# Discussion Paper

Discussion Paper No. 93-19

# Forecasting Volatility and Option Pricing for Exchange-Rate Dynamics: A Comparison of Models

Juergen Kaehler

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

International Finance Serie

# Forecasting Volatility and Option Pricing for Exchange-Rate Dynamics: A Comparison of Models

by
Juergen Kaehler

Z.E.W. and University of Mannheim

December 1993

**Abstract**

This paper explores the applicability of static and dynamic models to capture the stylized facts of exchange-rate dynamics. The static models (mixture of distributions, compound Poisson process, generalized Student distribution) are compatible with leptokurtosis and can be characterized as scale-compounded distributions. The dynmic models (GARCH, GARCH-t, EGARCH, Markov-switching model), on the other hand, are compatible with both leptokurtosis and heteroskedasticity. In a comparison of the candidate models, it is found that the dynamic models do indeed achieve a better fit to the data than the static models. However, in forecasting experiments the dynamic models can outperform a 'naive' model of constant variances only with respect to unbiasedness but not with respect to precision. Furthermore, the paper examines the implications of the static and dynamic models for the pricing of foreign-currency options by simple simulations. Static models show significang option-price effect only when the maturity is short. GARCH and EGARCH models, on the other hand, imply options prices which are higher than Black-Scholes prices for the full range of moneyness. Only the Markov-switching model is compatible with the observed 'smile effects' on option markets.

# 1. Introduction

The concept of decision making under uncertainty is central to the theory of finance. Therefore, the stochastic specification of financial models is of fundamental importance. It is common practice in finance to assume that rates of return and price dynamics in speculative markets follow a normal distribution. The assumption of normality is both convenient and natural. It is convenient because this assumption simplifies considerably theoretical analysis and empirical applications. It is also a natural assumption because the central limit theorem in probability theory gives a justification for the normal distribution under rather weak conditions. However, in the seminal papers of Mandelbrot (1963) and Fama (1965) strong evidence against the normal distribution was found for price dynamics in commodity markets and stock markets.

In the 1960's and in the first half of the 1970's their findings lead to much research on the distributional properties of stock returns and the implications for portfolio analysis. However, the interest into this area virtually ceased with the finding that daily and weekly stock returns exhibit strong non-normality but that monthly returns are only slightly non-normal. If one uses monthly data, it was argued, one would be again on safe grounds (see e.g. Fama (1976), Ch. 1).

More recently, a renewed interest in distributional properties of financial data emerged. This renewed interest emerged from the scrutiny of the assumptions underlying the Black-Scholes model of option pricing. The ubiquitous assumption of normality cannot as easily be maintained in option pricing as it can be in portfolio analysis because the natural time horizon in empirical option analysis is the short-run corresponding to the continuous-time models, i.e. one would typically use daily or perhaps weekly data in empirical option analysis.

Therefore, at least three questions come up in this context. First, what is the nature of the observed non-normality? Second, which model can capture the observed non-normality? And third, what are the consequences for the pricing of options?

The data to be analysed are the exchange rates of the dollar against the German mark, the British pound, the Swiss franc (sfr), and the Japanese yen. The data are on a daily basis, but also weekly, monthly, and quarterly data are used. In these cases, end-of-period data were derived from the daily exchange rates. The data range from July 1st, 1974 to December 31st, 1987. Due to differences in bank holidays between countries, there are different numbers of observations in the daily data: 3386 for the mark, 3417 for the pound, 3392 for the sfr and 3365 for the yen. For all currencies, the number of observations in the weekly series is 704, in the monthly series it is 161 and in the quarterly series it is 53. Data source is the IMF's International Financial Statistics and the monthly reports of the Swiss National Bank. The data are analysed in the form of first differences in the logarithm of exchange rates, i.e. $x_t = \Delta e_t = \log E_t - \log E_{t-1}$.

## 2. Statistical Properties of Exchange-Rate Data

A natural approach to test for normality is to compare theoretical moments with empirical ones. Since the normal distribution is symmetric, its odd central moments are zero. Symmetry may be tested by computing the third moment of the standardized variable $z_t = (x_t - \overline{x})/s$ , where $\overline{x}$ is the mean and $s$ is the standard deviation of $x$ . This gives the Chalier measure of skewness. Extensive statistical analysis shows that the null hypothesis of symmetry cannot be rejected (see Kaehler (1989)).

The fourth moment of $z_t$ is the kurtosis $\beta_2$. It can be shown that $\beta_2 \geq 1$ and that for the normal distrbution $\beta_2 = 3$. With respect to the kurtosis of the normal distribution, the excess $\beta_2^*$ is defined as $\beta_2^* = \beta_2 - 3$. Kurtosis is a location- and scale-free measure which increases when probability mass is shifted from the shoulders of the distribution into the tails and centre of the distribution, i.e. kurtosis measures both tail weight and peakedness (see Balanda and MacGillivray (1988)). This dual character of kurtosis is a consequence of the fact that any movement of mass from the shoulders to the centre of the distrbution must be accompanied by a simultaneous shift of mass into the tails (et vice versa) if the variance, by which $\beta_2$ is standardized, is to remain constant.

**Table 1**
**Test for mesokurtosis**

| | mark | pound | sfr | yen |
|---|---|---|---|---|
| day | 8.32 *** | 8.36 *** | 8.89 *** | 8.00 *** |
| week | 5.84 *** | 7.36 *** | 4.96 *** | 7.03 *** |
| month | 3.87 ** | 4.15 *** | 4.19 *** | 3.62 |
| quarter | 2.67 | 2.72 | 2.77 | 2.62 |

Significance levels: 1 percent (***), 5 percent (**), 10 percent (*)

A test of the null hypothesis $H_0$:$\beta_2 = 3$ is a test for mesokurtosis with the two-sided alternatives of platykurtic $(\beta_2 < 3)$ and leptokurtic $(\beta_2 > 3)$ distributions. The values of $\beta_2$ are reported in table 1 for the series of $x_t$. As the table shows, there is extremely strong leptokurtosis in the daily and weekly series. In the monthly series, the null hypothesis of mesokurtosis can be rejected at the 0.05 level for 3 exchange rates, whereas no rejection of $H_0$ is possible for any of the quarterly series. This means that leptokurtosis is essentially a property of short-run exchange-rate dynamics. It is only moderately inherent in monthly series and vanishes completely in quarterly data.

The other strong statistical property of short-run price dynamics (or returns) in speculative markets is heteroskedasticity. Here I measure heteroskedasticity by the autocorrelation function (ACF) for the squared data $y_t = x_t^2$. As a summary measure, I apply the Ljung-Box statistic

$$Q(K) = T(T-2) \sum_{k=1}^{K} \hat{r}^2(k)/(T-k) \tag{1}$$

where $\hat{r}(k)$ is the estimated autocorrelation coefficient at lag $k$, i.e.

$$\hat{r}(k) = \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) / \sum_{t=1}^{T} (y_t - \bar{y})^2. \tag{2}$$

The Ljung-Box test is a portmanteau test against white noise. It follows asymptotically a $\chi^2$ distribution with $(K-m)$ degrees of freedom, where $m$ is the number of estimated parameters.

The Ljung-Box statistic for all four exchange rates at four different time horizons each is reported in table 2. $Q$ is estimated at lag $K = 15$. It is evident that there is a strong rejection of the $H_0$ of white noise in daily and weekly data only. In the daily series, the ACF for squared exchange-rate movements is significant at all lags up to 15 for all four exchange rates. For weekly data, the estimated autocorrelation coefficients exceed the conventional confidence limits of $\pm 2\sqrt{T}$ at various lags $k$.

**Table 2**
**Ljung-Box statistic for squared data**

| | mark | pound | sfr | yen |
|---|---|---|---|---|
| day | 355.2 *** | 507.3 *** | 561.0 *** | 432.2 *** |
| week | 61.5 *** | 123.9 *** | 98.2 *** | 52.8 *** |
| month | 12.3 | 12.9 | 11.8 *** | 25.0 ** |
| quarter | 12.0 | 7.4 | 12.1 | 5.4 |

Significance levels: see table 1

The results of this analysis and of a more comprehensive study of the statistical properties of exchange rate data (see Kaehler, 1989) can be summarized as:

i) All series of exchange-rate dynamics show approximate serial independence and no periodicities. The series have a constant mean at zero and a symmetric distribution.

ii) Short-run exchange-rate dynamics (i.e. daily and weekly changes) are characterized by heteroskedasticity as well as peakedness and fat tails in distribution.

iii) Medium-run exchange-rate dynamics show no heteroskedasticity and have a frequency distribution which is approximately normal.

## 3. Stochastic Models of Exchange-Rate Dynamics

In this section I will introduce several stochastic models of exchange-rate dynamics which are supposed to capture the main empirical regularities of short-run exchange-rate data. The models can be classified into two groups. The first group consists of four models which are static in the sense that the conditional probability distribution for $x_t$ is the same for all $t$. The four models are the mixture of normal distributions, the compound Poisson process, the generalized Student distribution and the family of stable distributions. These models have very different probabilistic backgrounds but they can all be viewed as compound normal distributions where an independent probability distribution is attached to the variance of a normal variable. They can, therefore, be called scale-compounded distribution models. Table 3 illustrates how these models are usually characterized and how they can be described as scale-compounded distributions.

**Table 3**
**Static Models**

| *Model* | *Characterization* | *Variance function* |
|---|---|---|
| Mixture | $f(x \mid \Psi) = \sum_{j=1}^{J} p_j f_j(x \mid \mu, \sigma_j)$ | Multinomial |
| Compound Poisson | $X_t = Y_{1,t} + Y_{2,t} + \ldots + Y_{N,t} + V_t$<br>N ~ Poisson ( $\lambda dt$ ) | Poisson |
| Student | $\frac{d}{dx} f(x) = -\frac{(x-a)f(x)}{c_0 + c_1 x + c_2 x^2}$ | Inverted Gamma |
| Stable Paretian | $\log \Phi_X(u) = i\delta u - \mid \gamma u \mid^{\alpha}$ | Positive stable Paretian |

The mixture of normal distributions is a weighted sum of normal densities with different means and/or variances, where the weights are positive and sum to 1 (i.e. they can be interpreted as probabilities). Here I only consider two-component mixtures with $\mu_1 = \mu_2 = 0$.

The compound Poisson process can be described as a sum of N random variables where N is itself random with a Poisson distribution. The $Y_{i,t}$ are independent and identically distributed with a normal distribution and $V_t$ represents background noise also having a normal distribution.

The generalized (two parameter) Student distribution can be derived within the Pearson system of frequency curves which is characterized by the differential equation for the density funtion $f(x)$ given in table 3. The generalized Student distribution obtains within this system for $a = c_1 = 0$ and $c_0 > 0, c_2 > 0$.

Finally, the family of stable Paretian distribution is related to a generalization of the central limit theorem without the assumptions of finite means and variances. In general, closed form expressions for the density or distribution function for the members of this family are not available but symmetric stable distributions can be described by the log-characteristic function given in table 3.

A unifying framework for these seemingly ad-hoc models may be provided by viewing these models as scale-compounded normal distributions. Within this framework, these models differ only with respect to the distribution function which is attached to the variance of the normal distribution. As shown in table 3, the mixture model attaches a multinomial distribution to the variance, the compound Poisson process attaches a Poisson distribution, the generalized Student distribution attaches an inverted Gamma distribution and the symmetric stable Paretian distributions attach positive stable distributions with $\alpha < 1$ to the random variance. This unifying framework of scale-compounded distribution is also useful because it can be related to the stylized fact of leptokurtosis. It can be shown that every scale-compounded normal distribution is leptokurtic (see Kaehler (1993b)).

However, the static models assume that draws from these distributions are independent and, therefore, they cannot capture heteroskedasticity. Obviously, the modelling of heteroskedasticity requires dynamic models. Research in this area has recently been conducted along two lines. The first approach is based on the continuous-time modelling in finance and supplements the diffusion process for the price of the underlying asset (usually in the form of a geometric Brownian motion) by a diffusion process for the volatility in the form of a geometric Brownian motion or an Ornstein-Uhlenbeck process. This approach is surveyed by Taylor (1992) and Clewlow and Xu (1992).

In this paper I will only deal with the second approach which can be called the "econometric" approach. In recent years the modelling of financial volatility by ARCH models, which were introduced by Engle (1982), became very popular (see the survey of Bollerslev et al. (1992)). There is now a plethora of variants of ARCH-type models. Here I consider the following three variants: The GARCH (1,1) model is given by:

(3) $$x_t = \varepsilon_t \sqrt{h_t}$$

(4) $$h_t = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 h_{t-1}$$

where $\varepsilon_t$ is Gaussian white noise with unit variance and $h_t$ is the variance of $x_t$ conditional on information available at time $t$.

In the GARCH-t model, the standard normal distribution of $\varepsilon_t$ is replaced by a Student t-distribution. This modification, introduced by Bollerslev (1987), was motivated by the fact that the "residuals" $\varepsilon_t = x_t/\sqrt{h_t}$ of GARCH models often had significant leptokurtosis. The idea behind the GARCH-t model, therefore, is to capture very high leptokurtosis by fatter tails of the unconditional distribution.

The third variant is the EGARCH (1,1) model of Nelson (1991). It is based on (3) but replaces (4) by

(4') $$h_t = \exp\{a_0 + a_{1a}\varepsilon_{t-1} + a_{1b}(|\varepsilon_{t-1}| - E|\varepsilon_{t-1}|) + b_1 \log h_{t-1}\}.$$

Nelson (1991) suggested this functional form of the conditional variance equation to deal with the problems of negative variance estimates, of asymmetric variance effects and of non-stationary variances.

Finally, I will consider the Markov-switching model which is a simple extension of the mixture model. Analytically, the mixture model may be decomposed into two independent random variables where the first variable determines the component $j$, which is drawn with probability $p_j$, and the second variable has a conditional normal distribution with variance $\sigma_j^2$. The first variable may be regarded as a state variable $s_t$ which can take on values $j = 1,\ldots,J$, where $J$ is the number of components in the mixture. The Markov-switching model assumes that $s_t$ follows a time-homogeneous first-order Markov process characterized by the transition probabilities

(5) $$p(s_t = j \mid s_{t-1} = i) = p_{ij}.$$

It should be noted that the four dynamic models do not only incorporate heteroskedasticity but also leptokurtosis. Furthermore, they imply convergence to normality under time aggregation (see Bollerslev (1986), Nelson (1991), and Kaehler and Marnet (1993)).

## 4. Comparison of Candidate Models

In this section, I will compare the candidate models with respect to two general principles: their ability to capture the characteristics of short-run exchange-rate dynamics, i.e. their goodness of fit, and their ability to forecast exchange-rate volatility. From the eight candidate models, introduced in the previous section, one can dismiss the stable Paretian distributions because it is not compatible with the statistical property of convergence to normality, but also direct estimation of the model (see Kaehler (1993a)) shows that it is not appropriate for the exchange-rate data. This leaves us with seven candidate models: the two-component scale-mixture of normal distributions (mixture, for short), the compound Poisson process (Poisson, for short), the generalized Student distribution (Student, for short), the two-component Markov-switching model (Markov, for short), the GARCH (1,1) model, the GARCH-t(1,1) model, and the EGARCH (1,1) model. In addition, the Gaussian random walk (Gauss, for short) shall serve as a benchmark model to judge the performance of the seven candidate models.

As mentioned in the previous section, all candidate models have a leptokurtic distribution, but it is still interesting to examine whether the models underestimate or overestimate the magnitude of leptokurtosis in the data. Table 4 shows the kurtosis of the exchange-rate samples (as in table 1) and the implied kurtosis of the candidate models. For daily data, the actual kurtosis is between 8.00 and 8.89, and for weekly data it is between 4.96 and 7.36. The mixture model, the compound Poisson process, the Markov-switching model, and the EGARCH model underestimate in general the kurtosis of the data, the only exception being the weekly franc series where the implied kurtosis of the estimated compound Poisson process is larger than the kurtosis of the data. In general, the underestimation is stronger for daily than for weekly data.

The generalized Student distribution, the GARCH model, and the GARCH-t model lead to an overestimation of kurtosis. With the exception of the daily and weekly pound, the estimates of the GARCH model and all estimates of the GARCH-t model implied non-stationarity of variances. Therefore, kurtosis cannot be finite for those models. Table 4 shows that the GARCH models also imply non-existing kurtosis for the two pound series. For the generalized Student distribution, the condition for finite kurtosis is only violated for the daily pound series but the only two series for which the implied kurtosis has roughly the magnitude of the actual kurtosis are the daily mark series and the weekly pound series.

The fact that some models imply infinite kurtosis raises the more fundamental question whether the true data-generating process has a finite kurtosis. It is difficult to answer this question from the kurtosis of the data because every empirical kurtosis is necessarily finite. But there are some reasons to conjecture that the data-generating process has finite kurtosis. A data-generating process with infinite kurtosis would produce empirical values of kurtosis which would vary strongly and which would tend to increase with an increase of observations. However, the empirical values for the

**Table 4**
**Kurtosis and implied kurtosis of candidate models**

| | | mark | pound | sfr | yen |
|---|---|---|---|---|---|
| day | sample | 8.32 | 8.36 | 8.89 | 8.00 |
| | mixture | 5.81 | 4.86 | 6.60 | 5.65 |
| | Poisson | 4.84 | 5.07 | 5.76 | 4.74 |
| | Student | 8.62 | ∞ | 16.41 | 860.14 |
| | Markov | 5.28 | 4.49 | 4.72 | 7.28 |
| | GARCH | ∞ | ∞ | ∞ | ∞ |
| | GARCH-t | ∞ | ∞ | ∞ | ∞ |
| | EGARCH | 5.32 | 4.35 | 5.21 | 4.20 |
| week | sample | 5.84 | 7.36 | 4.96 | 7.03 |
| | mixture | 4.32 | 5.86 | 4.80 | 5.76 |
| | Poisson | 4.77 | 4.99 | 5.08 | 4.92 |
| | Student | 13.31 | 8.90 | 9.34 | 22.21 |
| | Markov | 3.65 | 3.70 | 4.39 | 3.78 |
| | GARCH | ∞ | ∞ | ∞ | ∞ |
| | GARCH-t | ∞ | ∞ | ∞ | ∞ |
| | EGARCH | 4.16 | 4.03 | 4.19 | 4.63 |

daily and weekly data are all in the same order of magnitude. Furthermore, other empirical studies of exchange-rate data produced the same order of magnitude for kurtosis statistics.

Besides leptokurtosis, heteroskedasticity is the other strong empirical regularity of short-run exchange-rate dynamics. Of course, only the dynamic models can depict heteroskedasticity but the question is: how much of the heteroskedasticity do these models capture? Table 5 summarizes the results on the residual heteroskedasticity and compares it to the heteroskedasticity of the data. Residual heteroskedasticity is here measured as the Ljung-Box statistic at lag 15 of the standardized data $x_t/\sqrt{h_t}$. As table 5 shows, the dynamic models exhibit residual heteroskedasticity which is drastically lower than the one in the data. It is only the Markov-switching model that has significant residual heteroskedasticity for all daily series. The ARCH-Type models capture heteroskedasticity in all series very well, with the exception of the daily sfr series. One may conclude, therefore, that the GARCH, the GARCH-t and the EGARCH models are superior to the Markov-switching model in depicting heteroskedasticity.

As a final criterion to judge the goodness of fit of the candidate models, the Schwarz information criterion (SIC) will be employed. A direct comparison of models by the likelihood-ratio statistic is not possible because the models are not nested but the SIC, defined by SIC = r logT - 2 L* (where r is the number of parameters estimated, T is the number of observations and L* is the value of the maximised likelihood), is

**Table 5**
**ACF of squared data and residual heteroskedasticity of dynamic models**

| | | mark | pound | sfr | yen |
|---|---|---|---|---|---|
| day | sample | 355.2 *** | 507.3 *** | 561.0 *** | 432.2 *** |
| | Markov | 50.0 *** | 215.4 *** | 132.6 *** | 155.3 *** |
| | GARCH | 22.9 * | 8.4 | 54.1 *** | 3.2 |
| | GARCH-t | 28.1 ** | 0.4 | 59.0 *** | 3.3 |
| | EGARCH | 24.1 * | 8.6 | 67.2 *** | 4.1 |
| week | sample | 61.5 *** | 123.9 *** | 98.2 *** | 52.8 *** |
| | Markov | 14.9 | 94.9 *** | 18.9 | 20.4 |
| | GARCH | 29.6 ** | 8.0 | 11.7 | 4.4 |
| | GARCH-t | 24.0 * | 8.1 | 13.1 | 4.8 |
| | EGARCH | 20.1 | 9.7 | 9.7 | 6.3 |

Significance levels: see Table 2

also based on likelihoods and it corrects for the number of estimated parameters. Table 6 reports the SIC of all candidate models together with the SIC of Gaussian white noise as a benchmark. The ranking of the models according to SIC is given in brackets. Several observations may be drawn from table 6. First, all seven candidate models are clearly superior to the benchmark model and this is especially evident in the daily series. Second, the dynamic models are superior to the static models for all daily and weekly series. Within the group of static models the mixture model has in general the highest value of SIC and hence the worst performance, whereas an overall ranking between the compound Poisson process and the generalized Student distribution is not possible. Third, within the group of dynamic models, the GARCH-t model achieves by far the best result. It has the lowest value of SIC for all series. The second best model seems to be the Markov-switching model.

Finally, I will compare the candidate models with respect to their ability to forecast volatility. There are at least two reasons why forecasting performance is important for model evaluation in this case. First, from an econometric point of view, poor forecasting performance of a model which fits well within the sample would indicate a lack of structural stability. Second, from an economic point of view, financial markets are most interested in good forecasts. Dealers in derivative markets often say that they "trade" volatility. Of course volatility is not a traded asset and, more important, it is not observable. What is meant by "trading volatility" is the fact that dealers buy options when the implicit volatility of the option, calculated from the Black-Scholes model, is smaller than the expected future volatility and they sell options when the implicit volatility is larger than the expected volatility.

**Table 6**
**Comparison of models by SIC**

| | | mark | pound | sfr | yen |
|---|---|---|---|---|---|
| day | Gauss | 7018.3 | 6873.3 | 8298.6 | 6315.5 |
| | mixture | 6664.5 (7) | 6300.6 (7) | 7811.5 (7) | 5805.7 (7) |
| | Poisson | 6630.1 (6) | 6207.7 (5) | 7794.8 (6) | 5724.9 (5) |
| | Student | 6610.0 (5) | 6295.5 (6) | 7765.0 (5) | 5781.4 (6) |
| | Markov | 6167.2 (4) | 5911.8 (2) | 7365.3 (4) | 5287.9 (2) |
| | GARCH | 6064.3 (2) | 5981.1 (4) | 7299.7 (3) | 5430.1 (4) |
| | GARCH-t | 5937.1 (1) | 5303.3 (1) | 7103.1 (1) | 4725.7 (1) |
| | EGARCH | 6071.6 (3) | 5966.1 (3) | 7297.0 (2) | 5355.1 (3) |
| week | Gauss | 2547.7 | 2512.7 | 2770.0 | 2358.1 |
| | mixture | 2482.0 (7) | 2448.6 (7) | 2715.2 (3) | 2271.2 (7) |
| | Poisson | 2467.1 (5) | 2443.4 (6) | 2714.4 (5) | 2266.8 (6) |
| | Student | 2476.4 (6) | 2433.8 (5) | 2716.9 (7) | 2263.3 (5) |
| | Markov | 2414.7 (2) | 2410.6 (3) | 2653.8 (2) | 2189.8 (2) |
| | GARCH | 2442.4 (4) | 2422.7 (4) | 2663.0 (3) | 2237.4 (4) |
| | GARCH-t | 2404.1 (1) | 2336.4 (1) | 2631.1 (1) | 2141.0 (1) |
| | EGARCH | 2439.2 (3) | 2408.5 (2) | 2664.7 (4) | 2219.6 (3) |

Since this study has concentrated on the modelling of variance effects and has neglected mean effects, the forecasting preformance will only be evaluated with respect to volatility. The benchmark of the forecasting performance is provided by a simple model which extrapolates the volatility of the past as a constant into the future, i.e. the "naive" volatility forecast at time $\tau$ for the next $k$ periods is given by

(6) $$\tilde{\sigma}^2_{\tau+k} = \frac{1}{\tau}\sum_{t=1}^{\tau} x_t^2 \qquad k = 1,\ldots,K,$$

where $x_t$ is the first difference in the logarithm of the exchange rate at time $t$. Note that it is assumed throughout that the mean is zero. These naive forecasts also serve to represent volatility forecasts from the static models which would also produce constant volatility forecasts. It would be possible to estimated each static model up to time $t$ and to compute the implied variances from the parameter estimates; It will show, however, in the next section that the implied variances of the static models are very close to the historical variances as defined in (6).

Volatility forecasts from the dynamic models, on the other hand, are non-trivial. For the Markov-switching model, they can be derived along the following lines. The volatility forecast at time $\tau$ for the $k$-th period in the future is given by

$$(7) \qquad \hat{\sigma}_{\tau+k} = \sigma_1^2 p(s_{\tau+k}=1 \mid x_\tau) + \sigma_2^2 p(s_{\tau+k}=2 \mid x_\tau)$$

$$= (\sigma_1^2 - \sigma_2^2) p(s_{\tau+k}=1 \mid x_\tau) + \sigma_2^2$$

where $\sigma_1^2$ and $\sigma_2^2$ are the variances in states 1 and 2, respectively, and $p(s_{\tau+k}=1 \mid x_\tau)$ is the probability of being in state 1 in $\tau+k$ given $x_\tau$. This probability may be decomposed into

$$(8) \qquad p(s_{\tau+k}=1 \mid x_\tau) = \sum_{i=1}^{2} p(s_{\tau+k}=1 \mid s_\tau = i) p(s_\tau = i \mid x_\tau)$$

where $p(s_\tau = i \mid x_\tau)$ is the filter probability of being in state $i$ and $p(s_{\tau+k}=1 \mid s_\tau = i)$ is a $k$-step transition probability. From the Markov-chain structure, one can compute this transition probability (see e.g. Chiang (1980), p. 160) to get after some arithmetic

$$(9) \qquad p(s_{\tau+k}=1 \mid x_\tau) = \{p(s_\tau = 1 \mid x_\tau)(2 - p_{11} - p_{22})(p_{11} + p_{22} - 1)^n$$

$$+ (1 - p_{22}) - (1 - p_{22})(p_{11} + p_{22} - 1)^n\}/(2 - p_{11} - p_{22})$$

where $p_{11}$ and $p_{22}$ are the estimated elements of the transition matrix.

The volatility forecasts of the three GARCH(1,1) variants can be derived in a simple recursive way. From the conditional variance equation

$$(10) \qquad h_\tau = a_0 + a_1 x_{\tau-1}^2 + b_1 h_{\tau-1}$$

one gets the first-period forecast

$$(11) \qquad \hat{h}_{\tau+1} = a_0 + a_1 x_\tau^2 + b_1 h_\tau$$

which involves only observable variables. For the periods $k \geq 2$, the forecasts are

$$(12) \qquad \hat{h}_{\tau+k} = a_0 + a_1 E(x_{\tau+k-1}^2) + b_1 E(h_{\tau+k-1})$$

$$= a_0 + (a_1 + b_1)\hat{h}_{\tau+k-1}.$$

Only minor changes to the first-period forecasts are necessary in the case of the EGARCH model.

The forecasting experiments were conducted by estimating the dynamic models on a "rolling basis". For the daily data, the models were first estimated for the observations from $t = 1$ to $\tau = 1000$. Volatility forecasts were made for the next 20 days and the forecasts were compared with $x_{\tau+k}^2$. In the next step, 100 observations were added, parameters were re-estimated and forecasts were again compared with observations. In this way, parameters and forecasts were computed 23 times for each daily series. For weekly data, the first estimation period includes observations up to $\tau = 220$

and on each step 20 observations were added to the previous subsample. This gives 24 forecast experiments for each of the weekly series. The forecast horizon includes each of the next 20 weeks for every forecast experiment.

The volatility forecasts of the dynamic models and of the "naive" model are compared with respect to mean errors and with respect to root mean square errors (RMSE). The mean error measures the bias of forecasts and RMSE measures the lack of precision of forecasts. The results are summarized in tables 7 and 8. Note that the mean errors and RMSE are averaged over all 20 forecast horizons.

Table 7 shows that the naive model and the Markov-switching model tend to underestimate future volatility since the entries for all eight series in the case of the naive model and for seven series in the case of the Markov-switching model are negative. The GARCH model and the GARCH-t model, on the other hand, tend to overestimate future volatility since all eight entries for GARCH models and seven entries for the GARCH-t models are positive.

**Table 7**
**Volatility forecasts of dynamic models: mean error**

| | | mark | pound | sfr | yen |
|---|---|---|---|---|---|
| day | Naive | -0.217 (5) | -0.184 (4) | -0.005 (1) | -0.269 (4) |
| | Markov | -0.086 (3) | -0.142 (2) | 0.100 (3) | -0.198 (2) |
| | GARCH | 0.116 (4) | 0.165 (3) | 0.464 (5) | 0.144 (1) |
| | GARCH-t | -0.060 (2) | 0.364 (5) | 0.127 (4) | 0.327 (5) |
| | EGARCH | 0.002 (1) | -0.050 (1) | 0.046 (2) | -0.233 (3) |
| week | Naive | -1.193 (5) | -1.060 (3) | -0.615 (3) | -0.765 (5) |
| | Markov | -1.014 (4) | -0.957 (2) | -0.108 (1) | -0.609 (4) |
| | GARCH | 0.077 (1) | 2.571 (4) | 0.985 (5) | 0.188 (1) |
| | GARCH-t | 0.101 (2) | 6.296 (5) | 0.913 (4) | 0.257 (2) |
| | EGARCH | -0.389 (3) | -0.453 (1) | 0.129 (2) | -0.277 (3) |

It is also interesting to compare the models with respect to the absolute mean error for each series. The resulting ranking is given in brackets. The EGARCH model seems to dominate the other models since it finishes first in three out of eight cases. If the rankings are aggregated over all eight series, the EGARCH model obtains an overall ranking of 16,[1] followed by the Markov-switching model with 21, the GARCH model with 24, the GARCH-t model with 29, and the naive model with 30.

[1] It is three times the best model, twice the second best, and three times the third best.

Table 8 reports the results for the RMSE criterion. The ranking among models is quite different. The Markov-switching model obtains the highest precision, i.e. the smallest RMSE, of volatility forecasts for five of the eight series and finishes twice in second place, whereas, quite surprisingly, the naive model is once the best model and five times the second best. With respect to the overall rank sums, the EGARCH model is the third best model with a sum of 20 followed by the GARCH model with 34 and the GARCH-t model with 37.

The overall picture, which emerges from these forecasting experiments, is that it is indeed possible to beat the naive model in the forecasting of volatility but the naive model is only clearly dominated with respect to the mean error where the EGARCH and Markov-switching model have smaller average biases. With respect to the RMSE, only the Markov-switching model performs better than the benchmark model of static variance but numerically this improvement is rather small. It is interesting to note that these results parallel in a way the results of Meese and Rogoff (1983) on the forecasting of exchange-rate levels. Meese and Rogoff (1983) found that asset-market models are not able to outperform a random-walk model in forecasting exchange-rate levels but this was more obvious with respect to the RMSE than with respect to mean errors. Since tables 7 and 8 show that it is less clear with respect to RMSE that the naive volatility model can be outperformed by the dynamic model than with respect to mean errors, there is some correspondence between their results and the results reported here.

**Table 8**
**Volatility forecasts of dynamic models: RMSE**

| | | mark | pound | sfr | yen |
|---|---|---|---|---|---|
| day | Naive | 1.207 (2) | 1.021 (2) | 1.274 (2) | 1.199 (3) |
| | Markov | 1.191 (1) | 1.025 (3) | 1.268 (1) | 1.183 (1) |
| | GARCH | 1.322 (5) | 1.156 (4) | 2.165 (5) | 1.355 (4) |
| | GARCH-t | 1.210 (3) | 1.597 (5) | 1.449 (4) | 1.623 (5) |
| | EGARCH | 1.218 (4) | 1.020 (1) | 1.280 (3) | 1.190 (2) |
| week | Naive | 4.747 (2) | 4.947 (3) | 5.390 (1) | 3.656 (2) |
| | Markov | 4.713 (1) | 4.933 (2) | 5.411 (2) | 3.625 (1) |
| | GARCH | 5.088 (4) | 7.633 (4) | 6.079 (4) | 3.872 (4) |
| | GARCH-t | 5.234 (5) | 13.084 (5) | 6.346 (5) | 3.951 (5) |
| | EGARCH | 5.061 (3) | 4.861 (1) | 5.419 (3) | 3.779 (3) |

In order to gain more insight into the forecasting performance, figure 1 plots the mean errors and RMSE at forecast horizons 1 to 20 for the daily Swiss franc. It is quite striking how similar the patterns of mean errors and RMSE are across forecast horizons. The plot of mean errors shows how the GARCH models tend to overestimate volatility. Recall from the previous chapter that the GARCH model of the daily Swiss franç implied non-stationarity of variances. The same is true for most subperiods and, therefore, the GARCH model tends to overestimate volatility, especially for longer forecast horizons. On the other hand, the naive model produces the smallest forecast errors for all forecast horizons and 9 of its 20 forecast errors are negative.

**Figure 1**
**Forecast errors of volatility at different time horizons: daily sfr**

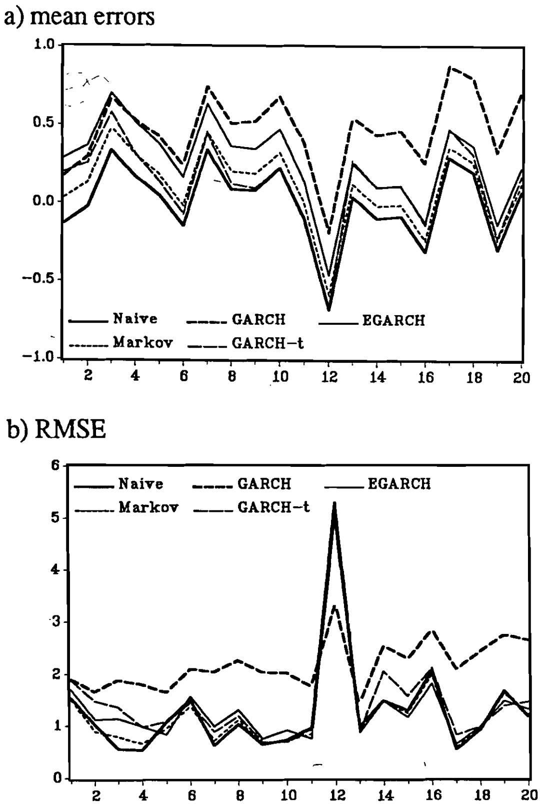



Figure 1 also shows that all models underestimate the volatility of 12 days in the future. This, however, is caused by a single outlier at $t= 1113$ in the second forecast experiment. On Monday, 20th November 1978, the Swiss franc depreciated against the dollar by 5.1 percent. This depreciation came quite unexpectedly and all models underestimate the value of $x_{1113}^2 = 25.85$. The historical variance at $t= 1100$ is 0.68, the Markov-switching model produces a volatility forecast of 1.40, the GARCH model predicts 10.10, the GARCH-t model predicts 1.61, and the EGARCH model predicts 2.31. The plot of RMSE also illustrates that the non-stationary GARCH model tends to give small precision of volatility forecasts but if an outlier occurs, the non-stationary model tends to perform better than stationary models.

To summarize, the three static models and the four dynamic models are clearly superior to a simple random-walk model of the exchange rate with Gaussian increments with respect to goodness-of-fit criteria. Furthermore, the dynamic models have a natural advantage over the static models because they do not only capture leptokurtosis but also heteroskedasticity. However, only the Markov-switching model and the EGARCH model, which do not violate stationarity conditions, are able to outperform a naive model in the forecasting of volatility, but this superiority holds only with respect to unbiasedness and not with respect to precision of forecasts.

## 5. Implications for the Pricing of Foreign-Currency Options

As noted before, the approaches to modelling stochastic volatility can be grouped under the headings "continuous-time-finance approach" and "econometric approach". The aim of the econometric approach is to find a specification of the volatility process which adequately represents the stylized facts of the financial data. A problem with this approach is that it is unclear under which conditions the specified volatility process is compatible with the risk-neutral valuation principle. Duan (1991), however, has established such conditions for the GARCH model.

In this section, I follow the econometric approach to study the impact of leptokurtosis and heteroskedasticity on option pricing. More specifically, I computed call option prices which would obtain under three static models (mixture of distributions, compound Poisson process and Student distribution) and under three dynamic models (Markov-switching model, GARCH model and EGARCH model). The GARCH-t model had to be dismissed from this list because parameter estimates of this model were numerically unstable and implied strong nonstationarity. The GARCH-t model produced also very erratic option prices. It should be stressed, however, that the results should be regarded as preliminary since it is not clear for all models at this stage how option pricing in the framework of the risk-neutral-valuation principle is possible.

Option prices were computed by simulation based on the expected value of the boundary condition, i.e. call option prices were computed as

$$C = \frac{1}{R} \sum_{r=1}^{R} \max\{E_r - B; 0\} \tag{13}$$

where $B$ is the exercise price and $R = 20{,}000$ is the number of repetition in every experiment. The simulations were based on the parameter estimates of the daily pound series which are shown in table 9 along with the stationary variances of the estimated models. It is noteworthy that the implied variances of the static models and the Markov-switching model are very close to the variance of the sample which is 0.437, but the implied variance of the EGARCH model exceeds this variance by about 20 percent and the implied variance of the GARCH model is almost tenfold the variance of the sample data.

**Table 9**
**Parameter estimates and implied variances: daily pound**

| *Model* | *Parameter estimates* | *Stationary variance* |
|---|---|---|
| Mixture | $p = 0.460$<br>$\sigma_1^2 = 0.061$<br>$\sigma_2^2 = 0.756$ | 0.437 |
| Poisson | $\lambda = 1.262$<br>$\sigma_Y^2 = 0.307$<br>$\sigma_V^2 = 0.027$ | 0.414 |
| Student | $\eta = 2.423$<br>$\gamma = 0.887$ | 0.426 |
| Markov | $p_{11} = 0.933$<br>$p_{22} = 0.952$<br>$\sigma_1^2 = 0.073$<br>$\sigma_2^2 = 0.697$ | 0.436 |
| GARCH | $\alpha_0 = 0.007$<br>$\alpha_1 = 0.135$<br>$\alpha_2 = 0.864$ | 4.299 |
| EGARCH | $a_0 = -0.038$<br>$a_{1a} = 0.026$<br>$a_{1b} = 0.300$<br>$b_1 = 0.940$ | 0.528 |

Figures 2 and 3 show the results from the simulation experiments when the current spot rate $E_t$ is varied between 1.40 and 2.20. The time to maturity is set to 20 days and the exercise price B is set to 1.80. For simplicity, the domestic and foreign interest rates are assumed to be zero. Note that, for each spot rate, the computed option prices are based on the same realizations of the random variable (with the exception of the Student distribution, of course) whereas the drawings are distinct for different spot rates.

For the understanding of price differences between Black-Scholes prices and simulated option prices, it is useful to decompose the price effects into different components. Following Jarrow and Rudd (1982), the option price under an arbitrary distribution A can be approximated by a generalized Edgeworth series expansion as:

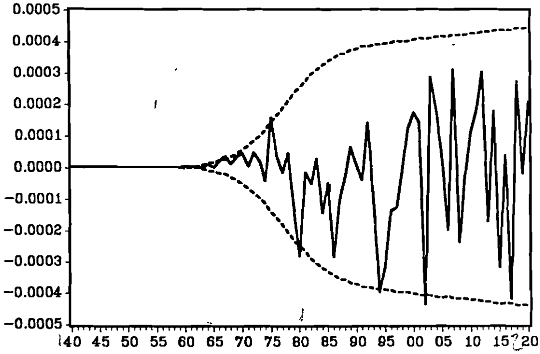
$$
\begin{aligned}
(14) \qquad C_A = C_L &+ \frac{1}{2e^{rt}}[\sigma^2(A) - \sigma^2(L)]f_L(B) \\
&- \frac{1}{6e^{rt}}[\mu_3(A) - \mu_3(L)]\frac{df_L(B)}{dx} \\
&+ \frac{1}{24e^{rt}}\left[\kappa_4(A) - \kappa_4(L) + 3(\sigma_2^2(A) - \sigma_2^2(L))^2\right]\frac{d^2 f_L(B)}{dx^2} \\
&+ \varepsilon(B)
\end{aligned}
$$

where $C_L$ is the Black-Scholes price (based on the log-normal distribution), $e^{-rt}$ is the discount factor, $\sigma^2(A)$ and $\sigma^2(L)$ are the variances of the alternative distribution and the log-normal distribution, respectively, $\mu_3$ is the third central moment (which is related to the skewness $\beta_1$ by $\mu_3 = \beta_1\sigma^3$), $\kappa_4$ is the 4-th cumulant, $f_L$ is the density of the log-normal distribution, and $\varepsilon$ is an error term. Note that $\kappa_4 = \mu_4 - 3\mu_2^2$, where $\mu_j$ is the j-th central moment, i.e. $\kappa_4 > 0$ if the distribution is leptokurtic.

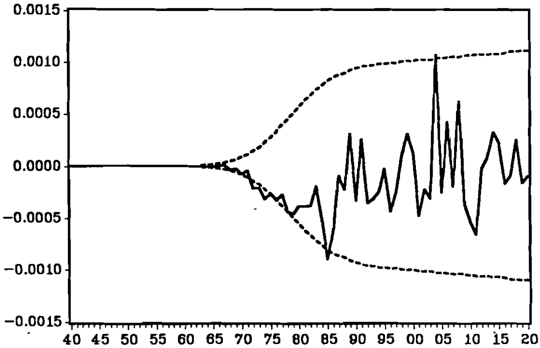
With reference to (14), option price biases can be decomposed into three components. First, option prices under an alternative distribution will, ceteris paribus, be higher if $\sigma^2(A) > \sigma^2(L)$. Second, there is a skewness effect which, however, can be neglected because all models, with the exception of the EGARCH model, have symmetric (stationary) distributions and the asymmetry of the EGARCH model is small and statistically insignificant. The third effect is related to kurtosis and has weights given by the second derivative of $f_L$ which is positive for in-the-money and out-of-the-money options and negative for at-the-money options. Under leptokurtosis, therefore, ceteris paribus $C_A > C_L$ for in-the-money options and out-of-the-money options, whereas $C_A < C_L$ for at-the-money options. Statistically, the at-the-money effect is caused by peakedness and out-of-the-money and in-the-money effects are caused by fat tails.

**Figure 2**
**Spot-rate effect of biases in option prices for static models**

a) Mixture distribution




b) Compound Poisson process




c) Generalized Student distribution

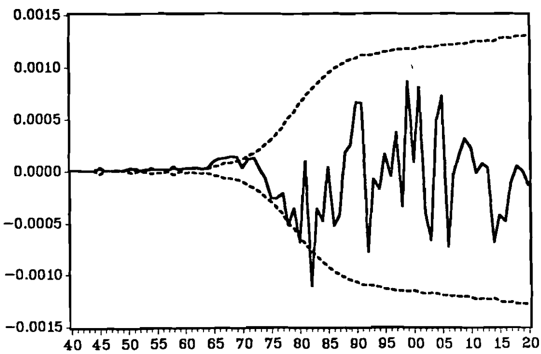
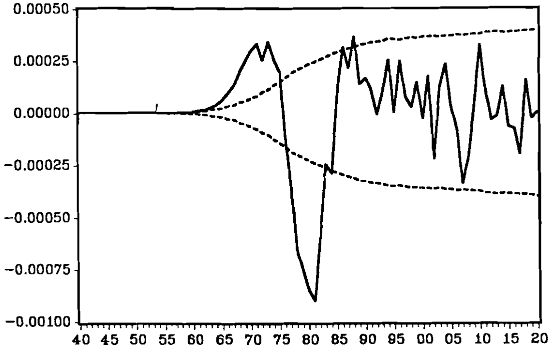
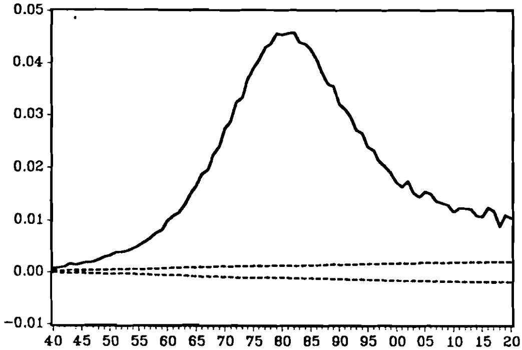


**Figure 3**
**Spot-rate effect of biases in option prices for dynamic models**

a) Markov-switching model




b) GARCH model




c) EGARCH model

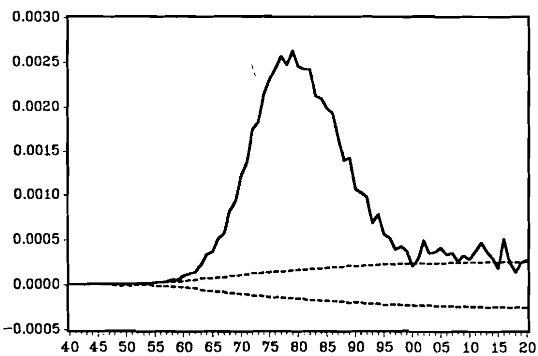


Figure 2 plots the differences between computed option prices of static models and Black-Scholes prices. A negative value indicates that the Black-Scholes price is larger than the simulated price of the corresponding model. The dotted lines give the 95 per cent confidence interval around zero. In general, the price effects of the static models are not very strong. Both the compound Poisson process and Student's distribution exhibit the peakedness effect of at-the-money options but only a few simulated biases are statictically significant. The fat-tail effect of out-of-the-money options shows up only for Student's distribution whereas there are no significant fat-tail effects of in-the-money options for any of the static models. Furthermore, there are no sizeable variance effects, because the variances of the static models are quite close to the sample variance of 0.437, and there are no skewness effects because the models have symmetric distributions.

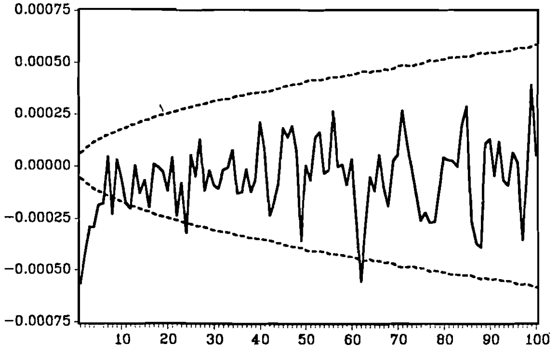
Turning next to the spot-rate effects of the dynamic models in figure 3, one clearly finds a fat-tail effect for out-of-the money options and a peakedness effect for at-the-money options in the case of the Markov-switching model. On the other hand, the fat-tail effect for in-the-money options is rather weak since only two biases for these options are statistically significant at the 5 per cent level. It is interesting to note that the Markov-switching model displays roughly a "smile effect" and that these smile effects also have been derived in the continuous-time approach of modelling stochastic volatility by diffusion processes (see Hull and White (1987). Note, too, that the smile effect is also a stylized fact on foreign-currency option markets. It was found that the implied variance of foreign-currency options is smaller for at-the-money options than for in-the-money and out-of-the-money options (see e.g. Shastri and Wethyavivorn (1987)).

In contrast, both the GARCH and EGARCH model imply that option prices should be significantly higher than Black-Scholes prices over the full range of moneyness. This can be attributed to the fact that both models have a stationary variance which is much larger than the sample variance. As table 9 shows, the stationary variance of the GARCH model is a multiple of the sample variance of 0.437 and the EGARCH's variance of 0.528 exceeds the sample variance by about 20 percent. Since, according to the decomposition of price biases in (14), the variance effect is weighted by the density $f_L$, the price effect is strongest at the money where the density has its maximum.

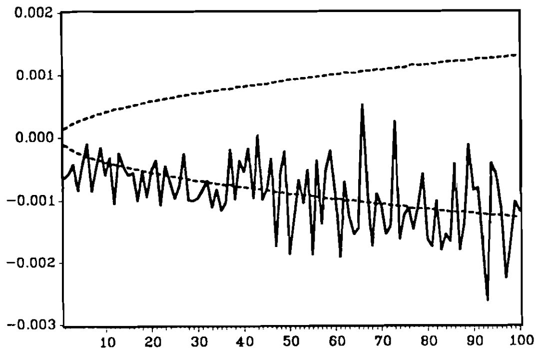
Results from experiments of varying the time to maturity are shown in figure 4 and 5. Call option prices were computed for at-the-money options with a spot rate and an exercise price of 1.80. The simulations were based on the same parameter estimates as in the previous experiment and the time to maturity was varied between 1 and 40 days. The maturity effects of the biases of static models are plotted in figure 4. The biases tend to be negative for the three models, i.e. Black-Scholes prices tend to be higher than simulated prices. However, for the Mixture model and Student's distribution, the negative biases gradually become statistically insignificant when maturity increases. For the compound Poisson process, the negative maturity biases become stronger when maturity increases but the standard errors of simulation increase with

**Figure 4**
**Maturity effect of biases in option prices for static models**

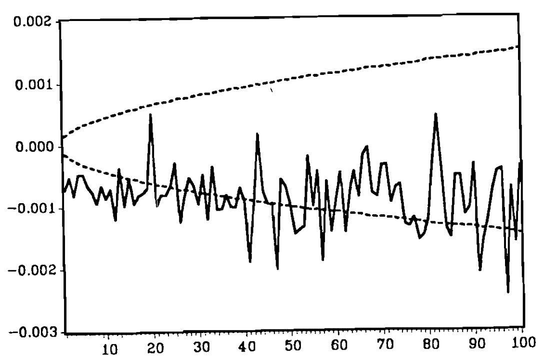
a) Mixture distribution




b) Compound Poisson process

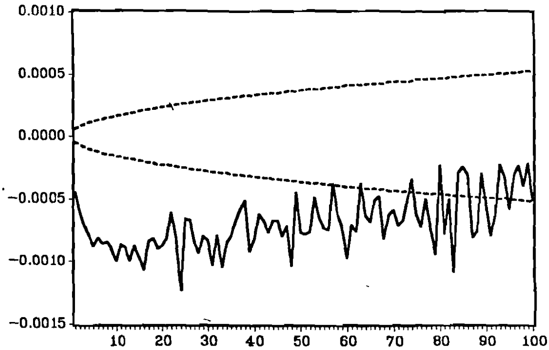



c) Generalized Student distribution

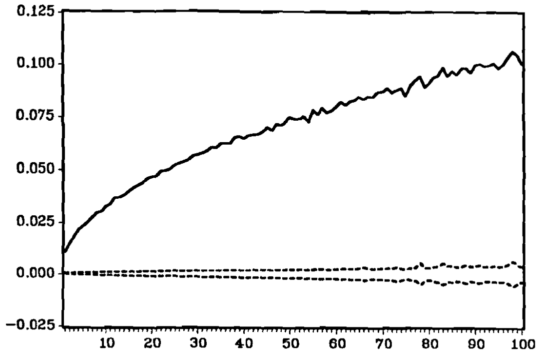


## Figure 5
## Maturity effect of biases in option prices for dynamic models

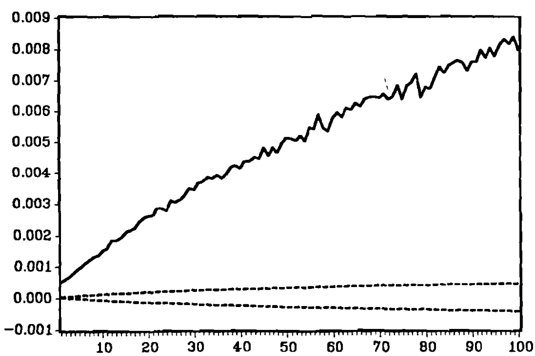
a) Markov-switching model




b) GARCH model




c) EGARCH model



the biases. The negative biases of the three models is caused by the peakedness effect and the convergence to Black-Scholes prices can be explained with the central limit theorem which implies that the three static models converge to a normal distribution under temporal aggregation.

The maturity effects for the dynamic models are shown in figure 5. As in figure 3, there is a positive bias for the GARCH model and the EGARCH model for at the money options. As explained above, this positive bias is due to the large stationary volatility implied by these models. The bias increases with maturity but the increase seems to level off for longer maturities. The Markov-switching model on the other hand, implies lower option prices, compared to Black-Scholes, for all maturities. However, the bias becomes statistically insignificant for long maturities and this is compatible with convergence to normality under temporal aggregation.

## 6. Conclusion

The three questions, raised in the Introduction, can now be answered in the following way. First, the most important stylized facts of the exchange-rate data are leptokurtosis and heteroskedasticity. It is important to note that leptokurtosis can be caused by fat tails or by peakedness. Furthermore, heteroskedasticity is linked to the property of leptokurtosis since the dynamic models, which have been introduced, imply leptokurtosis. It can also be shown that all scale-compounded normal distributions are leptokurtic.

Second, it is rather difficult to choose the best model among the seven candidates since the relative performance varies with the applied criterion. However, the dynamic models have a natural advantage over the static models because the latter do not capture heteroskedasticity. For practical purposes, the most important criterion of performance is presumably forecasting performance. On this account one would pick the Markov-switching model as the most satisfactory model.

Third, systematic option-price effects are, in general, rather small for static models when maturity is longer than 20 days. The convergence to Black-Scholes prices under time aggregation is in accordance with the central limit theorem. The GARCH and EGARCH models, on the other hand, imply a systematic and strong underpricing of call options by the Black-Scholes formula. This, however, is caused by the near non-stationarity of the models and the resulting large stationary variances. The Markov-switching model is the only model which is compatible with the observed smile effects on the foreign-currency option markets. It, therefore, appears that overall the Markov-switching model is the best candidate model to capture both the stylized facts of exchange-rate dynamics and the price biases of foreign-currency options.

## REFERENCES


Balanda, K.P. and MacGillivray, H.L. (1988), Kurtosis: A Critical Review, The American Statistician, Vol. 42, 111-119.

Bollerslev, T. (1987), A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, Review of Economics and Statistics, Vol. 69, 542-547.

Bollerslev, T., R.Y. Chou and K.F. Kroner (1992), ARCH Modeling in Finance, Journal of Econometrics, Vol. 52, 5-59.

Chiang, C.L. (1980), An Introduction to Stochastic Processes and Their Applications, Krieger.

Clewlow, L.J. and X. Xu (1992), A Review of Option Pricing with Stochastic Volatility. University of Warwick, FORC Preprint 92/35.

Duan, J.-C. (1991), The GARCH Option Pricing Model, McGill University, Faculty of Management, mimeo.

Engle, R.F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica, Vol. 50, 987-1007.

Fama, E.F. (1965), The Behavior of Stock Market Prices, Journal of Business, Vol. 38, 34-105.

Fama, E.F. (1976), Foundations of Finance, Basic Books.

Hull, J. and A. White (1987), The Pricing of Options on Assets with Stochastic Volatilities, Journal of Finance, Vol. 42, 281-300.

Jarrow, R. and A. Rudd (1982), Approximate Option Valuation for Arbitrary Stochastic Processes, Journal of Financial Economics, Vol. 10, 347-369.

Kaehler, J. (1989), Statistical Properties of Exchange Rates. Universität Mannheim, Institut für Volkswirtschaftslehre und Statistik, Discussion Paper No. 286-89.

Kaehler, J. (1993a), On the Modelling of Exchange-Rate Dynamics by Stable Paretian Distributions and Regularly Varying Tails, ZEW Discussion Paper No. 93-25.

Kaehler, J. (1993b), Stochastic Models of Exchange-Rate Dynamics and their Implications for the Pricing of Foreign-Currency Options, ZEW Mannheim, mimeo.

Kaehler, J. and V. Marnet (1993), Markov-Switching Models for Exchange-Rate Dynamics and the Pricing of Foreign Currency Options, forthcoming in J. Kaehler and P. Kugler (eds.), Econometric Analysis of Financial Markets. Heidelberg.

Mandelbrot, B. (1963), The Variation of Certain Speculative Prices, Journal of Business, Vol. 36, 394-419.

Meese, R.A., Rogoff, K. (1983), Empirical Exchange Rate Models of the Seventies: Do They Fit out of Sample? Journal of International Economics, Vol. 14, 3-24.

Nelson, D.B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica, Vol. 59, 347-370.

Shastri, K. and K. Wethyavivorn (1987), The Valuation of Currency Options for Alternate Stochastic Processes, Journal of Financial Research, Vol. 10, 283-392.

Taylor, S.J. (1992), Modelling Stochastic Volatility, Lancaster University, Department of Accounting and Finance, mimeo.