

Multiplicative Models for Implied Volatility

Katja Ahoniemi*

Helsinki School of Economics, FDPE, and HECER

January 15, 2007

Abstract

This paper estimates a mixture multiplicative error model for the implied volatilities of both call and put options on the Nikkei 225 index. Diagnostics show that the mixture multiplicative model is a good fit to the data, and it outperforms a multiplicative model with no mixture components. The forecast performance of the mixture model is superior to that of simpler models for Nikkei call implied volatility, but the directional forecast accuracy of an ARIMA model is slightly better than that of multiplicative models for Nikkei put implied volatility.

*Correspondence: Helsinki School of Economics, Department of Economics, P.O. Box 1210, 00101 Helsinki, Finland. E-mail: katja.ahoniemi@hse.fi. Financial support from the Finnish Doctoral Programme in Economics and the Yrjö Jahnsson Foundation is gratefully acknowledged.

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 stock portfolio weights or composition.

Implied volatility has traditionally been modeled with ARMA and other linear regression models (e.g. Harvey & Whaley (1992) and Brooks & Oozeer (2002)), or with ARMA models with exogenous regressors and GARCH errors (Ahoiemi (2006)). 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 compute one-month-ahead forecasts and use them as explanatory variables in an AR(1) model for the VIX index, an index of S&P 500 index option implied volatility. The conclusion is that the forecasts from the multiplicative specification have significant explanatory power in modeling the value of the VIX.

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.

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 research¹. So far, multiplicative modeling has not been applied to implied 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 allows for 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%. For the Nikkei call IV time series, multiplicative models outperform ARIMA models as directional forecasters, but for the put IV, ARIMA models fare just as well or even better than multiplicative models.

The paper is structured as follows. Section 2 presents the mixture multiplicative error model, and Section 3 describes the data used in the study, the estimation results, and diagnostics. Section 4 analyzes the forecasts from various competing models, and Section 5 concludes.

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

¹See e.g. Bauwens & Giot (2003), Ghysels et al. (2004), Manganeli (2005), Fernandes & Grammig (2006) and Meitz & Teräsvirta (2006).

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 :

$$IV_t = \mu_t \varepsilon_t \quad (1)$$

where IV_t is the implied volatility time series under analysis. Shocks can be assumed to be *iid* with mean 1 from a non-negative distribution. In this particular study, a mixture multiplicative error model (MMEM) similar to that in Lanne (2006) has been estimated. In such a specification, there are two possible mean equations:

$$\mu_{1t} = \omega_1 + \sum_{i=1}^{q_1} \alpha_{1i} IV_{t-i} + \sum_{j=1}^{p_1} \beta_{1j} \mu_{1,t-j} \quad (2)$$

$$\mu_{2t} = \omega_2 + \sum_{i=1}^{q_2} \alpha_{2i} IV_{t-i} + \sum_{j=1}^{p_2} \beta_{2j} \mu_{2,t-j} \quad (3)$$

Therefore, μ_t depends on q past observations of implied volatility and p past expected implied volatilities, and the model is specified as $MMEM(p_1, q_1; p_2, q_2)$. This autoregressive form for the mean equations can help to capture possible clustering in the data. Clustering is often present in financial time series, i.e. small (large) changes are more likely to be followed by small (large) changes. The mixture specification is also extended into the error term, with the error terms coming from two gamma distributions with possibly different shape and scale parameters. 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 mixture components 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 $(1 - \pi)$.

The conditional mean equations reveal that MEM (and ACD) models are similar in structure to GARCH models. Therefore, parameter constraints that apply to GARCH models also apply to MEM models (see Section 3.2 for further discussion). The shape and scale parameters of the gamma distributions are constrained so that with $\varepsilon_{1t} \sim \text{Gamma}(\gamma_1, \delta_1)$, $\delta_1 = 1/\gamma_1$ and with $\varepsilon_{2t} \sim \text{Gamma}(\gamma_2, \delta_2)$, $\delta_2 = 1/\gamma_2$, or so that the scale parameter is the inverse of the shape parameter, which ensures that the error term will have mean unity.

When employing maximum likelihood (ML) estimation for the MEM model with mixture components, the conditional distribution of IV_t is:

$$\begin{aligned}
f(IV_t; \theta) = & \pi \frac{1}{\mu_{1t} \Gamma(\gamma_1) \delta_1^{\gamma_1}} \left(\frac{IV_t}{\mu_{1t}} \right)^{\gamma_1 - 1} \exp\left(-\frac{IV_t}{\delta_1 \mu_{1t}} \right) + \\
& (1 - \pi) \frac{1}{\mu_{2t} \Gamma(\gamma_2) \delta_2^{\gamma_2}} \left(\frac{IV_t}{\mu_{2t}} \right)^{\gamma_2 - 1} \exp\left(-\frac{IV_t}{\delta_2 \mu_{2t}} \right)
\end{aligned} \tag{4}$$

where θ is the parameter vector and $\Gamma(\cdot)$ is the gamma function. The log-likelihood function can then be written as:

$$\ell(\theta) = \sum_{t=1}^T \ln[f_{t-1}(IV_t)] \tag{5}$$

3 Estimation

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 currency of denomination for the Nikkei 225 index is the Japanese Yen. The component stocks of the index are reviewed once a year. 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 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.

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 demand for call options.

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 1. The IV of puts has been slightly more volatile during the time period in question. Both series are skewed to the right and they display

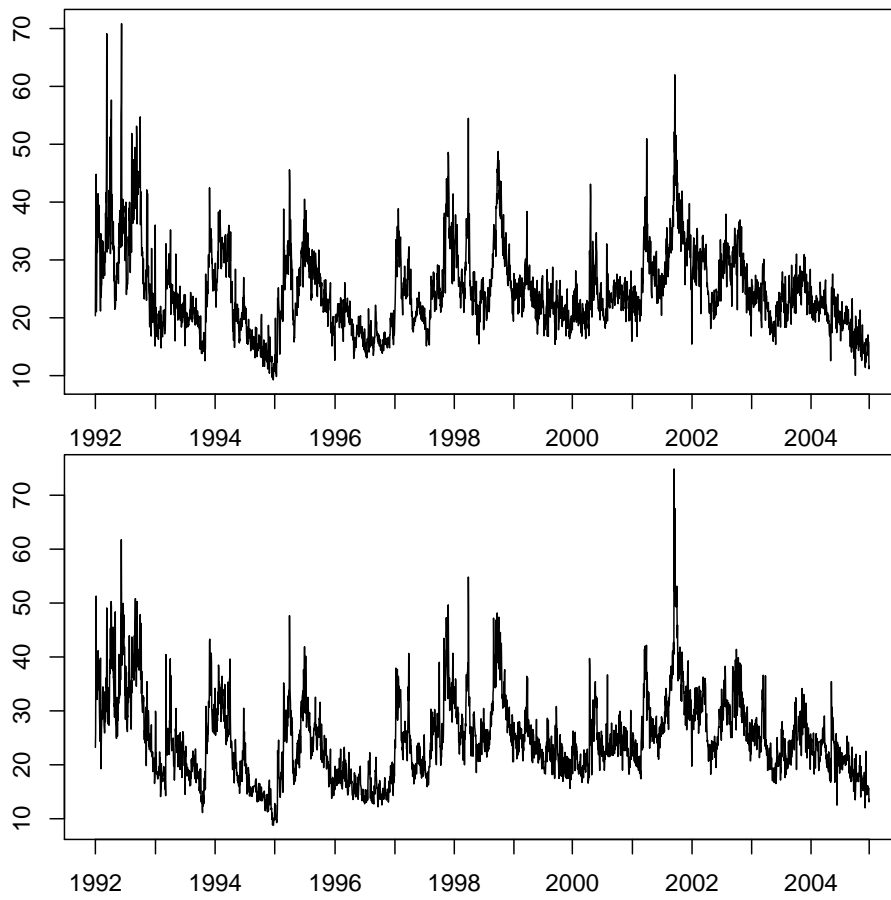


Figure 1: Nikkei 225 index call implied volatility (upper panel) and put implied volatility (lower panel) 1.1.1992 - 31.12.2004.

excess kurtosis. The autocorrelations for NIKC and NIKP are displayed in Figure 2, revealing the relatively high degree of persistence in the data. A unit root is rejected by the Augmented Dickey-Fuller test for both NIKC and NIKP.

	<i>NIKC</i>	<i>NIKP</i>
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
Excess kurtosis	2.42	1.79

Table 1: Descriptive statistics for NIKC and NIKP for the full sample of 1.1.1992 - 31.12.2004.

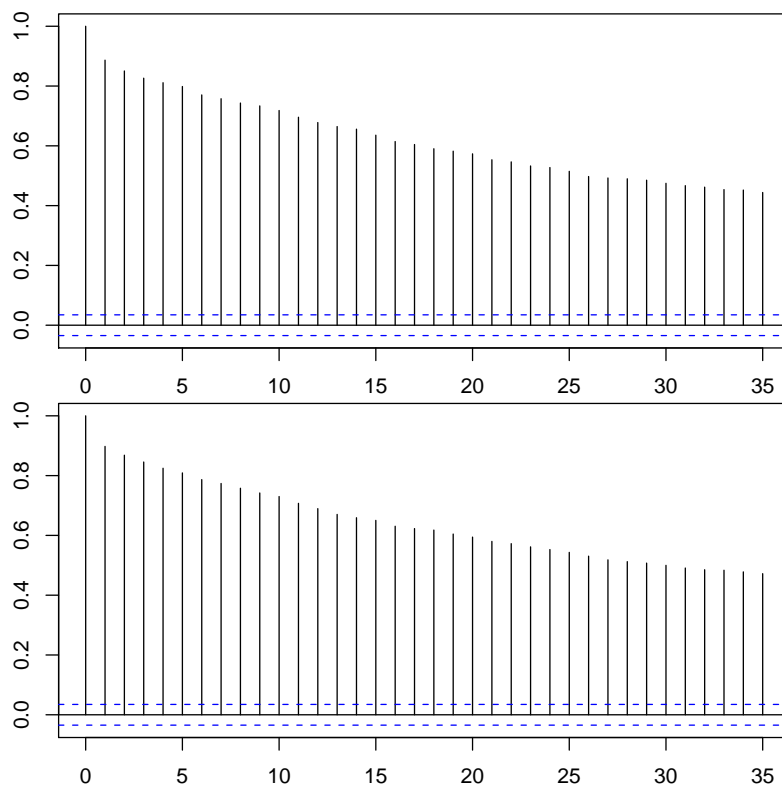


Figure 2: Autocorrelations for NIKC (upper panel) and NIKP (lower panel). The dashed lines mark the 95% confidence interval.

3.2 Model estimation

The in-sample period used in model estimation covered 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 the exchange rate realized volatility time series used by Lanne (2006). However, the coefficient for α_{22} was not statistically significant for NIKC or NIKP, so a (1,2;1,1) specification was also estimated for both time series.

Table 2 presents the coefficients and log-likelihoods of the estimated models. As the change in log-likelihoods is minimal, it can be concluded that the parameter α_{22} can be dropped from the models.

	<i>NIKC</i>		<i>NIKP</i>	
	<i>MMEM(1,2;1,2)</i>	<i>MMEM(1,2;1,1)</i>	<i>MMEM(1,2;1,2)</i>	<i>MMEM(1,2;1,1)</i>
Log likelihood	-6422.82	-6422.85	-6384.90	-6385.10
π	0.821 (0.000)	0.822 (0.000)	0.940 (0.000)	0.940 (0.000)
γ_1	145.689 (0.000)	145.495 (0.000)	113.177 (0.000)	113.300 (0.000)
ω_1	0.264 (0.002)	0.261 (0.002)	0.288 (0.000)	0.292 (0.000)
α_{11}	0.637 (0.000)	0.638 (0.000)	0.573 (0.000)	0.571 (0.000)
α_{12}	-0.255 (0.000)	-0.261 (0.000)	-0.183 (0.000)	-0.177 (0.000)
β_1	0.606 (0.000)	0.610 (0.000)	0.595 (0.000)	0.590 (0.000)
γ_2	26.553 (0.000)	26.541 (0.000)	19.437 (0.000)	19.346 (0.000)
ω_2	0.674 (0.092)	0.717 (0.060)	3.528 (0.240)	2.228 (0.148)
α_{21}	0.339 (0.000)	0.324 (0.000)	0.543 (0.021)	0.587 (0.005)
α_{22}	-0.031 (0.808)	- -	0.324 (0.491)	- -
β_2	0.6742 (0.000)	0.657 (0.000)	0.119 (0.810)	0.409 (0.001)

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

The estimated coefficients satisfy the constraints outlined in Nelson & Cao (1992) for GARCH models. For the (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, α_{i2} can be negative.

Figure 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 1, and the densities for the second mixture component are more dispersed and skewed to the right.

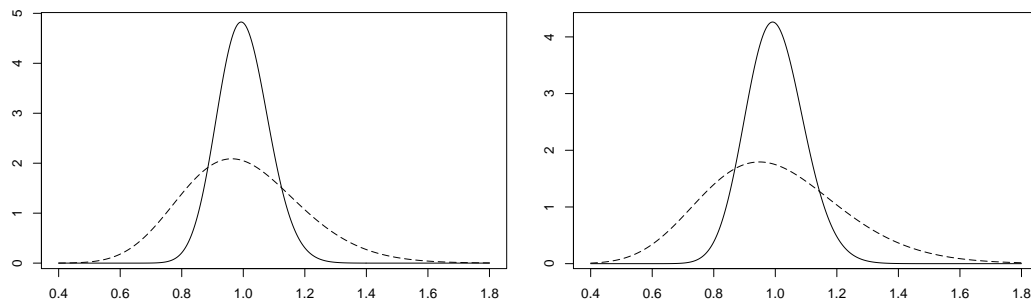


Figure 3: Estimated densities of error terms from the MMEM(1,2;1,1) model for NIKC (left) and NIKP (right). The solid line is the density of ε_{1t} and the dashed line is the density of ε_{2t} .

With financial market data, it may be that only the most recent history is relevant in modeling and forecasting, so the MMEM(1,2;1,2) and MMEM(1,2;1,1) models were also estimated for NIKC and NIKP using only the past 500 observations. This corresponds to an in-sample of 20.12.2000 - 30.12.2002. The estimation results for this specification are given in Table 3. As before, α_{22} is excluded from the models when choosing which models to use for forecasts.

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 results of this estimation are given in Table 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 4).

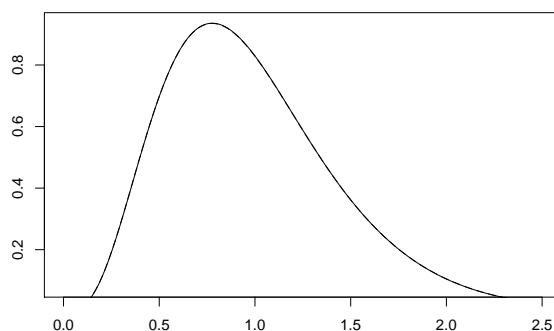


Figure 4: Estimated density of error terms from the MEM(1,2) model for NIKC and NIKP.

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. 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

	<i>NIKC</i>		<i>NIKP</i>	
	<i>MMEM(1,2;1,2)</i>	<i>MMEM(1,2;1,1)</i>	<i>MMEM(1,2;1,2)</i>	<i>MMEM(1,2;1,1)</i>
Log likelihood	-1227.03	-1227.37	-1200.32	-1200.69
π	0.871 (0.000)	0.870 (0.000)	0.928 (0.000)	0.930 (0.000)
γ_1	145.999 (0.000)	146.208 (0.000)	146.099 (0.000)	145.589 (0.000)
ω_1	0.644 (0.011)	0.654 (0.011)	0.894 (0.023)	0.880 (0.026)
α_{11}	0.505 (0.000)	0.491 (0.000)	0.458 (0.000)	0.452 (0.000)
α_{12}	-0.268 (0.000)	-0.248 (0.001)	0.035 (0.653)	0.042 (0.597)
β_1	0.741 (0.000)	0.735 (0.000)	0.475 (0.000)	0.475 (0.000)
γ_2	27.251 (0.000)	27.471 (0.000)	24.841 (0.004)	24.656 (0.004)
ω_2	0.000 (0.000)	0.000 (0.060)	1.495 (0.728)	1.522 (0.667)
α_{21}	0.487 (0.077)	0.633 (0.006)	0.367 (0.367)	0.548 (0.060)
α_{22}	0.469 (0.446)	- -	0.411 (0.456)	- -
β_2	0.025 (0.969)	0.353 (0.139)	0.223 (0.680)	0.427 (0.167)

Table 3: Estimation results for the *MMEM(1,2;1,2)* and *MMEM(1,2;1,1)* models for *NIKC* and *NIKP* with an in-sample of 500 observations. P-values for the significance of the coefficients are given in parentheses.

	<i>NIKC</i>	<i>NIKP</i>
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 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.

e.g. Bauwens et al. (2004) and Lanne (2006). The probability integral transforms are computed as:

$$z_t = \int_{-\infty}^{IV_t} f_{t-1}(u) du \quad (6)$$

where f_{t-1} is the conditional density of IV_t relating to the model under analysis. The framework of Diebold et al. (1998) was developed to evaluate density forecasts, but 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 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, so the model specification succeeds in taking account of the tails of the conditional distribution for both NIKC and NIKP². This holds true even when using only 500 observations in the estimation.

As a second diagnostic check, autocorrelation functions based on demeaned probability integral transforms and their squares were computed (see Figures 6 and 7). There is some autocorrelation in the demeaned probability integral transforms of NIKP, as well as in the squares 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 when using only 500 observations in the model estimation³.

²With a perfect model, z_t would be uniformly distributed and the columns of the histogram would all be of exactly the same height. The confidence interval is calculated without taking estimation error into account.

³The addition of the statistically insignificant parameter α_{22} to the diagnostic analysis does not improve the autocorrelations.

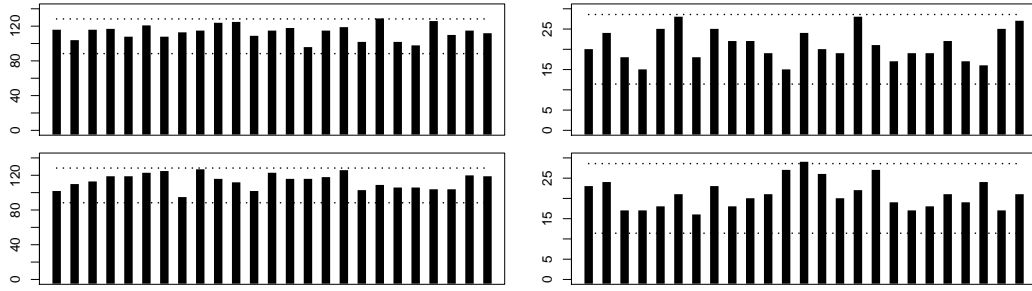


Figure 5: Histograms of probability integral transforms for NIKC (upper left panel) and NIKP (lower left panel) with the MMEM(1,2;1,1) model. Histograms for the MMEM(1,2;1,1) model estimated with 500 observations are given on the right. The dotted lines depict the boundaries of the 95% confidence interval.

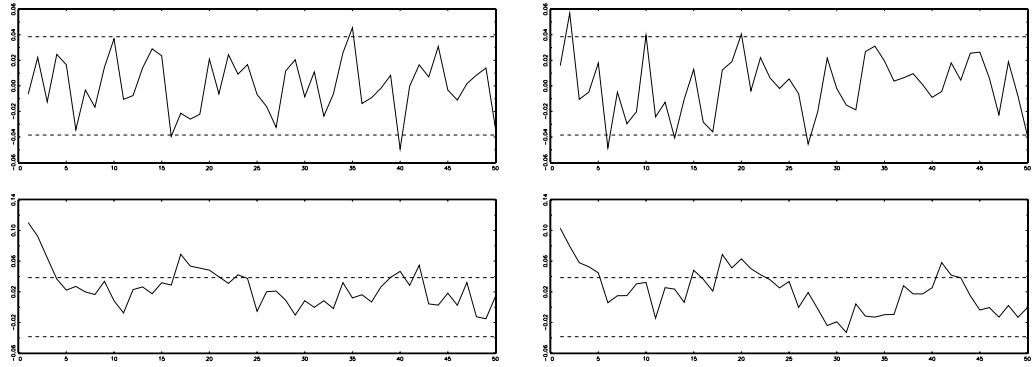


Figure 6: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) from the MMEM(1,2;1,1) model. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

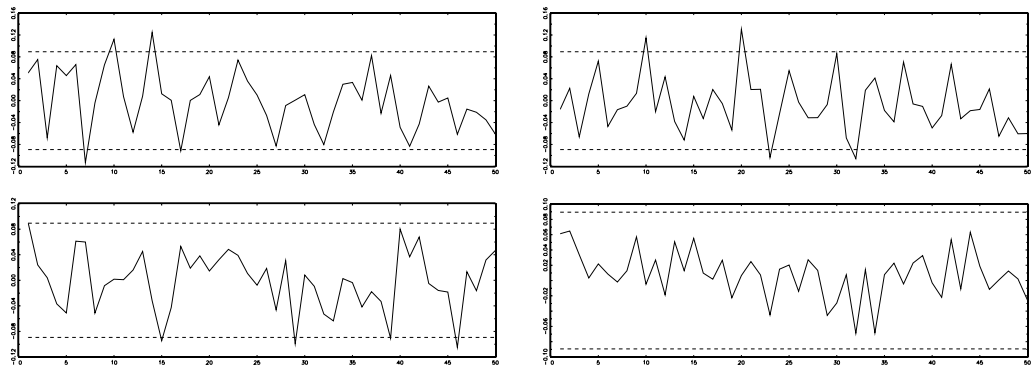


Figure 7: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) from the MMEM(1,2;1,1) model estimated with 500 observations. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

The necessity of using the mixture-MEM model specification is underscored when inspecting the histogram of probability integral transforms with the MEM(1,2) model (Figure 8). With no mixture component, 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 9), with autocorrelations from even the level series falling well beyond the confidence interval.

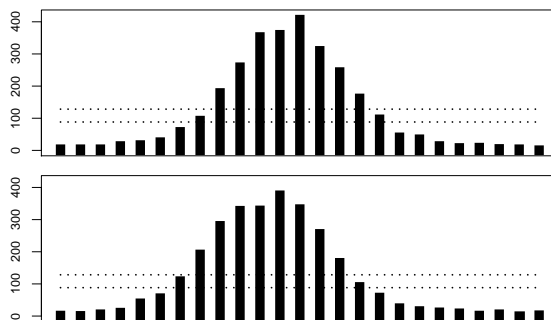


Figure 8: Histograms of probability integral transforms with the MEM(1,2) model for NIKC (upper panel) and NIKP (lower panel). The dotted lines depict the boundaries of the 95% confidence interval.

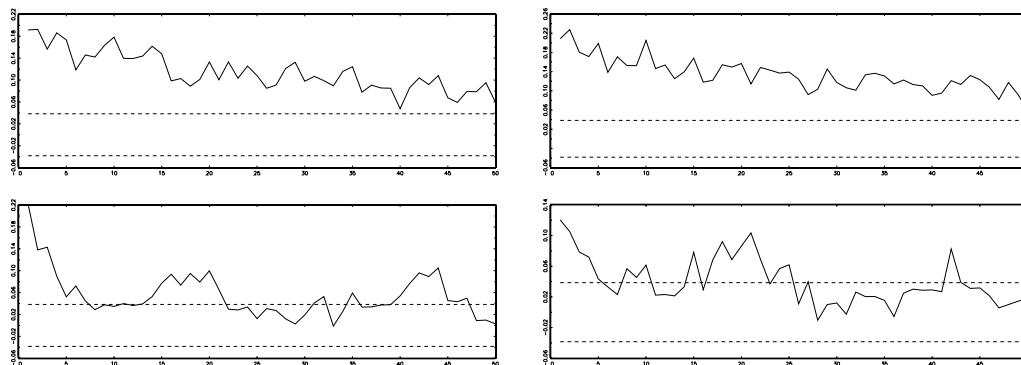


Figure 9: Autocorrelation functions of demeaned probability integral transforms (upper panels) and their squares (lower panels) with the MEM(1,2) model. NIKC on left and NIKP on right. The dotted lines depict the boundaries of the 95% confidence interval.

4 Forecasts

Forecasts were calculated from the chosen model specification of $\text{MMEM}(1,2;1,1)$ as well as from several competing models 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 $\text{MMEM}(1,2;1,1)$ model estimated from the entire in-sample, forecasts were calculated from the $\text{MMEM}(1,2;1,1)$ 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 practise, the forecasts from MMEM models are calculated as shown in Equation 7:

$$\widehat{IV}_{t+1} = \pi \hat{\mu}_{1,t+1} + (1 - \pi) \hat{\mu}_{2,t+1} \quad (7)$$

For comparison, ARIMA models were also estimated for the log time series of NIKC and NIKP. This was done in order to see if the MEM specification has added value over more traditional time series models. The chosen specifications are ARIMA(2,0,1) and ARIMA(1,1,1). The former is based on values of the Schwarz Information Criterion for models estimated from the level series and the latter on the previous finding that for the VIX Volatility Index, the ARIMA(1,1,1) specification is best suited for the differenced time series (Ahoniemi (2006)).

ARIMAX variants of both these models were also estimated, where the exogenous regressors are the positive and negative log returns of the Nikkei 225 index. The returns of the underlying index have been found to be significant explanatory variables for changes in IV, and the separation of positive and negative returns allows for asymmetric effects: negative shocks often raise volatility more than positive shocks. The ARIMAX model is estimated as in Equation 8, where POS and NEG are the positive and negative log returns of the Nikkei 225 index. All ARIMA and ARIMAX models were estimated with rolling samples, with coefficients updated each day.

$$IV_t = c_i + \sum_{i=1}^q \alpha_i IV_{t-i} + \sum_{i=1}^p \beta_i \epsilon_{t-i} + \delta_1 POS_{t-1} + \delta_2 NEG_{t-1} + \epsilon_t \quad (8)$$

The forecast performance of the various models is summarized in Table 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, 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. 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, the choice of 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 (2006) for the VIX index, whose sign was predicted accurately on 62% 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. Also, the daily

	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
ARIMA(2,0,1)	329	67.7%	4.71	321	66.0%	6.06
ARIMAX(2,0,1)	332	68.3%	4.58	323	66.5%	6.03
ARIMA(1,1,1)	327	67.3%	4.71	323	66.5%	6.12
ARIMAX(1,1,1)	326	67.1%	4.52	318	65.4%	6.12

Table 5: Correct sign predictions (out of 486 trading days) and mean squared errors. The best values within each column are in boldface. Mean squared errors for the ARIMA models are calculated by returning log forecasts to the original level.

updating of coefficients becomes important, as the directional accuracy improves if using updating rather than fixed coefficients. In other words, a sample period of 500 observations is sensitive to the choice of sample period, and parameter stability is not achieved.

The forecast performance of the MEM(1,2) model falls short of that of the MEMEM(1,2;1,1) estimated from the in-sample, but is no poorer than that of the MEMEM(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.

When including ARIMA models in the comparison, the MEMEM(1,2;1,1) model remains the best predictor for NIKC despite the fact that the ARIMAX models contain more information due to the inclusion of exogenous variables. However, for NIKP, the ARIMAX(2,0,1) and the ARIMA(1,1,1) model outperform the MEMEM(1,2;1,1) model, and the ARIMA(2,0,1) model is just as accurate.

		Actual outcome		
		<i>Up</i>	<i>Down</i>	<i>Total</i>
Forecast	<i>Up</i>	188	94	282
	<i>Down</i>	56	148	204
	<i>Total</i>	244	242	486

Table 6: NIKC 2x2 contingency table for the MEMEM(1,2;1,1) model with updating coefficients

		Actual outcome			
		<i>Up</i>	<i>Zero</i>	<i>Down</i>	<i>Total</i>
Forecast	<i>Up</i>	178	1	101	280
	<i>Down</i>	60	3	143	206
	<i>Total</i>	238	4	244	486

Table 7: NIKP 2x2 contingency table for the MEMEM(1,2;1,1) model with updating coefficients

Tables 6 and 7 show 2×2 contingency tables with forecasts from the MEMEM(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.

For the ARIMA models, the contingency tables are more balanced. Tables 8 and 9 and show 2×2 contingency tables for the ARIMA(1,1,1) model.

		Actual outcome		
		<i>Up</i>	<i>Down</i>	<i>Total</i>
Forecast	<i>Up</i>	161	76	237
	<i>Down</i>	83	166	249
	<i>Total</i>	244	242	486

Table 8: NIKC 2x2 contingency table for the ARIMA(1,1,1) model.

		Actual outcome			
		<i>Up</i>	<i>Zero</i>	<i>Down</i>	<i>Total</i>
Forecast	<i>Up</i>	162	1	83	246
	<i>Down</i>	76	3	161	240
	<i>Total</i>	238	4	244	486

Table 9: NIKP 2x2 contingency table for the ARIMA(1,1,1) model.

$$MSE = \frac{1}{N} \sum_N (\widehat{IV}_t - IV_t)^2 \quad (9)$$

When evaluating the mean squared errors (calculated as in Equation 9) of the various forecast series, values for NIKC are again superior to those for NIKP (see Table 5). From the family of MEM models, the MMEM(1,2;1,1) model estimated with 500 observations and updating coefficients emerges as the best specification. 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.

For NIKC, the ARIMA(2,0,1) and ARIMAX(1,1,1) models outperform all MEM models when compared with mean squared error. However, all mean squared errors for NIKP ARIMA models are higher than those for most MEM models.

Forecast diagnostics

The value of the obtained directional forecasts can be assessed with the market timing test developed by Pesaran and Timmermann (1992), and the 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 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 6. For NIKP, the days when the actual outcome was 0 are dropped from the analysis in order to run the test. The PT test statistic is computed as in Equation 10 (this version of the test statistic is due to Granger and Pesaran (2000)).

$$PT = \frac{\sqrt{N}KS}{\left(\frac{\hat{\pi}_f(1-\hat{\pi}_f)}{\hat{\pi}_a(1-\hat{\pi}_a)} \right)^{1/2}} \quad (10)$$

where

$$KS = \frac{N_{uu}}{N_{uu}+N_{du}} - \frac{N_{ud}}{N_{ud}+N_{dd}}$$

$$\hat{\pi}_a = \frac{N_{uu}+N_{du}}{N}$$

$$\hat{\pi}_f = \frac{N_{uu} + N_{ud}}{N}$$

N_{uu} is the number of days when both the actual outcome and forecast are up, N_{dd} is the number of days when both the actual outcome and forecast are down, N_{ud} is the number of days when the forecast is up but the actual outcome is down, N_{du} is the number of days when the forecast is down but the actual outcome is up, KS is the Kuiper score, $\hat{\pi}_a$ is the probability that the actual outcome is up, and $\hat{\pi}_f$ is the probability that the outcome is forecast to be up. The limiting distribution of the PT test statistic is $N(0, 1)$ when the null hypothesis is true.

	<i>NIKC</i>		<i>NIKP</i>	
	<i>PT statistic</i>	<i>p-value</i>	<i>PT statistic</i>	<i>p-value</i>
MMEM(1,2;1,1) - updating	8.617	0.000	7.515	0.000
MMEM(1,2;1,1) - fixed	8.701	0.000	7.534	0.000
MMEM(1,2;1,1) - updating; 500 obs.	7.653	0.000	6.642	0.000
MMEM(1,2;1,1) - fixed; 500 obs.	6.722	0.000	6.028	0.000
MEM(1,2) - updating	6.898	0.000	6.451	0.000
MEM(1,2) - fixed	6.546	0.000	5.820	0.000
ARIMA(2,0,1)	7.927	0.000	7.452	0.000
ARIMAX(2,0,1)	8.195	0.000	7.642	0.000
ARIMA(1,1,1)	7.748	0.000	7.614	0.000
ARIMAX(1,1,1)	7.686	0.000	7.149	0.000

Table 10: Pesaran-Timmermann test statistics and their p-values.

The PT test shows that all the evaluated directional forecast series are statistically significant, as the test statistic has a p-value of 0 for all forecast series. The values for the test statistic and their p-values are summarized in Table 10.

SPA test: to be added soon.

5 Conclusions

A multiplicative error model with two alternative mean equations and two alternative gamma distributions for the error term was estimated 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. Without the mixture components, the model is a much worse fit to the data.

Measured with directional accuracy, forecasts calculated from various MMEM models outperform those from ARIMA models for the time series of call IV. ARIMA models fare slightly better than MMEM models for the put-side implied volatility. Again, the lack of a mixture component leads to poorer forecasts. When mean squared errors are used for forecast evaluation, MMEM models are superior for put IV, with ARIMA models leading to lower values for call IV.

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. A mixture specification seems to be necessary in order to obtain the best possible results. 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 MMEM models.

REFERENCES

- Ahoniemi, K. (2006), 'Modeling and Forecasting Implied Volatility - an Econometric Analysis of the VIX Index', *HECER Discussion Paper No. 129*
- Bauwens, L. & Giot, P. (2003), 'Asymmetric ACD models: Introducing price information in ACD models', *Empirical Economics*, 28, 709-731
- 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
- Brooks, C. & Oozer, M.C. (2002), 'Modeling the Implied Volatility of Options on Long Gilt Futures', *Journal of Business Finance & Accounting*, 29, 111-137
- 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
- Diebold, F.X. & Mariano, R.S. (1995), 'Comparing Predictive Accuracy', *Journal of Business & Economic Statistics*, 13, 253-263
- Engle, R.F. (2002), 'New Frontiers for ARCH models', *Journal of Applied Econometrics*, 17, 425-446
- Engle, R.F. & Gallo, G.M. (2006), 'A multiple indicators model for volatility using intra-daily data', *Journal of Econometrics*, 131, 3-27
- Engle, R.F. & Russell, J.R. (1998), 'Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data', *Econometrica*, 66, 1127-1162
- Fernandes, M. & Grammig, J. (2006), 'A family of autoregressive conditional duration models', *Journal of Econometrics*, 130, 1-23
- Ghysels, E., Gouriéroux, C. & Jasiak, J. (2004), 'Stochastic volatility duration models', *Journal of Econometrics*, 119, 413-433
- Granger, C.W.J & Pesaran, M.H. (2000), 'Economic and Statistical Measures of Forecast Accuracy', *Journal of Forecasting*, 19, 537-560
- Hansen, P.R. (2005), 'A Test for Superior Predictive Ability', *Journal of Business & Economic Statistics*, 23, 365-380
- 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
- Lanne, M. (2006), 'A Mixture Multiplicative Error Model for Realized Volatility', *Journal of Financial Econometrics*, 4, 594-616
- Manganelli, S. (2005), 'Duration, volume, and volatility impact of trades', *Journal of Financial Markets*, 8, 377-399
- Meitz, M. & Teräsvirta, T. (2006), 'Evaluating Models of Autoregressive Conditional Duration', *Journal of Business & Economic Statistics*, 24, 104-124

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

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