Modeling and Forecasting the VIX Index

Katja Ahoniemi*
Helsinki School of Economics and HECER

July 17, 2008

Abstract

This paper estimates numerous time-series models of the VIX index, and finds that an ARIMA(1,1,1) model has predictive power regarding the directional change in the VIX. A GARCH(1,1) specification improves both directional and point forecasts, but augmenting the model with financial or macroeconomic explanatory variables such as S&P 500 returns does not produce further improvement. The direction of change in the VIX is predicted correctly on over 58 percent of trading days in an out-of-sample period of five years. An out-of-sample straddle trading simulation with S&P 500 index options lends further support to the forecasting model choice.

*Correspondence: Helsinki School of Economics, Department of Economics, P.O. Box 1210, 00101 Helsinki, Finland. E-mail: katja.ahoniemi@hse.fi. The author wishes to thank Markku Lanne, Pekka Ilmakunnas, Petri Jylhä, and participants at the European Financial Management Association 2006 annual meeting and at the 2006 European Meeting of the Econometric Society for helpful comments. Petri Jylhä also provided valuable assistance with the data. Financial support from the Finnish Doctoral Programme in Economics, the Jenny and Antti Wihuri Foundation, the Okobank Group Research Foundation, the Finnish Foundation for Share Promotion, and the Finnish Foundation for Advancement of Securities Markets is gratefully acknowledged.

1 Introduction

Professional option traders such as hedge funds and banks' proprietary traders are interested primarily in the volatility implied by an option's market price when making buy and sell decisions. If the implied volatility (IV) is assessed to be too high, the option is considered to be overpriced, and vice versa. Returns from volatility positions in options, such as straddles, depend largely on the movements in IV, and the trader does not need a directional view regarding the price of the option's underlying asset. As a measure of market risk, IV can also be seen as a useful tool in all asset pricing, and its value can assist in making portfolio management decisions. Due to these considerations, implied volatility forecasting can provide added value to practitioners and retail investors alike. Though traditionally solved by backing out volatility from an option pricing formula such as Black-Scholes, model-free implied volatility indices have emerged over the past years. The most popular and widely followed index is the VIX index, calculated by the CBOE from S&P 500 index option prices. Simon (2003) describes how markets tend to view extreme values of the VIX as trading signals. If the VIX is very high, markets are pessimistic, which could lead to a subsequent rally in stock prices. On the other hand, if IV is very low, the market may next face a disappointment stemming from a downward move in prices.

Due to its popularity, the VIX has been the topic of numerous studies.[1] The traditional vein of IV research investigates how well IV succeeds in forecasting future realized volatility: after all, IV should equal the market's expected future volatility. Existing research is in relatively unanimous agreement that the VIX index forecasts future volatility better than any historical volatility measure. Fleming et al. (1995) find that the VIX dominates historical volatility as a forecaster of future volatility. Blair et al. (2001) agree, concluding that volatility forecasts provided by the VIX index are more accurate than forecasts based on intra-day returns or GARCH models. Blair et al. (2001) also maintain that there is some incremental information in daily index returns when forecasting one day ahead, but the VIX index provides nearly all information relevant for forecasting. Corrado and Miller (2005) report that the VIX yields forecasts that are biased upward, but the forecasts are more efficient in terms of mean squared error than forecasts based on historical volatility. Dennis et al. (2006) find that daily VIX changes are very significant in predicting future index return volatility. Carr and Wu (2006) conclude that a GARCH-based variance does not provide incremental information over that contained in the VIX index. Giot and Laurent (2007) also conclude that the information content of the VIX is high: the explanatory power of a model for realized volatility with only lagged values of the VIX is nearly as good as that of a model with jump and continuous components of historical volatility included. Becker et al. (2006) provide evidence that contradicts the findings of the above-mentioned papers. They investigate the forecasting ability of the VIX and find that it is not an efficient forecaster of future S&P 500 realized volatility, particularly when data is sampled daily rather than at longer intervals. In other words, according to Becker et al. (2006), other historical volatility estimates can improve volatility forecasts based on the VIX alone.

Relatively little work has been done on modeling IV itself, compared with the ex-

[1] Note that VIX is used here to refer to both the current VIX and VXO indices. The VIX used to be based on S&P 100 options, but is based on S&P 500 options since September 2003. The index that is today calculated from S&P 100 option prices carries the ticker VXO. Research conducted before the switch speaks of the VIX, when today that index is the VXO.

tensive literature on modeling the volatility of the returns of various financial assets that exists today. Bollen and Whaley (2004) model changes in S&P 500 index implied volatility with variables such as returns and trading volume in the underlying, as well as net buying pressure variables. Mixon (2002) finds that contemporaneous domestic stock returns have significant explanatory power for changes in IV, but other observable variables, such as foreign stock returns and interest rate variables, can also be useful in IV modeling. Low (2004) uses the VIX as a proxy for option traders' risk perception, and investigates how changes in risk perceptions and changes in prices are linked. Gonçalves and Guidolin (2006) model the whole surface of S&P 500 option IV, thus incorporating the term structure of IV and option moneyness into the model.[2]

There are a few papers that are very close in nature to this study: Harvey and Whaley (1992) and Brooks and Oozeer (2002) are based on very much the same approach. Harvey and Whaley (1992) forecast the implied volatility of options on the S&P 100 index and find that changes in IV are indeed predictable. They forecast the direction of change in IV and trade accordingly in the option market. Brooks and Oozeer (2002) model the IV of options on Long Gilt futures, traded on the London International Financial Futures Exchange, calculate directional forecasts, and run an option trading simulation. Both studies suggest profits for a market maker, but not for a trader facing transaction costs. Guo (2000) is also a similar paper, containing a model for the changes in call and put IV of foreign exchange options. This model is used to forecast IV, with option trades executed based on the forecasts. The goal of this paper is somewhat different, as it seeks to compare whether IV or GARCH forecasts lead to larger option trading returns. Another closely related paper is Konstantinidi et al. (2008), who estimate a variety of models for several IV indices, including the VIX. However, the directional forecast accuracy they receive for the VIX is outperformed by the results in this study. The models in this study are augmented with both GARCH and day-of-the-week effects, elements which are absent from Konstantinidi et al. (2008), and may contribute to the superior forecast performance of this study.

There are a number of studies that explore option trading in connection with IV analysis. Harvey and Whaley (1992), Brooks and Oozeer (2002) and Corredor et al. (2002) employ simple buy or sell option trading strategies, whereas typical volatility trades involve various types of spreads, most commonly the straddle. Guo (2000) and Noh et al. (1994) simulate straddle trades, and in the trading simulation of Poon and Pope (2000), S&P 100 call options are bought and S&P 500 call options simultaneously sold, or vice versa. Kim and Kim (2003) trade exchange rate options on futures to take advantage of observed intraweek and announcement day effects on IV. Harvey and Whaley (1992) and Guo (2000) first convert their IV forecasts into option prices, and compare these predicted option prices to actual market quotes when choosing how to trade. Gonçalves and Guidolin (2006) solve for IV from option prices, compare that value to their IV forecasts, and trade accordingly. Both these approaches require an assumption on the option pricing model. In this study, IV forecasts are used directly as trading signals in the same manner as in Brooks and Oozeer (2002), so that if IV is forecast to rise (fall), options are bought (sold), as their value is expected to increase

[2]There are other examples of IV modeling from markets other than the stock market. Using data on currency futures options, Kim and Kim (2003) model changes in implied volatility with lagged IV changes and futures returns. Davidson et al. (2001) use data from various types of options on non-equity futures and explain changes in IV with lagged IV changes, futures returns, trading volume, and open interest.

(decrease).

This study estimates various traditional time series models for the VIX. An ARIMA(1,1,1) model augmented with GARCH errors is found to be a good fit for the time series of the VIX index.[3] The data clearly displays conditional heteroskedasticity, and GARCH modeling proves important. S&P 500 index returns have some explanatory power over VIX changes, but the trading volume of the underlying index does not. The ARIMA-GARCH model specification produces forecasts with a directional accuracy of up to 58.4 percent in a five-year out-of-sample period. An accuracy of above 50% can potentially lead to profits in option positions. A straddle trading simulation is run based on the forecasts from the estimated models. The goal of the simulation, however, is not to determine whether or not abnormal profits can be achieved. The option trading returns are used instead to confirm the choice of the best forecasting model. Option trades simulated on the basis of these forecasts indicate that adding the lagged returns of the S&P 500 index to the model can perhaps be beneficial. Both the forecast evaluation and straddle trading exercise show that it is better to estimate the forecasting models with a shorter in-sample period of 1,000 rather than a longer period of 3,279 observations. Conditions in the market have changed sufficiently during the years in question that more recent events are more relevant for parameter estimation.

This paper proceeds as follows. Section 2 describes the data used in this study, including various financial and macroeconomic variables that could have explanatory power over the VIX index. Section 3 presents the models estimated for the VIX time series and discusses their goodness of fit. Section 4 contains the analysis of forecasts, with forecast evaluation based on directional accuracy and mean squared errors. Section 5 contains the option trading simulation. Section 6 concludes the paper.

2 Data

2.1 The VIX index

The core data in this study consists of daily observations of the VIX Volatility Index calculated by the Chicago Board Options Exchange.[4] The VIX, introduced in 1993, is derived from the bid/ask quotes of options on the S&P 500 index. It is widely followed by financial market participants and is considered not only to be the market's expectation of the volatility in the S&P 500 index over the next month, but also to reflect investor sentiment and risk aversion. If investors grow more wary, the demand for above all put options will rise, thus increasing IV and the value of the VIX. The VIX is used as an indicator of market implied volatility in studies such as Blair et al. (2001) and Mayhew and Stivers (2003). The use of an implied volatility index such as the VIX considerably alleviates the problems of measurement errors and model misspecification. The simultaneous measurement of all variables required by an option pricing model is often difficult to achieve. When the underlying asset of an option is a stock index, infrequent trading in one of the component stocks of the index can lead to misvaluation of the index level. Also, there is no correct measure for the volatility required as an input in a pricing model such as the traditional Black-Scholes model.

[3]Fernandes et al. (2007) explore the long-memory properties of the VIX index, but find that when forecasting one day ahead, simple linear models perform as well as more sophisticated models.

[4]see www.cboe.com/micro/vix

The calculation method of the VIX was changed on September 22, 2003 to bring it closer to actual financial industry practices. From that day onwards, the VIX has been based on S&P 500 rather than S&P 100 options: the S&P 500 index is the most commonly used benchmark for the U.S. equity market, and the most popular underlying for U.S. equity derivatives (Jiang and Tian (2007)). A wider range of strike prices is included in the calculation, making the new VIX more robust. Also, the Black-Scholes formula is no longer used, but the methodology is independent of a pricing model. In practice, the VIX is calculated directly from option prices rather than solving implied volatility out of an option pricing formula.[5] Values for the VIX with the new methodology are available from the CBOE from 1.1.1990. The data set used in this study consists of daily observations covering eighteen years, from 1.1.1990 to 31.12.2007. Public holidays that fall on weekdays were omitted from the data set, resulting in a full sample of 4,537 observations.

2.2 Other data

Data on various financial and macroeconomic indicators was also obtained from the Bloomberg Professional service to test whether they could help explain the variations in the VIX. The data set contains the S&P 500 index, the trading volume of the S&P 500 index, the MSCI EAFE (Europe, Australasia, Far East) stock index, the three-month U.S. dollar LIBOR interest rate, the 10-year U.S. government bond yield, and the price of crude oil from the next expiring futures contract. Data on the S&P 500 trading volume is available only from 4.1.1993 onwards.

The variables outlined above were used to construct a number of explanatory variables for the VIX time series models. These variables are the returns on the S&P 500 index, the rolling one-month, or 22-trading-day, historical volatility of the S&P 500 index and its first difference, the spread between the VIX and the historical volatility of the S&P 500 index and its first difference, the trading volume of the S&P 500 index and its first difference, the returns of the MSCI EAFE (Europe, Australasia, Far East) stock index, the first difference of the three-month USD LIBOR interest rate, the first difference of the slope of the yield curve proxied by the 10-year rate less the 3-month rate, and the first difference of the price of oil from the next expiring futures contract.

Many variables similar to these have been used by e.g. Simon (2003), Harvey and Whaley (1992), and Franks and Schwartz (1991). The returns of the S&P 500 index, its trading volume, and returns in the MSCI EAFE index in particular could be assumed to have some effect on changes in the VIX. It is well documented that returns and volatility are linked, with large returns and high volatility going hand-in-hand. Carr and Wu (2006) provide correlation evidence that S&P 500 index returns predict changes in the VIX index. Fleming et al. (1995) note that spikes in the value of the VIX tend to coincide with large moves in the underlying stock index level. Although the U.S. market tends to dominate in global financial trends and events, it is also plausible that the returns of stock markets in other countries could have some spillover effects on the U.S. market. In fact, Simon (2003) finds that Nikkei 225 index returns are useful in a model for the term structure of IV. High trading volume could be a signal of e.g. panic selling, linked to rising IV. Also, the arrival of both positive and negative news, which introduces shocks to returns and volatility, can be reflected in higher trading volume.

[5] In investigations comparing implied to realized volatility, the use of model-free implied volatility allows the researcher to avoid a joint test of the option pricing model and market efficiency.

Interest rates are likely to affect stock markets, especially when they change rapidly. As day-to-day changes are quite small, the interest rate variables are not likely to be significant explanatory variables.[6] Oil prices are included in the data set to identify the significance of another important, non-equity market.

Logarithms were used for all variables except the slope of the yield curve. Use of the S&P 500 trading volume, the short-term interest rate, and the slope of the yield curve without differencing was ruled out by the Augmented Dickey-Fuller (ADF) test: the p-values from the ADF test indicate that the null hypothesis of a unit root cannot be rejected at the one-percent level for the above-mentioned time series. Also, returns must naturally be used for the S&P 500 and MSCI EAFE indices due to the same non-stationarity issues.

3 Modeling the VIX

The VIX index was relatively stable in the early 1990s, but more volatile from the last quarter of 1997 to the first quarter of 2003. The year 2007 was also characterized by large variability in the value of the VIX. Figure 1 shows the daily level of the VIX for the entire sample. Clear spikes in the value of the VIX coincide with the Iraqi invasion of Kuwait in late 1990, the allied attack on Iraq in early 1991, the Asian financial crisis of late 1997, the Russian and LTCM crisis of late summer 1998, and the 9/11 terrorist attacks. A visual inspection of the VIX first differences points to heteroskedasticity in the data, as can be seen in Figure 2.

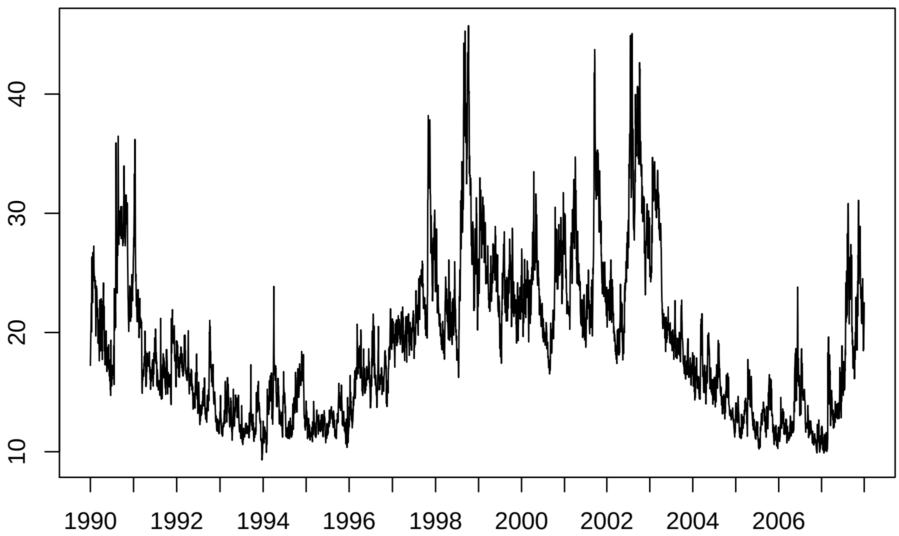

Figure 1: VIX index 1.1.1990 - 31.12.2007

The VIX is a very persistent time series, as ita daily levels display high autocorrelation (shown in Figure 3). The autocorrelations for the differenced time series are also provided in Figure 3. A unit root is rejected by the ADF test for both the level and differenced time series at the one-percent level of significance.[7]

[6]This would be consistent with the findings of Harvey and Whaley (1992). On the other hand, Simon (2003) reports (relatively weak) evidence to the contrary.

[7]The p-value of the test is 0.001 for the level series and less than 0.0001 for the differenced series.

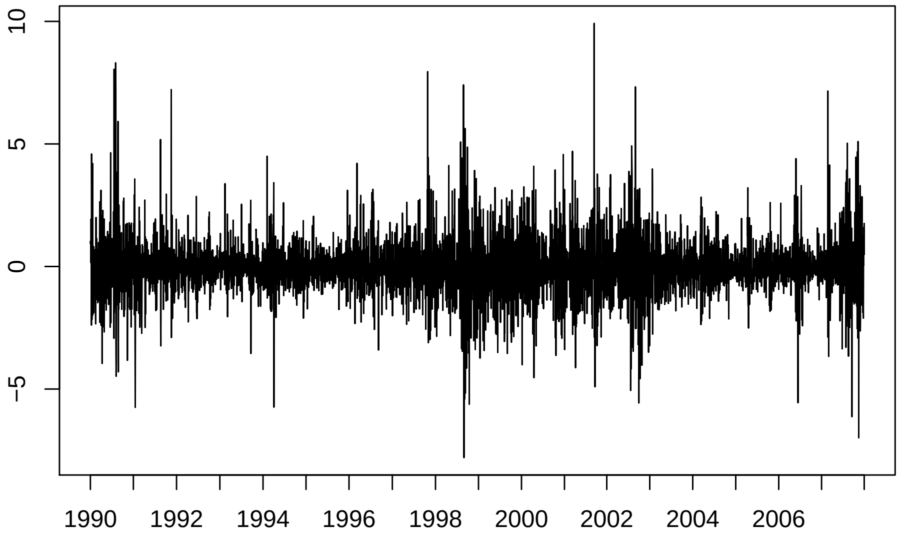

Figure 2: VIX first differences 1.1.1990 - 31.12.2007

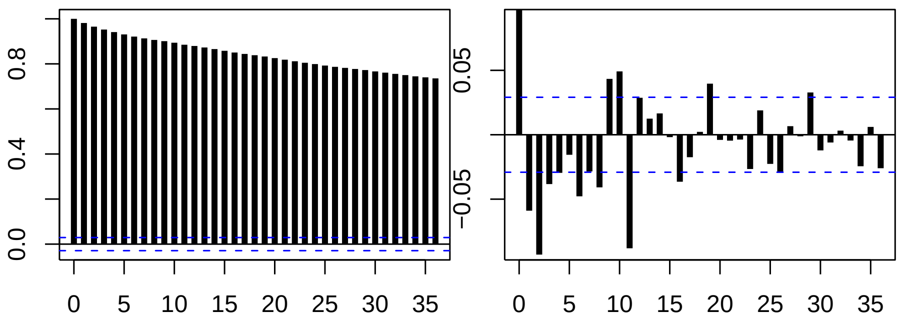

Figure 3: Autocorrelations for log VIX (above) and log VIX first differences (below)

Logarithms of the VIX observations were taken in order to avoid negative forecasts of volatility. As noted in Simon (2003), the use of logs is consistent with the positive skewness in IV data. ARIMA and ARFIMA models were estimated for the log VIX levels, but despite the high persistence in the time series, even the ARFIMA models failed to produce useful forecasts (directional accuracy of over 50 percent). Also, the value received for d in the ARFIMA modeling exceeded 0.49 when its value was restricted to be less than 0.5, and fell within the non-stationary range of $[0.5, 1]$ when no restrictions were placed on its value. In light of this evidence, the models in this study were built for log VIX first differences rather than levels.[8] This choice of modeling volatility changes is also supported by Fleming et al. (1995), who argue that academics and practitioners alike are interested primarily in changes in expected volatility. Also, given the high level of autocorrelation in the VIX level series, Fleming et al. (1995) remark that inference in finite samples could be adversely affected.

Descriptive statistics for the VIX index, its first differences, and log VIX first differences (the dependent variable in the models of this study) are provided in Table 1. The largest day-to-day changes in the VIX are quite large, a gain of 9.92 and a drop of 7.80. The largest gain takes place on the first trading day after 9/11, and the largest drop was witnessed during the turbulence of late summer and early fall 1998. The data is skewed to the right, and the differenced series display sizeable excess kurtosis, or fat tails.

	VIX	VIX change	ln(VIX change)
Mean	18.975	0.001	0.00006
Max	45.74	9.92	0.49601
Min	9.31	-7.8	-0.29987
Median	17.73	-0.04	-0.00224
St. dev.	6.384	1.226	0.05763
Skewness	0.974	0.522	0.64324
Excess kurtosis	0.794	6.188	4.5048

Table 1: Descriptive statistics for the VIX index, VIX first differences, and log VIX first differences. Statistics cover full sample of 1.1.1990-31.12.2007.

Two alternative in-sample periods were used in the model-building phase in order to determine the robustness of the results and stability of coefficients over time. The full in-sample of 1.1.1990 - 31.12.2002, or 3,279 observations, is the base case that is compared to an alternative of 1,000 observations. This corresponds to an in-sample period of 8.1.1999 - 31.12.2002. The second sample period is selected in order to see whether forecast performance would improve with only a short period of observations: conditions in financial markets can change rapidly, and perhaps only the most recent information is relevant for forecasting purposes. The minimum number of observations is set at 1,000 to ensure that GARCH parameters can be estimated reliably. Studies such as Noh et al. (1994) and Blair et al. (2001) use 1,000 observations when calculating forecasts from GARCH models. Also, Engle et al. (1993) calculate variance forecasts for

The lag lengths in the ADF test are eleven for the level series and ten for the differenced series, based on selection with the Schwarz Information Criterion.

[8] Probit models were also estimated for VIX first differences, with a move upward denoted by 1 and a move downward by 0. The forecast performance of these models was inferior to that of ARIMA models, and these results are therefore not reported.

an equity portfolio and find that ARCH models using 1,000 observations perform better than models with 300 or 5,000 observations.

Based on tests for no remaining autocorrelation and the p-values for the statistical significance of the coefficients, an ARIMA(1,1,1) specification was found to be the best fit for the log VIX time series from the family of ARIMA models. Konstantinidi et al. (2008) also settle on this specification. As IV has been found to exhibit weekly seasonality (see e.g. Harvey and Whaley (1992) and Brooks and Oozeer (2002)), the significance of day-of-the-week effects is investigated with dummy variables. The VIX index displays a clear weekly pattern, with the index level on average highest on Mondays and lowest on Fridays. More importantly for the modeling of changes in the VIX, the index level tends to rise on average on Mondays, fall slightly on Tuesdays, Wednesdays, and Thursdays, and experience a more pronounced drop on Fridays. Therefore, it can be deduced that a day-of-the-week dummy would most likely be significant for Mondays and Fridays.

Judging by the pattern of VIX returns in Figure 2, the VIX index could possibly display volatility of volatility, or conditional heteroskedasticity in shocks to the VIX. To further investigate this consideration, the ARIMA(1,1,1) model was augmented with GARCH errors.[9] In modeling realized volatility, Corsi et al. (2008) find that accounting for the volatility of realized volatility improves forecast performance. In light of the probably existence of GARCH effects, heteroskedasticity-robust standard errors are used in all model estimation. The financial and macroeconomic indicators outlined in Section 2.2 were also included in the regressions to see whether they could improve ARIMA models of the VIX. The estimated linear equation is

$$\Delta VIX_t = \omega + \phi_1 \Delta VIX_{t-1} + \theta_1 \epsilon_{t-1} + \sum_{i=1}^{r} \phi_i \mathbf{X}_{i,t-1} + \sum_{k=1}^{5} \gamma_k D_{k,t} + \epsilon_t \tag{1}$$

where ΔVIX_t is the log return of the VIX index, the weekday dummy variable $D_{k,t}$ receives the value of 1 on day k and zero otherwise, and all explanatory variables other than the AR and MA components are grouped in the vector $\mathbf{X}$. The model is a first-order model throughout, as no second lags turned out to be statistically significant. When augmenting the model of Equation 1 with a model for the conditional variance, a GARCH(1,1) specification is found to be sufficient and is parameterized as follows

$$\epsilon_t = N(0, h_t^2)$$

$$h_t^2 = \kappa + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}^2$$

The estimation results for four alternative model specifications and the two in-sample periods are presented in Table 2. The estimated models are an ARIMA(1,1,1) model, an ARIMA(1,1,1)-GARCH(1,1) model, and the same two models with other explanatory variables included: ARIMAX(1,1,1) and ARIMAX(1,1,1)-GARCH(1,1). Consistent with earlier evidence on weekly seasonality in implied volatility in equity markets (see e.g. Harvey and Whaley (1992) and Ahoniemi and Lanne (2007)), dummy variables for Monday and Friday are statistically significant. The only financial or macroeconomic variable that was found to be statistically significant was the lagged log return of the

[9] GARCH modeling was proposed by Bollerslev (1986), extending the work of Engle (1982).

S&P 500 index.[10] Data on the S&P 500 index trading volume was not available for the full in-sample period, so its significance was assessed only with the shorter sample period. Contrary to expectations, neither the volume nor its first difference proved to be significant in explaining changes in the VIX index. Brooks and Oozeer (2002) also found that trading volume was not significant in explaining changes in IV. The addition of the GARCH specification for the error term fits the chosen models, with statistically significant coefficients for both the ARCH and GARCH terms. The full value of GARCH modeling will be underscored in the forecasting application of Section 4.

When using the full in-sample period for estimation and comparing goodness-of-fit with the Schwarz Information Criterion (BIC), the ARIMA-GARCH model emerges as the favored specification. The AR and MA parameters are close in value for all four models. The parameters of the GARCH equation sum to 0.875 in both cases, meaning that stationarity is achieved. Also, this indicates that the volatility of volatility is not highly persistent. The weekday dummies have the expected signs: the positive Monday dummy (γ_1) is consistent with the VIX tending to rise on Mondays, and the negative Friday dummy (γ_5) is consistent with the drop that the VIX experiences on average on Fridays.

The coefficient for S&P 500 index returns is negative, indicating that a negative return over the previous trading day raises the VIX on the following day, and a positive return equivalently lowers the VIX. This result is comparable to those in Low (2004), who estimates a negative coefficient for contemporaneous index returns in a model for VIX changes. Bollen and Whaley (2004) also estimate a negative coefficient for contemporaneous returns of the S&P 500 index as an explanatory variable for changes in the IV of S&P 500 index options. Fleming et al. (1995) also find that the VIX and stock returns have a strong negative contemporaneous correlation, although they do estimate a small positive coefficient for lagged stock index returns as an explanatory variable of VIX changes. The use of lagged returns is essential in the context of this study, as the models will later be used for forecasting.

Simon (2003) provides further discussion on how to explain the finding that the VIX tends to fall after positive returns and rise after negative returns. A fall in the market can easily lead to more demand for puts as a form of portfolio insurance, thus raising IV through increased option demand. On the other hand, after a rise in the underlying index level, options with a higher strike price become at-the-money (ATM) options. Due to the well-documented volatility skew in equity option implied volatilities, these higher-strike options will have lower IVs than those options that were previously ATM. Hence, the level of ATM implied volatility will fall, even if the IV of the same-strike options does not change.

The coefficient for the S&P 500 index returns in the ARIMAX model receives a p-value of 0.11, and in the ARIMAX-GARCH model, the p-value of the coefficient is 0.29. Therefore, it would seem to be unnecessary to include the underlying index returns

[10] Davidson et al. (2001) create two explanatory variables from lagged returns by separating positive and negative returns. This can help capture whether negative return shocks have a larger effect on implied volatility than positive return shocks, a result that has often been reported for stock return volatility (see e.g. Schwert (1990) and Glosten et al. (1993)). With the data sample in this study, however, the positive and negative return series, when used separately, were not statistically significant. Simon (2003), who models the Nasdaq Volatility Index VXN, finds that the VXN responds in equal magnitude, but in opposite directions, to contemporaneous positive and negative returns of the Nasdaq index. An opposite approach to the relation between IV and stock returns is taken in Banerjee et al. (2007) and Giot (2005), who use the VIX to predict stock market returns.

	ARIMA		ARIMA-GARCH		ARIMAX		ARIMAX-GARCH	
Panel A - 3,279 observations								
BIC	4812.64		4872.43		4810.25		4868.97	
R^2	0.067		0.067		0.068		0.068	
ω	-0.004	(0.000)	-0.004	(0.000)	-0.003	(0.000)	-0.004	(0.000)
ϕ_1	0.812	(0.000)	0.856	(0.000)	0.788	(0.000)	0.843	(0.000)
θ_1	-0.902	(0.000)	-0.934	(0.000)	-0.891	(0.000)	-0.928	(0.000)
ψ_1	-		-		-0.190	(0.110)	-0.113	(0.286)
γ_1	0.027	(0.000)	0.027	(0.000)	0.027	(0.000)	0.027	(0.000)
γ_5	-0.008	(0.004)	-0.009	(0.001)	-0.008	(0.004)	-0.009	(0.001)
κ	-		0.002	(-)	-		0.002	(-)
α_1	-		0.085	(0.001)	-		0.084	(0.001)
β_1	-		0.790	(0.000)	-		0.791	(0.000)
Panel B - 1,000 observations								
BIC	1490.65		1491.27		1487.31		1487.89	
R^2	0.084		0.084		0.084		0.084	
ω	-0.002	(0.154)	-0.004	(0.038)	-0.002	(0.147)	-0.004	(0.035)
ϕ_1	0.803	(0.000)	0.777	(0.000)	0.788	(0.000)	0.765	(0.000)
θ_1	-0.883	(0.000)	-0.863	(0.000)	-0.875	(0.000)	-0.857	(0.000)
ψ_1	-		-		-0.070	(0.643)	-0.058	(0.675)
γ_1	0.030	(0.000)	0.031	(0.000)	0.030	(0.000)	0.031	(0.000)
γ_5	-0.015	(0.001)	-0.015	(0.001)	-0.015	(0.001)	-0.015	(0.001)
κ	-		0.001	(-)	-		0.001	(-)
α_1	-		0.059	(0.022)	-		0.059	(0.004)
β_1	-		0.889	(0.000)	-		0.889	(0.000)

Table 2: Estimation results for ARIMA, ARIMA-GARCH, ARIMAX, and ARIMAX-GARCH models. Models estimated with entire in-sample in upper panel, and models estimated with 1,000 observations in lower panel. P-values for the statistical significance of the coefficients are given in parentheses. Heteroskedasticity-robust standard errors were used throughout the analysis.

in the model specification. The ARIMAX and ARIMAX-GARCH specifications are nonetheless included in the set of models used in the forecasting application as there is mixed evidence regarding the significance of the index returns (see Sections 4 and 5 for further discussion).

With an in-sample of 1,000 observations, the magnitude of the estimated AR and MA parameters, as well as the dummy variables, remains more or less the same as with the longer sample period. The coefficient for S&P 500 index returns is negative as with the longer sample, but it is now clearly not statistically significant (p-values exceed 0.6). The sum of the GARCH parameters is higher than with the longer sample, 0.95 for both models. According to the BIC, the ARIMA-GARCH model continues to be the best fit to the data.

The residuals for the ARIMA model estimated with the entire in-sample are shown in Figure 4 and the conditional variances for the ARIMA-GARCH model are in Figure 5. The clustering in the residuals and the spikes in the conditional variances again point to the conditional heteroskedasticity in the time series of VIX first differences. These figures are provided as representative examples, and the equivalent figures from the other models are similar.

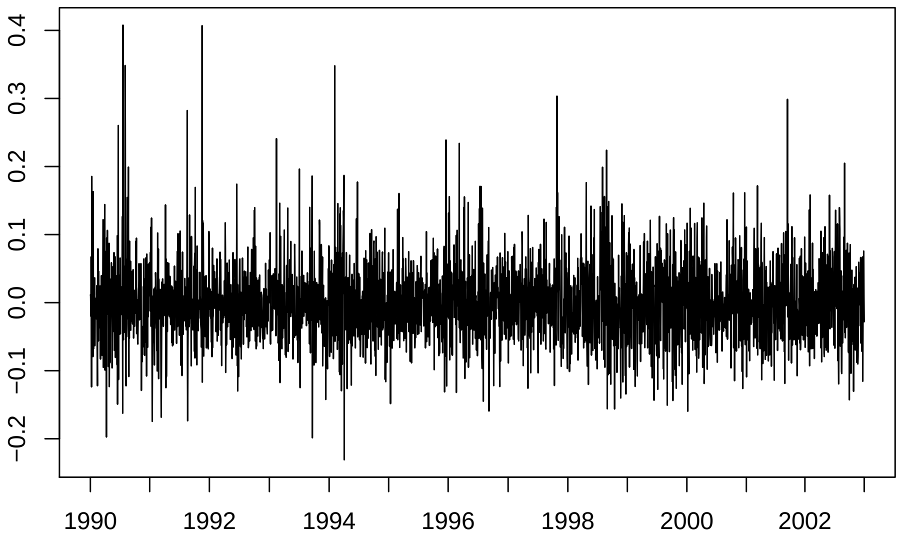

Figure 4: Residuals for ARIMA-GARCH model estimated with entire in-sample (3,279 observations).

Lagrange multiplier (LM) tests for autocorrelation, remaining ARCH effects, and heteroskedasticity were run for all the above model specifications. P-values of the tests are reported in Table 3. Five lags were used in the tests for autocorrelation and remaining ARCH. The test results differ somewhat depending on the length of the data sample. For the in-sample of 3,279 observations, the test for autocorrelation indicates that the null hypothesis of no autocorrelation cannot be rejected for any of the considered models. GARCH errors make a pronounced change to the LM test results regarding remaining ARCH effects and heteroskedasticity, so taking conditional heteroskedasticity into account appears to be warranted. The null hypotheses of no neglected ARCH and no heteroskedasticity cannot be rejected after adding the conditional variance specification to the models. For the ARIMA and ARIMAX models, the null of no remaining ARCH effects is rejected at the one-percent level, and the null of no heteroskedasticity at the five-percent level.

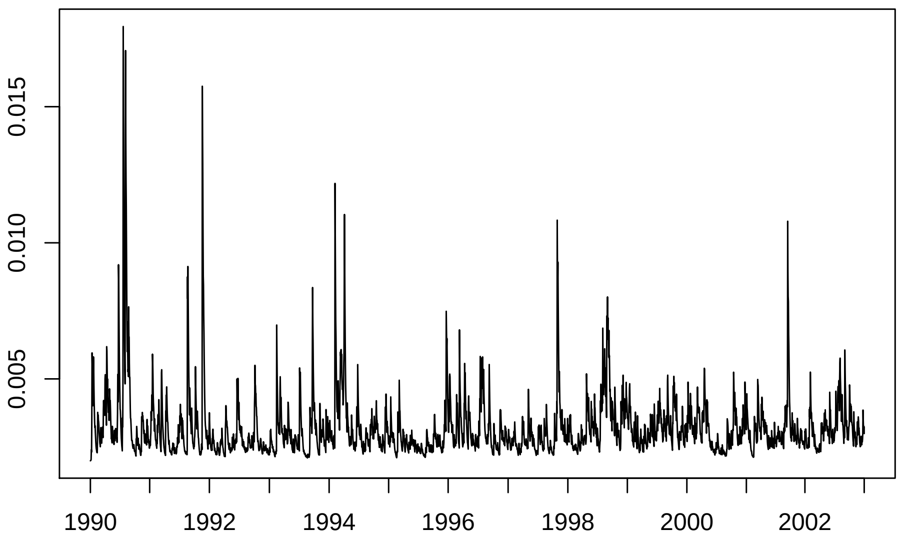

Figure 5: Conditional variances for ARIMA-GARCH model estimated with entire in-sample (3,279 observations).

When estimating the models with only 1,000 observations, the test results are in part contrary to expectations. The null of no autocorrelation is now rejected at the ten-percent level for the ARIMA and ARIMAX models, and the five-percent level for the ARIMA-GARCH and ARIMAX-GARCH models. For the remaining ARCH test, the results are similar to those received with the longer in-sample. The null hypothesis is rejected at the one-percent level when GARCH errors are not included in the model, but the null is not rejected when conditional heteroskedasticity is modeled. The null of no heteroskedasticity cannot be rejected for any model specification, including the the ARIMA and ARIMAX models. Though in part surprising, these tests results are not a cause for concern, as the shorter in-sample period proves to be a better forecaster.

	ARIMA	ARIMA-GARCH	ARIMAX	ARIMAX-GARCH
Panel A - 3,279 observations				
Autocorrelation	0.255	0.158	0.448	0.272
ARCH effects	0.000	0.920	0.000	0.914
Heteroskedasticity	0.021	0.778	0.013	0.771
Panel B - 1,000 observations				
Autocorrelation	0.094	0.038	0.051	0.018
ARCH effects	0.000	0.176	0.000	0.183
Heteroskedasticity	0.491	0.932	0.468	0.981

Table 3: P-values from Lagrange multiplier tests for remaining autocorrelation, remaining ARCH effects, and heteroskedasticity. Five lags are used in the first two tests.

4 Forecasts

The next step in the analysis is to obtain forecasts from the four model specifications described in Section 3. Out-of-sample, one-step-ahead daily forecasts were calculated for the VIX first differences for the period 1.1.2003-31.12.2007, spanning 1,258 trading days. The long out-of-sample period will help to ensure the robustness of the results. The forecasts were calculated from rolling samples, keeping the sample size constant each day. In other words, after calculating each forecast, the furthest observations are dropped, the observations for the most recent day are added to the sample, and the parameter values are re-estimated. In practice, the change in the VIX from day T-1 to day T is regressed on all specified variables with time stamps up to and including day T-1. The estimated parameter values are then used together with day T values to predict the change in the VIX from day T to day T+1. Only the dummy variables are treated differently, i.e. day T+1 dummy variables are used when forecasting the change in VIX from day T to day T+1. Both sample sizes are included in the analysis, i.e. 3,279 and 1,000 observations.

Successful forecasting of IV from a trader's point of view primarily involves forecasting the direction of IV correctly; a correct magnitude for the change is not as relevant. This is because option positions such as the straddle will generate a profit if the IV moves in the correct direction, ceteris paribus (the size of the profit is affected by the magnitude of change, however). The forecasting accuracy of the various models is first evaluated based on sign: how many times does the sign of the change in the VIX correspond to the direction forecasted by the model. However, the point forecasts are also evaluated based on mean squared errors (MSE). Accurate point forecasts can be valuable, for example, in risk management and asset pricing applications.

Table 4 presents the forecast performance of the various models, measured with the correct direction of change and mean squared errors. When using the longer sample window, the models succeed in predicting the direction of change correctly for 56.7-57.2% of the trading days, with accuracy improving to 57.3-58.4% with the shorter sample period. The best forecaster is the ARIMA-GARCH model estimated with 1,000 observations. A relevant result is also that the addition of GARCH errors improves forecast performance for all four pairs of models. On the other hand, the inclusion of S&P 500 index returns in the model does not improve the directional forecasts. The improved performance due to modeling conditional heteroskedasticity receives theoretical support from Christoffersen and Diebold (2006), who show that it is possible to predict the direction of change of returns in the presence of conditional heteroskedasticity, even if the returns themselves cannot be predicted. In fact, Christoffersen and Diebold (2006) contend that sign dynamics can be expected in the presence of volatility dynamics.

The choice of either the ARIMA-GARCH or ARIMAX-GARCH model coupled with the shorter sample period is robust in the sense that the models also emerge as the best directional forecasters if the out-of-sample period is broken down into two sub-samples of equal length, or 629 observations. In the first sub-sample, the accuracy of the ARIMA-GARCH model is highest with 58.8%, and in the second sub-sample, the ARIMAX-GARCH model fares best with correct predictions on 58.5% of trading days.

A directional accuracy of over 50% can be seen as valuable for option traders. Given the large trading volumes in U.S. markets, it also seems realistic that the achieved accuracy is not dramatically larger than 50%. Harvey and Whaley (1992) model the IV of S&P 100 options and achieve a directional accuracy of 62.2% for call options and 56.6% for puts. Their sample is much older, and it is feasible that predictability has declined over the years rather than improved. Also, the S&P 500 receives more attention among today's investors than the S&P 100 index. Konstantinidi et al. (2008) forecast the direction of change in the VIX at best on 54.7% of trading days in their out-of-sample period of roughly 640 days. In their similarly oriented study, Brooks and Oozeer (2002) report correct sign predictions for the IV of options on Long Gilt futures for 52.5% of trading days.

When assessing point forecasts with MSEs, it appears to be useful to use a longer in-sample period for model estimation. In fact, the ARIMA and ARIMAX models now perform better than the ARIMA-GARCH or ARIMAX-GARCH models. All values obtained when using 1,000 observations are greater than when using the entire in-sample period. When looking at the two equal-length sub-samples, the ARIMA and ARIMAX models estimated from 3,279 observations remain the top two performers, so this result is relatively robust. However, the issue of whether or not the differences in MSEs are statistically significant will be investigated shortly.

	3,279 observations			1,000 observations		
	Correct sign	*%*	*MSE*	*Correct sign*	*%*	*MSE*
ARIMA	716	56.9%	0.00328	721	57.3%	0.00330
ARIMA-GARCH	720	57.2%	0.00329	735	58.4%	0.00332
ARIMAX	713	56.7%	0.00328	721	57.3%	0.00329
ARIMAX-GARCH	720	57.2%	0.00329	733	58.3%	0.00331

Table 4: Directional forecast accuracy (out of 1,258 trading days) and mean squared errors for forecasts from all four models and both sample periods.

In the spirit of the particular nature of this study, the predictive ability of the models was tested using the market timing test for predictive accuracy developed by Pesaran and Timmermann (1992). The Pesaran-Timmermann test (henceforth PT test) was originally developed with the idea that an investor switches between stocks and bonds depending on the returns expected from each asset class, and can be used in the present context to ascertain that the directional forecasts outperform a coin flip. The PT test is calculated with the help of a contingency table that shows how many times the actual outcome was up if the forecast was up, how many times the outcome was down even if the forecast was up, and likewise for the other two combinations. In practice, the four models analyzed here forecast down too often, leading to more mistakes where the forecast was down but the true change was up than vice versa. The PT test confirms that the estimated models do possess market timing ability. The null hypothesis of predictive failure is rejected at the one-percent level for all models with both in-sample periods.

Returning to whether or not the differences between the MSEs are statistically significant, the Diebold-Mariano (DM) test is run next.[11] Comparing the ARIMA model estimated with 3,279 observations, which has the lowest MSE, to the ARIMA-GARCH model estimated with 1,000 observations, which was the best directional forecaster but has the highest MSE, the DM test does not reject the null hypothesis of equal predictive accuracy (the p-value of the test is 0.12). In other words, there is no statistically significant difference between the models when evaluated with mean squared errors, and model selection will be based on directional accuracy.

5 Option trading

An attractive application of the VIX forecasts is to provide useful information for option traders. The directional forecasts from the models presented above were used to simulate

[11]This test is described in Diebold and Mariano (1995).

option trades with S&P 500 option market prices. The trades that were simulated were straddles, which are spreads that involve buying or selling an equal amount of call and put options. The trading simulation provides a way to assess the forecasting ability of the above time series models with their economic significance.

Straddles are the leading strategy for trading volatility (Ni et al. (2008)). Their use in this simulation should be more realistic than the buy-or-sell strategies simulated by Brooks and Oozeer (2002) and Harvey and Whaley (1992). Bollen and Whaley (2004) note that buying straddles is the most effective way to trade if one expects volatility to rise. A long (short) straddle, i.e. an equal number of bought (sold) call and put options, yields a profit if IV rises (falls).

An out-of-sample option trading simulation, with trades executed based on forecasts for the VIX, requires daily close quotes of at-the-money (or in practice, near-the-money) S&P 500 index options. The S&P 500 option quotes were obtained from Commodity Systems, Inc. for the out-of-sample period of 1.1.2003 to 31.12.2007. Buraschi and Jackwerth (2001) note that at-the-money S&P 500 options have the highest trading volume. Ni et al. (2008) observe that investors with a view on volatility are more likely to trade with near-the-money options than in-the-money or out-of-the-money options, and Bollen and Whaley (2004) point out that at-the-money options have the highest sensitivity to volatility. Daily straddle positions were simulated with this data by utilizing the out-of-sample forecasts from the four models presented above, and with both in-sample period lengths.

The option positions are opened with the close quotes on day T and closed with the close quotes on day T+1, which is the day for which the directional forecast is made. This strategy allows for using options that are as close-to-the-money as possible on each given day. The strike price is chosen so that the gap between the actual closing quote of the S&P 500 index from the previous day (day T) and the option's strike price is the smallest available. The only exceptions to using the closest-to-the-money options come on days when there is zero trading volume in that series on either day T or day T+1. On such days, the next-closest contract was used in the simulation.

Options with the nearest expiration date were used, up to fourteen calendar days prior to the expiration of the nearby option, when trading was rolled over to the next expiration date. This is necessary as the IV of an option close to maturity may behave erratically. Poon and Pope (2000) analyze S&P 100 and S&P 500 option trading data for a period of 1,160 trading days and find that contracts with 5-30 days of maturity have the highest number of transactions and largest trading volume.

This analysis does not incorporate transaction costs, as the main purpose of the exercise is not to obtain accurate estimates of actual (possibly abnormal) profits, but to use the trading profits and losses to rank the forecast models. The deduction of a fee would naturally scale all profits downwards, but would not change the ranking of the models. The issue of transaction costs is addressed further in the discussion of filter use below.

The option positions are technically not delta neutral, which means that the trading returns are sensitive to large changes in the value of the options' underlying asset, or the S&P 500 index, during the course of the day. However, this problem was not deemed critical for this analysis. The deltas of at-the-money call and put options nearly offset each other,[12] so that the positions are close to delta neutral when they are opened at

[12] see Noh et al. (1994)

the start of each day. The deviations from delta neutrality in this study come primarily from the fact that strike prices are only available at certain fixed intervals. The positions are updated daily, so the strike price used can be changed each day. Also, Engle and Rosenberg (2000) and Ni et al. (2008) note that straddles are sensitive to changes in volatility but insensitive to changes in the price of the options' underlying asset. Driessen and Maenhout (2007) note that the correlation between ATM straddle returns and equity returns is only -0.07 for the S&P 500 index. In the trading simulation of Jackwerth (2000), unhedged and hedged strategies yield similar excess returns.

Although the straddle is a volatility trade, its returns are naturally not completely dependent on the changes in IV. Even a trader with perfect foresight about the direction of change in the VIX would lose on her straddle position on 493 days out of the 1,258 days analyzed, or on 39.2 percent of the days.

In practice, if the forecasted direction of the VIX was up, near-the-money calls and near-the-money puts were bought. Equivalently, if the forecast was for the VIX to fall, near-the-money calls and near-the-money puts were sold. The exact amounts to be bought or sold were calculated separately for each day so that 100 units (dollars) were invested in buying the options each day, or a revenue of 100 was received from selling the options. This same approach of fixing the investment outlay has been used in a number of studies, for example, in Harvey and Whaley (1992), Noh et al. (1994), Jackwerth (2000), and Ederington and Guan (2002). The return from a long straddle is calculated as in Equation 2, and the return from a short straddle is shown in Equation 3. In this analysis, the proceeds from selling a straddle are not invested during the day, but held with zero interest.

$$R_l = \frac{100}{C_t + P_t}(-C_t - P_t + C_{t+1} + P_{t+1}) \tag{2}$$

$$R_s = \frac{100}{C_t + P_t}(C_t + P_t - C_{t+1} - P_{t+1}) \tag{3}$$

In Equations 2 and 3, C_t is the close quote of a near-the-money call option, P_t is the close quote of a near-the-money put option (with same strike and maturity as the call option), and C_{t+1} and P_{t+1} are the respective closing quotes of the same options at the end of the next trading day.

Although the emphasis is on directional accuracy, filters have also been used in the option trading simulations. These three filters leave out the weakest signals, i.e. signals that predict the smallest percentage changes, as they may not be as reliable in the directional sense. Harvey and Whaley (1992) and Noh et al. (1994) employ two filters to leave out the smallest predictions of changes, and Poon and Pope (2000) use three filters in order to take transaction costs into account. This use of filters can indeed be seen as a way to account for transaction costs, as very small changes in the VIX may lead to option trading profits that are so small that they are eaten away by transaction costs. The three filters leave out a trading signal when the projected change in log VIX is smaller than 0.1%, 0.2%, or 0.5%.[13]

The returns from trading options on all days, and when employing a filter, are presented in Table 5. The returns are measured in daily percentage returns on the investment outlay of 100. One revision is done to the data before calculating the final

[13] A filter of 1.0% gave a trading signal on only 70 to 95 trading days out of 1,258, depending on the model.

returns. The trading returns for four days in the 1,258 sample period are removed from the returns in order to reduce noise. These four days, which all fall in the year 2007[14], produce daily straddle returns of well over 100%. Whether the return is positive or negative depends on the trading signal for that day. The range of values that the returns take on these six days is from 120.4% to 310.3%, which are deemed so large that they could easily bias the results upward or downward, depending on whether a model forecasts the signal correctly on these select four days. These four days happen to be days when the values of nearest-to-the-money options experienced relatively dramatic moves. Importantly, the S&P 500 index did not experience a large rise or drop on any of these days, so the outliers are not a consequence of unhedged exposure to moves in the options' underlying. The cut-off of 100% is by no means arbitrary, as the next largest daily return is 47.5%, well below the four outliers.

	Trading P/L	Filter I		Filter II		Filter III	
		Panel A - 3,279 observations					
ARIMA	0.23%	0.11%	(952)	0.05%	(722)	-0.04%	(347)
ARIMA-GARCH	0.23%	0.09%	(959)	0.00%	(747)	-0.06%	(358)
ARIMAX	0.27%	0.13%	(962)	0.07%	(727)	-0.04%	(347)
ARIMAX-GARCH 0.21%	0.16%	(973)	-0.01%	(752)	-0.09%	(361)	
		Panel B - 1,000 observations					
ARIMA	0.23%	0.22%	(911)	0.08%	(696)	-0.16%	(320)
ARIMA-GARCH	0.29%	0.28%	(1059)	0.17%	(850)	-0.02%	(379)
ARIMAX	0.26%	0.27%	(925)	0.05%	(707)	-0.10%	(330)
ARIMAX-GARCH	0.36%	0.28%	(1068)	0.17%	(853)	-0.02%	(379)

Table 5: Option trading returns for trading on all days in column Trading P/L. Returns expressed as percentage daily returns on an investment outlay of 100. Number of days traded out of 1,258 in parentheses for filtered returns. Filter I leaves out signals where the change in log VIX is projected to be less than 0.1%. The corresponding values for Filters II and III are 0.2% and 0.5%.

Turning now to the returns themselves, the model choice from Section 4 is in fact confirmed. The average daily return on the straddle positions is 0.36% with the ARIMAX-GARCH model estimated with the shorter in-sample period, and the second-best return is produced by the ARIMA-GARCH model, again coupled with 1,000 observations. This same top two emerged in the forecast evaluation. Although the series of S&P 500 returns was not statistically significant in the models when estimated with 1,000 observations, the option trading results indicate that there may be incremental information in the index returns as an explanatory variable for the changes in the VIX. This result receives further support from the analysis of sub-samples. As in Section 4, the returns are now evaluated looking at sub-samples of 629 observations. In the first sub-sample, the top returns (0.50%) come from the ARIMAX model with 1,000 observations, but the ARIMAX-GARCH model is not far behind with average daily returns of 0.48%. In the second sub-sample, the ARIMAX model's returns plummet and it is the weakest performer among all eight models with a daily return of only 0.01%. The ARIMAX-GARCH model is now the top performer with 0.24%. Regarding the use of the filters, it appears that it is not necessary to filter the returns in general. Returns improve in only one case (the ARIMAX model estimated with 1,000 observations) when using Filter I. With Filter II and Filter III, returns clearly suffer. However, if interpreting the filters as a way of incorporating transaction costs, it is clear that returns should indeed suffer

[14]The exact dates are Feb. 27, June 1, July 6, and Oct. 5.

when using the filters.

In light of the forecast evaluation in Section 4 and the above analysis of option returns, the recommendation of this study would be to trade based on the forecasts of an ARIMA-GARCH or an ARIMAX-GARCH model estimated with 1,000 observations. The evidence in favor of the two alternative specifications is nearly as strong. The use of a shorter history proved valuable for both forecast accuracy and option returns, indicating that the more recent market conditions are relevant for IV forecasters. The ARIMA-GARCH model is slightly better as a directional forecaster over the whole out-of-sample period, but the returns from the ARIMAX-GARCH model are slightly better in the straddle trading exercise.

6 Conclusions

The forecasting of implied volatility is of interest for option market practitioners, as well as for investors with portfolio risk management concerns. The size of this latter group in particular underscores the importance of the task at hand in this paper. This paper has sought to find well-fitting ARIMA and ARIMAX models for the VIX index, to analyze their predictive ability, and to calculate the returns from a straddle trading simulation based on forecasts from the models. An ARIMA(1,1,1) specification is the best fit to the data, and the returns of the S&P 500 index are a statistically significant explanatory variable for the first differences of the VIX when a longer in-sample period is used in model estimation. Despite the high persistence in the time series of the VIX index, ARFIMA modeling proved unsuccessful.

GARCH terms are statistically significant in ARIMA(1,1,1) models, and the conditional variance specification improves the directional forecast accuracy of the various models considered. The best model, ARIMA(1,1,1)-GARCH(1,1), forecasts the direction of change of the VIX correctly on 58.4 percent of the trading days in the 1,258-day out-of-sample period. This best performance comes when using a shorter in-sample period of 1,000 observations, indicating that only the most recent history is relevant in forecasting the VIX index. The Pesaran-Timmermann test confirms that the forecast accuracy of the best models is statistically significant. In the option trading application, straddle trades confirm the value of modeling conditional heteroskedasticity and using a shorter data sample. The primary purpose of the trading exercise is to provide additional evidence as to the best forecasting model, and no conclusion regarding market efficiency is thus drawn.

As was to be expected in light of earlier research, there seems to be a certain degree of predictability in the direction of change of the VIX index. This predictability can potentially be exploited profitably by option traders, at least for certain periods of time. The profitability of a trading strategy after transaction costs would most likely paint a different picture for market makers versus other investors. A more detailed investigation of trading returns based on forecasts from models such as those included in this study is left for future research.

References

[1] Ahoniemi, K. & Lanne, M. (2007). Joint Modeling of Call and Put Implied Volatility. HECER Discussion Paper, 198.

[2] Banerjee, P.S., Doran, J.S. & Peterson, D.R. (2007). Implied Volatility and Future Portfolio Returns. Journal of Banking & Finance, 31, 3183–3199.

[3] Becker, R., Clements, A.E. & White, S.I. (2006). On the Informational Efficiency of S&P500 Implied Volatility. North American Journal of Economics and Finance, 17, 139–153.

[4] Blair, B.J., Poon, S-H. & Taylor, S.J. (2001). Forecasting S&P 100 Volatility: the Incremental Information Content of Implied Volatilities and High-frequency Index Returns. Journal of Econometrics, 105, 5–26.

[5] Bollen, N.P.B. & Whaley, R.E. (2004). Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?. The Journal of Finance, 59, 711–753.

[6] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307–327.

[7] Brooks, C. & Oozeer, M.C. (2002). Modelling the Implied Volatility of Options on Long Gilt Futures. Journal of Business Finance & Accounting, 29, 111–137.

[8] Buraschi, A. & Jackwerth, J. (2001). The Price of a Smile: Hedging and Spanning in Option Markets. The Review of Financial Studies, 14, 495–527.

[9] Carr, P. & Wu, L. (2006). A Tale of Two Indices. Journal of Derivatives, 13:3, 13–29.

[10] Christoffersen, P.F. & Diebold, F.X. (2006). Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics. Management Science, 52, 1273–1287.

[11] Corrado, C.J., Miller, T.W., 2005. The Forecast Quality of CBOE Implied Volatility Indexes. The Journal of Futures Markets 25, 339–373.

[12] Corredor, P., Lechón, P. & Santamaría, R. (2002). Is it Possible to Obtain Profits Using Naïve Volatility Models?. Derivatives Use, Trading & Regulation, 8, 13–32.

[13] Corsi, F., Mittnik, S., Pigorsch, C. & Pigorsch, U. (2008). The Volatility of Realized Volatility. Econometric Reviews, 27, 46–78.

[14] Davidson, W.D., Kim, J.K., Ors, E. & Szakmary, A. (2001). Using Implied Volatility on Options to Measure the Relation Between Asset Returns and Variability. Journal of Banking & Finance, 25, 1245–1269.

[15] Dennis, P., Mayhew, S., & Stivers, C. (2006). Stock Returns, Implied Volatility Innovations, and the Asymmetric Volatility Phenomenon. Journal of Financial and Quantitative Analysis, 41, 381–406.

[16] Diebold, F.X. & Mariano, R.S. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13, 253–263.

[17] Driessen, J. & Maenhout, P. (2007). An Empirical Portfolio Perspective on Option Pricing Anomalies. Review of Finance, 11, 561–603.

[18] Ederington, L. & Guan, W. (2002). Why Are Those Options Smiling?. Journal of Derivatives, 10:2, 9–34.

[19] Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50, 987–1006.

[20] Engle, R.F., Hong, C-H., Kane, A. & Noh, J. (1993). Arbitrage Valuation of Variance Forecasts with Simulated Options. In: Chance, D.M. & Trippi, R.R. (Eds.), Advances in Futures and Options Research, Volume 6. Greenwich, Conn.: JAI Press, pp.393–415.

[21] Engle, R.F. & Rosenberg, J.V. (2000). Testing the Volatility Term Structure Using Option Hedging Criteria. Journal of Derivatives, 8:1, 10–28.

[22] Fernandes, M., Medeiros, M.C. & Scharth, M. (2007). Modeling and Predicting the CBOE Market Volatility Index. Working paper, Pontifical Catholic University of Rio de Janeiro.

[23] Fleming, J., Ostdiek, B. & Whaley, R.E. (1995). Predicting Stock Market Volatility: A New Measure. The Journal of Futures Markets, 15, 265–302.

[24] Franks, J.R. & Schwartz, E.S. (1991). The Stochastic Behaviour of Market Variance Implied in the Prices of Index Options. The Economic Journal, 101, 1460–1475.

[25] Giot, P. (2005). Relationships Between Implied Volatility Indexes and Stock Index Returns. Journal of Portfolio Management, 31, 92–100.

[26] Giot, P. & Laurent, S. (2007). The Information Content of Implied Volatility in Light of the Jump/Continuous Decomposition of Realized Volatility. Journal of Futures Markets, 27, 337–359.

[27] Glosten, L.R., Jagannathan, R. & Runkle, D.E. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance, 48, 1779–1801.

[28] Gonçalves, S. & Guidolin, M. (2006). Predictable Dynamics in the S&P 500 Index Options Implied Volatility Surface. Journal of Business, 79, 1591–1635.

[29] Guo, D. (2000). Dynamic Volatility Trading Strategies in the Currency Option Market. Review of Derivatives Research, 4, 133–154.

[30] Harvey, C.R. & Whaley, R.E. (1992). Market Volatility Prediction and the Efficiency of the S&P 100 Index Option Market. Journal of Financial Economics, 31, 43–73.

[31] Jackwerth, J. (2000). Recovering Risk Aversion from Option Prices and Realized Returns. Review of Financial Studies, 13, 433–451.

[32] Jiang, G.J. & Tian, Y.S. (2007). Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index. The Journal of Derivatives, 14:3, 35–60.

[33] Kim, M. & Kim, M. (2003). Implied Volatility Dynamics in the Foreign Exchange Markets. Journal of International Money and Finance, 22, 511–528.

[34] Konstantinidi, E., Skiadopoulos, G. & Tzagkaraki, E. (2008). Can the Evolution of Implied Volatility be Forecasted? Evidence from European and US Implied Volatility Indices. Journal of Banking & Finance, forthcoming.

[35] Low, C. (2004). The Fear and Exuberance from Implied Volatility of S&P 100 Index Options. Journal of Business, 77, 527–546.

[36] Mayhew, S. & Stivers, C. (2003). Stock Return Dynamics, Option Volume, and the Information Content of Implied Volatility. The Journal of Futures Markets, 23, 615–646.

[37] Mixon, S. (2002). Factors Explaining Movements in the Implied Volatility Surface. The Journal of Futures Markets, 22, 915–937.

[38] Ni, S.X., Pan, J. & Poteshman, A.M. (2008). Volatility Information Trading in the Option Market. The Journal of Finance, 63, 1059–1091.

[39] Noh, J., Engle, R.F. & Kane, A. (1994). Forecasting Volatility and Option Prices of the S&P 500 Index. Journal of Derivatives, 2, 17–30.

[40] Pesaran, M.H. & Timmermann, A.G. (1992). A Simple Non-parametric Test of Predictive Performance. Journal of Business & Economic Statistics, 10, 461–465.

[41] Poon, S-H. & Pope, P.F. (2000). Trading Volatility Spreads: a Test of Index Option Market Efficiency. European Financial Management, 6, 235–260.

[42] Schwert, G.W. (1990). Stock Volatility and the Crash of '87. The Review of Financial Studies, 3, 77–102.

[43] Simon, D.P. (2003). The Nasdaq Volatility Index During and After the Bubble. Journal of Derivatives, 11:2, 9–22.