Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 25 (2009) 239–258
international journal of forecasting
www.elsevier.com/locate/ijforecast

Joint modeling of call and put implied volatility

Katja Ahoniemi[a,b,*], Markku Lanne[c,b]

[a] *Helsinki School of Economics, Finland*
[b] *HECER, Finland*
[c] *University of Helsinki, Finland*

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.

Keywords: Implied volatility; Option markets; Volatility forecasting; MEM models; Impulse responses

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

* Corresponding author at: Helsinki School of Economics, Finland.
E-mail addresses: katja.ahoniemi@hse.fi (K. Ahoniemi), markku.lanne@helsinki.fi (M. Lanne).

0169-2070/$ - see front matter © 2009 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2009.01.012

be determined by no-arbitrage conditions, making the 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 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 body of 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 the 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, Mayhew, and Stivers (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.[1] Other studies that attempt to forecast IV or to utilize the information contained in IV to trade in option markets include Harvey and Whaley (1992), Noh, Engle, and Kane (1994), and Poon and Pope (2000). Reliable forecasts of implied volatility can benefit option traders, and many other market participants as well: all investors with risk management concerns could also 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 US 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 US 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.

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.[2] 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

[1] The data set used by 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.

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

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 of the above-mentioned MEM applications consider univariate models, but Cipollini, Engle, and Gallo (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 is 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 have elapsed. 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 2 discusses the differences in the markets for call and put options in more detail. Section 3 describes the bivariate mixture multiplicative error model estimated in this paper. Section 4 presents the data, model estimation results, and diagnostic checks of the chosen model specification. Impulse response functions are discussed in Section 5, and the forecasts are evaluated in Section 6. Section 7 concludes.

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; Bollen & Whaley, 2004; Dumas, Fleming, & Whaley, 1998). Garleanu, Pedersen, and Poteshman (2006) document that end users (non-market makers) of options have a net long position in S&P 500 index puts, and that the net demand for low-strike options (such as OTM puts) is higher than the demand for high-strike options. The results of Chan, Cheng, and Lung (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 the 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 by, e.g., Bollen & 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. However, 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 & Vishny, 1997). When market makers face unhedgeable risk, they must be compensated for bearing this risk through option prices. 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 period 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 the implied volatility.

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. Consider the following bivariate model

$$\mathbf{v}_t = \mu_t \odot \varepsilon_t, \quad t = 1, 2, \ldots, T,$$

where the conditional mean

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

ε_t is a stochastic positive-valued error term such that $E(\varepsilon_t|\mathfrak{F}_{t-1}) = \mathbf{1}$ 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 was outlined by 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 with achieving positivity arise.

The multiplicative structure of the model was suggested for volatility modeling in the univariate case

$$f_{t-1}(v_{1t}, v_{2t}; \theta) = f_{\varepsilon_1,\varepsilon_2}\left(v_{1t}\mu_{1t}^{-1}, v_{2t}\mu_{2t}^{-1}\right)\mu_{1t}^{-1}\mu_{2t}^{-1}$$
$$= \frac{\lambda^{(\lambda+1)}\left[v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}\right]^{(\lambda-1)/2}\exp\left\{-\frac{\lambda\left(v_{1t}\mu_{1t}^{-1}+v_{2t}\mu_{2t}^{-1}\right)}{1-\rho}\right\}}{\Gamma(\lambda)(1-\rho)\rho^{(\lambda-1)/2}} I_{\lambda-1}\left(\frac{2\lambda\sqrt{\rho v_{1t}v_{2t}\mu_{1t}^{-1}\mu_{2t}^{-1}}}{1-\rho}\right)\mu_{1t}^{-1}\mu_{2t}^{-1}$$

Box I.

by Engle (2002), who proposed using the exponential distribution. However, the gamma distribution nests the exponential distribution, among others, and is therefore more general. In addition, the findings of Lanne (2006, 2007) and Ahoniemi (2007) also lend support to the gamma distribution.

The error term ε_t is assumed to follow a bivariate gamma distribution, which is a natural extension of the univariate gamma distribution used in previous literature (Ahoniemi, 2007; Lanne, 2006, 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, Ouarda, & Bobée, 2001). This particular specification is quite tractable, and is thus well suited for our purposes. Collecting the parameters into the vector $\theta = (\tau_1, \tau_2, \lambda, \rho)$, the density function can be written as

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

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 from Ahoniemi (2007) based on univariate models 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 in Box I.

Consequently, the conditional log-likelihood function can be written as [3]

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

and the model can be estimated with the maximum likelihood method (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. 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 calmer and more volatile time periods. Different parameter values can be assumed to better describe periods of larger shocks compared to periods of smaller shocks, and error terms

[3] Specifically, for observation t,

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

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 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 ε_t is a mixture of $\varepsilon_t^{(1)}$ and $\varepsilon_t^{(2)}$ with mixing probability π, and that $\varepsilon_t^{(1)}$ and $\varepsilon_t^{(2)}$ follow the bivariate gamma distribution with parameter vectors θ_1 and θ_2, respectively. In other words, the error term is $\varepsilon_t^{(1)}$ with probability π and $\varepsilon_t^{(2)}$ 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

$$l_T(\theta) = \sum_{t=1}^{T} l_{t-1}(\theta) = \sum_{t=1}^{T} \ln\left[\pi f_{t-1}^{(1)}(v_{1t}, v_{2t}; \boldsymbol{\theta}_1) + (1-\pi) f_{t-1}^{(2)}(v_{1t}, v_{2t}; \theta_2)\right],$$

where $f_{t-1}^{(1)}(v_{1t}, v_{2t}; \theta_1)$ and $f_{t-1}^{(2)}(v_{1t}, v_{2t}; \theta_2)$ are given by the equation in Box I, with θ replaced by θ_1 and θ_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 θ. 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 4.3).

Table 1
Descriptive statistics for NIKC and NIKP for the full sample, 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

4. Estimation results

4.1. Data

The data set in this study covers 3194 daily closing observations from the period 1.1.1992–31.12.2004, and was obtained from Bloomberg Professional Service (see Fig. 1). The first eleven years of the full sample, or 1.1.1992–31.12.2002, comprise the in-sample period, with 2708 observations. The final two years, 2003 and 2004, are left as the 486-day out-of-sample period to be used for forecast evaluation.

The call-side (put-side) implied volatility time series is calculated as an unweighted average of the Black-Scholes implied volatilities from the 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 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 US 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 1% level. The correlation between NIKC and NIKP is 0.877 in the full sample, but the correlation has varied between 0.534 in 1996 and 0.924 in 1994.

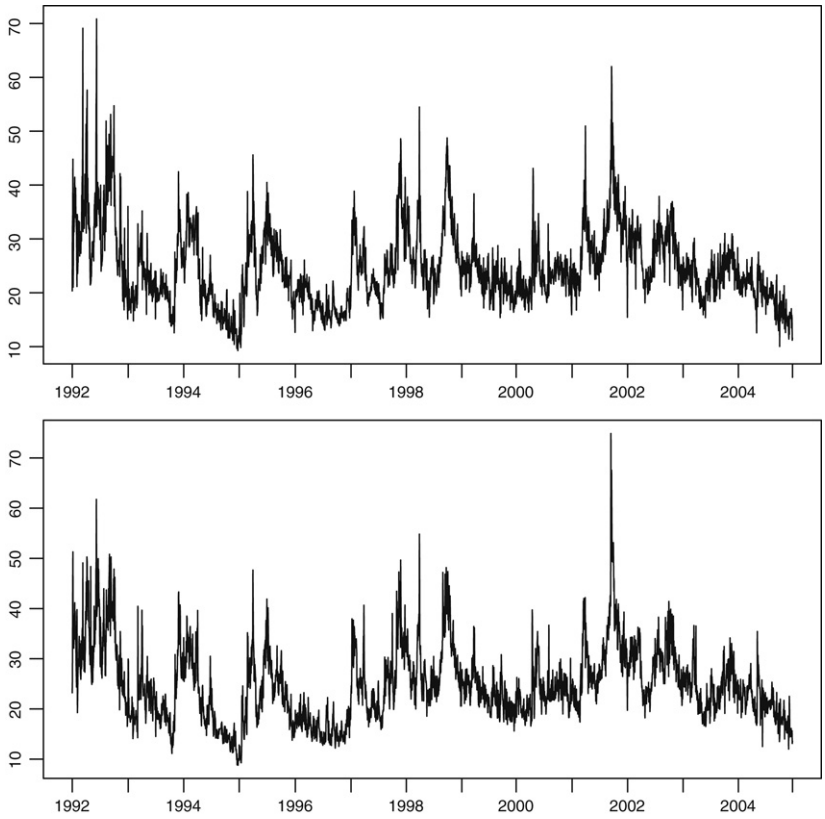

Fig. 1. Nikkei 225 index call (upper panel) and put (lower panel) implied volatility, 1.1.1992–31.12.2004.

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 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 US market, 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 by the addition (see Section 4.3 for more details on diagnostic checks). The level of IV is lowest on Fridays for both call and put options,[4] 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 smallest 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) show that the IV in the S&P 100 index option market tends to be lowest on Fridays and rise on Mondays. Peña, Rubio, and Serna (1999) find that in the Spanish stock index market,

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

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 3. Let μ_{mt} denote the conditional mean of the 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, respectively.

The specifications of the conditional means are

$$\mu^C_{mt} = \omega^C_m + \sum_{i=1}^{q_C} \alpha^C_{mi} v_{C,t-i} + \sum_{i=1}^{r_C} \psi^C_{mi} v_{P,t-i} + \sum_{i=1}^{s_C} \delta^{CP}_{mi} D_i v_{P,t-i} + \sum_{j=1}^{p_C} \beta^C_{mj} \mu^C_{m,t-j}$$

and

$$\mu^P_{mt} = \omega^P_m + \sum_{i=1}^{q_P} \alpha^P_{mi} v_{P,t-i} + \sum_{i=1}^{r_P} \psi^P_{mi} v_{C,t-i} + \sum_{i=1}^{s_P} \delta^{PP}_{mi} D_i v_{P,t-i} + \sum_{j=1}^{p_P} \beta^P_{mj} \mu^P_{m,t-j},$$

where the ψs are the coefficients of lagged cross terms, and D_i has a 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^{CP}_{mi} and δ^{PP}_{mi}). This specification is later referred to as the unrestricted model.[5]

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,

$$\mu^C_{mt} = \omega^C_m + \sum_{i=1}^{q_C} \alpha^C_{mi} v_{C,t-i} + \sum_{i=1}^{s_C} \delta^{CC}_{mi} D_i v_{C,t-i} + \sum_{j=1}^{p_C} \beta^C_{mj} \mu^C_{m,t-j}$$

and

$$\mu^P_{mt} = \omega^P_m + \sum_{i=1}^{q_P} \alpha^P_{mi} v_{P,t-i} + \sum_{i=1}^{s_P} \delta^{PP}_{mi} D_i v_{P,t-i} + \sum_{j=1}^{p_P} \beta^P_{mj} \mu^P_{m,t-j}.$$

The estimation results for both the unrestricted and restricted models are presented in Table 2. The lag structure is selected based on the statistical significance of the coefficients, as well as on autocorrelation diagnostics (see Section 4.3). The parameter values for all ωs, αs and βs meet the Nelson and Cao (1992) constraints discussed in Section 3. Also, the coefficients of the cross terms (ψs) and dummies (δs) are positive, so positivity is guaranteed in the model. The estimated values of β^C_{11}, ω^P_1, and α^P_{12} 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 8% of the trading days in the in-sample period. The estimated shape (and scale) parameters of the error distribution differ considerably between the two regimes, with residuals being more dispersed in the second regime. Fig. 2 shows the joint error density of the unrestricted model with the parameters

[5] 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 the p-values from likelihood ratio tests validated the constraining of the other dummies to zero.

Table 2
Estimation results for the BVMEM model. Standard errors calculated from the final Hessian matrix are given in parentheses. (**) indicates statistical significance at the 1% level, and (*) significance at the 5% level.

	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)

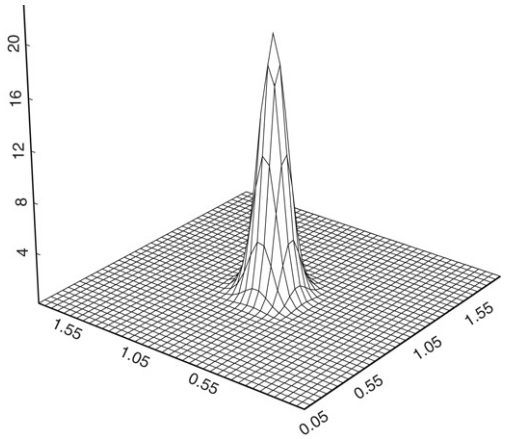

Fig. 2. Estimated density of the error terms in the first regime of the unrestricted model.

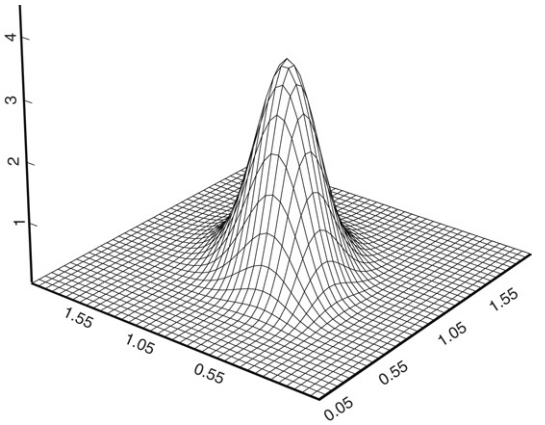

Fig. 3. Estimated density of the error terms in the second regime of the unrestricted model.

estimated for the first regime, while the error density for the second regime is depicted in Fig. 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, 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 Figs. 2 and 3. The coefficients of the cross terms ψ are significant at the 1% level in the first, more common regime, and jointly significant in the second regime (the p-value from the LR test was 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. The 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 βs 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 performance, 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 being 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 in 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 to those in Figs. 2 and 3, and are therefore not displayed.

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 transformations of the data. This method was suggested by Diebold, Gunther, and Tay (1998) and extended to the multivariate case by Diebold, Hahn, and Tay (1999). The probability integral transform in the univariate case (for one IV series) is obtained as

$$z_t = \int_0^{y_t} f_{t-1}(u)\mathrm{d}u, \tag{1}$$

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 as 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 transformations, which allows for determining uniformity, and autocorrelation functions of demeaned probability integral transforms and their squares. The graphical approach allows one to easily identify where a possible model misspecification arises. Fig. 4 presents the 25-bin histogram and autocorrelations for $z_t^{C|P}$, and Fig. 5 does the same for $z_t^{P|C}$ for the unrestricted model. Figs. 6 and 7 present the equivalent graphs for z_t^C and z_t^P, respectively.

Most columns of the histograms fall within the 95% confidence intervals, which are 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

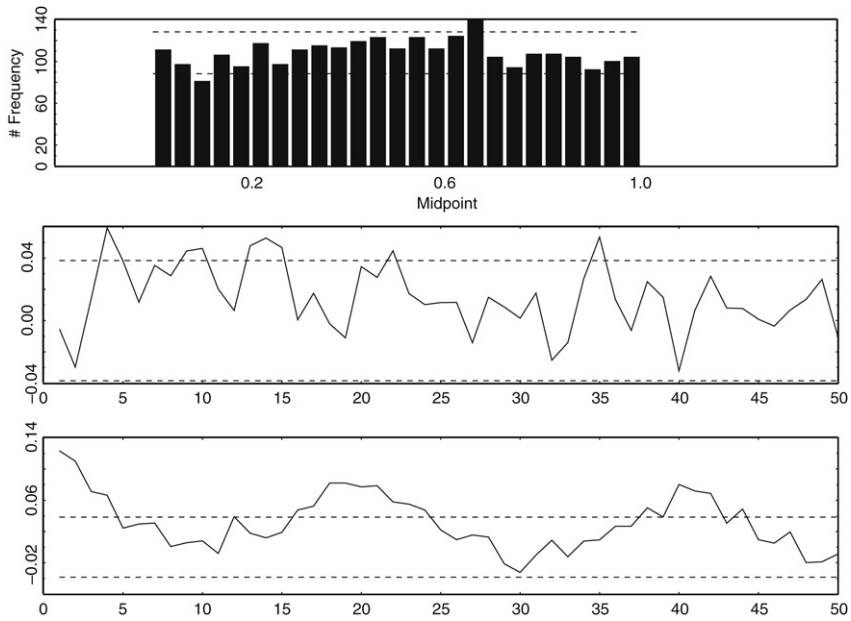

Fig. 4. Diagnostic evaluation of $z_t^{C|P}$: NIKC conditional on NIKP. Histograms of probability integral transforms (upper panel), and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines indicate the boundaries of the 95% confidence interval.

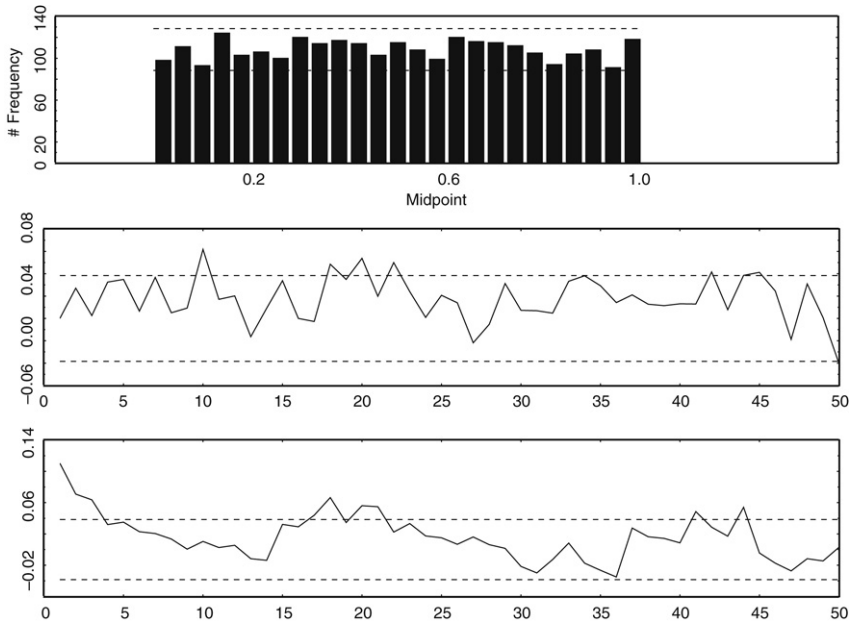

Fig. 5. Diagnostic evaluation of $z_t^{P|C}$: NIKP conditional on NIKC. Histograms of probability integral transforms (upper panel), and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines indicate the boundaries of the 95% confidence interval.

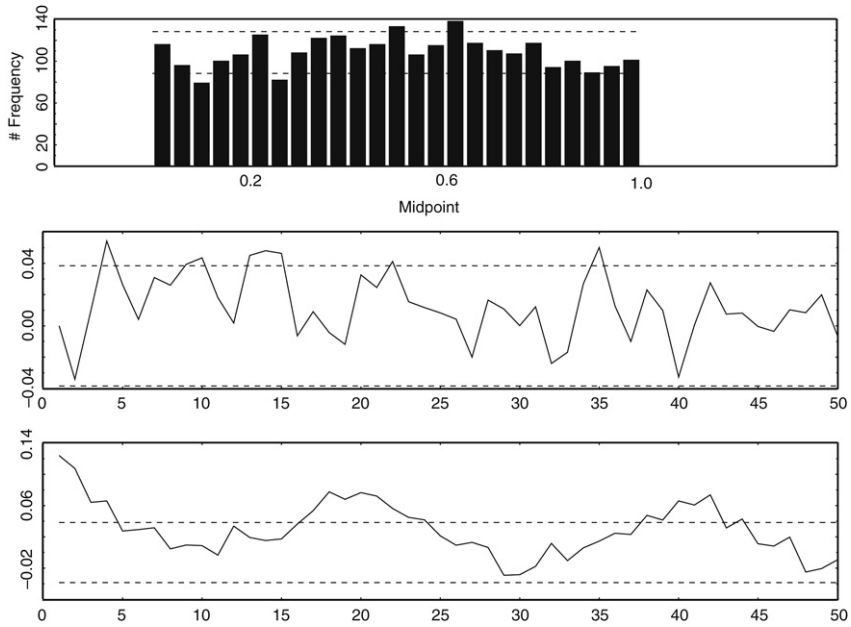

Fig. 6. Diagnostic evaluation of z_t^C: Marginal NIKC. Histograms of probability integral transforms (upper panel), and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines indicate the boundaries of the 95% confidence interval.

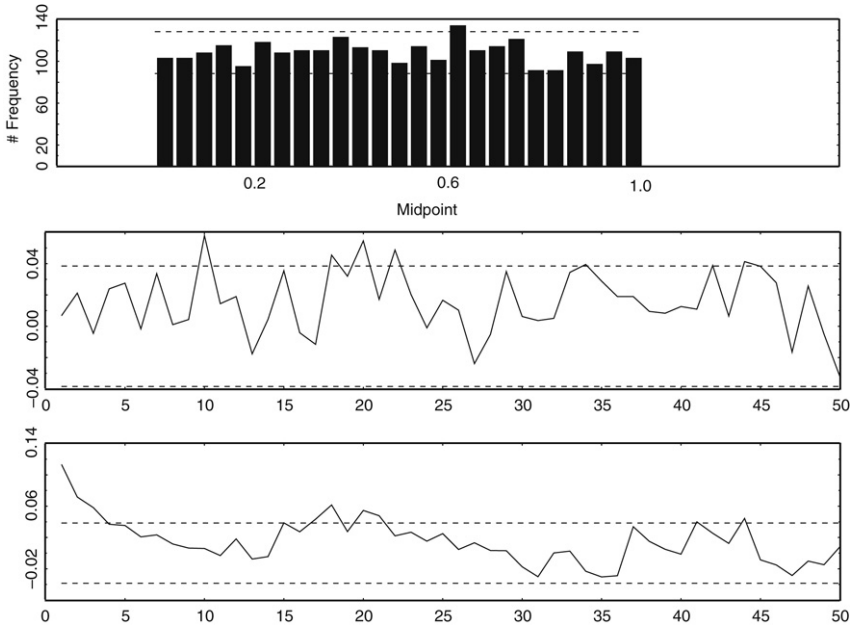

Fig. 7. Diagnostic evaluation of z_t^P: Marginal NIKP. Histograms of probability integral transforms (upper panel), and autocorrelation functions of demeaned probability integral transforms (middle panel) and their squares (lower panel). The dotted lines indicate the boundaries of the 95% confidence interval.

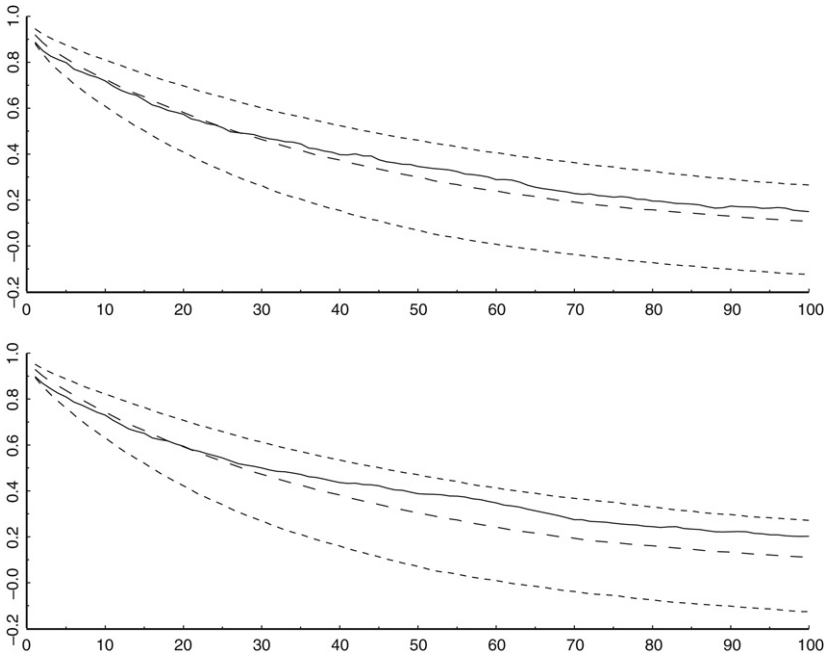

Fig. 8. NIKC (upper panel) and NIKP (lower panel) autocorrelation functions. The solid lines depict the ACFs estimated from the full sample of the data, the lines with long dashes are the ACFs implied by the mixture-BVMEM model with 100,000 simulated data points, and the lines with short dashes are the 95% confidence bands around the ACFs.

5% (but not at the 10%) level.[6] There seems to be some remaining autocorrelation in the squares of the demeaned probability integral transformations. 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 to capture the time-varying volatility of implied volatility, as autocorrelation in the squares is a sign of conditional heteroskedasticity.

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.[7] The improvement in diagnostics due to the removal of cross effects is surprising, as the cross terms are statistically significant and improve the forecasts (see Section 6 for discussion on forecasts).

In order to verify that our unrestricted model takes into account the high persistence in the data, we compare the autocorrelation functions (ACFs) estimated from the call and put IV data with those calculated from data simulated with our model. Fig. 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 3194 data points (equal to the full sample size), and forming a band that encompasses 95% of the autocorrelations pointwise at

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

[7] 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.

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.[8] However, the imbalance in the histograms is not as severe as with the univariate models of Ahoniemi (2007), indicating that, even for models without a mixture structure, joint modeling improves the fit to the Nikkei 225 IV data somewhat.

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 each of the 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.[9]

We generate the impulse responses by simulating data according to the conditional mean profiles method proposed by Gallant, Rossi, and Tauchen (1993). It turns out that after approximately 40 periods, the effects of all considered shocks go to zero. Therefore, we present impulse responses up to 40 periods (trading days) ahead. The calculation of the impulse response functions proceeds as follows: we generate 1000 series of 40 random error terms from gamma distributions with the shape and scale parameters estimated above. Also, we generate 1000 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 1000 paths, forty days into the future, are simulated from that point onwards with the random error terms, random regime indicators, and estimated parameter values. Another set of 1000 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 1000 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 Fig. 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 between 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 of 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 in the other series.[10] 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.[11]

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

[9] 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 the results more easily readable from graphs.

[10] 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.

[11] 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

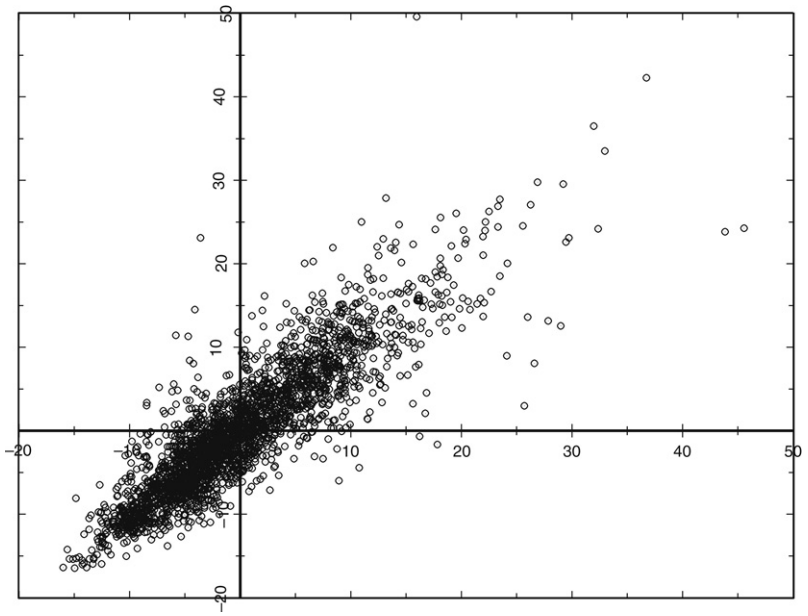

Fig. 9. Scatter plot of demeaned NIKC (x-axis) and demeaned NIKP (y-axis) for the in-sample period (1.1.1992–31.12.2002).

The impulse responses for the three shock combinations are presented in Figs. 10–12, 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 (Fig. 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 US markets). Also, trading volumes for puts are higher (measured by the 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 the evidence from Jorion (1995), who estimated the half-life of a variance shock to be 17 days for the IV of currency options.

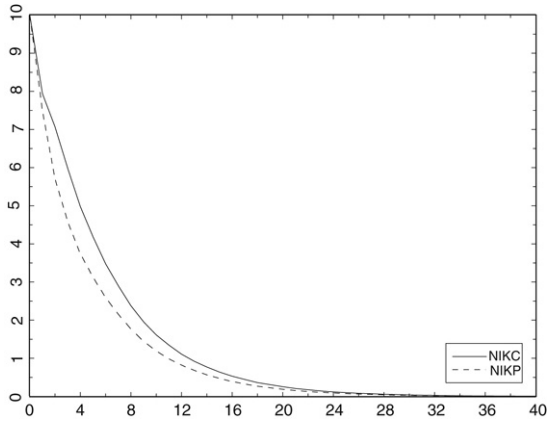

Fig. 10. Impulse response function for a shock of (10, 10) with the unrestricted model.

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

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 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.

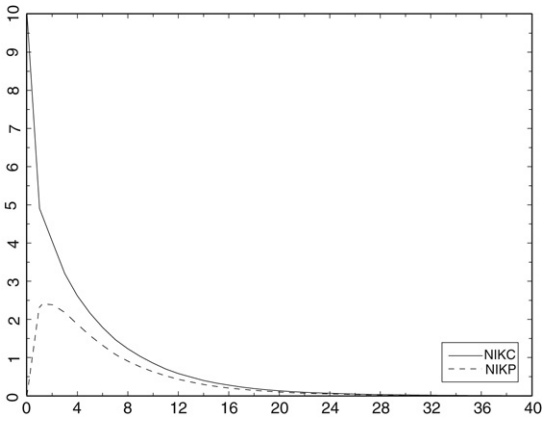

Fig. 11. Impulse response function for a shock of (10, 0) with the unrestricted model.

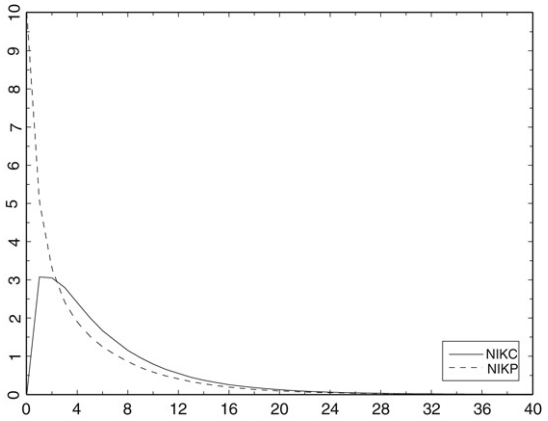

Fig. 12. Impulse response function for a shock of (0, 10) with the unrestricted model.

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 other 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, at times quite considerably, 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 trading 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.

6. Forecasts

In this section, we turn our attention to the forecasting ability of the two models outlined in Section 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 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 one of 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 the parameter values are not stable over time, there can be added value in updating them before calculating each new forecast. When the 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 period, namely 2708.

The one-step-ahead forecasts are evaluated in terms of both directional accuracy and MSE. Although

Table 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.

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

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 of the direction of change (up or down) is correct, ceteris paribus.

The forecast results are summarized in Table 3. The directional accuracy of the bivariate model in the two-year out-of-sample period is superior to the performance of univariate models. The BVMEM model predicts the direction of change correctly on 348 out of 486 days for NIKC, and on 351 days for NIKP. This is in contrast to the results of 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 in updating parameter values each day. It clearly improves directional accuracy for NIKC, 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 too often forecasts a move downwards, but the restricted model too often forecasts a move upwards.

A useful statistical test of the sign forecasting ability of the BVMEM models is the test statistic presented by Pesaran and 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 of the series of forecasts in Table 3, so the null hypothesis of predictive failure can be rejected at the 1% level for all four forecast series.

The values of the MSE in Table 3 indicate that more accurate forecasts can be obtained for NIKC. The mean squared errors are lower than with the univariate models of Ahoniemi (2007), even with the restricted model. The Diebold and 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 5% level for NIKC and the 1% 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.[12]

Overall, the results obtained for the Nikkei 225 index option market are superior to those obtained for, e.g., the US market. Ahoniemi (2008) finds that ARIMA models can predict the correct direction of change in the VIX index on 58.4% of trading days at best. 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

[12] 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 for NIKC is 332 out of 486 at best, and 318 for NIKP. The detailed results for this third model specification are available from the authors.

Table 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.

	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

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 presents the MSEs for the 482 five-step-ahead forecasts that could be calculated within the chosen out-of-sample period. 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 10% level when comparing the restricted and unrestricted models with updating coefficients for NIKC, but not at the 5% level. For the corresponding NIKP values (8.79 and 9.36), the null is not rejected. The results for NIKP are better than those of 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.

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 of 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. The diagnostics show that the joint model specification is a good fit to the data, and the 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 at 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.

Acknowledgements

The authors wish to thank two anonymous referees, and 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.

References

Ahoniemi, K. (2007). *Multiplicative models for implied volatility*. HECER Discussion Papers, No. 172.

Ahoniemi, K. (2008). *Modeling and forecasting the VIX index*. Unpublished working paper.

Bates, D. S. (1991). The crash of '87: Was it expected? The evidence from options markets. *Journal of Finance*, *46*, 1009–1044.

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.

Brooks, C., & Oozeer, M. C. (2002). Modelling the implied volatility of options on Long Gilt futures. *Journal of Business Finance and Accounting*, *29*, 111–137.

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.

Buraschi, A., & Jackwerth, J. (2001). The price of a smile: Hedging and spanning in option markets. *Review of Financial Studies*, *14*, 495–527.

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.

Cipollini, F., Engle, R. F., & Gallo, G. M. (2006). *Vector multiplicative error models: Representation and inference*. NBER Working Paper, No. 12690.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *64*, 247–254.

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.

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., 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.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

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

Ederington, L., & Guan, W. (2002). Why are those options smiling? *Journal of Derivatives*, *10*(2), 9–34.

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.

Figlewski, S. (1989). Options arbitrage in imperfect markets. *Journal of Finance*, *44*, 1289–1311.

Fleming, J. (1999). The economic significance of the forecast bias of S&P 100 index option implied volatility. In P. Boyle, G. Pennacchi, & P. Ritchken (Eds.), *Advances in futures and options research*: *Vol. 10* (pp. 219–251). Stamford, Conn: JAI Press.

Gallant, A. R., Rossi, P. E., & Tauchen, G. (1993). Nonlinear dynamic structures. *Econometrica*, *61*, 871–907.

Garleanu, N. B., Pedersen, L. H., & Poteshman, A. M. (2006). *Demand-based option pricing*. CEPR Discussion Papers, No. 5420.

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.

Hentschel, L. (2003). Errors in implied volatility estimation. *Journal of Financial and Quantitative Analysis*, *38*, 779–810.

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, *13*, 433–451.

Jorion, P. (1995). Predicting volatility in the foreign exchange market. *Journal of Finance*, *50*, 507–528.

Lanne, M. (2006). A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics*, *4*, 594–616.

Lanne, M. (2007). Forecasting realized exchange rate volatility by decomposition. *International Journal of Forecasting*, *23*, 307–320.

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.

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.

Mayhew, S., & Stivers, C. (2003). Stock return dynamics, option volume, and the information content of implied volatility. *Journal of Futures Markets*, *23*, 615–646.

Mo, H., & Wu, L. (2007). International capital asset pricing: Evidence from options. *Journal of Empirical Finance*, *14*, 465–498.

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].

Nelson, D. B., & Cao, C. Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics*, *10*, 229–235.

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.

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.

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

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

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance*, *49*, 771–818.

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *Journal of Finance*, *52*, 35–55.

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.

Katja Ahoniemi holds an M.Sc. in Economics from the Helsinki School of Economics and is preparing her Ph.D. dissertation on implied volatility modeling and forecasting. She is currently a graduate school fellow of the Finnish Doctoral Programme in Economics.

Markku Lanne is Professor of Economics at the University of Helsinki, Finland. He graduated from the University of Helsinki, and holds a Ph.D. in Economics. His current research interests are in the areas of financial and macroeconometrics.