Katja Ahoniemi

Modeling and Forecasting Implied Volatility

HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS

A-340

Katja Ahoniemi

Modeling and Forecasting Implied Volatility

HELSINKI SCHOOL OF ECONOMICS

ACTA UNIVERSITATIS OECONOMICAE HELSINGIENSIS

A-340

ISSN 1237-556X
ISBN 978-952-488-303-0

E-version:
ISBN 978-952-488-304-7

Helsinki School of Economics -
HSE Print 2009

For Petri

Contents

Acknowledgements

This dissertation would never have been completed in its current form without the help and support of numerous people. First and foremost, I want to thank my supervisor, Markku Lanne. His tireless support, patience, and enthusiasm kept me going through thick and thin, and always served to give me new optimism and confidence. Also, his willingness to meet and discuss my progress whenever I felt it was necessary was of immense value throughout this project. Considering the limited number of people who conduct research on volatility in Finland, I would probably never have been able to complete a dissertation on implied volatility without our shared interest in this research topic.

I have received valuable methodological and background support from Pekka Ilmakunnas and Pertti Haaparanta at my home department in HSE. I wish to also extend my thanks to my preliminary examiners, Giampiero Gallo and Charlotte Christiansen, for their encouraging assessment of my dissertation.

The relaxed and inspirational atmosphere we have had at the HSE Department of Economics made working day to day on this project very enjoyable. I have experienced plenty of unforgettable moments with my PhD student colleagues Anni Heikkilä, Antti Kauhanen, Hanna Pesola, Hanna Virtanen, Heli Virta, Jukka Ruotinen, Kirsti Kuisma, Lotta Väänänen, Matti Ilonen, Nelli Valmari, Olli Kauppi, Pekka Sääskilahti, Sami Napari, Satu Roponen, Suvi Vasama, Tero Kuusi, and Torsten Santavirta both at work and beyond. I have shared a room with Satu, Heli, Anni, and Hanna P. at different times during this project and although that may not always have contributed towards efficient time management, our at times ceaseless chatter has helped make this job so much more fun.

I am thankful to Päivi Kankaanranta for helping me survive the first-year core courses, and to Helinä Laakkonen for sharing my interest in financial econometrics

- this would have been a lonely field otherwise. I also want to thank my old, dear friends Ella Kirjasniemi, Kirsi Helsingius, and Tiina Heinonen for their support and understanding during this adventure.

The majority of this research was carried out while I held a graduate school fellowship of the Finnish Doctoral Programme in Economics (FDPE). I am very grateful to FDPE for giving me the opportunity to work on my dissertation full time. Before I started as a GS fellow, support from the Jenny and Antti Wihuri Foundation made it possible for me to start doing my research full time. I am also thankful to a number of foundations for the financial support they provided during this project: the Finnish Foundation for Advancement of Securities Markets, the Finnish Foundation for Share Promotion, the Okobank Group Research Foundation, and the Yrjö Jahnsson Foundation.

Last but not least, I want to thank my family. My parents, Eija and Pekka, have always helped me believe that I can achieve anything I set my mind to, and holding onto that belief has definitely helped see me through this project. My wonderful husband Petri has been there for me at all times, and made my life outside of work complete. My beautiful daughters Oona and Rosa came into my life fully only after this dissertation was written, but their imminent arrival helped set some important final deadlines.

Helsinki, January 2009
Katja Ahoniemi

List of Essays

This dissertation consists of an introduction and the following four essays:

Essay I: "Modeling and Forecasting the VIX Index", unpublished. An earlier version of this essay has been published in HECER Discussion Papers, No. 129, 2006.

Essay II: "Multiplicative Models for Implied Volatility", unpublished. An earlier version of this essay has been published in HECER Discussion Papers, No. 172, 2007.

Essay III: "Joint Modeling of Call and Put Implied Volatility", with Markku Lanne. *International Journal of Forecasting*, forthcoming.

Essay IV: "Implied Volatility with Time-Varying Regime Probabilities", with Markku Lanne. HECER Discussion Papers, No. 246, 2008.

Abstract

This dissertation contains four essays, all of which model time series of implied volatility (IV) and assess the forecast performance of the models. The overall finding is that implied volatility is indeed forecastable, and its modeling can benefit from a new class of time series models, so-called multiplicative error models. It is often beneficial to model IV with two (or more) regimes to allow for periods of relative stability and periods of higher volatility in markets.

The first essay uses a traditional ARIMA model to model the VIX index. As the data displays conditional heteroskedasticity, forecast performance improves when the model is augmented with GARCH errors. 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 option trading simulation with S&P 500 index options provides further evidence that an ARIMA-GARCH model works well in forecasting changes in the VIX.

The second essay estimates two-regime multiplicative error models for the implied volatility of options on the Nikkei 225 index. Diagnostics show that the model is a good fit to the data. The implied volatility of call options turns out to be more forecastable than the implied volatility of put options in a two-year out-of-sample period. A two-regime model forecasts better than a one-regime model. When the forecasts of the multiplicative models are used to trade options on the Nikkei 225 index, the value of using two regimes is again confirmed.

The third essay investigates the joint modeling of call and put implied volatilities with a two-regime bivariate multiplicative error model. The data set is the same as in the second essay. Forecast performance improves when call IV is used as an explanatory variable for put IV, and vice versa. Forecasts with the bivariate model outperform forecasts obtained with the univariate models in the second essay. An

impulse response analysis shows that put IV recovers faster from shocks, and the effect of shocks lasts for up to six weeks.

The fourth essay models the IV of options on the USD/EUR exchange rate. This essay extends the models of the second and third essays with a time-varying probability between regimes. The changes in the USD/EUR exchange rate are used as a regime indicator, with large changes in the exchange rate signifying a more volatile regime. Out-of-sample forecasts indicate that it is beneficial to jointly model the implied volatilities derived from call and put options: both mean squared errors and directional accuracy improve when employing a bivariate rather than a univariate model.

Keywords: Implied volatility, Forecasting, Option markets, VIX index, GARCH models, Multiplicative error models, Mixture models.

Chapter 1

Introduction

Implied volatility (IV) modeling and forecasting is a topic that has received much less attention in academia compared to volatility modeling and forecasting in general, particularly in light of the immense literature on the GARCH model and its various extensions. Volatility as such is a fundamental issue in financial markets, and it is significant in virtually all types of financial decision-making. Implied volatility is often considered to be the best available forecast for future volatility. Calculated from prevailing market prices for options, implied volatility is regarded as the market's expectation of volatility in the returns of the option's underlying asset during the remaining life of the option. The forward-looking view provided by IV should thus reflect investor sentiment regarding the future.

Harvey and Whaley (1992) describe the benefits of IV by pointing out that when using options to garner a volatility forecast, one does not need to specify a time-series model to link ex-post and ex-ante volatility. IV forecasts can be valuable for, broadly speaking, all market participants. A forecast of future volatility affects derivative pricing, but also all investors with risk management, asset allocation, and portfolio insurance concerns: an expected change in volatility may warrant altering portfolio weights and/or composition. Portfolio managers who are judged against a benchmark can also benefit from volatility forecasts. Fleming (1998) writes that IV can be used as an indicator of market sentiment, to evaluate asset pricing models, and also perhaps to forecast returns. Konstantinidi et al. (2008) also argue that IV and future returns can be linked. They assert that if

IV is predictable, it may contain evidence on how expected asset returns change. Asset price predictability stems from IV predictability as implied volatility is, in essence, a reparameterization of the market price of options.

Trading based on volatility assumes one has a view on the future direction of volatility, or that one feels the market's view (expectation) is not correct. Such a volatility trade requires no opinion on how the price of the option's underlying asset is going to develop in the future. Professional option traders such as hedge funds and proprietary traders in investment banks can potentially profit if their view on the direction of IV is correct, ceteris paribus. At-the-money option prices in particular are sensitive to changes in IV (Bollen and Whaley (2004)), so a trader with a correct view of the direction could enter into a position that stands to gain from the impending change in the market's volatility expectation. Ni et al. (2008) emphasize that if a trader possesses a directional view about the price of the underlying asset, she can trade in either the underlying directly, or then with options. With a volatility view, the trader is restricted to trading in options alone.

This dissertation seeks to fill gaps in existing literature by modeling implied volatility with a variety of time series models and assessing the quality of forecasts calculated from these models. The modeling itself answers questions on what type of time series models fit IV data, giving new insights on how both traditional and more recently developed models can be used in IV modeling. The goal of directional accuracy in forecasting receives support from Harvey and Whaley (1992), who find that IV changes are not unpredictable. This dissertation confirms that finding for three different time series of IV data. As the directional accuracy of the models in all four essays is well above 50% (between 58.4% and 72.2%), it can be argued that the forecasts can potentially be used to trade options profitably.

1.1 Calculating Implied Volatility

Implied volatility is calculated by inverting an option valuation formula when the prevailing market price for an option is known. Volatility is the only ambiguous input into e.g. the Black-Scholes option pricing formula,[1] which is shown below for a call option on a non-dividend paying stock:

$$C = SN(d_1) - Xe^{-rT}N(d_2) \tag{1.1}$$

where

$$d_1 = \frac{ln(S/X)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

C denotes the price of a European call option, S is the market price of the underlying asset, X is the strike price of the option, r is the risk-free interest rate, T is the time to maturity of the option, N is the cumulative normal distribution function, and σ is the volatility in the returns of the underlying asset (in this case, stock) during the life of the option. This last input to the model is thus easily confirmed to be the only one that is not easily observed in the markets. Extensions of the Black-Scholes model have been developed also for e.g. currency and futures options.

Three of the four essays in this dissertation model IV data that is calculated using the Black-Scholes model or one of its extensions. The data on Nikkei 225 index option IV that is used in the second and third essays is derived from the Black-Scholes model. The IV data from USD/EUR exchange rate options that is modeled in the fourth essay is calculated with the Garman-Kohlhagen model[2] for currency options, an extension of the Black-Scholes model. The shortcomings and merits of the Black-Scholes model thus warrant further discussion.

The Black-Scholes option pricing model makes two assumptions that are in contrast with what is observed in actual financial markets. First, the model assumes that the volatility in the price of the underlying asset will remain constant throughout the life of the option. Second, the logarithmic returns of the underlying asset are assumed to follow a normal distribution, whereas financial market returns exhibit both skewness and excess kurtosis. The non-normality of financial

[1] Black and Scholes (1973).
[2] Garman and Kohlhagen (1983).

returns is manifested in the fact that when IVs are calculated from prevailing market prices with the Black-Scholes model, the so-called volatility smile or volatility skew emerges: IVs vary with the strike price of options, even when the options have the same maturity date.

Due to the widely recognized shortcomings of the Black-Scholes model, an abundance of research concentrates on developing option pricing models that could account for stochastic volatility and thus reflect market prices of options more accurately.[3] However, the Black-Scholes model remains largely the market consensus choice for obtaining IVs, as no other model has conclusively gained ground.[4] Computationally and conceptually tractable, the Black-Scholes model is often used particularly in over-the-counter markets to calculate IVs, which are then provided directly as market quotes rather than the option prices themselves. Numerous studies use IVs calculated with the Black-Scholes model or one of its extensions: examples include Day and Lewis (1992), Jorion (1995), Jackwerth (2000), Bollen and Whaley (2004), and Ederington and Guan (2005).

In addition, existing research provides evidence that the Black-Scholes model may not perform at all poorly as an option pricing model and source of IVs. Hull and White (1987) find that Black-Scholes IVs are least biased for at-the-money (ATM),[5] short-term maturity options, meaning that errors in IV estimates stemming from model misspecification can be minimized by using ATM options. Mayhew (1995) also notes that the consensus view is that the Black-Scholes model performs reasonably well for ATM options with no more than one or two months to expiration. As the vast majority of IV literature, including the essays in this dissertation, focuses on ATM options with short times to maturity, these model selection concerns are mitigated. Also, Fleming (1998) maintains that Black-Scholes IV is the rational forecast for mean volatility over the life of an option. He also argues that using an alternative option pricing model involves uncertainty about the chosen volatility process. Even if the process is correctly specified, additional

[3]One of the first, classic examples of such a competing model is that of Hull and White (1987).

[4]Fleming et al. (1995) assert that no commonly accepted alternative model to the Black-Scholes model exists.

[5]At-the-money options are such that $S = X$. In-the-money call (put) options have a strike price that is lower (higher) than the prevailing market price. Out-of-the-money call (put) options have a strike price that is higher (lower) than the prevailing market price.

parameters must be estimated, which makes it harder to recover the true volatility forecast. Jorion (1995) also points out that the use of stochastic volatility models requires the estimation of additional parameters, which introduces an additional potential source of error.

Ederington and Guan (2002) provide evidence that the Black-Scholes model can be correct despite the existence of the volatility smile. The authors find that trading based on the volatility smile can yield substantial profits before transaction costs. In such a strategy, options with high Black-Scholes IVs are sold and options with low IVs bought. There should be zero profits in this strategy if the Black-Scholes model is inaccurate, as the options are not over- or underpriced, but correctly priced. As the returns are non-zero, the options are indeed over- or underpriced, and the Black-Scholes model identifies the mispricing correctly. Ederington and Guan (2002) conclude that a large part of the volatility smile must reflect other forces than wrong distributional assumptions of the Black-Scholes model. These forces could include different demand pressures for calls and puts, and the failure of the no-arbitrage assumption. Hentschel (2003) discusses how measurement errors in option prices can contribute to the existence of the volatility smile. As an example, for out-of-the-money (OTM) options, sensitivity to volatility is small in the sense that large changes in volatility bring about small changes in option prices. Conversely, if option prices are even slightly erroneous, implied volatility estimates can be far off. Measurement errors that contribute to errors in option prices include bid-ask spreads, finite quote precision (tick sizes), and non-synchronous observations (such as stale component prices of indices). Hentschel (2003) maintains that if prices are measured with error, implied volatility can smile even if all the Black-Scholes assumptions hold.

To tackle the problems associated with model selection in acquiring IV estimates, model-free methods of recovering implied volatilities directly from option price quotes have emerged in recent years. Most notably, the VIX Volatility Index of the Chicago Board Options Exchange, which is a measure of the IV of S&P 500 index options, has been calculated with a model-free methodology from September 2003 onwards.[6] The VIX is a staple for active investors around the world, and is

[6]For details on how the VIX is calculated, see www.cboe.com/micro/vix. Prior to September 2003, the Black-Scholes model was used to calculate the VIX.

considered to be a measure of investor sentiment and risk aversion. In the wake of the VIX, numerous other implied volatility indices have switched to the model-free methodology, including the VXN for options on the Nasdaq index, the VDAX for options on the DAX index, and the VSTOXX for options on the Dow Jones Euro Stoxx 50 index. The topic of the first essay of this dissertation is modeling and forecasting the VIX index.

The second, third, and fourth essays in this dissertation model two time series of IV data: one series recovered from call options, and the other from same-strike, same-maturity put options. Both types of options should provide the market's expectation of future volatility, in other words, the exact same value of IV. However, the values of call and put IV differ, which could potentially lead to arbitrage opportunities. Put-call parity is one of the key arbitrage bounds that apply to option prices. If option prices depart from such arbitrage bounds, traders could make riskless profits in the absence of transaction costs. However, the existence of transaction costs and other market imperfections allows option prices to deviate from their no-arbitrage values, which is then reflected in implied volatilities: the IVs recovered from otherwise identical calls and puts should be equal, but in practice, they tend to consistently differ from one another. The demand for puts as portfolio insurance creates demand pressure that is not matched by the demand for equivalent call options. Such demand pressure is one reason why option prices try to pull away from no-arbitrage bounds. Importantly, the gap in call and put IVs provides a valuable building block for the joint modeling of the two time series with bivariate models in the third and fourth essays.

1.2 Brief Review of IV Literature

Implied volatility forecasting in this dissertation is a separate issue from how well IV forecasts future realized volatility. This latter issue is the topic of the majority of IV literature to date. These studies seek to discover whether IV is an unbiased and efficient predictor of future realized volatility. Examples of the above-mentioned literature include Chiras and Manaster (1978), Day and Lewis (1992), Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), Fleming et al. (1995), Jorion (1995), Christensen and Prabhala (1998), Blair et al. (2001), and

Carr and Wu (2006). Broadly speaking, these studies conclude that IV is a biased forecast of future volatility, but it tends to be efficient in the sense that the information contained in IV subsumes all other available information. In other words, augmenting a forecasting model with e.g. a GARCH-based forecast does not improve the forecasts if IV is already a variable in the model. As examples of new lines of work in this area, Bandi and Perron (2006) look at the relation between realized and implied volatility and conclude that the time series are fractionally cointegrated. Giot and Laurent (2007) estimate a model for S&P 500 realized volatility with only lagged values of the VIX index as explanatory variables and conclude that it is nearly as good as a model with jump and continuous components of historical volatility included.

In contrast to the above studies, Harvey and Whaley (1992) argue that since true conditional volatility is not observable, it is impossible to assess the forecast accuracy of any models. Instead, they opt to model and forecast IV itself. Harvey and Whaley (1992) and Brooks and Oozeer (2002) share more or less the same structure as the first and second essays of the dissertation. Both select a time series model for an IV series, calculate IV forecasts, and also run a trading simulation with option price quotes based on the forecasts. The third and fourth essays differ from this basic structure in that they do not include option trading simulations. Harvey and Whaley (1992) model the IV of call and put options on the S&P 100 index, which was more important than the S&P 500 index at the time. Brooks and Oozeer (2002) use European data, or the IV of options on Long Gilt futures, which are traded on the London International Financial Futures Exchange. Both the above-mentioned studies find in their option trading simulations that profits disappear after taking transaction costs into account. The point of view in the trading exercises of the first and second essays is somewhat different, as they do not seek to find whether or not abnormal profits could be made. To that end, transaction costs are not accounted for. The purpose of the trading simulation is to confirm which model is the best forecaster. Also, the option trades that are simulated in the first and second essays are straddles, whereas Harvey and Whaley (1992) and Brooks and Oozeer (2002) simulate buy-or-sell strategies. The straddle is the most effective way to trade options if one has a view on the change in volatility (Bollen and Whaley (2004)).

Other studies also model implied volatility but have a different perspective. Guo (2000) models and forecasts the IV of currency options, but with the aim of comparing IV forecasts to GARCH-based forecasts in an option trading exercise. Mixon (2002) looks at the link between contemporaneous stock returns and changes in IV, but this approach cannot be used for IV forecasting: a forecasting model needs to contain lagged, not contemporaneous, explanatory variables. Bollen and Whaley (2004) model changes in S&P 500 index implied volatility, focusing on how net buying pressure drives changes in IV. Gonçalves and Guidolin (2006) model the S&P 500 implied volatility surface, thus modeling IVs across a multitude of strike prices and maturity dates. Konstantinidi et al. (2008) estimate various time series models for the VIX and other IV indices. However, their directional accuracy falls short of the results received for the VIX in the first essay of the dissertation. Their trading simulation is carried out with VIX futures rather than S&P 500 options; the conclusion is that no abnormal profits can be made with the VIX forecasts.

1.3 Methodology

The first essay of the dissertation uses traditional time series methodology, namely ARIMA and GARCH modeling. The modeling of conditional heteroskedasticity proves to be important for directional forecast accuracy. This result finds support in Christoffersen and Diebold (2006), who show that in the presence of conditional heteroskedasticity, direction can be forecast even in cases where reliable point forecasts cannot be obtained.

The remaining three essays estimate mixture multiplicative error models (MEM models), where the conditional mean μ_t of the IV time series is multiplied by the error term ε_t:

$$IV_t = \mu_t \varepsilon_t \tag{1.2}$$

This contrasts the additive structure of traditional linear models such as ARIMA models. All three essays with MEM models assume that the error term follows the gamma distribution, receiving only non-negative values. MEM models are suited for non-negative time series, and were first suggested by Engle (2002) for modeling volatility. It is not necessary to take logarithms of data when estimating MEM models, as is normally the case with financial time series. The structure of MEM models is the same as in autoregressive conditional duration (ACD) models.[7]

The basic MEM model has turned out to be inadequate for describing the behavior of IV time series. Therefore, the models that are estimated in this dissertation have a mixture structure. The existence of mixture components in the models means that the data is assumed to be drawn from two (or more) regimes. The regime-switching properties of e.g. exchange rate volatility have been widely studied (a relevant example is Bollen et al. (2000)), and Lanne (2006) finds that a two-regime mixture MEM model is a good fit to data on the realized volatility of exchange rates. Engle and Gallo (2006) provide a MEM application for the returns, high-low range, and realized volatility of the S&P 500 index.[8]

The two-regime MEM model specification in the second essay of the dissertation is drawn from Lanne (2006), and this model is extended to a bivariate setting in the third essay. The bivariate model specification is a new model in itself. Cipollini et al. (2006) estimate a multivariate MEM model, but using copula functions. The bivariate model specification in this dissertation uses a bivariate gamma distribution for the error term, which makes the use of copulas unnecessary. A fixed probability for both regimes is estimated in both the univariate and bivariate MEM models. The new ingredient in the fourth essay is a time-varying probability between regimes, which proves necessary for the data set in that study. The time-varying probability depends on a regime indicator, with the choice falling on the returns of the underlying asset. The regime thresholds are not strict; in other words, there is a range around the threshold value within which the regime switch occurs. The width of the range is estimated as one of the parameters of the model.

[7]ACD models were introduced by Engle and Russell (1998) and are used to model e.g. the duration between trades.

[8]The MEM model is not a good fit for the VIX index, however.

1.4 Summaries of Essays

1.4.1 Modeling and Forecasting the VIX Index

The VIX index is followed closely by financial market practitioners as well as retail investors. It is considered to be an indicator of market sentiment and risk aversion. Based on the market prices of options on the S&P 500 index, the VIX is also the market's expected volatility in the returns of the S&P 500 index over the next thirty days. The S&P 500 index, in turn, is the most important benchmark for U.S. equity markets.

The VIX has been the topic of plenty of academic research over the past years. This study extends the existing literature by estimating time series models for the VIX, and calculating forecasts from these models. This contrasts most existing studies, as they attempt to evaluate whether or not the VIX is an unbiased predictor of future realized volatility. The current methodology for calculating the VIX is model-free, which helps to avoid issues with option pricing model misspecification.

The models are built for log VIX first differences. An ARIMA(1,1,1) specification is found to be the best fit. The VIX also displays volatility persistence, or conditional heteroskedasticity. As a result, the parameters in a GARCH model for the error terms are statistically significant. The significance of a number of financial and macroeconomic explanatory variables is investigated, but only the returns of the S&P 500 index prove significant. For example, S&P 500 trading volume and the returns of a stock index covering key non-U.S. markets are not significant explanatory variables for changes in the VIX. Weekly seasonality is important in modeling the VIX, as both Monday and Friday dummy variables are significant. The VIX tends to rise on Mondays and fall on Fridays.

Two in-sample periods are used in the study. Both end on 31.12.2002, allowing for a five-year out-of-sample period from 1.1.2003 to 31.12.2007. The longer (shorter) in-sample period uses 3,279 (1,000) observations for parameter estimation. It turns out that the directional accuracy of the VIX forecasts improves when using the shorter sample period. The ARIMA-GARCH specification without the underlying index returns as an explanatory variable is the superior directional forecaster, with the correct direction of change forecasted on 58.4% of days in the out-of-sample period. Mean squared errors are slightly lower when the models are estimated with the longer in-sample, but the differences are not statistically significant.

A straddle trading simulation with S&P 500 index options is run in order to confirm whether or not the ARIMA-GARCH model, estimated with 1,000 observations, is indeed the forecasting model of choice. The goal of the trading exercise is not to uncover if the forecasts could lead to abnormal profits; the results are simply used to rank the models. When leaving out four outlier days that could make the results noisy, the ARIMAX-GARCH model (which includes S&P 500 returns as an explanatory variable) leads to the highest returns, followed by the ARIMA-GARCH model. Both these models are estimated with 1,000 observations. All in all, the evidence is thus somewhat mixed regarding the significance of including S&P 500 returns in the model, but the modeling of conditional heteroskedasticity is important.

1.4.2 Multiplicative Models for Implied Volatility

A new class of models, so-called multiplicative error models (MEM models), have been used successfully in recent years to model volatility. MEM models were first suggested by Engle (2002) for modeling financial time series. Due to the way they are set up, multiplicative models can be used for time series that always receive non-negative values, such as trading volume or volatility. MEM models differ from linear regression models in that the mean equation is multiplied with the error term.

In this particular study, a mixture multiplicative error model (MMEM) similar to that in Lanne (2006) has been estimated. The model has two mean equations and error distributions, with the error terms coming from gamma distributions. The time-varying conditional mean and possibility for a mixture of two gamma distributions can help model the fact that in financial time series, periods of business-as-usual alternate with periods of large shocks, which can be captured by the second mixture components of the model. A probability parameter dictates which state the model is in.

Daily data on the implied volatility (IV) of options on the Nikkei 225 index was obtained for both Nikkei 225 index call and put options for the time period 1.1.1992 - 31.12.2004. The use of separate time series of IV from calls (NIKC) and puts (NIKP) can offer new insights into the analysis, and e.g. benefit investors wishing to trade in only either call or put options. The in-sample period used in model estimation extends from 2.1.1992 to 30.12.2002. Statistical significance leads to a final choice of a MMEM(1,2;1,1) specification. In practice, the business-as-usual equation has two lags of implied volatility and one lag of the conditional mean, whereas the second mean equation contains one lag of each.

With financial market data, it may be that only the most recent history is relevant in modeling and forecasting, so the MMEM(1,2;1,1) model was also estimated for NIKC and NIKP using only the past 500 observations. In order to investigate the necessity of the mixture components of the model, a MEM(1,2) specification (i.e., a model with only one mean equation and error distribution) was also estimated. Coefficients are not statistically significant in this latter case.

The last 486 trading days of the full sample were left as an out-of-sample period. Forecast performance is evaluated primarily with directional accuracy, and secondly with mean squared error. In general, it appears to be somewhat easier to forecast NIKC than NIKP. At best, the MMEM(1,2;1,1) model forecasts the direction of change correctly on 69.1% of trading days for NIKC and on 66.0% of trading days for NIKP. For e.g. option traders, any level of accuracy over 50% can potentially be worth money. Also, comparing with the findings of the first essay of the dissertation for the VIX index, where the sign was predicted accurately on 58% of trading days at best, the directional accuracy is clearly better for the Nikkei 225 implied volatility.

When incorporating only the most recent information, or estimating the model with 500 observations, the forecast performance deteriorates considerably. The forecast performance of the one-regime MEM(1,2) model also falls short of that of the MMEM(1,2;1,1) estimated from the entire in-sample. Mean squared errors are lowest when using only 500 observations to estimate the model.

Mean squared errors from forecasts five days ahead into the future indicate that the chosen MMEM(1,2;1,1) specification fares best. In an option trading simulation with actual market prices of Nikkei 225 index options, trading straddles based on the forecasts from the MMEM(1,2;1,1) model also leads to the highest returns.

Therefore, the evidence leads to a recommendation of using the MMEM(1,2;1,1) model in order to forecast the implied volatility in Japanese option markets.

1.4.3 Joint Modeling of Call and Put Implied Volatility

This study introduces a bivariate mixture multiplicative error model (BVMEM). The model has two mean equations and error distributions, with the error terms coming from a bivariate gamma distribution. The two-regime structure is useful in financial market applications as financial time series have often been found to exhibit regime-switching properties. A probability parameter that indicates how much time is spent in each regime is estimated as one of the parameters in the model.

The data set consists of daily observations of Black-Scholes implied volatilities (IV) of options on the Nikkei 225 index for the time period 1.1.1992 - 31.12.2004. The joint modeling of separate time series of IV from calls (NIKC) and puts (NIKP) can be beneficial, as call-side IV is likely to affect put-side IV, and vice versa. These cross effects are captured by using lagged call (put) IV as an explanatory variable for put (call) IV. Evidence from prior literature indicates that trading in put options can potentially have a particularly strong effect: demand for index puts is often quite large, as institutional investors buy puts as portfolio insurance. These demand effects can also help explain why the IV calculated from calls and puts differs in the first place: limits to arbitrage let imbalances in demand and supply affect prices so that they depart from their no-arbitrage bounds.

The eleven-year in-sample period used in the model estimation covers the time period 2.1.1992 - 30.12.2002. The final model specification includes cross effects between NIKC and NIKP in both regimes, as well as slope dummy variables for put IV on Fridays. The weekly pattern of the IV data is that IV is highest on Mondays and lowest on Fridays. Error terms are much more dispersed around their mean of unity in the second regime. Diagnostic checks with probability integral transforms, as well as a simulation of autocorrelations, indicate that the model is a good fit to the data.

The paper includes an impulse response analysis that shows that the effect of shocks to implied volatility can last for up to six weeks, or thirty trading days.

Interestingly, put-side IV recovers much more quickly from shocks, even if only NIKP receives the shock - therefore, trading in puts may be more efficient. The point in time when the shock is introduced does not affect the impulse responses (i.e., whether it is a time of low or high IV).

The last two years, or 486 trading days, of the data set were used as an out-of-sample period for forecast evaluation. Forecasts were calculated from the chosen bivariate model specification as well as from a model with no cross effects. Forecasts were obtained with updating and fixed coefficients, so there are a total of four competing forecast series. Forecast performance is evaluated with both directional accuracy and with mean squared errors. It turns out to be slightly easier to forecast NIKP than NIKC. At best, the BVMEM model forecasts the direction of change correctly on 72.2% of trading days for NIKP and on 71.6% of trading days for NIKC. Both results are superior to those achieved with equivalent univariate models. Mean squared errors are lower for NIKC.

All in all, the model with cross effects and updating coefficients fares best as a forecasting model. More importantly, these results improve upon those in the second essay, which contains similar models but in a univariate setting. Therefore, there is clear added value to modeling call and put implied volatility jointly.

1.4.4 Implied Volatility with Time-Varying Regime Probabilities

Numerous studies document that exchange rates appear to exhibit regime-switching tendencies. For example, Bollen et al. (2000) estimate a two-regime Markov-switching model for log exchange rate changes, with both mean and variance regimes switching independently of each other. Their results show that the volatility of exchange rate returns is two to three times higher in the more volatile regime. In light of this evidence, it can be hypothesized that the implied volatility of foreign exchange options could also be drawn from two or more regimes.

This study models the at-the-money implied volatilities of call and put options on the USD/EUR exchange rate with two and three-regime multiplicative error models. A model with a fixed probability between regimes (similar to the models in the second and third essays) is not sufficient for this data set. Therefore, a time-varying property is added to the model so that the regime probabilities are estimated for each data point separately. In other words, with daily observations, regime probabilities are determined separately for each day in the data set.

Daily returns in the USD/EUR exchange rate are used as a regime indicator. The assumption behind this choice is that large moves in the underlying asset of the options will correspond to days that fall into the more volatile regime. The performance of this regime indicator is confirmed by comparing the time series of probabilities to the time series of IV data. This comparison indicates that the regime probabilities do in fact react when IV experiences large moves, or is at a relatively high level.

Additional flexibility is introduced into the model by allowing the regime switch to occur within a certain range rather than at a fixed threshold value. A threshold value of the exchange rate return is estimated, but an additional volatility parameter provides the width of the range within which a regime shift occurs.

Both univariate and bivariate MEM models are estimated for the call and put IV data. As there is no test for the number of regimes, several alternatives were evaluated and statistical significance and diagnostic tests used to select viable specifications. Both two and three-regime models are viable in the univariate setting, whereas two regimes are sufficient for the bivariate model.

As in the third essay, the bivariate model proves to be a superior forecaster when compared to the univariate models. Both the two and three-regime univariate forecasts have weaker directional accuracy and mean squared errors than the bivariate forecasts. Also in line with the results of the third essay, the implied volatility of USD/EUR puts is more forecastable than the IV of the equivalent calls.

Bibliography

[1] Bandi, F.M. & Perron, B. (2006). Long Memory and the Relation between Implied and Realized Volatility. Journal of Financial Econometrics, 4, 636–670.

[2] Black, F. & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81, 637–654.

[3] 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.

[4] Bollen, N.P.B., Gray, S.F. & Whaley, R.E. (2000). Regime Switching in Foreign Exchange Rates: Evidence from Currency Option Prices. Journal of Econometrics, 94, 239–276.

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

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

[7] Canina, L. & Figlewski, S. (1993). The Informational Content of Implied Volatility. Review of Financial Studies, 6, 659–681.

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

[9] Chiras, D.P. & Manaster, S. (1978). The Information Content of Option Prices and a Test of Market Efficiency. Journal of Financial Economics, 6, 213–234.

[10] Christensen, B.J. & Prabhala, N.R. (1998). The Relation between Implied and Realized Volatility. Journal of Financial Economics, 50, 125–150.

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

[12] Cipollini, F., Engle, R.F. & Gallo, G.M. (2006). Vector Multiplicative Error Models: Representation and Inference. NBER Working Paper, 12690.

[13] Day, T.E. & Lewis, C.M. (1992). Stock Market Volatility and the Information Content of Stock Index Options. Journal of Econometrics, 52, 267–287.

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

[15] Ederington, L. & Guan, W. (2005). The Information Frown in Option Prices. Journal of Banking & Finance, 29, 1429-1457.

[16] Engle, R.F. (2002). New Frontiers for ARCH Models. Journal of Applied Econometrics, 17, 425–446.

[17] Engle, R.F. & Gallo, G.M. (2006). A Multiple Indicators Model for Volatility Using Intra-daily Data. Journal of Econometrics, 131, 3–27.

[18] Engle, R.F. & Russell, J.R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica, 66, 1127-1162.

[19] Fleming, J. (1998). The Quality of Market Volatility Forecasts Implied by S&P 100 Index Option Prices. Journal of Empirical Finance, 5, 317–345.

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

[21] Garman, M.B. & Kohlhagen, S.W. (1983). Foreign Currency Option Values. Journal of International Money and Finance, 2, 231–237.

[22] 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.

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

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

[25] 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.

[26] Hentschel, L. (2003). Errors in Implied Volatility Estimation. Journal of Financial and Quantitative Analysis, 38, 779–810.

[27] Hull, J. & White, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. Journal of Finance, 42, 281–300.

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

[29] Jorion, P. (1995). Predicting Volatility in the Foreign Exchange Market. Journal of Finance, 50, 507–528.

[30] 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, 32, 2401–2411.

[31] Lamoureux, C.G. & Lastrapes, W.D. (1993). Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities. Review of Financial Studies, 6, 293–326.

[32] Lanne, M. (2006). A Mixture Multiplicative Error Model for Realized Volatility. Journal of Financial Econometrics, 4, 594–616.

[33] Mayhew, S. (1995). Implied Volatility. Financial Analysts Journal, 51, 8–20.

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

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

Chapter 2

Modeling and Forecasting the VIX Index

Katja Ahoniemi[1]

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.

[1]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.

2.1 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.

Relatively little work has been done on modeling IV itself, compared with the extensive 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]

[2]There are other examples of IV modeling from markets other than the stock market. Using

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

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.

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 (decrease).

Due to its popularity, the VIX has been the topic of numerous articles that differ somewhat in perspective from the present study.[3] 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. Giot and Laurent (2007) 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.

[3]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.

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.[4] 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 2.2 describes the data used in this study, including various financial and macroeconomic variables that could have explanatory power over the VIX index. Section 2.3 presents the models estimated for the VIX time series and discusses their goodness of fit. Section 2.4 contains the analysis of forecasts, with forecast evaluation based on directional accuracy and mean squared errors. Section 2.5 contains the option trading simulation. Section 2.6 concludes the paper.

[4]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.

2.2 Data

2.2.1 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.[5] 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.

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.[6] Values for the VIX with the new methodology are available from the CBOE from 1.1.1990.

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

[6] 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.

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.

2.2.2 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. 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.[7] 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. First differences of the S&P 500 trading volume, the short-term interest rate, and the slope of the yield curve are used, as p-values from the Augmented Dickey-Fuller (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.

2.3 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 2.1 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 2.2.

The VIX is a very persistent time series, as its daily levels display high autocorrelation (shown in Figure 2.3). The autocorrelations for the differenced time

[7] 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.

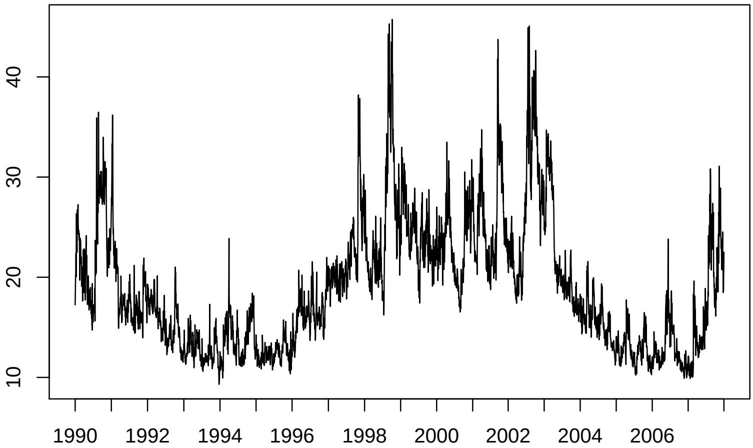

Figure 2.1: VIX index 1.1.1990 - 31.12.2007

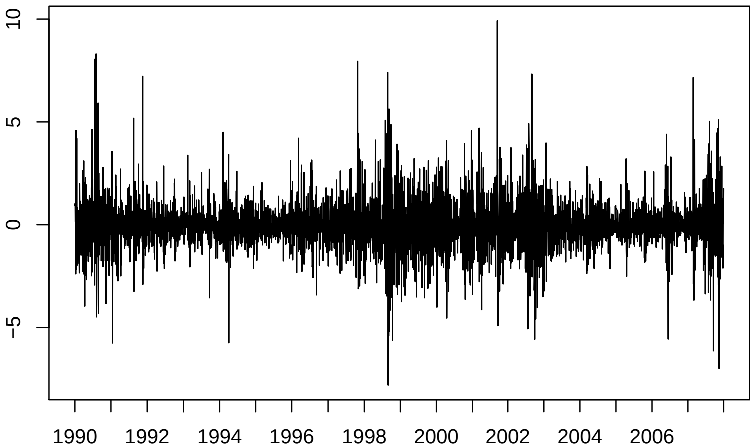

Figure 2.2: VIX first differences 1.1.1990 - 31.12.2007

series are also provided in Figure 2.3. A unit root is rejected by the ADF test for both the level and differenced time series at the one-percent level of significance.[8]

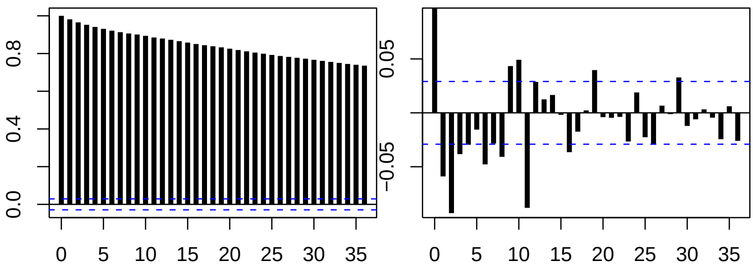

Figure 2.3: Autocorrelations for log VIX (above) and log VIX first differences (below)

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.[9] 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 2.1. The largest day-to-day changes in the VIX are quite large, a gain of

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

[9]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.

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 2.1: 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 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 2.2, 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.[10] In modeling realized volatility, Corsi et al. (2008) find that accounting for the volatility of realized volatility improves forecast performance. In light of the probable existence of GARCH effects, heteroskedasticity-robust standard errors are used in all model estimation. The financial and macroeconomic indicators outlined in Section 2.2.2 were also included in the regressions to see whether they could improve ARIMA models of the VIX. The estimated linear equation is

$$\Delta VIX_t = \omega + \phi_1 \Delta VIX_{t-1} + \theta_1 \epsilon_{t-1} + \sum_{i=1}^{r} \phi_i \mathbf{X}_{i,t-1} + \sum_{k=1}^{5} \gamma_k D_{k,t} + \epsilon_t \qquad (2.1)$$

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 2.1 with a model for the conditional variance, a GARCH(1,1) specification is found to be sufficient and is parameterized as follows

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

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

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

The estimation results for four alternative model specifications and the two in-sample periods are presented in Table 2.2. 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 (2009)), 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 S&P 500 index.[11] 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 2.4.

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 α_1 and β_1 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

[11] 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 2.2: 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.

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 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 2.4 and 2.5 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 α_1 and β_1 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 2.4 and the conditional variances for the ARIMA-GARCH model are in Figure 2.5. 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.

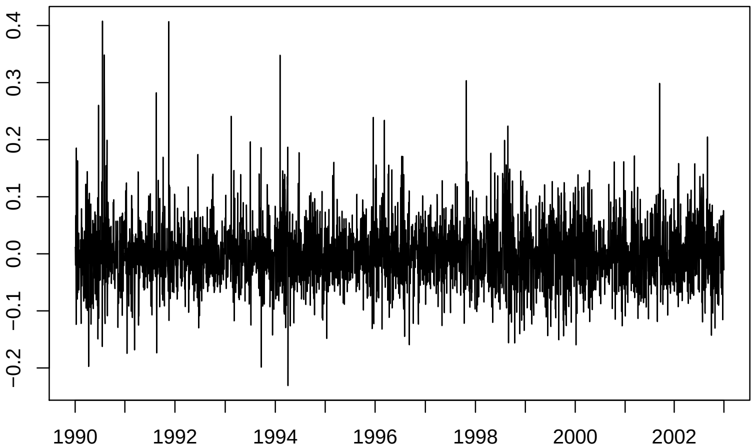

Figure 2.4: Residuals for ARIMA-GARCH model estimated with entire in-sample (3,279 observations).

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 2.3. 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

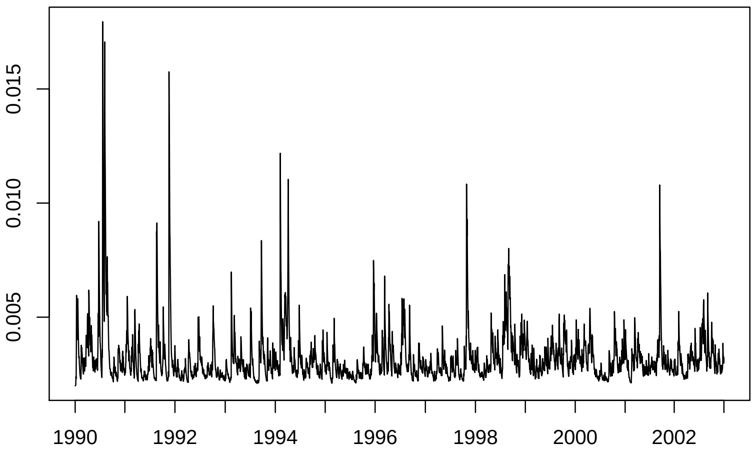

Figure 2.5: Conditional variances for ARIMA-GARCH model estimated with entire in-sample (3,279 observations).

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.

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

2.4 Forecasts

The next step in the analysis is to obtain forecasts from the four model specifications described in Section 2.3. 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 2.4 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.

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

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.

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.

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 similar values of the obtained MSEs raise the question of whether or not the differences in MSEs are statistically significant.

To evaluate the significance of the mean squared error ranking, the test for superior predictive ability (SPA) developed by Hansen (2005) was employed. The SPA test of Hansen allows for the simultaneous comparison of m series of forecasts, in contrast to e.g. the forecast accuracy test of Diebold and Mariano (1995), which evaluates two series at a time. In the SPA test, one series of forecasts is deemed to be the benchmark, and its MSE is compared to those of the competing forecast series. In this case, the benchmark is the ARIMA model estimated with 3,279 observations. The consistent p-value of the test, which is calculated with 1,000 bootstrap resamples, is 0.992. As the null hypothesis of the test is that the benchmark is not inferior to the alternative forecasts, it can thus not be rejected. However, the consistent p-value when using the model with the highest MSE, or the ARIMA-GARCH model estimated with 1,000 observations, as the benchmark is 0.149, which is also significant. 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.

2.5 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 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 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 2.2, and the return from a short straddle is shown in Equation 2.3. In this analysis, the proceeds from selling a straddle are not invested during the day, but held with zero interest.

$$R_l = \frac{100}{C_t + P_t}(-C_t - P_t + C_{t+1} + P_{t+1}) \tag{2.2}$$

[12]see Noh et al. (1994)

$$R_s = \frac{100}{C_t + P_t}(C_t + P_t - C_{t+1} - P_{t+1}) \quad (2.3)$$

In Equations 2.2 and 2.3, 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 2.5. 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 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 four 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

[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.

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

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 2.5: 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%, respectively.

Turning now to the returns themselves, the model choice from Section 2.4 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 2.4, 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 when using the filters.

In light of the forecast evaluation in Section 2.4 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.

2.6 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.

Bibliography

[1] Ahoniemi, K. & Lanne, M. (2009). Joint Modeling of Call and Put Implied Volatility. International Journal of Forecasting, forthcoming.

[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] Hansen, P.R. (2005). A Test for Superior Predictive Ability. Journal of Business & Economic Statistics, 23, 365-380.

[31] 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.

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

[33] 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.

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

[35] 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, 32, 2401–2411.

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

[37] 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.

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

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

[40] 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.

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

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

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

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

Chapter 3

Multiplicative Models for Implied Volatility

Katja Ahoniemi[1]

Abstract

This paper estimates a mixture multiplicative error model for the implied volatilities of call and put options on the Nikkei 225 index. Diagnostics show that a two-regime mixture multiplicative model is a good fit to the data. In an out-of-sample of two years, mixture multiplicative models correctly predict the direction of change in implied volatility on close to 70 percent of trading days at best. Forecast evaluation shows that a mixture model with two regimes outperforms a multiplicative model with no mixture components. An option trading simulation with Nikkei 225 index options also points to the predictive superiority of the mixture multiplicative model over a one-regime model.

[1]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 XII Spring Meeting of Young Economists in 2007 for valuable comments. Financial support from the Finnish Doctoral Programme in Economics, the Finnish Foundation for Advancement of Securities Markets, the Jenny and Antti Wihuri Foundation, and the Yrjö Jahnsson Foundation is gratefully acknowledged.

3.1 Introduction

Reliable volatility forecasts can greatly benefit professional option traders, market makers who need to price derivatives, and all investors with risk management concerns. Implied volatilities, which can be garnered from option markets, can be particularly useful in such contexts as they are forward-looking measures of the market's expected volatility during the remaining life of an option. A correct view of the direction of change in implied volatility can facilitate entering into profitable positions in option markets, and an expected change in the level of market volatility may lead to a need to change portfolio weights or composition.

The bulk of implied volatility research investigates how well implied volatility forecasts the volatility that is realized in the returns of an option's underlying asset during the remaining life of the option.[2] Implied volatility has traditionally been modeled with linear regression models (e.g. Harvey & Whaley (1992) and Brooks & Oozeer (2002)), or with linear models augmented with GARCH errors (Ahoniemi (2008)). However, a new class of models, so-called multiplicative models, have been used successfully in recent years to model volatility.

Engle & Gallo (2006), using data on the S&P 500 index, estimate a system of multiplicative error models for squared log returns, the square of the high-low price range, and realized volatility. They conclude that forecasts from the multiplicative specification have significant explanatory power in modeling the value of the VIX index. Lanne (2006) builds a mixture multiplicative error model for the realized volatility of the Deutsche Mark and Japanese Yen against the U.S. dollar. He finds that the in-sample fit of the model is superior to that of ARFIMA models, and forecasts outperform those from several competing models, including ARFIMA and GARCH models. Lanne (2007) uses the same model specification to forecast the realized volatility of exchange rates by decomposition. Multiplicative models are similar in structure to autoregressive conditional duration (ACD) models, which were introduced by Engle & Russell (1998) and have since led to an abundance of research.[3] So far, multiplicative modeling has not been applied to implied

[2] Examples include Day & Lewis (1992), Canina & Figlewski (1993), Jorion (1995), Christensen & Prabhala (1998), and Blair et al. (2001).

[3] See e.g. Bauwens & Giot (2003), Ghysels et al. (2004), Manganelli (2005), Fernandes & Grammig (2006) and Meitz & Teräsvirta (2006).

volatility.

This paper models the implied volatility (IV) time series of call and put options on the Nikkei 225 index with a mixture multiplicative model similar to that in Lanne (2006). The model contains two mean equations and two error distributions, allowing days of large shocks to be modeled separately from more average trading days. The model specification is a good fit to both the call and put IV time series, and produces forecasts with directional accuracy of up to 69.1% in a two-year out-of-sample period. These forecasts outperform those obtained with a one-regime model, so the two mixture components are indeed necessary.

Further analysis also points to the choice of a two-regime multiplicative specification for both call and put IV. When forecasting five trading days ahead, multiplicative models with two regimes again outperform one-regime models. An option trading simulation using market quotes of Nikkei 225 index options reveals that the best out-of-sample returns are yielded when trading based on the forecasts of mixture multiplicative models. The purpose of the option trading exercise, which is carried out by simulating straddle trades, is not to determine whether or not the forecasts can generate abnormal profits. The results are used simply to rank the forecast models.

This paper is structured as follows. Section 3.2 presents the mixture multiplicative error model. Section 3.3 describes the data used in the study, the estimation results, and diagnostics. Section 3.4 analyzes the forecasts from various competing models, with forecast evaluation based on directional accuracy and mean squared errors. Section 3.5 describes the returns from the option trading simulation. Section 3.6 concludes the paper.

3.2 The mixture-MEM model

Multiplicative error models (MEM) were first suggested by Engle (2002) for modeling financial time series. Due to the way they are set up, multiplicative models can be used for time series that always receive non-negative values, such as the time interval between trades, the bid-ask spread, trading volume, or volatility. In traditional regression models, logarithms are normally taken from time series data in order to avoid negative forecasts, but this is not necessary with MEM models.

MEM models differ from traditional, linear regression models in that the mean equation μ_t is multiplied with the error term ε_t:

$$y_t = \mu_t \varepsilon_t \tag{3.1}$$

Shocks are assumed to be independent and identically distributed (*iid*), to have mean unity, and come from a non-negative distribution. In this particular study, a mixture multiplicative error model (MMEM) similar to that in Lanne (2006) is estimated. In such a specification, there are two possible mean equations:

$$\mu_{1t} = \omega_1 + \sum_{i=1}^{q_1} \alpha_{1i} y_{t-i} + \sum_{j=1}^{p_1} \beta_{1j} \mu_{1,t-j} \tag{3.2}$$

$$\mu_{2t} = \omega_2 + \sum_{i=1}^{q_2} \alpha_{2i} y_{t-i} + \sum_{j=1}^{p_2} \beta_{2j} \mu_{2,t-j} \tag{3.3}$$

Therefore, μ_t depends on q past observations of implied volatility and p past expected implied volatilities, and the model is denoted MMEM$(p_1, q_1; p_2, q_2)$. This autoregressive form for the mean equations can help to capture possible clustering in the data. Clustering is often present in financial time series, meaning that high (or low) volatility can persist. The mixture specification is also extended into the error term, with the error terms assumed to come from two gamma distributions with possibly different shape and scale parameters. In other words, μ_{1t} is paired with ε_{1t}, and μ_{2t} is paired with ε_{2t}. Engle (2002) suggested the exponential distribution for the error term, but the gamma distribution is more general, as it nests e.g. the exponential distribution and the χ^2 distribution.

The time-varying conditional mean and possibility for a mixture of two gamma distributions bring considerable flexibility into the model. These elements can help model the fact that in financial time series, periods of business-as-usual alternate with periods of large shocks, which can be captured by the second regime of the model. The probability parameter π $(0 < \pi < 1)$ dictates which state the model is in, i.e. the conditional mean is μ_{1t} and errors ε_{1t} with probability π, and the conditional mean is μ_{2t} and errors ε_{2t} with probability $(1 - \pi)$. There could of course be even more mixture components in the model than just two, but the diagnostics presented in Section 3.3 indicate that two components is sufficient for this application. Unfortunately, there is no way to directly test for the number of necessary regimes.

The conditional mean equations reveal that MEM (and ACD) models are similar in structure to GARCH models, meaning that parameter constraints that apply to GARCH models also apply to MEM models. The estimated coefficients must therefore satisfy the positivity constraints outlined in Nelson & Cao (1992) for GARCH models. For both mixture components of a MMEM(1,2;1,2) model, the constraints are:

$$\omega_i \geq 0$$

$$\alpha_{i1} \geq 0$$

$$0 \leq \beta_i < 1$$

$$\beta_1 \alpha_{i1} + \alpha_{i2} \geq 0$$

with $i = 1, 2$. Therefore, in contrast to a (1,1) model, not all parameters need to be non-negative.

The shape and scale parameters of the gamma distributions are constrained so that with $\varepsilon_{1t} \sim Gamma(\gamma_1, \delta_1)$, $\delta_1 = 1/\gamma_1$ and with $\varepsilon_{2t} \sim Gamma(\gamma_2, \delta_2)$, $\delta_2 = 1/\gamma_2$, or so that the scale parameter is the inverse of the shape parameter. This ensures that the error term will have mean unity. Under these assumptions, the conditional distribution of y_t is:

$$\begin{aligned} f_{t-1}(y_t;\theta) \;=\; & \pi \frac{1}{\mu_{1t}\Gamma(\gamma_1)\delta_1^{\gamma_1}} \left(\frac{y_t}{\mu_{1t}}\right)^{\gamma_1 - 1} \exp\left(-\frac{y_t}{\delta_1 \mu_{1t}}\right) \;+ \\ & (1-\pi)\frac{1}{\mu_{2t}\Gamma(\gamma_2)\delta_2^{\gamma_2}} \left(\frac{y_t}{\mu_{2t}}\right)^{\gamma_2 - 1} \exp\left(-\frac{y_t}{\delta_2 \mu_{2t}}\right) \end{aligned} \tag{3.4}$$

where θ is the parameter vector and $\Gamma(\cdot)$ is the gamma function. The model can be estimated with conditional maximum likelihood (ML), and the log-likelihood function can then be written as:

$$\ell(\theta) = \sum_{t=1}^{T} \ln[f_{t-1}(y_t)] \tag{3.5}$$

3.3 Estimation

3.3.1 Data

The underlying asset for the option data used in this study is the Nikkei 225 index, which is a price-weighted average of 225 Japanese companies listed on the Tokyo Stock Exchange and likely to be the most closely followed stock index in Asian markets. The Nikkei 225 reached its all-time high in December 1989, topping 38,900 at the time. In the sample used in this study, the index value ranges from 7,608 to 23,801.

Data on the implied volatility of options on the Nikkei 225 index was obtained from the Bloomberg Professional Service for both Nikkei 225 index call and put options for the time period 1.1.1992 - 31.12.2004. The graphs of the time series of Nikkei 225 call and put IV are shown in Figure 3.1. The use of separate time series of IV from calls and puts can offer new insights into the analysis, and e.g. benefit investors wishing to trade in only either call or put options. Harvey and Whaley (1992) also choose the approach of modeling call and put IV separately and conclude, as does this study, that call-side IV is more predictable.

Although the implied volatility calculated from call and put options with equal expiration dates and strike prices should theoretically be the same, as both reflect the same market expectation, the empirical finding is that the two almost always differ slightly. This may be a reflection of differing demand and supply balances in the markets for calls and puts: for example, Bollen & Whaley (2004) note that in the S&P 500 index option market, puts account for 55% of trades, with institutional investors buying puts as portfolio insurance. They also find that the level of at-the-money implied volatility is largely driven by the demand for at-the-money puts. In the data set used in this study, the time series for put-side IV reacts particularly strongly on 9/11, which is a logical reflection of the plummet in stock prices and the ensuing panic selling that took place at the time. This high market uncertainty would have raised the demand for put options more than the

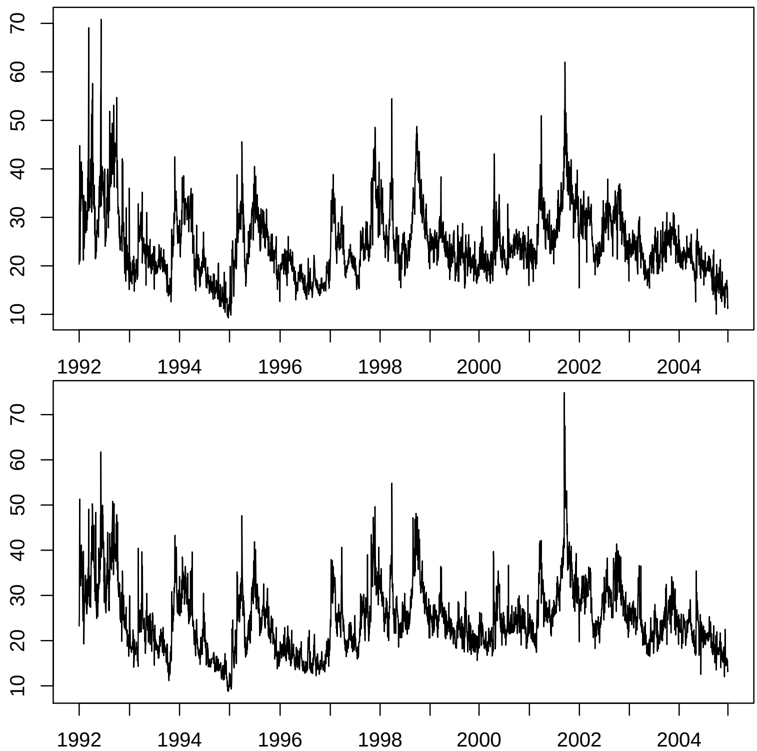

Figure 3.1: Nikkei 225 index call implied volatility (upper panel) and put implied volatility (lower panel) 1.1.1992 - 31.12.2004.

demand for call options. Differing demand and supply conditions and limits to arbitrage, as described by e.g. Garleanu et al. (2006), Liu and Longstaff (2004), and Figlewski (1989), can together explain how call and put IVs can differ without there being a profitable arbitrage opportunity in the market.

The IV time series are calculated daily as the unweighted average of the Black-Scholes implied volatilities of two near-term nearest-to-the-money options. Near-term options tend to be most liquid, and therefore have the most accurate prices. Options on the Nikkei 225 index are available with maturity dates for every month. Days when public holidays fall on weekdays, or when there was no change in the value of call or put implied volatility, were omitted from the data set. After this modification, the full sample contains 3,194 observations.

Descriptive statistics for the Nikkei 225 call (NIKC) and put (NIKP) implied volatility time series are given in Table 3.1. The IV of puts has been slightly more volatile during the time period in question. The range of values that NIKC and NIKP receive during the time period under study is remarkably wide: NIKC varied from 9.3 to 70.8, and NIKP from 8.8 to 74.9. Both series are skewed to the right and they display excess kurtosis. The autocorrelations for NIKC and NIKP are displayed in Figure 3.2, revealing the relatively high degree of persistence in the data. However, a unit root is rejected by the Augmented Dickey-Fuller test for both NIKC and NIKP at the one-percent level of significance. Lanne (2006) finds that with such slowly decaying autocorrelations, a multiplicative model with two mixture components can fit the data quite well.

	NIKC	NIKP
Maximum	70.84	74.87
Minimum	9.26	8.80
Mean	24.68	24.82
Median	23.42	23.84
Standard deviation	7.07	7.41
Skewness	1.10	0.94
Kurtosis	5.42	4.79

Table 3.1: Descriptive statistics for NIKC and NIKP for the full sample of 1.1.1992 - 31.12.2004.

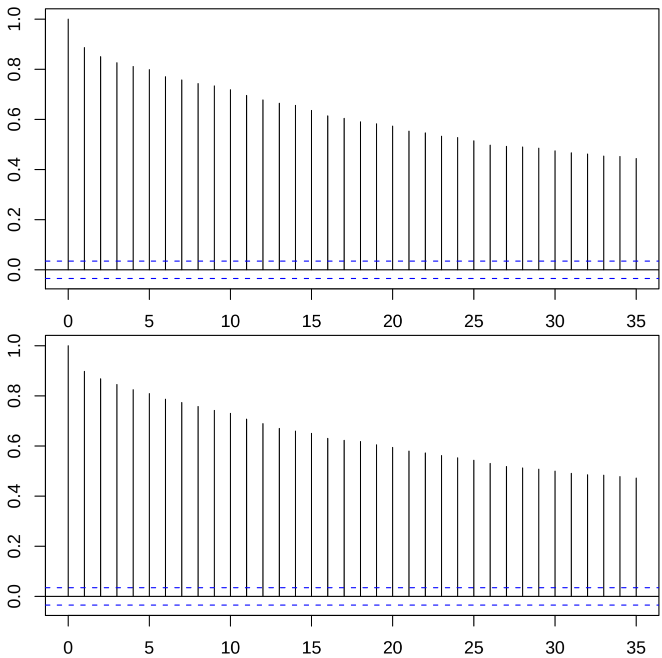

Figure 3.2: Autocorrelations for NIKC (upper panel) and NIKP (lower panel). The dashed lines mark the 95% confidence interval.

3.3.2 Model estimation

The in-sample period used in model estimation covers 2,708 observations from 2.1.1992 to 30.12.2002. The base case in the estimation was the MMEM(1,2;1,2) model, which was found to be the best specification for exchange rate realized volatility time series by Lanne (2006). However, the coefficient for second lags of IV in the more volatile regime (α_{22}) is not statistically significant for NIKC or NIKP, so the final specification choice is a (1,2;1,1) model for both time series.[4] Table 3.2 presents the coefficients and log-likelihoods of the estimated models. The estimated coefficients satisfy all the Nelson & Cao (1992) constraints.

	NIKC		NIKP	
Log likelihood	-6422.85		-6385.10	
π	0.822	(0.000)	0.940	(0.000)
γ_1	145.495	(0.000)	113.300	(0.000)
ω_1	0.261	(0.002)	0.292	(0.000)
α_{11}	0.638	(0.000)	0.571	(0.000)
α_{12}	-0.261	(0.000)	-0.177	(0.000)
β_1	0.610	(0.000)	0.590	(0.000)
γ_2	26.541	(0.000)	19.346	(0.000)
ω_2	0.717	(0.060)	2.228	(0.148)
α_{21}	0.324	(0.000)	0.587	(0.005)
β_2	0.657	(0.000)	0.409	(0.001)

Table 3.2: Estimation results for the MMEM(1,2;1,1) model for NIKC and NIKP. P-values for the statistical significance of the coefficients are given in parentheses.

The incidence of the business-as-usual regime is clearly higher for put-side implied volatility than for call-side IV, as the probability parameter receives the value of 0.94 for NIKP and 0.82 for NIKC. The constant terms of the model are higher in the second regime, which indicates that this regime most likely models days with larger shocks. The persistence of the model, reflected in the parameters

[4]The p-values from a likelihood ratio test for the restriction of $\alpha_{22} = 0$ are 0.81 for NIKC and 0.53 for NIKP.

β_1 and β_2, is almost equal in the first regime for both NIKC and NIKP. NIKC is somewhat more persistent in the second regime.

Figure 3.3 shows the estimated densities of the error terms from the MMEM(1,2;1,1) model for NIKC and NIKP. The densities for the more common, business-as-usual component of the model are more concentrated around unity, and the densities for the second mixture component are more dispersed and skewed to the right.

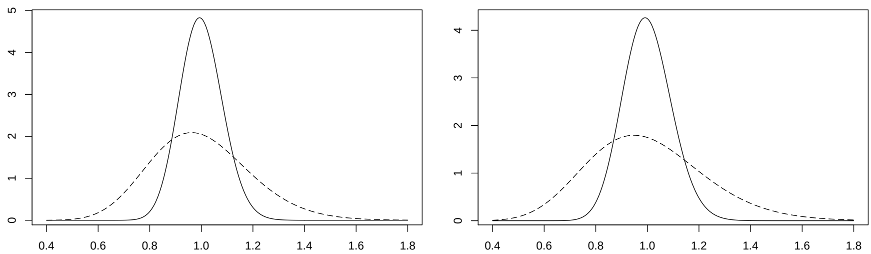

Figure 3.3: Densities of error terms from the MMEM(1,2;1,1) model for NIKC (left) and NIKP (right). The solid line is the density of ε_{1t} and the dashed line is the density of ε_{2t}.

With financial market data, it may be that only the most recent history is relevant in modeling and forecasting, so the MMEM(1,2;1,1) model was also estimated for NIKC and NIKP using only 500 observations from the end of the in-sample period. This corresponds to an in-sample of 20.12.2000 - 30.12.2002. The estimation results for this sample period are given in Table 3.3.[5]

Compared with the values in Table 3.2, the probability parameters are closer in value between NIKC and NIKP in the models estimated with 500 observations (0.87 and 0.93, respectively). Also, the gamma distribution shape parameters are now close to equal in the first regime, contrary to estimates with the entire in-sample. The coefficient for the second lag of NIKP is no longer negative in the first regime, nor is it statistically significant. The constant in the second regime falls to its lower limit, or 0.00001, for NIKC and is thus omitted from the model.

In order to investigate the necessity of the mixture components of the model, a MEM(1,2) specification (i.e., a model with only one mean equation and error distribution) was also estimated for both the call and put IV time series. The

[5]As before, α_{22} is constrained to zero in the models (the p-value for this restriction in an LR test was 0.41 for NIKC and 0.39 for NIKP).

	NIKC		NIKP	
Log likelihood	-1227.37		-1200.69	
π	0.870	(0.000)	0.930	(0.000)
γ_1	146.208	(0.000)	145.589	(0.000)
ω_1	0.654	(0.011)	0.880	(0.026)
α_{11}	0.491	(0.000)	0.452	(0.000)
α_{12}	-0.248	(0.001)	0.042	(0.597)
β_1	0.735	(0.000)	0.475	(0.000)
γ_2	27.471	(0.000)	24.656	(0.004)
ω_2	-		1.522	(0.667)
α_{21}	0.633	(0.006)	0.548	(0.060)
β_2	0.353	(0.139)	0.427	(0.167)

Table 3.3: Estimation results for the MMEM(1,2;1,1) model for NIKC and NIKP with an in-sample of 500 observations. P-values for the significance of the coefficients are given in parentheses.

results of this estimation are given in Table 3.4. Coefficients are not statistically significant, and the parameters of the gamma distribution are very different from those for the MMEM models (see also Figure 3.4). This lends support to the model specification with two mixture components. Also, this specification receives further support from likelihood ratio tests for the two mixture components having the same dynamics. The hypothesis of $\omega_1 = \omega_2$, $\alpha_{11} = \alpha_{21}$, $\alpha_{12} = \alpha_{22}$, and $\beta_1 = \beta_2$ is rejected with a p-value of less than 0.0001 for both NIKC and NIKP. Also, the error terms need different distributions in the two regimes, as the hypothesis that $\gamma_1 = \gamma_2$ is rejected with a p-value of less than 0.00001 for both call and put implied volatility.

3.3.3 Diagnostics

Due to the use of the gamma distribution, it is not possible to conduct many standard diagnostic tests for the MMEM models, as such tests assume a normal distribution. Also, as the model has two mixture components and switching between the two regimes is random, there is no straightforward way to obtain

	NIKC		NIKP	
Log likelihood	33.31		33.57	
γ_1	4.485	(0.000)	4.476	(0.000)
ω	1.487	(0.830)	1.558	(0.839)
α_1	0.464	(0.520)	0.484	(0.508)
α_2	-0.188	(0.894)	-0.166	(0.921)
β_1	0.663	(0.630)	0.616	(0.715)

Table 3.4: Estimation results for the MEM(1,2) model for NIKC and NIKP. P-values for the significance of the coefficients are given in parentheses.

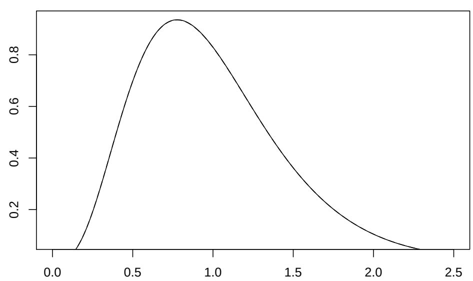

Figure 3.4: Density of error terms from the MEM(1,2) model for NIKC and NIKP.

residuals. In-sample diagnostic checks can be made by analyzing the so-called probability integral transforms of the data, as proposed by Diebold et al. (1998) and employed by e.g. Bauwens et al. (2004) and Lanne (2006). The probability integral transforms are computed as:

$$z_t = \int_0^{y_t} f_{t-1}(u)du \tag{3.6}$$

where $f_{t-1}(\cdot)$ is the conditional density of y_t relating to the model under analysis. The framework of Diebold et al. (1998) was developed to evaluate density forecasts, but it can be used for in-sample diagnostics as well. The diagnostics are based on the idea that the sequence of probability integral transforms of a model's density forecasts are *iid* uniform $U(0,1)$ if the model specification is correct. Diebold et al. (1998) recommend the use of graphical procedures to interpret the fit of the models, which makes the approach simple to use and also easily gives clues as to where a misspecification may lie.

Figure 3.5 plots 25-bin histograms of the probability integral transforms of both NIKC and NIKP with the MMEM(1,2;1,1) model for estimations from the entire in-sample as well as a sample of 500 observations. All columns fall within the 95% confidence interval based on Pearson's goodness-of-fit test, so the model specification succeeds in taking into account the tails of the conditional distribution for both NIKC and NIKP.[6] This holds true even when using only 500 observations in the estimation. The Pearson's goodness-of-fit test statistics, as well as the confidence intervals, are not valid as they are calculated without taking estimation error into account. However, this problem most likely leads to rejecting too frequently, and as the p-values from the test range from 0.78 to 0.90, the results lend support to the model specification.

As a second diagnostic check, autocorrelation functions based on demeaned probability integral transforms and their squares were computed (see Figures 3.6 and 3.7).[7] There is very little autocorrelation in the levels of the demeaned probability integral transforms, but there is noticeable autocorrelation in the squares

[6]With a perfect model, z_t would be uniformly distributed and the columns of the histogram would all be of exactly the same height.

[7]The confidence intervals in the autocorrelation figures are also not valid due to the same estimation error issues as with the Pearson's goodness-of-fit test.

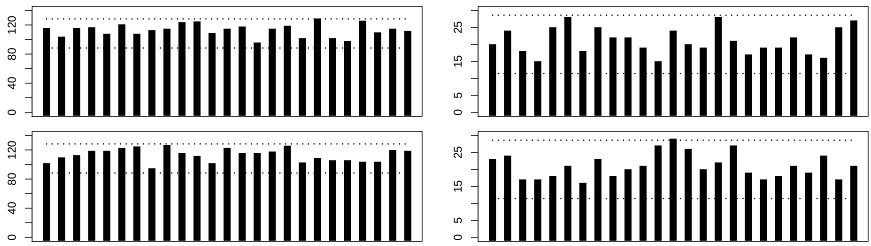

Figure 3.5: Histograms of probability integral transforms for NIKC (upper left panel) and NIKP (lower left panel) with the MMEM(1,2;1,1) model. Histograms for the MMEM(1,2;1,1) model estimated with the last 500 observations of the in-sample are given on the right. The dotted lines depict the boundaries of the 95% confidence interval.

of demeaned z_t for both NIKC and NIKP. The autocorrelation in squares was also present in the data of Lanne (2006). The situation improves clearly in the sub-sample consisting of only the last 500 observations of the in-sample period.[8] This autocorrelation in squares is most likely a reflection of conditional heteroskedasticity, or volatility of volatility, in the data.

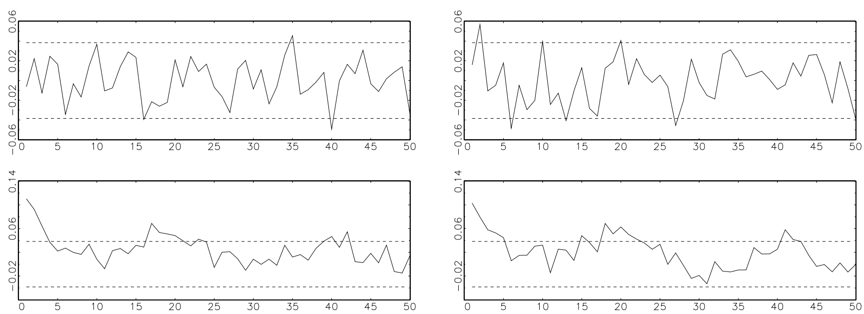

Figure 3.6: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) from the MMEM(1,2;1,1) model. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

The necessity of using the mixture-MEM model specification is underscored when inspecting the histogram of probability integral transforms calculated with

[8]The addition of the statistically insignificant parameter α_{22} to the diagnostic analysis does not improve the autocorrelations.

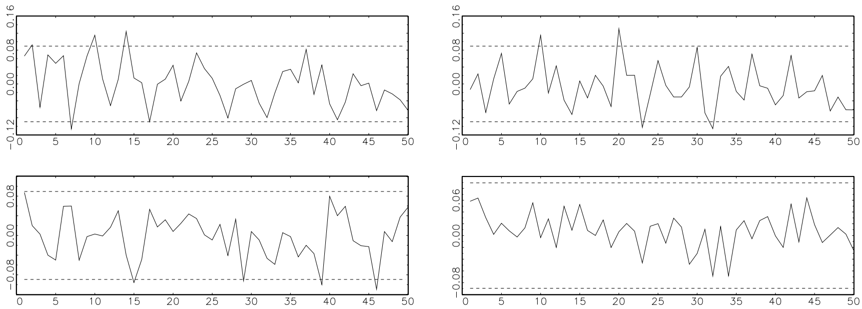

Figure 3.7: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) from the MMEM(1,2;1,1) model estimated with the last 500 observations of the in-sample. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

the MEM(1,2) model (Figure 3.8). With only one regime in the model, the tails of the conditional distribution are not modeled properly, with too much emphasis on the mid-range of the distribution. The poor fit of the MEM(1,2) model is also visible in autocorrelation functions (Figure 3.9), with autocorrelations from even the level series falling well beyond the confidence interval.

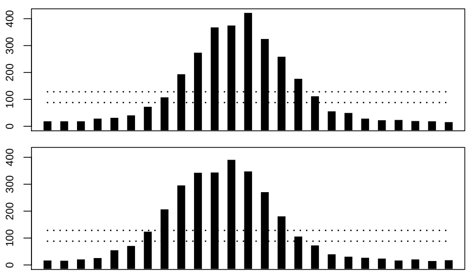

Figure 3.8: Histograms of probability integral transforms with the MEM(1,2) model for NIKC (upper panel) and NIKP (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

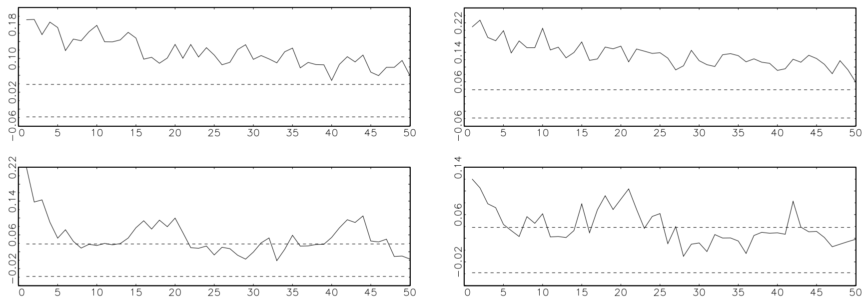

Figure 3.9: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) with the MEM(1,2) model. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

3.4 Forecasts

Forecasts were initially calculated from the chosen model specification of MMEM(1,2;1,1) in order to assess the value of this modeling approach for option traders and other investors. Of the 3,194 observations in the full sample, the last 486 trading days were left as an out-of-sample period. This corresponds to 1.1.2003 - 31.12.2004.

In addition to the MMEM(1,2;1,1) model estimated from the entire in-sample, forecasts were calculated from the MMEM(1,2;1,1) model using 500 observations, as well as from the MEM(1,2) model, which is expected to fare much worse in the forecast evaluation. Forecasts were calculated by keeping the estimated coefficients constant throughout the out-of-sample period as well as by updating the coefficients each day. In this case, the sample size was kept constant (2,708 or 500 observations), with the furthest observation dropped and the newest observation added each day. In this alternative, the most recent information is incorporated into the model estimation, which may result in added value if the coefficients are not stable over time.

In practice, the one-step forecasts $\widehat{y}_{t+1}$ from MMEM models are calculated according to Equation 3.7:

$$\widehat{y}_{t+1} = \pi \hat{\mu}_{1,t+1} + (1-\pi)\hat{\mu}_{2,t+1} \tag{3.7}$$

The forecast performance of the various models is summarized in Table 3.5. Performance is evaluated primarily with directional accuracy, and secondly with mean squared error. Option traders can potentially enter into profitable positions in the market if their expected directional change in IV (up or down) is correct. On the other hand, a forecast of the level of future volatility is of value from a risk management point of view.

In general, it appears to be somewhat easier to forecast NIKC than NIKP, regardless of the model choice. This could perhaps be due to the fact that trading volumes for put options are slightly higher, making put option prices slightly more accurate and predictions more difficult.

The results indicate that the coefficients of the MMEM(1,2;1,1) model are stable over time when using the entire in-sample for estimation. The directional accuracy of the model is exactly the same with fixed and updating coefficients for both NIKC and NIKP. Therefore, it would seem that when using a sample period that is sufficiently long, updating brings little benefit and the choice of the days covered by the sample period is not critical.

The MMEM(1,2;1,1) model forecasts the direction of change correctly on 69.1% of trading days for NIKC and on 66.0% of trading days for NIKP. For e.g. option traders, any level of accuracy over 50% can potentially be worth money. Also, comparing with the findings of Ahoniemi (2008) for the VIX index, whose sign was predicted accurately on 58.4% of trading days at best, the directional accuracy is clearly better for the Nikkei 225 implied volatility. With an older sample, Harvey and Whaley (1992) achieve a directional accuracy of 62.2% for the IV of S&P 100 call options, and 56.6% for put options. The conclusion of this study that call IV is more predictable is thus equivalent to the finding of Harvey and Whaley (1992). Brooks and Oozeer (2002) predict the direction of change of the IV of options on Long Gilt futures on 52.5% of trading days.

When incorporating only the most recent information, or estimating the model with 500 observations, the forecast performance deteriorates considerably. Also, the daily updating of coefficients becomes important, as the directional accuracy improves if using updating rather than fixed coefficients. Further analysis of the parameters estimated with the updating models reveals that a part of the parameter estimates vary considerably over the out-of-sample period. This indicates that

	NIKC			*NIKP*		
	Correct sign	%	*MSE*	*Correct sign*	%	*MSE*
MMEM(1,2;1,1) - updating	**336**	**69.1%**	4.68	**321**	**66.0%**	6.02
MMEM(1,2;1,1) - fixed	**336**	**69.1%**	4.69	**321**	**66.0%**	6.04
MMEM(1,2;1,1) - updating; 500 obs.	325	66.9%	**4.59**	310	63.8%	**5.95**
MMEM(1,2;1,1) - fixed; 500 obs.	309	63.6%	4.89	301	61.9%	6.09
MEM(1,2) - updating	314	64.6%	4.93	309	63.6%	6.04
MEM(1,2) - fixed	310	63.8%	4.98	302	62.1%	6.06

Table 3.5: Correct sign predictions (out of 486 trading days) and mean squared errors. The best values within each column are in boldface.

500 observations may be too few to reliably estimate the parameters.

The forecast performance of the MEM(1,2) model falls short of that of the MMEM(1,2;1,1) model estimated from 2,708 observations, but is no poorer than that of the MMEM(1,2;1,1) model with fixed coefficients estimated from 500 observations. All in all, each model performs at least slightly better in forecasting the direction of change of NIKC rather than NIKP.

		Actual outcome		
		Up	*Down*	*Total*
Forecast	*Up*	188	94	282
	Down	56	148	204
	Total	244	242	486

Table 3.6: NIKC 2×2 contingency table for the MMEM(1,2;1,1) model with updating coefficients

		Actual outcome			
		Up	*Zero*	*Down*	*Total*
Forecast	*Up*	178	1	101	280
	Down	60	3	143	206
	Total	238	4	244	486

Table 3.7: NIKP 2×2 contingency table for the MMEM(1,2;1,1) model with updating coefficients

The directional forecasts can be further broken down into contingency tables to investigate where the models make mistakes. Tables 3.6 and 3.7 show 2×2 contingency tables with forecasts from the MMEM(1,2;1,1) model with updating coefficients and actual outcomes. For both NIKC and NIKP, the true number of moves up and down is almost equal, but the model forecasts a move upwards too often. In other words, the model makes more mistakes where the prediction was up but the true change was down than vice versa. There were four days included in the out-of-sample when the change in NIKP was zero, but the change in NIKC non-zero.

When evaluating the mean squared errors of the various forecast series, values for NIKC are again superior to those for NIKP (see Table 3.5). The MMEM(1,2;1,1) model estimated with 500 observations and updating coefficients emerges as the best specification for both NIKC and NIKP. This is perhaps due to the small sample including observations that are relatively near in value to the current level of IV, whereas the entire in-sample contains observations that are tens of percentage points apart.

The statistical significance of the forecasts is evaluated next with two sets of tests. The value of the obtained directional forecasts can be assessed with the market timing test developed by Pesaran and Timmermann (1992). The Pesaran-Timmermann test (PT test) stems from the case of an investor who switches between stocks and bonds. The test statistic is computed from contingency tables like the one in Table 3.6. For NIKP, the days when the actual outcome was 0 are dropped from the analysis in order to run the test. The limiting distribution of the PT test statistic is $N(0,1)$ when the null hypothesis that actual outcomes and forecasts are distributed independently is true. The PT test shows that all the evaluated directional forecast series are statistically significant, as the test statistic has a p-value of less than 0.00001 for all forecast series. In other words, the null hypothesis of predictive failure can be rejected at the one-percent level of significance.

Mean squared errors can be used in the test for superior predictive ability (SPA) due to Hansen (2005) to check that the forecasts outperform a forecast series of zero change for each day. The SPA test allows for the simultaneous comparison of n series of forecasts, as opposed to the Diebold-Mariano test (Diebold & Mariano, 1995), which compares forecast series one pair at a time. This makes the SPA test useful in the context of this study, as all forecast models can be evaluated at the same time. In the SPA test, one series of forecasts is defined to be the benchmark, and MSE is then used as a loss function to determine whether that benchmark is the best forecast series. The null hypothesis of the test is that the benchmark is not inferior to the other forecast series.

Two benchmarks are used for the purposes of this paper. The first is zero change, meaning that today's value for NIKC or NIKP is used as the forecast for the next day. The consistent p-value of the test, calculated using 1,000 bootstrap resamples, is reported as less than 0.00001 for both NIKC and NIKP. Therefore, there is at least one other forecast series that is a better forecaster than the benchmark of zero change. In other words, the forecasts calculated in this study can clearly add value to such a naive forecast.

The second benchmark is the MMEM(1,2;1,1) model estimated with the entire in-sample and updating coefficients, which produces the second-lowest MSE for both time series. Although being only second best, this model yields much better directional accuracy for both NIKC and NIKP than the model with the lowest MSE. The SPA test reports consistent p-values of 0.448 for NIKC and 0.675 for NIKP with this benchmark, meaning that there is little evidence against the null hypothesis that the benchmark is not inferior to the other series of forecasts.[9]

Investors seeking volatility forecasts in order to evaluate the risks in their portfolios may be more interested in longer-term forecasts than simply one day ahead. To address this issue, forecasts up to five trading days ahead in time have been calculated for both NIKC and NIKP from the models presented above. Table 3.8 summarizes the mean squared errors for these forecasts.

Model	MSE (NIKC)	MSE (NIKP)
MMEM(1,2;1,1) - updating	**6.29**	**9.33**
MMEM(1,2;1,1) - fixed	6.39	9.34
MMEM(1,2;1,1) - updating; 500 obs.	6.54	9.76
MMEM(1,2;1,1) - fixed; 500 obs.	8.33	10.89
MEM(1,2) - updating	8.35	9.87
MEM(1,2) - fixed	8.65	9.90

Table 3.8: Mean squared errors from forecasts five trading days ahead. Best results for each column in boldface.

The longer horizon advocates the use of the mixture multiplicative model with updating coefficients and the longer in-sample period. The mean squared error for the MMEM(1,2;1,1) model with updating coefficients is the lowest for both NIKC and NIKP. However, the MSEs of the same model with fixed coefficients are not much higher. The SPA test confirms the statistical significance of the ranking, giving a consistent p-value of 0.942 to the MMEM(1,2;1,1) model with the lowest mean squared error for NIKC and 0.884 for NIKP. The one-regime model fares worst with the five-step-ahead forecasts for NIKC, and for NIKP, there is only one

[9]The consistent p-values of the SPA test when using the third model as the benchmark, or the model with the lowest MSE, are 0.907 and 0.909 for NIKC and NIKP, respectively.

model (MMEM(1,2;1,1) with 500 observations and fixed coefficients) that produces a higher MSE than the one-regime specification.

To summarize the results from the forecast evaluation, it appears to be beneficial to forecast Nikkei 225 index option implied volatility with a MMEM(1,2;1,1) model specification. The use of a long in-sample period to estimate coefficients provides added value. Two regimes fare better than one regime, and this ranking holds whether the forecasts are calculated one or five days ahead.[10]

3.5 Option trading

In order to further assess the value of the forecasts calculated in Section 3.4, an option trading simulation was carried out with market quotes for Nikkei 225 index options. If an option trader can correctly forecast the direction of change in implied volatility over the next trading day, with all other factors held equal, it is possible to enter into profitable positions in option markets. The purpose of this simulation is not to determine whether trading based on these forecasts could lead to abnormal returns, but to use the returns to rank the forecast models. Therefore, no conclusions regarding market efficiency will be drawn.

The straddle is an option spread that is particularly useful for traders with a view on the future movements in IV. Straddles are volatility trades where a long position benefits from a rise in IV, and a short position benefits from a fall. Noh et al. (1994) use straddles in their option trading simulation, and Bollen & Whaley (2004) note that the most effective way to trade given an impending upward move in volatility is to buy straddles. In a long straddle, an equal number of call and put options are purchased, with a short position involving selling an equal number of calls and puts. The options are selected so that their strike prices and maturity dates are equal.

Option quotes for Nikkei 225 index options were obtained for the entire out-of-sample period, or 1.1.2003 - 31.12.2004. Straddle returns were calculated based on the directional forecasts of all the forecast series reported above. Positions were

[10]The favored model specification of MMEM(1,2;1,1) also outperforms various ARIMA models as a forecaster. ARIMA models were estimated and used to calculate forecasts in order to determine the value of using a multiplicative model rather than a traditional linear model.

entered into at the closing quote of each day, and closed with the closing quote of the following day. A theoretical sum of 1,000 Japanese Yen is invested in long positions, and 1,000 Yen is received from short positions, so that the absolute value of the options does not affect the returns. Harvey and Whaley (1992), Noh et al. (1994), and Ederington and Guan (2002) also use such a fixed investment outlay.

The data on Nikkei 225 options that was used in this analysis contains strike prices 500 index points apart. On each day, the strike price that was as close-to-the-money as possible was selected. Trading volumes are usually highest for at-the-money options and, as noted by Bollen & Whaley (2004), at-the-money options are most sensitive to volatility. Ni et al. (2008) also observe that investors who have a view regarding volatility are more likely to trade with close-to-the-money options than with in-the-money or out-of-the-money options.

The straddle positions are not delta-hedged. However, at-the-money straddles are close to delta-neutral, with the deltas from the call and put options nearly offsetting each other. Options with the next maturity date were used up until two calendar weeks before expiration, with trading then rolled on to the next maturity date. This switch ensures that any odd movements in option quotes and volatility close to expiration do not affect the results.

In practise, the return from a long straddle (R_l) was calculated as in Equation 3.8, and the return from a short straddle (R_s) as in Equation 3.9. The proceeds from the short position are held with zero interest.

$$R_l = \frac{1000}{C_{t-1} + P_{t-1}}(-C_{t-1} - P_{t-1} + C_t + P_t) \tag{3.8}$$

$$R_s = \frac{1000}{C_{t-1} + P_{t-1}}(C_{t-1} + P_{t-1} - C_t - P_t) \tag{3.9}$$

C_{t-1} is the close quote of the near-the-money call option that is used to enter into the straddle, P_{t-1} is the respective quote for the put option with the same strike price and maturity date, and C_t and P_t are the closing quotes of the same options at the end of the next trading day.

It may not be beneficial to trade every single day, particularly if the signal for change in IV is very weak. Three filters have been used to leave out the forecasts for the smallest changes in IV, i.e. treat a part of the forecasts as zero-change forecasts. This approach is similar to those used by Harvey and Whaley (1992) and Noh et al. (1994), who each used two filters in their trading simulations. Poon and Pope (2000) use three filters to account for transaction costs. The three filters employed in this study leave out forecasts for a change of under 1.0%, under 2.0%, and under 5.0%, respectively. No transaction costs are deducted from the returns, as the aim of the trading exercise is not to determine the exact returns from the trades. The inclusion of transaction costs would naturally scale all the reported returns downwards.

The option trading returns are presented in Table 3.9 as average daily percentage returns to the position of 1,000 Yen. For NIKC, the forecasts from the MMEM(1,2;1,1) model estimated with the full in-sample period of 2,708 observations yield the highest returns. Also, the use of a filter would not seem necessary. When looking at the returns from NIKP forecasts, the highest value is again received with the MMEM(1,2;1,1) model, but with the use of fixed coefficients. However, the same model with updating coefficients is by far the second best alternative. As with NIKC, the use of a filter decreases the returns for the MMEM(1,2;1,1) model when it is estimated using 2,708 observations. Interestingly, the use of a filter is profitable for half of the MEM models (those with weakest directional forecast ability). This lends further support to using the MMEM(1,2;1,1) specification with a long in-sample as the forecast model of choice: the forecasts are reliable enough to not need filtering.[11] It should be noted that the daily percentage returns are large, over four percent at best. This is explained in part by a few outliers: for example, the highest daily return contributing to the best average of 4.48% for NIKC was 132.6%. This return occurred on a day when the Nikkei 225 index rose by 2.0%, but the values of the nearest-to-the-money options more than doubled. Leaving out the 25 largest positive and negative daily returns, the average return for the MMEM(1,2;1,1) with updating coefficients and the longer in-sample period falls to 1.33%.

The SPA test was employed in order to assess the statistical significance of

[11] Even with perfect foresight, i.e. by entering into straddles knowing what direction the IV will take over the coming day, the trading return would be positive only on 59.5% (61.3%) of the days for NIKC (NIKP) in this particular out-of-sample. This reflects the fact that many other market parameters that affect option prices also change continually (in addition to volatility).

	NIKC							NIKP						
	Trading P/L	Filter I		Filter II		Filter III		Trading P/L	Filter I		Filter II		Filter III	
MMEM(1,2;1,1) - updating	**4.48**	4.00	(353)	3.90	(288)	2.29	(120)	4.07	3.37	(386)	3.55	(296)	3.01	(113)
MMEM(1,2;1,1) - fixed	4.40	3.96	(359)	3.92	(288)	2.28	(115)	**4.11**	3.30	(380)	3.53	(301)	2.84	(109)
MMEM(1,2;1,1) - updating; 500 obs.	4.36	4.24	(393)	3.71	(312)	3.05	(174)	3.79	3.68	(414)	3.58	(348)	3.47	(178)
MMEM(1,2;1,1) - fixed; 500 obs.	4.17	4.03	(419)	3.81	(347)	3.18	(185)	3.33	3.76	(416)	3.91	(355)	3.42	(188)
MEM(1,2) - updating	3.88	3.63	(418)	3.70	(358)	3.07	(200)	3.44	3.93	(407)	3.40	(343)	3.31	(182)
MEM(1,2) - fixed	3.80	3.60	(417)	3.72	(360)	3.06	(203)	3.62	3.79	(404)	3.70	(346)	3.29	(184)

Table 3.9: Option trading returns expressed as average daily percentage returns from the out-of-sample period of 486 trading days. The return from trading on all 486 days is given in columns Trading P/L. The best values for NIKC and NIKP are given in boldface. Filter I gives no trading signal when the projected change in IV is under 1.0%, Filter II when the signal is under 2.0%, and Filter III when the signal is under 5.0%. The number of days traded out of 486 is given in parentheses for each filtered return.

the differences in the unfiltered option trading returns. In this application, the opposite numbers of the returns achieved when trading every day were used as the loss function in the test. For NIKC, the logical benchmark is the MMEM(1,2;1,1) model with updating coefficients and the full in-sample. The consistent p-value in this case is 0.988, indicating that there is little evidence against the null hypothesis that this is the best forecast model. For NIKP, the same model was used as the benchmark model despite its being only second best in option trading returns. The use of this benchmark is supported by its consistent p-value of 0.753.[12] Therefore, the evidence from the option trading simulation as a whole is not in conflict with the choice of the MMEM(1,2;1,1) model with updating coefficients.

3.6 Conclusions

The modeling of implied volatility can gain new dimensions from multiplicative error models, which allow the econometrician to work directly with observed data rather than logarithms. Implied volatility forecasting is of value in derivatives pricing, as well as in the pricing of other assets due to the nature of IV as a measure of market risk. Portfolio management is also an area that can benefit from accurate IV forecasts. The forecasting of IV is a separate issue from the oft-researched question of how well IV forecasts future realized volatility.

A multiplicative error model with two alternative mean equations and two alternative gamma distributions for the error term was estimated in this paper for time series of implied volatilities derived from call and put options on the Nikkei 225 index. The mixture-MEM model was found to be a good fit, possessing statistically significant coefficients and satisfactory in-sample diagnostics. With only one regime, the model is a much worse fit to the data. In other words, the IV time series appear to both be drawn from two distinct regimes, one corresponding to stable market conditions and the other to a more volatile state.

Measured with directional accuracy, one-step-ahead forecasts calculated from a mixture-MEM model outperform those from a multiplicative model with no mixture components. Therefore, the successful modeling and forecasting of Nikkei

[12]The consistent p-value when using the MMEM(1,2;1,1) model with fixed coefficients as the benchmark is 0.999.

225 implied volatility requires taking the two separate regimes into account. This result holds for both call and put option implied volatility in a two-year out-of-sample period. When mean squared errors are used for forecast evaluation, a two-regime MEM model estimated with a relatively short sample period (500 observations) is superior for both call and put IV.

Further analysis lends support to recommending the use of a mixture-MEM model with a longer in-sample period for forecasting the implied volatility of the Nikkei 225 index. Two-regime models with an eleven-year in-sample period produce lower mean squared errors when forecasting five trading days ahead in time. Also, an option trading simulation that calculates returns from long and short straddle positions based on the signals of the forecasts indicates that the highest average returns can be achieved when trading based on the same mixture-MEM model. Statistical tests confirm this superiority.

These results indicate that option traders and others interested in forecasting the direction of change of implied volatility in the Japanese market can benefit from using the new class of multiplicative models, as directional accuracy is well over 50 percent. The directional forecasts are also statistically significant. Investors looking to forecast the future level of volatility implied by Nikkei 225 options or the future level of volatility in the returns of the Nikkei 225 index can also receive added value from the forecasts of mixture-MEM models, at least up to five trading days into the future. Trading in the Nikkei 225 index option market can also be shown to result in positive returns before transaction costs if entering into straddle positions based on the forecasts from multiplicative error models - at least in the particular out-of-sample used in this study. These returns may not be positive after transaction costs, however, and no evidence on market efficiency is thus provided. A detailed option trading exercise seeking to uncover the true returns for market makers and other investors is left for future research.

Bibliography

[1] Ahoniemi, K. (2008). Modeling and Forecasting the VIX Index. Unpublished working paper.

[2] Bauwens, L. & Giot, P. (2003). Asymmetric ACD models: Introducing price information in ACD models. Empirical Economics, 28, 709-731.

[3] Bauwens, L., Giot, P., Grammig, J., & Veredas, D. (2004). A comparison of financial duration models via density forecasts. International Journal of Forecasting, 20, 589-609.

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

[7] Canina, L., & Figlewski, S. (1993). The Informational Content of Implied Volatility. The Review of Financial Studies, 6, 659-681.

[8] Christensen, B.J., & Prabhala, N.R. (1998). The relation between implied and realized volatility. Journal of Financial Economics, 50, 125-150.

[9] Day, T.E. & Lewis, C.M. (1992). Stock market volatility and the information content of stock index options. Journal of Econometrics, 52, 267-287.

[10] Diebold, F.X., Gunther, T.A., & Tay, A.S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, 39, 863-883.

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

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

[13] Engle, R.F. (2002). New Frontiers for ARCH models. Journal of Applied Econometrics, 17, 425-446.

[14] Engle, R.F. & Gallo, G.M. (2006). A multiple indicators model for volatility using intra-daily data. Journal of Econometrics, 131, 3-27.

[15] Engle, R.F. & Russell, J.R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica, 66, 1127-1162.

[16] Fernandes, M. & Grammig, J. (2006). A family of autoregressive conditional duration models. Journal of Econometrics, 130, 1-23.

[17] Figlewski, S. (1989). Options Arbitrage in Imperfect Markets. The Journal of Finance, 44, 1289–1311.

[18] Garleanu, N.B., Pedersen, L.H. & Poteshman, A.M. (2006). Demand-based Option Pricing. CEPR Discussion Paper, 5420.

[19] Ghysels, E., Gouriéroux, C. & Jasiak, J. (2004). Stochastic volatility duration models. Journal of Econometrics, 119, 413-433.

[20] Hansen, P.R. (2005). A Test for Superior Predictive Ability. Journal of Business & Economic Statistics, 23, 365-380.

[21] 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.

[22] Jorion, P. (1995). Predicting Volatility in the Foreign Exchange Market. The Journal of Finance, 50, 507-528.

[23] Lanne, M. (2006). A Mixture Multiplicative Error Model for Realized Volatility. Journal of Financial Econometrics, 4, 594-616.

[24] Lanne, M. (2007). Forecasting realized exchange rate volatility by decomposition. International Journal of Forecasting, 23, 307-320.

[25] Liu, J. & Longstaff, F.A. (2004). Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities. The Review of Financial Studies, 17, 611–641.

[26] Manganelli, S. (2005). Duration, volume, and volatility impact of trades. Journal of Financial Markets, 8, 377-399.

[27] Meitz, M. & Teräsvirta, T. (2006). Evaluating Models of Autoregressive Conditional Duration. Journal of Business & Economic Statistics, 24, 104-124.

[28] Nelson, D.B. & Cao, C.Q. (1992). Inequality Constraints in the Univariate GARCH Model. Journal of Business & Economic Statistics, 10, 229-235.

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

[30] 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.

[31] Pesaran, M.H. & Timmermann, A.G. (1992). A simple non-parametric test of predictive performance. Journal of Business & Economic Statistics, 10, 461-465.

[32] Poon, S-H., & Pope, P.F. (2000). Trading volatility spreads: a test of index option market efficiency. European Financial Management, 6, 235-260.

Chapter 4

Joint Modeling of Call and Put Implied Volatility

Katja Ahoniemi and Markku Lanne[1]

Abstract

This paper exploits the fact that implied volatilities calculated from identical call and put options have often been empirically found to differ, although they should be equal in theory. We propose a new bivariate mixture multiplicative error model and show that it is a good fit to Nikkei 225 index call and put option implied volatility (IV). A good model fit requires two mixture components in the model, allowing for different mean equations and error distributions for calmer and more volatile days. Forecast evaluation indicates that in addition to jointly modeling the time series of call and put IV, cross effects should be added to the model: put-side implied volatility helps forecast call-side IV, and vice versa. Impulse response functions show that the IV derived from put options recovers faster from shocks, and the effect of shocks lasts for up to six weeks.

[1]Correspondence: katja.ahoniemi@hse.fi and markku.lanne@helsinki.fi. The authors wish to thank seminar participants at the Nonlinear Economics and Finance Research Community in February 2008, at the XVI Annual Symposium of the Society for Nonlinear Dynamics and Econometrics in April 2008, at the European University Institute in May 2008, and at the European Meeting of the Econometric Society in August 2008 for valuable comments. Financial support from the Okobank Group Research Foundation is gratefully acknowledged. Katja Ahoniemi also thanks the Finnish Doctoral Programme in Economics, the Finnish Foundation for Advancement of Securities Markets, and the Yrjö Jahnsson Foundation for financial support.

4.1 Introduction

In theory, the implied volatilities derived from a call option and a put option with the same underlying asset, strike price, and expiration date should be equal - both reflect the market's expectation of the volatility of the returns of the underlying asset during the remaining life of the two options. However, empirical research suggests that when call and put implied volatilities (IV) are backed out of option prices using an option pricing formula, they often deviate from each other.

The reason behind the inequality of put and call implied volatilities may lie in the different demand structure for calls and puts. There is an inherent demand for put options that does not exist for similar calls, as institutional investors buy puts regularly for purposes of portfolio insurance. There are often no market participants looking to sell the same options to offset this demand, meaning that prices may need to be bid up high enough for market makers to be willing to become counterparties to the deals. With no market imperfections such as transaction costs or other frictions present, option prices should always be determined by no-arbitrage conditions, making implied volatilities of identical call and put options the same. However, in real-world markets the presence of imperfections may allow option prices to depart from no-arbitrage bounds if there is, for example, an imbalance between supply and demand in the market. References to existing literature and more details on this topic are provided in Section 4.2.

Despite the fact that call and put-side implied volatilities differ, they must be tightly linked to one another at all times - after all, they both represent the same market expectation, and the driving forces behind their values are common. Therefore, it can be argued that there is potential value added in jointly modeling time series of implied volatilities, one derived from call option prices and the other from put option prices. Further, the interactions between the two variables can be studied with cross effects, i.e. allowing call IV to depend on lagged values of put IV, and vice versa.

The modeling of IV provides a valuable addition to the extensive literature on volatility modeling. IV is truly a forward-looking measure: implied volatility is the market's expectation of the volatility in the returns of an option's underlying asset during the remaining life of the option in question. Examples of IV modeling literature include Ahoniemi (2008), who finds that there is some predictability in the direction of change of the VIX Volatility Index, an index of the IV of S&P 500 index options. Dennis et al. (2006) find that daily innovations in the VIX Volatility Index contain very reliable incremental information about the future volatility of the S&P 100 index.[2] Other studies that attempt to forecast IV or utilize the information contained in IV to trade in option markets include Harvey and Whaley (1992), Noh et al. (1994), and Poon and Pope (2000). Reliable forecasts of implied volatility can benefit option traders, but many other market participants as well: all investors with risk management concerns can benefit from accurate forecasts of future volatility.

The implied volatility data used in this study are calculated separately from call and put options on the Japanese Nikkei 225 index. Separate time series for call and put-side IV offer a natural application for the bivariate multiplicative model presented below. In their analysis of implied volatilities of options on the S&P 500 index, the FTSE 100 index, and the Nikkei 225 index, Mo and Wu (2007) find that U.S. and UK implied volatilities are more correlated with each other than with Japanese implied volatilities, indicating that the Japanese market exhibits more country-specific movements. Therefore, it is interesting to analyze the Japanese option market and its implied volatility in this context, as investors may be presented with possibilities in the Japanese index option market that are not available elsewhere. Mo and Wu (2007) also report that the implied volatility skew is flatter in Japan than in the U.S. or UK markets. They conclude that in Japan, the risk premium for global return risks is smaller than in the other two countries. The developments in the Japanese stock market during the late 1990s in particular are very different from Western markets, with prices declining persistently in Japan. This characteristic also makes the Japanese market unique. Mo and Wu (2007) observe that out-of-the-money calls have relatively higher IVs in Japan, as investors there expect a recovery after many years of economic downturn. Investors in Japan seem to price more heavily against volatility increases than against market crashes.

[2]The data set in Dennis et al. (2006) ends at the end of 1995, when options on the S&P 100 index were used to calculate the value of the VIX. The Chicago Board Options Exchange has since switched to S&P 500 options.

In this paper, we introduce a new bivariate multiplicative error model (MEM). MEM models have gained ground in recent years due to the increasing interest in modeling non-negative time series in financial market research.[3] The use of MEM models does not require logarithms to be taken of the data, allowing for the direct modeling of variables such as the duration between trades, the bid-ask spread, volume, and volatility. Recent papers that successfully employ multiplicative error modeling in volatility applications include Engle and Gallo (2006), Lanne (2006, 2007), Brunetti and Lildholdt (2007), and Ahoniemi (2007). Lanne (2006) finds that the gamma distribution is well suited for the multiplicative modeling of the realized volatility of two exchange rate series, and Ahoniemi (2007), using the same data set as in the present study, finds that MEM models together with a gamma error distribution are a good fit to data on Nikkei 225 index implied volatility. All the above-mentioned MEM applications consider univariate models, but Cipollini et al. (2006) build a multivariate multiplicative error model using copula functions instead of directly employing a multivariate distribution. In our application, we use a bivariate gamma distribution to model the residuals.

Our results show that it is indeed useful to jointly model call and put implied volatilities. The chosen mixture bivariate model with a gamma error distribution is a good fit to the data, as shown by coefficient significance and diagnostic checks. The addition of lagged cross effects turns out to be important for one-step-ahead daily forecast performance. Our model correctly forecasts the direction of change in IV on over 70% of trading days in an out-of-sample analysis. Impulse response functions are also calculated, and they reveal that there is considerable persistence in the data: shocks do not fully disappear until thirty trading days elapse. Also, put-side IV recovers more quickly from shocks than call-side IV, indicating that the market for put options may price more efficiently due to larger demand and trading volumes.

This paper proceeds as follows. Section 4.2 discusses the differences in the markets for call and put options in more detail. Section 4.3 describes the bivariate mixture multiplicative error model estimated in this paper. Section 4.4 presents the data, model estimation results, and diagnostic checks of the chosen model

[3]A special case of multiplicative error models is the autoregressive conditional duration (ACD) model, for which an abundant literature has emerged over the past ten years.

specification. Impulse response functions are discussed in Section 4.5, and forecasts are evaluated in Section 4.6. Section 4.7 concludes.

4.2 The Markets for Call and Put Options

This section provides evidence on the differences between call and put option implied volatilities, and discusses how demand pressures and limits to arbitrage can lead to these theory-contradicting differences. For example, Bollen and Whaley (2004) have documented that put options account for 55 % of trades in S&P 500 index options, and that the level of implied volatility calculated from at-the-money (ATM) options on the S&P 500 index is largely driven by the demand for ATM index puts. Buraschi and Jackwerth (2001), using an earlier data set of S&P 500 index options, report that put volumes are around three times higher than call volumes. There is also evidence that out-of-the-money (OTM) puts in particular can be overpriced, at least part of the time (Bates (1991), Dumas et al. (1998), Bollen and Whaley (2004)). Garleanu et al. (2006) document that end users (non-market makers) of options have a net long position in S&P 500 index puts, and that net demand for low-strike options (such as OTM puts) is higher than the demand for high-strike options. The results of Chan et al. (2004) from Hang Seng Index options in Hong Kong are similar to those of Bollen and Whaley (2004) in that net buying pressure is more correlated with the change in implied volatility of OTM put options than in-the-money put options. Also, trading in Hang Seng Index puts determines the shape of the volatility smile to a greater degree than trading in calls.

If OTM puts are consistently overpriced, investors who write such options could earn excess returns (empirical evidence in support of this is provided in e.g. Bollen and Whaley (2004)). On the other hand, Jackwerth (2000) finds that it is more profitable to sell ATM puts than OTM puts in the S&P 500 index option market. Fleming (1999) compares ATM S&P 100 index calls and puts, and finds that selling puts is more profitable than selling calls. Further evidence on different market mechanisms for calls and puts is provided by Rubinstein (1994), who notes that after the stock market crash of October 1987, prices of OTM puts were driven upwards, changing the volatility smile into the now-observed volatility skew. He hypothesizes that the crash led to OTM puts being more highly valued in the eyes of investors. Ederington and Guan (2002) also remark that the volatility smile may be caused in part by hedging pressures which drive up the prices of out-of-the-money puts. They point out that this notion is supported by both trading volume evidence and the fact that in equity markets, implied volatilities calculated from options with low strike prices have been found to be higher than ex-post realized volatilities.

Even if the demand for a put option causes its price (and implied volatility) to rise, no-arbitrage conditions should ensure that the price of a call option with the same strike price and maturity date yields an implied volatility that is equal to the one derived from the put counterpart. But as Fleming (1999) writes,

> "..., transaction costs and other market imperfections can allow option prices to deviate from their "true" values without signaling arbitrage opportunities."

The possibility that option prices can depart from no-arbitrage bounds, thus allowing call and put IV to differ, has been documented numerous times in earlier work. Hentschel (2003) points out that noise and errors in option prices stemming from fixed tick sizes, bid-ask spreads, and non-synchronous trading can contribute to miscalculated implied volatilities, and to the volatility smile. Garleanu et al. (2006) develop a model for option prices that allows for departures from no-arbitrage bounds. These arise from the inability of market makers to perfectly hedge their positions at all times, which in turn allows option demand to affect option prices. Empirical evidence lends support to this theory: market makers require a premium for delivering index options. Even market makers cannot fully hedge their exposures due to issues such as transaction costs, the indivisibility of securities, and the impossibility of executing rebalancing trades continuously (Figlewski (1989)), and capital requirements and sensitivity to risk (Shleifer and Vishny (1997)). When market makers face unhedgeable risk, they must be compensated through option prices for bearing this risk. In fact, Garleanu et al. (2006) find that after periods of dealer losses, the prices of options are even more sensitive to demand. Other impediments to arbitrage include the fact that a stock index portfolio is difficult and costly to trade, but if an investor uses futures, she must bear basis and possibly tracking risk (Fleming (1999)): spot and futures prices may not move hand-in-hand at all times, and the underlying asset of the futures contracts may not be identical to the asset being hedged. Liu and Longstaff (2004) demonstrate that it can often be optimal to underinvest in arbitrage opportunities, as mark-to-market losses can be considerable before the values of the assets involved in the trade converge to the values that eventually produce profits to the arbitrageur. When it is suboptimal to fully take advantage of an arbitrage opportunity, there is no reason why the arbitrage could not persist for even a lengthy amount of time. Bollen and Whaley (2004), in their analysis of the S&P 500 option market, find support for the hypothesis that limits to arbitrage allow the demand for options to affect implied volatility.

4.3 The Model

In this section, we present the bivariate mixture multiplicative error model (BVMEM) that will be used to model the two time series of implied volatilities described in Section 4.4. Consider the following bivariate model

$$\mathbf{v}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t, \qquad t = 1, 2, ..., T,$$

where the conditional mean

$$\boldsymbol{\mu}_t = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} = \begin{pmatrix} \omega_1 + \sum_{i=1}^{q_1} \alpha_{1i} v_{1,t-i} + \sum_{j=1}^{p_1} \beta_{1j} \mu_{1,t-j} \\ \omega_2 + \sum_{i=1}^{q_2} \alpha_{2i} v_{2,t-i} + \sum_{j=1}^{p_2} \beta_{2j} \mu_{2,t-j} \end{pmatrix},$$

$\boldsymbol{\varepsilon}_t$ is a stochastic positive-valued error term such that $E\left(\boldsymbol{\varepsilon}_t | \mathfrak{F}_{t-1}\right) = \mathbf{1}$ with $\mathfrak{F}_{t-1} = \{\mathbf{v}_{t-j}, j \geq 1\}$, and $\odot$ denotes element-by-element multiplication. In what follows, this specification will be called the BVMEM($p_1, q_1; p_2, q_2$) model. As the conditional mean equations of the model are essentially the same as the conditional variance equations in the GARCH model in structure, the constraints on parameter values that guarantee positivity in GARCH models also apply to each of the equations of the BVMEM model. As outlined in Nelson and Cao (1992), the parameter values in a first-order model must all be non-negative. In a higher-order model, positivity of all parameters is not necessarily required. For example, in a model with $p_i = 1$ and $q_i = 2$, $i = 1, 2$, the constraints are $\omega_i \geq 0$, $\alpha_{i1} \geq 0$, $0 \leq \beta_i < 1$, and $\beta_1 \alpha_{i1} + \alpha_{i2} \geq 0$. It should be noted that this basic conditional mean specification must often be augmented with elements such as cross effects between the variables and seasonality effects. In these cases, one must ensure that positivity continues to be guaranteed. For example, if the coefficients for lagged cross terms are positive, no problems in achieving positivity arise.

The multiplicative structure of the model was suggested for volatility modeling in the univariate case by Engle (2002), who proposed using the exponential distribution. However, the gamma distribution nests, among others, the exponential distribution, and is therefore more general. Also, the findings of Lanne (2006, 2007) and Ahoniemi (2007) lend support to the gamma distribution.

The error term $\boldsymbol{\varepsilon}_t$ is assumed to follow a bivariate gamma distribution, which is a natural extension of the univariate gamma distribution used in previous literature (Lanne (2006, 2007) and Ahoniemi (2007)). Of the numerous bivariate distributions having gamma marginals, the specification suggested by Nagao and Kadoya (1970) is considered (for a discussion on alternative bivariate gamma densities, see Yue et al. (2001)). This particular specification is quite tractable and thus well suited for our purposes. Collecting the parameters into vector $\boldsymbol{\theta} = (\tau_1, \tau_2, \lambda, \rho)$, the density function can be written as

$$f_{\varepsilon_1,\varepsilon_2}\left(\varepsilon_{1t}, \varepsilon_{2t}; \boldsymbol{\theta}\right) = \frac{\left(\tau_1 \tau_2\right)^{(\lambda+1)/2} \left(\varepsilon_{1t}\varepsilon_{2t}\right)^{(\lambda-1)/2} \exp\left\{-\frac{\tau_1\varepsilon_{1t}+\tau_2\varepsilon_{2t}}{1-\rho}\right\}}{\Gamma\left(\lambda\right)\left(1-\rho\right)\rho^{(\lambda-1)/2}} I_{\lambda-1}\left(\frac{2\sqrt{\tau_1\tau_2\rho\varepsilon_{1t}\varepsilon_{2t}}}{1-\rho}\right),$$

where $\Gamma(\cdot)$ is the gamma function, ρ is the Pearson product-moment correlation coefficient, and $I_{\lambda-1}(\cdot)$ is the modified Bessel function of the first kind. The marginal error distributions have distinct scale parameters τ_1 and τ_2, but the shape parameter, λ, is the same for both. However, since the error term needs to have mean unity, we impose the restriction that the shape and scale parameters are equal, i.e. $\tau_1 = \tau_2 = \lambda$. In other words, we will also restrict the scale parameters to be equal. This is not likely to be very restrictive in our application, as earlier evidence based on univariate models in Ahoniemi (2007) indicates that the shape and scale parameters for the time series used in this study, the implied volatilities of Nikkei 225 call and put options, are very similar.

Incorporating the restrictions discussed above and using the change of variable theorem, the conditional density function of $\mathbf{v}_t = (v_{1t}, v_{2t})'$ is obtained as

$$\begin{aligned} f_{t-1}\left(v_{1t}, v_{2t}; \boldsymbol{\theta}\right) &= f_{\varepsilon_1,\varepsilon_2}\left(v_{1t}\mu_{1t}^{-1}, v_{2t}\mu_{2t}^{-1}\right)\mu_{1t}^{-1}\mu_{2t}^{-1} \\ &= \frac{\lambda^{(\lambda+1)}\left[v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}\right]^{(\lambda-1)/2}\exp\left\{-\frac{\lambda\left(v_{1t}\mu_{1t}^{-1}+v_{2t}\mu_{2t}^{-1}\right)}{1-\rho}\right\}}{\Gamma\left(\lambda\right)\left(1-\rho\right)\rho^{(\lambda-1)/2}} \times \\ &\quad I_{\lambda-1}\left(\frac{2\lambda\sqrt{\rho v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}}}{1-\rho}\right)\mu_{1t}^{-1}\mu_{2t}^{-1}. \end{aligned} \tag{4.1}$$

Consequently, the conditional log-likelihood function can be written as[4]

$$l_T\left(\boldsymbol{\theta}\right) = \sum_{t=1}^{T} l_{t-1}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln\left[f_{t-1}\left(v_{1t}, v_{2t}; \boldsymbol{\theta}\right)\right],$$

and the model can be estimated with the method of maximum likelihood (ML) in a straightforward manner. Although the gamma distribution is quite flexible in describing the dynamics of implied volatilities, in our empirical application it turned out to be inadequate as such. In particular, it failed to capture the strong persistence in the implied volatility time series. As an extension, we consider a mixture specification that allows for the fact that financial markets experience different types of regimes, alternating between calm and more volatile periods of time. Different parameter values can be assumed to better describe periods of larger shocks compared with periods of smaller shocks, and error terms are allowed to come from two gamma distributions whose shape and scale parameters can differ. Earlier evidence from Lanne (2006) and Ahoniemi (2007) indicates that

[4]Specifically, for observation t,

$$\begin{aligned} l_{t-1}\left(\boldsymbol{\theta}\right) &= \left(\lambda+1\right)\ln\left(\lambda\right) + \frac{1}{2}\left(\lambda-1\right)\left[\ln\left(v_{1t}\right)+\ln\left(v_{2t}\right)-\ln\left(\mu_{1t}\right)-\ln\left(\mu_{2t}\right)\right] \\ &\quad -\frac{\lambda\left(v_{1t}\mu_{1t}^{-1}+v_{2t}\mu_{2t}^{-1}\right)}{1-\rho} - \ln\left[\Gamma\left(\lambda\right)\right] - \ln\left(1-\rho\right) - \frac{1}{2}\left(\lambda-1\right)\ln\left(\rho\right) \\ &\quad +\ln\left[I_{\lambda-1}\left(\frac{2\lambda\sqrt{\rho v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}}}{1-\rho}\right)\right] - \ln\left(\mu_{1t}\right) - \ln\left(\mu_{2t}\right). \end{aligned}$$

the use of a mixture specification improves the fit of a multiplicative model as well as the forecasts obtained from the models.

We will assume that the error term $\boldsymbol{\varepsilon}_t$ is a mixture of of $\boldsymbol{\varepsilon}_t^{(1)}$ and $\boldsymbol{\varepsilon}_t^{(2)}$ with mixing probability π, and that $\boldsymbol{\varepsilon}_t^{(1)}$ and $\boldsymbol{\varepsilon}_t^{(2)}$ follow the bivariate gamma distribution with parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. In other words, the error term is $\boldsymbol{\varepsilon}_t^{(1)}$ with probability π and $\boldsymbol{\varepsilon}_t^{(2)}$ with probability $1-\pi$ $(0<\pi<1)$. The model based on this assumption will subsequently be called the mixture-BVMEM model. The conditional log-likelihood function becomes

$$l_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} l_{t-1}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln\left[\pi f_{t-1}^{(1)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_1) + (1-\pi) f_{t-1}^{(2)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_2)\right],$$

where $f_{t-1}^{(1)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_1)$ and $f_{t-1}^{(2)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_2)$ are given by (4.1) with $\boldsymbol{\theta}$ replaced by $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. As mentioned previously, the conditional mean parameters can differ between regimes.

The asymptotic properties of the ML estimator for our model are not known, and their derivation lies outside the scope of this paper. However, assuming that v_t is stationary and ergodic, it is reasonable to apply standard asymptotic results in statistical inference. In particular, approximate standard errors can be obtained from the diagonal elements of the matrix $-\left[\partial^2 l_T(\hat{\theta})/\partial\theta\partial\theta'\right]^{-1}$, where $\hat{\theta}$ denotes the ML estimate of $\boldsymbol{\theta}$. Similarly, Wald and likelihood ratio (LR) tests for general hypotheses will have the conventional asymptotic χ^2 null distributions. Note, however, that hypotheses restricting the number of mixture components do not have the usual χ^2 distributions due to the problem of unidentified parameters (see e.g. Davies (1977)). We will not attempt such tests, but assume throughout that there are two mixture components. The adequacy of the assumption will be verified by means of diagnostic procedures (see Section 4.4.3).

4.4 Estimation Results

4.4.1 Data

The data set in this study covers 3,194 daily closing observations from the period 1.1.1992 - 31.12.2004, and was obtained from Bloomberg Professional Service (see Figure 4.1). The first eleven years of the full sample, or 1.1.1992 - 31.12.2002, comprise the in-sample of 2,708 observations. The final two years, 2003 and 2004, are left as the 486-day out-of-sample to be used for forecast evaluation.

The call-side (put-side) implied volatility time series is calculated as an unweighted average of Black-Scholes implied volatilities from two nearest-to-the-money call (put) options from the nearest maturity date. Rollover to the next maturity occurs two calendar weeks prior to expiration in order to avoid possibly erratic behavior in IV close to option expiration. ATM options are typically used to estimate the market's expected volatility for the remainder of the option's maturity, as trading volumes are usually high for ATM options. Also, ATM options have the highest sensitivity to volatility.

Table 4.1 provides descriptive statistics on both the call-side IVs (NIKC) and put-side IVs (NIKP). The average level of put-side implied volatility is higher in the sample of this study, a phenomenon which has also been documented in the U.S. markets by Harvey and Whaley (1992). Tests indicate that the means and medians of the two series are not significantly different (p-values exceed 0.4 in both cases). On the other hand, the equality of the standard deviations is rejected at the one-percent level. The correlation between NIKC and NIKP is 0.877 in the full sample, but the correlation has varied from 0.534 in 1996 to 0.924 in 1994.

4.4.2 Model Estimation

Given the clear linkages between the implied volatilities of call and put options on the same underlying asset outlined above, call-side (put-side) IV can be expected to be a significant predictor of future put-side (call-side) IV. Therefore, the model presented in Section 4.3 is augmented with lagged cross terms, so that call (put) implied volatility depends on its own history as well as on the history of put (call) implied volatility. Bollen and Whaley (2004) find that in the U.S. market,

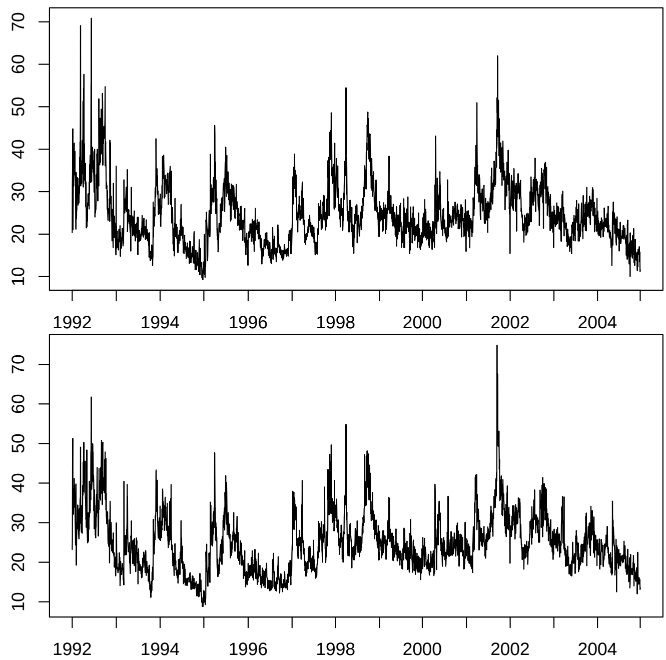

Figure 4.1: Nikkei 225 index call implied volatility (upper panel) and put implied volatility (lower panel) 1.1.1992 - 31.12.2004.

	NIKC	NIKP
Maximum	70.84	74.87
Minimum	9.26	8.80
Mean	24.68	24.82
Median	23.42	23.84
Standard deviation	7.07	7.41
Skewness	1.10	0.94
Kurtosis	5.42	4.79

Table 4.1: Descriptive statistics for NIKC and NIKP for the full sample of 1.1.1992 - 31.12.2004.

the demand for ATM index puts drives both the changes in ATM put implied volatility and the changes in ATM call implied volatility. Therefore, we expect that for our Japanese implied volatility data, lagged put IV will be more significant in explaining call IV than lagged call IV will be in explaining put IV.

Dummy variables for Friday effects of put-side IV are also added due to the improvement in diagnostics achieved after the addition (see Section 4.4.3 for more details on diagnostic checks). The level of IV is lowest on Fridays for both call and put options,[5] but trading volumes are highest on Fridays. An analysis of trading volumes of close-to-the-money, near-term maturity call and put options on the Nikkei 225 index reveals that during the two-year out-of-sample period used in this study, put options account for 52.0 % of trading volume (measured with number of contracts traded). The share of puts is lowest on Mondays (50.4%) and largest on Fridays (53.6%).

Findings similar to ours concerning weekly seasonality have been reported in previous studies. Harvey and Whaley (1992) document that the IV in the S&P 100 index option market tends to be lowest on Fridays and rise on Mondays. Peña et al. (1999) find that in the Spanish stock index market, the curvature of the volatility smile at the beginning of the week is statistically significantly different from the smile at the end of the week. Lehmann and Modest (1994) report that trading volumes on the Tokyo Stock Exchange are substantially lower on Mondays than on other days of the week. They hypothesize that this is due to reduced demand by liquidity traders due to the risk of increased information asymmetry after the weekend. Also, bid-ask spreads are largest on Mondays, making transaction costs highest at the start of the week. The significance of trading volumes is highlighted by Mayhew and Stivers (2003), who find that implied volatility performs well when forecasting individual stock return volatility, but only for those stocks whose options have relatively high trading volumes.

In order to take cross effects and the observed seasonal variation into account, we need to modify the basic model presented in Section 4.3. Let μ_{mt} denote the conditional mean of mixture component m ($m = 1, 2$), and $\mu_{mt} = (\mu^C_{mt}, \mu^P_{mt})'$, where μ^C_{mt} and μ^P_{mt} are the conditional means of the call and put implied volatilities,

[5]The level of IV is highest on Mondays. However, dummies for Monday effects were not statistically significant.

respectively.

The specifications of the conditional means are

$$\mu_{mt}^{C} = \omega_m^C + \sum_{i=1}^{q_C} \alpha_{mi}^C v_{C,t-i} + \sum_{i=1}^{r_C} \psi_{mi}^C v_{P,t-i} + \sum_{i=1}^{s_C} \delta_{mi}^{CP} D_i v_{P,t-i} + \sum_{j=1}^{p_C} \beta_{mj}^C \mu_{m,t-j}^C$$

and

$$\mu_{mt}^{P} = \omega_m^P + \sum_{i=1}^{q_P} \alpha_{mi}^P v_{P,t-i} + \sum_{i=1}^{r_P} \psi_{mi}^P v_{C,t-i} + \sum_{i=1}^{s_P} \delta_{mi}^{PP} D_i v_{P,t-i} + \sum_{j=1}^{p_P} \beta_{mj}^P \mu_{m,t-j}^P$$

where the ψ's are the coefficients for lagged cross terms, and D_i receives the value of 1 on Fridays, and zero otherwise. As mentioned above, the dummy variable in both the call and put mean equations is for put-side Friday effects (coefficients δ_{mi}^{CP} and δ_{mi}^{PP}). This specification is later referred to as the unrestricted model.[6]

In order to fully understand the value of including cross effects between NIKC and NIKP in the model, an alternative specification with no cross terms was also estimated. In this model, dummies for Friday effects are also included, but due to the elimination of cross effects, the dummy in the equation for NIKC captures the Friday effect of call-side, not put-side, implied volatility. In the second model specification, or the restricted model,

$$\mu_{mt}^{C} = \omega_m^C + \sum_{i=1}^{q_C} \alpha_{mi}^C v_{C,t-i} + \sum_{i=1}^{s_C} \delta_{mi}^{CC} D_i v_{C,t-i} + \sum_{j=1}^{p_C} \beta_{mj}^C \mu_{m,t-j}^C$$

and

$$\mu_{mt}^{P} = \omega_m^P + \sum_{i=1}^{q_P} \alpha_{mi}^P v_{P,t-i} + \sum_{i=1}^{s_P} \delta_{mi}^{PP} D_i v_{P,t-i} + \sum_{j=1}^{p_P} \beta_{mj}^P \mu_{m,t-j}^P.$$

The estimation results for both the unrestricted and the restricted model are presented in Table 4.2. The lag structure is selected based on the statistical significance of coefficients as well as autocorrelation diagnostics (see Section 4.4.3).

[6] The model originally included six dummies: both first-regime equations had Friday-effect dummies for the intercept, own lagged value, and the lagged value of the other variable. Only the put-side Friday effects were statistically significant, and p-values from likelihood ratio tests validated the constraining of the other dummies to zero.

	Unrestricted Model		Restricted Model	
Log likelihood	-12370.0		-12653.4	
π	0.919**	(0.012)	0.883**	(0.018)
λ_1	126.594**	(3.745)	127.3036**	(4.614)
ρ_1	0.094**	(0.027)	0.019	(0.033)
ω_1^C	1.264**	(0.164)	0.298**	(0.085)
α_{11}^C	0.514**	(0.020)	0.617**	(0.025)
α_{12}^C	0.104**	(0.019)	-0.223**	(0.048)
ψ_{11}^C	0.321**	(0.017)	-	
δ_{11}^{CC}	-		0.043**	(0.006)
δ_{11}^{CP}	0.045**	(0.005)	-	
β_{11}^C	-		0.584**	(0.046)
ω_1^P	-		0.254**	(0.077)
α_{11}^P	0.529**	(0.020)	0.611**	(0.023)
α_{12}^P	-		-0.215**	(0.053)
ψ_{11}^P	0.247**	(0.018)	-	
δ_{11}^{PP}	0.043**	(0.005)	0.046**	(0.006)
β_{11}^P	0.210**	(0.026)	0.581**	(0.048)
λ_2	20.043**	(2.003)	23.413**	(2.288)
ρ_2	0.360**	(0.063)	0.378**	(0.058)
ω_2^C	1.401	(0.772)	0.718	(0.368)
α_{21}^C	0.185**	(0.070)	0.216**	(0.044)
ψ_{21}^C	0.150	(0.101)	-	
β_{21}^C	0.627**	(0.108)	0.769**	(0.048)
ω_2^P	0.835	(0.688)	0.933*	(0.403)
α_{21}^P	0.244**	(0.087)	0.270**	(0.050)
ψ_{21}^P	0.146	(0.090)	-	
β_{21}^P	0.612**	(0.111)	0.717**	(0.053)

Table 4.2: Estimation results for the BVMEM model. Standard errors calculated from the final Hessian matrix are given in parentheses. (**) indicates statistical significance at the one-percent level, and (*) at the five-percent level.

The parameter values for all ω's, α's and β's meet the Nelson and Cao (1992) constraints discussed in Section 4.3. Also, the coefficients of cross terms (ψ's) and dummies (δ's) are positive, so positivity is guaranteed in the model. The estimated values of β_{11}^C, ω_1^P, and α_{12}^P are not reported because the estimates hit their lower bound of zero (which is required for positivity).

The probability parameter π is quite high for the unrestricted model, close to 0.92. Therefore, the second regime, which displays larger shocks, occurs on only some eight percent of the trading days in the in-sample. The estimated shape (and scale) parameters of the error distribution differ considerably between the two regimes, with residuals more dispersed in the second regime. Figure 4.2 shows the joint error density of the unrestricted model with the parameters estimated for the first regime, while the error density for the second regime is depicted in Figure 4.3. It should be noted that the scale of the z-axis is different in the two figures. The errors are much more tightly concentrated around unity in the first, more commonly observed, regime, whereas the tail area is emphasized in the second regime.

The correlation of errors, or ρ, is higher in the second regime, making changes in call and put IV more correlated when volatility is high. This is also clearly visible in Figures 4.2 and 4.3. The coefficients of cross terms ψ are significant at the one-percent level in the first, more common regime, and jointly significant in the second regime (p-value from LR test equal to 0.007). The coefficients of the cross terms are higher in the first regime, making the cross effects more pronounced. In other words, the cross effects are smaller when volatility is high. For both regimes, the effect put-side IV has on call-side IV is larger than the effect call IV has on put IV, although the difference in coefficients is quite small in the second regime. Friday dummies for the first lag of put IV are also significant and positive, indicating that the effect of the lagged put IV is larger on Fridays, when trading volumes are highest. Values of intercepts are higher in the second regime, consistent with the notion that this regime occurs on days when shocks are larger. The clearly greater β's in the second regime indicate higher persistence in that regime. This can be interpreted as a sign that once the second regime is entered, it is likely that large shocks persist, i.e. there is volatility clustering present.

As we are interested in seeing the relevance of cross terms for forecast perfor-

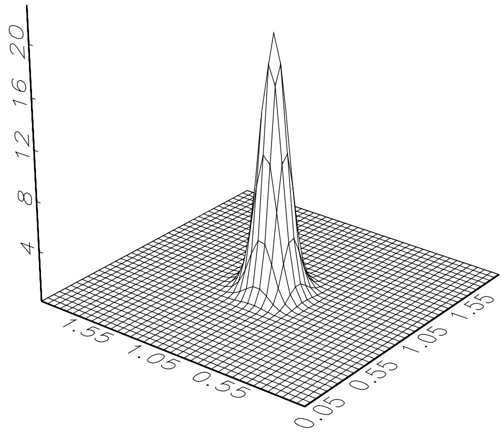

Figure 4.2: Estimated density of error terms in the first regime of the unrestricted model.

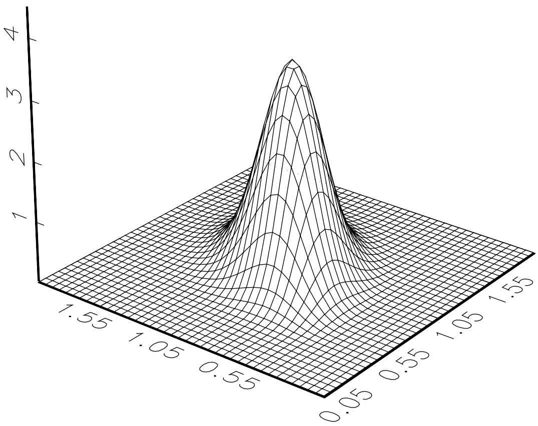

Figure 4.3: Estimated density of error terms in the second regime of the unrestricted model.

mance, we also present the results for the restricted model without these cross effects. It should be noted that the null hypothesis of all coefficients of cross terms equal to zero is rejected by an LR test at all reasonable significance levels (p-value of less than 0.00001). In the restricted model, the estimate of π is smaller than in the unrestricted model, but the first regime remains clearly more prevalent. The parameters of the error distribution are very similar, but the correlation of the residuals is lower in the first regime than it was with the unrestricted model. One notable difference to the parameter values of the unrestricted model is that the coefficients of the second lags are both significant in the first regime, and have a negative sign. This suggests that the exclusion of cross effects results in biased estimates of these parameters. The dummies for Friday effects are significant, indicating that the data behaves somewhat differently when the trading volume is at its highest. As the shape (and scale) parameters of the error distribution that are estimated for the restricted model are very close in value to those for the unrestricted model, the graphs for error densities are qualitatively similar as those in Figures 4.2 and 4.3 and are therefore not displayed.

4.4.3 Diagnostics

Most standard diagnostic tests are based on a normal error distribution, which renders these tests unfeasible for our purposes due to the use of the gamma distribution. Also, as our model specification has two mixture components and switching between the regimes is random, there is no straightforward way to obtain residuals.

In order to investigate the goodness-of-fit of our model, diagnostic evaluations can nevertheless be conducted by means of so-called probability integral transforms of the data. This method was suggested by Diebold et al. (1998) and extended to the multivariate case by Diebold et al. (1999). The probability integral transform in the univariate case (for one IV series) is obtained as

$$z_t = \int_0^{y_t} f_{t-1}(u)du \qquad (4.2)$$

where $f_{t-1}(\cdot)$ is the conditional density of the implied volatility with the chosen model specification. The transforms are independently and identically uniformly distributed in the range [0,1] if the model is correctly specified. Although commonly employed in the evaluation of density forecasts, this method is also applicable to the evaluation of in-sample fit. In the bivariate case, Diebold et al. (1999) recommend evaluating four sets of transforms, which we denote z_t^C, z_t^P, $z_t^{C|P}$, and $z_t^{P|C}$. The transforms z_t^C and z_t^P are based on the marginal densities of the call and put implied volatilities, respectively. Similarly, $z_t^{C|P}$ is based on the density of call IV conditional on put IV, and vice versa for $z_t^{P|C}$.

Graphical analyses of the probability integral transforms are commonplace. These involve both a histogram of the transforms, that allows for determining uniformity, as well as autocorrelation functions of demeaned probability integral transforms and their squares. The graphical approach allows for easily identifying where a possible model misspecification arises. Figure 4.4 presents the 25-bin histogram and autocorrelations for $z_t^{C|P}$, and Figure 4.5 for $z_t^{P|C}$ for the unrestricted model. Figures 4.6 and 4.7 present the equivalent graphs for z_t^C and z_t^P, respectively.

Most columns of the histograms fall within the 95 % confidence interval, which is based on Pearson's goodness-of-fit test. Although there are some departures from the confidence bounds (between zero and four, depending on the case), there is no indication that the model would not be able to capture the tails of the conditional distribution properly. It must be noted that Pearson's test statistics and confidence interval are not exactly valid, as their calculation does not take estimation error into account. However, this omission most likely leads to rejecting too frequently.

The autocorrelations of the demeaned probability integral transforms also provide encouraging evidence, although some rejections do occur at the five percent (but not at the ten percent) level.[7] There clearly seems to be some remaining autocorrelation in the squares of the demeaned probability integral transforms. This same finding has been made previously with univariate models for volatility data (see Ahoniemi (2007) and Lanne (2006, 2007)). A potential explanation is that the model is not quite sufficient in capturing the time-varying volatility of implied volatility, as autocorrelation in the squares is a sign of conditional heteroskedasticity.

[7] The confidence bands of the autocorrelations are also calculated without estimation error accounted for. Adding fourth, fifth, or tenth lags of IV to the model does not improve the autocorrelation diagnostics.

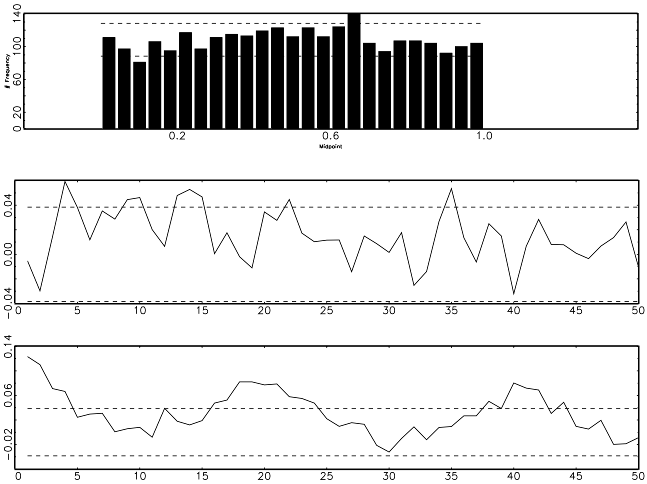

Figure 4.4: Diagnostic evaluation of $z_t^{C|P}$: NIKC conditional on NIKP. Histograms of probability integral transforms in the upper panel, and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

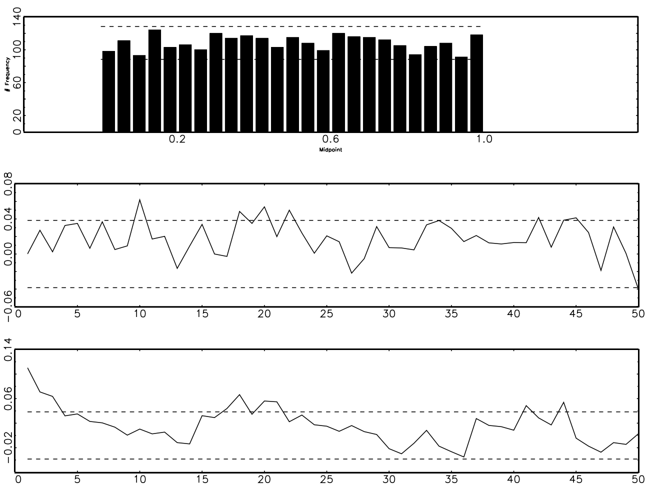

Figure 4.5: Diagnostic evaluation of $z_t^{P|C}$: NIKP conditional on NIKC. Histograms of probability integral transforms in the upper panel, and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

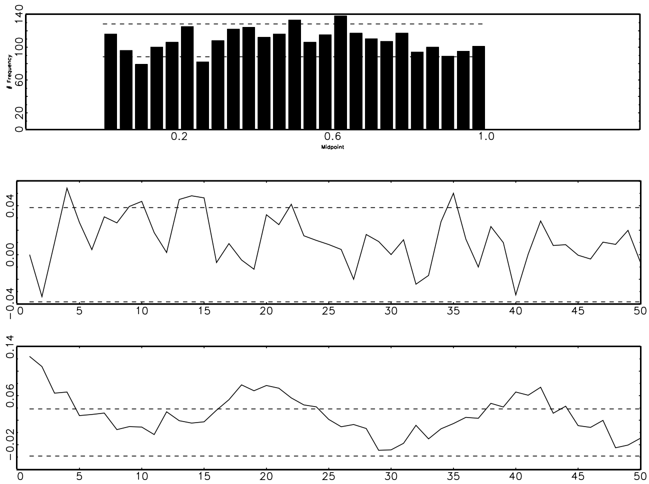

Figure 4.6: Diagnostic evaluation of z_t^C: Marginal NIKC. Histograms of probability integral transforms in the upper panel, and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

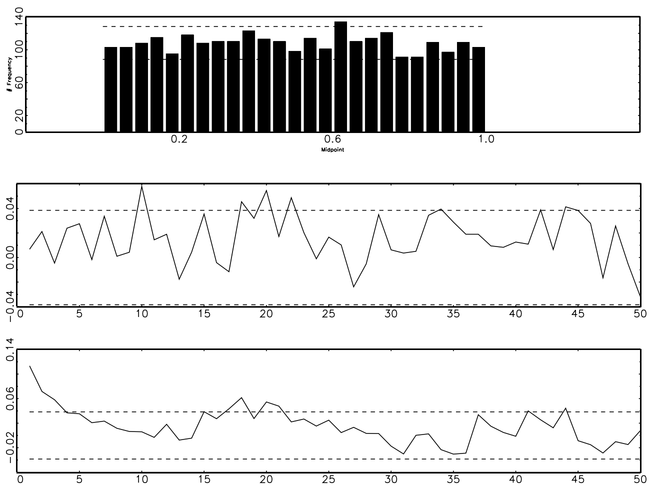

Figure 4.7: Diagnostic evaluation of z_t^P: Marginal NIKP. Histograms of probability integral transforms in the upper panel, and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

The removal of dummy variables from the unrestricted model results in a clear deterioration in the autocorrelation diagnostics, and consequently, we have deemed the inclusion of weekly seasonality effects relevant for our model. The diagnostics for the restricted model, or the model without cross effects, are somewhat better than those for the unrestricted model, especially where autocorrelations are concerned.[8] The improvement in diagnostics due to the removal of cross effects is surprising, as the cross terms are statistically significant and improve forecasts (see Section 4.6 for discussion on forecasts).

In order to verify that our unrestricted model takes the high persistence in the data into account, we compare the autocorrelation functions (ACF) estimated from the call and put IV data to those calculated from data simulated with our model. Figure 4.8 depicts the autocorrelation functions of NIKC and NIKP, as well as the autocorrelation functions generated by the unrestricted mixture-BVMEM model after simulating 100,000 data points. A 95% confidence band is drawn around the estimated autocorrelation functions. The band is obtained by simulating 10,000 series of 3,194 data points (equal to the full sample size), and forming a band that encompasses 95% of the autocorrelations pointwise at each lag. The fact that the autocorrelation coefficients generated by our model fall within the band at each lag lends support to the observed ACFs having been generated by our mixture-BVMEM model.

The diagnostics underscore the necessity of using a mixture model in this case. We also estimated a BVMEM model with only one regime, and the diagnostic checks clearly reveal its inadequacy. In particular, the four histograms for that model show that this specification fails to account for the tails of the conditional distribution, giving too little weight to values close to zero and unity and too much weight to the mid-range of the distribution.[9] However, the imbalance in the histograms is not as severe as with the univariate models in Ahoniemi (2007), thus indicating that even for models without a mixture structure, joint modeling improves the fit to the Nikkei 225 IV data somewhat.

[8] To save space, the diagnostic graphs for the unrestricted model without dummy variables and for the restricted model are not presented in the paper, but are available from the authors upon request.

[9] The estimation results and diagnostic evaluation for the no-mixture model are available from the authors upon request.

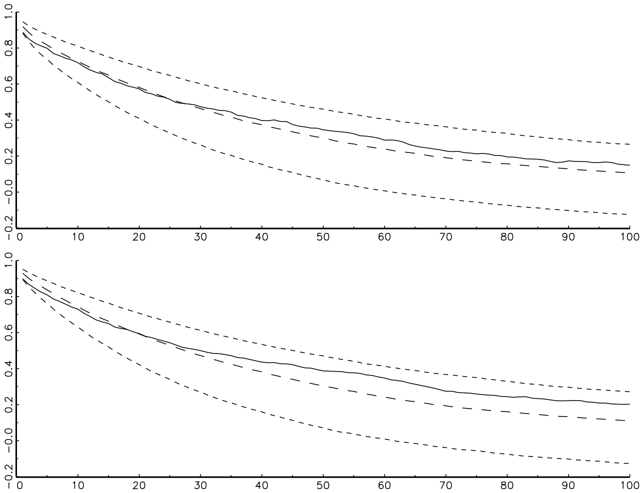

Figure 4.8: NIKC (upper panel) and NIKP (lower panel) autocorrelation functions. The solid lines depict the ACFs estimated from the full sample of the data, the lines with long dashes are the ACFs implied by the mixture-BVMEM model with 100,000 simulated data points, and the lines with short dashes draw 95% confidence bands around the ACFs.

4.5 Impulse Response Analysis

The bivariate nature of our model allows for a further analysis of how the variables adjust dynamically to shocks. In order to investigate this issue, impulse responses of various types are calculated with both model specifications presented above. This analysis should also uncover more evidence pertaining to the persistence of the data. The more interesting specification for this purpose is naturally the unrestricted model, which includes lagged cross terms in the first regime. Also, as the coefficients of the cross terms are significant, the unrestricted model specification is favored over the restricted version.[10]

We generate the impulse responses by simulating data according to the conditional mean profiles method proposed by Gallant et al. (1993). It turns out that after approximately 40 periods, the effects of all considered shocks go to zero.

[10]Dummy variables are removed from the mean equations in the impulse response analysis. The removal of weekly seasonality does not affect the general shape of the impulse response functions, but makes results more easily readable from graphs.

Therefore, we present impulse responses up to 40 periods (trading days) ahead. The calculation of the impulse response functions proceeds as follows: we generate 1,000 series of 40 random error terms from gamma distributions with the shape and scale parameters estimated above. Also, we generate 1,000 series of 40 random numbers that are uniformly distributed on the interval [0,1]. These series are used in each period to determine which regime the model is in: if the value of the random number exceeds the value of π, the mean equation for the second regime is used. To get initial values, a starting point in the data set is chosen, and then 1,000 paths, forty days ahead into the future, are simulated from that point onwards with the random error terms, random regime indicators, and estimated parameter values. Another set of 1,000 paths are also simulated, this time with a shock added to the values of NIKC, NIKP, or both in time period 0. The baseline value and the value affected by the shock are calculated simultaneously, so that the same random error terms and regime indicators are used for both. The averages of the 1,000 realizations are taken for each of the forty days, and the impulse response function is then obtained as the difference between the series affected by the shocks and the baseline series without the shocks.

In order to select a realistic magnitude for the shocks, we follow Gallant et al. (1993) and study a scatter plot of demeaned NIKC and NIKP. The scatter plot, shown in Figure 4.9, helps to identify perturbations to NIKC and NIKP that are consistent with the actual data. As expected, the scatter plot reveals a strong correlation in the two time series. On the basis of the graphical analysis, three different plausible shock combinations are selected: (10,10), (10,0), and (0,10). All these points appear in the scatter plot, and moreover, 10 appears to be the highest magnitude that can be realistically paired with a shock of zero to the other series.[11] In other words, the shock is introduced directly into the value of NIKC or NIKP (or both), rather than into the error terms of the model.[12]

[11]Our model is linear in the sense that it does not allow positive and negative shocks to have effects of differing magnitude. Therefore, only the impulse responses to positive shocks are presented. The equivalent graphs with negative shocks are mirror images.

[12]We cannot retrieve the residuals from our model due to the mixture structure. Therefore, we use the demeaned observations in selecting the appropriate combinations of shocks. Because the implied volatilities are autocorrelated, obtaining the shocks this way should be seen as an approximation only. However, as a robustness check, we have also tried shocks of other magnitudes than the ones presented here, and the results are qualitatively the same. Also, as the

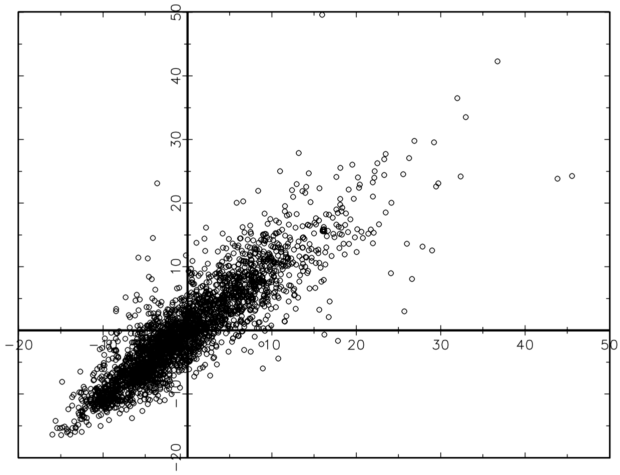

Figure 4.9: Scatter plot of demeaned NIKC (x-axis) and demeaned NIKP (y-axis) for in-sample period (1.1.1992 - 31.12.2002).

The impulse responses for the three shock combinations are presented in Figures 4.10, 4.11, and 4.12, respectively. The starting point in the data was July 11, 1996, a time when both NIKC and NIKP were historically quite low. Three noteworthy conclusions can be drawn from the analysis. Most importantly, put-side IV recovers from shocks more rapidly than call-side IV, which is evident in all three figures - even when the shock affects only put-side IV (Figure 4.12). This result could be based on the phenomenon documented by Bollen and Whaley (2004): the demand for ATM index puts drives the level of ATM implied volatility (in U.S. markets). Also, trading volumes for puts are higher (measured with number of contracts). Therefore, the pricing of puts may be somewhat more efficient, allowing shocks to persist for shorter periods of time than in a less efficient market.

Second, the effects of shocks take a relatively long time to disappear entirely: some thirty trading days, or six weeks, seem to elapse before the effect of a shock is completely wiped out. This finding gives further support to the existence of considerable persistence in the data, and is in line with evidence in Jorion (1995),

results are not sensitive to the point in time when the simulation is started, the scatter plot can be considered a reasonable approximation of the distribution of the true shocks.

where the half-life of a variance shock is estimated to be 17 days for the IV of currency options.

Third, the impulse responses are similar regardless of the starting point that is selected from the data. Four different starting points were in fact considered: a moment when both IVs were low, a moment when both were high, a moment when NIKC was considerably higher than NIKP, and a moment when NIKP was considerably higher than NIKC. This third result is to be expected as the nonlinearity in the mixture-BVMEM model arises primarily through the mixture of two regimes, with the selection of the mixture component being random rather than dependent on the past values of the implied volatilities.

Without the evidence on historical values provided by the scatter plot, it could be argued that a shock of the type (10,0), or any shock with a clearly different magnitude for call and put IVs, is not realistic. As outlined above, both IVs represent the market's expectation of future volatility, and should thus be equal. However, empirical analyses again lend support to the fact that call and put IV can differ even considerably at times, due to market imperfections and demand shocks. As an example, the difference between NIKP and NIKC is greatest on Sept. 12, 2001, or immediately after the 9/11 terrorist attacks, when the demand for put options was extremely high. On that day, the difference between the put and call implied volatilities was 33.6.

With the restricted model, or the model without cross effects, the impulse responses look very different. The effect of a shock lasts for less than ten days, or less than two weeks. As there are no lagged cross terms in this model specification, if a shock affects only one variable, the other is (naturally) entirely unaffected and the impulse response is flat.

4.6 Forecasts

In this section, we turn our attention to the forecasting ability of the two models outlined in Section 4.4.2. Forecast evaluation is based on two separate criteria: the direction of change in IV as well as the traditional forecast accuracy measure mean squared error (MSE). It is of particular interest whether the inclusion of cross effects between the two time series can improve the earlier forecast performance

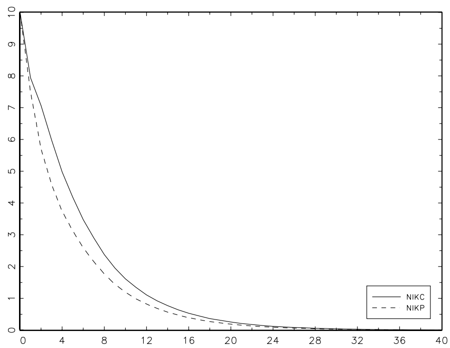

Figure 4.10: Impulse response function for a shock of (10,10) with the unrestricted model.

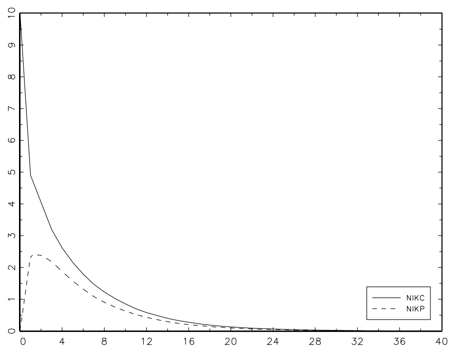

Figure 4.11: Impulse response function for a shock of (10,0) with the unrestricted model.

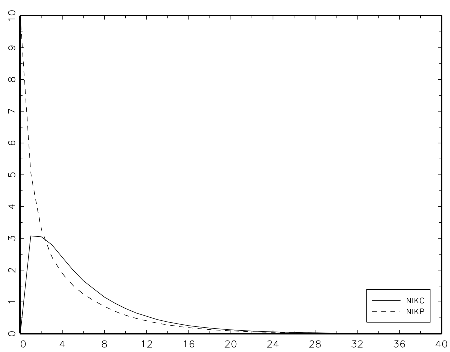

Figure 4.12: Impulse response function for a shock of (0,10) with the unrestricted model.

of univariate models for NIKC and NIKP investigated by Ahoniemi (2007).

Both daily (one-step-ahead) and five-step-ahead forecasts were calculated with the mixture-BVMEM model specifications outlined above for the 486-day out-of-sample period of 1.1.2003 - 31.12.2004. Days when public holidays fall on weekdays and the observed value of implied volatility does not change were omitted from the data set. Parameter values are treated in two ways: they are either estimated once using the data from the in-sample period and then kept fixed, or re-estimated each day. If parameter values are not stable over time, there can be added value in updating them before calculating each new forecast. When parameter values are updated daily, the forecasts are calculated from rolling samples. In other words, the first observation is dropped and a new one added each day, in order to include information that is as relevant as possible. In this case, the number of observations remains the same as in the in-sample, or 2,708.

The one-step-ahead forecasts are evaluated in terms of both directional accuracy and MSE. Although an accurate forecast of the future level of IV can be valuable to all market participants with risk management concerns, a correct forecast of the direction of change in implied volatility can be useful for option traders. Various option spreads, such as the straddle, can yield profits for the trader if the view on direction of change (up or down) is correct, ceteris paribus.

The forecast results are summarized in Table 4.3. The directional accuracy of the bivariate model in the two-year out-of-sample is superior to the performance of univariate models. The BVMEM model predicts the direction of change correctly on 348 days out of 486 for NIKC, and on 351 days for NIKP. This is in contrast to the results in Ahoniemi (2007), where the best figures from multiplicative models were 336 and 321 for NIKC and NIKP, respectively. There would appear to be some value to updating parameter values each day. This improves directional accuracy for NIKC clearly, and yields a lower mean squared error for both series of forecasts. However, the direction of change is predicted correctly for NIKP on one day more when daily updating is not employed. Both the unrestricted and the restricted model make more upward mistakes for NIKC, i.e. the models make a prediction of an upward move too often. When predicting the direction of change of NIKP, the unrestricted model forecasts a move downwards too often, but the restricted model a move upwards too often.

A useful statistical test of the sign forecasting ability of the BVMEM models is the test statistic presented in Pesaran & Timmermann (1992). This market timing test can help confirm that the percentage of correct sign forecasts is statistically significant. The p-values from the test are below 0.00001 for all the series of forecasts in Table 4.3, so the null hypothesis of predictive failure can be rejected at the one-percent level for all four forecast series.

The values for MSE in Table 4.3 indicate that more accurate forecasts can be obtained for NIKC. Mean squared errors are lower than with the univariate models in Ahoniemi (2007), even with the restricted model. The Diebold-Mariano (1995) test (henceforth the DM test) confirms that the improvement upon the equivalent univariate models is statistically significant for the unrestricted model, with the null hypothesis of equal predictive accuracy rejected at the five-percent level for NIKC and the one-percent level for NIKP. Again, this lends support to the joint modeling of the time series, and the inclusion of cross effects. The forecast accuracy of NIKC with updating coefficients is significantly better than that with fixed coefficients. For NIKP, the difference between the two alternative treatments of parameter values is not statistically significant.[13]

	NIKC			NIKP		
	Correct sign	*%*	*MSE*	*Correct sign*	*%*	*MSE*
Unrest., updating	**348**	**71.6%**	**4.21**	350	72.0%	**5.24**
Unrest., fixed	341	70.2%	4.31	**351**	**72.2%**	5.26
Rest., updating	332	68.3%	4.34	336	69.1%	5.49
Rest., fixed	332	68.3%	4.39	332	68.3%	5.55

Table 4.3: Correct sign predictions (out of 486 trading days) and mean squared errors for forecasts from the BVMEM model with both updating and fixed parameter values. The best values within each column are in boldface.

Overall, the results obtained for the Nikkei 225 index option market are superior to those obtained for e.g. the U.S. market. Ahoniemi (2008) finds that ARIMA

[13] A bivariate model specification without cross effects and dummy terms is not a better forecaster than univariate models, regardless of whether directional accuracy or MSE is used as the measure of forecast performance. For example, the directional accuracy of this model is 332 out of 486 at best for NIKC, and 318 for NIKP. The detailed results for this third model specification are available from the authors.

models can predict the correct direction of change in the VIX index on 58.4 percent of trading days at best with an identical out-of-sample period. Harvey and Whaley (1992) achieve a directional accuracy of 62.2% for IV from call options on the S&P 100 index, and 56.6% for the corresponding put IV. Brooks and Oozeer (2002) model the implied volatility of options on Long Gilt futures that are traded in London. Their model has a directional accuracy of 52.5 %. Our earlier discussion on the effects of limits to arbitrage could perhaps explain why Japanese IV is more predictable in sign than the IV in other markets. If arbitrage is more difficult to carry out in Japan, option prices can depart from their true values to a greater degree, making the market more forecastable.

Table 4.4 presents the MSEs for the 482 five-step-ahead forecasts that could be calculated within the chosen out-of-sample. The unrestricted model continues to be the better forecaster for NIKP, but surprisingly, the simpler model, or the specification without cross effects, yields lower MSEs for NIKC. The Diebold-Mariano test also rejects the null of equal forecast accuracy at the ten-percent level when comparing the restricted and unrestricted models with updating coefficients for NIKC, but not at the five-percent level. For the corresponding NIKP values (8.79 and 9.36), the null is not rejected. The results for NIKP are better than in Ahoniemi (2007), but for NIKC, the univariate models provide lower mean squared errors. The DM test does not reject the null when the best MSEs from univariate and bivariate models are compared (this applies to both NIKC and NIKP). Therefore, no conclusive evidence is provided regarding the best forecast model for a five-day horizon, but in statistical terms, the BVMEM model is at least as good as univariate models.

	NIKC	NIKP
Unrestricted Model, updating	7.21	8.79
Unrestricted Model, fixed	7.65	**8.78**
Restricted Model, updating	**6.48**	9.36
Restricted Model, fixed	6.60	9.43

Table 4.4: Mean squared errors for 482 five-step-ahead forecasts from the BVMEM model with both updating and fixed parameter values. The best values within each column are in boldface.

4.7 Conclusions

Empirical research on implied volatilities typically seeks to determine whether or not implied volatility is an unbiased and efficient forecaster of future realized volatility. Time series modeling and forecasting of implied volatility is a less explored, but clearly relevant topic, as implied volatility is widely accepted to be the market's best forecast for future volatility. Also, professional option traders can potentially benefit from accurate IV forecasts.

It has often been empirically observed that implied volatilities calculated from otherwise identical call and put options are not equal. Market imperfections and demand pressures can make this phenomenon allowable, and this paper seeks to answer the question of whether call and put IVs can be jointly modeled, and whether joint modeling has any value for forecasters.

We show that the implied volatilities of Nikkei 225 index call and put options can be successfully jointly modeled with a mixture bivariate multiplicative error model, using a bivariate gamma error distribution. Diagnostics show that the joint model specification is a good fit to the data, and coefficients are statistically significant. Two mixture components are necessary to fully capture the characteristics of the data set, so that days of large and small shocks are modeled separately. There are clear linkages between the implied volatilities calculated from call and put option prices, as lagged cross terms are statistically significant. The IV derived from put options is a more important driving factor in our model than the IV from calls, as dummy variables for Friday effects of put-side IV are revealed to be significant and to improve the diagnostics of the joint model.

Impulse response analysis indicates that put-side IV recovers more quickly from shocks than call-side IV. Shocks persist for a relatively lengthy period of time (thirty trading days), which is consistent with good forecastability. Also, as the nonlinear feature of our model is primarily the random switching between regimes, the point of time in which a shock is introduced does not affect the behavior of the impulse response functions.

The BVMEM model provides better one-step-ahead forecasts than its univariate counterparts. Both directional accuracy and mean squared errors improve when jointly modeling call and put implied volatility. The direction of change in implied volatility is correctly forecast on over 70% of the trading days in our two-year out-of-sample period. When forecasting five trading days ahead, the BVMEM model is at least as good as univariate models in statistical terms. Based on the combined evidence from all forecast evaluations, we conclude that joint modeling and the inclusion of cross effects improves the forecastability of Nikkei 225 index option implied volatility, and can provide added value to all investors interested in forecasting future Japanese market volatility.

The results of this paper suggest several viable alternatives for future research. An extended version of this model could be fit to the implied volatility of stock index options from other markets. Also, an option trading simulation using Nikkei 225 index option quotes could be run based on the directional forecasts of the model in order to determine whether or not the model's trading signals could lead to abnormal returns.

Bibliography

[1] Ahoniemi, K. (2007). Multiplicative Models for Implied Volatility. HECER Discussion Papers, No. 172.

[2] Ahoniemi, K. (2008). Modeling and Forecasting the VIX Index. Unpublished working paper.

[3] Bates, D.S. (1991). The Crash of '87: Was It Expected? The Evidence from Options Markets. Journal of Finance, 46, 1009–1044.

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

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

[6] Brunetti, C. & Lildholdt, P.M. (2007). Time Series Modeling of Daily Log-Price Ranges for CHF/USD and USD/GBP. Journal of Derivatives, 15:2, 39–59.

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

[8] Chan, K.C., Cheng, L.T.W. & Lung, P.P. (2004). Net Buying Pressure, Volatility Smile, and Abnormal Profit of Hang Seng Index Options. Journal of Futures Markets, 24, 1165–1194.

[9] Cipollini, F., Engle, R.F. & Gallo, G.M. (2006). Vector Multiplicative Error Models: Representation and Inference. NBER Working Paper, No. 12690.

[10] Davies, R.B. (1977). Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative. Biometrika, 64, 247–254.

[11] 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.

[12] Diebold, F.X., Gunther, T.A. & Tay, A.S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, 39, 863–883.

[13] Diebold, F.X., Hahn, J. & Tay, A.S. (1999). Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-frequency Returns on Foreign Exchange. Review of Economics and Statistics, 81, 661–673.

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

[15] Dumas, B., Fleming, F. & Whaley, R.E. (1998). Implied Volatility Functions: Empirical Tests. Journal of Finance, 53, 2059–2106.

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

[17] Engle, R.F. (2002). New Frontiers for ARCH Models. Journal of Applied Econometrics, 17, 425–446.

[18] Engle, R.F. & Gallo, G.M. (2006). A Multiple Indicators Model for Volatility Using Intra-daily Data. Journal of Econometrics, 131, 3–27.

[19] Figlewski, S. (1989). Options Arbitrage in Imperfect Markets. Journal of Finance, 44, 1289–1311.

[20] Fleming, J. (1999). The Economic Significance of the Forecast Bias of S&P 100 Index Option Implied Volatility. In: Boyle, P., Pennacchi, G. & Ritchken, P. (Eds), Advances in Futures and Options Research, Vol. 10. Stamford, Conn: JAI Press, pp.219–251.

[21] Gallant, A.R., Rossi, P.E. & Tauchen, G. (1993). Nonlinear Dynamic Structures. Econometrica, 61, 871–907.

[22] Garleanu, N.B., Pedersen, L.H. & Poteshman, A.M. (2006). Demand-based Option Pricing. CEPR Discussion Papers, No. 5420.

[23] 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.

[24] Hentschel, L. (2003). Errors in Implied Volatility Estimation. Journal of Financial and Quantitative Analysis, 38, 779–810.

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

[26] Jorion, P. (1995). Predicting Volatility in the Foreign Exchange Market. Journal of Finance, 50, 507–528.

[27] Lanne, M. (2006). A Mixture Multiplicative Error Model for Realized Volatility. Journal of Financial Econometrics, 4, 594–616.

[28] Lanne, M. (2007). Forecasting Realized Exchange Rate Volatility by Decomposition. International Journal of Forecasting, 23, 307–320.

[29] Lehmann, B.N. & Modest, D.M. (1994). Trading and Liquidity on the Tokyo Stock Exchange: A Bird's Eye View. Journal of Finance, 49, 951–984.

[30] Liu, J. & Longstaff, F.A. (2004). Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities. Review of Financial Studies, 17, 611–641.

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

[32] Mo, H. & Wu, L. (2007). International Capital Asset Pricing: Evidence from Options. Journal of Empirical Finance, 14, 465–498.

[33] Nagao, M. & Kadoya, M. (1970). Study on Bivariate Gamma Distribution and its Engineering Application (1). Annals of the Disaster Prevention Research Institute of Kyoto University, 13B, 105–115 (in Japanese with English synopsis).

[34] Nelson, D.B. & Cao, C.Q. (1992). Inequality Constraints in the Univariate GARCH Model. Journal of Business & Economic Statistics, 10, 229–235.

[35] 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.

[36] Peña, I., Rubio, G. & Serna, G. (1999). Why do we smile? On the Determinants of the Implied Volatility Function. Journal of Banking & Finance, 23, 1151–1179.

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

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

[39] Rubinstein, M. (1994). Implied Binomial Trees. Journal of Finance, 49, 771–818.

[40] Shleifer, A. & Vishny, R.W. (1997). The Limits of Arbitrage. Journal of Finance, 52, 35–55.

[41] Yue, S., Ouarda, T.B.M.J. & Bobée, B. (2001). A Review of Bivariate Gamma Distributions for Hydrological Application. Journal of Hydrology, 246, 1–18.

Chapter 5

Implied Volatility with Time-Varying Regime Probabilities

Katja Ahoniemi and Markku Lanne[1]

Abstract
This paper presents a mixture multiplicative error model with a time-varying probability between regimes. We model the implied volatility derived from call and put options on the USD/EUR exchange rate. The daily first difference of the USD/EUR exchange rate is used as a regime indicator, with large daily changes signaling a more volatile regime. Separate mean equations and error distributions are estimated for each regime. Forecasts indicate that it is beneficial to jointly model the implied volatilities derived from call and put options: both mean squared errors and directional accuracy improve when employing a bivariate rather than a univariate model. In a two-year out-of-sample period, the direction of change in implied volatility is correctly forecast on two thirds of the trading days.

[1]Correspondence: katja.ahoniemi@hse.fi and markku.lanne@helsinki.fi. Financial support from the Okobank Group Research Foundation is gratefully acknowledged. Katja Ahoniemi also thanks the Finnish Doctoral Programme in Economics, the Finnish Foundation for Advancement of Securities Markets, and the Yrjö Jahnsson Foundation.

5.1 Introduction

Implied volatility (IV) forecasting in foreign exchange markets is a relevant issue for numerous market participants. Not only are accurate implied volatility forecasts valuable for option traders and market makers who price derivatives, they can benefit investors facing risk management issues in international portfolios. As argued by Pong et al. (2004), option traders may possess additional information about future events on top of what is provided by historical data, making implied volatilities potentially more accurate forecasts of future volatility than those based on history alone.

The traditional vein of foreign exchange IV literature investigates how well IV forecasts future realized volatility. Pong et al. (2004) report that historical forecasts contain substantial incremental information beyond that included in IV for short forecast horizons of exchange rate realized volatility, crediting this result to the use of high-frequency data. Jorion (1995) and Xu and Taylor (1995) provide the opposite result, with IV from foreign exchange options yielding more accurate forecasts. In another study using high-frequency returns, Martens and Zein (2004) conclude that for options on USD/JPY futures, long-memory forecasts from fractionally integrated models and implied volatility both possess information that the other does not contain. Guo (2000) links volatility forecasting to the option market by investigating returns from trading foreign exchange options, with option strategies based on conditional volatility predictions stemming from either IV or a GARCH model specification. This study concludes that trading is more profitable when IV is used to forecast future volatility.

The level and behavior of asset prices and their volatilities can be very different when comparing times of relative calm to times of distress or panic in the market. Both exchange rate returns and their volatility have been found to exhibit regime-switching properties. It is therefore easy to surmise that the implied volatility of options on foreign exchange could also possess the same property. If a time series follows two or more distinct regimes, good model fit and forecasting performance will require a specification that allows for different conditional means and error distributions for each regime.

Studies that document regime shifts in exchange rates include Engel and Hamilton (1990), Bekaert and Hodrick (1993), and Engel and Hakkio (1996). Engel and Hamilton (1990) observe that exchange rate returns vary more when the U.S. dollar is appreciating, and that a regime-switching model outperforms a random-walk model as a forecaster. Bekaert and Hodrick (1993) estimate a two-regime model and find that the variances of forward premiums are nine to ten times larger in the more volatile regime. In Engel and Hakkio (1996), the difference between variances of exchange rate returns in the stable and volatile states is even larger.

Bollen et al. (2000) estimate a Markov-switching model with two regimes for log exchange rate changes and their variances, but with mean and variance regimes allowed to switch independently. This modification allows for high volatility in times of both appreciation and depreciation of an exchange rate. The model produces more accurate variance forecasts than competing models, and estimation results indicate that the standard deviations of exchange rate returns are two to three times higher in the more volatile regime. Klaassen (2002) shows that a regime-switching GARCH model outperforms a single-regime GARCH model in terms of mean squared forecast errors.

Both regime-switching and mixture models introduce separate regimes for times of business as usual and times of higher volatility. In essence, a model with various regimes is an alternative to the GARCH class of models: volatility clustering is now modeled through changes in the variance of the underlying data generating process, with switches in regime corresponding to these changes (see Bollen et al. (2000)).[2] Mixture multiplicative error models with two regimes have been found to fit time series of the realized volatility of exchange rates (Lanne (2006)) and the implied volatility of Nikkei 225 index options (Ahoniemi (2007) and Ahoniemi and Lanne (2009)).

In this study, we present an extension of a mixture multiplicative error model that allows for a time-varying probability between regimes, with the mixing probabilities depending on an observed variable that functions as a regime indicator.

[2]The literature is rife with examples of GARCH modeling of exchange rate series, Baillie and Bollerslev (1991) and Hansen and Lunde (2005) among them. Rapach and Strauss (2008) find strong evidence of structural breaks in GARCH models for seven exchange rate series of the U.S. dollar against other currencies, with substantial differences in unconditional variance between regimes. In their out-of-sample forecast exercise, models with structural breaks provide lower mean squared errors than competing models. Klaassen (2002) and Haas et al. (2004) present Markov-switching GARCH models with promising results for exchange rate series.

Our data set consists of the implied volatilities calculated from call and put options on the USD/EUR exchange rate, and USD/EUR exchange rate returns serve as our indicator variable. Although implied volatilities calculated from call and put options with the same strike prices and maturity dates should be equal, in practice we observe that that is rarely the case. Differences in call and put IVs can arise due to differing demand pressures and limits to arbitrage, as explained in e.g. Garleanu et al. (2006), Bollen and Whaley (2004), and Figlewski (1989).

The aim of this study is to find a good model fit with time-varying regime probabilities and to generate forecasts of the IV of USD/EUR currency options that succeed both in predicting the direction of change of implied volatility and its value. We model call and put IV separately with two-regime and three-regime model specifications, but also estimate a bivariate specification that jointly models both call and put implied volatilities. In this latter model, we include cross terms that allow call-side IV to affect put-side IV, and vice versa. The bivariate specification proves to be valuable, as the directional accuracy and mean squared errors of forecasts both improve with the bivariate model. The correct direction of change in IV is forecast on two thirds of the 514 trading days in the two-year out-of-sample period. Two regimes are sufficient in the bivariate model, but it is not evident whether two or three regimes should be preferred in the univariate setting.

This paper proceeds as follows. Section 5.2 describes the univariate and bivariate multiplicative error models with time-varying regime probabilities. Section 5.3 presents the data, the model estimation results, and diagnostics. Section 5.4 discusses forecast results and their evaluation, with performance measured with directional accuracy and mean squared errors. Section 5.5 concludes the paper.

5.2 The Model

In this section, we describe univariate and bivariate mixture multiplicative error models, and present an extension with multiple mixture components and time-varying mixing probabilities. Henceforth, this new model will be called the TVMEM model.

Multiplicative error models are suitable for modeling any time series that always receives non-negative values. This particular study focuses on volatility, but the same type of model could be used in, for example, duration or trading volume applications. Engle (2002) first suggested applying MEM models to volatility modeling, pointing out that with a MEM model, there is no need to take logarithms of the data. Since then, promising results with MEM models in volatility applications have been obtained by e.g. Ahoniemi (2007), Ahoniemi and Lanne (2009), Engle and Gallo (2006), and Lanne (2006). In a standard multiplicative error model volatility v_t is specified as

$$v_t = \mu_t \varepsilon_t, \qquad t = 1, 2, ..., T \tag{5.1}$$

where ε_t is a stochastic non-negative error term, and the conditional mean μ_t is specified as

$$\mu_t = \omega + \sum_{i=1}^{q} \alpha_i v_{t-i} + \sum_{j=1}^{p} \beta_j \mu_{t-j}. \tag{5.2}$$

As the mean equation is structured in the same way as in GARCH models, the parameter values must comply with the restrictions for GARCH models outlined in Nelson and Cao (1992) in order to ensure positivity. In the simplest setting, or with a first-order model, all parameter values must be non-negative. With higher-order models, non-negativity is not always required.[3]

We extend the basic MEM model by considering a mixture-MEM model. In a mixture model, the data is allowed to follow i regimes with $i \geq 2$. This type of model specification lends itself particularly well to financial time series data, as financial markets typically alternate between periods of high and low volatility. With possibly different mean equations and error distributions for each mixture component, the variations in the market in calm and more volatile time periods can be more thoroughly captured. When estimating a mixture-MEM model, one must also estimate the probability of each regime (π_i) along with the other parameters.

Lanne (2006), Ahoniemi (2007), and Ahoniemi and Lanne (2009) all estimate a mixture-MEM model with a fixed mixing probability. In other words, one value of π_i is estimated in their models, and the probability of each regime is thus the same

[3]For example, in a model with $p = 1$ and $q = 2$, the constraints are $\omega \geq 0$, $\alpha_1 \geq 0$, $0 \leq \beta < 1$, and $\beta\alpha_1 + \alpha_2 \geq 0$.

every day. We extend the basic mixture-MEM model by allowing the probability of the regimes, π_i, to vary over time. In practice, the values of the parameters that determine π_i are estimated from the data, and a value for π_i is generated for every time period.

The regime probability varies in time according to a regime indicator. We reason that when modeling implied volatility, the returns of the options' underlying asset would work well as an indicator: a large change in the value of the underlying can be taken as a signal of higher volatility in the market. We later confirm the functioning of such an indicator graphically by plotting the IV data and the regime probabilities together (see Section 5.3.2). It is also important to note that absolute values of the returns are used when estimating a model with two regimes: this allows both large positive and negative returns to be properly accounted for. When estimating a model with three regimes, the first differences are used as such, with one threshold separating large negative shocks, and a second threshold separating large positive shocks. In a two-regime model, we expect the largest probability mass to fall into the first regime of calmer days, whereas with three regimes, the middle regime should capture the largest number of trading days.

When using an indicator variable to determine the thresholds between regimes, strict thresholds that allow for no gray area around the value of the threshold are often estimated. However, we introduce an additional error term, η_t, into the indicator function

$$I(c_{i-1} + \eta_t \leq y_{t-d} < c_i + \eta_t) \tag{5.3}$$

with $\eta_t \sim NID(0, \sigma_\eta^2)$ and independent of ε_t. This idea was previously applied to autoregressive models by Lanne and Saikkonen (2003). The indicator function divides the observations onto both sides of the thresholds, but with the addition of the unobservable white noise term η_t, variation around the threshold value is introduced. In other words, it is not clear which regime the model is in when the indicator variable receives a value that is close to the threshold. Also, a switch in regime does not always occur when the value of the indicator variable crosses the threshold. This element brings additional flexibility into the model, which is beneficial as models with regime switches based on a strict threshold may not always be realistic (Lanne and Saikkonen (2003)). The volatility of η_t, σ_η, is later estimated as one of the parameters of our model and used as follows when determining the daily regime probabilities $\pi_{i,t-d}$

$$\pi_{i,t-d} = \begin{cases} 1 - \Phi((y_{t-d} - c_1)/\sigma_\eta), & i = 1 \\ \Phi((y_{t-d} - c_{i-1})/\sigma_\eta) - \Phi((y_{t-d} - c_i)/\sigma_\eta), & i = 2, \ldots, m-1 \\ \Phi((y_{t-d} - c_{m-1})/\sigma_\eta), & i = m \end{cases} \tag{5.4}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (cdf) of the standard normal distribution, m is the number of mixture components (regimes), σ_η denotes the standard deviation of η_t and satisfies $\sigma_\eta > 0$, and $c_1, \ldots, c_{m-1}$ denote the values of the threshold parameters and satisfy $c_1 < \ldots c_{m-1}$. The larger the volatility parameter σ_η, the larger is the range around the estimated threshold where a regime switch is probable. One common value of σ_η is estimated for all thresholds. In practice, the time index $t-d$ is always $t-1$ in our application, as further lags would not be relevant in a financial application.

We assume that the error term of our TVMEM model follows the gamma distribution. The gamma distribution has the benefit that it is very flexible and nests e.g. the exponential and χ^2 distributions. Also, earlier work by Lanne (2006), Ahoniemi (2007) and Ahoniemi and Lanne (2009) shows that the gamma distribution is suitable for volatility modeling purposes. Our requirement that the error terms have mean unity means we must impose the restriction that the shape and scale parameters of the gamma distribution are reciprocals of each other. Under this restriction, the conditional density of v_t with two mixture components (regimes) is

$$\begin{aligned} f_{t-1}(v_t;\theta) \;=\; & \pi_1 \frac{1}{\mu_{1t}\Gamma(\lambda_1)\delta_1^{\lambda_1}} \left(\frac{v_t}{\mu_{1t}}\right)^{\lambda_1 - 1} \exp\left(-\frac{v_t}{\delta_1 \mu_{1t}}\right) \;+ \\ & (1-\pi_1) \frac{1}{\mu_{2t}\Gamma(\lambda_2)\delta_2^{\lambda_2}} \left(\frac{v_t}{\mu_{2t}}\right)^{\lambda_2 - 1} \exp\left(-\frac{v_t}{\delta_2 \mu_{2t}}\right) \end{aligned} \tag{5.5}$$

where θ is the vector of parameters, $\Gamma(\cdot)$ is the gamma function, λ_1 and λ_2 are the shape parameters of the gamma distribution, and δ_1 and δ_2 are the scale parameters.

As a further extension of the univariate TVMEM model, we also consider a bivariate model that allows us to jointly model the implied volatilities garnered from call and put options. As the evidence in Ahoniemi and Lanne (2009) indicates, there can be clear added value to modeling call and put IVs jointly. In the bivariate case, the mean equations include cross terms that allow call-side IV to affect put-side IV, and vice versa. The mean equations in our bivariate TVMEM model are

$$\mu_{mt}^{C} = \omega_m^C + \sum_{i=1}^{q_C} \alpha_{mi}^C v_{C,t-i} + \sum_{i=1}^{r_C} \psi_{mi}^C v_{P,t-i} + \sum_{j=1}^{p_C} \beta_{mj}^C \mu_{m,t-j}^C$$

and

$$\mu_{mt}^{P} = \omega_m^P + \sum_{i=1}^{q_P} \alpha_{mi}^P v_{P,t-i} + \sum_{i=1}^{r_P} \psi_{mi}^P v_{C,t-i} + \sum_{j=1}^{p_P} \beta_{mj}^P \mu_{m,t-j}^P$$

where μ_{mt}^C and μ_{mt}^P are the conditional means of call-side and put-side implied volatility, respectively, and the ψ's are the coefficients for lagged cross terms.

There are a number of bivariate gamma distributions with gamma marginals (see Yue et al. (2001) for a review). From the available bivariate distributions, we select the specification suggested by Nagao and Kadoya (1970), as it is a tractable alternative. The density function in this case can be written as

$$f_{\varepsilon_1,\varepsilon_2}\left(\varepsilon_{1t},\varepsilon_{2t};\boldsymbol{\theta}\right) = \frac{\left(\tau_1\tau_2\right)^{(\lambda+1)/2}\left(\varepsilon_{1t}\varepsilon_{2t}\right)^{(\lambda-1)/2}\exp\left\{-\frac{\tau_1\varepsilon_{1t}+\tau_2\varepsilon_{2t}}{1-\rho}\right\}}{\Gamma\left(\lambda\right)\left(1-\rho\right)\rho^{(\lambda-1)/2}} I_{\lambda-1}\left(\frac{2\sqrt{\tau_1\tau_2\rho\varepsilon_{1t}\varepsilon_{2t}}}{1-\rho}\right), \tag{5.6}$$

where ρ is the Pearson product-moment correlation coefficient, and $I_{\lambda-1}(\cdot)$ is the modified Bessel function of the first kind. The ρ's reflect the correlation of shocks between the two time series and are constrained to be between zero and unity. The fact that ρ cannot be negative should not be restrictive, as the underlying asset of the options in our application is the same. In the bivariate application, both error distributions have a distinct scale parameter. However, the shape parameter is the same for both time series, which makes the scale parameters also equal: with the parametrization of Equation 5.6, the scale and shape parameters must be equal (rather than reciprocals) to ensure mean unity. Given the earlier evidence in Ahoniemi (2007) and the results for univariate models in this paper, this requirement is not restrictive: while the shape (and scale) parameters tend to differ clearly between regimes, they are very close in value between the call and put IV time series.

For the conditional density function of $\mathbf{v}_t = (v_{1t}, v_{2t})'$, we use the change of variable theorem and make the substitution $\tau_1 = \tau_2 = \lambda$ to obtain

$$\begin{aligned} f_{t-1}\left(v_{1t}, v_{2t}; \boldsymbol{\theta}\right) &= f_{\varepsilon_1,\varepsilon_2}\left(v_{1t}\mu_{1t}^{-1}, v_{2t}\mu_{2t}^{-1}\right)\mu_{1t}^{-1}\mu_{2t}^{-1} \\ &= \frac{\lambda^{(\lambda+1)}\left[v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}\right]^{(\lambda-1)/2}\exp\left\{-\frac{\lambda\left(v_{1t}\mu_{1t}^{-1}+v_{2t}\mu_{2t}^{-1}\right)}{1-\rho}\right\}}{\Gamma\left(\lambda\right)\left(1-\rho\right)\rho^{(\lambda-1)/2}} \times \\ & \quad I_{\lambda-1}\left(\frac{2\lambda\sqrt{\rho v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}}}{1-\rho}\right)\mu_{1t}^{-1}\mu_{2t}^{-1}. \end{aligned} \tag{5.7}$$

The conditional log-likelihood function for the two-regime case can thus be written as

$$l_T\left(\boldsymbol{\theta}\right) = \sum_{t=1}^{T} l_{t-1}\left(\boldsymbol{\theta}\right) = \sum_{t=1}^{T} \ln\left[\pi_{i,t-d} f_{t-1}^{(1)}\left(v_{1t}, v_{2t}; \boldsymbol{\theta}_1\right) + \left(1-\pi_{i,t-d}\right) f_{t-1}^{(2)}\left(v_{1t}, v_{2t}; \boldsymbol{\theta}_2\right)\right],$$

where $f_{t-1}^{(1)}\left(v_{1t}, v_{2t}; \boldsymbol{\theta}_1\right)$ and $f_{t-1}^{(2)}\left(v_{1t}, v_{2t}; \boldsymbol{\theta}_2\right)$ are given by (5.7) with $\boldsymbol{\theta}$ replaced by $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. All the TVMEM models described in this section can be estimated with the method of maximum likelihood (ML). The asymptotic properties of the ML estimators are not known; however, assuming stationarity and ergodicity, it is reasonable to apply standard asymptotic results in statistical inference.

5.3 Estimation Results

5.3.1 Data

The data set used in this study includes two time series of at-the-money (ATM), 30-day maturity implied volatility data: one calculated from call options on the USD/EUR exchange rate, and the other from put options (see Figure 5.1). Each observation is interpolated from the implied volatilities of four different options: two nearest-to-the-money strikes are used, one above and one below the current spot rate, and maturities are selected from both sides of the 30-day interval. As noted by Ederington and Guan (2005), at-the-money options are typically the most liquid, have the smallest bid-ask spreads, and are most commonly used in empirical research. The use of ATM options helps in avoiding the effects of the volatility smile. Also, at-the-money options have the highest sensitivity to volatility (Bollen & Whaley (2004)), and IV is more likely to equal the mean expected volatility over the remaining life of the option when ATM options are used (Day and Lewis (1992)). A 30-day maturity is commonly used in established IV indices such as the VIX index.

The implied volatilities of the options are calculated using the Black-Scholes extension for currency options, namely the Garman-Kohlhagen model.[4] The data is obtained from Datastream, and includes daily observations for the time period 1.8.2000-31.12.2007. Days when public holidays fall on weekdays are omitted from the data set. The full sample covers 1,904 trading days, 1,390 of which are treated as the in-sample. The last two years, or 514 days, are left as an out-of-sample period and used for forecast assessment. Our data set also includes the WM/Reuters daily closing spot rates for the EUR/USD exchange rate, which are obtained from Datastream. The reciprocals of these rates are used, as the underlying asset of the options is the USD/EUR rate. The raw IV data is multiplied by 100 before performing any estimations, and log exchange rate returns are also multiplied by 100.

The correlation of the call (USD/EUR C) and put (USD/EUR P) implied volatility time series is 84.4%. Basic descriptive statistics for the series are provided

[4]Campa et al. (1998) report that in the over-the-counter currency option market, the convention is to quote prices as Garman-Kohlhagen implied volatilities.

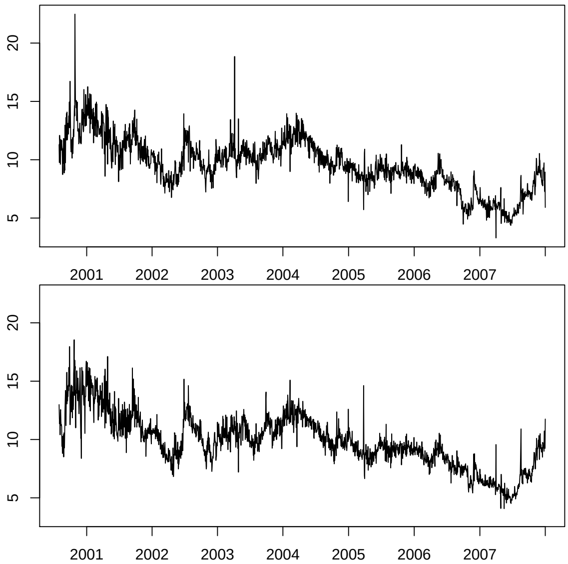

Figure 5.1: USD/EUR exchange rate call option implied volatility (upper panel) and put option implied volatility (lower panel) 1.8.2000 - 31.12.2007.

in Table 5.1, and autocorrelation functions in Figure 5.2. Table 5.1 shows that the values for the two implied volatilities clearly differ: although there is no great difference in the means, the maximum values of the two time series are far apart: 22.5 for USD/EUR C and 18.5 for USD/EUR P. The maximum value of USD/EUR C is higher, but its minimum value is also lower. Both time series are slightly skewed to the right.

Table 5.1 also provides descriptive information on the USD/EUR exchange rate return series. The largest positive daily return in our full sample is 3.3%, and the largest negative return is -2.3%. The return distribution is more kurtotic than the IV distributions. The autocorrelation functions of USD/EUR C and USD/EUR P reveal that there is high persistence in the data, although augmented Dickey-Fuller tests indicate that both series are stationary (a unit root is rejected at the one-percent level for USD/EUR C and the five-percent level for USD/EUR P).

	USD/EUR C	USD/EUR P	USD/EUR return
Maximum	22.481	18.549	3.321
Minimum	3.296	4.075	-2.266
Mean	9.604	9.872	0.024
Median	9.583	9.790	0.018
Standard deviation	2.225	2.308	0.578
Skewness	0.144	0.225	0.113
Kurtosis	3.536	3.169	4.052

Table 5.1: Descriptive statistics for USD/EUR C, USD/EUR P, and the log first differences of the USD/EUR exchange rate for the full sample of 1.8.2000 - 31.12.2007. Raw IV data and log exchange rate returns are all multiplied by 100.

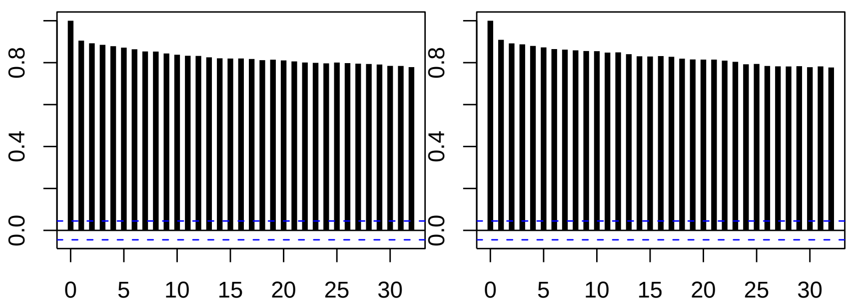

Figure 5.2: Autocorrelation function of USD/EUR exchange rate call option implied volatility (left panel) and put option implied volatility (right panel).

5.3.2 Model Estimation

We estimate three competing model specifications: two-regime and three-regime univariate models as well as a two-regime bivariate model.[5] In the two-regime models, the regime indicator variable is the absolute value of the lagged log return of the USD/EUR exchange rate, and in the three-regime models, we use the lagged log first difference as such. Thus, in the two-regime case we hypothesize that both large positive and large negative shocks have the same dynamics, whereas in the three-regime case, both tails of the return distribution are allowed to provide separate mean equations and error distributions. Kim and Kim (2003) corroborate our view that exchange rate returns could function well as a regime indicator by providing evidence that the implied volatility of currency options on futures is higher when currency futures returns are large. They also note that both positive and negative movements in the currency futures price contribute to increases in IV. However, there is evidence that IV reacts differently to positive and negative return shocks, at least in equity markets. In a study on S&P 500 index options, Bakshi and Kapadia (2003) look at implied volatility changes after large positive and large negative index returns. After a large negative return shock, the relative change in IV was 10.6% on average, and after a large positive return shock, the average change was -1.5%.

Table 5.2 shows the parameter values for the two-regime univariate models. The estimate of the threshold parameter c_1 is very similar for both call and put implied volatilities: 1.83 for the call IV model and 1.81 for the put IV model. The volatility parameter σ_η equals 0.6 for both models. In other words, the time-varying regime probability, which is modeled by using exchange rate returns as a regime indicator, behaves in a very similar fashion for both call-side and put-side IV. The range within which a regime switch is likely to occur is centered on 1.8% returns of the exchange rate. In a hypothetical model with $\sigma_\eta = 0$, when the absolute return is greater than 1.8%, the data is assumed to be drawn from

[5] Estimation of a four-regime univariate model shows that three regimes is sufficient: the first threshold reaches its lower limit of 2.26, which is the smallest exchange rate return in our sample. Thus, no observations would be left for a fourth regime. Similarly, the results for a three-regime bivariate model indicated that a two-regime model is sufficient in the bivariate case: none of the coefficients of the two high-volatility regimes were statistically significant.

the second regime. However, given our estimated range around the threshold, the regime switch does not automatically occur when the absolute exchange rate return exceeds the estimated threshold value. The standard errors of σ_η are very small, which lends support to this specification over the alternative where $\sigma_\eta = 0$.

The error distributions of the two regimes are also nearly identical for two-regime univariate call and put IV models, as the estimated shape parameters are very close in value for both time series. Figure 5.3 depicts the error distributions of the two-regime call IV model (the figure is qualitatively the same for the put IV model). The residuals of the models are more concentrated around their mean in the first, more common regime, and more wide-tailed in the more volatile regime.

The constant terms are, somewhat surprisingly, higher in the more common regime for both the call and put models. In fact, there are no constants in the second regime as their values are estimated to be zero (in practice, 1e-5, the lower limit). As there are no radical differences in the α or β coefficients either, the largest difference between the two regimes arises from the difference in error distributions.

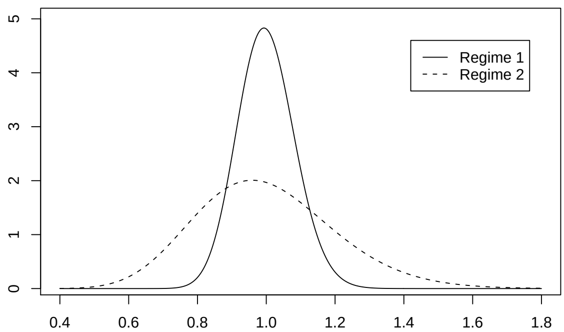

Figure 5.3: Estimated density of error terms of two-regime univariate model for USD/EUR exchange rate call option implied volatility.

We next estimate a univariate model specification with three regimes. We wish to check whether the use of more than two regimes is necessary, and in the absence of a formal test for the optimal number of regimes, explore the issue by estimating both alternative specifications. Parameter values for the three-regime models are presented in Table 5.3. The estimated threshold values are now

	USD/EUR C		USD/EUR P	
Log likelihood	-1669.52		-1713.31	
c_1	1.826**	(0.178)	1.806**	(0.190)
σ_η	0.601**	(0.093)	0.600**	(0.093)
λ_1	145.72**	(0.003)	145.67**	(0.003)
ω_1	0.142**	(0.054)	0.304**	(0.056)
α_{11}	0.333**	(0.032)	0.238**	(0.019)
α_{12}	-0.080	(0.045)	-	
β_{11}	0.733**	(0.034)	0.727**	(0.022)
λ_2	24.46**	(7.870)	25.90**	(8.440)
ω_2	-		-	
α_{21}	0.393	(0.269)	0.288	(0.228)
β_{21}	0.625*	(0.257)	0.715**	(0.225)

Table 5.2: Estimation results for two-regime univariate TVMEM models. Standard errors calculated from the final Hessian matrix are given in parentheses. ** indicates statistical significance at the one-percent level, and * at the five percent level. Note that for σ_η, zero is on the boundary of the parameter space so that statistical significance cannot be interpreted in the standard way. Constants are excluded from the second regime as their estimated value was at the lower limit (1e-5). Second lags of IV values are excluded on the basis of likelihood ratio tests with the exception of the more common regime for USD/EUR C.

distinctly different for USD/EUR C and USD/EUR P, with the regime of large negative shocks receiving clearly more weight for call IV (greater value of c_1 for call side). On the other hand, the regime of large positive shocks receives more weight with put IV (smaller value of c_2 for put side). These large weights coincide with gamma distribution shape parameters that are similar in size to those of the most common regimes. In other words, a more dispersed error distribution occurs only when the threshold value leaves a relatively small amount of days into a regime (in practice, the positive-shock regime of USD/EUR C and the negative-shock regime of USD/EUR P).

	USD/EUR C		USD/EUR P	
Log likelihood	-1676.77		-1705.64	
c_1	-0.932	(0.676)	-1.714**	(0.103)
c_2	1.997**	(0.230)	1.201**	(0.278)
σ_η	0.989**	(0.144)	0.641**	(0.072)
λ_1	146.47**	(0.070)	20.28**	(6.660)
ω_1	0.009	(0.058)	-	
α_{11}	0.109	(0.058)	0.412	(0.305)
β_{11}	0.889**	(0.059)	0.591	(0.303)
λ_2	145.67**	(0.450)	147.30**	(0.003)
ω_2	0.251*	(0.121)	0.386**	(0.072)
α_{21}	0.370**	(0.062)	0.264**	(0.022)
β_{21}	0.606**	(0.069)	0.691**	(0.027)
λ_3	22.81**	(7.610)	145.47**	(1.450)
ω_3	-		-	
α_{31}	0.462	(0.288)	0.109**	(0.034)
β_{31}	0.549	(0.282)	0.892**	(0.035)

Table 5.3: Estimation results for three-regime univariate TVMEM models. Standard errors calculated from the final Hessian matrix are given in parentheses. ** indicates statistical significance at the one-percent level, and * at the five percent level. Note that for σ_η, zero is on the boundary of the parameter space so that statistical significance cannot be interpreted in the standard way.

The estimate of the volatility parameter σ_η is larger for USD/EUR C in the three-regime model, with the same value of σ_η applying to both thresholds. As with the two-regime models, constants are highest in the most common regime, and omitted due to the lower limit being reached in three of the four remaining cases. The highest persistence in the models emerges in the negative-shock regime for USD/EUR C and the positive-shock regime for USD/EUR P, with β coefficient values of 0.89.[6] In general, the parameter values of all three regimes differ clearly from one another for both the call and put IV models, so that the use of three regimes rather than two seems justified. However, we will calculate forecasts from the two-regime models as well, as we are interested in discovering the best model specification for forecasting purposes.

The coefficients for the two-regime bivariate model specification are shown in Table 5.4. The threshold value c_1 is similar to that of univariate two-regime models, but the volatility parameter σ_η is somewhat smaller. Due to the distributional properties of the bivariate gamma distribution, the upper bound of the shape parameter is lower than in the univariate case, making the error distribution slightly less concentrated around unity than in the univariate case (see Figure 5.4 for the densities of the residuals with the bivariate model). The correlation of shocks, ρ, is at its lower bound for both regimes, which yields the somewhat surprising interpretation that the shocks of USD/EUR C and USD/EUR P are quite uncorrelated with each other.[7] The constant term is higher in the higher-volatility regime for USD/EUR P, which is in line with the result we originally expected to see in all cases. The more volatile regime is more persistent, with β coefficients clearly higher in the second regime for both call and put IV. Cross terms are significant only in the more common regime.

Based on the coefficients, call-side IV affects put-side IV slightly more than the other way around. This result is in contrast to results achieved earlier for equity

[6]The sum of the α and β coefficients is slightly greater than unity in nearly all cases for the more volatile regimes, indicating explosive dynamics. However, a visual inspection of very long simulated data series shows that the data process with our models is not explosive. Alos, the means and medians of the simulated data series are close in value to those of the original data set.

[7]This may be a consequence of the fact that the underlying asset of the options in this study is an exchange rate, which naturally involves two currencies. Investors with positions in either the USD or the EUR may have differing exposures in the option market, with demand for calls and demand for puts originating from different sets of investors. Depending on which currency a shock affects more, the shock can thus affect call and put options (and their implied volatilities) differently.

index option implied volatility (see e.g. Ahoniemi and Lanne (2009), Bollen and Whaley (2004)), as in those markets, trading in put options tends to dominate developments in the overall market. With currency options, the demand of hedgers is not as clearly concentrated on put options, so the demand and supply balances in the markets of exchange rate call and put options may be more even than for equity index call and put option markets.[8]

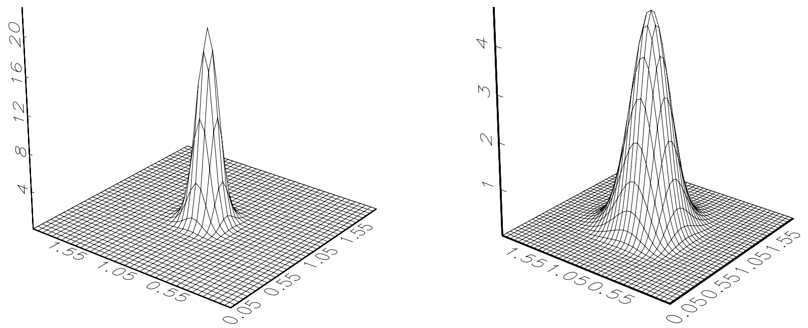

Figure 5.4: Estimated density of error terms of two-regime bivariate model: first regime in left panel and second regime in right panel.

Figures 5.5 and 5.6 plot the estimated regime probabilities and IV data in the same graphs. This setting allows us to see how the probabilities vary over time, but also to verify that the regime indicator is functioning properly: the regime probability, though calculated solely from exchange rate returns, will preferably have a stronger reaction on days when there is a large change in the value of IV. The regime probability should also react when the level of IV is relatively high. As can be seen from Figures 5.5 and 5.6, this relation holds quite well, so we are satisfied with the performance of the exchange rate returns as our regime indicator.[9]

[8]The significance of Friday and Monday dummy variables for lagged IV observations was also investigated. Earlier evidence indicates that there is a weekly pattern in IV (see e.g. Ahoniemi and Lanne (2009), Kim and Kim (2003), Harvey and Whaley (1992)). In particular, Kim and Kim (2003) use exchange rate options on futures data and discover that IV is low on Mondays and tends to rise by Wednesday, remaining high for the rest of the week. No weekday effects were found for the data sample used in this study, however.

[9]As regards the necessity of time-varying regime probabilities, it must be noted that with this particular data set, a fixed regime probability that would not vary over time was not sufficient.

	Bivariate, 2 regimes	
Log likelihood	-3323.84	
c_1	1.551**	(0.080)
σ_η	0.412**	(0.043)
λ_1	127.29	(-)
ρ_1	0.01	(-)
ω_1^C	0.312**	(0.114)
α_{11}^C	0.325**	(0.034)
ψ_{11}^C	0.238**	(0.029)
β_{11}^C	0.401**	(0.056)
ω_1^P	0.036	(0.085)
α_{11}^P	0.369**	(0.038)
ψ_{11}^P	0.270**	(0.035)
β_{11}^P	0.364**	(0.062)
λ_2	28.72**	(6.65)
ρ_2	0.01	(-)
ω_2^C	-	
α_{21}^C	0.138	(0.101)
ψ_{21}^C	-	
β_{21}^C	0.868**	(0.097)
ω_2^P	0.833	(0.939)
α_{21}^P	0.248	(0.208)
ψ_{21}^P	-	
β_{21}^P	0.696**	(0.224)

Table 5.4: Estimation results for the bivariate TVMEM model. Standard errors calculated from the final Hessian matrix are given in parentheses. Standard errors of (-) indicate that a parameter value is at a predefined limit. λ_1 reaches its upper limit, beyond which the likelihood function is no longer numerically tractable. Both ρ_1 and ρ_2 reach their lower limit of 0.01, which was set at slightly above the theoretical lower limit of zero. ** indicates statistical significance at the one-percent level. Note that for σ_η, zero is on the boundary of the parameter space so that statistical significance cannot be interpreted in the standard way.

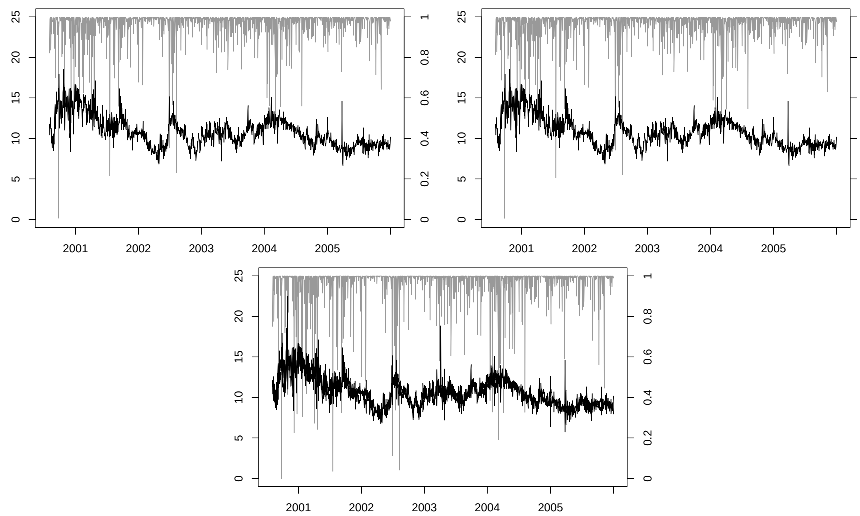

Figure 5.5: Implied volatility on left axis and probability of first regime on right axis. Two-regime model for USD/EUR C on left in upper row, two-regime model for USD/EUR P on right in upper row, and two-regime bivariate model in lower row.

Figure 5.5 depicts the probabilities of the first, more common regime together with the original IV data for all three two-regime models (univariate and bivariate). The average probability of the more common regime is 96.4% for USD/EUR C, 96.2% for USD/EUR P, and 96.0% for the bivariate model. Looking at the average probability alone, all models would seem to leave a relatively small share of days for the more volatile regime. However, the graphs show that on days of large moves in the market, the probability of the more common regime falls significantly, even to close to zero.

The probabilities obtained from the three-regime models are detailed in Figure 5.6. A more pronounced difference between the call and put IV time series now emerges, as was already evident from the threshold values presented in Table 5.3. With call-side IV, the regime of negative shocks receives an average probability of 20.8%, the regime of positive shocks has an average probability of 4.5%, with 74.7% left for the most common regime. For put-side IV, the corresponding averages are

For example with the bivariate specification, a fixed-probability model was not able to distinguish separate regimes from the data.

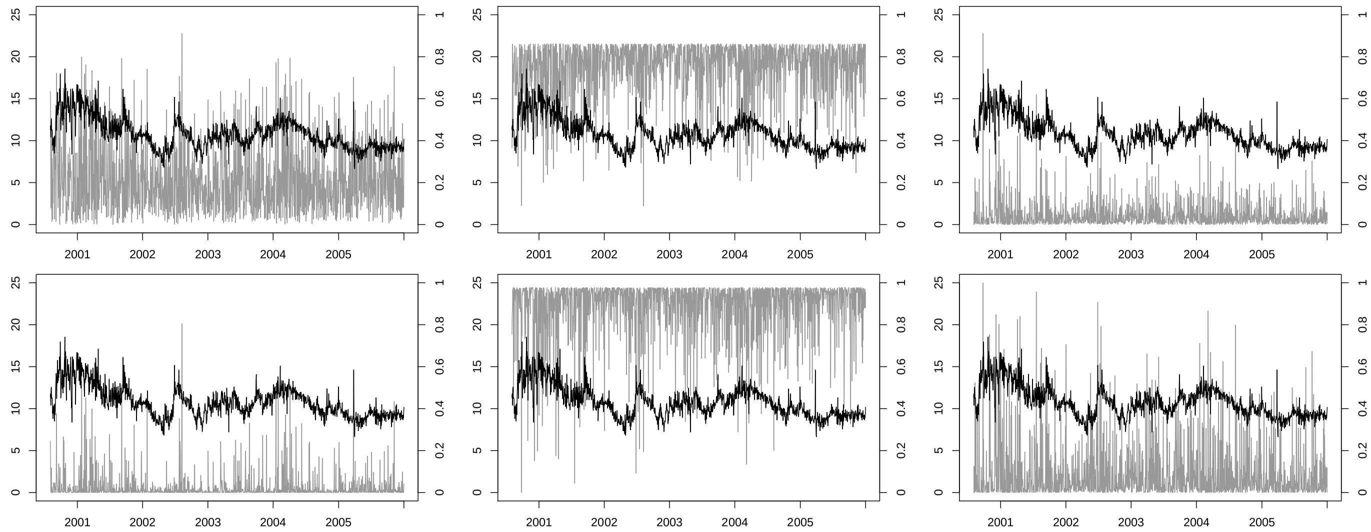

Figure 5.6: Three-regime model for USD/EUR C (top row) and USD/EUR P (bottom row): implied volatility on left axis and probability on right axis. Probability of first regime in left panel, probability of second regime in middle panel, and probability of third regime in right panel.

2.7%, 9.1%, and 88.2%.

5.3.3 Diagnostics

Due to the fact that our models have two or three mixture components and thus do not produce conventional (Pearson) residuals, we cannot perform standard diagnostic checks for the models' goodness-of-fit. As residuals cannot be obtained in any straightforward way, we use probability integral transforms of the data to evaluate the models' in-sample fit through autocorrelation diagnostics. This approach was suggested by Diebold et al. (1998) and extended to the multivariate case by Diebold et al. (1999). For a univariate model, the probability integral transforms are calculated as

$$z_t = \int_0^{v_t} f_{t-1}(u)du \tag{5.8}$$

where $f_{t-1}(\cdot)$ is the conditional density of the data with the chosen model specification. This procedure transforms each data point into a value between zero and unity. Following Diebold et al. (1999), we calculate four series of probability integral transforms for the bivariate case: z_t^C and z_t^P (based on the marginal densities of call and put IV), and $z_t^{C|P}$ and $z_t^{P|C}$ (call IV conditional on put IV, and vice versa).

Graphs of the autocorrelation functions of demeaned probability integral transforms and their squares are presented in Figure 5.7 for the two-regime and three-regime univariate models and in Figure 5.8 for the bivariate model. The autocorrelations of the demeaned probability integral transforms in Figure 5.7 are extremely satisfactory for USD/EUR C, but leave some to be desired for USD/EUR P.[10] In line with previous research, there clearly seems to be some remaining autocorrelation in the squares of the demeaned probability integral transforms (this same finding has been made for volatility data in Ahoniemi and Lanne (2009) and Lanne (2006, 2007)). This autocorrelation in squares is a signal of conditional heteroskedasticity in the data, so it may be that the multiplicative model structure

[10]Note that the confidence bands of the autocorrelations are not exactly valid, as estimation error is not taken into account. However, this most likely leads to too frequent rejections, so the diagnostics might in fact be slightly better than those shown here.

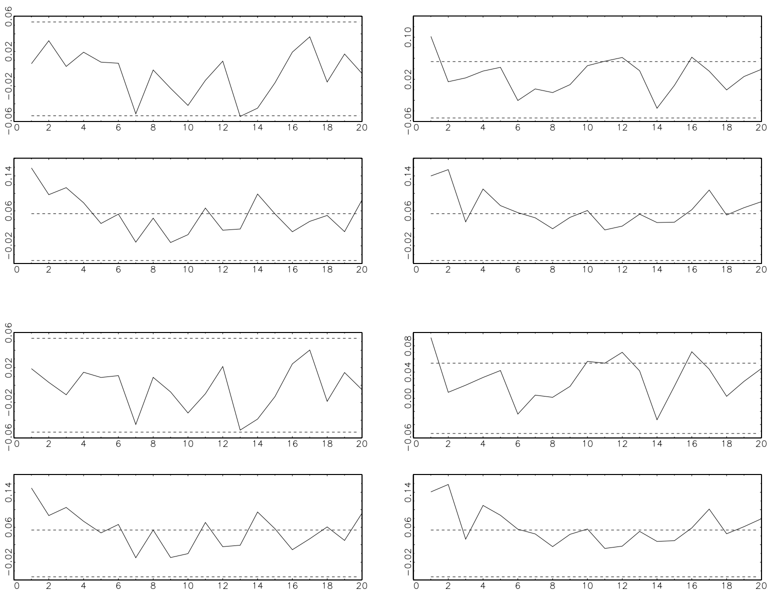

Figure 5.7: Diagnostics for two-regime model of USD/EUR C (top left), two-regime model of USD/EUR P (top right), three-regime model of USD/EUR C (bottom left), and three-regime model of USD/EUR P (bottom right). Autocorrelation functions of demeaned probability integral transforms in the upper panels and of their squares in the lower panel. The dotted lines depict the boundaries of the 95% confidence interval.

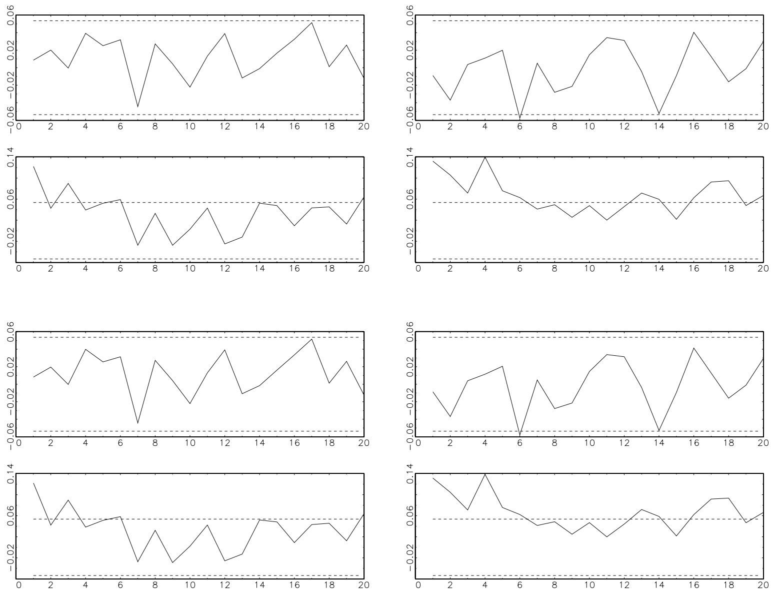

Figure 5.8: Diagnostics for two-regime bivariate model. Top left: z_t^C, top right: z_t^P, bottom left: $z_t^{C|P}$, bottom right: $z_t^{P|C}$. Autocorrelation functions of demeaned probability integral transforms in upper panel and of their squares in lower panel. The dotted lines depict the boundaries of the 95% confidence interval.

is not sufficient in capturing the time-varying volatility of USD/EUR exchange rate option implied volatility.[11] We witness an improvement when turning our attention to the bivariate model diagnostics in Figure 5.8. All autocorrelations of demeaned probability integral transforms now nearly fall within the confidence bands.

5.4 Forecasts

In this section, we proceed to calculate forecasts from the three models presented above. Daily one-step-ahead forecasts are obtained for the out-of-sample period

[11] As Corsi et al. (2008) note, realized volatility models may be subject to heteroskedastic errors due to time-varying volatility in the realized volatility estimator. This evidence can possibly also be extended to implied volatility.

of 1.1.2006-31.12.2007, amounting to 514 forecasts. Forecasts are calculated using the parameter values presented in Section 5.3.2. Forecast evaluation involves two separate paths: directional accuracy and mean squared errors (MSE). We are interested in the direction of change of implied volatility due to volatility's important role in option pricing. Option traders can potentially profit if they have the correct view on the direction of IV. We use MSEs to assess the relative performance of the models in generating point forecasts, which are of interest in e.g. portfolio risk management.

Table 5.5 summarizes the forecast results. The results clearly indicate that the model of choice is the bivariate model: both directional accuracy and MSEs improve when switching from a univariate to a bivariate model. For USD/EUR C, the number of days with a correct prediction of sign improves from 301 to 326 when employing the bivariate model, providing a correct sign on 63.4% of trading days. For USD/EUR P, the result is even more impressive: an improvement from 318 to 346, which translates to the correct sign on 67.3% of trading days in the out-of-sample period. Any result above 50% could potentially yield profits to a trader. Mean squared errors are also lowest with the bivariate model specification.

	USD/EUR C			USD/EUR P		
	Correct sign	*%*	*MSE*	*Correct sign*	*%*	*MSE*
Univariate, 2 regimes	295	57.4%	0.358	318	61.9%	0.383
Univariate, 3 regimes	301	58.6%	0.357	313	60.9%	0.389
Bivariate, 2 regimes	**326**	**63.4%**	**0.335**	**346**	**67.3%**	**0.303**

Table 5.5: Correct directional forecasts (out of 514 trading days) and mean squared errors for forecasts from both univariate and bivariate TVMEM models. The best values within each column are in boldface.

When comparing the two-regime and three-regime univariate models, it is not obvious which model specification is superior. For USD/EUR C, the three-regime model is a better forecaster of the direction of change, but for USD/EUR P, the two-regime model dominates. The rank of the models is also the same when using MSE as the criterion. In general, it can be noted that the TVMEM models forecast up too often: the models make more mistakes by predicting a move up when the true direction was down than vice versa. This tendency could exaggerate forecast performance in periods of consistently rising IV, but in our particular out-of-sample period, there are 243 moves up and 271 down for call IV, and 249 moves up and 265 down for put IV. The best balance by far is achieved with the bivariate model for USD/EUR P, with 279 forecasts up and 235 down. The bivariate model is also the most balanced for USD/EUR C, but the hits are more off: 358 up forecasts and 156 down forecasts.

We next run two types of tests in order to determine the statistical significance of the models' differences in predictive ability. First, we calculate Pesaran-Timmermann test statistics for the directional forecasts.[12] This test allows us to verify that the sign predictions of the forecast series outperform a coin flip. The null hypothesis of predictive failure can be rejected at all relevant levels of significance for all of our models, as the p-values of the test statistic are all less than 0.00001.

The Diebold-Mariano test (from Diebold and Mariano (1995)) is next used to see whether the improvement in mean squared errors achieved with the bivariate model is statistically significant. When comparing the MSE from the bivariate model to the best univariate model (three regimes for USD/EUR C and two regimes for USD/EUR P), the null hypothesis of equal predictive accuracy can be rejected at the one-percent level for USD/EUR P. However, for USD/EUR C, the test delivers a p-value of 0.16. Therefore, in statistical terms, the bivariate model specification does not deliver a better MSE than the best univariate model.

Compared to earlier studies that explore directional forecast accuracy, these results fall short of only those of Ahoniemi and Lanne (2009), who calculate the correct direction of change for Nikkei 225 index option implied volatility on at best 72% of trading days. In other work on IV modeling, Ahoniemi (2008) obtains the correct sign for the VIX index on 58.4% of trading days with an ARIMA model, and Harvey and Whaley (1992) achieve an accuracy of 62.2% (56.6%) for the implied volatility of S&P 100 index call (put) options. Brooks and Oozeer (2002), who model the implied volatility of options on Long Gilt futures, report a correct sign prediction on 52.5% of trading days. The fact that put IV is more predictable than call IV for USD/EUR currency options contrasts the results of

[12] This test statistic is presented in Pesaran and Timmermann (1992).

Harvey and Whaley (1992), whereas Ahoniemi and Lanne (2009) have the same result.

In another comparison to earlier work, we choose to look at relative MSEs. For this purpose, we calculate root mean squared errors (RMSE). The average value of USD/EUR C in the out-of-sample period is 7.12, and the best RMSE is some 8.1% of that. For USD/EUR P, the corresponding percentage is 7.5%. In the bivariate model for Nikkei 225 index option implied volatilities of Ahoniemi and Lanne (2009), call-side RMSE amounted to 9.6% of the average out-of-sample value, and put-side IV had a relative RMSE of 10.2%. Therefore, the point forecasts generated by the TVMEM model seem to be quite accurate.

5.5 Conclusions

Existing research has documented that exchange rate returns and volatilities appear to be drawn from several regimes. It is therefore plausible that the modeling of implied volatilities of currency options would benefit from allowing for two or more regimes. The results of this paper lend support to that assumption: for the IV of options on the USD/EUR exchange rate, a mixture multiplicative error model can be successfully fit to the data. Moreover, the regime probabilities vary in time, which proves to be a valuable feature of the model. Both two and three-regime models are viable in a univariate setting, but when the IV time series are jointly modeled, two mixture components are sufficient. The daily returns of the underlying exchange rate series function well as a regime indicator: days of large moves in the exchange rate seem to coincide with days of high levels of IV and days of large shifts in IV.

The bivariate model specification emerges as the model of choice for a forecaster. Both directional accuracy and point forecasts improve when moving from a univariate to a bivariate setting. The bivariate model predicts the direction of change in implied volatility correctly on 63% (67%) of trading days for call (put) options. This information can be particularly useful for traders in currency options, as the hit ratio is well over 50%. Also, at-the-money options, on which the data is based, are particularly sensitive to changes in volatility. Whether or not the forecast results could be profitably exploited in an out-of-sample option trading exercise is left for future research.

Bibliography

[1] Ahoniemi, K. (2007). Multiplicative Models for Implied Volatility. HECER Discussion Paper, 172.

[2] Ahoniemi, K. (2008). Modeling and Forecasting the VIX Index. Unpublished working paper.

[3] Ahoniemi, K. & Lanne, M. (2009). Joint Modeling of Call and Put Implied Volatility. International Journal of Forecasting, forthcoming.

[4] Baillie, R.T. & Bollerslev, T. (1991). Intra-Day and Inter-Market Volatility in Foreign Exchange Rates. Review of Economic Studies, 58, 565–585.

[5] Bakshi, G. & Kapadia, N. (2003). Delta-Hedged Gains and the Negative Market Volatility Risk Premium. Review of Financial Studies, 16, 527–566.

[6] Bauwens, L., Giot, P., Grammig, J. & Veredas, D. (2004). A Comparison of Financial Duration Models via Density Forecasts. International Journal of Forecasting, 20, 589–609.

[7] Bekaert, G. & Hodrick, R.J. (1993). On Biases in the Measurement of Foreign Exchange Risk Premiums. Journal of International Money and Finance, 12, 115–138.

[8] Bollen, N.P.B., Gray, S.F. & Whaley, R.E. (2000). Regime Switching in Foreign Exchange Rates: Evidence from Currency Option Prices. Journal of Econometrics, 94, 239–276.

[9] 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.

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

[11] Campa, J.M., Chang, P.H.K. & Reider, R.L. (1998). Implied Exchange Rate Distributions: Evidence from OTC Option Markets. Journal of International Money and Finance, 17, 117–160.

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

[13] Day, T.E. & Lewis, C.M. (1992). Stock market volatility and the information content of stock index options. Journal of Econometrics, 52, 267-287.

[14] Diebold, F.X., Gunther, T.A., & Tay, A.S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, 39, 863-883.

[15] Diebold, F.X., Hahn, J, & Tay, A.S. (1999). Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-frequency Returns on Foreign Exchange. The Review of Economics and Statistics, 81, 661–673.

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

[17] Ederington, L. & Guan, W. (2005). The Information Frown in Option Prices. Journal of Banking & Finance, 29, 1429–1457.

[18] Engel, C. & Hakkio, C.S. (1996). The Distribution of Exchange Rates in the EMS. International Journal of Finance & Economics, 1, 55–67.

[19] Engel, C. & Hamilton, J.D. (1990). Long Swings in the Dollar: Are They in the Data and Do Markets Know It?. American Economic Review, 80, 689–713.

[20] Engle, R.F. (2002). New Frontiers for ARCH Models. Journal of Applied Econometrics, 17, 425–446.

[21] Engle, R.F. & Gallo, G.M. (2006). A Multiple Indicators Model for Volatility Using Intra-daily Data. Journal of Econometrics, 131, 3–27.

[22] Figlewski, S. (1989). Options Arbitrage in Imperfect Markets. The Journal of Finance, 44, 1289–1311.

[23] Garleanu, N.B., Pedersen, L.H. & Poteshman, A.M. (2006). Demand-based Option Pricing. CEPR Discussion Paper, 5420.

[24] Gençay, R. (1998). Optimization of Technical Trading Strategies and the Profitability in Securities Markets. Economics Letters, 59, 249–254.

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

[26] Hansen, P.R. & Lunde, A. (2005). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?. Journal of Applied Econometrics, 20, 873–889.

[27] Haas, M., Mittnik, S. & Paolella, M.S. (2004). A New Approach to Markov-Switching GARCH Models. Journal of Financial Econometrics, 2, 493–530.

[28] 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.

[29] Jorion, P. (1995). Predicting Volatility in the Foreign Exchange Market. The Journal of Finance, 50, 507–528.

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

[31] Klaassen, F. (2002). Improving GARCH Volatility Forecasts with Regime-Switching GARCH. Empirical Economics, 27, 363–394.

[32] Lanne, M. (2006). A Mixture Multiplicative Error Model for Realized Volatility. Journal of Financial Econometrics, 4, 594–616.

[33] Lanne, M. & Saikkonen, P. (2003). Modeling the U.S. Short-Term Interest Rate by Mixture Autoregressive Processes. Journal of Financial Econometrics, 1, 96–125.

[34] Martens, M. & Zein, J. (2004). Predicting Financial Volatility: High-frequency Time-series Forecasts vis-à-vis Implied Volatility. Journal of Futures Markets, 24, 1005–1028.

[35] Nagao, M., & Kadoya, M. (1970). The Study on Bivariate Gamma Distribution and its Applicability (1). Annals of the Disaster Prevention Research Institute of Kyoto University 13B, 105–115 (in Japanese with English synopsis).

[36] Nelson, D.B. & Cao, C.Q. (1992). Inequality Constraints in the Univariate GARCH Model. Journal of Business & Economic Statistics, 10, 229–235.

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

[38] Pesaran, M.H., Timmermann, A.G. (1995). Predictability of Stock Returns: Robustness and Economic Significance. The Journal of Finance, 50, 1201–1228.

[39] Pong, S., Shackleton, M.B., Taylor, S.J. & Xu, X. (2004). Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models. Journal of Banking & Finance, 28, 2541–2563.

[40] Rapach, D.E. & Strauss, J.K. (2008). Structural Breaks and GARCH Models of Exchange Rate Volatility. Journal of Applied Econometrics, 23, 65–90.

[41] Xu, X. & Taylor, S.J. (1995). Conditional Volatility and the Informational Efficiency of the PHLX Currency Options Market. Journal of Banking & Finance, 19, 803–821.

[42] Yue, S., Ouarda, T.B.M.J. & Bobée, B. (2001). A Review of Bivariate Gamma Distributions for Hydrological Application. Journal of Hydrology, 246, 1–18.

HELSINGIN KAUPPAKORKEAKOULUN JULKAISUJA
Publications of the Helsinki School of Economics

A-SARJA: VÄITÖSKIRJOJA - DOCTORAL DISSERTATIONS. ISSN 1237-556X.

A:287. TIMO JÄRVENSIVU: Values-driven management in strategic networks: A case study of the influence of organizational values on cooperation. 2007. ISBN-10: 952-488-081-4, ISBN-13: 978-952-488-081-7.

A:288. PETRI HILLI: Riskinhallinta yksityisen sektorin työeläkkeiden rahoituksessa. 2007. ISBN-10: 952-488-085-7, ISBN-13: 978-952-488-085-5. E-version: ISBN 978-952-488-110-4.

A:289. ULLA KRUHSE-LEHTONEN: Empirical Studies on the Returns to Education in Finland. 2007. ISBN 978-952-488-089-3, E-version ISBN 978-952-488-091-6.

A:290. IRJA HYVÄRI: Project Management Effectiveness in Different Organizational Conditions. 2007. ISBN 978-952-488-092-3, E-version: 978-952-488-093-0.

A:291. MIKKO MÄKINEN: Essays on Stock Option Schemes and CEO Compensation. 2007. ISBN 978-952-488-095-4.

A:292. JAAKKO ASPARA: Emergence and Translations of Management Interests in Corporate Branding in the Finnish Pulp and Paper Corporations. A Study with an Actor-Network Theory Approach. 2007. ISBN 978-952-488-096-1, E-version: 978-952-488-107-4.

A:293. SAMI J. SARPOLA: Information Systems in Buyer-supplier Collaboration. 2007. ISBN 978-952-488-098-5.

A:294. SANNA K. LAUKKANEN: On the Integrative Role of Information Systems in Organizations: Observations and a Proposal for Assessment in the Broader Context of Integrative Devices. 2006. ISBN 978-952-488-099-2.

A:295. CHUNYANG HUANG: Essays on Corporate Governance Issues in China. 2007. ISBN 978-952-488-106-7, E-version: 978-952-488-125-8.

A:296. ALEKSI HORSTI: Essays on Electronic Business Models and Their Evaluation. 2007. ISBN 978-952-488-117-3, E-version: 978-952-488-118-0.

A:297. SARI STENFORS: Strategy tools and strategy toys: Management tools in strategy work. 2007. ISBN 978-952-488-120-3, E-version: 978-952-488-130-2.

A:298. PÄIVI KARHUNEN: Field-Level Change in Institutional Transformation: Strategic Responses to Post-Socialism in St. Petersburg Hotel Enterprises. 2007. ISBN 978-952-488-122-7, E-version: 978-952-488-123-4.

A:299. EEVA-KATRI AHOLA: Producing Experience in Marketplace Encounters: A Study of Consumption Experiences in Art Exhibitions and Trade Fairs. 2007. ISBN 978-952-488-126-5.

A:300. HANNU HÄNNINEN: Negotiated Risks: The Estonia Accident and the Stream of Bow Visor Failures in the Baltic Ferry Traffic. 2007. ISBN 978-952-499-127-2.

A-301. MARIANNE KIVELÄ: Dynamic Capabilities in Small Software Firms. 2007.
ISBN 978-952-488-128-9.

A:302. OSMO T.A. SORONEN: A Transaction Cost Based Comparison of Consumers' Choice between Conventional and Electronic Markets. 2007. ISBN 978-952-488-131-9.

A:303. MATTI NOJONEN: Guanxi – The Chinese Third Arm. 2007. ISBN 978-952-488-132-6.

A:304. HANNU OJALA: Essays on the Value Relevance of Goodwill Accounting. 2007.
ISBN 978-952-488-133-3, E-version: 978-952-488-135-7.

A:305. ANTTI KAUHANEN: Essays on Empirical Personnel Economics. 2007.
ISBN 978-952-488-139-5.

A:306. HANS MÄNTYLÄ: On "Good" Academic Work – Practicing Respect at Close Range. 2007. ISBN 978,952-488-1421-8, E-version: 978-952-488-142-5.

A:307. MILLA HUURROS: The Emergence and Scope of Complex System/Service Innovation. The Case of the Mobile Payment Services Market in Finland. 2007.
ISBN 978-952-488-143-2

A:308. PEKKA MALO: Higher Order Moments in Distribution Modelling with Applications to Risk Management. 2007. ISBN 978-952-488-155-5, E-version: 978-952-488-156-2.

A:309. TANJA TANAYAMA: Allocation and Effects of R&D Subsidies: Selection, Screening, and Strategic Behavior. 2007. ISBN 978-952-488-157-9, E-version: 978-952-488-158-6.

A:310. JARI PAULAMÄKI: Kauppiasyrittäjän toimintavapaus ketjuyrityksessä. Haastattelututkimus K-kauppiaan kokemasta toimintavapaudesta agenttiteorian näkökulmasta.
2008. Korjattu painos. ISBN 978-952-488-246-0, E-version: 978-952-488-247-7.

A:311. JANNE VIHINEN: Supply and Demand Perspectives on Mobile Products and Content Services. ISBN 978-952-488-168-5.

A:312. SAMULI KNÜPFER: Essays on Household Finance. 2007. ISBN 978-952-488-178-4.

A:313. MARI NYRHINEN: The Success of Firm-wide IT Infrastructure Outsourcing: an Integrated Approach. 2007. ISBN 978-952-488-179-1.

A:314. ESKO PENTTINEN: Transition from Products to Services within the Manufacturing Business. 2007. ISBN 978-952-488-181-4, E-version: 978-952-488-182-1.

A:315. JARKKO VESA: A Comparison of the Finnish and the Japanese Mobile Services Markets: Observations and Possible Implications. 2007. ISBN 978-952-488-184-5.

A:316. ANTTI RUOTOISTENMÄKI: Condition Data in Road Maintenance Management. 2007.
ISBN 978-952-488-185-2, E-version: 978-952-488-186-9.

A:317. NINA GRANQVIST: Nanotechnology and Nanolabeling. Essays on the Emergence of New Technological Fields. 2007. ISBN 978-952-488-187-6, E-version: 978-952-488-188-3.

A:318. GERARD L. DANFORD: INTERNATIONALIZATION: An Information-Processing Perspective. A Study of the Level of ICT Use During Internationalization. 2007.
ISBN 978-952-488-190-6.

A:319. TIINA RITVALA: Actors and Institutions in the Emergence of a New Field: A Study of the Cholesterol-Lowering Functional Foods Market. 2007. ISBN 978-952-488-195-1.

A:320. JUHA LAAKSONEN: Managing Radical Business Innovations. A Study of Internal Corporate Venturing at Sonera Corporation. 2007.
ISBN 978-952-488-201-9, E-version: 978-952-488-202-6.

A:321. BRETT FIFIELD: A Project Network: An Approach to Creating Emergent Business. 2008.
ISBN 978-952-488-206-4, E-version: 978-952-488-207-1.

A:322. ANTTI NURMI: Essays on Management of Complex Information Systems Development Projects. 2008. ISBN 978-952-488-226-2.

A:323. SAMI RELANDER: Towards Approximate Reasoning on New Software Product Company Success Potential Estimation. A Design Science Based Fuzzy Logic Expert System.
2008. ISBN 978-952-488-227-9.

A:324. SEPPO KINKKI: Essays on Minority Protection and Dividend Policy. 2008.
ISBN 978-952-488-229-3.

A:325. TEEMU MOILANEN: Network Brand Management: Study of Competencies of Place Branding Ski Destinations. 2008. ISBN 978-952-488-236-1.

A:326. JYRKI ALI-YRKKÖ: Essays on the Impacts of Technology Development and R&D Subsidies. 2008. ISBN 978-952-488-237-8.

A:327. MARKUS M. MÄKELÄ: Essays on software product development. A Strategic management viewpoint. 2008. ISBN 978-952-488-238-5.

A:328. SAMI NAPARI: Essays on the gender wage gap in Finland. 2008.
ISBN 978-952-488-243-9.

A:329. PAULA KIVIMAA: The innovation effects of environmental policies. Linking policies, companies and innovations in the Nordic pulp and paper industry.
2008. ISBN 978-952-488-244-6.

A:330 HELI VIRTA: Essays on Institutions and the Other Deep Determinants of Economic Development. 2008. ISBN 978-952-488-267-5.

A:331 JUKKA RUOTINEN: Essays in trade in services difficulties and possibilities.
2008. ISBN 978-952-488-271-2, E-version: ISBN 978-952-488-272-9.

A:332 IIKKA KORHONEN: Essays on commitment and government debt structure.
2008. ISBN 978-952-488-273-6, E-version: ISBN 978-952-488-274-3.

A:333. MARKO MERISAVO: The interaction between digital marketing communication and customer loyalty. 2008. ISBN 978-952-488-277-4, E-version 978-952-488-278-1.

A:334. PETRI ESKELINEN: Reference point based decision support tools for interactive multiobjective optimization. 2008. ISBN 978-952-488-282-8.

A:335. SARI YLI-KAUHALUOMA: Working on technology: a study on collaborative R&D work in industrial chemistry. 2008. ISBN 978-952-488-284-2

A:336. JANI KILPI: Sourcing of availability services - case aircraft component support. 2008. ISBN 978-952-488-284-2, 978-952-488-286-6 (e-version).

A:337 HEIDI SILVENNOINEN: Essays on household time allocation decisions in a collective household model. 2008. ISBN 978-952-488-290-3, ISBN 978-952-488-291-0 (e-version).

A:338 JUKKA PARTANEN: Pk-yrityksen verkostokyvykkyydet ja nopea kasvu - case: Tiede- ja teknologiavetoiset yritykset. 2008. ISBN 978-952-488-295-8.

A:339 PETRUS KAUTTO: Who holds the reins in Integrated Product Policy? An individual company as a target of regulation and as a policy maker. 2008. ISBN 978-952-488-300-9, 978-952-488-301-6 (e-version).

A:340 KATJA AHONIEMI: Modeling and Forecasting Implied Volatility. 2009. ISBN 978-952-488-303-0, E-version: 978-952-488-304-7.

B-SARJA: TUTKIMUKSIA - RESEARCH REPORTS. ISSN 0356-889X.

B:77. MATTI KAUTTO – ARTO LINDBLOM – LASSE MITRONEN: Kaupan liiketoiminta-osaaminen. 2007. ISBN 978-952-488-109-8.

B:78. NIILO HOME: Kauppiasyrittäjyys. Empiirinen tutkimus K-ruokakauppiaiden yrittäjyysasenteista. Entrepreneurial Orientation of Grocery Retailers – A Summary. ISBN 978-952-488-113-5, E-versio: ISBN 978-952-488-114-2.

B:79. PÄIVI KARHUNEN – OLENA LESYK – KRISTO OVASKA: Ukraina suomalaisyritysten toimintaympäristönä. 2007. ISBN 978-952-488-150-0, E-versio: 978-952-488-151-7.

B:80. MARIA NOKKONEN: Näkemyksiä pörssiyhtiöiden hallitusten sukupuolikiintiöistä. Retorinen diskurssianalyysi Helsingin Sanomien verkkokeskusteluista. Nasta-projekti. 2007. ISBN 978-952-488-166-1, E-versio: 978-952-488-167-8.

B:81. PIIA HELISTE – RIITTA KOSONEN – MARJA MATTILA: Suomalaisyritykset Baltiassa tänään ja huomenna: Liiketoimintanormien ja -käytäntöjen kehityksestä. 2007. ISBN 978-952-488-177-7, E-versio: 978-952-488-183-8.

B:82. OLGA MASHKINA – PIIA HELISTE – RIITTA KOSONEN: The Emerging Mortgage Market in Russia: An Overview with Local and Foreign Perspectives. 2007. ISBN 978-952-488-193-7, E-version: 978-952-488-194-4.

B:83. PIIA HELISTE – MARJA MATTILA – KRZYSZTOF STACHOWIAK: Puola suomalais-yritysten toimintaympäristönä. 2007. ISBN 978-952-488-198-2, E-versio: 978-952-488-199-9.

B:84. PÄIVI KARHUNEN – RIITTA KOSONEN – JOHANNA LOGRÉN – KRISTO OVASKA: Suomalaisyritysten strategiat Venäjän muuttuvassa liiketoimintaympäristössä. 2008. ISBN 978-953-488-212-5, E-versio: 978-952-488-241-5.

B:85. MARJA MATTILA – EEVA KEROLA – RIITTA KOSONEN: Unkari suomalaisyritysten toi-mintaympäristönä. 2008. ISBN 978-952-488-213-2, E-versio: 978-952-488-222-4.

B:86 KRISTIINA KORHONEN – ANU PENTTILÄ – MAYUMI SHIMIZU – EEVA KEROLA – RIITTA KOSONEN: Intia suomalaisyritysten toimintaympäristönä.2008. ISBN 978-952-488-214-9, E-versio: 978-952-488-283-5

B:87. SINIKKA VANHALA – SINIKKA PESONEN: Työstä nauttien. SEFE:en kuuluvien nais- ja miesjohtajien näkemyksiä työstään ja urastaan. 2008. ISBN 978-952-488-224-8, E-versio: 978-952-488-225-5.

B:88. POLINA HEININEN – OLGA MASHKINA – PÄIVI KARHUNEN – RIITTA KOSONEN: Leningradin lääni yritysten toimintaympäristönä: pk-sektorin näkökulma. 2008. ISBN 978-952-488-231-6, E-versio: 978-952-488-235-4.

B:89. Ольга Машкина – Полина Хейнинен: Влияние государственного сектора на развитие малого и среднего предпринимательства в Ленинградской области: взгляд предприятий.2008. ISBN 978-952-488-233-0, E-version: 978-952-488-240-8.

B:90. MAI ANTTILA – ARTO RAJALA (Editors): Fishing with business nets – keeping thoughts on the horizon Professor Kristian Möller. 2008. ISBN 978-952-488-249-1, E-version: 978-952-488-250-7.

B:91. RENÉ DE KOSTER – WERNER DELFMANN (Editors): Recent developments in supply chain management. 2008. ISBN 978-952-488-251-4, E-version: 978-952-488-252-1.

B:92 KATARIINA RASILAINEN: Valta orkesterissa. Narratiivinen tutkimus soittajien kokemuksista ja näkemyksistä. 2008. ISBN 978-952-488-254-5, E-versio: 978-952-488-256-9.

B:93. SUSANNA KANTELINEN: Opiskelen, siis koen. Kohti kokevan subjektin tunnistavaa korkeakoulututkimusta. 2008. ISBN 978-952-488-257-6, E-versio: 978-952-488-258.

B:94 KATRI KARJALAINEN – TUOMO KIVIOJA – SANNA PELLAVA: Yhteishankintojen kustannusvaikutus. Valtion hankintatoimen kustannussäästöjen selvittäminen. 2008. ISBN 978-952-488-263-7, E-versio: ISBN 978-952-488-264-4.

B:95 ESKO PENTTINEN: Electronic Invoicing Initiatives in Finland and in the European Union – Taking the Steps towards the Real-Time Economy. 2008. ISBN 978-952-488-268-2, E-versio: ISBN 978-952-488-270-5.

B:96 LIISA UUSITALO (Editor): Museum and visual art markets. 2008. ISBN 978-952-488-287-3, E-version: ISBN 978-952-488-288-0.

B:97 EEVA-LIISA LEHTONEN: Pohjoismaiden ensimmäinen kauppatieteiden tohtori Vilho Paavo Nurmilahti 1899-1943. 2008. ISBN 978-952-488-292-7, E-versio: ISBN 978-952-488-293-4.

B:98 ERJA KETTUNEN – JYRI LINTUNEN – WEI LU – RIITTA KOSONEN: Suomalaisyritysten strategiat Kiinan muuttuvassa toimintaympäristössä. 2008 ISBN 978-952-488-234-7, E-versio: ISBN 978-952-488-297-2.

B:99 SUSANNA VIRKKULA – EEVA-KATRI AHOLA – JOHANNA MOISANDER – JAAKKO ASPARA – HENRIKKI TIKKANEN: Messut kuluttajia osallistavan markkinakulttuurin fasilitaattorina: messukokemuksen rakentuminen Venemessuilla. 2008. ISBN 978-952-488-298-9, E-versio: ISBN 978-952-488-299-6

N-SARJA: HELSINKI SCHOOL OF ECONOMICS. MIKKELI BUSINESS CAMPUS PUBLICATIONS.
ISSN 1458-5383

N:63. SOILE MUSTONEN – ANNE GUSTAFSSON-PESONEN: Oppilaitosten yrittäjyyskoulutuksen kehittämishanke 2004–2006 Etelä-Savon alueella. Tavoitteiden, toimenpiteiden ja vaikuttavuuden arviointi. 2007. ISBN: 978-952-488-086-2.

N:64. JOHANNA LOGRÉN – VESA KOKKONEN: Pietarissa toteutettujen yrittäjäkoulutusohjelmien vaikuttavuus. 2007. ISBN 978-952-488-111-1.

N:65. VESA KOKKONEN: Kehity esimiehenä – koulutusohjelman vaikuttavuus. 2007. ISBN 978-952-488-116-6.

N:66. VESA KOKKONEN – JOHANNA LOGRÉN: Kaupallisten avustajien – koulutusohjelman vaikuttavuus. 2007. ISBN 978-952-488-116-6.

N:67. MARKKU VIRTANEN: Summary and Declaration. Of the Conference on Public Support Systems of SME's in Russia and Other North European Countries. May 18 – 19, 2006, Mikkeli, Finland. 2007. ISBN 978-952-488-140-1.

N:68. ALEKSANDER PANFILO – PÄIVI KARHUNEN: Pietarin ja Leningradin läänin potentiaali kaakkoissuomalaisille metallialan yrityksille. 2007. ISBN 978-952-488-163-0.

N:69. ALEKSANDER PANFILO – PÄIVI KARHUNEN – VISA MIETTINEN: Pietarin innovaatiojärjestelmä jayhteistyöpotentiaali suomalaisille innovaatiotoimijoille. 2007. ISBN 978-952-488-164-7.

N:70. VESA KOKKONEN: Perusta Oma Yritys – koulutusohjelman vaikuttavuus. 2007. ISBN 978-952-488-165-4.

N:71. JARI HANDELBERG – MIKKO SAARIKIVI: Tutkimus Miktech Yrityshautomon yritysten näkemyksistä ja kokemuksista hautomon toiminnasta ja sen edelleen kehittämisestä. 2007. ISBN 978-952-488-175-3.

N:72. SINIKKA MYNTTINEN – MIKKO SAARIKIVI – ERKKI HÄMÄLÄINEN: Mikkelin Seudun yrityspalvelujen henkilökunnan sekä alueen yrittäjien näkemykset ja suhtautuminen mentorointiin. 2007. ISBN 978-952-488-176-0.

N:73. SINIKKA MYNTTINEN: Katsaus K-päivittäistavarakauppaan ja sen merkitykseen Itä-Suomessa. 2007. ISBN 978-952-488-196-8.

N:74. MIKKO SAARIKIVI: Pk-yritysten kansainvälistymisen sopimukset. 2008. ISBN 978-952-488-210-1.

N:75. LAURA TUUTTI: Uutta naisjohtajuutta Delfoi Akatemiasta – hankkeen vaikuttavuus. 2008. ISBN 978-952-488-211-8.

N:76. LAURA KEHUSMAA – JUSSI KÄMÄ – ANNE GUSTAFSSON-PESONEN (ohjaaja): StuNet -Business Possibilities and Education - hankkeen arviointi. 2008. ISBN 978-952-488-215-6.

N:77. PÄIVI KARHUNEN – ERJA KETTUNEN – VISA MIETTINEN – TIINAMARI SIVONEN: Determinants of knowledge-intensive entrepreneurship in Southeast Finland and Northwest Russia. 2008. ISBN 978-952-488-223-1.

N:78. ALEKSANDER PANFILO – PÄIVI KARHUNEN – VISA MIETTINEN: Suomalais-venäläisen innovaatioyhteistyön haasteet toimijanäkökulmasta. 2008. ISBN 978-952-488-232-3.

N:79. VESA KOKKONEN: Kasva Yrittäjäksi – koulutusohjelman vaikuttavuus. 2008. ISBN 978-952-488-248-4.

N:80. VESA KOKKONEN: Johtamisen taidot - hankkeessa järjestettyjen koulutusohjelmien vaikuttavuus. 2008. ISBN 978-952-488-259-0.

N:81. MIKKO SAARIKIVI: Raportti suomalaisten ja brittiläisten pk-yritysten yhteistyön kehittämisestä uusiutuvan energian sektorilla. 2008. ISBN 978-952-488-260-6.

N:82. MIKKO SAARIKIVI – JARI HANDELBERG – TIMO HOLMBERG – ARI MATILAINEN: Selvitys lujitemuovikomposiittituotteiden mahdollisuuksista rakennusteollisuudessa. 2008. ISBN 978-952-488-262-0.

N:83. PÄIVI KARHUNEN – SVETLANA LEDYAEVA – ANNE GUSTAFSSON-PESONEN – ELENA MOCHNIKOVA – DMITRY VASILENKO: Russian students' perceptions of entrepreneurship. Results of a survey in three St. Petersburg universities. Entrepreneurship development –project 2. 2008. ISBN 978-952-488-280-4.

W-SARJA: TYÖPAPEREITA - WORKING PAPERS . ISSN 1235-5674.
ELECTRONIC WORKING PAPERS, ISSN 1795-1828.

W:412. LOTHAR THIELE – KAISA MIETTINEN – PEKKA J. KORHONEN – JULIAN MOLINA: A Preference-Based Interactive Evolutionary Algorithm for Multiobjective Optimization. 2007. ISBN 978-952-488-094-7.

W:413. JAN-ERIK ANTIPIN – JANI LUOTO: Are There Asymmetric Price Responses in the Euro Area? 2007. ISBN 978-952-488-097-8.

W:414. SAMI SARPOLA: Evaluation Framework for VML Systems. 2007. ISBN 978-952-488-097-8.

W:415. SAMI SARPOLA: Focus of Information Systems in Collaborative Supply Chain Relationships. 2007. ISBN 978-952-488-101-2.

W:416. SANNA LAUKKANEN: Information Systems as Integrative Infrastructures. Information Integration and the Broader Context of Integrative and Coordinative Devices. 2007. ISBN 978-952-488-102-9.

W:417. SAMULI SKURNIK – DANIEL PASTERNACK: Uusi näkökulma 1900-luvun alun murroskauteen ja talouden murrosvaiheiden dynamiikkaan. Liikemies Moses Skurnik osakesijoittajana ja -välittäjänä. 2007. ISBN 978-952-488-104-3.

W:418. JOHANNA LOGRÉN – PIIA HELISTE: Kymenlaakson pienten ja keskisuurten yritysten Venäjä-yhteistyöpotentiaali. 2001. ISBN 978-952-488-112-8.

W-419. SARI STENFORS – LEENA TANNER: Evaluating Strategy Tools through Activity Lens. 2007. ISBN 978-952-488-120-3.

W:420. RAIMO LOVIO: Suomalaisten monikansallisten yritysten kotimaisen sidoksen heikkeneminen 2000-luvulla. 2007. ISBN 978-952-488-121-0.

W:421. PEKKA J. KORHONEN – PYRY-ANTTI SIITARI: A Dimensional Decomposition Approach to Identifying Efficient Units in Large-Scale DEA Models. 2007. ISBN 978-952-488-124-1.

W:422. IRYNA YEVSEYEVA – KAISA MIETTINEN – PEKKA SALMINEN – RISTO LAHDELMA: SMAA-Classification - A New Method for Nominal Classification. 2007. ISBN 978-952-488-129-6.

W:423. ELINA HILTUNEN: The Futures Window – A Medium for Presenting Visual Weak Signals to Trigger Employees' Futures Thinking in Organizations. 2007. ISBN 978-952-488-134-0.

W:424. TOMI SEPPÄLÄ – ANTTI RUOTOISTENMÄKI – FRIDTJOF THOMAS: Optimal Selection and Routing of Road Surface Measurements. 2007. ISBN 978-952-488-137-1.

W:425. ANTTI RUOTOISTENMÄKI: Road Maintenance Management System. A Simplified Approach. 2007. ISBN 978-952-488-1389-8.

W:426. ANTTI PIRJETÄ – VESA PUTTONEN: Style Migration in the European Markets 2007. ISBN 978-952-488-145-6.

W:427. MARKKU KALLIO – ANTTI PIRJETÄ: Incentive Option Valuation under Imperfect Market and Risky Private Endowment. 2007. ISBN 978-952-488-146-3.

W:428. ANTTI PIRJETÄ – SEPPO IKÄHEIMO – VESA PUTTONEN: Semiparametric Risk Preferences Implied by Executive Stock Options. 2007. ISBN 978-952-488-147-0.

W:429. OLLI-PEKKA KAUPPILA: Towards a Network Model of Ambidexterity. 2007. ISBN 978-952-488-148-7.

W:430. TIINA RITVALA – BIRGIT KLEYMANN: Scientists as Midwives to Cluster Emergence. An Interpretative Case Study of Functional Foods. 2007. ISBN 978-952-488-149-4.

W:431. JUKKA ALA-MUTKA: Johtamiskyvykkyyden mittaaminen kasvuyrityksissä. 2007. ISBN 978-952-488-153-1.

W:432. MARIANO LUQUE – FRANCISCO RUIZ – KAISA MIETTINEN: GLIDE – General Formulation for Interactive Multiobjective Optimization. 2007. ISBN 978-952-488-154-8.

W:433. SEPPO KINKKI: Minority Protection and Information Content of Dividends in Finland. 2007. ISBN 978-952-488-170-8.

W:434. TAPIO LAAKSO: Characteristics of the Process Supersede Characteristics of the Debtor Explaining Failure to Recover by Legal Reorganization Proceedings. 2007. ISBN 978-952-488-171-5.

W:435. MINNA HALME: Something Good for Everyone? Investigation of Three Corporate Responsibility Approaches. 2007. ISBN 978-952-488-189.

W:436. ARTO LAHTI: Globalization, International Trade, Entrepreneurship and Dynamic Theory of Economics.The Nordic Resource Based View. Part One. 2007. ISBN 978-952-488-191-3.

W:437. ARTO LAHTI: Globalization, International Trade, Entrepreneurship and Dynamic Theory of Economics.The Nordic Resource Based View. Part Two. 2007 ISBN 978-952-488-192-0.

W:438. JANI KILPI: Valuation of Rotable Spare Parts. 2007. ISBN 978-952-488-197-5.

W:439. PETRI ESKELINEN – KAISA MIETTINEN – KATHRIN KLAMROTH – JUSSI HAKANEN: Interactive Learning-oriented Decision Support Tool for Nonlinear Multiobjective Optimization: Pareto Navigator. 2007. ISBN 978-952-488-200-2.

W:440. KALYANMOY DEB – KAISA MIETTINEN – SHAMIK CHAUDHURI: Estimating Nadir Objective Vector: Hybrid of Evolutionary and Local Search. 2008. ISBN 978-952-488-209-5.

W:441. ARTO LAHTI: Globalisaatio haastaa pohjoismaisen palkkatalousmallin. Onko löydettävissä uusia aktiivisia toimintamalleja, joissa Suomi olisi edelleen globalisaation voittaja? 2008. ISBN 978-952-488-216-3.

W:442. ARTO LAHTI: Semanttinen Web – tulevaisuuden internet. Yrittäjien uudet liiketoimintamahdollisuudet. 2008. ISBN 978-952-488-217-0.

W:443. ARTO LAHTI: Ohjelmistoteollisuuden globaali kasvustrategia ja immateriaalioikeudet. 2008. ISBN 978-952-488-218-7.

W:444. ARTO LAHTI: Yrittäjän oikeusvarmuus globaalisaation ja byrokratisoitumisen pyörteissä. Onko löydettävissä uusia ja aktiivisia toimintamalleja yrittäjien syrjäytymisen estämiseksi? 2008. ISBN 978-952-488-219-4.

W:445. PETRI ESKELINEN: Objective trade-off rate information in interactive multiobjective optimization methods – A survey of theory and applications. 2008. ISBN 978-952-488-220-0.

W:446. DEREK C. JONES – PANU KALMI: Trust, inequality and the size of co-operative sector – Cross-country evidence. 2008. ISBN 978-951-488-221-7.

W:447. KRISTIINA KORHONEN – RIITTA KOSONEN – TIINAMARI SIVONEN – PASI SAUKKONEN: Pohjoiskarjalaisten pienten ja keskisuurten yritysten Venäjäyhteistyöpotentiaali ja tukitarpeet. 2008. ISBN 978-952-488-228-6.

W:448. TIMO JÄRVENSIVU – KRISTIAN MÖLLER: Metatheory of Network Management: A Contingency Perspective. 2008. ISBN 978-952-488-231-6.

W:449. PEKKA KORHONEN: Setting "condition of order preservation" requirements for the priority vector estimate in AHP is not justified. 2008. ISBN 978-952-488-242-2.

W:450. LASSE NIEMI – HANNU OJALA – TOMI SEPPÄLÄ: Misvaluation of takeover targets and auditor quality. 2008. ISBN 978-952-488-255-2.

W:451. JAN-ERIK ANTIPIN – JANI LUOTO: Forecasting performance of the small-scale hybrid New Keynesian model. 2008. ISBN 978-952-488-261-3.

W:452. MARKO MERISAVO: The Interaction between Digital Marketing Communication and Customer Loyalty. 2008. ISBN 978-952-488-266-8.

W:453. PETRI ESKELINEN – KAISA MIETTINEN: Trade-off Analysis Tool with Applicability Study for Interactive Nonlinear Multiobjective Optimization. 2008. ISBN 978-952-488-269-9.

W:454. SEPPO IKÄHEIMO – VESA PUTTONEN – TUOMAS RATILAINEN: Antitakeover provisions and performance – Evidence from the Nordic countries. 2008. ISBN 978-952-488-275-0.

W:455. JAN-ERIK ANTIPIN: Dynamics of inflation responses to monetary policy in the EMU area. 2008. ISBN 978-952-488-276-7.

W:456. KIRSI KOMMONEN: Narratives on Chinese colour culture in business contexts. The Yin Yang Wu Xing of Chinese values. 2008. ISBN 978-952-488-279-8.

W:457. MARKKU ANTTONEN – MIKA KUISMA – MINNA HALME – PETRUS KAUTTO: Materiaalitehokkuuden palveluista ympäristömyötäistä liiketoimintaa (MASCO2). 2008. ISBN 978-952-488-279-8.

Z-SARJA: HELSINKI SCHOOL OF ECONOMICS.
CENTRE FOR INTERNATIONAL BUSINESS RESEARCH. CIBR WORKING PAPERS. ISSN 1235-3931.

Z:16. PETER GABRIELSSON – MIKA GABRIELSSON: Marketing Strategies for Global Expansion in the ICT Field. 2007. ISBN 978-952-488-105-0.

Z:17. MIKA GABRIELSSON – JARMO ERONEN – JORMA PIETALA: Internationalization and Globalization as a Spatial Process. 2007. ISBN 978-952-488-136-4.

Kaikkia Helsingin kauppakorkeakoulun julkaisusarjassa ilmestyneitä julkaisuja voi tilata osoitteella:

KY-Palvelu Oy
Kirjakauppa
Runeberginkatu 14-16
00100 Helsinki
Puh. (09) 4313 8310, fax (09) 495 617
Sähköposti: kykirja@ky.hse.fi

Helsingin kauppakorkeakoulu
Julkaisutoimittaja
PL 1210
00101 Helsinki
Puh. (09) 4313 8579, fax (09) 4313 8305
Sähköposti: julkaisu@hse.fi

All the publications can be ordered from

Helsinki School of Economics
Publications officer
P.O.Box 1210
FIN-00101 Helsinki
Phone +358-9-4313 8579, fax +358-9-4313 8305
E-mail: julkaisu@hse.fi