

ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS



MASTER THESIS IN QUANTITATIVE FINANCE
MASTER ECONOMETRICS & MANAGEMENT SCIENCE

Forecasting the implied volatility surface using dynamic factor models with GARCH disturbances

Edwin J. van Vliet (342783)

Supervisor:
Xun Gong

Co-reader:
Prof.dr. Dick J.C. van Dijk

August 25, 2017

Abstract

The implied volatility surface (IVS) explains the dynamics between different option contracts by representing the total set of implied volatilities across moneyness and maturity dimensions. In this thesis, we implement dynamic factor models to study the dynamics of the IVS. In particular, we examine whether we can improve the fit of the IVS estimated by dynamic factor models by integrating additional volatility disturbances onto their residuals. In general, we provide four key findings. First, including GARCH disturbances appears to at least mitigate the problem of poorly fitting corner IVS groups by correcting for heteroskedasticity and autocorrelation in the error terms. Second, although our extended setups have a better in-sample fit, they are outperformed by the general dynamic factor model in terms of statistical and economical forecasting performances. Third, all our dynamic factor models for the IVS only have economic value when excluding transaction costs. Fourth, we hardly report significant differences between our dynamic factor models including GARCH disturbances.

Keywords: Implied volatility surface, Dynamic factor model, State space model, Kalman filtering, Maximum likelihood estimation

Table of contents

1	Introduction	1
2	Implied Volatility Surface Data	5
2.1	Data	5
2.2	Constructing the Volatility Surface	5
2.3	Summary Statistics and Preliminary Analysis	6
3	Modeling the Implied Volatility Surface	11
3.1	Dynamic Factor Models	11
3.1.1	General Dynamic Factor Model (DFM)	11
3.1.2	Implementing GARCH Disturbances	12
3.1.3	Restricted Economic Dynamic Factor Model (RFM)	14
3.2	Other Benchmark Models for the Implied Volatility Surface	15
3.3	Estimation Procedure	16
4	Statistical Evaluation	19
4.1	Measuring Significance of GARCH Effects	19
4.1.1	Evaluation GARCH Effects Statistics	20
4.2	Statistical Measures of Predictability	21
4.3	Statistical Evaluation Results	22
4.3.1	Estimation Results	23
4.3.2	In-Sample Fit	29
4.3.3	Out-of-Sample Forecasting Performance	33
5	Economic Evaluation	39
5.1	Constructing the Trading Strategies	39
5.1.1	Trading Strategies	39
5.1.2	Transaction Costs	42
5.2	Economic Results	43
5.2.1	Trading Results before Transaction Costs	43
5.2.2	Trading Results after Transaction Costs	47
6	Conclusion	49
	Bibliography	51
	Appendix	54

1 Introduction

Option prices on financial markets contain implicit information on the volatility of the underlying expected by traders and investors. Using option pricing models like Black and Scholes (1973) for European and a binomial tree model introduced by Cox et al. (1979) for American options, expected volatilities can be derived from those option prices. Hence, these implied volatilities can be obtained by matching observed market prices with theoretical option prices and subsequently extracting associated volatilities. Due to strongly varying option prices caused by differences in strike price and time-to-maturity, it is difficult to compare and interpret option contracts based on their prices. Therefore, corresponding implied volatilities are rather used to compare and interpret these contracts more easily. If the widely used Black and Scholes (1973) option pricing model is correctly specified, the implied volatility would be the same for all available option contracts for a particular underlying. However, in practice we continuously observe varying implied volatilities depending on both maturity and moneyness levels. The implied volatility surface (IVS) is the three-dimensional collection of volatilities that are indirectly determined by a range of option contracts with different strike prices and time-to-maturities. Within this empirically non-flat surface, Rubinstein (1994) states that the volatility smile can be seen as the common pattern for a given time-to-maturity over different strike prices. Likewise, the pattern for given moneyness over different time-to-maturities is referred to as the volatility term structure. Moreover, Heston and Nandi (2000) show that due to varying market beliefs, the IVS appears to dynamically fluctuate over time. Among others, Poterba and Summers (1984) and Fleming (1998) consider the understanding of these dynamics to be highly relevant, while in many situations implied volatilities are used to gain information on expected market volatility. In option pricing applications for example, they characterize the future beliefs of market participants, whereas risk managers analyze implied volatilities in order to regulate their risk exposure. Hence, accurately forecasting the IVS could lead to better performances on the portfolios of risk and investment managers.

In case of option contracts with the same underlying, the payoffs of these contracts all depend on the performances of that specific underlying. Hence, different option contracts that share a common underlying are expected to have interrelated movements in their prices. Knowing that implied volatilities are being derived by corresponding option prices, these co-movements are also expected to be present in corresponding implied volatilities. In order to capture these co-movements, one could argue to absorb them into a model with common factors. Existing literature already introduces several different types of factor models to capture the dynamics of the IVS. For example, Dumas et al. (1998) and Goncalves and Guidolin (2006) link implied volatility to moneyness and maturity by fitting linear parametric specifications. Heston and Nandi (2000) aim to exploit predictability in the IVS using their GARCH(1,1) option pricing model. More recently, Van der Wel et al. (2016) apply maximum likelihood estimation including a collapsed filtering approach in an attempt to examine the in-sample performances of three different dynamic factor models (DFM) on S&P 500 index options. In particular, they show that in-sample plain dynamic factor models fit the IVS fairly well in the center of the surface.

However, Van der Wel et al. (2016) also report that fitting the corners of the IVS with dynamic factor models turns out to be problematic, as the residuals do not seem to behave like white noise processes. These findings might indicate misspecification of the estimated dynamic factor model, resulting in heteroskedasticity and autocorrelation in the error terms. In case of heteroskedasticity, the standard errors of a dynamic factor model are expected to be biased and differ systematically between economical stable and uncertain times. Due to this varying variability across all observations, considering a general dynamic factor model evidently results in inefficient estimates of the IVS. Hence, although the overall fit of the IVS seems to be promising, inconsistent estimates in particular IVS groups could have negative

impact on the forecasting performances of dynamic factor models. Therefore, in an attempt to force the residuals in all IVS groups to be white noise processes, one could argue to model the residuals separately using additional volatility disturbances. For example, Harvey et al. (1992) suggest to incorporate GARCH disturbances in unobserved component time series models like dynamic factor models. Due to its ability to correct for heteroskedasticity in the error terms, this additional GARCH model on the residuals might be a valuable extension to dynamic factor models for the IVS.

The contribution of this thesis to literature is twofold. On the one hand, we extend the work of Van der Wel et al. (2016) by examining out-of-sample performances of their likelihood-based dynamic factor models for the IVS. On the other hand, we attempt to further develop forecasting abilities of these likelihood-based dynamic factor models for the IVS by including additional GARCH disturbances. By implementing GARCH models on the residuals in several ways, we are the first to explore whether these extensions can have a positive impact on predicting the dynamics of the IVS. In particular, we examine whether including simple GARCH(1,1) models on the residuals of both observation and state equation could correct for autocorrelation and particularly heteroskedasticity. In case of model improvements, we might find a way to increase the percentage of correctly predicted implied volatilities, which can potentially lead to profitable strategies for traders and investors. Hence, the research question is:

Could we integrate additional volatility models onto the residuals of a dynamic factor model in order to improve its in-sample fit and out-of-sample forecasting performances of the implied volatility surface?

In this paper we use a daily data set consisting of implied volatilities on European S&P 500 index options traded on the U.S. markets over the period January 2002 until August 2015. In line with Van der Wel et al. (2016), we construct the IVS by splitting the data into 24 different moneyness-maturity groups. Moneyness is divided into six groups based on Δ values¹, whereas a division of time-to-maturity splits maturity into four groups. On each day, we select option contracts closest to the midpoint of each of the 24 IVS groups. The daily balanced panel of these selected contracts is considered in an attempt to capture the dynamics of the IVS.

In order to find the best performing predictor of the IVS, we consider three different DFM-GARCH setups along the lines of Harvey et al. (1992). First, we extend the general dynamic factor model of Van der Wel et al. (2016) by incorporating GARCH disturbances in the observation equation to correct for heteroskedasticity in its residuals. However, these phenomena in the residuals of the observation equation might be better absorbed indirectly through the residuals in the state equation. Therefore, our second DFM-GARCH setup consists of a general dynamic factor model with GARCH disturbances incorporated in the state equation. Likewise, a combination of GARCH disturbances in both the observation and state equations is considered as our third DFM-GARCH setup. For comparison purposes, we also consider two basic dynamic factor model setups. First, we use the general DFM of Van der Wel et al. (2016) for which only identification restrictions are applied. Second, we adopt their restricted economic dynamic factor model (RFM), designed to capture the key features of the surface along the moneyness and maturity dimensions. All five dynamic factor models can be presented in formulation of a basic state space model. In order to estimate these models, we adopt the work of Jungbacker and Koopman (2014) by using maximum likelihood estimation with a recursive collapsed Kalman filtering procedure. Besides, we use two benchmark models for forecasting the movements of the IVS. First, we use the

¹The Δ value measures the rate of change of the theoretical option value with respect to changes in the price of the underlying. Because the Δ is an approximation of the probability that an option ends in-the-money, this Greek can be seen as a measure for moneyness. We use this measure for moneyness while Van der Wel et al. (2016) show significantly improved results when replacing the strike price relative to the spot price measure with this Δ .

two-step vector-autoregression model from Goncalves and Guidolin (2006). As a second benchmark we use a simple random walk model for the implied volatility, in line with among others Chalamandaris and Tsekrekos (2010). Comparing the three DFM-GARCH setups with our benchmark models provides us valuable information on their forecasting performances of the IVS. All mentioned models will be examined using statistical and economical evaluation methods. At statistical level we evaluate both the in-sample fit and out-of-sample forecasting performances using root mean squared error (RMSE) and mean correct prediction of direction of change (MCP) measures. This latter measure is also of great importance in the economic evaluation, where we follow Bernales and Guidolin (2014) in using virtual trading strategies to empirically examine profitability of the models.

We report the following main conclusion. In general, we find strong indications of improved estimates of the IVS after including GARCH disturbances into a general dynamic factor model in any way. This conclusion is based on four main findings. First, we still find significant heteroskedasticity and autocorrelation in the error terms of extended dynamic factor models including GARCH disturbances. However, the results of these extended setups do show weaker significances than in our general DFM setup. Hence, although a GARCH model is originally designed to capture heteroskedasticity, we conclude that including GARCH disturbances into a dynamic factor model can mitigate its problem of heteroskedasticity and even autocorrelation in the error terms. Second, we document a better in-sample fit for our extended DFM-GARCH setups compared to the general DFM model. In particular, our DFM-GARCH setups show time series of residuals that look a bit more on white noise processes, especially for corner groups of the IVS. On the contrary, the general setup outperforms the models including GARCH disturbances significantly in an out-of-sample setting. Due to the inclusion of additional GARCH parameters, forecasts based on our DFM-GARCH setups are possibly affected by overfitting issues. Hence, although we not succeed in finding improvements in out-of-sample forecasts of the IVS, our improved in-sample estimation of the IVS indicates potential value of including GARCH disturbances into a general dynamic factor model. Third, without taking transaction costs into account in our economic simulation, all considered dynamic factor models prove to have value in their predictability of the IVS. However, after a realistic implementation of transaction costs, these potential profits disappear and are converted into great losses. In addition, we find confirming evidence that out-of-sample our DFM-GARCH setups are also outperformed by the general DFM model economically. In our economic evaluation section, we further report strongly deviating performances when applying our trading strategies within individual IVS groups. In particular, we document extremely high risks within corner groups of the IVS, explained by their poorer fit and relatively less liquid option contracts with more erratic trading patterns compared to groups in the center of the IVS. Hence, our economic evaluation results confirm that our extended DFM-GARCH setups are less effective in economically exploiting their out-of-sample forecasts compared to the general DFM model. Fourth, by comparing individual DFM-GARCH setups we hardly find any differences. In most cases, including GARCH disturbances in only the observation equation shows slightly better results than the other two variants. Hence, although we generally report insignificant differences, including GARCH disturbances in the observation equation of dynamic factor models turns out to be our preferred way to neutralize heteroskedasticity in the error terms.

A few existing papers are closely related to this thesis. Bedendo and Hodges (2009) explore the IVS dynamics with a Kalman filtering approach, but their IVS extension is only a limited analysis of different smile levels instead of a dynamic analysis over the entire surface. Hence, Van der Wel et al. (2016) lay the foundation of this thesis by offering a framework to estimate a subset of our models using a likelihood-based dynamic factor approach. They examine the in-sample performances of three different models on S&P 500 index options: a general DFM model, a restricted economic RFM model and a flexible

spline-based DFM model. Yet, they do not investigate out-of-sample performances of these promising models for the IVS. Several related papers also show other methods to examine the dynamics of the IVS. For example, Skiadopoulos et al. (2000) and Cont et al. (2002) use principal component analysis in order to explain the dynamics, whereas for similar purposes Fengler et al. (2007) and Christoffersen et al. (2009) implement semi-parametric and stochastic volatility models respectively. The economic restricted RFM model in this thesis can be traced back to the work Dumas et al. (1998), which has been subsequently extended by Goncalves and Guidolin (2006). Furthermore, together with Bernales and Guidolin (2014) and Chalamandaris and Tsekrekos (2010) this latter paper serves as an example for our out-of-sample economic evaluation section. Using a two-step OLS approach, they economically evaluate the forecasting performances of DFMs with virtual trading strategies on empirical data sets. In this two-step approach the common factors are modeled using vector-autoregressions after they have first been retrieved by OLS regression. Although their approach shows some similarities with the research of Van der Wel et al. (2016), the one-step likelihood-based DFM approach we consider seems to be a more efficient way to examine predictability in the dynamics of the IVS than their two-step OLS approach. Finally, we adopt the ideas of Harvey et al. (1992) on how to include GARCH disturbances in a time series model in order to correct for heteroskedasticity in the residuals. In their research, they show the implications these GARCH disturbances have for Kalman filtering estimation and introduce improved estimation procedures.

We contribute to literature in various ways. To begin with, and to the best of our knowledge, we are the first to introduce GARCH disturbances on factor models that forecast the dynamics of the IVS. In particular, we find a reliable way to correct for heteroskedasticity in the residuals of specifically the corners of the IVS. Hence, we provide a powerful method to transform the residuals into white noise processes, resulting in improved estimation and forecasting performances of the DFMs. Moreover, we deliver extensive out-of-sample evaluations on the general and economic restricted DFMs for the IVS recently introduced by Van der Wel et al. (2016). In contrast to common inefficient two-step approaches, we use maximum likelihood estimation with a collapsed filtering approach introduced by Jungbacker and Koopman (2014). We demonstrate our results using both statistical and economical measures, where we report empirical evidence of the profitability of these promising models.

This thesis is organized as follows. In section 2, the data and corresponding construction of the IVS are described and analyzed. Section 3 introduces the modeling setup and estimation procedures of various dynamic factor models and additional benchmark models for the IVS. In particular, we discuss a general dynamic factor model, extended dynamic factor models with GARCH disturbances and a restricted economic dynamic factor model. Furthermore, we evaluate the significance of GARCH effects in our dynamic factor models by testing the residuals on autocorrelation and heteroskedasticity in section 4. In addition, we discuss statistical evaluation methods and corresponding results for both in-sample and out-of-sample evaluation in this section. Thereafter, section 5 documents the economic profitability prospects of our modeling setups by simulating several trading strategies based on the dynamic factor models. Finally, the conclusion and discussion topics are provided in section 6.

2 Implied Volatility Surface Data

In order to test the forecasting performances of our dynamic factor models in a reliable way, we use empirical implied volatility data from the recent past. This section starts with an extensive description of the data set we use. Further, we illustrate the construction procedure of the implied volatility surface (IVS) by introducing the moneyness-maturity buckets in which the option contracts are being segmented. Eventually, section 2.3 reports key summary statistics and performs preliminary analysis on the data.

2.1 Data

In this thesis we use a daily data set consisting of implied volatilities on European S&P 500 index options traded on the Chicago Board Options Exchange (CBOE) over the period January 2, 2002 until August 31, 2015. In total, this covers 3,440 observations over a time period of nearly 14 years, including the worldwide turbulent financial crisis in 2008. The choice of 2002 as starting year is motivated by avoiding the years that have limited data due to a less active global option market or events like 9/11. By using one of the most actively traded derivative securities, we aim to provide a realistic representation of the dynamics in the total U.S. option market. Within the daily data set, OptionMetrics provides a detailed overview of several relevant characteristics of each option contract that has been on the market. These characteristics include end-of-day values for bid and ask quotes, time-to-maturity, Δ , implied volatility and strike price.

Using common knowledge that out-of-the-money (OTM) options are more frequently traded than in-the-money (ITM) options, we only select OTM options from the data set. Due to the put-call parity² this selection has no consequences for our analyses, as also pointed out by Van der Wel et al. (2016). Furthermore, we filter the data using the similar four restrictions as Barone-Adesi et al. (2008). In detail, options are deleted when time-to-maturity is lower than 10 or higher than 360 days, when implied volatility is above 70%, when option price is below \$0.05, or when any values for implied volatility or Δ are missing. By removing incomplete and inactive options, these filters make sure that our model estimates only consider relevant option contracts and are not affected by unrealistic or extreme trades.

Although actual prices and transactions costs of individual options are unknown, we approximate them using the difference between bid and ask quotes defined as the bid-ask spread. To determine the price of an option, we calculate the average of both corresponding quotes. In order to set up economic trading strategy benchmarks, we consider an additional second data set from OptionMetrics. This daily data set contains prices of the S&P 500 index fund. Finally, on account of the economic evaluation section we consider the data library of *French* (2017) to select daily global riskfree rates. On average, the yearly riskfree rate over our full sample period amounts 1.31%.

2.2 Constructing the Volatility Surface

By spanning the data over both moneyness and maturity levels, we construct a three-dimensional IVS. In order to arrange the large cross-section of options, among others Bollen and Whaley (2004) and Barone-Adesi et al. (2008) group the options into several moneyness-maturity categories. In line with their ideas, and following the approach of Van der Wel et al. (2016), we construct the IVS by dividing the data over 24 groups. Hence, we adopt their attempts to strike a balance between forming completely filled IVS buckets and representing common movements of options in the large cross-section of data. The maturity dimension is split up into four groups based on time-to-maturity: 10-45 days, 45-90 days,

²Under the assumption that call option x matches put option y on moneyness and maturity levels, Δ_x is per definition equal to $1 + \Delta_y$.

90-180 days and 180-360 days. In order to select a proper moneyness measure, we again follow Bollen and Whaley (2004) in their choice for the option's Δ . They report that the more often used strike price relative to spot price ratio fails to account for the fact that the volatility of the underlying asset also affects the option's likelihood to be in-the-money. On the contrary, an option's Δ can be interpreted as the risk-neutral probability that the option will be in-the-money when it expires. Hence, this proves to be a better measure for moneyness. Moreover, Van der Wel et al. (2016) confirm this statement by showing that factor models with Δ as moneyness measure outperform factor models with strike price relative to spot price ratios as moneyness measure. The moneyness criteria form at-the-money (ATM), out-of-the-money (OTM) and deep out-of-the-money (DOTM) categories for both put and call options. In detail, we form the following moneyness categories for call options: $0.375 < \Delta < 0.5$ for ATM call, $0.125 < \Delta < 0.375$ for OTM call and $0 < \Delta < 0.125$ for DOTM call. Likewise, put option categories are as follows: $-0.5 < \Delta < -0.375$ for ATM put, $-0.375 < \Delta < -0.125$ for OTM put and $-0.125 < \Delta < 0$ for DOTM put.

In line with Van der Wel et al. (2016), on each day in each group we select an option contract nearest to the midpoint. In detail, from all available contracts within a group we select the option that is closest to the center regarding both moneyness and maturity dimensions.³ Clearly, these selected options have different moneyness and maturity measures over time due to variation in Δ and seasonality in expiry dates. An alternative could be to construct time series for exactly the midpoint implied volatilities of the IVS groups by smoothing the observed implied volatilities. However, we do not expect our implemented construction approach to have negative impact on the in-sample estimation evaluation. Moreover, our generated time series for statistical and economical out-of-sample evaluation are based on the total data set of all available individual option contracts. Therefore, we are able to select the same contract on two consecutive trading days. More specifically, we make sure that we only consider option contracts that are available on subsequent trading days by tracking and matching corresponding option IDs. Hence, our out-of-sample evaluation procedures do not suffer from inconsistent implied volatility surfaces as well. Within each group, we daily select an option nearest to the midpoint from on average 18 available contracts. In case of groups with no available contracts to select from, we fill groups by selecting the closest contract from all available contracts that day. By concatenating all selected contracts, we form a surface that represents the complete IVS over the full time period.

2.3 Summary Statistics and Preliminary Analysis

Prior to defining and estimating the dynamic factor models, we first investigate the data by analyzing the constructed surfaces and corresponding statistics in various ways. First, in order to show stylized facts of the IVS, we present plots of two different days within the sample period in figures 1 and 2. Both figures are evidently in contradiction with the ideas of Black and Scholes (1973), who assume constant volatility across all moneyness and maturity groups. Compared to each other, the figures display varying patterns due to related economic (in)stable times. An example of the IVS during the worldwide financial crisis in 2008 is shown in figure 1. During these uncertain economic times the S&P 500 index options prove to have high volatility levels. Meanwhile, the volatility levels of days in more flourishing economic times as in figure 2 are considerably lower. Hence, these figures show confirm differences in volatility levels between stable and crisis periods. Furthermore, both figures present evidence of a common pattern in implied volatilities across the moneyness dimensions, also referred to as the volatility smile.⁴ In detail, for

³The distance between center of the group and an individual contract is defined as the sum of squared deviations of both time-to-maturity and Δ . Due to scale differences, we put ten times more weight on the distance of Δ in order to compare both deviations appropriately.

⁴To be precise, this continuously diminishing asymmetric pattern of the IVS across the moneyness groups is referred to as the volatility smirk. The volatility smile is actually defined as the symmetric pattern in which the implied volatilities

a given maturity category the IVS is downward sloping along the moneyness dimension, where the IVS is lowest for the DOTM-Call categories. Correspondingly, the pattern across the maturity dimension is referred to as the volatility term structure. However, contrary to the volatility smile this pattern differs in slope between stable and unstable economic times. Where figure 1 shows downward slopes across the maturity dimension in times of high volatility, figure 2 shows upward slopes in times of low volatility. Following Van der Wel et al. (2016), these term structure dynamics can be explained by the mean-reversion property of volatilities.

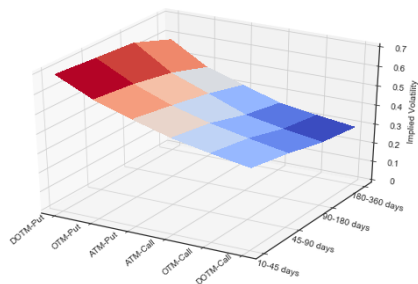


Figure 1: IVS on November 13, 2008

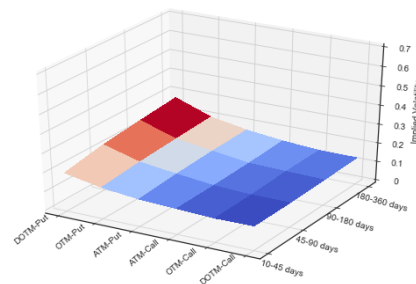


Figure 2: IVS on March 27, 2015

Notes: These figures show examples of implied volatility surfaces (IVS) on two days in the sample period, consisting of implied volatilities from S&P 500 index options. The implied volatilities are displayed across the six moneyness and four maturity groups, as set up in section 2.2.

Furthermore, an overview of the summary statistics of all selected option contracts is given in table 1. Here, the sample means and standard deviations of four different variables within each IVS group are given. The table provides us several insights regarding the data. First, we observe two patterns in the dynamics of the option prices by considering the mid-quotes of the selected contracts. ATM options are traded at higher prices than OTM and particularly DOTM options. This pattern of higher prices for options closer to being in-the-money is intimately related to our choice for Δ as measure for moneyness. Since an option's Δ can be considered as a Greek for the probability that the option will end up in-the-money at expiration, deeper out-of-the-money options generally have lower prices. Besides, we observe a pattern that longer maturities are accompanied by higher option prices. This pattern can be simply declared by the time value of money, which implies that options with longer maturities have higher probabilities to end up in-the-money and are therefore traded at higher prices. Next, considering the implied volatility variables the statistics are in line with the general findings of figures 1 and 2. In particular, the volatility smile effect turns out to be present over the full sample period, as implied by the downward sloping average implied volatilities across the moneyness groups. In addition, the table shows Δ and maturity (in days) are close to the midpoints of corresponding groups. Therefore, the table provides evidence that our selection of contracts closest to the midpoints of the IVS groups is working properly.

By extracting the surface over the full sample period, we provide indications of strong co-movements across the surface by demonstrating time series of the average implied volatilities within the 24 IVS groups in figure 3. In addition, the average implied volatility across all groups is highlighted in blue. Volatility peaks can be observed due to the presence of several striking events, like the Gulf War in 2002-2003, the financial crisis in 2008, the European sovereign debt crisis in 2010 and the debt-ceiling crisis in 2011. Since 2013, the average volatility no longer exhibits high peaks due to little uncertainty for both deeper out-of-the-money call and put options slope upward. However, in literature the volatility smile is widely accepted as definition for this asymmetric pattern. Hence, we adopt this interpretation of the volatility smile.

Table 1: Summary Statistics

	10-45 days		45-90 days		90-180 days		180-360 days		
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	
DOTM Put	Mid-Quote	2.68	3.32	4.49	3.47	6.56	7.21	9.79	3.82
	IV	0.281	0.106	0.291	0.097	0.296	0.089	0.298	0.079
	Δ Maturity	-0.0638 27.7	0.0218 7.9	-0.0641 67.4	0.0181 8.4	-0.0633 133.6	0.0221 22.9	-0.0640 268.2	0.0118 25.6
OTM Put	Mid-Quote	10.71	4.95	18.09	6.42	27.03	9.23	41.03	13.18
	IV	0.219	0.090	0.226	0.082	0.231	0.074	0.234	0.065
	Δ Maturity	-0.2503 27.7	0.0166 7.9	-0.2492 67.4	0.0166 8.3	-0.2497 133.9	0.0245 22.6	-0.2497 269.4	0.0169 24.8
ATM Put	Mid-Quote	20.61	8.48	34.63	11.77	51.36	16.42	77.87	23.53
	IV	0.195	0.082	0.198	0.072	0.201	0.065	0.204	0.056
	Δ Maturity	-0.4369 27.7	0.0177 8.0	-0.4374 67.4	0.0182 8.3	-0.4370 133.9	0.0203 22.7	-0.4373 269.0	0.0169 24.9
ATM Call	Mid-Quote	18.07	6.75	29.08	8.77	41.27	11.79	58.70	15.61
	IV	0.183	0.076	0.184	0.066	0.186	0.058	0.189	0.051
	Δ Maturity	0.4374 27.7	0.0187 8.0	0.4381 67.5	0.0202 8.4	0.4383 133.9	0.0221 22.7	0.4383 269.1	0.0189 24.9
OTM Call	Mid-Quote	7.66	2.85	12.25	3.83	17.36	5.16	24.91	7.00
	IV	0.169	0.070	0.167	0.060	0.168	0.053	0.170	0.046
	Δ Maturity	0.2498 27.7	0.0189 7.9	0.2507 67.4	0.0248 8.3	0.2507 133.9	0.0244 22.6	0.2521 269.0	0.0237 24.7
DOTM Call	Mid-Quote	1.30	0.58	2.29	1.79	3.28	2.09	5.50	5.83
	IV	0.158	0.064	0.152	0.054	0.152	0.047	0.152	0.040
	Δ Maturity	0.0627 27.6	0.0148 7.9	0.0689 67.1	0.0380 8.5	0.0705 133.9	0.0365 22.8	0.0806 266.5	0.0575 27.7

Notes: This table provides summary statistics for the full data set, consisting of selected implied volatilities across moneyness and maturity categories as set up in section 2.2. Within each IVS group, the table shows both mean and standard deviations (Std) of the Mid-Quote (in US Dollars), implied volatility (IV), Δ and maturity (in days) of the selected S&P 500 index options. The sample period is January 2, 2002 - August 31, 2015.

compared to the years before. Hence, we evidently observe larger implied volatilities during extreme events compared to during stable times.

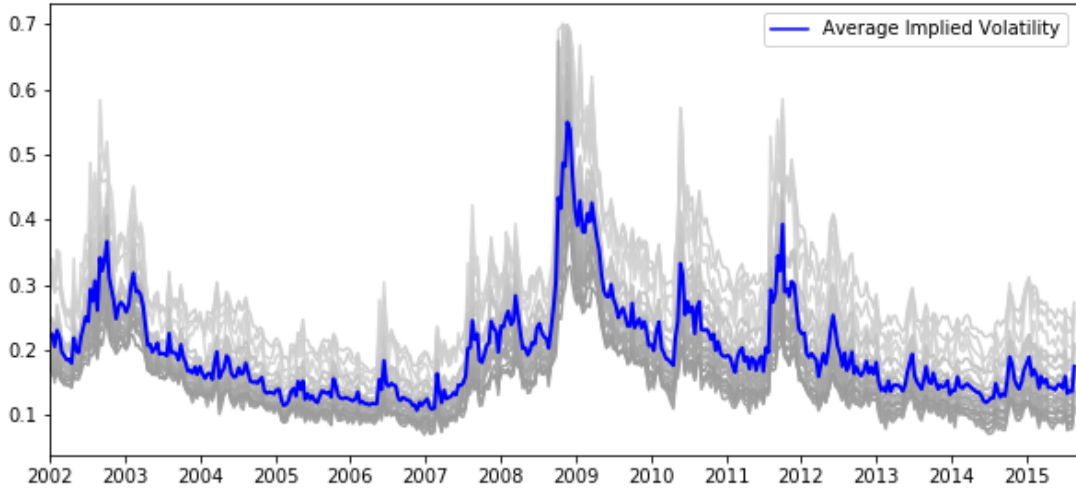


Figure 3: Average Implied Volatility

Notes: This figure shows time series of the average implied volatilities, both for the entire data set across all groups (highlighted in blue) as well as within each IVS group (displayed in shades of grey). For clarity purposes, we compress space by showing snapshots of every 8 observations.

These strong co-movements are confirmed by figure 4, which displays all available slopes of the volatility smile and term structure. Here, the slope of the volatility smile is defined as the implied volatility of selected DOTM-Put options minus the implied volatility of selected DOTM-Call options within each maturity group. Likewise, the slope of the term structure within each moneyness group is defined as the implied volatility of selected options with the longest maturity minus the implied volatility of selected options with the shortest maturity. As can be seen in the upper figure, the slope of the volatility smile is positive for all maturity groups, with higher values during economically uncertain periods. Even during these events, the slopes of all maturity groups show similar movements. Across the moneyness groups, the bottom figure displays the slopes of the volatility term structure. These slopes contain both positive and negative values, which is in line with our mean-reversion findings from figures 1 and 2. But consistent with the slopes of the volatility smile, they also show little differences and many similar patterns. Hence, figure 4 displays evidence of strong co-movements in both the volatility smile and term structure.

Moreover, by analyzing the cross-correlations of the IVS groups in table 12 in the appendix, we find substantially high cross-correlations in all cases. In particular, all cross-correlations are greater than 0.83, whereas cross-correlations between groups with similar moneyness or maturity measures are even greater than 0.9. Logically, we also observe a trend that cross-correlations between groups become smaller when their moneyness or maturity measures move further apart. Note that for clarity reasons, we drop the most intermediate moneyness groups and only present the ATM and DOTM categories.

In consequence, all previous findings present evidence of strong co-movements in the average implied volatility and slopes of volatility smile and term structure. This suggests that it might be appropriate to use a model with common factors in our attempt to model and forecast the IVS. Hence, in order to be able to determine this with certainty, we eventually perform principal component analysis on the data.

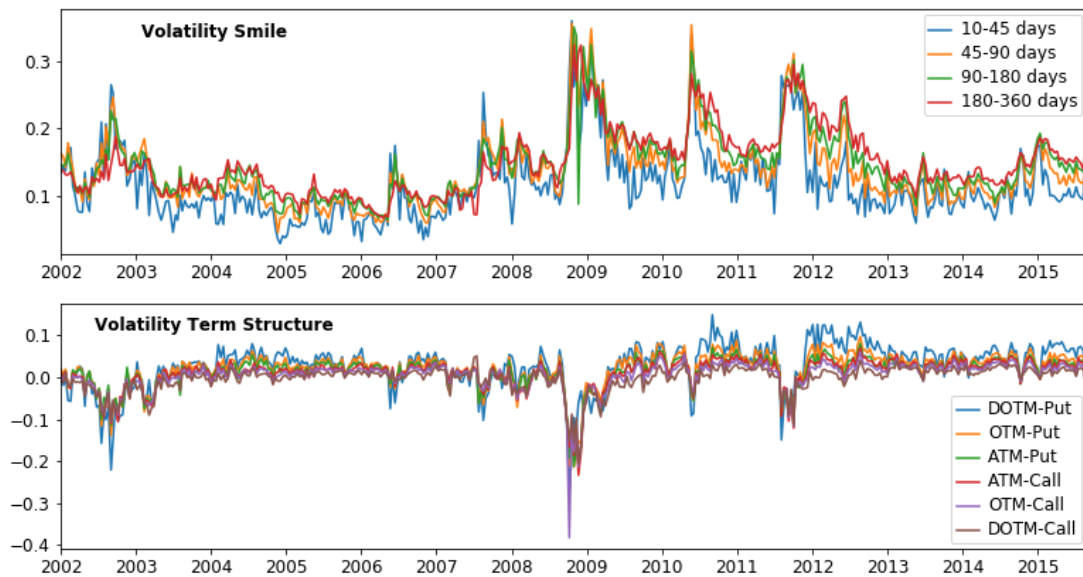


Figure 4: Slopes of Volatility Smile and Volatility Term Structure

Notes: These figures show time series of the slope of the volatility smile and the volatility term structure from selected S&P 500 index options. Here, the slope of the volatility smile is defined as the implied volatility of selected DOTM-Put options minus the implied volatility of selected DOTM-Call options within each maturity group. Likewise, the slope of the term structure within each moneyness group is defined as the implied volatility of selected options with the longest maturity (180-360 days) minus the implied volatility of selected options with the shortest maturity (10-45 days). For clarity purposes, we compress space by showing snapshots of every 8 observations.

Table 11 in the appendix provides percentages of variances explained by the principal components that result from this principal component analysis on our S&P 500 data. For the first five out of 24 principal components, the table reports both the variation explained by each individual principal component as well as the cumulative percentages. Evidently, the first principal component already explains a majority of almost 96% of the total variation. Furthermore, second and third principal components are responsible for more than 2% and 1% respectively, resulting in a total explained variation of 99%. Hence, the presence of various common factors in the IVS data are confirmed by these findings. In addition, table 11 in the appendix also provides corresponding autocorrelations for lags 1, 5 and 10, and partial autocorrelations for lags 1, 2 and 3. Again, we observe significantly high persistence in the first three principal components, motivating the use of a vector-autoregressive model. Besides, the partial autocorrelations strongly decline for lags greater than 1, indicating best practice for selecting the order equal to one. Hence, consistent with the ideas of among others Dumas et al. (1998) and Goncalves and Guidolin (2006), these findings suggest that the first three principal components are persistent and explain the majority of the total variation. In conclusion, the results of our principal component and other preliminary analyses support the presence of a strong factor structure in the IVS and validates our approach of using dynamic factor models with three factors.

3 Modeling the Implied Volatility Surface

In this thesis, we test if we can find better predictions of the IVS. To do so, we aim to improve the fit of dynamic factor models by adding GARCH disturbances and compare corresponding results with several benchmark models. We start this section by introducing dynamic factor models and the additional implementation of GARCH disturbances in section 3.1. In section 3.2, definitions of various benchmark models are given. Eventually, we discuss corresponding estimation procedures of the dynamic factor models in section 3.3.

3.1 Dynamic Factor Models

As stated in section 2.3, fitting dynamic factor models is a convenient way to capture the dynamics of the IVS. In particular, existing literature shows that co-movements in the cross-sections of the IVS can be successfully captured by models with common factors. In this section, a detailed overview of the dynamic factor models we consider can be found. First, we provide a setup of the general dynamic factor model (DFM) adopted from Van der Wel et al. (2016) in section 3.1.1. Next, in section 3.1.2 we introduce additional volatility disturbances for the residuals of dynamic factor models. More specifically, we follow Harvey et al. (1992) by adding GARCH disturbances to our time series models in various ways. To ensure a better understanding of this extension to dynamic factor models, we eventually consider the restricted economic dynamic factor model (RFM) with time-varying loading matrix in section 3.1.3.

3.1.1 General Dynamic Factor Model (DFM)

In order to cross-sectionally fit our dynamic factor models, we first stack implied volatilities of all 24 IVS groups in the (24×1) observation vector y_t , defined by

$$y_t = \begin{pmatrix} IV_{\tau_1, m_1, t} \\ \vdots \\ IV_{\tau_T, m_1, t} \\ IV_{\tau_1, m_2, t} \\ \vdots \\ \vdots \\ IV_{\tau_T, m_M, t} \end{pmatrix} \quad (1)$$

where on day t the implied volatility for a contract with time-to-maturity τ_i and moneyness m_j is given by $IV_{\tau_i, m_j, t}$. As stated in section 2.2, our IVS is constructed along the lines of Van der Wel et al. (2016) by using $i = 1, 2, \dots, T$ and $j = 1, 2, \dots, M$ with $T = 4$ and $M = 6$. Van der Wel et al. (2016) plug this vector into the observation equation of their general DFM model, given by

$$\begin{aligned} y_t &= \Lambda f_t + \epsilon_t & \epsilon_t &\sim N(0, \Sigma_\epsilon) \\ f_t &= \mu + \Phi(f_{t-1} - \mu) + \eta_t & \eta_t &\sim N(0, \Sigma_\eta) \end{aligned} \quad (2)$$

where these equations are referred to as observation equation and state equation, respectively. Here, f_t denotes the vector of latent dynamic factors, which is included in a vector-autoregressive (VAR) model of order one as a result of our conclusions from the (partial) autocorrelations in section 2.3. In particular, supported by our findings regarding principal component analysis in section 2.3, we select three as the number of factors to include in our dynamic factor models. Hence, this results in a (24×3) loading matrix Λ , a (3×1) vector of latent factors f_t and a (24×24) covariance matrix Σ_ϵ of the normal distributed vector

of measurement errors ϵ_t . Likewise, the state equation contains a (3×1) vector of factor intercepts μ , a (3×3) transition matrix Φ and (3×3) covariance matrix Σ_η of the normal distributed vector of factor innovations η_t .

Furthermore, the general DFM model requires identification restrictions on its loading matrix in order to enable proper estimation. In their research, Geweke and Zhou (1996) propose to restrict the top (3×3) part of the loading matrix Λ to an identity matrix. However, Van der Wel et al. (2016) argues that in case of implied volatility surfaces, it seems more sufficient to restrict specific elements of the loading matrix in order to strengthen their interpretations. Specifically, they propose restrictions to force the three latent factors to capture the level, term structure effect en volatility smile, respectively. This can be accomplished by setting certain loading elements λ of shortest (longest) maturity and call (put) moneyness categories to minus (plus) one. Hence, we adopt this idea of restricting four rows of Λ by setting

$$\begin{bmatrix} \lambda_{2T+1,1} & \lambda_{2T+1,2} & \lambda_{2T+1,3} \\ \lambda_{3T,1} & \lambda_{3T,2} & \lambda_{3T,3} \\ \lambda_{3T+1,1} & \lambda_{3T+1,2} & \lambda_{3T+1,3} \\ \lambda_{4T,1} & \lambda_{4T,2} & \lambda_{4T,3} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \quad (3)$$

where we still have $T = 4$. Similar to all upcoming models, further details on the estimation procedure of this general DFM model can be found in section 3.3.

Although Van der Wel et al. (2016) show that this model generally fits the IVS fairly well, they also show certain limitations. Particularly, they report that fitting the corners of the IVS appears to be complicated due to residuals that do not seem to match with white noise processes. Moreover, section 4.3.2 confirms these findings by showing heteroskedasticity and autocorrelation in the residuals of the corners of the IVS. Normally if a model is properly specified, corresponding residuals follow white noise processes. Hence, we can conclude that the general DFM model is not fully effective in fitting the IVS and consequently provides room for improvement. Therefore, in section 3.1.2 we introduce a dynamic factor model in combination with an additional volatility model for the residuals.

3.1.2 Implementing GARCH Disturbances

In this thesis, we report supporting evidence for the statement of Van der Wel et al. (2016) that fitting the corners of the IVS with a general DFM model turns out to be problematic. These signs of heteroskedasticity and autocorrelation might indicate that the residuals have to be modeled individually. In this way, we attempt to improve our estimation and forecasting performances by converting the residuals to white noise processes. For this purpose, we adopt the ideas of Harvey et al. (1992) by including a GARCH(1,1) model onto the residuals of our dynamic factor model. Initially, Bollerslev (1986) introduced a generalized autoregressive conditional heteroskedasticity (GARCH) model in order to make the unconditional variance constant.⁵ In particular, Bollerslev (1986) provides a GARCH(p,q) framework which allows for a longer memory with q residuals and a more flexible lag structure using p autoregressive lags. However, motivated by our significance tests on lagged autocorrelations of the error terms and corresponding GARCH errors⁶, we select a GARCH(1,1) model that only takes the first lags into account. Harvey et al. (1992) already provides an extensive framework on how to incorporate this GARCH(1,1) model into unobserved component time series models and how to deal with corresponding

⁵The GARCH model is a generalized version of the autoregressive conditional heteroskedasticity (ARCH) model as proposed by Engle (1982).

⁶Detailed significance test statistics are not reported in this thesis and are available upon request.

implications during estimation procedures. Therefore, we follow their ideas by integrating additional GARCH(1,1) model specifications and estimation methods into our DFM framework. In an attempt to find the best possible fit of the IVS, we attempt to explore various setups. For instance, Koopman et al. (2010) specify the overall volatility as a GARCH process by decomposing the disturbance vectors of both the observation and state equations. Inspired by their work, we therefore select three different setups. First, we add GARCH disturbances directly to the implied volatility residuals in the observation equation. Motivated by the possibility that heteroskedasticity in the implied volatility residuals is passed through via the latent factors, we add GARCH disturbances to the factor residuals in the second setup. The third setup consists of a combination of GARCH disturbances on both the observation error terms and the state error terms. Hence, by incorporating these varying GARCH disturbances we provide extended DFM-GARCH models that are considered to be able to correct for heteroskedasticity in the residuals and subsequently improve estimation and forecasting performances of the IVS.

Our extended DFM-GARCH model is given by

$$\begin{aligned} y_t &= \Lambda f_t + \Gamma \epsilon_t + \epsilon_t^* & \epsilon_t^* &\sim NID(0, \Sigma_\epsilon^*) \\ f_t &= \mu + \Phi(f_{t-1} - \mu) + \Psi \eta_t + \eta_t^* & \eta_t^* &\sim NID(0, \Sigma_\eta^*) \end{aligned} \quad (4)$$

where GARCH disturbances are incorporated in both observation and state equations. Specifically, our initial residuals ϵ_t and η_t are modeled separately by subdividing them into $\Gamma \epsilon_t + \epsilon_t^*$ and $\Psi \eta_t + \eta_t^*$. Here, the disturbance vectors ϵ_t^* (24×1) and η_t^* (3×1) are assumed to be normally and serial independently distributed. In order to avoid identification issues, we again assume the covariance matrix Σ_ϵ^* to be diagonal. The GARCH effects are introduced via the scalar disturbances ϵ_t and η_t and corresponding loading matrices Γ (24×1) and Ψ (3×1). Across the entire IVS, option prices and corresponding implied volatilities depend on the volatility of the underlying. Hence, a shock in the volatility of the underlying affects the implied volatilities of all IVS groups. Due to the high-dimensionality of the observation vector y_t , we therefore expect these common scalar disturbances to be proper instruments to include GARCH disturbances into the observation equation. Although common scalar disturbances for the state equation enforces the same volatility dynamics on the latent factors, we also introduce GARCH by means of common scalar disturbances in the state equation due to the factor structure of its residuals. In particular, we perform PCA on the residuals of the state equation η_t in order to determine whether these residuals exhibit a common factor.⁷ Evidently, the first principal component explains more than 84% of the total variation, indicating a factor structure for the residuals and motivating our inclusion of GARCH disturbances in the state equation by means of common scalar disturbances. In detail, the scalar disturbances are defined by

$$\epsilon_t = h_t^{1/2} \epsilon_t^\dagger \quad \text{and} \quad \eta_t = q_t^{1/2} \eta_t^\dagger \quad (5)$$

where $\epsilon_t^\dagger \sim NID(0, 1)$ and $\eta_t^\dagger \sim NID(0, 1)$ are white noise processes. Furthermore, the origin of the GARCH effects can be found in the variables h_t and q_t , defined by

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 h_{t-1} \quad (6)$$

$$q_t = \gamma_0 + \gamma_1 \eta_{t-1}^2 + \gamma_2 q_{t-1} \quad (7)$$

where we assume that $\alpha_1 + \alpha_2 < 1$ and $\gamma_1 + \gamma_2 < 1$. In consequence, these GARCH disturbances are included in both observation and state equations of the DFM-GARCH model from equation 4. For

⁷Detailed results of the principal component analysis are not reported in this thesis and are available upon request.

future estimation and filtering purposes, it seems useful to rewrite our DFM-GARCH model into state space formulation. Hence, the observation equation can be given by

$$y_t = \begin{bmatrix} \Lambda & 0 & \Gamma \end{bmatrix} f_t^x + \epsilon_t^* \quad \epsilon_t^* \sim NID(0, \Sigma_\epsilon^*) \quad (8)$$

whereas the state equation can be rewritten as

$$f_t^x = \begin{bmatrix} f_t \\ \eta_t \\ \epsilon_t \end{bmatrix} = \begin{bmatrix} \Phi & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ \eta_{t-1} \\ \epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} I & \Psi & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t^* \\ \eta_t \\ \epsilon_t \end{bmatrix} \quad (9)$$

with $\begin{bmatrix} \eta_t^* \\ \eta_t \\ \epsilon_t \end{bmatrix} \sim NID \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_\eta^* & 0 & 0 \\ 0 & q_t & 0 \\ 0 & 0 & h_t \end{bmatrix} \right)$

3.1.3 Restricted Economic Dynamic Factor Model (RFM)

In previous dynamic factor models, we consider the latent factors as unidentified components. However, regarding the setup of DFM models in relation to the three-dimensional structure of the IVS, it is a common thought to give these factors certain interpretations. In particular, in related literature it is widely used to consider the latent factors as representations of the volatility smile and the slope of the volatility term structure. For example, Christoffersen et al. (2015) propose a framework that cross-sectionally regresses implied volatility on moneyness and maturity dimensions in order to obtain level, volatility smile and term structure components. This regression, frequently adopted by among others Van der Wel et al. (2016), is given by

$$IV_{\tau_i, m_j, t} = l_t + \tau_i c_t + m_j s_t + \epsilon_{i, j, t} \quad (10)$$

where l_t denotes the implied volatility level, c_t denotes the volatility term structure effect and s_t denotes the volatility smile. Related literature provides several alternatives to implement a restricted dynamic factor model for the IVS, including additional factors for squared moneyness s_t^2 and the interaction between moneyness and time-to-maturity $c_t s_t$. However, for comparison reasons regarding both the work of Van der Wel et al. (2016) and our general three-factor DFM model, we consider a basic RFM model containing three factors to represent the level, term structure and volatility smile effects. Hence, by stacking these variables into the latent state vector $f_t = (l_t, c_t, s_t)'$, we are able to adopt this regression into our general DFM setup. Therefore, we restrict the factor loading matrix from equation 2 by setting

$$\Lambda_t = \begin{pmatrix} 1 & \tau_{1,t} & m_{1,t} \\ \vdots & \vdots & \vdots \\ 1 & \tau_{T,t} & m_{1,t} \\ 1 & \tau_{1,t} & m_{2,t} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & \tau_{T,t} & m_{M,t} \end{pmatrix}$$

where Λ_t is specified time-varyingly by taking actual time-to-maturity values τ_i and moneyness measures Δ of the selected contracts within all groups.⁸ For out-of-sample forecasting purposes in a later

⁸Due to its strong outperformance of the RFM setup with constant loading matrix in the research of Van der Wel et al. (2016), we only consider the RFM model with time-varying loading matrix in this thesis.

stage, we simply consider current loading matrices in order to avoid complex forecasting procedures for one-day-ahead loading matrices. Accordingly, we force the first factor to capture movements in the level of the IVS by de-meaning second and third column on a daily basis. Hence, this enables us to allocate the latent factors and interpret the dynamics of the IVS more specifically. Another main advantage of this restricted economic dynamic factor model (RFM) is the relatively low number of parameters to be estimated. By restricting the full loading matrix we obtain a deterministic matrix Λ_t , which makes it able to eliminate up to 72 parameters in our general DFM setup. Hence, compared to previous DFM(-GARCH) models this RFM model is expected to mitigate the natural problem of overfitting due to numerous parameters. However, Van der Wel et al. (2016) report that in-sample this RFM model is strongly outperformed by the general DFM model, motivating our choice of adding GARCH disturbances to the general DFM model. Therefore, and to efficiently obtain a better understanding of our GARCH extensions considering limited time, we only compare RFM model to the DFM-GARCH models in future sections.

3.2 Other Benchmark Models for the Implied Volatility Surface

In this thesis, we focus on the IVS forecasting performances of general and extended dynamic factor models. However, in order to properly analyze and compare their results, we select two additional benchmark models for the IVS. First, we include a simple random walk (RW) model on the individual implied volatilities, which is frequently used in related literature by among others Konstantinidi et al. (2008) and Bernales and Guidolin (2014). In general, this model assumes that for each individual contract today's implied volatility is the best forecast of tomorrow's implied volatility. Although this model seems fairly naive, Harvey et al. (1992) state that discussions with practitioners reveal that the RW model is widely used by traders in index option markets. Hence, we consider this as a useful standard to compare our DFM-GARCH models with.

As a second benchmark for the IVS, we select a modified version of the two-step framework from Goncalves and Guidolin (2006), where they first obtain the factors by cross-sectionally performing ordinary least squares (OLS) and then model these factors using a vector-autoregression (VAR) model. In detail, this comes down to a first step in which we use OLS to estimate a cross-sectional model for the IVS on a daily basis, given by

$$y_k = \beta_0 + \beta_1 m_k + \beta_2 \tau_k + e_k \quad (11)$$

where we have that $k = 1, \dots, K$, with k denoting a specific option of all K available contracts in each daily cross-section. Again, m_k and τ denote the maturity and moneyness (Δ) measures of option k , and e is the random error term. Meanwhile, we assemble the OLS coefficients into a vector of coefficients $\beta = (\beta_0, \beta_1, \beta_2)'$. By considering this three factor setup, we attempt to provide a proper benchmark model for our likelihood-based dynamic factor models. As a second step, we then fit a VAR model to the time series of OLS estimates β . Specifically, we set up a VAR model given by

$$\beta_t = \nu + \sum_{z=1}^p \Theta_z \beta_{t-z} + v_t \quad \text{with} \quad v_t \sim NID(0, \Omega) \quad (12)$$

where Θ_z denotes the VAR loading matrix at lag z , and v independently follows a normal distribution with covariance matrix Ω . In order to realize appropriate comparability with our DFM(-GARCH) and RFM models, we consider the lag length p to be fixed at 1. Except the two-step framework we consider to estimate the model, this setup can be conceptually seen as an equivalent of the likelihood-based RFM model introduced in section 3.1.3. Hence, this benchmark model is an alternative way to include factor

dynamics in the IVS and serves as proper comparison material for our DFM-GARCH setups.

3.3 Estimation Procedure

So far, comprehensive explanations of our basic and extended models are discussed. However, in order to apply these models properly to forecast the dynamics of the IVS, a detailed outline of complementary estimation procedures is required. Therefore, in this section we provide several methods to evaluate the latent factors and to estimate the dynamic factor models efficiently. In particular, we describe our estimation approach consisting of Kalman filtering and maximum likelihood estimation. In addition, we implement the collapsed approach of Jungbacker and Koopman (2014) in order to improve efficiency of our estimation procedures. Finally, we clarify our setup further by explaining identification restrictions and applied methods to obtain starting values for the parameters.

To begin with, we rewrite our DFM and DFM-GARCH models from equations 2 and 4 into general state space form, given by

$$\begin{aligned} y_t &= Zf_t + \epsilon_t & \epsilon_t &\sim N(0, H) \\ f_t &= d + Tf_{t-1} + \eta_t & \eta_t &\sim N(0, Q) \end{aligned} \quad (13)$$

where Z represents loading matrix Λ from the general DFM model in equation 2. Likewise, in our DFM-GARCH models Z represents the combined loading matrix $[\Lambda \ 0 \ \Gamma]$ from equation 4. For both cases, T represents transition matrix Φ , H and Q are covariance matrices Σ_ϵ and Σ_η , and d is defined as $d = (I - \Phi)\mu$. This state space model is estimated by applying a Kalman filter, which is a recursive formula running forwards through time in order to estimate latent factors from past observations. By adopting the Kalman filtering framework of Durbin and Koopman (2012), we define the vector of all observations up to time s as $Y_s = \{y_1, \dots, y_s\}$ and initial state distribution $f_1 \sim \mathcal{N}(a_1, P_1)$. Then, subsequent means and variances of the latent factors are defined by $a_{t+1} = \mathbb{E}(f_{t+1}|Y_t)$ and $P_{t+1} = \text{Var}(f_{t+1}|Y_t)$. Hence, the optimal filtered estimates $a_{t|t}$ and $P_{t|t}$ and optimal predicted estimates a_{t+1} and P_{t+1} can be achieved by following the recursive Kalman filtering procedure, given by

$$\begin{aligned} v_t &= y_t - Za_t & F_t &= ZP_tZ' + H \\ a_{t|t} &= a_t + P_tZ'F_t^{-1}v_t & P_{t|t} &= P_t - P_tZ'F_t^{-1}ZP_t \\ a_{t+1} &= Ta_{t|t} + d & P_{t+1} &= TP_{t|t}T' + Q \end{aligned} \quad (14)$$

where the Kalman Gain is defined as $K_t = TP_tZ'F_t^{-1}$. Because of the inclusion of prediction error decomposition, we are also capable to evaluate corresponding log-likelihood function simultaneously. Hence, if we define the data vector $y = (y'_1, \dots, y'_T)$, the log-likelihood function $\ell(y|\psi)$ for the Gaussian density can be evaluated by

$$\ell(y|\psi) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log |F_t| - \frac{1}{2} \sum_{t=1}^T v'_t F_t^{-1} v_t \quad (15)$$

where N is the dimension of y_t , T is the length of the sample, and the definitions of F_t and v_t are provided within the Kalman filtering recursion in equation 14. By maximizing this log-likelihood function using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm of Battiti and Masulli (1990), our procedure results in optimal estimates for all unknown parameters ψ . As a typical quasi-Newton method, this BFGS maximum likelihood method solves our nonlinear optimization problems by considering three stopping criteria. Besides reaching the maximum number of iterations, the optimizer stops when either the decrease of the objective function or the norm of the projected gradient becomes marginal.

Initially, we evaluate our dynamic factor models by performing the standard Kalman filtering algorithm of Koopman and Durbin (2000). However, Jungbacker and Koopman (2014) state that in many existing DFM applications, high-dimensional panels of time series and resulting large numbers of parameters make such an approach infeasible. In particular, within our dynamic factor models consisting of 24 groups, the initial filtering approach requires to invert the (24×24) matrix F_t for the log-likelihood evaluation in equation 15. Hence, in addition to the standard Kalman filtering approach we implement the collapsed filtering approach, recently developed by Jungbacker and Koopman (2014). By splitting our time series into a low-dimensional vector series and a high-dimensional vector series, the number of dimensions in the log-likelihood function diminishes and we manage to improve computational efficiency sufficiently. More specifically, we start with carrying out signal extraction for f_t in two steps. First, we cross-sectionally project observation vector y_t onto the latent factors by defining

$$\hat{f}_t = (Z'H^{-1}Z)^{-1}Z'H^{-1}y_t \quad (16)$$

Second, we use Kalman filtering methods in order to evaluate the low-dimensional model. After defining $C = (Z'H^{-1}Z)^{-1}$, this low-dimensional model can be presented by

$$\hat{f}_t = f_t + u_t \quad u_t \sim NID(0, C) \quad (17)$$

In this way, we are able to consider an updated log-likelihood function belonging to the collapsed filtering approach, given by

$$\ell(y|\psi) = c + \ell(\hat{f}|\psi) - \frac{T}{2} \log \frac{|H|}{|C|} - \frac{1}{2} \sum_{t=1}^T e_t' H^{-1} e_t \quad (18)$$

where c is a constant independent of both y and ϕ , $e_t = y_t - Z\hat{f}_t$, and $\ell(\hat{f}|\psi)$ is the log-likelihood of the low-dimensional model from equation 16. Here, the inversion of H seems the only time-consuming computation that is left. However, due to a diagonal structure inverting H is relatively straightforward. In consequence, by adopting the collapsed filtering approach of Jungbacker and Koopman (2014), we provide a convenient framework to efficiently estimate dynamic factor models for the IVS.

As stated in section 3.1.2, the three extended DFM-GARCH models can also be rewritten to general state space forms by using equations 8 and 9. Hence, for the greater part their estimation procedures are equivalent to the Kalman filtering approach of the general DFM model as discussed before. However, the addition of GARCH disturbances and extra parameters requires to apply certain adjustments to the estimation procedure. In particular, Harvey et al. (1992) argues that, although the models are not conditionally Gaussian because knowledge of past observations does not imply knowledge of past GARCH disturbances, we may treat the models as though they are conditionally Gaussian. For that matter, in presence of GARCH effects the Kalman filter can be regarded as quasi-optimal filter instead of optimal. Hence, the additional diagonal elements q_t and h_t in the covariance matrix of the state error term from equation 9 can be filled in by conditional variance terms. Again, Harvey et al. (1992) provides definitions of these conditional variance terms for both ϵ_t and η_t , given by

$$\text{var}_{t-1}(\epsilon_t) = \alpha_0 + \alpha_1(\hat{\epsilon}_{t-1}^2 + p_{t-1}^\epsilon) + \alpha_2 \text{var}_{t-2}(\epsilon_{t-1}) \quad (19)$$

$$\text{var}_{t-1}(\eta_t) = \gamma_0 + \gamma_1(\hat{\eta}_{t-1}^2 + p_{t-1}^\eta) + \gamma_2 \text{var}_{t-2}(\eta_{t-1}) \quad (20)$$

where p_{t-1}^ϵ and p_{t-1}^η represent the conditional variances of ϵ_{t-1} and η_{t-1} , respectively. Hence, by ap-

plying these calculations to a time-varying covariance matrix Q_t in equation 13, we are able to estimate our supposed conditionally Gaussian DFM-GARCH models appropriately. Besides, the collapsed approach is only effective for the GARCH extension where only the observation equation includes GARCH disturbances. In case of GARCH disturbances within the state equation, the loading matrices Z do not have full rank, resulting in a lack of impact from splitting the time series into a low-dimensional and a high-dimensional series. Hence, in these cases we consider a basic univariate Kalman filtering approach.

Furthermore, our dynamic factor models require identification restrictions on their representations matrices in order to enable proper estimation. Besides earlier restrictions on the factor loading matrix from equation 3, we also impose identification restrictions on the covariance matrices of the error terms. On the one hand, by forcing the covariance matrix of the observation residuals Σ_ϵ to be diagonal, we imply all cross-sectional co-movements to be attributed to the latent factors f_t . On the other hand, we follow Van der Wel et al. (2016) by estimating the parameters of the state residual covariance matrix Σ_η with LDL-decomposition. Here, Σ_η is defined by LDL' with L a (3×3) lower unit triangular and D a (3×3) diagonal matrix, both containing three parameters to be estimated. Moreover, for both covariance matrices we force variances on the diagonals to be positive by taking exponentials before and subsequently logarithms after running the BFGS optimizer. As discussed in section 3.1.2, we are dealing with six additional α and γ parameters and 27 additional Γ and Ψ parameters in case of extended GARCH disturbances. In order to enable their identification appropriately, we fix both α_0 and γ_0 equal to 0.0001.

In addition, we provide insights on how to initialize starting values for the unknown parameters. We start by setting up a three-step procedure for the general DFM parameters, involving Λ , Φ , μ , Σ_ϵ and Σ_η from equation 2. First, we again adopt the framework of Christoffersen et al. (2015) from equation 10 by regressing implied volatility on both moneyness and maturity dimensions. As a second step, by using principal component analysis we extract latent factors from this regression. Third, we use these principal components to perform OLS regression on the observation and state equations in order to obtain starting values for all unknown DFM parameters. Supplementary, we rotate the loading matrix Λ by means of corresponding restrictions given in equation 3. Next, we set up a comparable framework for the additional GARCH loading parameters Γ and Ψ . Knowing that the GARCH terms $\Gamma\epsilon_t + \epsilon_t^*$ and $\Psi\eta_t + \eta_t^*$ from equation 4 can be regarded as individual volatility models, we are able to determine their starting values in a same way. Specifically, we perform principal component analysis and OLS regression directly on the estimated residuals instead of onto the implied volatilities. Hence, we manage to obtain starting values for the additional loading matrices Γ and Ψ . For the remaining α and γ parameters we adopt the constraints of Harvey et al. (1992), given by $\alpha_1 + \alpha_2 < 1$ and $\gamma_1 + \gamma_2 < 1$. As displayed in equations 6 and 7, α_1 and γ_1 represent the conditional variances of ϵ_{t-1} and η_{t-1} , whereas α_2 and γ_2 represent the lagged variances h_{t-1} and q_{t-1} . Hence, given that these latter terms are more persistent, we set the starting values of α_1 and γ_1 equal to 0.09 and the starting values of α_2 and γ_2 equal to 0.9. In consequence, we conduct a detailed framework on how to find starting values for the unknown parameters in order to successfully initiate model estimations.

4 Statistical Evaluation

In our attempt to find strong performing models to forecast the IVS, we introduced various dynamic factor models with corresponding estimation procedures and benchmarks in section 3. In this section, we evaluate these models statistically by investigating both their in-sample fit and out-of-sample forecasting performances. To begin with, we introduce two methods to test the significance of GARCH effects in dynamic factor models by measuring autocorrelation and heteroskedasticity in the error terms in section 4.1. Next, we provide definitions of several statistical measures for testing estimation and forecasting performances in section 4.2. Subsequently, we evaluate both in-sample and out-of-sample results extensively in section 4.3.

4.1 Measuring Significance of GARCH Effects

In this thesis, we attempt to find an improved specification of dynamic factor models on the IVS in order to mitigate the problem of heteroskedasticity and autocorrelation in the error terms. In subsequent sections, we introduce statistical methods to compare and benchmark the in-sample fit and out-of-sample performances of those extended dynamic factor models. However, merely using these evaluation methods will not provide us clear conclusions on whether potential improvements of our models are significantly due to the presence of GARCH effects. Therefore, we introduce additional testing methods to measure the significance of GARCH effects directly. Hence, although a GARCH model is originally designed to correct for heteroskedasticity rather than for autocorrelation, we evaluate the impact of GARCH disturbances on both aspects in order to fully explore potential of additional volatility disturbances to dynamic factor models for the IVS. More specifically, we compare our dynamic factor models on the levels of both autocorrelation and heteroskedasticity in the (squared) observation error terms.⁹

First, we test the significance of autocorrelations in the error terms of our dynamic factor models by evaluating corresponding Ljung-Box statistics introduced by Ljung and Box (1978). These statistics, closely connected to the Box-Pierce test from Box and Pierce (1970), assess the existence of GARCH effects in our fitted models by examining whether the autocorrelations of the (squared) residuals are different from zero. In case of testing for GARCH effects with using P autocorrelations, McLeod and Li (1983) report that the Ljung-Box statistic has a chi-square distribution with P degrees of freedom $\chi^2_{(P)}$ under the null hypothesis of no GARCH effects. In our specific case with GARCH(1,1) disturbances, the null hypothesis normally imposes $H_o : \alpha_1 = \alpha_2 = \gamma_1 = \gamma_2 = 0$. Inspired by the work of Harvey et al. (1992), we select six autocorrelations ($P = 6$). Hence, the Ljung-Box statistic for the residuals $Q(6)$ and squared residuals $Q^2(6)$ are defined by

$$Q(6) = n(n+2) \sum_{k=1}^6 \frac{(\hat{\rho}_6)^2}{n-k} \quad Q^2(6) = n(n+2) \sum_{k=1}^6 \frac{(\hat{\rho}_6^*)^2}{n-k} \quad (21)$$

where n is the sample size and $\hat{\rho}_6$ ($\hat{\rho}_6^*$) is the sample autocorrelation of the (squared) residuals at lag six.

Likewise, we use the Breusch-Pagan Lagrange multiplier test introduced by Breusch and Pagan (1980) to measure the significance of heteroskedasticity in the observation error terms of our dynamic factor models. In general this method tests whether the variances of the (squared) error terms are dependent of

⁹In addition to our reported results involving GARCH effects in the observation equation, we also statistically test the presence of GARCH effects in the state equation by means of Ljung-Box and Breusch-Pagan Lagrange multiplier statistics. After performing these tests, we find similar results compared to the tests for GARCH effects in the observation equation. For clarity purposes, we therefore leave the evaluation of GARCH effects in the state equation out of consideration.

the latent factors. In case of dependency, heteroskedasticity is significantly present. Hence, we attempt to find insights on whether the inclusion of GARCH disturbances in dynamic factor models at least reduces this dependency. In detail, we start by regressing the residuals ϵ on the latent factors, given by

$$\epsilon = \beta_0 + \beta_1 f + v \quad (22)$$

where v is the regression error term. Specifically, we test for homoscedasticity in ϵ which implies that coefficient β_1 is equal to zero. Following standard practice, the Lagrange multiplier statistics are therefore calculated by multiplying the resulting coefficient of determination with sample size n , given by

$$LM(\epsilon) = nR^2 \quad (23)$$

In line with the Ljung-Box statistics, this test statistic again has a chi-square distribution with P degrees of freedom $\chi^2_{(P)}$. In this case however, it holds that $P = 3$ due to the amount of three latent factors within f . Similar to these definitions for residuals, we also perform the Breusch-Pagan Lagrange multiplier test for squared residuals ϵ^2 based on equations 22 and 23. By testing the (squared) residuals of our main dynamic factor models, we attempt to find statistical insights on whether including GARCH disturbances mitigates the problem of heteroskedasticity and autocorrelation in the residuals of dynamic factor models.

4.1.1 Evaluation GARCH Effects Statistics

In this section, we evaluate whether the inclusion of GARCH disturbances can mitigate the problem of autocorrelation and heteroskedasticity in the residuals of dynamic factor models for the IVS. In particular, we analyze Ljung-Box statistics in order to examine autocorrelation in the (squared) residuals of the observation equation, whereas heteroskedasticity in these (squared) residuals is tested by analyzing Breusch-Pagan Lagrange multiplier statistics as introduced in previous section. Table 2 reports corresponding summary statistics of the average test statistics over all IVS groups.¹⁰ A full overview of the statistics per IVS group is provided in the appendix in tables 14 and 15. In general, we find extremely high statistics for all models and tests. In addition, nearly all individual IVS groups show significant statistics for both autocorrelation and heteroskedasticity in the (squared) residuals. On average, the lowest Ljung-Box and Breusch-Pagan statistics are even above 2,527 and 174, respectively. Therefore, all average test statistics for autocorrelation and heteroskedasticity in the (squared) residuals are significant at a 1% significance level. Nonetheless, we observe several differences between our dynamic factor models. Although we observe similar results for the DFM-GARCH setup involving GARCH disturbances in the state equation, the table also shows slight improvements in both setups involving GARCH disturbances in the observation equation. For example, the average Ljung-Box statistic for squared residuals for the general DFM model is 3,276, whereas the two DFM-GARCH setups show sufficiently lower statistics below 2,553. Likewise, corresponding Breusch-Pagan Lagrange multiplier statistics for the DFM-GARCH setups are below 186, in comparison to 204 for the general DFM model. Hence, our DFM-GARCH extensions seems to be more or less effective to correct for both autocorrelation and heteroskedasticity in the error terms, even though GARCH disturbances are naturally more appropriate to correct for heteroskedasticity rather than for autocorrelation. In conclusion, although we are not able to solve the problem of a poor fit in the corner groups of the IVS entirely, we do find that including GARCH disturbances in the observation equation of dynamic factor models can mitigate this problem. More specifically, including GARCH disturbances proves to correct for heteroskedasticity and autocorrelation

¹⁰This table reports statistics for testing GARCH effects in the observation equation. Testing for GARCH effects in the state equation results in similar findings.

in the error terms of dynamic factor models to some extent. Hence, these results might indicate that our GARCH extensions improve the fit of dynamic factor models, potentially resulting in improvements for practitioners when forecasting the IVS based on dynamic factor models in option pricing or risk exposure applications. In order to explore potential improvements of the inclusion of GARCH disturbances more extensively, we evaluate the in-sample fit of our extended dynamic factor models in section 4.3.2.

Table 2: **Significance of GARCH Effects**

		Autocorrelation		Heteroskedasticity	
		Ljung-Box		Breusch-Pagan	
		Q(6)	Q ² (6)	LM(ϵ)	LM(ϵ^2)
DFM-GARCH	Observation	4,437 (24/24)	2,553 (23/24)	743 (24/24)	186 (23/24)
	State	5,241 (24/24)	3,295 (24/24)	753 (24/24)	208 (23/24)
	Obs+State	4,407 (24/24)	2,527 (24/24)	740 (24/24)	174 (23/24)
Benchmarks	DFM	5,279 (24/24)	3,276 (24/24)	746 (24/24)	204 (23/24)
	RFM	5,813 (24/24)	3,963 (24/24)	1,117 (24/24)	331 (23/24)

Notes: This table provides summary statistics regarding the significance of additional GARCH disturbances in the observation equation of the general dynamic factor model. For this purpose, we compare presence of autocorrelation and heteroskedasticity in the observation error terms of our main dynamic factor models. In particular, we provide significance tests of autocorrelation in the residuals (Q(6)) and squared residuals (Q²(6)) using Ljung-Box statistics. Likewise, we provide significance tests of heteroskedasticity in the residuals (LM(ϵ)) and squared residuals (LM(ϵ^2)) using Breusch-Pagan Lagrange multiplier statistics. For clarity purposes, we combine the total set of IVS groups by taking average statistics. The number of individual IVS groups that show significant test statistics using an $\alpha = 1\%$ level are reported in parentheses. Full documentation of autocorrelation and heteroskedasticity statistics can be found in the appendix in tables 14 and 15.

4.2 Statistical Measures of Predictability

Following related literature like among others Chalamandaris and Tsekrekos (2010) and Bernales and Guidolin (2014), we provide several methods to evaluate statistical performances of our dynamic factor models. For this purpose, we consider the full sample period running from January 2002 until August 2015. First, we evaluate in-sample fit by analyzing the factor dynamics and whether the GARCH coefficients show persistence. Next, we compare the performances of our dynamic factor models by means of their maximized log-likelihood values and correlated likelihood-ratio (LR) tests. Further, in order to verify our findings, we also report Akaike (AIC) and Bayesian (BIC) information criteria.

Furthermore, we perform a recursive back-testing exercise in order to evaluate the out-of-sample forecasting performances systematically. If these forecasts turn out to be accurate, our dynamic factor models can have a positive impact on risk and portfolio management decisions. We set up our forecasting evaluation by initializing a rolling window of 1,000 days, resulting in an estimation time span of almost four years. Then, we recursively estimate our models on a daily basis by imposing forecast horizon $h = 1$. This choice is motivated by the findings of among others Goncalves and Guidolin (2006), who conclude that larger forecast horizons result in similar IVS outcomes compared to one-day-ahead predictions. Hence, for all our models we recursively compute one-day-ahead predictions of implied volatility from day 1,001 (December 2005) until the end of the sample period (August 2015).

In addition, we set up three measures to evaluate both in-sample and out-of-sample fit of the models of interest. Following standard practice, we define RMSE, adjusted R^2 and MCP as follows.

RMSE The *root mean squared error* is defined as the square root of the mean squared deviations from the model’s predicted implied volatilities \hat{y}_t compared to the actual observed implied volatilities y_t . This measure can be used both in-sample and out-of-sample and is given by

$$RMSE_{\tau_i, m_j} = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_{\tau_i, m_j, t} - y_{\tau_i, m_j, t})^2} \quad (24)$$

Adjusted R^2 The *adjusted R^2* , also referred to as *adjusted coefficient of determination*, is defined as a measure for the fit of a model adjusted for its number of variables. This measure can be obtained by performing a univariate regression of the actual implied volatilities y_t on the model’s predicted implied volatilities \hat{y}_t . In detail, the adjusted R^2 can be calculated by

$$\bar{R}^2 = 1 - \frac{SS_{res}/df_e}{SS_{tot}/df_y} \quad (25)$$

where SS_{res} and SS_{tot} represent the residual and total sum of squares, respectively. The degrees of freedom of the residuals variances and dependent variable variances are referred to as df_e and df_y .

MCP The *mean correct prediction of direction of change in implied volatilities* is defined as the percentage of the IVS for which the model correctly predicts the sign of change in implied volatility h days ahead. An analytical formulation is given by

$$MCP_{\tau_i, m_j} = \frac{1}{T} \sum_{t=1}^T \mathbf{1}_{sign[\hat{y}_{\tau_i, m_j, t+h} - y_{\tau_i, m_j, t}] = sign[y_{\tau_i, m_j, t+h} - y_{\tau_i, m_j, t}]} \quad (26)$$

For our in-sample evaluation purposes we consider RMSE and adjusted R^2 measures, whereas for out-of-sample forecasting evaluation purposes we consider the RMSE and MCP measures. Although we base our in-sample evaluation on the virtually constructed cross-sectional IVS, we generate out-of-sample forecasts for all available individual options. These forecasts of individual option contracts are most relevant for practitioners, as they implement implied volatility forecasts into their decision making by focusing on individual options rather than the entire IVS. The one-day-ahead forecasts of individual option contracts are obtained by interpolating the implied volatility forecasts of the entire surface on a daily basis. More specifically, we start our interpolation procedure by considering the implied volatility forecasts of the 24 selected option contracts. On each day, we then construct a two-dimensional grid of IVS forecasts based on the moneyness and maturity dimensions of these selected contracts, following a similar (6×4) framework as the construction of the IVS from section 2.2. Thereafter, by using actual Δ and time-to-maturity values of the individual option contracts, we interpolate all option contracts over these grids in order to determine their one-day-ahead implied volatility forecasts. Moreover, we further utilize MCP statistics in order to economically evaluate the IVS models in section 5.

4.3 Statistical Evaluation Results

In section 4.2 we discussed several ways to evaluate the in-sample fit and out-of-sample forecasting performances of the dynamic factor models and their benchmarks. Hence, we report the results of these methods extensively for both in-sample estimates and out-of-sample forecasts in the following sections.

4.3.1 Estimation Results

First of all, we analyze the in-sample estimation process by reporting the (first) three estimated dynamic factors of our five main likelihood-based dynamic factor models in figure 5. As expected, these models clearly follow a similar structure for their latent factors due to a consistent basic setup. More specifically, our inclusion of GARCH disturbances to the general DFM model only affects its ultimate results while leaving the conditional mean unchanged. Hence, our DFM-GARCH setups can be simply considered as extended versions of the general DFM model. In contrast, the RFM model already captures the level, term structure and smile dynamics to its factors. Similarities between all first factors suggest that the level of the IVS indeed explains the major part of the co-movements in the IVS. However, the figure shows different scales for particularly the second and third factor of the RFM model compared to the unrestricted models. Apparently, the remaining two latent factors of the unrestricted models do not fully correspond with the economically plausible smile and term structure effects of the RFM model. To examine this more extensively, we provide the estimated loading matrix of the general DFM model in table 13 in the appendix. In case of factors that fully resemble the volatility smile and term structure effects, we would expect the loadings to display corresponding patterns. However, the estimated loadings for the second and third factor in table 13 show inconsistent patterns across moneyness and maturity dimensions. Therefore, this table confirms that the volatility smile and term structure effects are not fully explained by the two latent factors of our DFM(-GARCH) setups. Moreover, the structures of the estimated dynamic factors of our unrestricted dynamic factor models show similar patterns, whereas the economically plausible RFM model shows reasonable differences for particularly second and third factors.

These relationships between the models are confirmed by estimates of the factor dynamics, provided in table 3. The diagonal elements of the factor transition matrix Φ are close to one, signifying that the (first) three dynamic factors are highly persistent. This persistence indicates potential danger of non-stationary latent factors. Dealing with similar signs of persistent factors, Van der Wel et al. (2016) already check robustness of their dynamic factor models by imposing a random walk for (one of) the three latent factors. On the one hand, they report evidence of rejected random walk restrictions on the factor dynamics by means of LR-tests, indicating the value of modeling the factors following an unrestricted first-order autoregressive specification. But on the other hand, they show strong similarities in VAR-coefficients and covariance matrices across all cases, signaling that the latent factors are at least close to being non-stationary. Hence, studying non-stationary dynamic factor models for the IVS would be an interesting alternative practice for further research. Meanwhile, we observe off-diagonal elements closer to zero than the diagonal elements. In addition, the table shows remarkable differences between the basic DFM/RFM setups and our extended DFM-GARCH setups. In particular, some of the off-diagonal transition elements within the DFM-GARCH setups are relatively high in absolute terms. This can be explained by the adjusted model setup in equation 9, where in state space form the state equation exhibits additional dynamics from the extra GARCH factors. Considering the estimated covariance matrices of the state error terms η , we also observe notable differences. As comparable basic dynamic factor model setups, the DFM and RFM models show similar scales and patterns in their covariance matrices. In contrast, all DFM-GARCH setups show some deviations due to the inclusion of additional GARCH disturbances. Particularly, we find a strong relation between the covariance matrices of both setups in which we added GARCH disturbances in the state equation. Including GARCH disturbances in only the observation equation results in different deviations compared to the benchmark models. These differences between our basic and DFM-GARCH setups are also visible in the state equation intercepts μ , where we observe minor differences in magnitude and proportions.

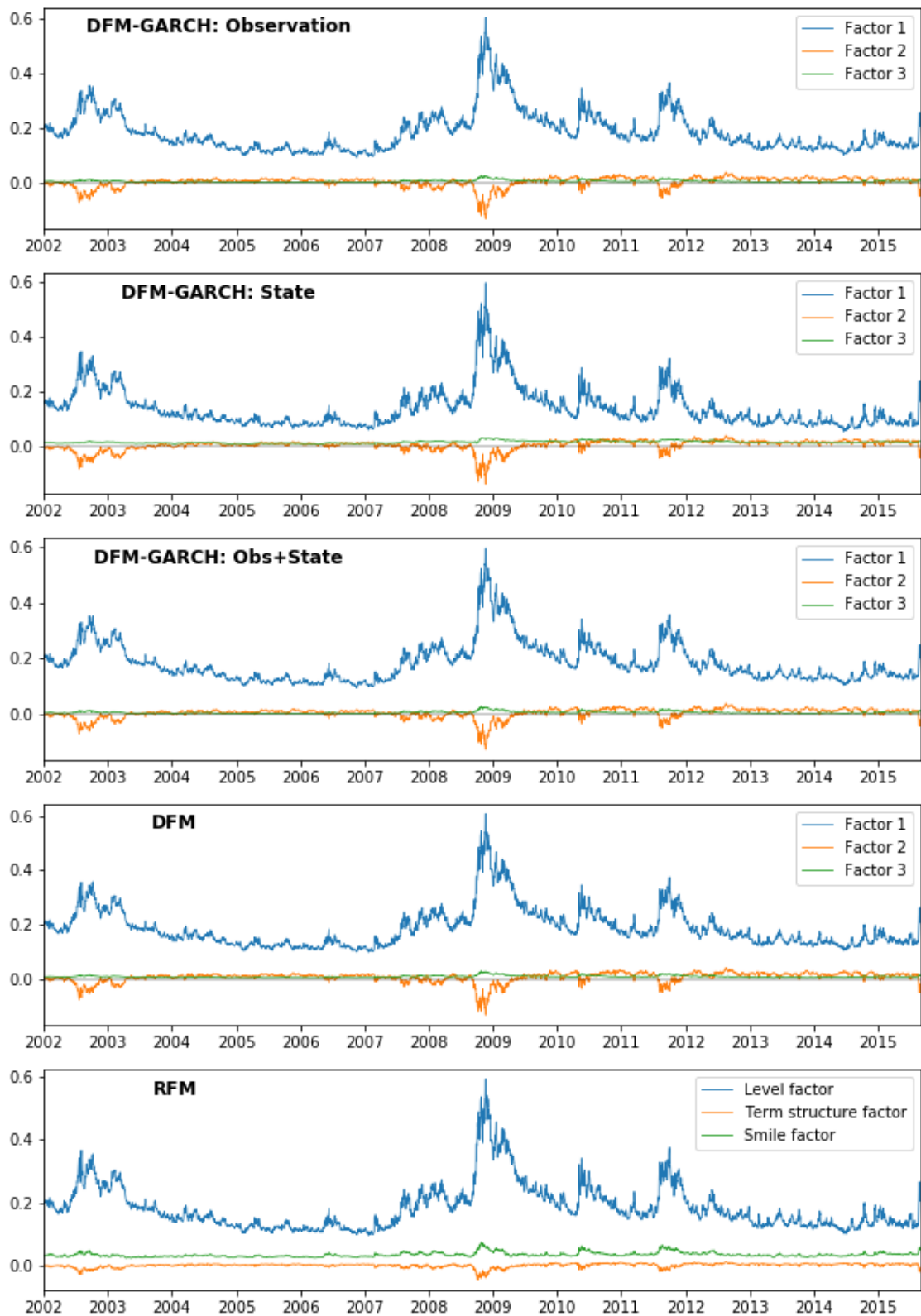


Figure 5: Estimated Dynamic Factors

Notes: These figures show the (first) three dynamic factors of our main dynamic factor models, estimated using cross-sectional implied volatility data of S&P 500 index options over the full sample period from January 2002 until August 2015. Specifically, we report the first three latent factors of the extended DFM-GARCH models and the general DFM model. Besides, we provide the specified level, term structure and smile factors of the RFM model.

Table 3: Factor Dynamics

Panel A: DFM-GARCH - Observation

		Φ			$\Sigma_{\eta}^* (\times 10^{-4})$		
	μ	$f_{1,t-1}$	$f_{2,t-1}$	$f_{3,t-1}$	$f_{1,t}$	$f_{2,t}$	$f_{3,t}$
$f_{1,t}$	0.109	1.041	0.180	-0.351	1.602	-0.545	0.006
$f_{2,t}$	0.012	-0.022	0.909	0.168	-0.545	0.192	0.000
$f_{3,t}$	0.018	-0.001	-0.005	1.003	0.006	0.000	0.002

Panel B: DFM-GARCH - State

		Φ			$\Sigma_{\eta}^* (\times 10^{-4})$		
	μ	$f_{1,t-1}$	$f_{2,t-1}$	$f_{3,t-1}$	$f_{1,t}$	$f_{2,t}$	$f_{3,t}$
$f_{1,t}$	0.098	1.032	0.138	-0.408	0.298	-0.087	0.022
$f_{2,t}$	0.012	-0.020	0.921	0.202	-0.087	0.032	-0.005
$f_{3,t}$	0.013	0.000	-0.003	0.996	0.022	-0.005	0.002

Panel C: DFM-GARCH - Observation+State

		Φ			$\Sigma_{\eta}^* (\times 10^{-4})$		
	μ	$f_{1,t-1}$	$f_{2,t-1}$	$f_{3,t-1}$	$f_{1,t}$	$f_{2,t}$	$f_{3,t}$
$f_{1,t}$	0.140	1.003	0.019	-0.242	0.254	-0.106	0.026
$f_{2,t}$	0.013	-0.007	0.977	0.168	-0.106	0.060	-0.012
$f_{3,t}$	0.006	0.001	0.002	0.972	0.026	-0.012	0.003

Panel D: General DFM

		Φ			$\Sigma_{\eta} (\times 10^{-4})$		
	μ	$f_{1,t-1}$	$f_{2,t-1}$	$f_{3,t-1}$	$f_{1,t}$	$f_{2,t}$	$f_{3,t}$
$f_{1,t}$	0.187	0.998	0.033	-0.042	1.087	-0.469	0.051
$f_{2,t}$	0.005	-0.004	0.967	0.075	-0.469	0.219	-0.022
$f_{3,t}$	0.008	0.000	0.001	0.983	0.051	-0.022	0.003

Panel E: RFM

		Φ			$\Sigma_{\eta} (\times 10^{-4})$		
	μ	$f_{1,t-1}$	$f_{2,t-1}$	$f_{3,t-1}$	$f_{1,t}$	$f_{2,t}$	$f_{3,t}$
$f_{1,t}$	0.184	0.997	0.031	-0.035	1.140	-0.513	0.070
$f_{2,t}$	0.004	-0.003	0.965	0.065	-0.513	0.254	-0.034
$f_{3,t}$	0.006	0.000	0.002	0.981	0.070	-0.034	0.007

Notes: This table provides estimated values of the factor dynamics of all our main dynamic factor models. Panels A-C provide factor dynamics of the extended DFM-GARCH models, whereas Panels D and E show estimated values of the factor dynamics for the general DFM and RFM models. For each model, we report the intercept of the state equation μ , the VAR coefficient matrix Φ , and the covariance matrix of the state residual Σ_{η} . For clarity purposes, we only report the upper left (3×3) parts of Φ and Σ_{η} in case of extended DFM-GARCH models.

So far, we considered the (first) three estimated dynamic factors. By zooming in on our extended DFM-GARCH setups, we provide additional time series of the GARCH factors in figure 6. For all three setups, we plot the observation GARCH factor ϵ and/or the state GARCH factor η . As expected, the extra factors from models in which GARCH disturbances are imposed in only one of the two state space equations follow similar patterns as the factors in the combined setup. Moreover, we observe highest deviations in the time series of η , with estimates varying between -0.5 and 2. Especially around the financial crisis in 2008 the observation GARCH factor reacts strongly. Meanwhile, ϵ shows relatively less varying estimates between -0.2 and 0.8 over the full sample period. In addition, we provide further insights on the estimation process of our DFM-GARCH models by showing time series of the additional GARCH variances h_t and q_t in figure 7. In general, we observe relatively fluctuating patterns in q_t , which is the additional variance term in case of including GARCH disturbances in the state equation. This time series consists of several peaks, with a maximum value of around 0.8. In contrast, the variance involving GARCH disturbances in the observation equation shows less fluctuation, with a maximum peak of below 0.2. Again, both variables show comparable patterns in case of including GARCH disturbances in only equation compared to the combined setup.

In addition, table 4 provides the dynamics of these additional GARCH factors by reporting parameter coefficients of associated α 's and γ 's. As discussed in section 3.3, the intercepts α_0 and γ_0 are fixed at 0.0001 in order to enable appropriate identification of the other parameters. These remaining parameters are restricted based on the work of Harvey et al. (1992), implying $\alpha_1 + \alpha_2 < 1$ and $\gamma_1 + \gamma_2 < 1$. Based on the table, we find optimally estimated GARCH factors with $\alpha_1 + \alpha_2$ and $\gamma_1 + \gamma_2$ very close to one in all cases. In general, the coefficients of the latter parameters for the lagged GARCH terms are greater and more significant than the coefficients of the squared scalar disturbance terms α_1 and γ_1 . Particularly the observation disturbances exhibit this pattern. In the meantime, the effects of the state disturbances are closer related, with an even higher and significant coefficient for γ_1 compared to γ_2 in case of GARCH disturbances in both equations.

Table 4: **Dynamics of GARCH Disturbances**

		Observation Disturbances			State Disturbances		
		α_0	α_1	α_2	γ_0	γ_1	γ_2
DFM-GARCH	Observation	0.0001	0.1434*	0.8399***			
	State				0.0001	0.2245*	0.7717*
	Obs+State	0.0001	0.2262*	0.7718**	0.0001	0.5117*	0.4713

Notes: This table provides estimated values of the dynamics of the additional GARCH disturbance terms in case of our extended DFM-GARCH models. Specifically, we report estimated parameter coefficients of additional GARCH disturbances in the observation equation, referred to as α_0 , α_1 and α_2 . Likewise, dynamics of GARCH disturbances in the state equation are given by reporting estimates for γ_0 , γ_1 and γ_2 . For identification purposes, α_0 and γ_0 are fixed at 0.0001. Detailed definitions of all GARCH parameters are provided in section 3.1.2 in equations 6 and 7. */**/** denotes statistically significant results at an $\alpha = 10\%/5\%/1\%$ significance level.

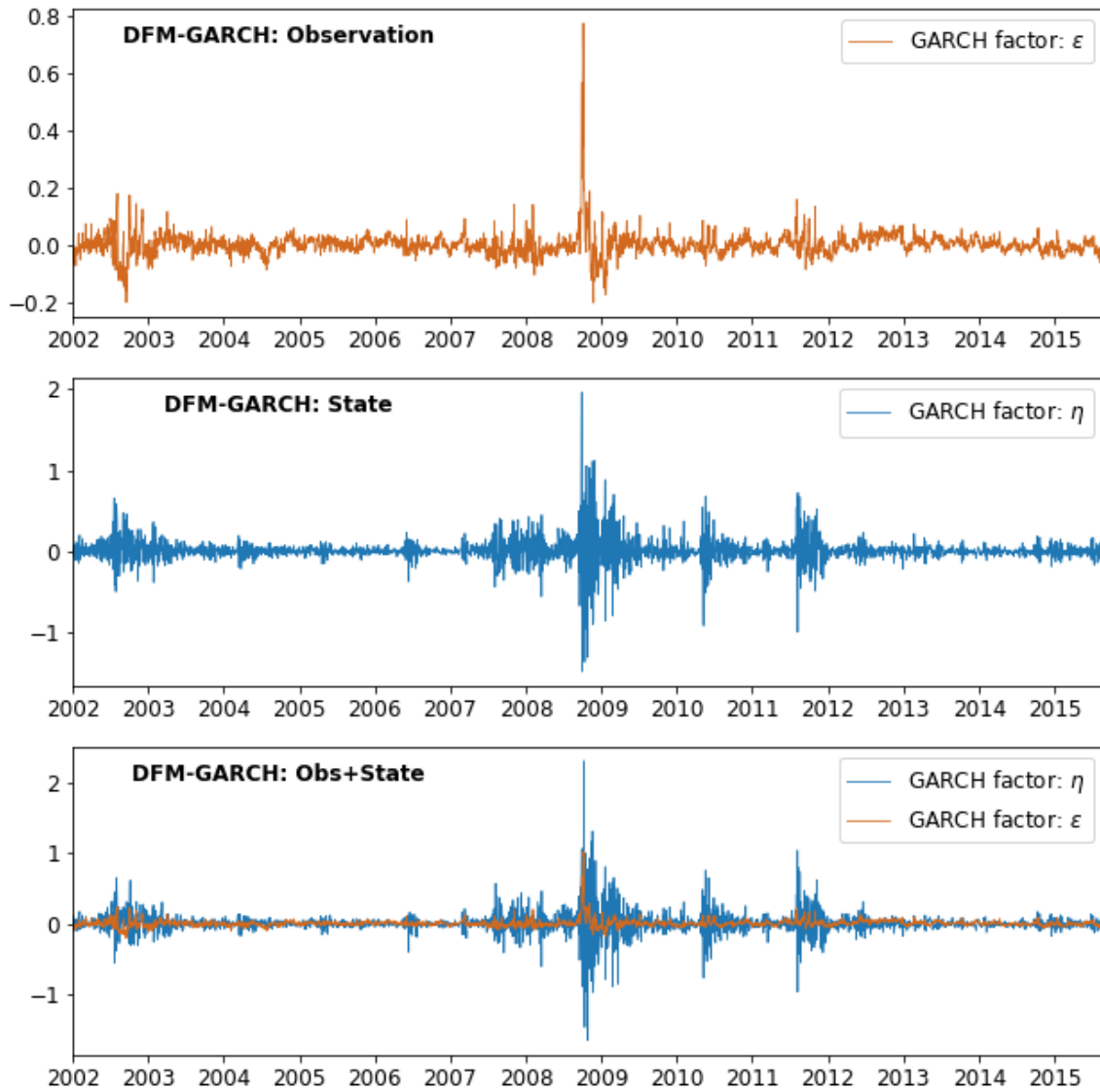


Figure 6: **Estimated Latent Factors of GARCH Disturbances**

Notes: These figures show the additional GARCH factors of our extended DFM-GARCH models, estimated using cross-sectional implied volatility data of S&P 500 index options over the full sample period from January 2002 until August 2015. Specifically, we report the latent factors of the GARCH disturbances within the observation equation (ϵ) and/or state equation (η). The upper two figures display the DFM-GARCH models that only includes GARCH disturbances in the observation or state equation, whereas the bottom figure provides the DFM-GARCH variant in which GARCH disturbances are present in both state space equations.

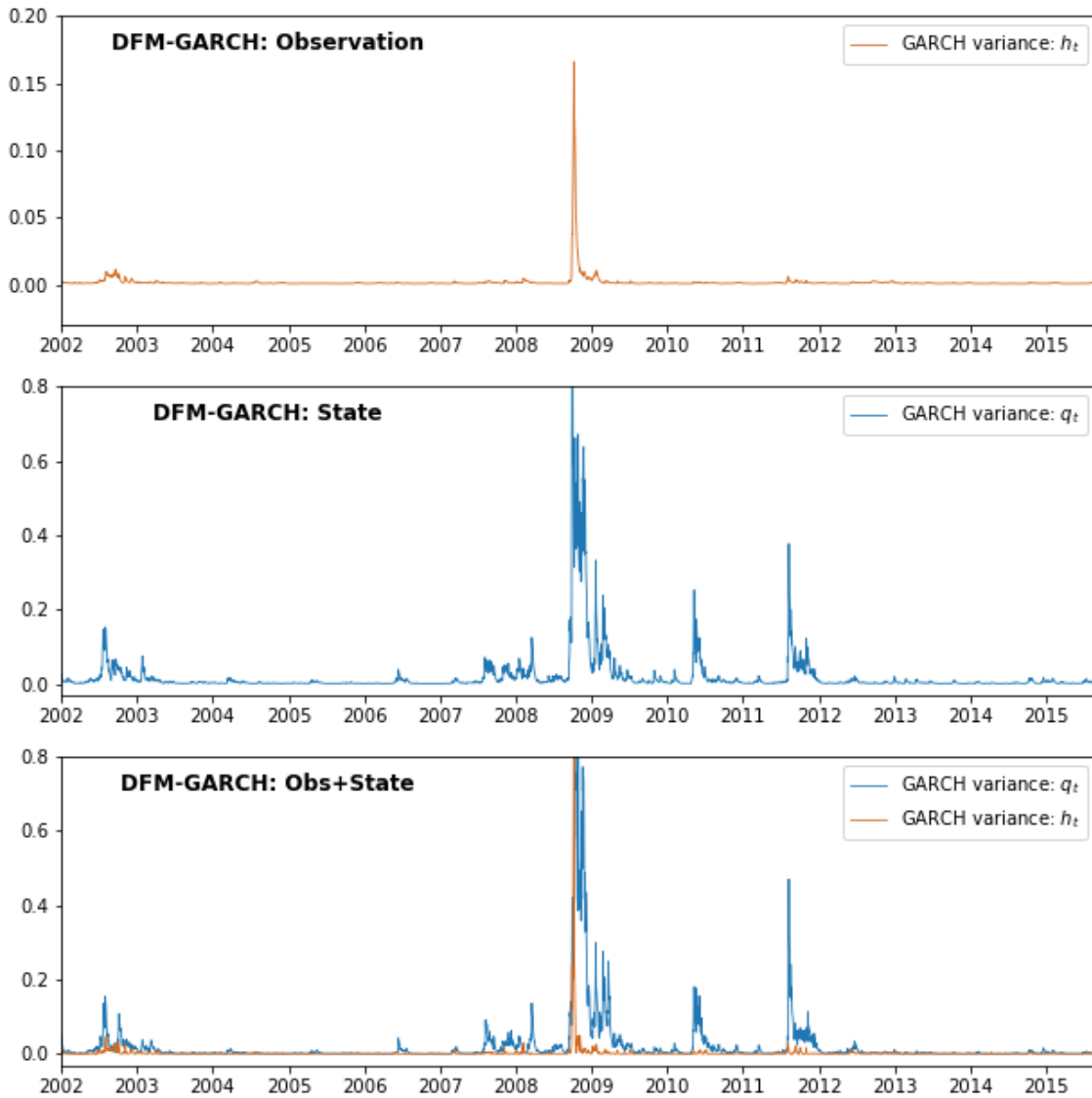


Figure 7: **Estimated Variances of GARCH Components**

Notes: These figures show the GARCH variances of our extended DFM-GARCH models, estimated using cross-sectional implied volatility data of S&P 500 index options over the full sample period from January 2002 until August 2015. Specifically, we report the variances of the GARCH components within the observation equation (h_t) and/or state equation (q_t). The upper two figures display the DFM-GARCH models that only includes GARCH disturbances in the observation or state equation, whereas the bottom figure provides the DFM-GARCH variant in which GARCH disturbances are present in both state space equations.

4.3.2 In-Sample Fit

After having analyzed the estimated dynamic factors extensively, we provide general results of the in-sample estimation performances in this section. First, we document key statistics of the in-sample fit of our main dynamic factor models in table 5. Here, the general DFM model is considered as baseline model for the likelihood-ratio tests (LR-tests). This benchmark contains 102 parameters and reports a log-likelihood of 288,385.¹¹ The restricted RFM model reports a sufficiently lower log-likelihood of 257,416, even though this already concerns the best performing RFM variant from Van der Wel et al. (2016) with time-varying loading matrix Λ_t . Although the RFM model contains only 42 parameters and suffers less risk of overfitting, this model is even outperformed considerably on statistical measures like AIC and BIC, which take the number of parameters into account. Evidently, the remaining co-movements in the IVS that are not explained by the implied volatility level, do not fully match with the economically plausible volatility smile and term structure effects as imposed by the RFM model. Hence, the DFM model significantly outperforming the RFM model motivates our choice of adding GARCH disturbances to the better performing general DFM model. In addition, these extended DFM-GARCH models even show improvements compared to the general DFM model, with log-likelihoods varying from 289,586 to 298,894. On the one hand, the setup with GARCH disturbances included in the state equation only shows a minor improvement in both log-likelihood and AIC/BIC. This model extension therefore suggest little impact from its five extra parameters. On the other hand, the setups with GARCH disturbances in either the observation equation or both equations show significantly improved log-likelihoods. Even with taking all their extra parameters (128 and 133) into account, these models deliver best performances on AIC and BIC criteria. In consequence, we can conclude that these summary statistics confirm the potential of our GARCH extensions to dynamic factor models. Moreover, including GARCH disturbances in both state space equations appears to be best practice, but differences with the setup including GARCH disturbances in the observation equation only are minimal. Hence, it seems useful to explore these relationships more extensively by evaluating other statistical measures.

Table 5: Comparing Dynamic Factor Models

		Loglike	LR-test	AIC	BIC	# Params
DFM-GARCH	Observation dist.	296,261	-15,752	-592,241	-591,449	128
	State dist.	289,586	-2,401	-578,978	-578,314	107
	Obs+State dist.	298,894	-21,017	-597,571	-596,748	133
DFM	General model	288,385	NA	-576,542	-575,909	102
RFM	Time-varying Λ_t	257,416	61,938	-514,724	-514,459	42

Notes: This table provides fundamental statistics regarding the fit of all key dynamic factor models, estimated using cross-sectional S&P 500 implied volatility data over the full sample period January 2002 until August 2015. First, statistics are provided for our extended DFM-GARCH models with additional GARCH disturbances incorporated in either the observation or state equations, and with GARCH disturbances included in both equations. Next, statistics regarding the fit of the general dynamic factor model (DFM) and the restricted economic dynamic factor model (RFM) with time-varying loading matrix Λ_t are given. For each model, we report log-likelihood values (Loglike), statistics of a likelihood-ratio test relative to the general DFM (LR-test), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and the number of parameters (# Params).

In particular, table 6 reports the in-sample fit of our main dynamic factor models and their benchmarks in terms of RMSE and adjusted- R^2 . In comparison to the RFM model and additional random

¹¹In order to guarantee solid comparisons to the research of Van der Wel et al. (2016), we test our general DFM model on the distinctive data set they consider. Overall, we find similar results with a log-likelihood of 302,132, confirming an accurate adoption of their estimation framework.

walk and two-step VAR benchmarks, the DFM(-GARCH) models show a considerably better fit on both measures. Evidently, the likelihood-based dynamic factor models prove to be more efficient and valuable in estimating the IVS accurately than the dynamic factor model benchmark estimated with a two-step OLS approach. Besides, by comparing our extended setups with the baseline DFM model, we observe similar trends as in table 5. Again, the variant in which only the state equation is adjusted shows minimal differences with the general DFM. In addition, the setups in which GARCH disturbances are included in the observation equation show substantial improvements once more. For example, RMSE has been reduced from 0.92 to 0.83 or 0.84, whereas the coefficient of determination is 0.003 higher in both DFM-GARCH cases involving GARCH disturbances in the observation equation. Hence, we can conclude that in general our DFM-GARCH setups again show accurate in-sample fits compared to their benchmarks.

Table 6: **In-Sample Fit**

		RMSE ($\times 10^{-2}$)	Adjusted- R^2
DFM-GARCH	Observation	0.83	0.987
	State	0.92	0.985
	Obs+State	0.84	0.987
Benchmarks	DFM	0.92	0.984
	RFM	1.35	0.967
	2-step VAR	1.48	0.959
	RW	1.41	0.964

Notes: This table provides statistics regarding the in-sample fit of our DFM-GARCH models relative to their benchmark models. Specifically, we report average root mean squared errors (RMSE) and adjusted- R^2 . In addition to our main dynamic factor models, we provide in-sample fit statistics for the random walk (RW) and the two-step vector-autoregressive (2-step VAR) benchmark models.

Previously analyzed tables provide general information on the average in-sample fit. However, in order to gain further insights on the dynamics of the IVS, we provide additional analyses of statistical measures applied to individual IVS groups. For example, figure 8 displays the in-sample RMSEs per IVS group in both absolute (left figures) and percentage (right figures) terms. In general, we observe minor differences between the DFM-GARCH setups and their general DFM benchmark. On the contrary, the restricted RFM model generally performs worse than the unrestricted setups, in particular within the two corner groups DOTM-Put with 10-45 days and DOTM-Call with 180-360 days. Hence, this confirms that the outperformance of the restricted model by the DFM setups is based on the entire surface. Overall, both the general DFM model and extended DFM-GARCH models show slightly convex curved RMSEs, whereas likewise the levels of the curves for shortest and longest maturities are slightly higher than the curves of centered maturity groups. Based on these results, one could argue that in general all dynamic factor models have most troubles with estimating corner groups of the IVS. This can be explained by the fact that contracts in corner groups of the IVS are less liquid and exhibit more erratic trading patterns than option contracts in center groups. Furthermore, we spot an additional remarkable pattern in the call option groups with shortest maturities. In particular, both DFM-GARCH setups involving GARCH disturbances in the observation equation score extremely low RMSEs within these IVS groups in comparison to the other DFM(-GARCH) setups. Hence, in specific groups on the edge of the surface, the DFM-GARCH setups with GARCH disturbances in the observation equation and in both equations show a better fit. In combination with similar RMSEs in nearly all other IVS groups, we

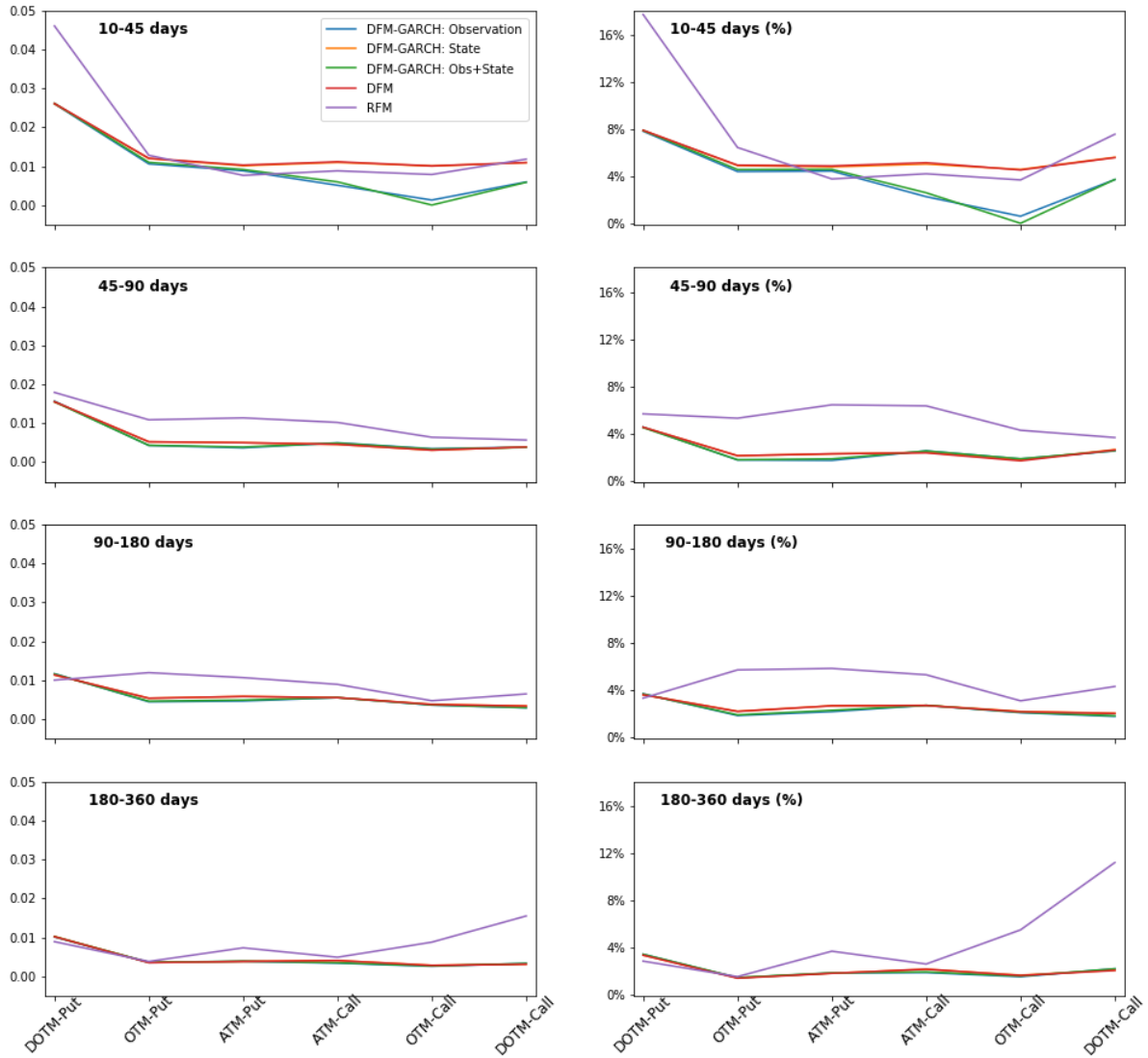


Figure 8: **In-Sample Root Mean Squared (Percentage) Errors**

Notes: These figures show the absolute and proportional in-sample fitting errors per IVS group of our main dynamic factor models. Specifically, the left figures display the absolute in-sample root mean squared errors (RMSE), whereas the right figures display corresponding fitting errors in terms of percentage. For all four maturity categories (from top to bottom), the fitting errors are plotted along the six moneyness categories.

therefore can conclude that the better average in-sample fit of these two models in tables 5 and 6 are mainly caused by a better fit in the call option IVS groups with shortest maturities. In consequence, this might indicate the first signals of improvements in the fit of the corner groups of the IVS after including GARCH disturbances in dynamic factor models.

In an attempt to find further confirmation of the relevance of our GARCH extensions to dynamic factor models, we compare time series plots of the fit of our main dynamic factor models. First, in line with Van der Wel et al. (2016) we plot general DFM fits from six of the 24 IVS groups in figure 9, where we provide time series of both the actual and fitted implied volatilities and corresponding residuals for two center groups and four corner groups. By analyzing the upper two subfigures, we can verify an accurate fit of the general DFM model in center IVS groups. However, the other four plots confirm the warnings of Van der Wel et al. (2016) that the DFM fits of corner IVS groups are problematically worse. In particular, we observe observation residuals that do not look like merely white noise, with sufficient peaks and volatility clustering.¹² However, previously reported findings might indicate that these deviations from white noise processes can be partly mitigated by including GARCH disturbances.

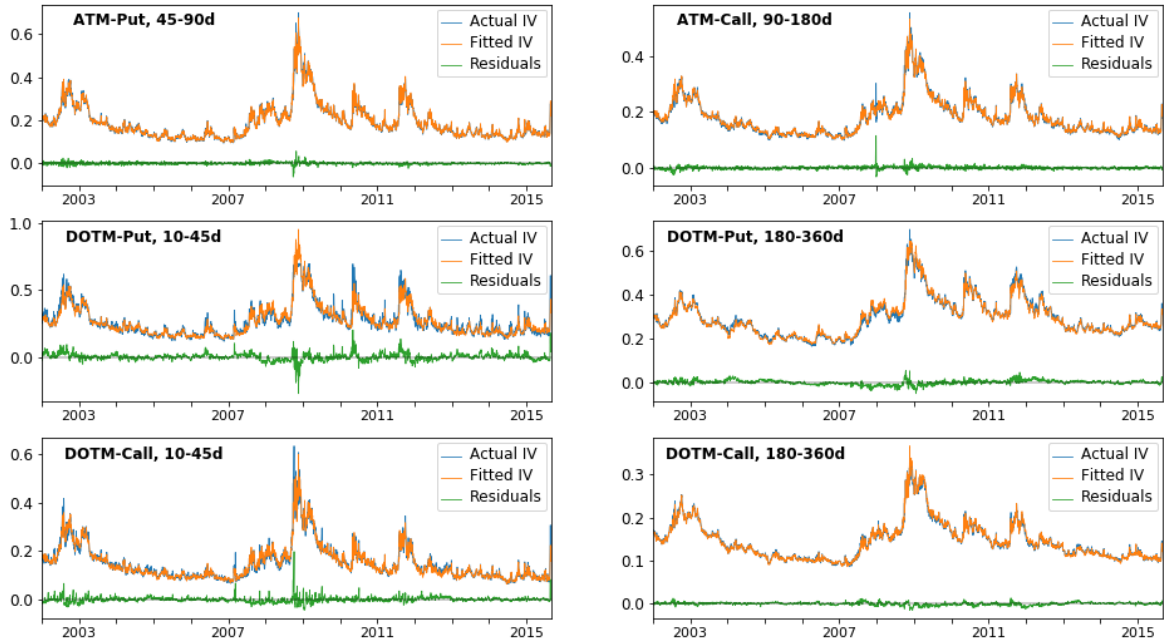


Figure 9: **Fit of General Dynamic Factor Model (DFM)**

Notes: These figures show the fit of the IVS estimated with the general dynamic factor model (DFM). In particular, we display time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals. In total, we document six different IVS groups. The upper two figures present two groups in the center of the IVS, whereas in the bottom four figures plots of the corner groups of the IVS are provided.

To examine whether these potential improvements are valid, we plot fits of a specific corner group for each of the four main dynamic factor models in figure 10. By considering the plots of the DFM-GARCH models, we still observe residuals that do not seem to behave like white noise. However, in comparison to the general DFM plot we observe slight improvements. The residuals of particularly the extended

¹²In addition to the residuals of the observation equation, we also checked time series of the residuals in the state equation and find that these residuals do not look like white noise as well. These analyses are left out of consideration and are available upon request.

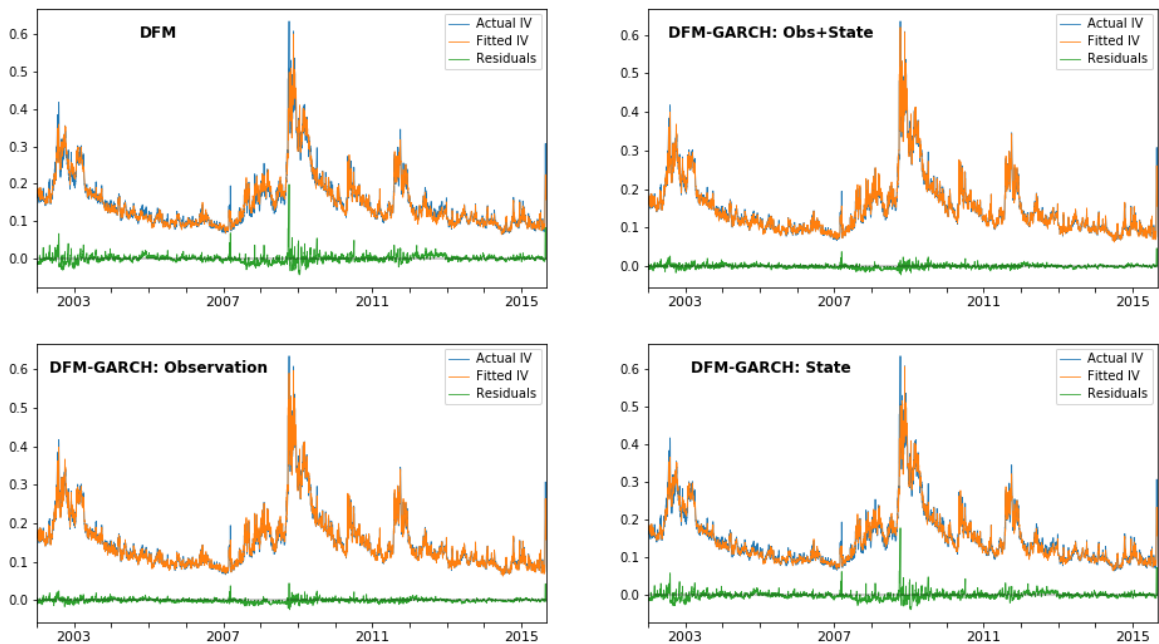


Figure 10: **Comparing Fit of Corner IVS Group (DOTM-Call, 10-45 days)**

Notes: These figures compare the fit of a specific corner group of the IVS (DOTM-Call, 10-45 days) estimated with four of our main dynamic factor models. In particular, we provide the fit of the IVS estimated with our three DFM-GARCH extensions relative to the fit of the IVS based on the general DFM model. The figures present time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals.

models involving GARCH disturbances in the observation equation show a more constant pattern and less volatility clustering. For other corner groups provided in the appendix in figures 16 to 18, we find similar but less convincing patterns. Full reports of the fit of all extended and restricted models can also be found in the appendix in figures 21 to 24. Hence, we again find indications that including GARCH disturbances in the observation equation of dynamic factor models can mitigate the problem of a weaker fit in corner groups of the IVS.

4.3.3 Out-of-Sample Forecasting Performance

In section 4.3.2 we already discussed that in-sample our DFM-GARCH setups including GARCH disturbances in the observation equation show an improved fit compared to their benchmarks. However, we are curious to find further insights on whether these models have predicting value in an out-of-sample setting based on individual option forecasts, which are more relevant for practitioners compared to delimited forecasts of the constructed IVS. In particular, it is interesting to explore whether the additional GARCH parameters cause overfitting issues in the dynamic factor models, possibly causing decreases in forecasting performances. In order to analyze the nature of the forecasting abilities of our DFM-GARCH models, we therefore report out-of-sample forecasting performances in table 7. Here, we explore one-day-ahead forecasts for individual options running from December 2005 until August 2015 by considering a moving window of 1,000 observations, which comes down to a time span of approximately 4 years. Specifically, we report corresponding average RMSE and MCP statistics for our DFM-GARCH setups and their benchmarks. In order to assess whether these differences in forecasting performances are statistically significant, we additionally perform the equal predictive ability test proposed by Diebold and Mariano (2002). In particular, we compare the one-day-ahead forecasting performances of the general DFM model compared to the other dynamic factor and benchmark models. For this purpose, we adopt

the heteroskedasticity and autocorrelation consistent variance estimator from Newey and West (1986). Inspired by among others Goncalves and Guidolin (2006), we consider performance indicators for both statistical evaluation measures. More specifically, for testing the differences in RMSE statistics we compare squared forecast errors of the general DFM model with the squared forecast errors of the other models. Similarly, for testing the MCP significances we consider the differences in correctly predicted directions of change as described in section 4.2. As null hypotheses, we assume the RMSE and MCP statistics of the compared models to be equal. Overall, the general DFM model outperforms all other models significantly on the RMSE measure with a 1% confidence interval. More specifically, the general DFM model has the lowest out-of-sample RMSE of 0.0481, whereas it scores the highest MCP of 54.11%. In comparison, all DFM-GARCH setups only predict the direction of change in implied volatility correctly for around 53% of the cases, from which only the MCP of the setup with GARCH disturbances in both equations is significantly different from the general DFM model. In addition, corresponding RMSEs are significantly higher than the general DFM model, with values ranging from 0.0605 to 0.0652. Still, these extended DFM-GARCH setups fairly outperform the restricted RFM model. Although all other likelihood-based dynamic factor models show better performances than the additional benchmark models, the RFM model is even outperformed by the random walk model (RMSE) and the two-step VAR model (MCP). Hence, in line with our in-sample findings the RFM model is less suitable for predicting the IVS compared to general DFM(-GARCH) models.

Table 7: **Out-of-Sample Forecasting Performance**

		RMSE		MCP	
		Value ($\times 10^{-2}$)	DM	Value (%)	DM
DFM-GARCH	Observation	6.45	-4.65***	53.26	1.02
	State	6.05	-3.92***	53.10	1.16
	Obs+State	6.52	-4.81***	52.69	1.59*
Benchmarks	DFM	4.81	NA	54.11	NA
	RFM	9.02	-17.78***	51.39	3.47***
	2-step VAR	9.28	-18.32***	51.84	2.81***
	RW	7.72	-12.63***	NA	NA

Notes: This table provides statistics regarding the out-of-sample forecasting performance of our DFM-GARCH models relative to their benchmark models for our full out-of-sample period from December 2005 until August 2015. Specifically, we report average root mean squared errors (RMSE) and mean correct predicted directions of change (MCP) for forecast horizon $h = 1$, including corresponding Diebold-Mariano (DM) statistics. In addition to our main dynamic factor models, we provide out-of-sample forecasting performance statistics for the random walk (RW) and the two-step vector-autoregressive (2-step VAR) benchmark models. In particular, */**/** denotes statistically significant Diebold-Mariano results at an $\alpha = 10\%/5\%/1\%$ significance level.

By comparing the three DFM-GARCH setups to each other, we observe different patterns compared to in-sample estimation. On the one hand, including GARCH disturbances in both equations scores the highest RMSE, whereas GARCH disturbances in only the state equation even performs better than the observation equation variant in terms of RMSE. On the other hand, based on the MCP statistics we observe a distinctive pattern. Here, the observation equation variant performs best, whereas the model with GARCH disturbances in both equations show lowest MCP of our extended setups. However, these differences in MCP are relatively small, especially between the state equation and observation equation variants. The conflicting patterns between those two setups in terms of RMSE and MCP can be explained by the fact that, although it predicts the direction of change correctly more often, the

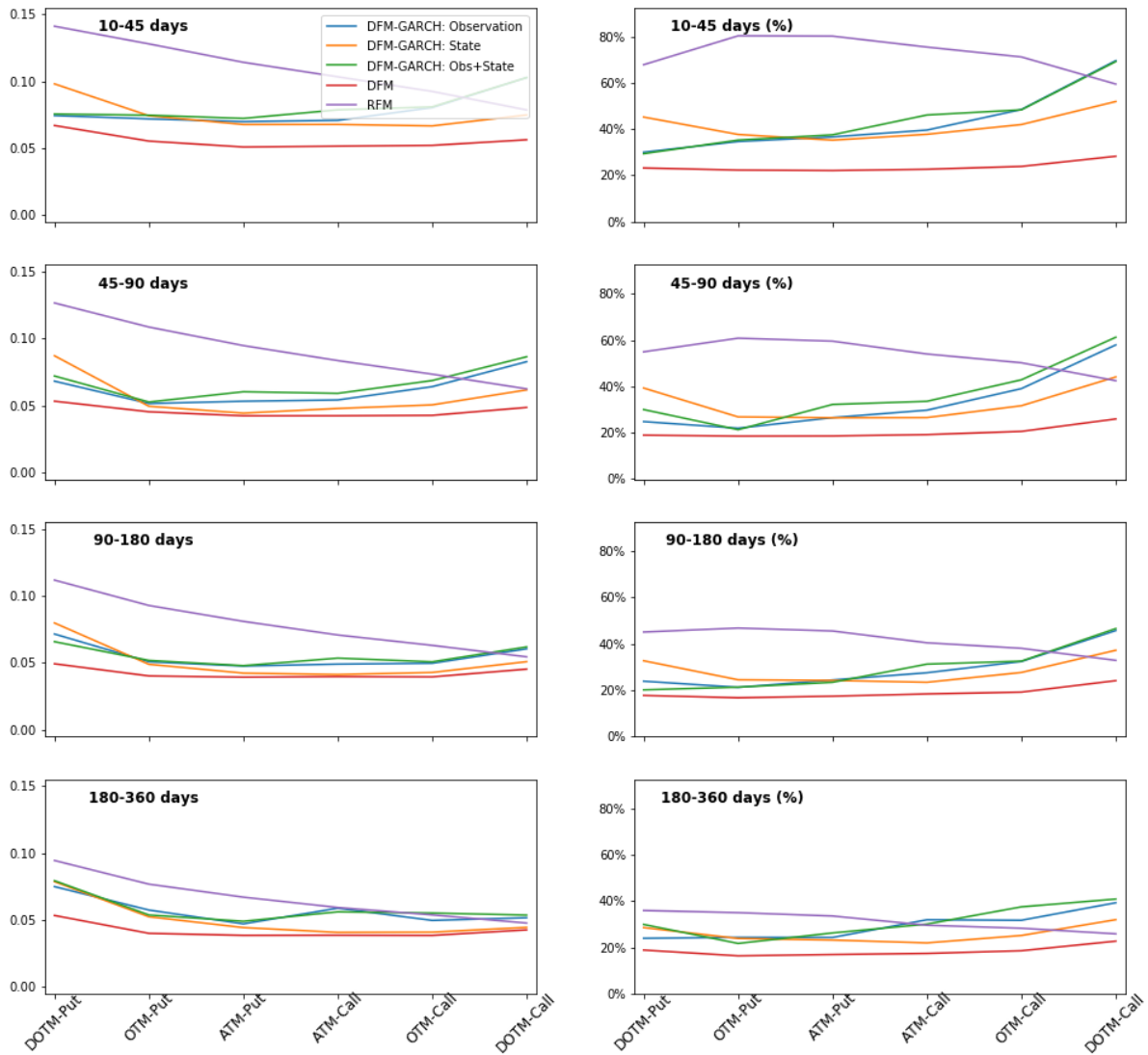


Figure 11: Out-of-Sample Root Mean Squared (Percentage) Errors

Notes: These figures show the absolute and proportional out-of-sample forecasting errors per IVS group of our main dynamic factor models for our full out-of-sample period from December 2005 until August 2015. Specifically, the left figures display the absolute out-of-sample root mean squared errors (RMSE), whereas the right figures display corresponding forecasting errors in terms of percentage. For all four maturity categories (from top to bottom), the forecasting errors are plotted along the six moneyness categories.

predictions of the observation equation variant are more volatile or more excessive when the direction of change is predicted incorrectly. In an attempt to find most profitable trading strategies in section 5, the MCP statistics are considered to be more relevant for our purposes. Overall, these findings indicate that including GARCH disturbances in only the observation equation slightly outperforms the other DFM-GARCH setups economically. However, in a plain out-of-sample setting these extended models are in general still outperformed significantly by the general DFM model.

Similar to our in-sample evaluation procedure, we proceed with performance evaluation per group in an attempt to gain further insights on the predictability of the entire IVS. Specifically, figure 11 displays the RMSE in absolute (left subfigures) and percentage (right subfigures) terms for all IVS groups. In general, we again observe a trend that center groups perform better than groups in the corner of the IVS, both in terms of absolute and percentage RMSEs. More specifically, we observe higher RMSEs in groups with maturities 10-45 days and 180-360 days, whereas within each subfigure we detect convex arches. In nearly all IVS groups the general DFM model shows lowest RMSEs, followed by a relatively good fitting DFM-GARCH setup involving GARCH disturbances in the state equation. The remaining DFM-GARCH setups perform worse, especially in case of DOTM-Call options. Once more, we find worst results for the RFM model, confirming that the remaining co-movements in the IVS do not fully match with the economically plausible volatility smile and term structure effects as imposed by the RFM model.

So far, we analyzed forecasting performances by considering average statistics and statistics per IVS group over the full out-of-sample period. However, to gain more insights on whether our models have predicting value in real life, it seems useful to evaluate performances dynamically by plotting the statistical measures over time. Therefore, figure 12 displays time series of RMSE and MCP statistics for our main likelihood-based dynamic factor models, using a one-year moving window consisting of 252 observations. By analyzing the plots, we find multiple remarkable things standing out. First, the relatively weak performances of the RFM model are not valid over the full time period. In particular, during the worldwide financial crisis in 2008 the RMSE of this model grows strongly compared to the other models. At the same time, the bottom figure displays great improvements in terms of MCP, from around 48% by mid-2007 to above 56% by mid-2008. Afterwards, the performances worsen again with negative peaks around relatively calm periods like 2013-2015. On the contrary, the general DFM model and all DFM-GARCH setups show reversed patterns. In detail, during the financial crisis we observe worst performances in terms of RMSE, whereas these performances show improvements after 2012. In contrast, the MCP statistics of these DFM(-GARCH) models show a weaker relationship with the state of the global economy. However, the figures do show some overlap between maximum RMSE peaks and minimum MCP peaks. Compared to each other, the four DFM(-GARCH) models exhibit notable differences. For example, in terms of MCP these models alternate over time in being the best performing model. The DFM-GARCH setup involving an adjusted state equation only show highest MCP in 2008 and lowest MCP in 2012/2013, whereas including GARCH disturbances in both equations causes completely reversed results. Hence, except for the dynamical performance of the RFM model these results do barely point out additional insights on differences between our DFM-GARCH setups and the general DFM model. Therefore, practitioners have a lack of preference between the unrestricted dynamic factor models when they attempt to gain information on expected market volatility by using implied volatility forecasts. Given that these relative accuracy and performance measures of our DFM(-GARCH) models vary over time, it might be of interest to construct forecast combinations in order to improve overall forecasting performances of dynamic factor models. However, these forecast combinations are beyond the scope of this thesis and can be relevant for further research.

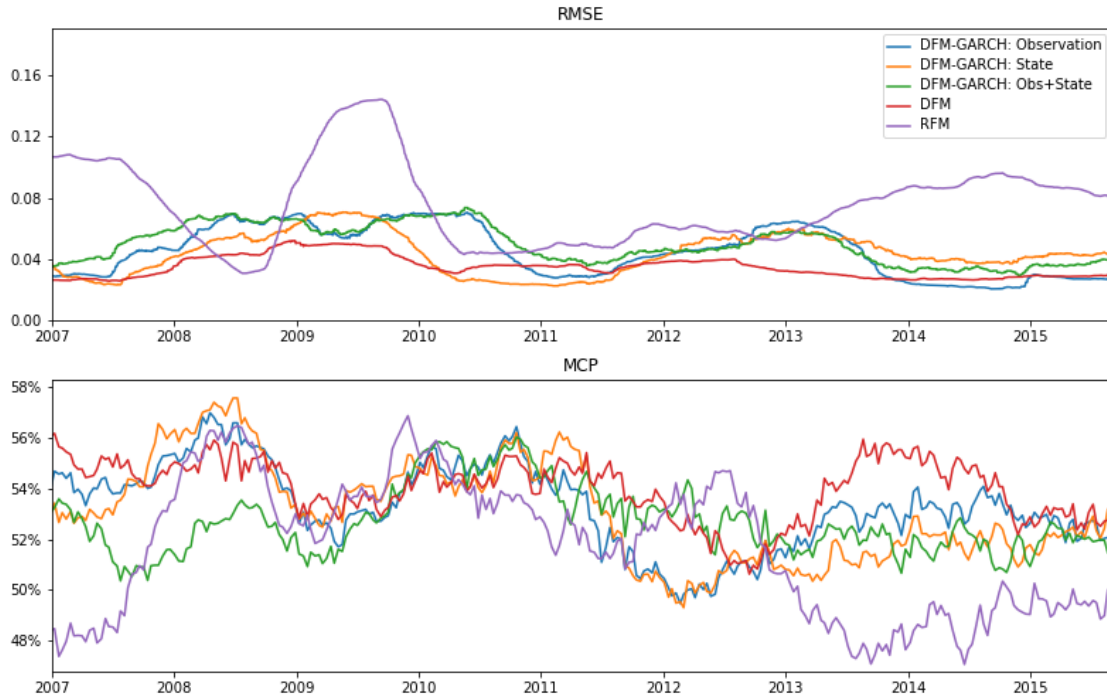


Figure 12: **Dynamic Out-of-Sample Forecasting Performance**

Notes: These figures show the dynamic out-of-sample forecasting performances over time for our main dynamic factor models. In particular, the upper figure plots the average root mean squared error (RMSE) and the bottom figure plots the mean correct predicted direction of change (MCP), using a one-year moving window consisting of 252 observations.

Currently, we focus on forecast horizons equal to $h = 1$ day. However, to gain a better understanding of predicting value of dynamic factor models, we provide additional average out-of-sample statistics for longer forecast horizons in figure 13. In case of using implied volatility forecasts within option pricing applications, practitioners particularly can benefit from these predicted market volatility further ahead in the future. Specifically, we plot RMSE and MCP statistics for our main models and benchmark models

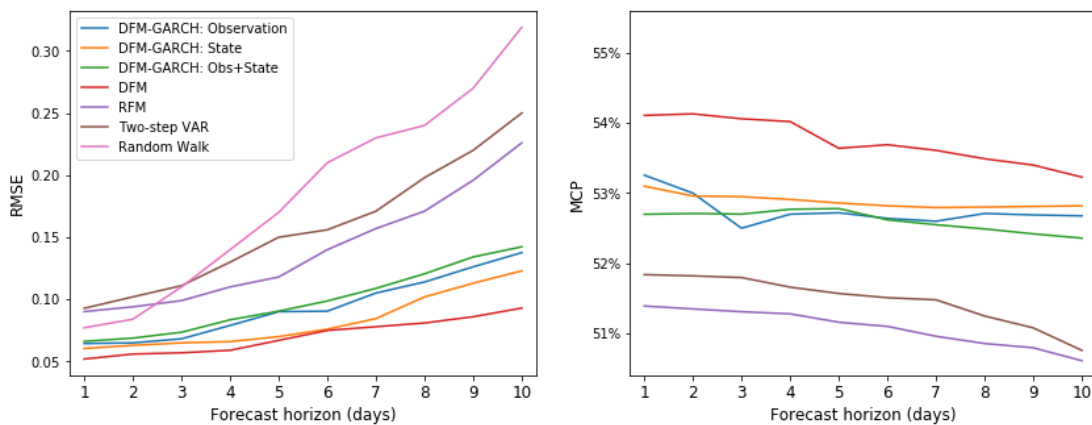


Figure 13: **Performance for Longer Forecast Horizons**

Notes: These figures show the out-of-sample forecasting performances of our main dynamic factor models and their benchmarks for different forecast horizons. In particular, the left figure plots the average root mean squared error (RMSE) and the right figure plots the mean correct predicted direction of change (MCP) for our full out-of-sample period from December 2005 until August 2015. Both figures provide results for forecast horizons $h = 1, 2, \dots, 10$ days.

as functions of the horizon h . As expected, for all models the RMSE increases for expanding forecast horizons, whereas MCP slightly declines accordingly. This is in line with our expectations based on the fact that, due to more insecurity, longer forecast horizons exhibit more volatility in the implied volatility forecasts. Besides, the general DFM model dominates both rankings based on RMSE and MCP at all horizons. Moreover, the differences between all models increase with the horizon. Hence, these figures broaden our findings on the relative differences between the basic and extended dynamic factor models over longer forecast horizons. Therefore, in their attempt to gain insights on future market volatility for option pricing or risk management purposes, practitioners will have similar preferences for dynamic factor model setups in case of predicting implied volatility further ahead compared to one-day-ahead forecasts.

5 Economic Evaluation

In section 4 we reported evidence of statistical significance in the predictability of the IVS by estimating general and extended dynamic factor models. However, these statistical findings do not directly imply that our models are valuable for traders and investors in real life. For example, it could be the case that our models appear to be successful based on statistical evaluation methods, while applying these same models in actual trading strategies can result in losses due to overfitting caused by the large number of estimated parameters. Therefore, in this section we explore whether predictability of the IVS is not only a statistical fact, but indeed signals imperfection of the S&P 500 option market. More specifically, we evaluate economic significance of the predictability by examining whether trading strategies involving our dynamic factor models can potentially generate profitable results. By evaluating economic potential of our GARCH extensions to dynamic factor models for the IVS, we also attempt to determine whether the implied volatility forecasts are economically affected by overfitting issues due to the additional GARCH parameters. In particular, section 5.1 illustrates how we economically evaluate the various dynamic factor models. Thereafter, related results regarding both with and without transaction costs are provided in section 5.2.

5.1 Constructing the Trading Strategies

Related literature shows several ways to simulate trading strategies in order to examine profitability of dynamic factor models. Section 5.1.1 outlines the trading strategies we implement in this thesis extensively. By evaluating these simple strategies, we report lower bounds of the actual profits traders and investors theoretically can realize. However, evaluating trading strategies without including additional costs does not provide realistic outcomes of actual profitability. Therefore, in section 5.1.1 we provide methods to incorporate transaction costs in our trading strategies in order to gain more insights on whether these strategies could be profitable in real life.

5.1.1 Trading Strategies

Evaluating economic performances of estimated models by simulating trading strategies is not new in option markets. For example, Day and Lewis (1992), Harvey and Whaley (1992) and Noh et al. (1994) already consider this additional evaluation procedure by introducing several different simulation setups. Afterwards, Goncalves and Guidolin (2006) and Bernales and Guidolin (2014) adopt similar setups in more recent developed dynamic factor models regarding predictability in implied volatilities. In line with their best practices, we exploit one-day-ahead forecasts of all models by investing a fixed amount of \$1.000 in analytically created portfolios on a daily basis. By analyzing associated average returns and their standard deviations, we eventually provide insights on whether our IVS dynamic factor models can theoretically be profitable for investors. For this purpose, we make use of the MCP measure introduced in section 4.2, which evaluates the accuracy of the correctly predicted directions of change in implied volatilities. Overall, this measure is widely used in numerous trading strategies and is therefore crucial in financial decision making. Correct predictions enable investors to implement well-considered decisions successfully. In case of incorrect predictions, traders usually make conflicting decisions and are therefore exposed to undesirable risks. Hence, success of a model's trading strategy largely depends on this MCP accuracy.

Inspired by Bernales and Guidolin (2014), we first introduce a plain-vanilla straddle trading strategy that is free of risk caused by changes in the prices of the underlying. This strategy requires us to trade combinations of both put and call option contracts with the same time-to-maturity and strike price.

Here, a long straddle involves buying these combinations of option contracts and can be regarded as a pure bet on higher future volatility levels. Likewise, a short straddle involves selling these contracts and can be regarded as a pure bet on lower future volatility levels. We determine whether we buy or sell options by evaluating whether our models predict an increase or decrease in one-day-ahead implied volatility, respectively. Again, these one-day-ahead forecasts of individual option contracts are obtained by interpolating the implied volatility forecasts of the entire surface on a daily basis. Besides, we follow Bernales and Guidolin (2014) once more by adopting their delta-hedged portfolios as a second trading strategy. Here, we combine our investments in put and call options with certain trades in the underlying stock. By selecting these specific trading volumes on the basis of the option's Δ , we manage to form delta-hedged positions. By daily re-balancing portfolios in both cases, we retain our constant investments of \$1.000 over the full back-testing period. Initially, we illustrate economic performances of our models by using full sets of option contracts over the entire IVS. In order to gain more insights on potential improvements when focusing our trading strategies on specific IVS groups, we also set up an additional framework to apply similar trading strategies within single categories.

In order to illustrate our trading strategies analytically, we start with introducing our straddle strategy by defining $Q_{t,+}$ ($Q_{t,-}$) as the subset of call and put option contracts that should be purchased (sold) following constructed buy (sell) signals. Next, we define $C_{m,t}$ ($P_{m,t}$) as the call (put) price of a specific option contract m within these subsets. In addition, we only select a call (put) contract if all relevant data is available for its current value $C_{m,t}$ ($P_{m,t}$), the value of its counterpart $P_{m,t}$ ($C_{m,t}$), and next day's values of both itself $C_{m,t+1}$ ($P_{m,t+1}$) and its counterpart $P_{m,t+1}$ ($C_{m,t+1}$). Overall, at day t the total value of all formed straddle positions $V_t^{straddle}$ is given by

$$V_t^{str} = \sum_{m \in Q_{t,+}} (C_{m,t} + P_{m,t}) - \sum_{m \in Q_{t,-}} (C_{m,t} + P_{m,t}) \quad (27)$$

where we consider all m options within the subsets of respectively the buy and sell contracts $Q_{t,+}$ and $Q_{t,-}$. Next, in case of a positive portfolio value $V_t^{str} > 0$, we define the one-day net gain of a corresponding long straddle G_{t+1}^{str-L} by

$$G_{t+1}^{str-L} = X_t^{str} \left[\sum_{m \in Q_{t,+}} ((C_{m,t+1} + P_{m,t+1}) - (C_{m,t} + P_{m,t})) \right] + X_t^{str} \left[\sum_{m \in Q_{t,-}} (-(C_{m,t+1} + P_{m,t+1}) + (C_{m,t} + P_{m,t})) \right] \quad (28)$$

where we buy a number of X_t^{str} portfolio units. Alternatively, in case of a negative portfolio value $V_t^{str} < 0$, we sell a number of X_t^{str} portfolio units. Here, the number of portfolio units in both cases is defined by

$$X_t^{str} = \frac{\$1.000}{|V_t^{str}|} \quad (29)$$

However, in the latter case we retain our untouched capital due to selling portfolios instead of purchasing. Hence, we define a specific one-day net gain for short straddle positions given by

$$G_{t+1}^{str-S} = G_{t+1}^{str-L} + \$2.000 * \left(\exp\left(\frac{r_t}{252}\right) - 1 \right) \quad (30)$$

where we invest our initial \$1.000 plus an additional \$1.000 from the sale at the one-day riskfree rate r_t as discussed in section 2.1. Moreover, in the event that none of the available options are indicated by the model's predictions to either buy or sell, we invest our initial capital fully at this riskfree rate.

Likewise, to illustrate our delta-hedged trading strategy we define the total daily value of all delta-hedged positions V_t^{DH} by

$$\begin{aligned}
V_t^{DH} = & \sum_{m \in Q_{t,+}^{call}} (C_{m,t} - S_t \Delta_{m,t}^C) + \sum_{m \in Q_{t,+}^{put}} (P_{m,t} + S_t \Delta_{m,t}^P) \\
& - \sum_{m \in Q_{t,-}^{call}} (C_{m,t} - S_t \Delta_{m,t}^C) - \sum_{m \in Q_{t,-}^{put}} (P_{m,t} + S_t \Delta_{m,t}^P)
\end{aligned} \tag{31}$$

where S_t is the price of the underlying S&P 500 index fund at time t , and $\Delta_{m,t}^C$ ($\Delta_{m,t}^P$) is the absolute value of the call (put) option's Δ . In line with our straddle definitions, we also define $Q_{t,+}^{call}$ ($Q_{t,+}^{put}$) as the subset of call (put) option contracts that should be purchased following constructed buy signals, whereas $Q_{t,-}^{call}$ ($Q_{t,-}^{put}$) denotes the subset of call (put) option contracts that should be sold following constructed sell signals. In addition, corresponding one-day net gain G_{t+1}^{DH-L} in case of going long due to a positive portfolio value $V_t^{DH} > 0$, is given by

$$\begin{aligned}
G_{t+1}^{DH-L} = & X^{DH} \left[\sum_{m \in Q_{t,+}^{call}} ((C_{m,t+1} - S_{t+1} \Delta_{m,t}^C) - (C_{m,t} - S_t \Delta_{m,t}^C)) \right] \\
& + X^{DH} \left[\sum_{m \in Q_{t,+}^{put}} ((P_{m,t+1} + S_{t+1} \Delta_{m,t}^P) - (P_{m,t} + S_t \Delta_{m,t}^P)) \right] \\
& + X^{DH} \left[\sum_{m \in Q_{t,+}^{call}} -(C_{m,t+1} - S_{t+1} \Delta_{m,t}^C) + (C_{m,t} - S_t \Delta_{m,t}^C) \right] \\
& + X^{DH} \left[\sum_{m \in Q_{t,+}^{put}} -(P_{m,t+1} + S_{t+1} \Delta_{m,t}^P) + (P_{m,t} + S_t \Delta_{m,t}^P) \right]
\end{aligned} \tag{32}$$

Similar to our straddle trading strategy, we calculate the number of delta-hedged portfolios X_t^{DH} and one-day net gain for short delta-hedged positions G_{t+1}^{DH-S} using equations 29 and 30.

In order to avoid extreme portfolio positions for both trading strategies, we implement two general constraints. First, we attempt to avoid unreliable forecasts by demanding the absolute predicted change in implied volatility to be higher than 1%. Besides, it may occur that our total portfolio value V_t is extremely small due to almost equally valued partial buying and selling portfolios. Then, our trading strategy generally imposes to buy or sell an extremely large number of portfolio units X_t , resulting in excessive risks and unrealistic trading positions. Therefore, as a second constraint we restrict our strategies to only perform a trade only if the total portfolio value is $V_t^{straddle} > \$1$.

In section 3.2 we already provided two benchmark models for forecasting the IVS in order to examine and compare the performances of our dynamic factor models. To be able to compare the profitability results of these models to an even better extent, we also introduce two additional trading benchmark strategies. First, we consider a passive benchmark strategy consisting of an effortless investment of \$1.000 at the riskfree rate over the entire time period. Here, we use the global Fama and French riskfree rate indicators as discussed in section 2.1 once more, resulting in a benchmark that only yields the time value of money. Associated one-day net gain is given by

$$G_{t+1}^{RFR} = \$1.000 * \left(\exp\left(\frac{r_t}{252}\right) - 1 \right) \tag{33}$$

As a second trading strategy benchmark, we implement the underlying buy and hold strategy along the lines of Chalamandaris and Tsekrekos (2010). This strategy daily invests \$1.000 fully in the underlying, which in our case is the S&P 500 index fund with current value S_t . Motivated by its illustrative representation of the S&P 500 market developments, this strategy is referred to as market representation and can be evaluated by analyzing its one-day net gain given by

$$G_{t+1}^{Market} = \$1.000 * \frac{(S_{t+1} - S_t)}{S_t} \quad (34)$$

Accordingly, we report the average daily profits and corresponding standard deviations for all models in following results sections. In addition, we provide further insights on the value of these models by providing t -test statistics and Sharpe ratios. The two-tailed t -test statistics naturally examine whether the returns of the trading strategies implemented on all models are significantly different from zero. Besides, Sharpe ratios are proper measures of profitability when investors have mean-variance preferences. To obtain Sharpe ratios we calculate risk-adjusted returns, where we consider similar daily riskfree rates as in our first benchmark strategy. In detail, Sharpe ratios are defined by

$$SR_p = \frac{\bar{r}_p - r_f}{\sigma_p} \quad (35)$$

where \bar{r}_p and σ_p denote expected return and standard deviation of portfolio p , and r_f again denotes the average daily riskfree rate. In contrast to meaningful positive Sharpe ratios, we do not report negative Sharpe ratios due to its misleading interpretation.¹³ By evaluating and comparing our models by means of these criteria, we gain a better understanding on whether the models are attractive for traders and investors.

5.1.2 Transaction Costs

Initially, we attempt to evaluate economic performances of our dynamic factor models by applying basic trading strategies. However, these trading strategies suffer from limitations. In particular, we leave out the impact of transaction costs, resulting in unrealistic potential outcomes for traders and investors in real financial markets. Hence, we introduce methods to incorporate transaction costs in order to test sensitivity of our trading strategies to these additional costs. Related literature already shows several ways to incorporate these costs. For example, Goncalves and Guidolin (2006) suggest to apply two fixed levels of unit costs, namely \$0.05 and \$0.125 per traded contract. However, a more realistic inclusion of transaction costs is proposed by Bernales and Guidolin (2014), who suggest to use effective bid-ask spreads. Although Battalio et al. (2004) support the benefits of this idea, they also state that the effective bid-ask spread is generally different from the quoted spread available in our data set. Therefore, we adopt their findings by using a conservative effective bid-ask spread equal to 0.5 times the quoted spread in line with Bernales and Guidolin (2014). As a robustness check, we impose \$0.05 fixed transaction costs per trade as suggested by Goncalves and Guidolin (2006) as an alternative approach to include transaction costs. Still, we find similar conclusions as for imposing transaction using the bid-ask spread. Hence, we leave this alternative transaction costs approach out of consideration. In both cases, transaction costs are incorporated by subtracting them from our one-day net gains on a daily basis. Hence, in section 5.2.2 we provide additional results with transaction costs included to gain further insights on whether our dynamic factor models could be used by traders and investors to gain profits theoretically.

¹³When the mean return is negative, the Sharpe ratio improves when volatility is higher. This results in misleading impressions of economic performances when analyzing negative Sharpe ratios.

5.2 Economic Results

After having illustrated our trading strategies and additional implementation of transaction costs in section 5.1, we report whether these strategies can potentially be profitable in this section. Specifically, the results of plain trading strategies without transaction costs are discussed in section 5.2.1, whereas section 5.2.2 provides corresponding trading results after including transaction costs.

5.2.1 Trading Results before Transaction Costs

Table 8 reports summary statistics of trading results before transaction costs for all dynamic factor models and corresponding benchmarks, obtained by applying both straddle and delta-hedged trading strategies. Overall, we observe that without taking transaction costs into account, all trading strategies

Table 8: **Trading Results before Transaction Costs**

		Before Transaction Costs			
		Mean profit (%)	Std. profit (%)	t -test	Sharpe ratio (%)
Panel A: Straddle Portfolios					
DFM-GARCH	Observation	1.31	14.78	4.38***	8.76
	State	1.26	14.77	4.20***	8.42
	Obs+State	1.15	14.32	3.97***	7.94
IVS Benchmarks	DFM	1.47	16.51	4.39***	8.80
	RFM	0.33	9.77	1.69*	3.27
	2-step VAR	0.54	12.19	2.18**	4.31
Trading Benchmarks	Market	0.03	1.31	1.02	0.97
	T-bill	0.01	0.01	87.07***	0
Panel B: Delta-Hedged Portfolios					
		Before Transaction Costs			
		Mean profit (%)	Std. profit (%)	t -test	Sharpe ratio (%)
DFM-GARCH	Observation	0.83	7.39	5.52***	10.99
	State	0.85	7.21	5.80***	11.55
	Obs+State	0.65	7.58	4.21***	8.33
IVS Benchmarks	DFM	0.95	7.79	6.02***	12.00
	RFM	0.64	5.79	5.43***	10.81
	2-step VAR	0.48	7.02	3.37***	6.63
Trading Benchmarks	Market	0.03	1.31	1.02	0.97
	T-bill	0.01	0.01	87.07***	0

Notes: This table provides economic summary statistics excluding transaction costs for our recursive out-of-sample implementation of trading strategies on S&P 500 index options over the full out-of-sample period from December 2005 until August 2015. In particular, we provide average returns and corresponding standard deviations of straddle portfolio strategies (Panel A) and delta-hedged portfolio strategies (Panel B) applied to our DFM-GARCH models and associated benchmark models. In addition, we document trading results of two trading benchmarks, consisting of the S&P 500 index fund (Market) and a plain T-bill investment. In order to be able to compare risk-adjusted returns of all trading strategies, we report additional t -test statistics including significance levels and Sharpe ratios as defined in section 5.1.1. In particular, */**/** denotes statistically significant t -test results at an $\alpha = 10\%/5\%/1\%$ significance level.

based on dynamic factor models for the IVS result in positive average returns. Hence, all our focus and benchmark models outperform both trading benchmarks significantly. In terms of mean profit for both trading strategies, we find the highest average returns for the general DFM model. However, this model also exhibits highest standard deviations, resulting in more risky decisions when implementing these strategies in real life. Therefore, corresponding risk-adjusted returns in terms of t -test statistics (4.39 and 6.02) and Sharpe ratios (8.80 and 12.00) only show minimal differences compared to other dynamic factor models. In particular, our DFM-GARCH setups with GARCH disturbances in only one of the state space equations also perform relatively well. In case of straddle portfolio strategies, including GARCH disturbances in the observation equation results in the second highest daily average return of 1.31%, with a significant t -test statistic of 4.38 and Sharpe ratio equal to 8.76. Including GARCH disturbances in the state equation is the best alternative in case of delta-hedged portfolio strategies, with an even more significant t -test statistic of 5.80 and Sharpe ratio equal to 11.55. Straddle portfolio strategies generally show substantially higher profits compared to delta-hedged portfolio strategies, potentially being caused by different decision criteria and the inclusion of the index price S_t in the formulation of the delta-hedged strategy from equation 31. More precisely, the straddle strategy is expected to incorporate all available information of the IVS, whereas success of the delta-hedged strategy partially depends on forecasts of the price of the underlying index. Although our DFM-GARCH setups differ in both mean returns and standard deviations, their general trading results are very similar compared to each other. Hence, we can conclude that for economic purposes, it barely matters how to include GARCH disturbances in dynamic factor models for the IVS. Even after comparing these models to the general DFM model, we hardly find significant economic differences. Therefore, the general DFM benchmark seems to be the best alternative in case of economically exploiting dynamic factor models. But although risk-adjusted returns of all our DFM-GARCH setups are only slightly lower than the general DFM model, these extended models still outperform the market and a riskfree T-bill investment significantly. For that reason, we can also conclude that without taking transaction costs into account, our DFM-GARCH models have sufficient economic value in predicting the dynamics of the IVS.

In line with our previously documented statistical out-of-sample evaluation, we analyze dynamic forecasting performance of all dynamic factor models by evaluating average returns over time in figure 14. Again, we consider a one-year moving window for both trading strategies, running from December 2005 until August 2015. Evidently, all unrestricted dynamic factor models follow similar patterns. More specifically, in times of economic uncertainty these four models show declining performances, like during the financial crisis in 2008 and the debt-ceiling crisis in 2011. On the contrary, the models show performance improvements in relatively prosperous times like 2013-2015. Interestingly, we also observe a relatively flat curve for the restricted RFM model. We expect this to be caused by less risky trading decisions due to less sensitivity in the implied volatility forecasts. For example, in turbulent economic times the DFM(-GARCH) models react strongly with suggestive forecasts, whereas the RFM model absorbs relatively less disturbances in its forecasts. This pattern was already visible statistically in figure 12, where the RMSE of the RFM model increases strongly in times of crises.

Initially, we evaluate economic potential of our dynamic factor models by applying trading strategies on the full set of available option contracts. However, a full surface consisting of both center and corner groups exhibits various dynamics. Based on the statistical differences between the IVS groups, we expect these dynamics to cause a wide variety in economic performances of the models. Therefore, in order to gain a better understanding of the dynamics of the IVS in relation to our dynamic factor models, we also evaluate all IVS groups individually in an economic setting. More specifically, we apply our trading

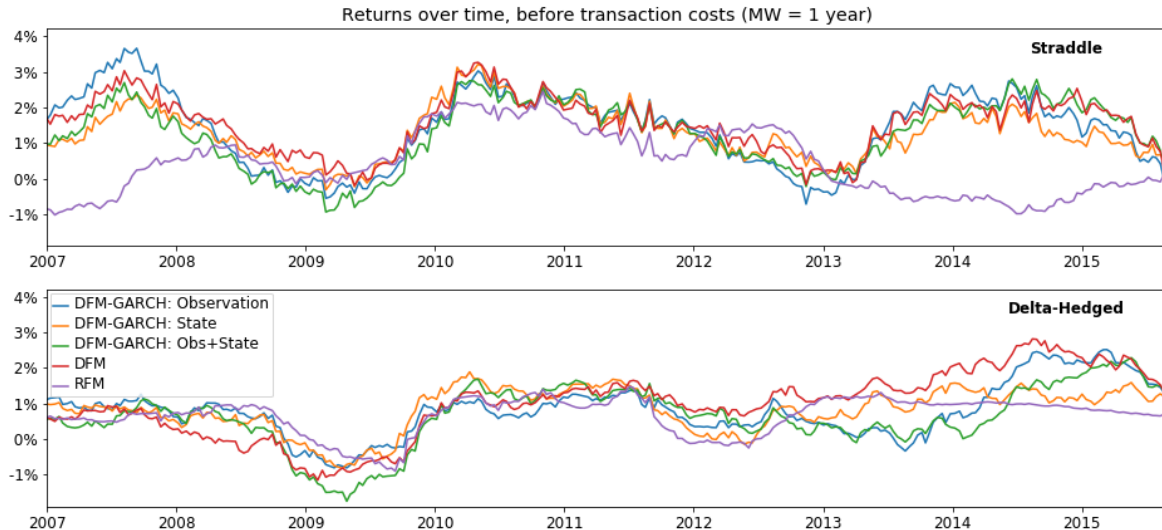


Figure 14: **Dynamic Profitability Performance before Transaction Costs**

Notes: These figures show the dynamic economic profitability performances over time for our main dynamic factor models, before including transaction costs. In particular, the upper figure plots the average returns obtained by applying straddle portfolio strategies, whereas the bottom figure plots the average returns obtained by applying delta-hedged portfolio strategies, using a one-year moving window consisting of 252 observations.

strategies within single IVS groups and report associated trading results in table 9. The table contains Sharpe ratios with corresponding t -test significance levels, and reveals several findings. First, applying straddle portfolio strategies within (D)OTM-Put categories, based on unrestricted dynamic factor models, results in highly significant profits. In the corner of the IVS where we consider DOTM-Put options with short maturities, we even document Sharpe ratios larger than 18. Hence, within these specific IVS groups our straddle trading strategy indicates to be most profitable. However, by analyzing corresponding call options we find a reversed pattern. Here, our models show insignificant profits in most cases with longer maturities, whereas trading within shortest maturity categories results in extreme losses.¹⁴ Exactly the opposite pattern is visible when we apply a delta-hedged trading strategy. This strategy performs extremely well for call options with shortest maturities, whereas corresponding put option categories show extreme losses. In center groups, we again find risk-adjusted returns that are not significant different from zero in most cases. Overall, we can conclude that applying our trading strategies within corner groups of the IVS is extremely risky and causes excessive results. Although reported ratios already take associated risks into account, we find extremely positive Sharpe ratios for some of these corner groups, whereas other corner groups show extremely negative average returns. This in line with our statistical findings, which suggest that our dynamic factor models are not able to fit the IVS fairly well in corner groups. Besides, option contracts in corner groups are already expected to be less liquid than contracts in center IVS groups. In an economic setting, we therefore find extreme profits and losses when applying trading strategies within these high risk corner groups.

Hence, without taking transaction costs into account we report the following four economic findings. First, all considered likelihood-based dynamic factor models show economic value in their predictability of the IVS. Second, our extended DFM-GARCH setups do not differ significantly in their potential to be exploited in profitable trading strategies. Even compared to the general DFM model, we hardly find significant differences. Hence, these findings indicate that including GARCH disturbances in dynamic

¹⁴In table 9, only Sharpe ratios and significance levels of associated t -tests are reported. For clarity purposes, corresponding average returns and standard deviations are left out of consideration and are available upon request.

Table 9: Trading Results per Group, before Transaction Costs

	Straddle Portfolios - Sharpe Ratio's					Delta-Hedged Portfolios - Sharpe Ratio's				
	DFM-GARCH					DFM-GARCH				
	Obs	State	Obs+State	DFM	RFM	Obs	State	Obs+State	DFM	RFM
DOTM- Put	10-45	18.13***	21.07***	20.52***	25.85***	NA	NA	NA	NA	NA
	45-90	20.59***	20.71***	19.20***	22.70***	8.89***	NA	NA	NA	NA
	90-180	15.39***	14.97***	12.18***	16.10***	13.17***	NA	NA	NA	NA
	180-360	7.89***	7.82***	3.17*	8.68***	5.08***	NA	2.07*	NA	0.60
OTM- Put	10-45	3.37**	5.93***	7.77***	5.72***	NA	2.92*	NA	1.20	11.82***
	45-90	8.84***	7.86***	9.09***	9.33***	NA	NA	NA	NA	NA
	90-180	8.45***	10.08***	5.52***	9.74***	2.12*	0.67	0.70	NA	0.16
	180-360	7.24***	7.93***	2.32*	6.90***	2.88*	3.50**	2.25*	2.88*	1.60
ATM- Put	10-45	NA	0.04	NA	NA	NA	11.74***	10.97***	11.65***	18.57***
	45-90	NA	2.15*	0.23	NA	NA	1.60	1.71	0.59	NA
	90-180	NA	3.14*	0.63	NA	NA	2.07*	1.00	NA	0.39
	180-360	2.07*	2.48*	NA	0.15	0.37	2.02*	2.11*	NA	2.29*
ATM- Call	10-45	NA	NA	NA	NA	NA	13.58***	9.25***	14.51***	22.01***
	45-90	NA	NA	NA	NA	0.19	NA	0.42	NA	0.95*
	90-180	NA	NA	NA	0.90	0.60	NA	NA	NA	1.82*
	180-360	NA	NA	NA	NA	NA	NA	1.44*	NA	NA
OTM- Call	10-45	NA	NA	NA	NA	NA	15.40***	13.74***	17.89***	24.45***
	45-90	NA	NA	NA	NA	NA	NA	NA	NA	NA
	90-180	NA	NA	NA	NA	2.47*	NA	NA	NA	NA
	180-360	0.14	1.74	NA	1.49	1.77	NA	NA	1.18*	NA
DOTM- Call	10-45	NA	NA	NA	NA	NA	11.99***	4.25***	12.68***	15.60***
	45-90	NA	NA	NA	NA	NA	NA	NA	NA	NA
	90-180	NA	NA	NA	NA	NA	NA	NA	NA	NA
	180-360	NA	NA	NA	NA	NA	1.82*	0.37	1.45	0.95

Notes: This table provides economic summary statistics before including transaction costs, for our recursive out-of-sample implementation of trading strategies per group over the full out-of-sample period running from December 2005 until August 2015. Here, we implement straddle portfolio and delta-hedged portfolio strategies on selected S&P 500 index options within specific IVS groups. For both trading strategies based on our main dynamic factor models, we provide risk-adjusted returns by means of Sharpe ratios and associated significance levels of corresponding t -test statistics. In particular, */**/** denotes statistically significant t -test results at an $\alpha = 10\%/5\%/1\%$ significance level. Due to its misleading interpretation, negative Sharpe ratios are omitted.

factor models has little impact on their economic results. Third, our trading results propose the RFM model to be a sound alternative for low risk investments. Although the co-movements in the IVS prove to mismatch with the economically restricted smile and term structure factors, this RFM model is apparently able to limit risk within investment decisions. Fourth, due to a weaker fit of the corner groups of the IVS, our dynamic factor models bring along high risks when applying trading strategies within these groups.

5.2.2 Trading Results after Transaction Costs

In previous section we discussed trading results before taking transaction costs into account. We provided several findings, including significant economic value of our constructed dynamic factor models in

Table 10: **Trading Results after Transaction Costs**

		After Transaction Costs			
		Mean profit (%)	Std. profit (%)	t -test	Sharpe ratio (%)
Panel A: Straddle Portfolios					
DFM-GARCH	Observation	-27.10	26.25	-51.0***	NA
	State	-27.71	25.52	-53.6***	NA
	Obs+State	-28.42	23.96	-58.7***	NA
IVS Benchmarks	DFM	-26.87	29.74	-44.6***	NA
	RFM	-29.22	15.91	-90.7***	NA
	2-step VAR	-28.89	18.42	-77.1***	NA
Trading Benchmarks	Market	0.03	1.31	1.0	1.0
	T-bill	0.01	0.01	87.1***	0.0
Panel B: Delta-Hedged Portfolios					
		After Transaction Costs			
		Mean profit (%)	Std. profit (%)	t -test	Sharpe ratio (%)
DFM-GARCH	Observation	-28.99	10.98	-130.4***	NA
	State	-28.97	10.77	-132.9***	NA
	Obs+State	-29.18	11.71	-123.2***	NA
IVS Benchmarks	DFM	-28.84	11.94	-119.2***	NA
	RFM	-29.25	7.54	-191.6***	NA
	2-step VAR	-29.56	9.92	-146.5***	NA
Trading Benchmarks	Market	0.03	1.31	1.0	1.0
	T-bill	0.01	0.01	87.1***	0.0

Notes: This table provides economic summary statistics including transaction costs for our recursive out-of-sample implementation of trading strategies on S&P 500 index options over the full out-of-sample period from December 2005 until August 2015. In particular, we provide average returns and corresponding standard deviations of straddle portfolio strategies (Panel A) and delta-hedged portfolio strategies (Panel B) applied to our DFM-GARCH models and associated benchmark models. In addition, we document trading results of two trading benchmarks, consisting of the S&P 500 index fund (Market) and a plain T-bill investment. In order to be able to compare risk-adjusted returns of all trading strategies, we report additional t -test statistics including significance levels and Sharpe ratios as defined in section 5.1.1. In particular, */**/** denotes statistically significant t -test results at an $\alpha = 10\%/5\%/1\%$ significance level. Due to its misleading interpretation, negative Sharpe ratios are omitted.

their predictability of the IVS. However, in real life these strategies suffer from the limitation of leaving transaction costs out of consideration. Hence, this section reports corresponding trading results after including transaction costs in table 10, by using a conservative effective bid-ask spread equal to 0.5 times the quoted spread. After including transaction costs, we find extremely negative returns for all models and trading strategies. In particular, we document daily average returns ranging from -26% to -30%. Hence, in these cases the trading strategies based on all our dynamic factor models are highly outperformed by the market and T-bill benchmarks. Relative to each other, our dynamic factor models show similar relationships as before including transaction costs. This is expected, as we consider similar option contracts and trading capitals in all cases. Again, we find the general DFM model to be the least poorly performing model for both strategies. The DFM-GARCH setups involving GARCH disturbances in only one of the two equations show similar results, whereas adjusting both equations results in a slightly worse performance. Hence, we can conclude that after taking transaction costs into account, all trading strategies based on our DFM(-GARCH) models are substantially outperformed by their trading strategy benchmarks. Therefore, although our dynamic factor models prove to be valuable in predicting the dynamics of the IVS in section 5.2.1, they appear to be ineffective when exploiting them directly into simple trading strategies due to the impact of transaction costs. However, practitioners might still be able to make use of their predicting value by implementing our dynamic factor models indirectly into option pricing and risk management applications.

In addition, we provide dynamic performances of all trading strategies after including transaction costs in figure 15. Once more, we find similar patterns as before including transaction costs. In general, we observe highly negative returns over the full out-of-sample period for both straddle and delta-hedged portfolios. Our general DFM and extended DFM-GARCH setups show similar patterns, with increasing performances in stable times and decreasing performances during crises. The restricted RFM model again exhibits relatively constant performances over the full sample. Hence, we can conclude that, in contrast to our findings before taking transaction costs into account, we are not able to directly exploit our dynamic factor models into profitable trading strategies in real life.

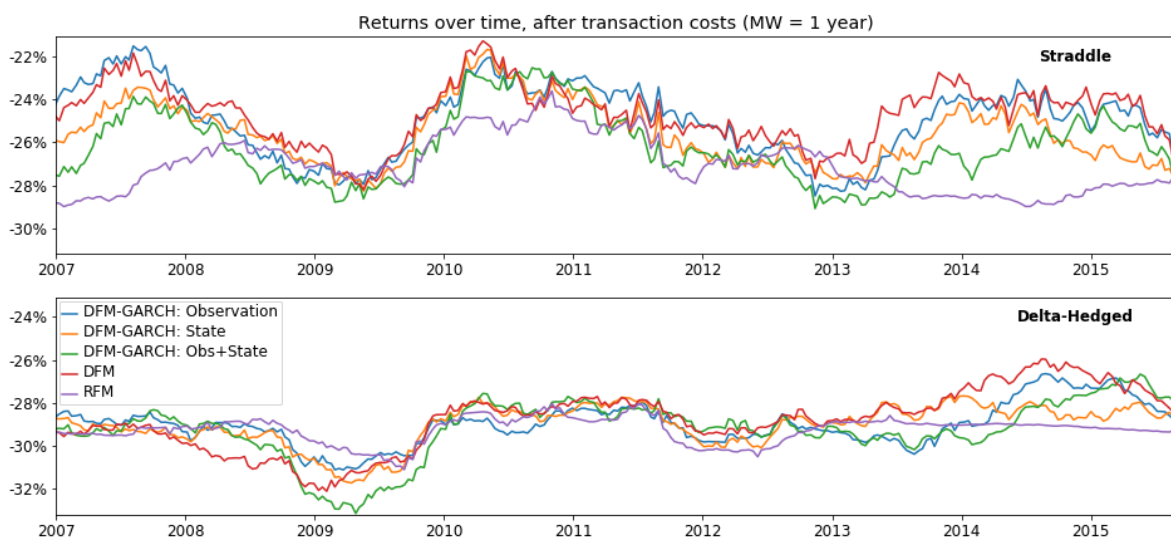


Figure 15: **Dynamic Profitability Performance after Transaction Costs**

Notes: These figures show the dynamic economic profitability performances over time for our main dynamic factor models, after including transaction costs. In particular, the upper figure plots the average returns obtained by applying straddle portfolio strategies, whereas the bottom figure plots the average returns obtained by applying delta-hedged portfolio strategies using a one-year moving window consisting of 252 observations.

6 Conclusion

The implied volatility surface explains the dynamics between different option contracts by representing the total set of implied volatilities across moneyness and maturity dimensions. In recent literature, dynamic factor models have been investigated extensively in order to find an appropriate way to capture these dynamics. Although the overall fit of these dynamic factor models for the IVS seems promising, the corner groups of the IVS generally show a poor fit in terms of heteroskedasticity and autocorrelation in the error terms. In this thesis, we examine whether we can improve this fit of dynamic factor models by integrating an additional volatility model onto their residuals. More specifically, we aim to find improvements in the predictability of the dynamics of implied volatility surfaces, by including GARCH(1,1) disturbances in one or both equations of dynamic factor models in state space form. Due to its ability to mitigate the problem of heteroskedasticity in the error terms, we consider this additional GARCH model on the residuals as a promising extension to dynamic factor models for the IVS. We implement an efficient estimation approach by using maximum likelihood estimation including a collapsed Kalman filtering approach. By analyzing statistical and economic evaluation methods, we compare performances of our extended DFM-GARCH setups in relation to general dynamic factor models. In addition, we explore the relevance of our extension by testing the dynamic factor models on significant GARCH effects using statistical tests for heteroskedasticity and autocorrelation.

We provide four key findings. First, although we still find significant heteroskedasticity and autocorrelation in the error terms, including GARCH disturbances into dynamic factor models can mitigate this problem of heteroskedasticity and autocorrelation in particularly the corner groups of the IVS. Second, the extended DFM-GARCH setups are able to outperform general dynamic factor models regarding the in-sample fit. In particular, including GARCH disturbances results in residuals that look a bit more on white noise processes, especially for corner groups of the IVS. However, in an out-of-sample setting all three DFM-GARCH variants are rejected as improved extensions of the general dynamic factor model in terms of statistical measures. This is possibly due to overfitting caused by the additional GARCH parameters in the DFM-GARCH setups. Hence, although we do not succeed in finding improvements in out-of-sample forecasts of the IVS, our improved in-sample estimation of the IVS indicates potential value of including GARCH disturbances into a general dynamic factor model. Third, all extended and general dynamic factor models prove to have economic value in their predictability of the IVS, before taking transaction costs into account. However, in a more realistic simulation including the impact of transaction costs, these potential profits disappear and transform into great losses. Besides, applying our constructed trading strategies within corner groups of the IVS causes extremely risky investment decisions, due to corresponding poorer fit and relatively illiquid options contracts. Again, the general DFM model economically performs best based on these out-of-sample forecasts of individual option contracts. Hence, our economic evaluation results confirm that our extended DFM-GARCH setups are less effective in economically exploiting their out-of-sample forecasts compared to the general DFM model. Fourth, performances of all three DFM-GARCH setups for the IVS are roughly similar and show fairly small differences compared to each other, both in-sample and out-of-sample. Hence, our findings suggest no significantly preferred way to incorporate GARCH disturbances into dynamic factor models for the IVS. However, since including GARCH disturbances in only the observation equation appears to cause marginally better results, we slightly recommend to consider this variant in further research. In conclusion, we can address our research question by confirming that integrating additional volatility models into the residuals of dynamic factor models can indeed improve the in-sample fit of the IVS. However, considering out-of-sample forecasting performances these extended models are outperformed significantly by the general dynamic factor model. Nonetheless, we find strong indications of potential

improvements of a general dynamic factor model for the IVS, after including GARCH disturbances to correct for heteroskedasticity in its error terms.

Finally, we suggest seven main directions for further research. First, it seems valuable to extend our work on improving dynamic factor models to obtain a better fit for the corner IVS groups. Considering promising improvements of the in-sample fit and corresponding out-of-sample limitations of our simplified GARCH extension setup, we recommend to further explore the potential of including additional volatility models into dynamic factor models for the IVS. For example, our GARCH extension to the residuals is naturally more efficient to correct for heteroskedasticity rather than for autocorrelation. Therefore, we suggest to consider other stochastic volatility models for the residuals like an ARMA model. Due to its modeling of the moving average, an ARMA model might be a powerful alternative to correct for autocorrelation in the residuals of dynamic factor models specifically. Besides, including GARCH disturbances to the restricted RFM setup can also be considered as interesting alternative in further research. Second, we recommend to explore alternative methods to fully correct for heteroskedasticity and autocorrelation in the error terms of dynamic factor models. In particular, although we initially impose linearity in volatility smile and term structure effects, in practice we are dealing with logistic patterns in both moneyness and maturity dimensions. In an alternative attempt to force the residuals to be white noise processes, one could therefore include additional factors to capture these logistic curve dynamics. Third, we examine the impact of our GARCH extensions on predictability in the dynamics of S&P 500 index options. In order to examine the potential of our extensions and similarities in the IVS dynamics of for instance equity options, implementation of our DFM-GARCH setups onto other types of options could also be explored in further research. As a fourth direction, we suggest to investigate the impact of GARCH disturbances into other types of dynamic factor models accordingly. For example, one could include GARCH disturbances into the spline-based model from Van der Wel et al. (2016), which combines the flexibility of a general dynamic factor model with the economically plausible factor interpretation of a restricted dynamic factor model. Fifth, we study basic trading strategies that highly suffer from the impact of transaction costs due to the large amount of selected trades. In addition to our basic trading strategy setups, we therefore suggest to explore more complex trading strategies based on our extended dynamic factor models. For example, one could attempt to reduce the impact of transaction costs by only selecting a limited amount of investments on days potentially high profits are expected. Sixth, we recommend to explore the impact of combining forecasts of our various dynamic factor models. In particular, we show that the relative forecasting performances of these models vary over time, indicating the potential benefit of combining IVS forecasts based on our various DFM(-GARCH) models. Seventh, we examine the dynamics of the IVS by considering a daily balanced panel of 24 groups differing across moneyness and maturity dimensions. As an alternative approach, we propose to investigate predictability in the dynamics of the entire cross-section of individual option contracts. Hence, we consider this alternative IVS construction and previously discussed directions as main suggestions for further research regarding predictability in the dynamics of the implied volatility surface.

Bibliography

- Barone-Adesi, G., Engle, R. F. and Mancini, L. (2008), ‘A garch option pricing model with filtered historical simulation’, *Review of Financial Studies* **21**(3), 1223–1258.
- Battalio, R., Hatch, B. and Jennings, R. (2004), ‘Toward a national market system for us exchange-listed equity options’, *The Journal of Finance* **59**(2), 933–962.
- Battiti, R. and Masulli, F. (1990), Bfgs optimization for faster and automated supervised learning, in ‘International neural network conference’, Springer, pp. 757–760.
- Bedendo, M. and Hodges, S. D. (2009), ‘The dynamics of the volatility skew: A kalman filter approach’, *Journal of Banking & Finance* **33**(6), 1156–1165.
- 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* pp. 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.
- Bollerslev, T. (1986), ‘Generalized autoregressive conditional heteroskedasticity’, *Journal of econometrics* **31**(3), 307–327.
- Box, G. E. and Pierce, D. A. (1970), ‘Distribution of residual autocorrelations in autoregressive-integrated moving average time series models’, *Journal of the American statistical Association* **65**(332), 1509–1526.
- Breusch, T. S. and Pagan, A. R. (1980), ‘The lagrange multiplier test and its applications to model specification in econometrics’, *The Review of Economic Studies* **47**(1), 239–253.
- Chalamandaris, G. and Tsekrekos, A. E. (2010), ‘Predictable dynamics in implied volatility surfaces from otc currency options’, *Journal of Banking & Finance* **34**(6), 1175–1188.
- Christoffersen, P., Fournier, M. and Jacobs, K. (2015), ‘The factor structure in equity options’, *Rotman School of Management Working Paper* (2224270).
- Christoffersen, P., Heston, S. and Jacobs, K. (2009), ‘The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well’, *Management Science* **55**(12), 1914–1932.
- Cont, R., Da Fonseca, J. et al. (2002), ‘Dynamics of implied volatility surfaces’, *Quantitative finance* **2**(1), 45–60.
- Cox, J. C., Ross, S. A. and Rubinstein, M. (1979), ‘Option pricing: A simplified approach’, *Journal of financial Economics* **7**(3), 229–263.
- Day, T. E. and Lewis, C. M. (1992), ‘Stock market volatility and the information content of stock index options’, *Journal of Econometrics* **52**(1-2), 267–287.
- Diebold, F. X. and Mariano, R. S. (2002), ‘Comparing predictive accuracy’, *Journal of Business & economic statistics* **20**(1), 134–144.

- 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. J. (2012), *Time series analysis by state space methods*, Vol. 38, OUP Oxford.
- Engle, R. F. (1982), ‘Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation’, *Econometrica: Journal of the Econometric Society* pp. 987–1007.
- Fengler, M. R., Härdle, W. K. and Mammen, E. (2007), ‘A semiparametric factor model for implied volatility surface dynamics’, *Journal of Financial Econometrics* **5**(2), 189–218.
- Fleming, J. (1998), ‘The quality of market volatility forecasts implied by s&p 100 index option prices’, *Journal of empirical finance* **5**(4), 317–345.
- French (2017), Kenneth R. French - Data Library - Tuck - http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Accessed: 2017-05-12.
- Geweke, J. and Zhou, G. (1996), ‘Measuring the pricing error of the arbitrage pricing theory’, *The review of financial studies* **9**(2), 557–587.
- 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.
- Harvey, A., Ruiz, E. and Sentana, E. (1992), ‘Unobserved component time series models with arch disturbances’, *Journal of Econometrics* **52**(1-2), 129–157.
- Harvey, C. R. and Whaley, R. E. (1992), ‘Market volatility prediction and the efficiency of the s & p 100 index option market’, *Journal of Financial Economics* **31**(1), 43–73.
- Heston, S. L. and Nandi, S. (2000), ‘A closed-form garch option valuation model’, *Review of Financial Studies* **13**(3), 585–625.
- Jungbacker, B. and Koopman, S. J. (2014), ‘Likelihood-based dynamic factor analysis for measurement and forecasting’, *The Econometrics Journal* .
- Konstantinidi, E., Skiadopoulos, G. and Tzagkaraki, E. (2008), ‘Can the evolution of implied volatility be forecasted? evidence from european and us implied volatility indices’, *Journal of Banking & Finance* **32**(11), 2401–2411.
- Koopman, S. J. and Durbin, J. (2000), ‘Fast filtering and smoothing for multivariate state space models’, *Journal of Time Series Analysis* **21**(3), 281–296.
- Koopman, S. J., Mallee, M. I. and Van der Wel, M. (2010), ‘Analyzing the term structure of interest rates using the dynamic nelson–siegel model with time-varying parameters’, *Journal of Business & Economic Statistics* **28**(3), 329–343.
- Ljung, G. M. and Box, G. E. (1978), ‘On a measure of lack of fit in time series models’, *Biometrika* **65**(2), 297–303.
- McLeod, A. I. and Li, W. K. (1983), ‘Diagnostic checking arma time series models using squared-residual autocorrelations’, *Journal of Time Series Analysis* **4**(4), 269–273.
- Newey, W. K. and West, K. D. (1986), ‘A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix’.

- Noh, J., Engle, R. F. and Kane, A. (1994), 'Forecasting volatility and option prices of the s&p 500 index', *The Journal of Derivatives* **2**(1), 17–30.
- Poterba, J. M. and Summers, L. H. (1984), 'The persistence of volatility and stock market fluctuations'.
- Rubinstein, M. (1994), 'Implied binomial trees', *The Journal of Finance* **49**(3), 771–818.
- Skiadopoulos, G., Hodges, S. and Clewlow, L. (2000), 'The dynamics of the s&p 500 implied volatility surface', *Review of derivatives research* **3**(3), 263–282.
- Van der Wel, M., Ozturk, S. R. and van Dijk, D. (2016), 'Dynamic factor models for the volatility surface', *Dynamic Factor Models (Advances in Econometrics, Volume 35)* Emerald Group Publishing Limited **35**, 127–174.

Appendix

Table 11: **Principal Component Analysis - Explained Variation and Persistence**

	Explained Variation		ACF			PCF		
	Percentage	Cum. Perc.	1	5	10	1	2	3
PC 1	95.98%	95.98%	0.978	0.923	0.884	0.978	0.112	0.069
PC 2	2.13%	98.11%	0.955	0.864	0.804	0.955	0.107	0.120
PC 3	1.04%	99.15%	0.859	0.653	0.566	0.858	0.144	0.148
PC 4	0.26%	99.40%	0.790	0.523	0.335	0.789	0.149	0.101
PC 5	0.17%	99.58%	0.706	0.522	0.459	0.706	0.304	0.146

Notes: This table provides various results from principal component analysis on the initial S&P 500 index options data. First, the table reports both the variation explained by each individual principal component as well as the cumulative percentages for the first five principal components. Second, the table provides (partial) autocorrelations of those principal components in order to determine its persistence. Specifically, the table shows autocorrelations for lags 1,5 and 10, whereas partial autocorrelations are reported for lags 1, 2 and 3.

Table 13: Estimated Loading Matrix Λ of General DFM Model

		Λ from General DFM		
		Factor 1	Factor 2	Factor 3
DOTM	10-45	1.19	-0.51	6.48
	45-90	1.18	0.34	7.79
	Put 90-180	1.14	1.21	9.51
	180-360	1.08	2.17	10.76
OTM	10-45	0.99	-1.01	4.04
	45-90	1.04	-0.25	3.42
	Put 90-180	1.04	0.51	3.82
	180-360	1.02	1.31	4.43
ATM	10-45	1	-1	1
	45-90	1.02	-0.33	0.33
	Put 90-180	1.02	0.32	0.57
	180-360	1	1	1
ATM	10-45	1	-1	-1
	45-90	1.04	-0.28	-2.15
	Call 90-180	1.02	0.36	-1.62
	180-360	1	1	-1
OTM	10-45	1.06	-0.94	-4.13
	45-90	1.05	-0.28	-4.58
	Call 90-180	1.02	0.27	-3.89
	180-360	0.99	0.82	-3.16
DOTM	10-45	1.15	-0.78	-7.85
	45-90	1.09	-0.22	-7.70
	Call 90-180	1.04	0.24	-6.79
	180-360	1.00	0.69	-5.84

Notes: This table provides estimated values of the loading matrix Λ based on the general DFM model. Elements of the loading matrix that are fixed for identification purposes, are highlighted in bold.

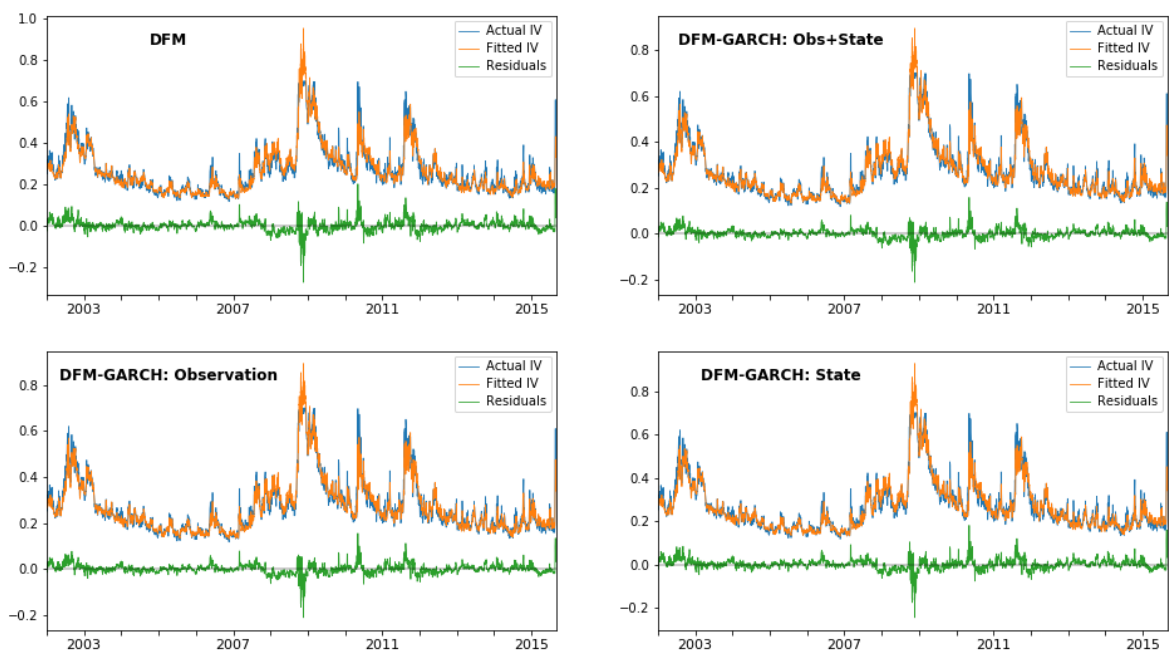


Figure 16: Comparing Fit of Corner IVS Group (DOTM-Put, 10-45 days)

Notes: These figures compare the fit of a specific corner group of the IVS (DOTM-Put, 10-45 days) estimated with four of our main dynamic factor models. In particular, we provide the fit of the IVS estimated with our three DFM-GARCH extensions relative to the fit of the IVS based on the general DFM model. The figures present time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals.

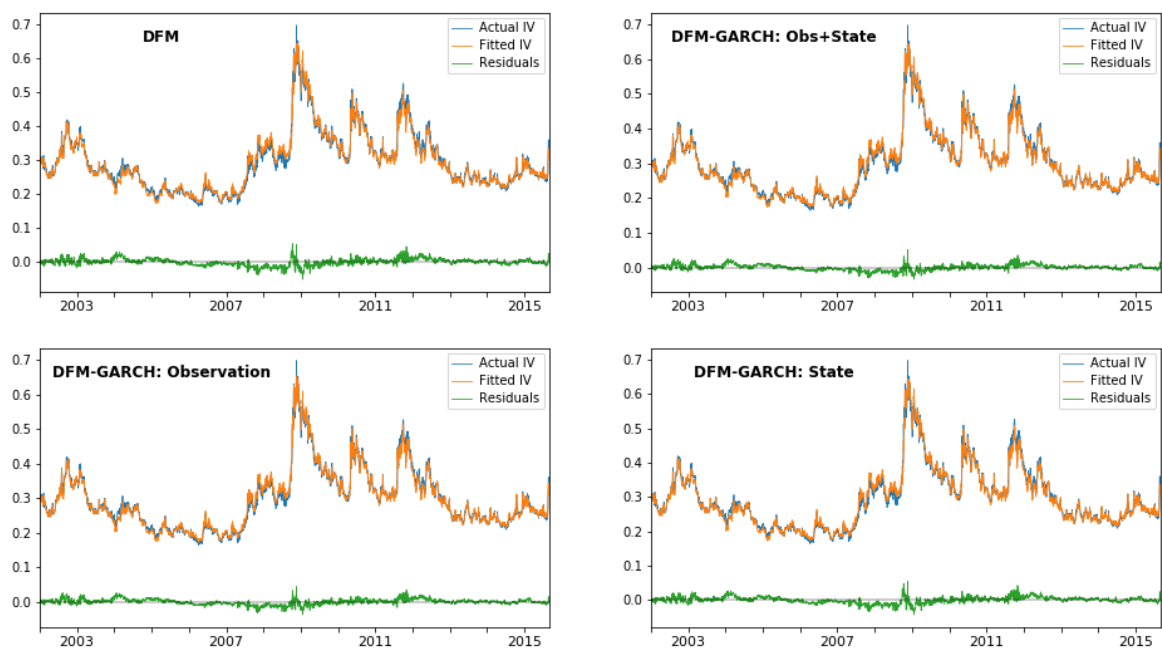


Figure 17: **Comparing Fit of Corner IVS Group (DOTM-Put, 180-360 days)**

Notes: These figures compare the fit of a specific corner group of the IVS (DOTM-Put, 180-360 days) estimated with four of our main dynamic factor models. In particular, we provide the fit of the IVS estimated with our three DFM-GARCH extensions relative to the fit of the IVS based on the general DFM model. The figures present time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals.

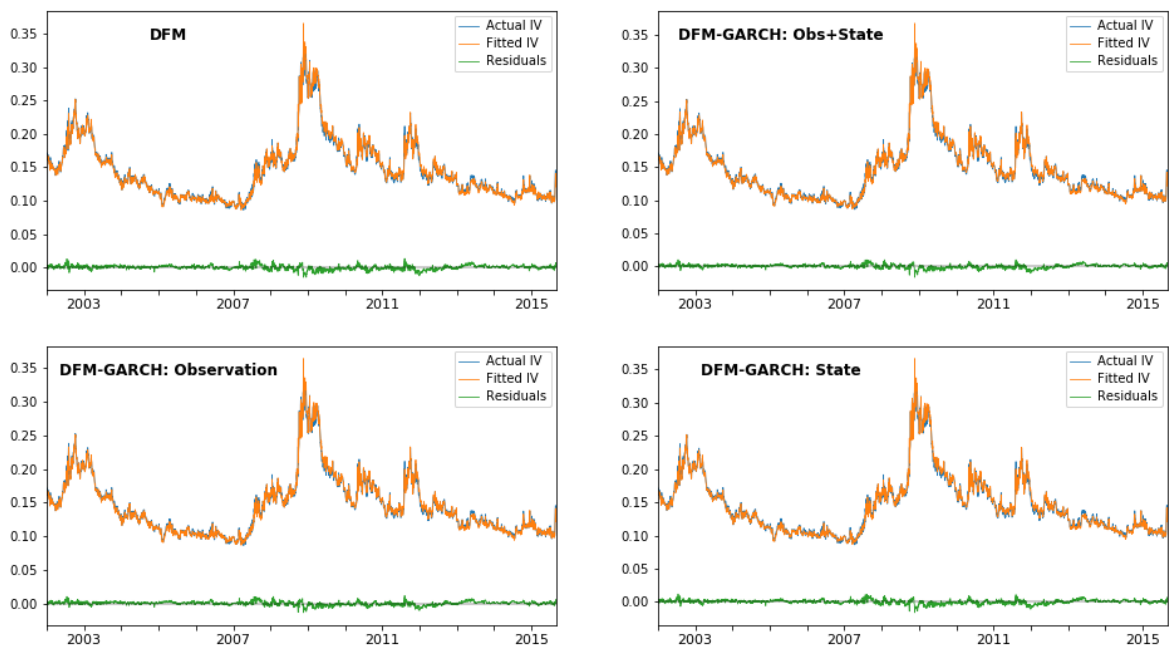


Figure 18: **Comparing Fit of Corner IVS Group (DOTM-Call, 180-360 days)**

Notes: These figures compare the fit of a specific corner group of the IVS (DOTM-Call, 180-360 days) estimated with four of our main dynamic factor models. In particular, we provide the fit of the IVS estimated with our three DFM-GARCH extensions relative to the fit of the IVS based on the general DFM model. The figures present time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals.

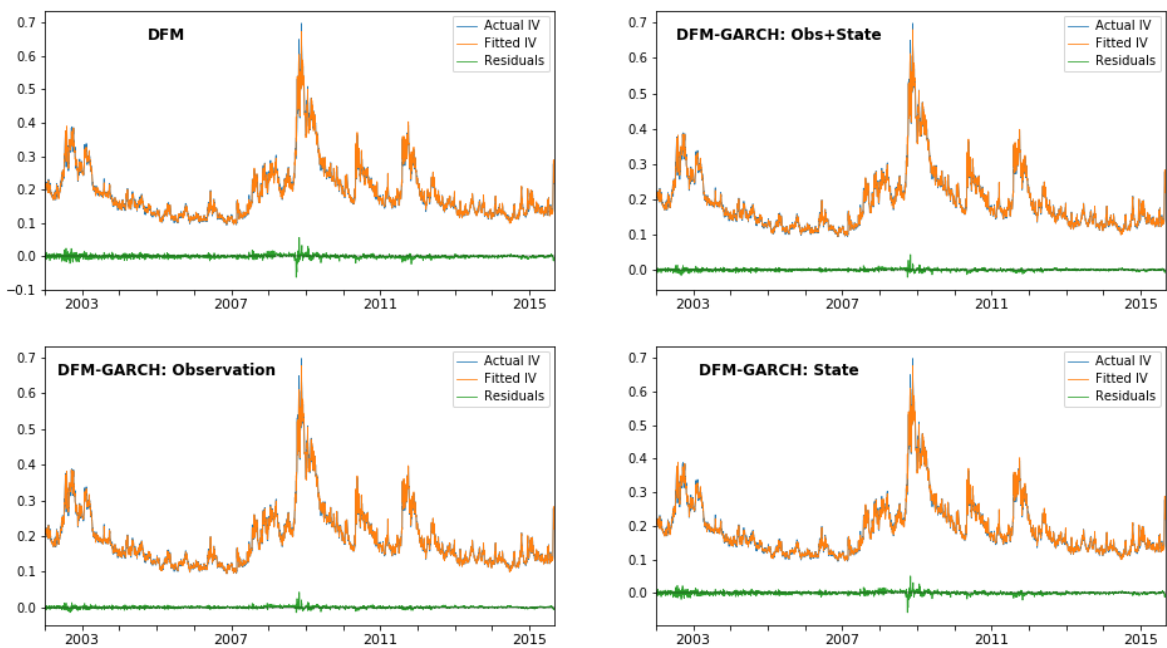


Figure 19: Comparing Fit of Middle IVS Group (ATM-Put, 45-90 days)

Notes: These figures compare the fit of an arbitrary middle group of the IVS (ATM-Put, 45-90 days) estimated with four of our main dynamic factor models. In particular, we provide the fit of the IVS estimated with our three DFM-GARCH extensions relative to the fit of the IVS based on the general DFM model. The figures present time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals.

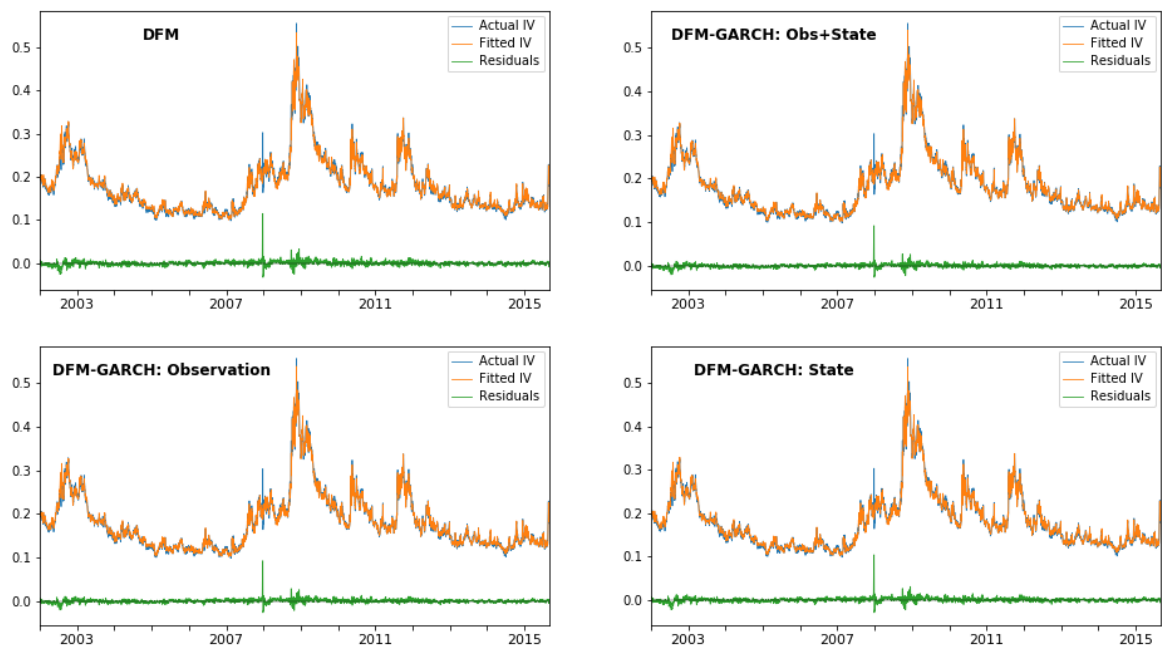


Figure 20: Comparing Fit of Middle IVS Group (ATM-Call, 90-180 days)

Notes: These figures compare the fit of an arbitrary middle group of the IVS (ATM-Call, 90-180 days) estimated with four of our main dynamic factor models. In particular, we provide the fit of the IVS estimated with our three DFM-GARCH extensions relative to the fit of the IVS based on the general DFM model. The figures present time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals.

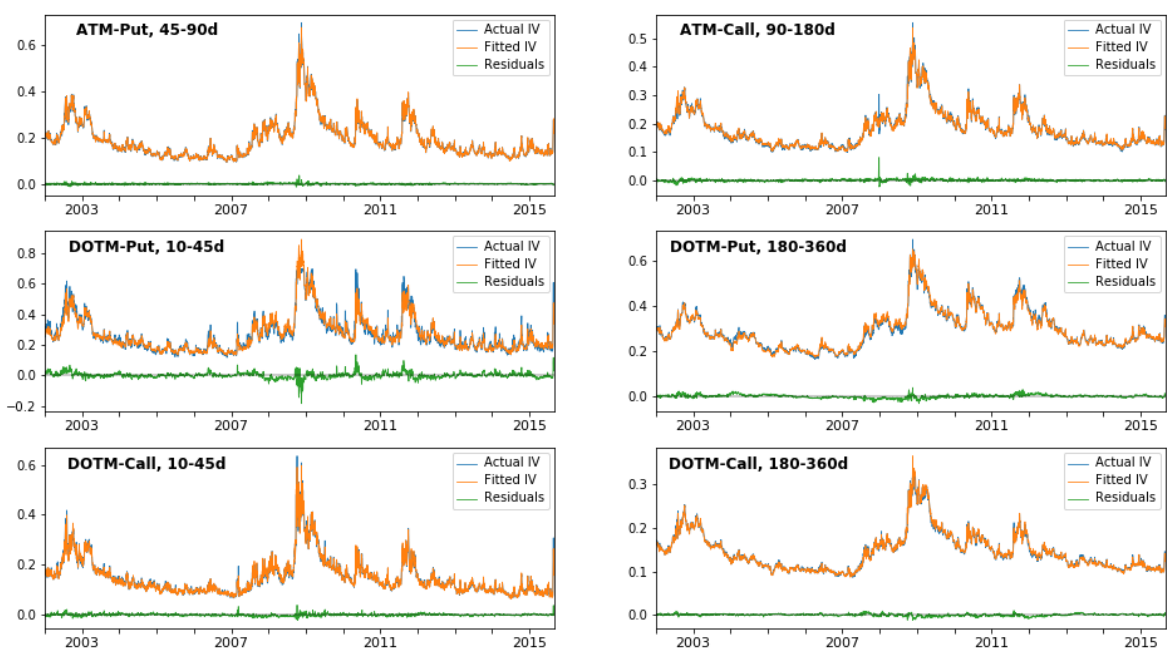


Figure 21: **Fit of DFM-GARCH (Observation) Model**

Notes: These figures show the fit of the IVS estimated with the extended DFM-GARCH model with GARCH disturbances incorporated in the observation equation only. In particular, we display time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals. In total, we document six different IVS groups. The upper two figures present two groups in the center of the IVS, whereas in the bottom four figures plots of the corner groups of the IVS are provided.

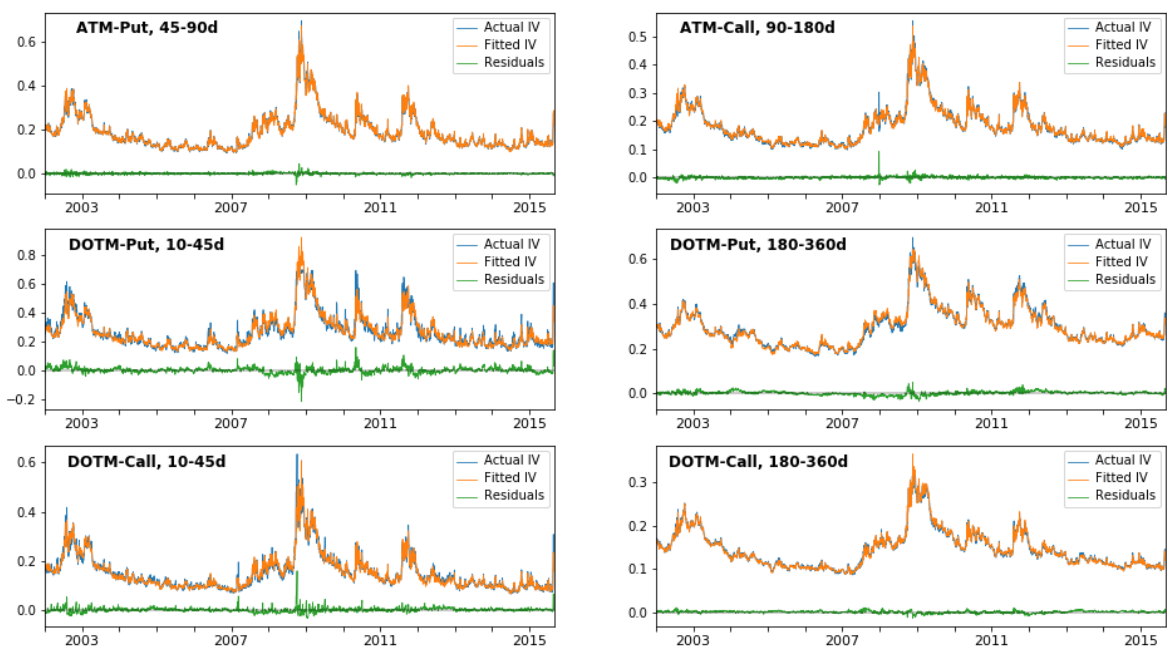


Figure 22: **Fit of DFM-GARCH (State) Model**

Notes: These figures show the fit of the IVS estimated with the extended DFM-GARCH model with GARCH disturbances incorporated in the state equation only. In particular, we display time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals. In total, we document six different IVS groups. The upper two figures present two groups in the center of the IVS, whereas in the bottom four figures plots of the corner groups of the IVS are provided.

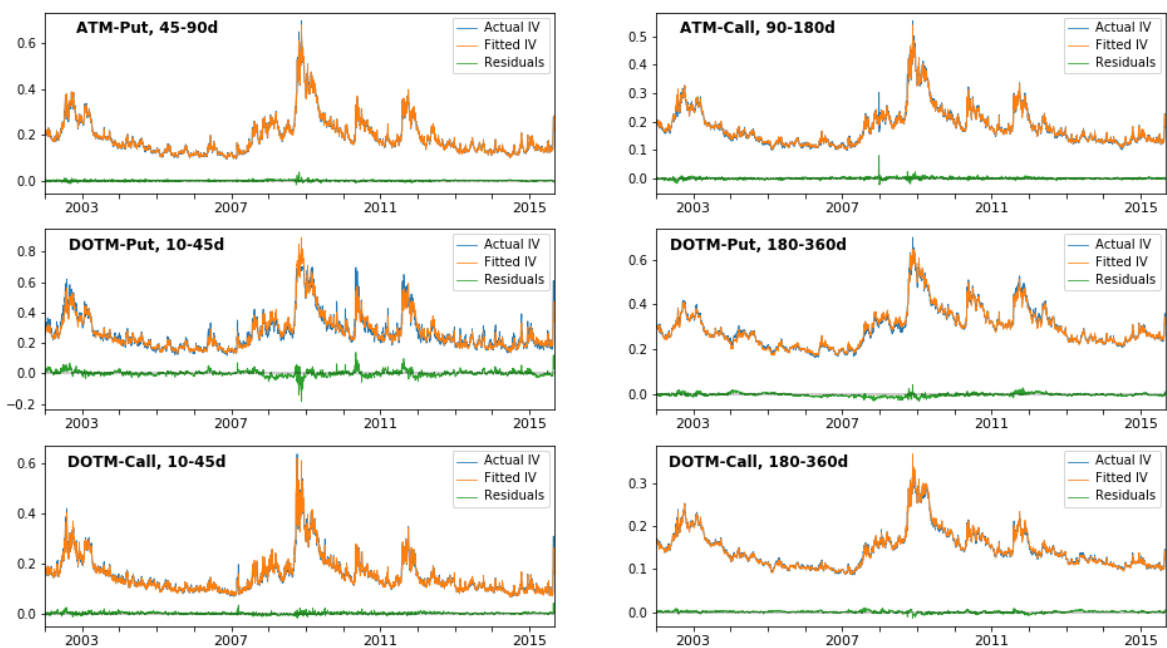


Figure 23: **Fit of DFM-GARCH (Observation+State) Model**

Notes: These figures show the fit of the IVS estimated with the extended DFM-GARCH model with GARCH disturbances incorporated in both the observation and state equation. In particular, we display time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals. In total, we document six different IVS groups. The upper two figures present two groups in the center of the IVS, whereas in the bottom four figures plots of the corner groups of the IVS are provided.

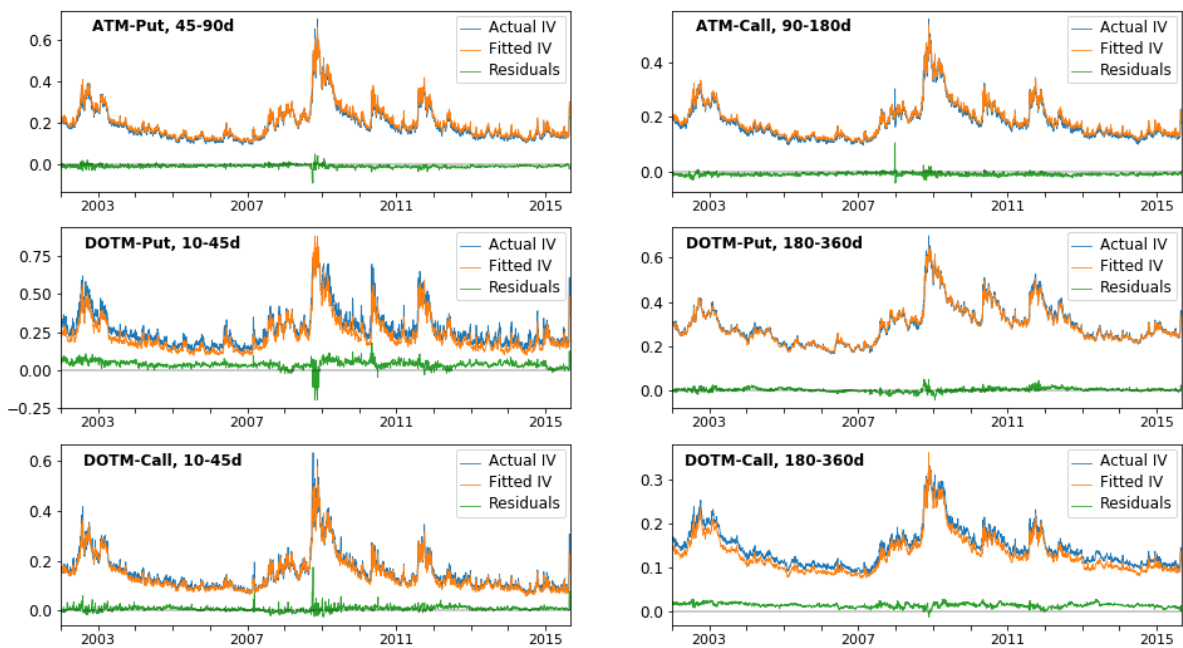


Figure 24: **Fit of Restricted Economic Dynamic Factor Model (RFM)**

Notes: These figures show the fit of the IVS estimated with the restricted economic dynamic factor model (RFM). In particular, we display time series of the actual implied volatilities, the fitted implied volatilities and corresponding residuals. In total, we document six different IVS groups. The upper two figures present two groups in the center of the IVS, whereas in the bottom four figures plots of the corner groups of the IVS are provided.

Table 14: Testing Residuals on Autocorrelation with Ljung-Box Statistics

		DFM-GARCH						Benchmarks					
		Observation		State		Obs+State		DFM		RFM			
		Q(6)	Q ² (6)	Q(6)	Q ² (6)	Q(6)	Q ² (6)	Q(6)	Q ² (6)	Q(6)	Q ² (6)		
DOTM- Put	10-45	8,610	3,963	8,736	4,200	8,759	4,021	8,616	4,029	9,443	5,433		
	45-90	11,147	5,063	10,886	4,498	11,175	5,219	10,794	4,346	10,250	4,612		
	90-180	10,271	4,737	10,064	5,253	10,308	5,026	10,005	4,840	7,957	4,437		
	180-360	11,696	5,716	11,475	5,589	11,568	5,156	11,588	6,015	9,184	4,604		
OTM- Put	10-45	2,744	970	4,253	1,260	2,867	910	4,290	1,264	1,231	787		
	45-90	1,940	1,238	1,956	2,182	1,531	1,081	1,951	2,022	4,569	4,151		
	90-180	2,812	3,941	4,564	4,854	2,969	4,064	4,554	4,697	6,424	4,612		
	180-360	2,144	1,904	1,944	1,843	2,122	1,614	1,983	2,062	3,021	2,081		
ATM- Put	10-45	1,025	467	2,597	1,666	960	414	2,689	1,651	700	891		
	45-90	1,688	827	1,825	1,404	1,368	710	1,786	1,425	4,232	3,874		
	90-180	4,551	3,295	5,917	4,662	4,714	3,741	5,894	4,730	7,050	4,309		
	180-360	5,464	5,603	4,799	4,265	5,407	5,404	4,823	4,302	7,193	6,921		
ATM- Call	10-45	314	<u>4</u>	3,830	2,016	250	10	3,753	2,044	2,752	1,542		
	45-90	1,558	1,630	1,405	876	1,576	1,222	1,401	735	2,449	3,005		
	90-180	1,721	1,065	1,956	1,064	1,759	1,066	1,965	1,064	1,662	1,279		
	180-360	5,074	1,839	6,610	4,770	5,084	1,800	6,604	4,718	8,872	4,862		
OTM- Call	10-45	528	39	4,420	2,462	461	37	4,467	2,456	3,410	1,299		
	45-90	868	2,203	897	2,440	867	1,966	918	2,151	1,509	3,949		
	90-180	3,027	2,985	4,114	4,026	3,142	2,968	4,267	4,226	3,926	2,183		
	180-360	3,721	3,106	4,474	4,286	3,708	3,180	4,501	4,146	7,091	5,253		
DOTM- Call	10-45	3,509	590	4,210	1,671	3,402	515	4,228	1,688	3,998	726		
	45-90	6,754	2,148	8,003	3,019	6,732	2,591	8,233	2,584	8,858	4,267		
	90-180	5,308	1,906	7,145	5,118	5,194	2,363	7,484	5,082	10,804	7,631		
	180-360	10,011	6,045	9,712	5,664	9,848	5,560	9,908	6,346	12,919	12,416		

Notes: This table provides statistics regarding the presence of autocorrelation in the error terms of our main dynamic factor models. In particular, we provide significance tests of autocorrelation in the residuals ($Q(6)$) and squared residuals ($Q^2(6)$) of all individual IVS groups. For this purpose, we perform Ljung-Box tests as introduced in section 4.1. Nearly all reported statistics are significant. Insignificant results at a 5% significance level are highlighted in bold and underlined.

Table 15: Testing Residuals on Heteroskedasticity with Breusch-Pagan Statistics

		DFM-GARCH						Benchmarks			
		Observation		State		Obs+State		DFM		RFM	
		LM(ϵ)	LM(ϵ^2)	LM(ϵ)	LM(ϵ^2)	LM(ϵ)	LM(ϵ^2)	LM(ϵ)	LM(ϵ^2)	LM(ϵ)	LM(ϵ^2)
DOTM- Put	10-45	1,059	294	986	267	1,044	291	983	257	2,026	478
	45-90	938	223	908	191	946	230	882	184	1,184	257
	90-180	1,056	250	1,001	251	1,088	295	985	211	930	225
	180-360	1,123	350	1,143	384	1,122	264	1,147	402	1,117	408
OTM- Put	10-45	503	160	492	133	509	153	490	132	893	168
	45-90	588	96	706	155	558	93	695	143	806	134
	90-180	949	329	992	339	927	275	969	303	1,491	286
	180-360	785	144	873	224	722	126	873	241	951	280
ATM- Put	10-45	412	29	469	76	381	33	503	78	581	58
	45-90	423	92	497	78	430	84	509	87	897	144
	90-180	913	341	887	258	920	308	889	264	1,129	194
	180-360	999	305	947	211	990	285	930	199	1,476	436
ATM- Call	10-45	43	20	234	52	38	15	219	52	287	61
	45-90	764	108	708	79	701	75	656	63	1,648	445
	90-180	53	3	53	3	51	3	53	3	280	4
	180-360	1,003	135	857	192	999	127	864	194	1,205	399
OTM- Call	10-45	70	20	220	84	91	16	221	85	250	96
	45-90	702	215	777	246	658	170	729	222	1,069	327
	90-180	1,094	448	1,134	465	1,093	463	1,134	449	1,056	221
	180-360	819	199	919	321	835	211	902	309	2,185	709
DOTM- Call	10-45	517	47	265	68	521	22	270	69	317	76
	45-90	790	139	793	173	840	185	736	132	889	243
	90-180	1,065	101	1,039	342	1,156	134	1,042	308	1,391	369
	180-360	1,172	430	1,174	392	1,137	320	1,217	511	2,758	1,931

Notes: This table provides statistics regarding the presence of heteroskedasticity in the error terms of our main dynamic factor models. In particular, we provide significance tests of heteroskedasticity in the residuals (LM(ϵ)) and squared residuals (LM(ϵ^2)) of all individual IVS groups. For this purpose, we perform Breusch-Pagan Lagrange multiplier tests as introduced in section 4.1. Nearly all reported statistics are significant. Insignificant results at a 5% significance level are highlighted in bold and underlined.