-524.
- Litterman, R. B. (1979). Techniques of forecasting using vector autoregressions. Federal Reserve Bank of Minneapolis, (115):1–95.
- Nelson, C. and Siegel, A. (1987). Parsimonious modeling of yield curves. Journal of Business, 60(4):473–489.
- Peña, I., Rubio, G., and G., S. (1999). Why do we smile? on the determinants of the implied volatility function. Journal of Banking and Finance, 23(8):1151–1179.
- van der Wel, M., Ozturk, S. R., and Dijk, D. (2016). Dynamic factor models for the volatility surface. *Emerald Group Publishing Limited*, 35(25):127—174.

A Robustness Check for Lambda

For robustness check purposes, Figure 5 shows the average root mean squared error (RMSE) of the crosssectional regression of Model 2 for different lambda values. It shows a U-shape with a 10 percent difference between minimum and maximum values. Therefore, the index implied volatility surfaces are not as flat for different lambda values as currency options. Contrary to Chalamandaris and Tsekrekos (2011), we do not take the median of λ after using non-linear least squares on all parameters, which results in λ values of -0.192, -0.294, and 0.037 for SPX, DJX, and MNX, respectively, but we use the λ 's that minimize the average RMSE, which are equal to -10.013, -12.868, and -12.164 for SPX, DJX, and MNX, respectively (also shown in the figure).



Figure 5: Average RMSE for Model 2 against different values for λ , for each index

B Out-of-Sample Performances State Space Model

Table 7: Descriptive Statistics for the Index Options by Maturity and Moneyness January 2, 2003 - December 29, 2017

Horizon		4SE			М	AE	MCPDC					
h	7-VAR(1)	7-SSM	5-VAR (1)	5-SSM	7-VAR(1)	7-SSM	5-VAR(1)	5-SSM	7-VAR(1)	7-SSM	5-VAR(1)	5-SSM
S&P 500:												
1	0.0431	0.0403	0.0442	0.0429	0.0215	0.0195	0.0235	0.0223	52.82	53.57	52.85	53.42
	(0.0329)	(0.0291)	(0.0366)	(0.0350)	(0.0167)	(0.0143)	(0.0195)	(0.0179)	(53.01)	(53.56)	(52.97)	(53.54)
2	0.0436	0.0431	0.0457	0.0451	0.0223	0.0218	0.0248	0.0242	53.85	53.93	53.61	53.76
	(0.0362)	(0.0356)	(0.0400)	(0.0394)	(0.0192)	(0.0185)	(0.0218)	(0.0212)	(53.90)	(53.90)	(53.72)	(53.80)
5	0.0479	0.0468	0.0488	0.0484	0.0256	0.0255	0.0276	0.0275	54.42	54.40	54.03	54.08
	(0.0440)	(0.0428)	(0.0454)	(0.0450)	(0.0239)	(0.0238)	(0.0258)	(0.0258)	(54.51)	(54.48)	(54.10)	(54.17)
10	0.0525	0.0512	0.0537	0.0519	0.0295	0.0290	0.0310	0.0307	54.35	54.36	54.14	54.03
	(0.0534)	(0.0491)	(0.0515)	(0.0496)	(0.0284)	(0.0279)	(0.0299)	(0.0296)	(54.36)	(54.39)	(54.18)	(54.06)
15	0.0616	0.0603	0.0632	0.0567	0.0330	0.0320	0.0339	0.0332	54.37	54.38	53.94	53.94
	(0.0597)	(0.0528)	(0.0616)	(0.0548)	(0.0320)	(0.0310)	(0.0330)	(0.0324)	(54.40)	(54.41)	(53.97)	(53.97)
20	0.0711	0.0630	0.0812	0.0642	0.0359	0.0347	0.0363	0.0353	54.32	54.32	54.08	54.04
	(0.0700)	(0.0619)	(0.0801)	(0.0628)	(0.0351)	(0.0338)	(0.0356)	(0.0345)	(54.34)	(54.33)	(54.10)	(54.07)
Dow Jones:												
1	0.0537	0.0509	0.0536	0.0521	0.0248	0.0228	0.0248	0.0238	53.73	53.97	53.87	54.15
	(0.0370)	(0.0355)	(0.0385)	(0.0370)	(0.0184)	(0.0168)	(0.0193)	(0.0176)	(53.86)	(53.79)	(54.03)	(53.94)
2	0.0533	0.0532	0.0543	0.0540	0.0252	0.0248	0.0261	0.0256	53.85	54.09	54.07	54.19
	(0.0421)	(0.0419)	(0.0431)	(0.0428)	(0.0214)	(0.0210)	(0.0220)	(0.0215)	(53.80)	(54.07)	(53.89)	(54.02)
5	0.0558	0.0546	0.0575	0.0570	0.0285	0.0284	0.0290	0.0290	53.86	53.92	53.84	53.92
	(0.0506)	(0.0492)	(0.0503)	(0.0498)	(0.0263)	(0.0261)	(0.0268)	(0.0267)	(53.93)	(53.97)	(53.79)	(53.89)
10	0.0631	0.0614	0.0638	0.0615	0.0319	0.0312	0.0329	0.0325	53.72	53.60	53.83	53.70
	(0.0561)	(0.0545)	(0.0587)	(0.0563)	(0.0309)	(0.0304)	(0.0313)	(0.0310)	(53.79)	(53.66)	(53.84)	(53.70)
15	0.0789	0.0722	0.0822	0.0743	0.0363	0.0355	0.0372	0.0365	53.16	53.12	53.14	53.17
	(0.0754)	(0.0663)	(0.0785)	(0.0703)	(0.0345)	(0.0343)	(0.0358)	(0.0351)	(53.20)	(53.16)	(53.15)	(53.19)
20	0.0993	0.0798	0.1273	0.1070	0.0411	0.0393	0.0415	0.0404	53.18	53.23	53.22	53.27
	(0.0954)	(0.0781)	(0.1252)	(0.1044)	(0.0398)	(0.0383)	(0.0403)	(0.0391)	(53.23)	(53.27)	(53.25)	(53.29)
CBOE Mini-NDX:												
1	0.0386	0.0342	0.0395	0.0373	0.0183	0.0160	0.0196	0.0184	56.15	55.46	55.99	55.07
	(0.0318)	(0.0310)	(0.0338)	(0.0328)	(0.0154)	(0.0145)	(0.0173)	(0.0163)	(56.12)	(55.24)	(56.05)	(54.95)
2	0.0394	0.0359	0.0397	0.0387	0.0185	0.0173	0.0201	0.0196	55.34	55.07	55.08	54.77
	(0.0342)	(0.0338)	(0.0361)	(0.0356)	(0.0171)	(0.0166)	(0.0189)	(0.0184)	(55.22)	(54.92)	(54.98)	(54.67)
5	0.0409	0.0401	0.0428	0.0422	0.0208	0.0208	0.0228	0.0227	54.88	54.74	54.68	54.55
	(0.0395)	(0.0387)	(0.0410)	(0.0404)	(0.0208)	(0.0204)	(0.0222)	(0.0221)	(54.84)	(54.70)	(54.64)	(54.50)
10	0.0515	0.0469	0.0517	0.0478	0.0248	0.0245	0.0263	0.0260	56.65	54.61	54.60	54.55
	(0.0505)	(0.0458)	(0.0506)	(0.0465)	(0.0245)	(0.0242)	(0.0260)	(0.0257)	(54.64)	(54.60)	(54.59)	(54.54)
15	0.0801	0.0648	0.0879	0.0661	0.0291	0.0287	0.0301	0.0293	54.24	54.20	54.19	54.18
	(0.0796)	(0.0642)	(0.0872)	(0.0652)	(0.0286)	(0.0280)	(0.0298)	(0.0291)	(54.14)	(54.20)	(54.18)	(54.17)
20	0.2123	0.1488	0.2394	0.1503	0.0342	0.0325	0.0350	0.0333	54.11	54.18	54.05	54.17
	(0.2076)	(0.1470)	(0.02392)	(0.1499)	(0.0340)	(0.0322)	(0.0348)	(0.0331)	(54.10)	(54.18)	(54.04)	(54.16)

Notes to Table: The table reports three statistics: root mean squared prediction error (RMSE), mean absolute forecast error (MAE), and correct predicted direction of change statistic (MCPDC), for each of the three indices. The table reports these statistics for a vector autoregressive model with lag order one and for the restricted dynamic factor State Space Model. Note that 7-SSM denotes the State Space Model applied on the 7-factor IVS specification. The same holds for 5-SSM, but then for the 5-factor IVS specification. VAR(1) denotes the vector autoregressive process with lag order one.