Erasmus School of Economics

BSc Thesis Financial Econometrics

How do we smile?

A study on predicting index option implied volatility and capturing its time-varying non-linear surface

Raoul Martin - 407423

Supervisor: Xun Gong

Second assessor: dr. Xiao Xiao

This paper examines the feasibility of modelling and predicting the implied volatility surface of index options. Based on the seven-factor model of Chalamandaris and Tsekrekos (2011), we propose a nine-factor model which explicitly models the option contract moneyness in the corner regions. We find evidence that this model best describes the surface in terms of in-sample fit. Its forecasting performance does, however, fail to improve on previous literature. Both the seven-factor model and our proposed model significantly improve the forecasting accuracy as compared to the model of Goncalves and Guidolin (2006), which has previously been regarded as the best deterministic model for describing the implied volatility surface. We analyse the economic value of our deterministic models and find that they deliver risk-adjusted abnormal returns using a simple delta-hedge trading strategy before transaction costs. Additionally, we show that trading short-term In-The-Money (ITM) call options and Out-Of-The-Money (OTM) put options is most profitable.

Keywords: Implied volatility surface, forecasting, moneyness, delta-hedge, put-call parity

July 8, 2018

Contents

List of Tables

List of Figures

1 Introduction

Antithetical to the Black and Scholes (1973) assumption on the constant implied volatility of option contracts, in reality, it tends to vary across strike prices and maturity. Empirical evidence shows that this results in a phenomenon which we refer to as the implied volatility surface (IVS hereafter).[1] The non-linear pattern of the surface across the moneyness and term-structure of option contracts is often referred to as the "smile", due to its characteristic U-shape. The IVS may also present a skewed U-shape, which is, in turn, referred to as "smirk". Secondly, the IVS tends to vary over time while reflecting investors' beliefs and incorporating market information. It is therefore highly valuable for traders, hedgers and risk managers to understand the underlying dynamics which drive this non-linearity and time-variability. The ability to model the ever-evolving shape of the implied volatility surface could be exploited to the benefit of their respective fields.

While a vast amount of research has been dedicated in an attempt to describe the IVS, much improvement is still to be made to increase the accuracy of the models as hitherto, they are over-simplified while imposing unrealistic assumptions. Previous research has mainly focused on the general exploitability of the dynamics of index options, and some to equity options. Kim and Kim (2003) focus on short-term, close-to-money implied volatility from currency options. Pena et al. (1999) examine the determinants of implied volatility "smiles" by using data from close-to-maturity Spanish IBEX 35 index options. Dumas et al. (1998) use a simple parametric specification that establishes the relationship between time to maturity and strike prices, using S&P 500 index options. They were the first to report promising results of using this linkage and showing the opportunities of capitalising on the implied volatility over time. Goncalves and Guidolin (2006) and Bernales and Guidolin (2014) use parametric specifications from Dumas et al. (1998) and Pena et al. (1999) in order to find a cross-sectional description of the S&P 500 daily implied volatility surface. More specifically, Bernales and Guidolin (2014) investigate predictable patterns and a dynamic linkage of equity and S&P 500 index options. The extent to which the non-linearity is captured within the models of these authors is, however, still limited. More sophisticated manners to approximate the IVS shape may prove to be useful in obtaining better forecasts and consequently higher returns.

Chalamandaris and Tsekrekos (2011) show that their seven-factor model specification which accounts for asymmetry in short and long-term maturities consistently outperforms simpler models in their ability to describe the surface.[2] Their model is based on the exponential components framework of Nelson and Siegel (1987). Diebold et al. (2008) already used this framework in their analysis for a dynamic approach to modelling global yield curve dynamics and interactions of government bonds. By explicitly modelling the term structure of foreign exchange option contracts, Chalamandaris and Tsekrekos (2011) have significantly improved on existing literature. Although the authors succeed at capturing the U-shape of the term-structure, the moneyness dimension remains under-investigated. It is, therefore, of the essence to walk this untrodden path in light of further improvements on capturing the implied volatility surface. Consequently, this leads to the question of whether we can develop a model that augments existing descriptions of the time-varying non-linear implied volatility surfaces by explicitly modelling the corner regions of both moneyness and maturity dimensions.

This paper follows the works of Chalamandaris and Tsekrekos (2011) and extends their analysis by proposing an alternative approach to their model using index option data. More specifically, we examine whether we can improve current published research on the matter in terms of predictability, using more sophisticated modelling of the moneyness dimension of implied volatility surfaces. Our proposed model is additive to the model of Chalamandaris and Tsekrekos (2011), which focuses on capturing the "smile/smirk" shape of the option term structure. The nine-factor model that we propose namely focuses on an enhancement of the moneyness structure which may better capture the non-linear extremities in the corresponding dimension of the IVS shape. Additionally, we correct for our still relatively simple model by explicitly modelling a lagged error term, while distinguishing between put options and call options. This inclusion adequately corrects for the non-white noise residuals and substantially increases our predictive accuracy.

We find that both our nine-factor deterministic model and the seven-factor model proposed by Chalamandaris and Tsekrekos (2011) significantly improve the predictive accuracy and general fit of the index option implied volatility surface as compared to the five-factor model of Goncalves and Guidolin (2006). Our model that focuses on both moneyness and maturity performs best in the in-sample description of our data. It does, however, produce neither significantly superior nor inferior forecasting accuracy results compared to the model of Chalamandaris and Tsekrekos (2011). Furthermore, we obtain significant risk-adjusted abnormal returns across all our models before transaction costs. Although imposing transaction costs eradicates these returns, our simple trading strategy serves as producing lower-bound returns as practitioners may have access to more professional techniques and strategies.

We contribute to two strands of literature. First, we provide an innovative perspective on modelling the implied volatility surface. Although we do not significantly improve on the forecasting accuracy of existing research, our model provides a remarkable in-sample fit of the IVS. We believe that further research may provide more elegant manners in modelling the moneyness dimension of the option contracts in light of forecasting efficiency and accuracy. Secondly, we show strong dissimilarities in trading returns across both moneyness and maturity dimensions. We believe traders

[1] See Canina and Figlewski (1993), Heynen et al. (1994), Xu and Taylor (1994) and Campa and Chang (1995) for the evidence of the non-linear structure of the implied volatility surface across moneyness and time to maturity

[2] They have obtained inspiration from the works of Dumas et al. (1998), Pena et al. (1999), Kim and Kim (2003), Goncalves and Guidolin (2006) and Diebold et al. (2008).

may find these results highly beneficial in the construction of their own, more sophisticated, trading algorithms.

This paper is constructed as follows. First, we analyse the main properties of our options trading data and discuss their descriptive statistics in Section 2. Secondly, Section 3 motivates and explains the used methods. Thirdly, the implementation of these methods and results are reported and discussed in section 4. Finally, we conclude in section 5.

2 Data

For this research, we use data on daily index option prices extracted from the OptionMetrics database, covering the period between January 01, 2003, and December 31, 2017. This time frame is chosen such that it includes periods with and without financial turmoil. The dataset consists of four index option tickers, namely the Dow Jones industrial average (DJX), Nasdaq (NDX), Standard and Poor's 500 (SPX), and the Russell 2000 index (RUT) in Panels (a)-(d), respectively. Option specific information includes implied volatility, volume, strike price, bid and ask price, delta, and the unique contract identification number.

Figure 1 shows plots of the surface of each of the four datasets on a given day. We clearly identify a "smile" shape in the moneyness dimension for the Dow Jones and Nasdaq index options. The S&P 500 and Russell 2000 index options exhibit a skewed structure in this dimension for the given date, which we refer to as a "smirk" shape. In the maturity dimension, we observe a relatively less pronounced non-linear shape contrary to the currency options investigated by Chalamandaris and Tsekrekos (2011). Finally, we notice an overall level shift of the IVS during periods of turmoil, such as during the financial crisis around 2008. This level shift is clearly visible in the IVS plot for NDX options in Figure 1. We also plot the daily standard deviation of the implied volatility surface, which is shown in Figure 6 in Appendix A. These plots display the variability of the implied volatility surfaces, which tends to be more severe for higher

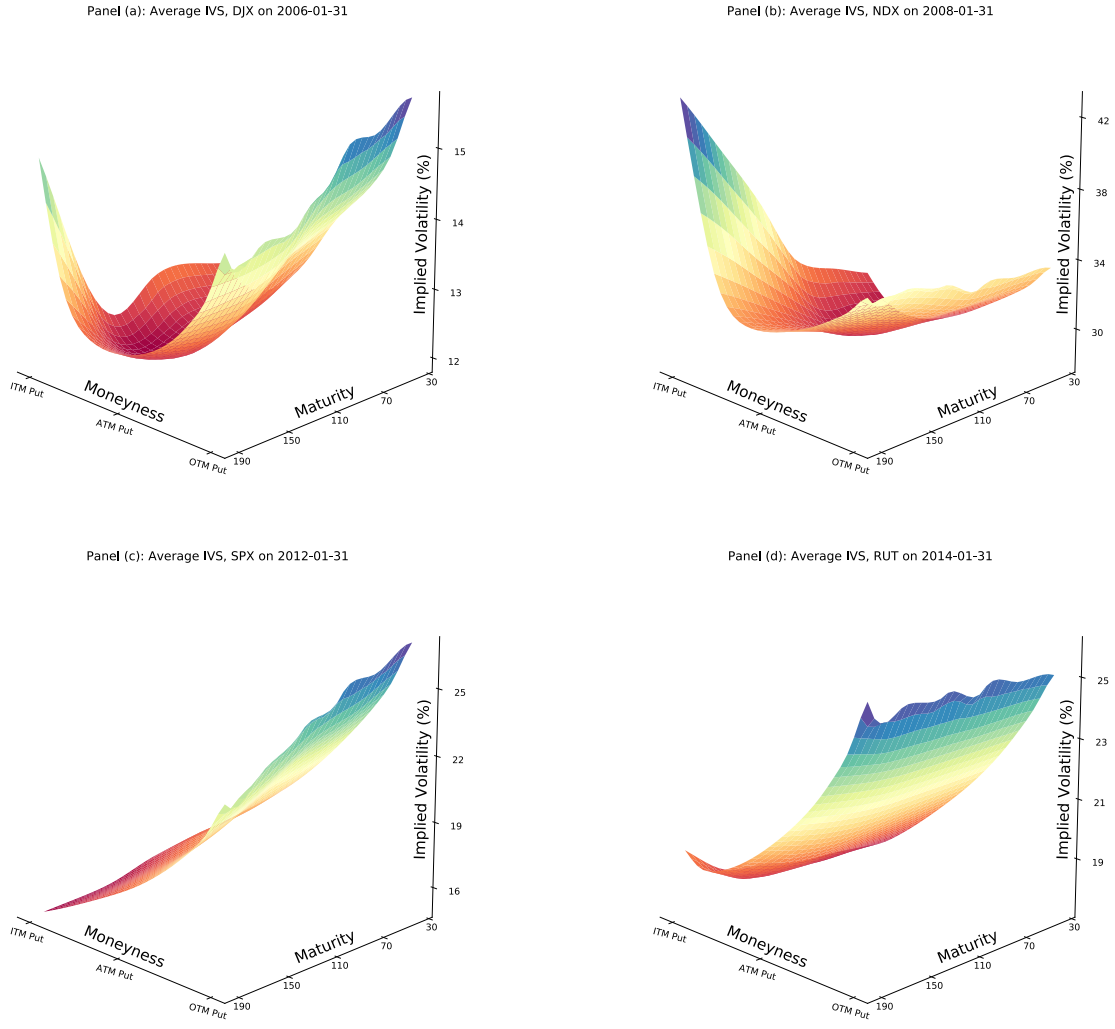

Figure 1: This figure shows four plots of the implied volatility surface (IVS) for each of our index option datasets on a given date. From Panel (a) to Panel (d): Dow Jones (DJX), Nasdaq (NDX), Standard and Poor's 500 (SPX), and Russell 2000 (RUT). The IVS plots are given on a different date for each index. In the same order: 2006-01-31, 2008-01-31, 2012-01-31, and 2014-01-31.

Table 1: Summary Statistics of Implied Volatilities across Moneyness and Maturities

	Short-term ($6 < \tau \leq 120$)			Medium-term ($120 < \tau \leq 240$)			Long-term ($240 > \tau$)		
Moneyness	Av. IV (%)	Std.dev. IV (%)	Av. Trading Volume	Av. IV (%)	Std.dev. IV (%)	Av. Trading Volume	Av. IV (%)	Std.dev. IV (%)	Av. Trading Volume
Panel (a): DJX Options									
Deep OTM Call/Deep ITM Put	27.07	18.07	99.70	22.12	10.54	103.27	19.02	7.11	75.08
Call OTM/ITM Put	16.21	8.31	199.85	17.31	6.67	117.60	18.31	5.86	103.03
Left ATM Call/ATM Put	17.38	8.56	292.71	18.97	6.99	102.66	19.46	5.91	134.81
Right ATM Call/ATM Put	18.62	9.01	305.32	20.40	7.35	119.60	20.96	6.20	145.46
ITM Call/OTM Put	21.55	9.85	183.01	23.87	8.11	112.64	24.41	6.92	78.12
Deep ITM Call/ Deep OTM Put	31.43	14.22	161.85	31.15	9.70	103.78	29.70	7.68	72.33
Panel (b): NDX Options									
Deep OTM Call/Deep ITM Put	19.80	12.76	176.64	18.10	7.43	51.36	16.19	5.57	35.13
Call OTM/ITM Put	16.31	6.37	130.15	18.10	5.47	46.89	17.91	4.02	29.23
Left ATM Call/ATM Put	17.41	6.38	125.67	19.53	5.26	37.98	19.72	3.64	23.13
Right ATM Call/ATM Put	18.33	6.58	137.86	20.40	5.15	53.18	21.33	3.78	30.24
ITM Call/OTM Put	20.76	7.52	113.45	23.16	6.11	45.09	24.30	4.66	25.60
Deep ITM Call/ Deep OTM Put	29.13	11.66	158.80	30.25	7.66	47.55	30.55	5.89	34.68
Panel (c): SPX Options									
Deep OTM Call/Deep ITM Put	15.25	10.45	623.81	16.49	9.14	403.26	15.76	6.31	273.55
Call OTM/ITM Put	13.56	6.82	811.62	15.55	6.03	605.77	16.33	4.97	396.76
Left ATM Call/ATM Put	15.18	7.22	1173.26	17.22	6.03	592.88	18.15	4.76	367.86
Right ATM Call/ATM Put	16.36	7.47	1657.90	18.87	6.38	833.25	19.70	4.93	493.15
ITM Call/OTM Put	18.99	7.89	825.23	21.98	6.74	744.72	23.05	5.46	421.04
Deep ITM Call/ Deep OTM Put	26.40	10.29	657.99	29.15	7.72	463.52	30.94	8.77	421.38
Panel (d): RUT Options									
Deep OTM Call/Deep ITM Put	21.90	13.04	314.61	19.60	8.78	27.56	18.76	6.93	23.84
Call OTM/ITM Put	18.76	7.52	218.45	19.69	6.37	56.40	20.26	5.77	48.31
Left ATM Call/ATM Put	20.23	7.72	246.70	21.50	6.34	59.36	22.37	5.59	91.23
Right ATM Call/ATM Put	21.45	8.02	300.24	23.15	6.67	76.06	24.09	5.63	74.00
ITM Call/OTM Put	24.24	8.70	199.59	26.51	7.42	73.35	27.54	6.23	85.70
Deep ITM Call/ Deep OTM Put	31.33	10.90	233.48	33.00	9.00	39.69	34.99	8.90	35.93

Notes: The table contains summary statistics for implied volatilities across moneyness (in terms of the option delta) and time-to-maturity (calendar days to the expiration). The first column determines the moneyness for call and put options: (Deep) Out-Of-The-Money (OTM), At-The-Money (ATM), and (Deep) In-the-Money (ITM). The moneyness measure is divided into six regions: $-50 < \Delta \leq -37.5$*;* $-37.5 < \Delta \leq -12.5$*;* $-12.5 < \Delta \leq 0$*;* $0 < \Delta \leq 12.5$*;* $12.5 < \Delta \leq 37.5$*;* $37.5 < \Delta \leq 50$*. The following three columns correspond to short-term maturities between 6 and 120 days. Medium-term maturities are those between 120 and 240 days, and long0term maturities over 240 days. Panels (a), (b), (c), and (d) report statistics for the index options of respectively DJX, NDX, SPX, and RUT. IV is the implied volatility, as well as the strike price, and the underlying asset price. The table presents trading means, standard deviations, and volumes. The trading volume is defined as the average volume over all actual trades with specific characteristics (given by the moneyness and the time-to-maturity). The data cover the period between January 1, 2003, and December 31, 2017.*

volatilities. As corner regions often have relatively high implied volatilities, the fact that their variability is also higher results in comparatively greater uncertainty. It may thus be more challenging to predict the implied volatility of options that fall into these outer regions.

We apply four exclusionary criteria to our data, akin to Bernales and Guidolin (2014). Observations which are not likely to represent traded prices in well-functioning and liquid option markets are consequently omitted. We first eliminate option data that violate basic no-arbitrage conditions by setting upper and lower bounds for call and put prices. Secondly, we exclude observations with less than one week, as their prices usually contain little information regarding the implied volatility surface (Dumas et al., 1998). Contrary to Bernales and Guidolin (2014) but similar to Chalamandaris and Tsekrekos (2011), we do not impose an upper bound on the maturity of the option contracts as we specifically aim to investigate the term structure in our cross-sectional model. The lower bound in maturity is seven days. Thirdly, we drop option contracts with prices lower than 3/8 to bypass the effect of price discreteness.[3] For this very reason, we finally exclude contracts with absolute deltas above 0.98 or below 0.02.

Analogously to Chalamandaris and Tsekrekos (2011) and Bollen and Whaley (2004), we use the delta of each contract Δ_i^{raw} as moneyness measure and transform it as follows:

$$\Delta_i = (\Delta_i^{raw} - 0.50) * 100, \tag{1}$$

where Δ_i is the transformed delta of option contract i and lies between [-50,50]. Bollen and Whaley (2004) explain that, usually, moneyness is measured as the relative difference between the forward price of the underlying asset and the exercise price of the option contract. They show that the volatility rate of the underlying asset strongly affects the probability of the option contract to expire In-The-Money. Using the typical moneyness metric may, therefore, be problematic. The adjusted moneyness metric Δ_i as described by the above equation is sensitive to the volatility of the underlying asset as well as the time to expiration of the option contract. Moreover, it has several advantages as depicted by Sheldon (1994). First, we avoid the need for interpolating strikes of different liquidity as the metric is based on actual contracts. Secondly, the shape of the IVS is conserved, and thirdly, it reduces the flattening effect.[4]

Summary statistics of our resulting data is given in Table 1, which is separated into six moneyness regions and

[3] As options prices are adjusted with a pre-specified 'tick-size', one tick in any direction may influence the IVS of an option with a relatively low price too strongly. This is called the effect of price discreteness.

[4] While it is known that volatility "smiles/skews" flatten at longer maturities, Δ_i produces a scaling which is independent of the volatility hence reducing this effect (Yang et al., 2010).

three maturity regions for each dataset. The moneyness (Δ) regions are: $-50 < \Delta \leq -37.5$; $-37.5 < \Delta \leq -12.5$; $-12.5 < \Delta \leq 0$; $0 < \Delta \leq 12.5$; $12.5 < \Delta \leq 37.5$; $37.5 < \Delta \leq 50$. The option expiration times are divided into the regions of 6 up to 120 days, 120 up to 240 days, and over 240 days. These regions correspond to respectively short-, medium-, and long-term maturities.

The given statistics are average implied volatility (IV), its average standard deviation as well as the average trading volume. Days in which no trade is executed are excluded in calculating the latter metric.

From this table, we immediately notice the non-linearity in average implied volatility over the moneyness region. The "smile" shape is most visible for the short-term Dow Jones index options. The other index options display on average an upwards sloping "skew" towards Deep In-The-Money (ITM) calls and Out-Of-The-Money (OTM) puts.

Also, while the variation of the average implied volatility is observed to be less pronounced in terms of maturity, the non-linearity is still present and of importance to our investigation. This is in line with our visual observation from Figure 1.

Interestingly, the standard deviation of the IV seems positively correlated with an increasing average IV over all regions, while observing the opposite for average trading volumes. More specifically, the relatively large standard deviations in the corner regions are most apparent for short-term maturities. Andersen et al. (2017) explain that this phenomenon occurs due to a larger jump-risk for short-term option contracts in moneyness corner regions.[5]

Clearly, the distribution of the implied volatility over both maturity and moneyness dimensions is highly non-linear and ought not to be ignored. Its complexity in shape and time-variation inclines to be described through a simplified parametric model.

3 Methodology

3.1 Modelling the Implied Volatility Surface

A method to describe the volatility surface in a model that broadly captures the key characteristics of the non-linearity of the IVS shape is by application of a cross-sectional regression with a select set of parameters. Both Bernales and Guidolin (2014) and Chalamandaris and Tsekrekos (2011) follow a five-factor specification proposed by Goncalves and Guidolin (2006). The former authors have concluded that this model performs best in describing the index option IVS. The resulting deterministic model, which is referred to as the five-factor GLS model for the implied volatility for contract i is given by the following equation:

$$\begin{aligned}\sigma_{i,t} &= \beta_{0,t} + \beta_{1,t} M_{i,t} + \beta_{2,t}\tau_{i,t} + \beta_3 M_{i,t}^2 + \\ &\beta_{4,t}(M_{i,t} \cdot \tau_{i,t}) + \epsilon_{i,t},\end{aligned} \tag{2}$$

where M_i is the time-adjusted moneyness, τ_i is the time-to-maturity, and ϵ_i the residual term. $i = 1, \ldots, N$, with N the number of option contracts available at day t, $t = 1, \ldots, 3776$. Note that, contrary to Chalamandaris and Tsekrekos (2011), we do not take the natural logarithm of the implied volatilities in order to allow for model comparisons. Similar to Bollen and Whaley (2004), M_i is defined by the delta of the corresponding option contract as motivated in Section 2.

As a result, β_0 of our deterministic five-factor GLS model given by Equation 2 is defined as the level coefficient, β_1 captures the moneyness ("smile/skew") slope of the IVS, and β_2 reflects the maturity (term-structure) slope. The curvature of the IVS in the moneyness dimension is captured by β_3. Finally, β_4 describes possible interactions between moneyness and time-to-maturity dimensions.

Similar to Chalamandaris and Tsekrekos (2011) and Bernales and Guidolin (2014), we apply recursive cross-sectional generalised least squares (GLS) to the deterministic model given by Equation (2) instead of ordinary least squares (OLS) to gain asymptotic efficiency.

A common feature of the IVS of (index) options, as established in the previous section, is that it takes a non-linear shape in both the moneyness and maturity directions. As Equation 2 only includes a linear term for maturity, one may wonder whether the non-linearity in the term structure is sufficiently captured. Chalamandaris and Tsekrekos (2011) indeed show that this relatively simple model does not suffice in the sense that the term structure conveys valuable information for describing the implied volatility of foreign exchange option contracts. Since they have found evidence of better predictability, we apply their seven-factor deterministic model to our data set of index options in order to exploit this information as well. The model, which we refer to as the seven-factor GLS model is given by the following equation:

$$\begin{aligned}\sigma_{i,t} &= \beta_{1,t} I_{1,i,t} + \beta_2 I_{2,i,t} + \beta_{3,t} I_{3,i,t} + \beta_{4,t} I_{4,i,t} + \\ &\beta_{5,t} I_{5,i,t} + \beta_{6,t} I_{6,i,t} + \beta_{7,t} I_{7,i,t} + \varepsilon_{i,t},\end{aligned} \tag{3}$$

where $i = 1, \ldots, N$, with N the number of option contracts available at day t, $t = 1, \ldots, 3776$.

$$\begin{aligned}
I_{1,i,t} &= 1 && Level \\
I_{2,i,t} &= 1_{\{\Delta_{i,t}>0\}}\Delta_{i,t}^2 && Right\ \text{“}smile\text{”} \\
I_{3,i,t} &= 1_{\{\Delta_{i,t}<0\}}\Delta_{i,t}^2 && Left\ \text{“}smile\text{”} \\
I_{4,i,t} &= \frac{1-e^{-\lambda\tau_{i,t}}}{\lambda\tau_{i,t}} && Short-term\ \ structure \\
I_{5,i,t} &= \frac{1-e^{-\lambda\tau_{i,t}}}{\lambda\tau_{i,t}} - e^{-\lambda\tau_{i,t}} && Medium-term\ \ structure \\
I_{6,i,t} &= 1_{\{\Delta_{i,t}>0\}}\Delta_{i,t}\tau_{i,t} && Right\ \text{“}smile\text{”}\ attenuation \\
I_{7,i,t} &= 1_{\{\Delta_{i,t}<0\}}\Delta_{i,t}\tau_{i,t} && Left\ \text{“}smile\text{”}\ attenuation
\end{aligned} \tag{4}$$

The seven indicators can be interpreted in a natural manner. $I_{1,i,t}$ corresponds to the constant level factor and thus $\beta_{1,t}$ describes the mean level of the IVS of contract i at time t. The following two factors, $I_{2,i,t}$ and $I_{3,i,t}$, describe the "smile" shape of the moneyness measure. Their corresponding coefficients, therefore, capture a potential asymmetry within this shape between contracts with a negative or positive delta.

Indicators $I_{4,i,t}$ and $I_{5,i,t}$ account for the non-linearity in the maturity of option contracts. They describe

[5] Jump-risk is the risk of sudden but infrequent movements of large magnitude in price (Yan, 2011).

the non-linear shape within the IVS through short- and medium-term structure asymmetry. The functions that determine the values of these indicators are calculated analogously to Diebold and Li (2006) who apply a factorization of the Nelson and Siegel (1987) parsimonious term structure model. This factorization proved to be useful in forecasting the yield curve of government bonds. In this parameterization, parameter λ determines the rate of the exponential decay rate of maturities.

Finally, $I_{6,i,t}$ and $I_{7,i,t}$ account for the flattening effect of the IVS "smile" shape as the option contract approaches maturity, as described by Chalamandaris and Tsekrekos (2011). These indicators allow for asymmetry between In-The-Money calls and puts in this effect.

Similarly to Diebold and Li (2006), we fix λ to be constant. However, as Chalamandaris and Tsekrekos (2011) show that the optimal λ varies over different datasets, we apply a two-step estimation approach. We first estimate λ via non-linear least squares after which we apply a recursive cross-sectional GLS regression to Equation 3.

Most importantly, our aim is to improve the extent to which our model captures the "smile" shape in the moneyness dimension. Specifically, we should capture the strong increase in the implied volatility for tail regions as it has been established in Section 2 that they strongly deviate from centre regions in terms of both the level of the surface as well as its variability. We consider contracts with adjusted absolute moneyness higher than 37.5 to be Deep In- and Out-Of-The-Money and separate their effect on the IVS from more At-The-Money (ATM) contracts. This results in the following equation that is referred to as the nine-factor GLS model:

$$\begin{aligned}\sigma_{i,t} = {} & \beta_{1,t}I_{1,i,t} + \beta_{2a,t}I_{2a,i,t} + \beta_{2b,t}I_{2b,i,t} + \beta_{3a,t}I_{3a,i,t} + \\ & \beta_{3b,t}I_{3b,i,t} + \beta_{4,t}I_{4,i,t} + \beta_{5,t}I_{5,i,t} + \beta_{6,t}I_{6,i,t} + \beta_{7,t}I_{7,i,t} + \varsigma_{i,t},\end{aligned} \tag{5}$$

where $i = 1, \ldots, N$, with N the number of option contracts available at day t, $t = 1, \ldots, 3776$.

$$\begin{aligned} I_{2a,i,t} &= 1_{\{0<\Delta_{i,t}<37.5\}}\Delta_{i,t}^2 && Right\ \ ATM\ \ \text{“}smile\text{”} \\ I_{2b,i,t} &= 1_{\{\Delta_{i,t}\geq 37.5\}}\Delta_{i,t}^2 && Right\ \ OTM\ \ \text{“}smile\text{”} \\ I_{3a,i,t} &= 1_{\{-37.5<\Delta_{i,t}<0\}}\Delta_{i,t}^2 && Left\ \ ATM\ \ \text{“}smile\text{”} \\ I_{3b,i,t} &= 1_{\{\Delta_{i,t}\leq -37.5\}}\Delta_{i,t}^2 && Left\ \ OTM\ \ \text{“}smile\text{”}, \end{aligned} \tag{6}$$

where $I_{2a,i,t}$ and $I_{3a,i,t}$ describe the "smile" shape of the moneyness measure for more ATM contracts. $I_{2b,i,t}$ and $I_{3b,i,t}$ describe the "smile" shape of the moneyness measure for Deep OTM/ITM contracts. Their corresponding coefficients, therefore, capture a possible asymmetry within this shape between contracts with a negative or positive delta as well as being Deep In-and Out-Of-The-Money or more ATM. It is important to note that the names given to these indicators are for simplicity and serve as a reference. For example, a call option with a positive delta larger than 37.5 is still Deep ITM. The name of the indicator should not suggest otherwise.

The interpretation of the remaining five indicators and their respective coefficients are identical to Equation (3).[6] As our proposed model strongly resembles the seven-factor model, we expect relatively similar results. However, we believe that our model should be able to better capture the variability corner regions within the moneyness regions. The proposed set of factors may thus prove to be a sensible addition.

3.1.1 Flattening the Dynamically Dependent Residuals

Although the seven-factor GLS model has been proven to produce better forecasting results than the five-factor model according to Chalamandaris and Tsekrekos (2011), we claim that neither model fully captures the non-linear dynamics of the IVS. For this reason, we adjust our forecasts using a threshold-AR(1) model for the residuals obtained from the cross-sectional deterministic regressions (2)-(5). Contributing to existing literature, we set the threshold to distinguish between residuals for respectively put and call contracts, as they may render asymmetric residual distributions. The threshold-AR(1) equation is estimated using the following equation:

$$\upsilon_{i,t}^{w} = \delta^{\upsilon,w}\upsilon_{i,t-1}^{w} + \eta_{i,t}^{w} \qquad \upsilon \in \{\epsilon, \varepsilon, \varsigma\}, w \in \{put, call\}, \tag{7}$$

where $i = 1, \ldots, N$, with N the number of option contracts available at day t, $t = 1, \ldots, 3776$. The error term from the cross-sectional regression for option contract i at day t on its respective deterministic model is denoted by $\upsilon_{i,t}$.

Using our estimated residual coefficients, we correct our in-sample fit of the IVS by including the lagged error term in the following equation:

$$\begin{aligned}\hat{\sigma}_{i,t} = {} & \sum_{j_m=1}^{J_m} \hat{\beta}_{j_m,t} I_{j_m,i,t} + \hat{\delta}^{\upsilon,w}\upsilon_{i,t-1} \\ & m \in \{(2), (3), (5)\}, \upsilon \in \{\varepsilon, \varsigma\} \ \ and \ \ w \in \{put, call\},\end{aligned} \tag{8}$$

where j_m is the factor index of the J_m-factor deterministic GLS model denoted by Equation m, for example for $j_{(3)} = 1, \ldots, 7$. Moreover, $t = 1, \ldots, 3776$ and $\hat{\delta}^{\upsilon,w}$ is estimated by the threshold-AR(1) model given by Equation (7).

The estimated coefficients are subsequently used in our forecasting procedures, which will be described in the following section.

3.2 Modelling the Dynamics of Index Option IVS

There exists a vast amount of literature that proves strong predictability of the IVS shape using simplified deterministic models, as mentioned in Section 1. Most relevant literature has focused on relatively simple dynamic models as these often seem to outperform more sophisticated models. Moreover, it is of importance to fully grasp the extent to which our deterministic factors can indeed capture the underlying dynamics of the IVS shape. For this reason, we implement relatively straightforward models to establish the linkage between our estimated factor coefficients and the implied volatility of index option contracts for forecasting purposes.

The first dynamic model to be adopted is a univariate AR(p) representation of the coefficients extracted from the

[6] Note that we have investigated another approach where the five-factor model is only adjusted in terms of moneyness in a similar fashion as in the nine-factor model, resulting in an eight-factor model. As the results generally did not outperform our other specifications, we have excluded it from this report.

cross-sectional GLS regressions. The autoregressive model is given by the following equation:

$$\hat{\beta}_{j_m,t} = c_{j_m} + \sum_{q=1}^{p} \phi_{j_m,q}\hat{\beta}_{j_m,t-q} + u_{j_m,t} \quad j_m = 1,\dots,J_m \;\; and \;\; m \in \{(2),(3),(5)\} \tag{9}$$

where $t = 1,\dots,3776$ and j_m is as previously defined. This model, which we denote as $\hat{\beta} - AR$, is among the best performing in forecasting the implied volatility of foreign exchange rates options and could too prove to be useful in forecasting the IVS of index options (Chalamandaris and Tsekrekos, 2011). Contrary to them, we apply the Bayes-Schwarz Information Criterion (BIC) to determine the number of lags to be included in the model. To prevent over-fitting our parameters, we allow for a maximum number of three lags.

The second dynamic model that we impose will be denoted as $\hat{\beta} - VAR$ and corresponds to a vector autoregressive model of order p for the estimated coefficients from the deterministic models. The equation is structured as follows:

$$\hat{\beta}_{m,t} = \mathbf{c} + \sum_{q=1}^{p} \mathbf{\Phi}_{m,q}\hat{\beta}_{m,t-q} + \mathbf{u}_{m,t} \qquad m \in \{(2),(3),(5)\}, \tag{10}$$

for $t = 1,\dots,3776$ and m denoting the respective deterministic model. This model allows for interdependence between the subsequent factors. Here, we also determine the number of included lags using the BIC, with a maximum of three.

For an alternative approach, which may render valuable insights, we also investigate whether including a market sentiment factor to our model significantly improves our results in terms of forecasting accuracy. More specifically, we investigate whether deviations from put-call parity help improve forecasting results.

In theory, every pair of puts and calls of the same underlying asset, strike price, and expiration date should be equally priced and thus have the same implied volatility at any given day. In practice, however, deviations from the equality occur due to market imperfections and investor preference towards a certain contract. One could argue that these imperfections, which are mostly short-lived, are caused by investors' beliefs and expectations towards the value of the underlying asset. Therefore, we view any deviation from the put-call parity as in indicator for market sentiment. A natural derivation from this logic is that lagged deviations could very well influence implied volatilities of options in the future.

It should be noted that for American-style option contracts, put-call parity does not hold, as they allow for early exercise. Nevertheless, we claim that this style of option contracts still provide useful information in grasping the time-variation of their implied volatilities. Cremers and Weinbaum (2010) have previously shown that deviations from put-call parity for American-style option contracts contain relevant predictable power for the underlying equity return. Since put-call parity does not hold for this style of options, it should be evident that we do not view any deviation as an unexploited arbitrage opportunity. Hence, to avoid ambiguity, we refer to deviations instead of violations from put-call parity.

Using interpolated option price data obtained from OptionMetrics, we calculate the deviations in put-call parity similarly to Amin et al. (2004) as follows:

$$VS_{i,t} = IV^{call}_{i,t} - IV^{put}_{i,t}, \tag{11}$$

where i refers to pairs of call and put option contracts and $IV^{o}_{i,t}$ ($o \in \{put, call\}$) denotes the Black and Scholes (1973) implied volatility. If a deviation is positive, we may conclude that investors value calls higher than puts which could be due to the expectation of a price increase in the underlying asset and vice versa.

Thereafter, we apply principal component analysis to reduce the dimensions and extract maximum variation from the 100 contracts. This results in the 3776x1 eigenvector β^{DPCP}, which will henceforth be referred to as DPCP (deviation from put-call parity).

The resulting model where we include DPCP to the $\hat{\beta} - VAR$ model is consequently referred to as $\hat{\beta} - VARX$ and is given by:

$$\hat{\beta}_{m,t} = \mathbf{c} + \sum_{q=1}^{p} \mathbf{\Phi}_{m,q}\hat{\beta}_{m,t-q} + \sum_{r=1}^{s} \mathbf{\Theta}_{r}\beta^{DPCP}_{t-r} + \mathbf{z}_{m,t} \quad m \in \{(2),(3),(5)\}, \tag{12}$$

where the parameters are defined as in Equation (10) including a maximum of three lags, determined by the BIC.

Finally, we adopt a Bayesian vector autoregressive model, which we refer to as $\hat{\beta} - BVAR$. The equation is as follows:

$$\hat{\beta}_{m,t,B} = \mathbf{c} + \sum_{q=1}^{p} \mathbf{\Phi}_{m,q,B}\hat{\beta}_{m,t-q} + \mathbf{u}_{m,t} \qquad m \in \{(2),(3),(5)\}, \tag{13}$$

where the parameters are defined as in Equation (10) and we use the prior proposed by Doan et al. (1984).

3.2.1 The Statistical Value of Predictability

Using the provided models, it is of importance to assess the predictability of the index option IVS. Each of the models given by Equations (9)-(12) are estimated by ordinary least squares (OLS) using a rolling window of 125 trading days for forecast horizons of 1, 3, 5, and 10 days. The predicted implied volatility of option contract i for forecast horizon h is subsequently calculated, as denoted by the following equation:

$$\hat{\sigma}_{i,t+h} = \sum_{j_m=1}^{J_m} \hat{\beta}_{j_m,t+h} I_{j_m,i,t} + \hat{\delta}^{\upsilon,w}\upsilon_{i,t} \quad m \in \{(2),(3),(5)\}, \upsilon \in \{\varepsilon, \varsigma\} \;\; and \;\; w \in \{put, call\}, \tag{14}$$

where the subscripts are as previously defined, $\upsilon_{i,t}$ is the error term from the cross-sectional regression for option contract i on its respective deterministic model, and $\hat{\delta}^{\upsilon,w}$ is estimated by the threshold-AR(1) model given by Equation (7) using a rolling window of 125 trading days.

To thoroughly assess the predictive power of the proposed models, we compare their forecasting performance to two benchmark models. The first benchmark we employ is a "Strawman" random walk model as used by Christoffersen et al. (2013) and Diebold and Li (2006). This model is a naive approach in the sense that tomorrow's best prediction is that of today in terms of factor coefficients and is denoted by:

$$\begin{aligned} &\hat{\beta}_{m,t} = \hat{\beta}_{m,t-1} \\ &m \in \{(2),(3),(5)\}. \end{aligned} \tag{15}$$

The second benchmark is a pure random walk (RW) model of the implied volatility:

$$\hat{\sigma}_{i,t+h} = \sigma_{i,t}, \tag{16}$$

where i, t, and h are as previously defined.

The statistical value of predictability of our models is evaluated using three metrics. The first two metrics are the root-mean-square error (RMSE) and the mean absolute error (MAE) of the predicted implied volatilities calculated by Equation (14). The third metric is the mean correct prediction of the direction of change (MCPDC), which essentially measures the percentage of correct predictions in terms of the direction of change in the implied volatility for each option contract between time t and $t-1$.[7]

3.2.2 The Economic Value of Predictability

To offer a complete analysis, we evaluate the economic implications of our models' forecasting abilities. In order to achieve this objective, we apply a basic trading strategy, which is similar to Bernales and Guidolin (2014). The trading strategy is based on the rule that when our models forecast a decrease (increase) of the implied volatility of option contract i at time $t+1$, we sell (purchase) that contract at the current day to potentially profit from this change in its IV.

We proceed by constructing a trading portfolio based on delta-hedged option strategies as we are not exposed to any dangers caused by price changes in the underlying index. Our delta-hedged positions are constructed by trading appropriate volumes of the underlying index based on the option delta.

We invest \$1000 in each delta-hedged portfolio which are re-balanced every day such that the \$1000 investment remains constant over time. Profits and losses are consequently calculated and analysed. Note that these profits are at best a lower bound for the actual profits that traders would make (before transaction costs), due to the sheer simplicity of our trading strategy.

Similar to Goncalves and Guidolin (2006), we assume today's index prices and interest rates to calculate option price forecasts, since we lack one-day ahead predictions of these. This assumption is fairly weak in our application, as our trading portfolios are completely hedged against the effects of changes in the prices of the underlying index.

In constructing the delta-hedge portfolio, let Q_t be the number of option contracts written on the same underlying index that should be traded following the trading rule introduced above. Furthermore, let V_t^{DH} be the total value of all delta-hedged positions in the portfolio on day t, which also depends on Q_t. We can, therefore, write V_t^{DH} as:

$$\begin{aligned} V_t^{D-H} = &\sum_{m\in Q_{t,+}^{call}} (C_{m,t} - S_t\Delta_{m,t}^C) + \sum_{m\in Q_{t,+}^{put}} (P_{m,t} + S_t\Delta_{m,t}^P) - \\ &\sum_{m\in Q_{t,-}^{call}} (C_{m,t} - S_t\Delta_{m,t}^C) + \sum_{m\in Q_{t,-}^{put}} (P_{m,t} + S_t\Delta_{m,t}^P), \end{aligned} \tag{17}$$

where $Q_{t,+}^{call}(Q_{t,-}^{call})$ is the subset of call contracts that are purchased (sold), $Q_{t,+}^{put}(Q_{t,-}^{put})$ is the subset of put contracts that are purchased (sold), S_t is the price of the underlying index and $\Delta_{m,t}^C(\Delta_{m,t}^P)$ is the absolute value of the call (put) option delta. When the net value of the delta-hedged portfolio is positive (i.e., $V_t^{DH} > 0$), we purchase the quantity $X_t^{DH} = \frac{\$1000}{V_t^{DH}}$ in units of the delta-hedged portfolio, for a total cost of \$1000. Consequently, the one-day net gain (G_{t+1}^{DH}) is:

$$\begin{aligned} G_{t+1}^{DH} = &X_t^{DH}\left[\sum_{m\in Q_{t,+}^{call}} ((C_m^{t+1} - S_{t+1}\Delta_{m,t}^C) - (C_{m,t} - S_t\Delta_{m,t}^C))\right] \\ &+X_t^{DH}\left[\sum_{m\in Q_{t,+}^{put}} ((P_m^{t+1} + S_{t+1}\Delta_{m,t}^P) - (P_{m,t} + S_t\Delta_{m,t}^P))\right] \\ &+X_t^{DH}\left[\sum_{m\in Q_{t,-}^{call}} (-(C_m^{t+1} - S_{t+1}\Delta_{m,t}^C) + (C_{m,t} - S_t\Delta_{m,t}^C))\right] \\ &+X_t^{DH}\left[\sum_{m\in Q_{t,-}^{put}} (-(P_m^{t+1} + S_{t+1}\Delta_{m,t}^P) + (P_{m,t} + S_t\Delta_{m,t}^P))\right] \end{aligned} \tag{18}$$

In case the net cost of the portfolio is negative (i.e., $V_t^{DH} < 0$), we sell the quantity $X_t^{DH} = \frac{\$1000}{|V_t^{DH}|}$ in units of the delta-hedged portfolio, which generates a cash inflow of \$1000. This sum is then invested at the daily risk-free rate in addition to the initially available \$1000. The net gain results in $G_{t+1}^{DH} + \$2000\cdot(e^{r_t/252} - 1)$, where G_{t+1}^{DH} is obtained from Equation (18).

The price data of the underlying indices that are used in constructing this trading strategy are obtained from OptionMetrics. The risk-free rate is the one-month Treasury bill rate at a daily frequency.

For comparative purposes, we construct benchmark portfolios in our analysis. The first benchmark is the Buy-and-Hold strategy, where we invest \$1000 daily in each of the four indices. The second is constructed by passively holding an investment of \$1000 at the risk-free interest rate, which thus represents the time value of money.

Finally, we investigate the effect of transaction costs on our average daily abnormal returns. Battalio and Schultz (2006) show that the effective transaction cost of equity options are on average 0.8 times the quoted bid-ask spread. We extrapolate this finding to index options. This means that for each trade that we make in our strategy, we impose a transaction cost of 0.8 times the bid-ask spread of that trade.

[7] $RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$; $MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t|$; $MCPDC = \frac{1}{n}\sum_{t=1}^{n} 1_{t,(sign(\sigma_{i,t+1}-\sigma_{i,t})=sign(\hat{\sigma}_{i,t+1}-\sigma_{i,t}))}$

Table 2: Results of Cross-Sectional Regressions on Deterministic Model Specifications

5-factor GLS	Av. R^2	Max. R^2	Min. R^2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
DJX	71.70	97.82	25.59	99.84	93.96	84.67	99.92	80.11
NDX	73.24	98.22	15.67	100.00	95.68	84.16	99.95	83.85
SPX	72.42	97.47	15.25	100.00	97.91	88.22	100.00	88.00
RUT	73.67	98.86	4.27	99.97	91.60	77.97	99.39	81.70

7-factor GLS	Avg. R^2	Max. R^2	Min. R^2	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
DJX	77.03	97.89	38.15	100.00	100.00	67.19	91.84	76.46	61.47	92.11
NDX	78.72	98.64	19.88	100.00	99.89	76.67	89.49	82.44	67.69	91.21
SPX	76.66	98.27	18.63	100.00	100.00	78.71	92.80	79.82	71.29	96.56
RUT	77.87	98.91	5.23	98.81	98.12	77.15	80.32	70.47	56.14	91.10

9-factor GLS	Avg. R^2	Max. R^2	Min. R^2	$\hat{\beta}_1$	$\hat{\beta}_{2a}$	$\hat{\beta}_{2b}$	$\hat{\beta}_{3a}$	$\hat{\beta}_{3b}$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
DJX	77.77	97.99	41.31	100.00	99.84	100.00	47.72	67.80	90.55	75.45	61.12	88.74
NDX	79.17	98.75	28.84	100.00	98.60	99.95	68.06	78.89	88.29	80.96	66.26	88.90
SPX	77.31	98.49	20.20	100.00	99.92	100.00	63.67	79.05	91.45	79.50	72.09	92.53
RUT	78.38	98.92	10.50	98.73	95.84	98.17	57.73	77.41	78.36	69.76	55.22	86.31

Notes: The table reports the average, maximum, and minimum R^2 obtained by applying generalised least squares (GLS) to our 5-, 7-, and 9-factor deterministic models given by Equations (2)*,* (3)*, and* (5)*, respectively. In these regressions, λ of Equations* (3) *and* (5) *is fixed after applying non-linear least squares to our seven- and nine-factor models. The chosen λ varies around 5.00 for each dataset. The table also reports the percentage of significant factor coefficients within each model and index option dataset. The data cover the period between January 1, 2003, and December 31, 2017.*

4 Results and Discussion

4.1 Modelling the Implied Volatility Surface

Our deterministic models are evaluated using recursive cross-sectional generalised least squares (GLS). In these regressions, λ of Equations (3) and (5) is fixed after applying non-linear least squares to our seven- and nine-factor models. The chosen λ varies around 5.00 for each dataset.

In-sample statistics are given in Table 2, which consists of measures for the goodness of fit and the percentage of significant coefficients for each of the factors within each deterministic model.

Similarly to the findings of Chalamandaris and Tsekrekos (2011) for currency options, their seven-factor model renders a significantly higher average in-sample R^2. Moreover, both the maximum and minimum variance explained are higher for all index options compared to the five-factor model. Our newly proposed nine-factor model does increase the R^2 of the seven-factor model as expected. The model even significantly outperforms the simpler models in terms of minimum R^2. Impressively, the minimum R^2 for the Russell 2000 index options given this model is more than double that of the five-factor model. A slight increase of the maximum R^2 is observed compared to the seven-factor model.

Among all models and datasets, we observe that all coefficients are significant at a 10% significant levels most of the time. As expected, the level coefficient is most significant of all, as it is the principal constituent of the implied volatility level.

The right "smile" seems to exhibit a more significant effect on the determination of implied volatility than the left "smile" as shown by the seven-factor coefficient results. When separating the corner moneyness regions from the more ATM regions in the nine-factor model, we find an interesting imbalance between ATM and OTM option contracts. Table 2 shows that the reduction in the significance of the left "smile" is mainly due to ATM option contracts as $\hat{\beta}_{3a}$ is on average less significant than $\hat{\beta}_{3b}$. Still, we notice that the right "smile" factor is on average more significant. Also, the difference between OTM and ITM contracts is less pronounced for this region. From these observations, we may conclude that the "smile" factor is on average more pronounced for ITM calls and OTM puts.

The term-structure effects also become apparent from this table, where we find a strong difference in significance between short- and medium-term maturities. The implied volatility is on average more significantly determined by the former. This could, however, be influenced by the choice of λ.

Interestingly, the interaction term between maturities and "smile" shows a strong difference between left and right "smiles" in both the seven- and nine-factor model. One could investigate how results would change when introducing interaction terms by separating terms for several maturity regions as well.

We winsorize the estimated coefficients by 0.50 percent to correct for the strong influence of outliers in our dynamic forecasting procedures. Figure 7 in Appendix B plots the daily evolution of these coefficients, averaged over our four index option datasets. Generally, we find similar patterns of the estimated coefficients between all models, while they follow closely together for the seven- and nine-factor models. It becomes apparent that all factors are significantly time-varying, as expected. Panels (b) and (c) show a convolution between the factors for ATM and OTM options while identifying a distinct difference in their values. Panels (d) and (e) also show strong asymmetry between short- and medium term contracts, as the values of their factor coefficients differ significantly. From these observations, it becomes evident that our nine-factor model specification is reasonable.

To further evaluate the in-sample fit of our deterministic models, we construct plots of the percentage root-mean-square error (RMSE) for both moneyness regions and maturity regions. The moneyness (Δ) regions, which range from region 1 through 6 are as follows: $-50 < \Delta \leq -37.5$; $-37.5 < \Delta \leq -12.5$; $-12.5 < \Delta \leq 0$;

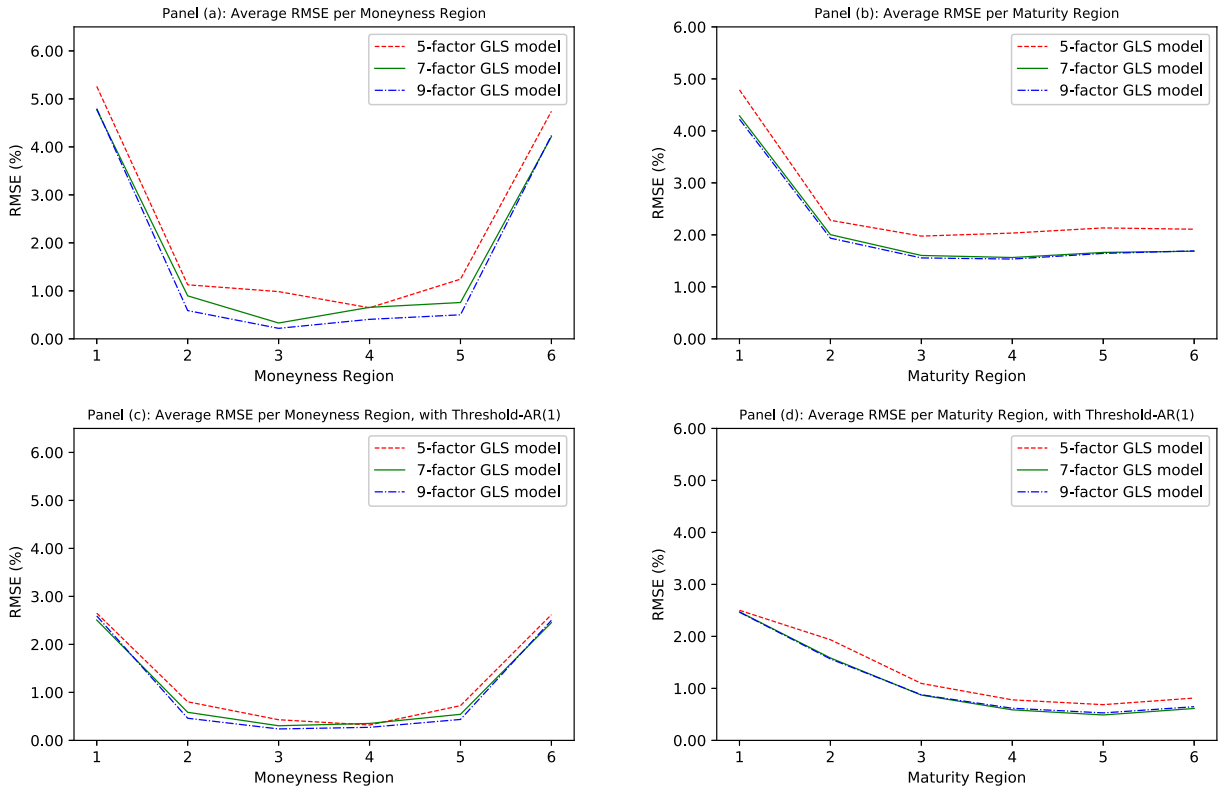

Figure 2: This figure shows four plots of the average root-mean-square error (RMSE); two over six moneyness regions (Panels (a) and (c)) and two over six maturity regions (Panels (b) and (d)), obtained from fitting the 5-, 7-, and 9-factor deterministic models given by Equations (2)*,* (3)*, and* (5) *using generalised least squares (GLS). Panels (a) and (b) present the RMSEs before correcting for our in-sample error terms with the use of a threshold-AR(1) model, given by Equation* (7)*. Panels (c) and (d) shows the RMSEs after this correction. The RMSEs are averaged over our different index options datasets, which cover the period between January 1, 2003, and December 31, 2017, is given. The moneyness* (Δ) *and maturity* (τ) *regions are depicted as follows:* $-50 < \Delta \leq -37.5$*;* $-37.5 < \Delta \leq -12.5$*;* $-12.5 < \Delta \leq 0$*;* $0 < \Delta \leq 12.5$*;* $12.5 < \Delta \leq 37.5$*;* $37.5 < \Delta \leq 50$*.* $6 < \tau \leq 60$*;* $60 < \tau \leq 120$*;* $120 < \tau \leq 180$*;* $180 < \tau \leq 240$*;* $240 < \tau \leq 320$*;* $320 < \tau$*.*

$0 < \Delta \leq 12.5$; $12.5 < \Delta \leq 37.5$; $37.5 < \Delta \leq 50$. The six maturity (τ) regions are separated in a similar equally spaced manner: $6 < \tau \leq 60$; $60 < \tau \leq 120$; $120 < \tau \leq 180$; $180 < \tau \leq 240$; $240 < \tau \leq 320$; $320 < \tau$. We distinguish between the three deterministic models but take the average RMSE of our four datasets for each region.

These RMSEs are visualised in Figure 2, where the average RMSE per moneyness region and maturity region are given respectively in Panel (a) and Panel (b). From this figure, we notice a clear "smile" shape in the RMSEs over the moneyness regions for all deterministic models, showing that they clearly fail to capture the non-linearity in the corner regions in this dimension. Nevertheless, we perceive lower RMSEs when including more factors in our models. The nine-factor model specifically produces the lowest errors across all moneyness regions. The figure also displays relatively low RMSEs between regions two and five, meaning that the models do capture a large extent of the IVS shape for these regions. Furthermore, we observe relatively large RMSEs for very short-term maturities, while flattening out for larger days to maturity. In terms of maturity, the extended models produce a better overall fit in terms of residuals compared to simpler models. Given the maturity regions, the nine-factor model obtains the lowest average root-mean-square errors. Both plots taken together still show however an alarmingly non-linear shape, for which inevitably has to be corrected.

As discussed in Section 3.1, we therefore explicitly model the error terms from the GLS regressions while separating put and call option contracts, as specified by Equation (7). The effective residuals of the adjusted IVS fit now equal to $\eta^{w}_{i,t}$, as denoted by the stated equation. Panel (c) and Panel (d) in Figure 2 show that modelling the GLS residuals via our threshold-AR(1) model drastically improves our fit in the corner regions. Specifically, the OTM/ITM regions present significantly lower average root-mean-square errors. In the outer corner regions, we namely achieve a decrease of over 40 percent. Also, the middle regions seem to flatten out somewhat as well, along with an overall decrease in RMSE. Taken together, we find a less significant "smile" shape along the moneyness regions. The in-sample correction using our dynamic error term model also improves the fit along the maturity regions. We clearly observe straighter lines in Panel (d) compared to Panel (b), especially for the short-term maturities. We thus find a strong flattening effect over the fitted implied volatility surfaces when using our threshold-AR(1) model for the residuals. As this correction renders such significant improvements, which are in line with previous research, we apply it to all of our forecasting results.[8] It must be noted that, due to this strong correction, the differences in performance across our three deterministic models are less apparent. Nevertheless, performance comparisons remain feasible.

In conclusion, we find that our nine-factor model produces the best fit among the deterministic models,

[8] We have investigated the effect of including the lagged error term in our forecasts and concluded that it always significantly improves our results.

having the highest average, minimum, and maximum R^2 for all indices. Moreover, the model produces the lowest root-mean-square errors for all moneyness and maturity regions. Finally, the coefficients for moneyness corner regions seem to differ significantly from more centred regions. In sum, we may argue that our nine-factor model delivers the best in-sample description of the implied volatility surface among previously proposed models.

4.2 Modelling the Dynamics of Index Option IVS

4.2.1 The Statistical Value of Predictability

We assess the statistical forecasting performance of our dynamic models for each of the three deterministic GLS models. The results of our discussed methodology of Section 3.2 are presented in Table 3. The table shows the percentage RMSE, MAE, and MCPDC for every model in four panels. For each statistical measure, the percentage of decrease compared to the relatively simpler model is given in parentheses. The seven-factor model is thus compared to the five-factor model and the nine-factor model is subsequently compared to the seven-factor model. A negative percentage can be considered an improvement in terms of RMSE and MAE, while the converse is true for the MCPDC as it corresponds to a relative decrease in the calculated statistic. Panels A, B, C, and D display results for respectively Dow Jones, Nasdaq, S&P 500, and Russell 2000 index options.

Most noticeable is the fact that the RMSE is significantly smaller for the seven- and nine-factor models compared to the five-factor model of Goncalves and Guidolin (2006) across all tickers, showing the effectiveness of our term structure specification. The difference in RMSE between the former two models is, however, less pronounced. Only the Bayesian VAR for all tickers and the AR model for SPX and RUT deliver a lower RMSE for the 9-factor model compared to the model of Chalamandaris and Tsekrekos (2011). We must note, however, that the AR model performs worse for the seven-factor model compared to its simpler specification. The nine-factor model does not exhibit such a large difference and even performs better than the five-factor model in this case. Among our dynamic models, the Strawman random walk performs best in terms of our root-mean-square errors, with only one exception.

The five-factor model is also beaten by our other

Table 3: Statistical Measures of Predictability

	5-factor GLS			7-factor GLS			9-factor GLS		
	RMSE (%)	MAE (%)	MCPDC (%)	RMSE (%)	MAE (%)	MCPDC (%)	RMSE (%)	MAE (%)	MCPDC (%)
Panel (a): DJX Options									
AR	3.60	1.86	53.02	2.69(-25.30)	1.75 (-5.77)	52.74 (-0.53)	2.70 (0.57)	1.75 (0.01)	52.72 (-0.03)
VAR	3.57	1.86	54.35	2.33(-34.84)	1.30(-29.70)	53.78 (-1.04)	2.36 (1.51)	1.32 (1.17)	53.63 (-0.28)
VARX	3.57	1.86	54.31	2.33(-34.78)	1.31(-29.75)	53.86 (-0.82)	2.36 (1.54)	1.32 (1.14)	53.72 (-0.26)
Bayesian VAR	3.93	2.14	53.96	3.45(-12.17)	2.35 (9.62)	53.54 (-0.78)	3.34 (-3.29)	2.24 (-4.53)	53.31 (-0.43)
Strawman random walk	3.31	1.86	52.77	1.88(-43.16)	1.04(-44.33)	52.65 (-0.21)	1.88 (0.24)	1.03 (-0.27)	52.81 (0.30)
Pure random walk	3.29	1.27	56.01						
Panel (b): NDX Options									
AR	2.74	1.44	53.02	2.28(-16.95)	1.34 (-6.72)	53.88 (1.63)	2.26 (-0.83)	1.32 (-1.63)	53.85 (-0.07)
VAR	2.73	1.43	53.24	1.92(-29.61)	1.02(-28.93)	53.30 (0.12)	1.95 (1.81)	1.04 (2.65)	52.72 (-1.09)
VARX	2.73	1.43	53.22	1.93(-29.43)	1.02(-28.60)	53.15 (-0.14)	1.96 (1.91)	1.05 (2.58)	52.61 (-1.01)
Bayesian VAR	3.07	1.70	52.30	2.94 (-3.93)	1.82 (7.22)	52.70 (0.76)	2.88 (-2.25)	1.77 (-3.05)	52.44 (-0.49)
Strawman random walk	2.66	1.46	51.89	1.51(-43.11)	0.80(-45.17)	51.95 (0.11)	1.52 (0.42)	0.80 (-0.34)	52.24 (0.55)
Pure random walk	3.11	1.26	52.83						
Panel (c): SPX Options									
AR	3.13	1.73	53.03	3.77 (20.74)	2.92 (68.29)	52.77 (-0.50)	2.17(-42.51)	1.39(-52.53)	54.64 (3.55)
VAR	3.13	1.75	53.44	2.02(-35.38)	1.20(-31.42)	54.35 (1.70)	2.05 (1.51)	1.21 (0.87)	54.09 (-0.49)
VARX	3.14	1.75	53.40	2.03(-35.23)	1.20(-31.34)	54.28 (1.66)	2.06 (1.39)	1.22 (0.97)	54.05 (-0.42)
Bayesian VAR	3.49	2.03	53.89	2.91(-16.55)	1.92 (-4.97)	54.27 (0.71)	2.78 (-4.65)	1.81 (-5.84)	53.90 (-0.68)
Strawman random walk	2.55	1.76	51.87	1.63(-36.16)	0.92(-47.73)	51.71 (-0.31)	1.62 (-0.90)	0.90 (-1.55)	52.05 (0.66)
Pure random walk	3.47	1.41	53.49						
Panel (d): RUT Options									
AR	2.72	1.51	53.97	2.43(-10.57)	1.53 (1.66)	53.75 (-0.42)	2.41 (-1.10)	1.51 (-1.59)	53.72 (-0.05)
VAR	2.65	1.45	54.48	1.97(-25.63)	1.10(-24.28)	54.22 (-0.48)	2.00 (1.67)	1.11 (0.67)	54.27 (0.10)
VARX	2.66	1.46	54.46	1.98(-25.38)	1.11(-24.15)	54.18 (-0.53)	2.01 (1.62)	1.12 (0.79)	54.29 (0.21)
Bayesian VAR	3.00	1.75	53.54	3.11 (3.62)	2.01 (14.49)	53.42 (-0.22)	3.03 (-2.67)	1.95 (-2.84)	53.41 (-0.03)
Strawman random walk	2.53	1.49	52.78	1.65(-34.85)	0.91(-39.08)	52.65 (-0.23)	1.65 (0.30)	0.91 (-0.32)	52.92 (0.50)
Pure random walk	2.86	1.10	55.66						

Notes: The table contains the out-of-sample statistical measures of predictability to evaluate the forecasting performance of five dynamic models for Dow Jones, Nasdaq, S&P 500, and Russell 2000 index options across three different deterministic models. The dynamic models include an AR, VAR, VARX, Bayesian VAR, given by Equations (9) - (13)*. The AR, VAR, and Bayesian VAR models take into account the lagged dynamics of the implied volatility surface of our index options. The VARX model extends these dynamics by including a factor representing the global dynamics of deviations in put-call parity obtained by principal component analysis. The Strawman random walk and pure random walk models of Equations* (15) *and* (16) *are included as benchmarks. The three deterministic models include the five-, seven-, and nine-factor deterministic GLS models given by Equations* (2)*,* (3)*, and* (5)*, respectively. The statistical measures include the root-mean-square error (RMSE), mean absolute error (MAE), and the mean correct prediction of the direction of change (MCPDC).* $RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$; $MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t|$; $MCPDC = \frac{1}{n}\sum_{t=1}^{n} 1_{t,(sign(\sigma_{i,t+1}-\sigma_{i,t})=sign(\hat{\sigma}_{i,t+1}-\sigma_{i,t}))}$. *We calculate these statistics for both the GLS while correcting for our residuals obtained from applying generalised least squares (GLS) to our deterministic model. This correction is applied by modelling these residuals using a threshold-AR(1) model given by Equation* (7)*. The data cover the period between January 1, 2003, and December 31, 2017.*

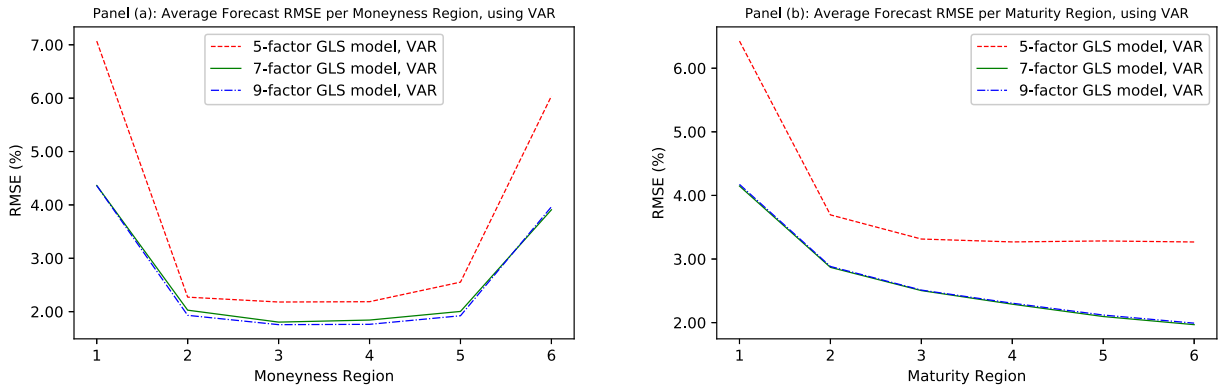

Figure 3: This figure shows two plots of the average root-mean-square error (RMSE); one over six moneyness regions (Panel (a)) and one over six maturity regions (Panels (b)), obtained from forecasting the implied volatility surface using a dynamic VAR model on the estimated coefficients obtained from the 5-, 7-, and 9-factor deterministic models given by Equations (2)*,* (3)*, and* (5) *using generalised least squares (GLS). The RMSEs are averaged over our different index options datasets, which cover the period between January 1, 2003, and December 31, 2017, is given. The moneyness* (Δ) *and maturity* (τ) *regions are depicted as follows:* $-50 < \Delta \leq -37.5$*;* $-37.5 < \Delta \leq -12.5$*;* $-12.5 < \Delta \leq 0$*;* $0 < \Delta \leq 12.5$*;* $12.5 < \Delta \leq 37.5$*;* $37.5 < \Delta \leq 50$*.* $6 < \tau \leq 60$*;* $60 < \tau \leq 120$*;* $120 < \tau \leq 180$*;* $180 < \tau \leq 240$*;* $240 < \tau \leq 320$*;* $320 < \tau$*.*

specifications in terms of the MAE, except for the $\hat{\beta} - AR$ models in Panels (c) and (d). The MAE of the $\hat{\beta} - BVAR$ model is also slightly larger in Panel (d). The extended models thus show value in terms of relative improvement in forecasting precision. Again, we find the generally best performing dynamic model to be the Strawman random walk. The Bayesian VAR model is, however, beaten by all other dynamic models. It must be noted that the pure random walk outperforms our dynamic models for the five-factor model across all tickers.

Finally, the MCPDC shows inconclusive results. The seven-factor model has a worse MCPDC in Panel (a) compared to the simpler deterministic model. Our nine-factor model does not generally improve this result, as only a larger MCPDC is observed for the Strawman random walk of a 0.50 percent increase on average across all panels. A noteworthy result is the one percent increase in MCPDC in panel B for the seven-factor model compared to the five-factor model. Interestingly, among our dynamic models, the MCPDC is highest for our $\hat{\beta} - VAR$ and $\hat{\beta} - VARX$. Some exceptions are noticed for the $\hat{\beta} - AR$ and $\hat{\beta} - BVAR$. A case in point is the MCPDC 53.89 percent for the latter dynamic model in Panel (c) using the estimated factor coefficients of the five-factor model. The pure random walk outperforms all models in Panels (a) and (d). Half of the time, however, this model is defeated by the dynamic models for NDX and SPX index options.

Although we do not observe a stronger forecasting performance by including the DPCP factor within our $\hat{\beta} - VARX$ model, the variance explained within the first principal component is higher than 81 percent for all tickers. One would, therefore, expect that there exists a strong global dynamic across all contracts in terms of these deviations from put-call parity.

Furthermore, Figure 3 shows the root-mean-square errors averaged over all dataset for six moneyness and maturity regions. Due to the ease of comprehension and superior performance among the dynamic models, we focus only on $\hat{\beta} - VAR$. The average RMSE per moneyness region and maturity region are given respectively in Panel (a) and Panel (b). Pre-eminently, we find that both the seven- and nine-factor model outperforms the five-factor model for all regions in both panels. Panel (a) presents a slight overall lower RMSE for the nine-factor model, compared to the model of Chalamandaris and Tsekrekos (2011). This shows that our proposed model indeed somewhat improves on capturing the dynamics of the moneyness dimension. It does not, however, clearly outperform the seven-factor model in terms of maturity, as would be expected due to its similar specification. From the panel (a), we observe a clear "smile" shape in the RMSEs over the moneyness regions for all deterministic models, similar to our in-sample results of Section 3.1. We find similar patterns in residuals obtained from our one-day ahead forecasts compared to our in-sample fit. Hence, our three models still fail to accurately predict the corner regions compared to ATM option contracts. Finally, the difference between the five-factor model and our more sophisticated models has become more apparent in our forecasts. However, our models are still to be improved if we aim to describe and forecast the corner regions of the IVS shape well.

To grasp the forecasting performance for larger horizons, we report the percentage RMSE for each ticker and deterministic model for a one-day up until and including ten-day horizons. Focusing only on $\hat{\beta} - VAR$, the results are visualised in Figure 4. From this figure, we immediately notice the fact that all deterministic models mostly outperform the pure random walk for higher forecasting horizons. Some exceptions are observed, especially for the seven-factor model using NDX options and the nine-factor model using RUT options. Both cases seem to diverge from the other models around a forecasting horizon of five days. The fact that the five-factor model is outperformed by our extended models until a horizon of seven days for DJX and SPX options is very promising. This also applies to our NDX and RUT options, although until a horizon of three days.

In sum, we find the model proposed by Chalamandaris and Tsekrekos (2011) to significantly improve our forecasting results in terms of RMSE and MAE compared to the model of Goncalves and Guidolin (2006). Moreover, the performance of our proposed nine-factor specification matches closely to the seven-factor model and in some cases significantly improves forecasting results for larger forecasting horizons.

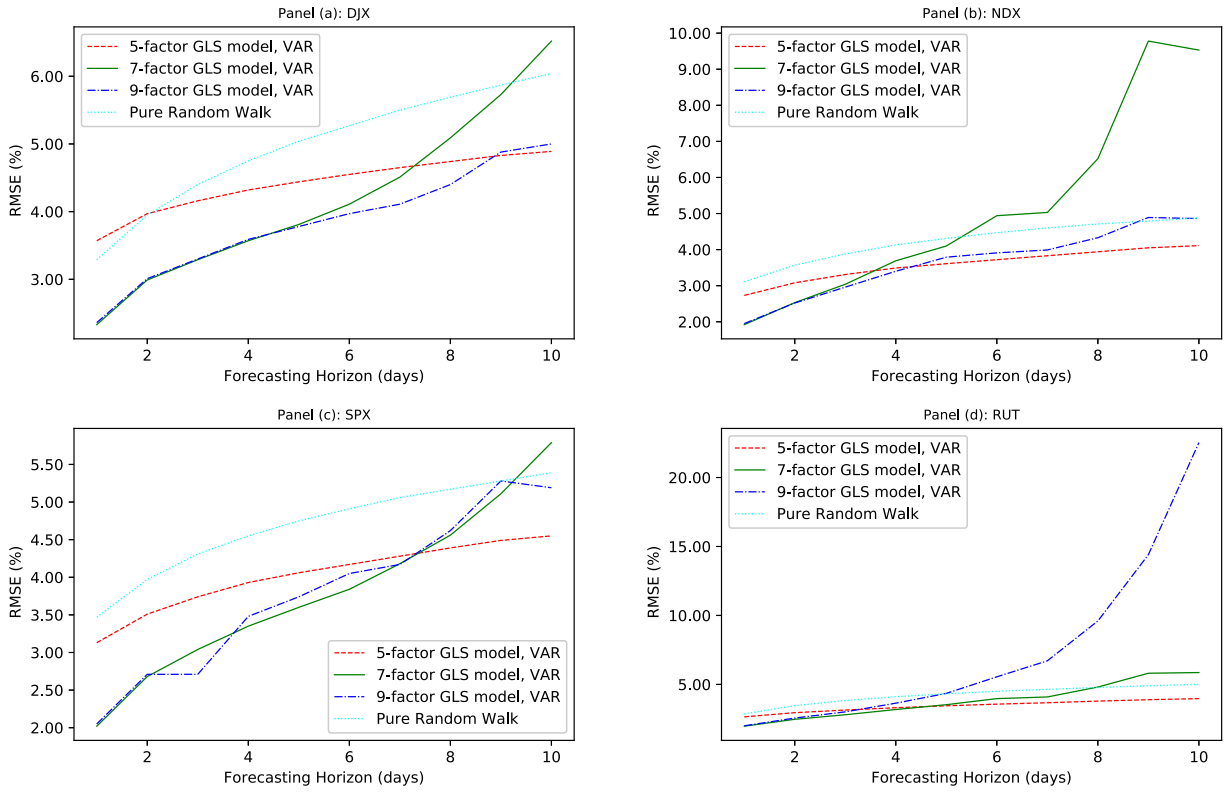

Figure 4: This figure shows four plots of the root-mean-square error (RMSE) obtained by the use of a VAR (Equation (13)*) and pure random walk (Equation* (16)*) model for each of our three deterministic IVS models. The latter include the five-, seven-, and nine-factor deterministic GLS models given by Equations* (2), (3), *and* (5)*, respectively. Results are given for each of our four index option datasets. From Panels (a)-(d), these include Dow Jones (DJX), Nasdaq (NDX), Standard and Poor's 500 (SPX), and Russell 2000 (RUT). The RMSEs are given for forecast horizons one through ten days.*

4.2.2 The Economic Value of Predictability

Building upon our findings in a statistical sense, we investigate the economic value of our proposed models. For this, we apply the methodology of Section 3.2.2. Our delta-hedged trading results, along with our benchmarks are reported in Table 4. The results consist of the average daily profit, its standard deviation, and the Sharpe ratio of our portfolios. The table shows the superiority of our dynamic models over both the Buy-and-Hold and Treasury bill benchmark portfolios. As expected, both the time value of money and the average risk-less rate are close to zero. Across all deterministic and dynamic models, we obtain significant abnormal returns, which is in line with findings of previous literature (e.g. Bernales and Guidolin (2014)).

The mean profit is generally highest for the seven-factor GLS model. While mostly showing similar returns as our nine-factor model, it significantly improves the result of the five-factor model of Goncalves and Guidolin (2006). The nine-factor model outperforms its simpler models merely for Russell 2000 options, among several exceptions. The Strawman random walk mostly renders the lowest average daily returns among all dynamic models. Our conclusions on the other dynamic models remain mixed, as none consistently outperforms their alternatives.

Conversely, the five-factor model is not outperformed by the other deterministic models in terms of risk-adjusted returns. Most notably, the Bayesian VAR model performs worst for our extended models. $\hat{\beta} - BVAR$ also renders the lowest average daily returns among all dynamic models. Both the VAR and VARX models stand out in terms of their risk-adjusted returns. Interestingly, $\hat{\beta} - VARX$ has on average a higher Sharpe ratio than the $\hat{\beta} - VAR$. Involving the market sentiment factor, therefore, seems moderately valuable in obtaining risk-adjusted abnormal returns.

To further evaluate the economic value of our deterministic models and investigate the most profitable trading regions, we construct plots of the percentage daily average profit for both moneyness regions and maturity regions. These regions are identical to those of Section 4.1 and, consistent with Figure 4, we report results obtained from our VAR model. The average daily profits are averaged over our four index option datasets while distinguishing between our three deterministic GLS models. From Panel (a) in Figure 5, we find that ITM calls and OTM puts are most profitable given our trading strategy. For ITM puts and OTM calls, however, we obtain on average negative results, *ante* transaction costs. Although the plot exhibits strong non-linearity, we observe no "smile" shape. Panel (b) of this figure, given the maturity regions, displays a similar shape as in Panel (b) of Figure 2. The short-term option contracts show significantly larger average daily profits than for longer maturities, while these have been recognised to be subject to the largest in-sample root-mean-square errors. This could be due to the comparatively larger variance, as previously shown in Table 1.

Clearly, our seven-factor and nine-factor models outperform the simplest deterministic model for nearly all moneyness and maturity regions. We may, therefore,

Table 4: Economic Measures of Predictability (before transaction costs)

	5-factor GLS			7-factor GLS			9-factor GLS		
	Mean Profit (%)	St.dev. Profit (%)	Sharpe Ratio (%)	Mean Profit (%)	St.dev. Profit (%)	Sharpe Ratio (%)	Mean Profit (%)	St.dev. Profit (%)	Sharpe Ratio (%)
Panel (a): DJX Options									
AR	1.45	11.89	12.19	3.02	24.40	12.38	3.11	24.40	13.28
VAR	2.73	16.88	16.12	2.98	18.48	16.12	2.57	17.11	15.00
VARX	2.65	16.30	16.20	3.30	18.85	17.48	2.79	18.59	15.00
Bayesian VAR	2.37	14.77	15.99	2.34	40.32	5.80	3.12	41.94	7.42
Strawman random walk	1.52	15.06	10.09	1.85	12.55	14.71	2.15	14.41	14.87
Buy-and-hold	0.03	1.07	2.79						
Panel (b): NDX Options									
AR	5.35	34.80	15.37	9.50	62.56	15.18	6.89	48.82	14.10
VAR	5.68	30.90	18.37	4.39	28.97	15.14	3.78	28.00	13.49
VARX	4.88	24.84	19.64	4.07	29.01	14.00	4.11	32.46	12.63
Bayesian VAR	4.93	26.62	18.51	5.64	49.23	11.44	4.99	47.98	10.39
Strawman random walk	3.72	8.42	13.93	2.68	23.27	11.48	2.80	22.63	12.36
Buy-and-hold	0.06	1.31	4.08						
Panel (c): SPX Options									
AR	4.64	23.36	19.86	4.79	102.34	4.68	5.16	37.84	13.62
VAR	6.36	30.82	20.62	3.44	19.42	17.69	2.99	23.11	12.93
VARX	5.10	24.54	20.78	4.40	25.41	17.30	3.07	21.64	14.16
Bayesian VAR	3.84	22.99	16.67	3.18	36.01	8.80	3.32	34.93	9.49
Strawman random walk	4.51	11.40	18.85	2.90	17.91	16.17	2.67	15.97	16.72
Buy-and-hold	0.04	1.16	3.29						
Panel (d): RUT Options									
AR	2.28	19.18	11.88	4.98	38.32	12.98	6.63	48.56	13.64
VAR	3.81	22.60	16.84	3.67	21.95	16.70	4.02	25.16	14.18
VARX	3.74	23.60	15.83	3.50	21.56	16.23	4.34	26.52	16.33
Bayesian VAR	2.76	18.55	14.88	4.21	79.80	5.27	4.32	84.83	5.08
Strawman random walk	2.37	22.83	10.34	2.09	17.05	12.25	2.47	17.37	14.18
Buy-and-hold	0.05	1.49	2.91						
Panel (e): Benchmark									
T-Bill	0.05	0.06	0.00						

Notes: The table contains the out-of-sample delta-hedge trading results before transaction costs to evaluate the economic value of five dynamic models for Dow Jones, Nasdaq, S&P 500, and Russell 2000 index options across three different deterministic models (see Section 3.2.2). The data cover the period between January 1, 2003, and December 31, 2017. The dynamic models include an AR, VAR, VARX, Bayesian VAR, given by Equations (9) - (13)*. The AR, VAR, and Bayesian VAR models take into account the lagged dynamics of the implied volatility surface of our index options. The VARX model extends these dynamics by including a factor representing the global dynamics of deviations in put-call parity obtained by principal component analysis. The Strawman random walk of* (15) *is included as the dynamic benchmark. Two other benchmarks are included: a riskless daily investment in the index options which we denote as the Buy-and-Hold strategy and an initial investment worth 1000$ in Treasury bills. The latter results in the time value of money. The three deterministic models include the five-, seven-, and nine-factor deterministic GLS models given by Equations* (2)*,* (3)*, and* (5)*, respectively. The economic measures include the average daily percentage return, its standard deviation, and the risk-adjusted return (Sharpe ratio). In obtaining our trading results, we use our implied volatility forecasts which are corrected using residuals obtained from applying generalised least squares (GLS) to our deterministic model. This correction is applied by modelling these residuals using a threshold-AR(1) model given by Equation* (7)*.*

conclude that the more extensive models are of higher economic value than the model of Goncalves and Guidolin (2006).

Despite the fact that all our models produce significant risk-adjusted abnormal returns, we have yet to establish whether these are due to the exclusion of transaction costs. As suggested by Battalio and Schultz (2006) we thus include transaction costs in our trading strategies of 0.8 times the quoted bid-ask spread. From Table 5 in Appendix C we immediately notice that both our average and risk-adjusted returns suffer significantly from these transaction costs. As expected, we now fail to deliver positive abnormal returns. The comparisons between our models stay roughly similar. Nonetheless, we deliver trading results of a highly simplified strategy. These should, therefore, serve as a mere lower-bound for the expected results obtained by experienced traders who have more sophisticated algorithms and means to process information.

5 Conclusion and Further Research

Currently, there exists a plethora of literature revolving around the modelling of the highly non-linear shape of implied volatility surfaces. The ability to exploit this non-flat shape to the benefit of academic knowledge, risk management, or augmenting trading performance seems to attract professionals in their respective fields to investigate the matter at hand.

Within this research, we have built upon previous literature focusing on applying parametric models in order to describe the "smile/smirk" shape of the surface in terms of option moneyness and expiration time. More specifically, we applied the deterministic models of Goncalves and Guidolin (2006) and Chalamandaris and Tsekrekos (2011) to four index option datasets. Additionally, we proposed a nine-factor model, that extends the seven-factor model of the latterly named authors. Alongside modelling the outer regions of maturity, it focuses on capturing the strong asymmetries in implied volatility between At-The-Money and Deep In(Out)-The-Money put and call options.

We have shown that both extended models result in a better overall fit of the surface in terms of variance

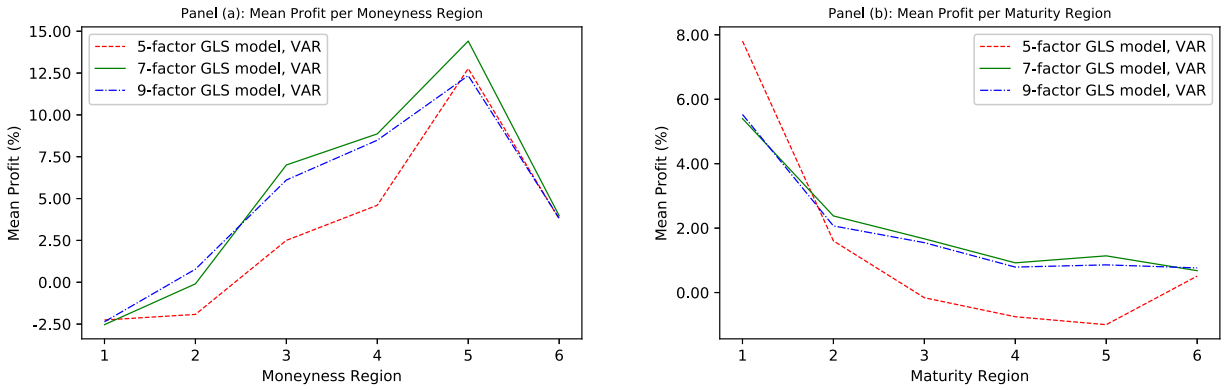

Figure 5: This figure shows two plots of the percentage average daily return from our delta-hedged portfolios, as obtained by our dynamic VAR model (Equation (10)*). These profits are plotted against six moneyness regions (Panel (a)) and over six maturity regions (Panel (b)). The calculated returns result from applying our delta-hedge strategy to our $\hat{\beta}$ – VAR forecasts after fitting the 5-, 7-, and 9-factor deterministic models given by Equations* (2)*,* (3)*, and* (5) *using generalised least squares (GLS). These forecasts are constructed by use of a threshold-AR(1) model, given by Equation* (7)*. The average daily profits are averaged over four index options datasets, which cover the period between January 1, 2003, and December 31, 2017. The four datasets consist of Dow Jones, Nasdaq, S&P 500, and Russell 2000 index options. The moneyness (Δ) and maturity (τ) regions are depicted as follows: $-50 < \Delta \leq -37.5$; $-37.5 < \Delta \leq -12.5$; $-12.5 < \Delta \leq 0$; $0 < \Delta \leq 12.5$; $12.5 < \Delta \leq 37.5$; $37.5 < \Delta \leq 50$. $6 < \tau \leq 60$; $60 < \tau \leq 120$; $120 < \tau \leq 180$; $180 < \tau \leq 240$; $240 < \tau \leq 320$; $320 < \tau$.*

explained and average residuals. As the nine-factor model has been proven to best describe the implied volatility surface in terms of in-sample fit, we may conclude that our specification is sensible and highly promising. It seems to describe the time-variation of the IVS in terms of moneyness somewhat better than the model of Chalamandaris and Tsekrekos (2011). In obtaining these results, we have applied generalised least squares (GLS) after fixing the decay rate for maturities within our seven- and nine-factor models. None of the models can, however, fully capture the non-linearity in the IVS in corner regions. Therefore, we have explicitly modelled the residuals with their lag while accounting for asymmetry in dynamics between put and call options, which drastically improved both in- and out-of-sample results.

Several popular dynamic models have been used to determine the out-of-sample statistical performance across our three deterministic models, along with simple benchmarks. We have shown that the seven-factor model significantly improves the general performance of the five-factor model for one-day ahead forecasts and larger horizons. The nine-factor model has shown to render similar results. Moreover, our dynamic models mostly outperform the benchmarks in terms of forecasting accuracy.

Our models generally prove to be promising in terms of their economic applicability. We have illustrated this by applying a simple delta-hedge strategy based on our IVS forecasts. Before transaction costs, all our models produce significantly positive abnormal risk-adjusted returns, outperforming well-known benchmarks. While these returns all become negative after imposing transaction costs, we may conclude that our models produce strong results. In reality, traders have more means and experience to process information and stronger trading algorithms that may produce better forecasts and strategies. Our results are, therefore, mainly illustrative of general model performance. Interestingly, we have found significant differences in average daily returns over both moneyness regions and maturity regions. ITM call options, OTM put options, and short-term maturity options have shown to be most profitable.

In conclusion, our results offer valuable contributions to existing literature. Firstly, we may conclude that it makes sense to explicitly model both the term structure and moneyness of index options as it provides better overall results than the simple five-factor model. Although the nine-factor model does not significantly improve on existing literature in terms of forecasting accuracy, it provides the best in-sample performance. Using our nine-factor model as a new baseline, one could investigate further refinements in modelling the moneyness dimension of the IVS. Secondly, ensuing research could focus on the exploitation of the asymmetry of corner regions in both maturity and moneyness dimensions, as this may be specifically helpful in obtaining abnormal returns. Finally, one could investigate further the constituents of the time-varying non-linear shape of the surface as these may originate from the dynamics of both macroeconomic factors and individual underlying stocks.

Acknowledgements We would like to thank mister Xun Gong from the Tinbergen Institute for his guidance in conducting this research and the thoughtful insights that he provided into the world of implied volatility.

References

Amin, K., Coval, J. D., and Seyhun, H. N. (2004). Index option prices and stock market momentum. *The Journal of Business*, 77(4):835–874.

Andersen, T. G., Fusari, N., and Todorov, V. (2017). Short-term market risks implied by weekly options. *The Journal of Finance*, 72(3):1335–1386.

Battalio, R. and Schultz, P. (2006). Options and the bubble. *Journal of Finance*, 61(5):2071–2102.

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. and Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654.

Bollen, N. P. and Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *The Journal of Finance*, 59(2):711–753.

Campa, J. and Chang, P. H. K. (1995). Testing the expectations hypothesis on the term structure of volatilities in foreign exchange options. *Journal of Finance*, 50(2):529–47.

Canina, L. and Figlewski, S. (1993). The informational content of implied volatility. *Review of Financial Studies*, 6(3):659–81.

Chalamandaris, G. and Tsekrekos, A. (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.

Christoffersen, P., Fournier, M., and Jacobs, K. (2013). The factor structure in equity options. Creates research papers, Department of Economics and Business Economics, Aarhus University.

Cremers, M. and Weinbaum, D. (2010). Deviations from put-call parity and stock return predictability. *Journal of Financial and Quantitative Analysis*, 45(2):335–367.

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

Diebold, X., Li, C., and Yue, V. Z. (2008). Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach. *Journal of Econometrics*, 146(2):351–363.

Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, 3(1):1–100.

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

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

Heynen, R., Kemna, A., and Vorst, T. (1994). Analysis of the term structure of implied volatilities. *The Journal of Financial and Quantitative Analysis*, 29(1):31–56.

Kim, M. and Kim, M. (2003). Implied volatility dynamics in the foreign exchange markets. *Journal of International Money and Finance*, 22(4):511–528.

Nelson, C. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–89.

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.

Sheldon, N. (1994). Option volatility & pricing : Advanced trading strategies and techniques.

Xu, X. and Taylor, S. J. (1994). The term structure of volatility implied by foreign exchange options. *The Journal of Financial and Quantitative Analysis*, 29(1):57–74.

Yan, S. (2011). Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics*, 99(1):216–233.

Yang, S.-H., Lee, J.-W., and Han, G.-S. (2010). Modeling implied volatility surfaces using two-dimensional cubic spline with estimated grid points. *Industrial Engineering and Management Systems*, 9(4):323–338.

Appendices

A Data

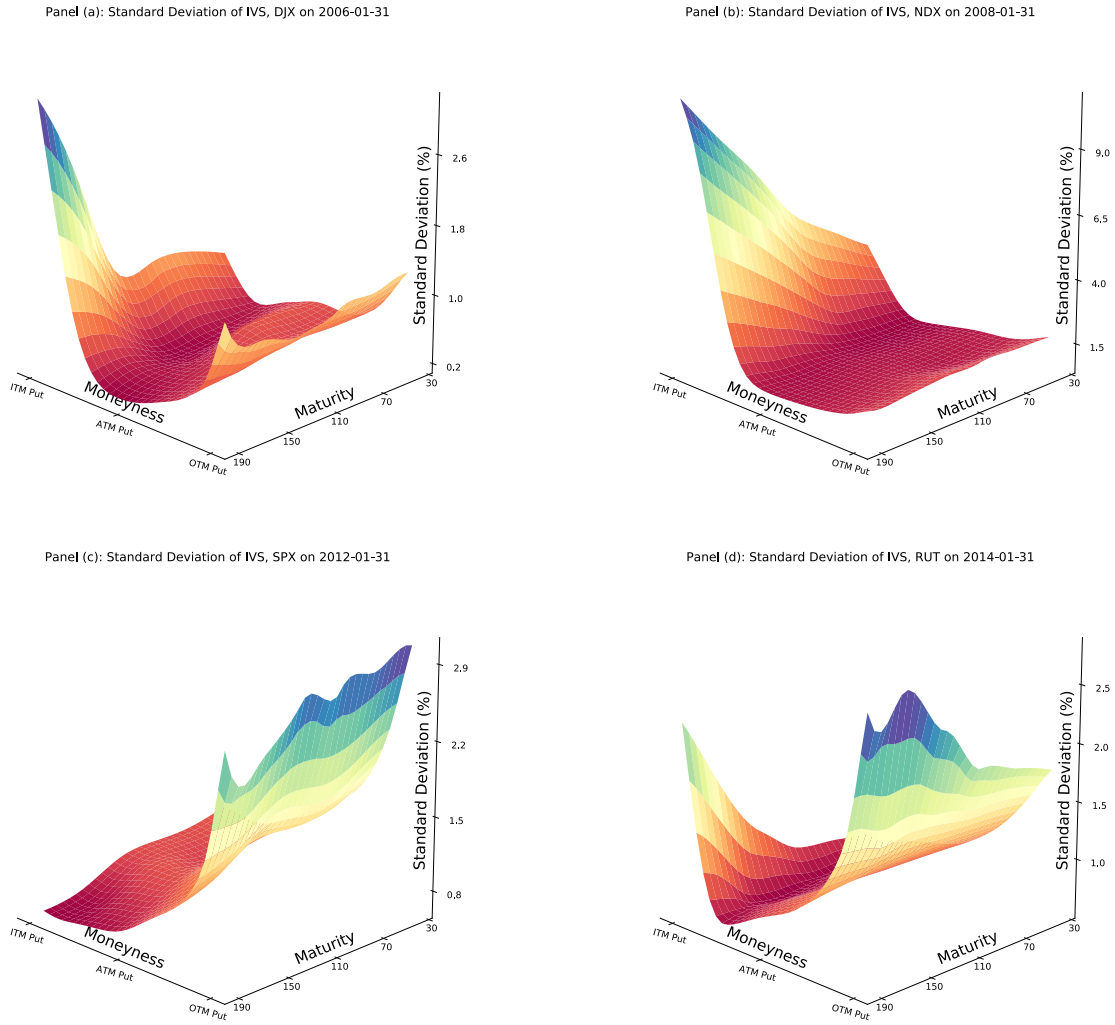

Figure 6: This figure shows four plots of the standard deviation of the implied volatility surface (IVS) for each of our index option datasets on a given date. From Panel (a) to Panel (d): Dow Jones (DJX), Nasdaq (NDX), Standard and Poor's 500 (SPX), and Russell 2000 (RUT). The IVS plots are given on a different date for each index. In the same order: 2006-01-31, 2008-01-31, 2012-01-31, and 2014-01-31.

B Modelling the Implied Volatility Surface

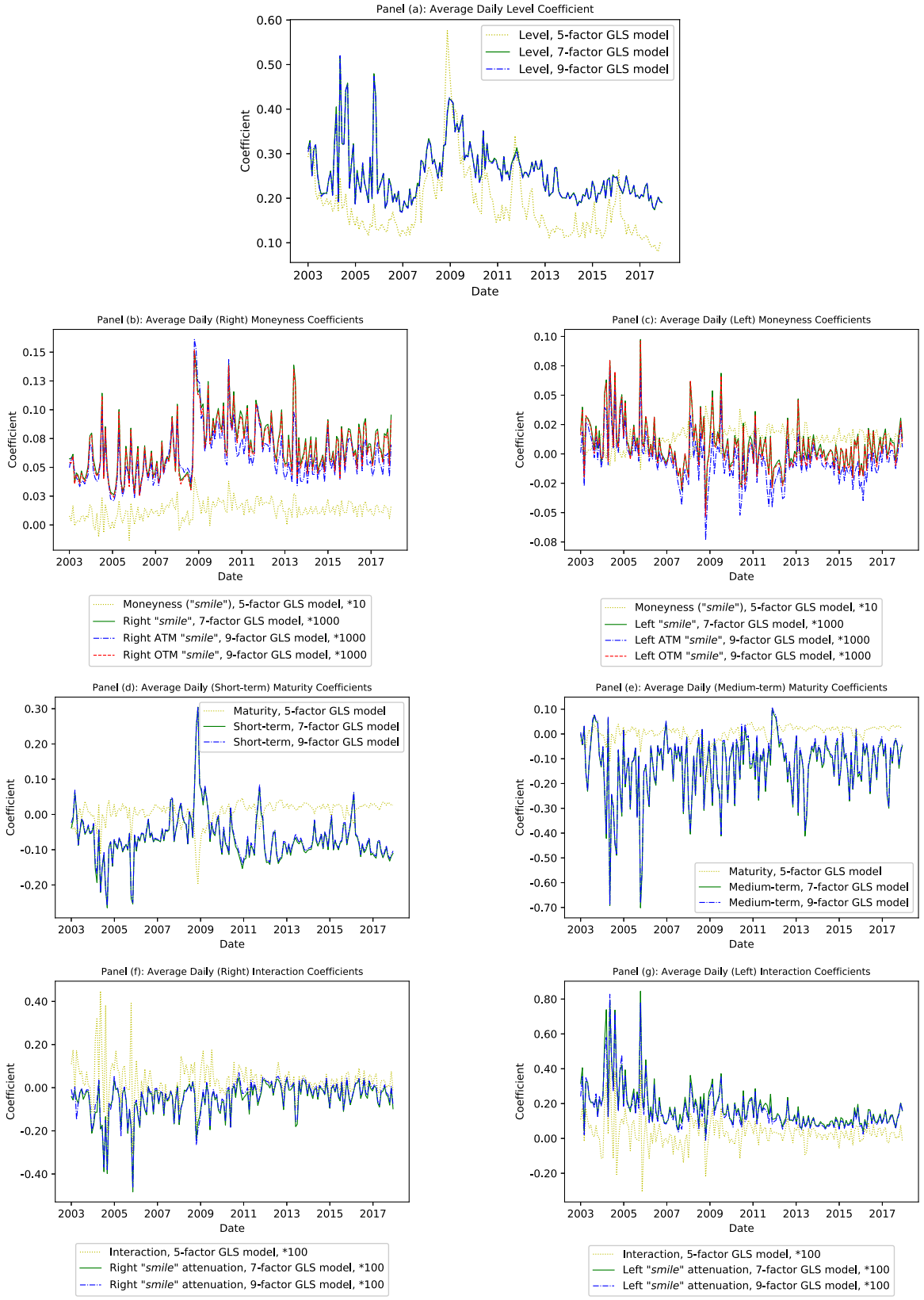

Figure 7: This figure plots the evolution of the estimated and 0.50 percent winsorized factor coefficients obtained by applying generalised least squares (GLS) to our 5-, 7-, and 9-factor deterministic models given by Equations (2), (3), *and* (5), *respectively. In these regressions, λ of Equations* (3) *and* (5) *is fixed after applying non-linear least squares to our seven- and nine-factor models. The chosen λ varies around 5.00 for each dataset. Each panel is constructed to show similar patterns among similar factors. The plots show the mean daily estimated coefficients, averaged over our four datasets consisting of Dow Jones, Nasdaq, S&P 500, and Russell 2000 index options. The data cover the period between January 1, 2003, and December 31, 2017.*

C The Economic Value of Predictability

Table 5: Economic Measures of Predictability (after transaction costs)

	5-factor GLS			7-factor GLS			9-factor GLS		
	Mean Profit (%)	St.dev. Profit (%)	Sharpe Ratio (%)	Mean Profit (%)	St.dev. Profit (%)	Sharpe Ratio (%)	Mean Profit (%)	St.dev. Profit (%)	Sharpe Ratio (%)
Panel (a): DJX Options									
AR	-2.22	14.35	-15.51	-1.99	26.68	-7.47	-1.81	25.08	-7.23
VAR	-11.14	266.55	-4.18	-10.30	115.49	-8.92	-7.72	66.60	-11.59
VARX	-34.81	1642.11	-2.12	-8.65	82.51	-10.48	-11.69	173.04	-6.76
Bayesian VAR	-11.75	174.05	-6.75	-19.07	217.85	-8.76	-18.13	176.26	-10.29
Strawman random walk	-11.96	144.87	-8.26	-8.63	156.34	-5.52	-5.73	31.27	-18.32
Panel (b): NDX Options									
AR	-4.79	76.00	-6.31	-12.22	291.36	-4.20	-10.99	252.28	-4.36
VAR	-3.84	77.03	-4.99	-4.84	94.69	-5.12	-4.85	61.37	-7.91
VARX	-3.31	76.46	-4.33	-5.14	81.35	-6.32	-6.66	121.74	-5.47
Bayesian VAR	-9.87	269.63	-3.66	-8.98	175.22	-5.13	-9.23	131.43	-7.03
Strawman random walk	-5.40	59.65	-9.07	-8.35	244.37	-3.42	-6.18	97.97	-6.32
Panel (c): SPX Options									
AR	-16.54	632.30	-2.62	-27.14	333.65	-8.14	-5.30	62.19	-8.53
VAR	-6.60	129.31	-5.11	-6.55	157.58	-4.16	-4.05	55.27	-7.33
VARX	-5.13	118.87	-4.32	-8.39	232.81	-3.60	-5.01	92.04	-5.45
Bayesian VAR	-9.20	238.04	-3.87	-32.30	1210.65	-2.67	-10.12	135.46	-7.47
Strawman random walk	-7.92	177.33	-4.47	-2.12	28.17	-7.53	-3.02	55.39	-5.46
Panel (d): RUT Options									
AR	-5.32	83.53	-6.38	-4.59	46.30	-9.93	-9.69	201.90	-4.80
VAR	-4.26	53.10	-8.02	-6.05	234.37	-2.59	-6.64	132.05	-5.03
VARX	-7.87	303.37	-2.59	-2.51	29.38	-8.55	-23.34	1109.07	-2.10
Bayesian VAR	-4.95	67.55	-7.33	-28.38	450.51	-6.30	-49.34	892.21	-5.53
Strawman random walk	-6.77	65.54	-10.34	-8.02	210.50	-3.81	-2.82	32.98	-8.55

Notes: The table contains the out-of-sample delta-hedge trading results after transaction costs to evaluate the economic value of five dynamic models for Dow Jones, Nasdaq, S&P 500, and Russell 2000 index options across three different deterministic models (see Section 3.2.2). The data cover the period between January 1, 2003, and December 31, 2017. The dynamic models include an AR, VAR, VARX, Bayesian VAR, given by Equations (9) - (13)*. The AR, VAR, and Bayesian VAR models take into account the lagged dynamics of the implied volatility surface of our index options. The VARX model extends these dynamics by including a factor representing the global dynamics of deviations in put-call parity obtained by principal component analysis. The Strawman random walk of* (15) *is included as the dynamic benchmark. Two other benchmarks are included: a riskless daily investment in the index options which we denote as the Buy-and-Hold strategy and an initial investment worth 1000$ in Treasury bills. The latter results in the time value of money. The three deterministic models include the five-, seven-, and nine-factor deterministic GLS models given by Equations* (2)*,* (3)*, and* (5)*, respectively. The economic measures include the average daily percentage return, its standard deviation, and the risk-adjusted return (Sharpe ratio). In obtaining our trading results, we use our implied volatility forecasts which are corrected using residuals obtained from applying generalised least squares (GLS) to our deterministic model. This correction is applied by modelling these residuals using a threshold-AR(1) model given by Equation* (7)*. The imposed transaction costs consist of 0.8 times the bid-ask spread, as suggested by Battalio and Schultz (2006).*