STOCHASTIC MODELS OF EXCHANGE-RATE DYNAMICS AND THEIR IMPLICATIONS FOR THE PRICING OF FOREIGN-CURRENCY OPTIONS

Juergen Kaehler

Doctor of Philosophy

London School of Economics and Political Science

July 1994

TABLE OF CONTENTS

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Abstract

The aim of this study is to find a suitable approach to model econometrically exchange-rate dynamics. In the first chapter, I examine the empirical properties of four exchange rates. The data used are daily, weekly, monthly and quarterly exchange rates of the German mark, the British pound, the Swiss franc, and the Japanese yen against the U.S. dollar from July 1974 to December 1987. I study the moment properties and time-series properties of these exchange rates and find in daily and weekly data leptokurtosis and heteroskedasticity. On the other hand, the hypotheses of no serial correlation, of a constant mean of zero, and of a symmetric distribution cannot be rejected. The fact that the daily and weekly data are not strictly equi-distant does not have a strong impact on these empirical regularities.

In chapter 2, static distributional models (mixture of distributions, compound Poisson process, Student distribution, and stable Paretian distributions) are estimated. Chi-squared goodness-of-fit tests reject these models. Direct inferential evidence against stable distributions is found by estimating the characteristic exponent by FFT and by estimating the exponent of regularly varying tails.

In chapter 3, dynamic models of heteroskedasticity (ARCH and Markov-switching models) are introduced. Quite satisfactory results are obtained for the EGARCH model and the Markov-switching model whereas the ARCH, GARCH and GARCH-t models are in conflict with stationarity conditions for the variance.

Chapter 4 compares the static and dynamic models with respect to goodness-of-fit and forecasting performance. With respect to goodness-of-fit criteria, the dynamic models appear to be superior to the static models. Furthermore, the dynamic models outperform a naive model of constant variance with respect to unbiasedness but not with respect to precision.

Chapter 5 studies the option-price implications of the static and dynamic models. The spot-rate effects of static models are rather small and they disappear, as expected, under temporal aggregation. GARCH and EGARCH models, on the other hand, imply higher option prices compared to Black-Scholes option prices along the whole spectrum of moneyness. Only the Markov-switching model is compatible with observed smile effects.

INTRODUCTION

The current state of the art in exchange-rate economics has a most puzzling feature. The theory of exchange-rate determination has been developed into a well-established branch of economic theory along the lines of the asset-market approach during the 1970's, whereas empirical testing has led to a complete rejection of these models. In 1979, Mussa wrote: "A model that was able to explain more than 50 percent of quarter-to-quarter changes in exchange rates should either be rejected on the grounds that it is too good to be true or should be reported to the Vatican as a miracle justifying the canonization of a new saint" (Mussa (1979), p. 50).

The situation has not changed very much since. In a very influencial paper by Meese and Rogoff (1983), the random-walk model

$$e_t = e_{t-1} + u_t \tag{1}$$

(where e_t is the logarithm of the exchange rate at time t and u_t is a white-noise variable)

emerged as the champion among competing empirical exchange-rate models. The random-walk model, however, is a champion in the sense of the one-eyed among the blind. In fact, the random-walk model is the confession of total ignorance. It does not relate exchange-rate fluctuations to fundamental determinants, nor has it any statistical structure which could be exploited for non-trivial exchange-rate forecasts.[1] The random-walk model simply predicts the current exchange rate to be the most probable exchange rate for the whole future, from tomorrow until infinity. The disillusioning result of Meese and Rogoff (1983) has been confirmed by a number of authors, see Backus (1984) and Leventakis (1987) and Pentecost (1991).

[1] Sometimes,the claim is made, that a random walk implies market effiency (or vice versa) in the sense of Fama (1970). However, a random walk is neither a necessary nor a sufficient condition for market efficiency, see Levich (1985).

To a large extent, this empirical failure of exchange-rate theories is explained by the fact that driving forces behind exchange rates are expectations about future values of relevant economic or political variables. It is obvious from reports about foreign-exchange markets in the popular press and it is incorporated into asset-market theories of exchange-rate determination (see e.g. Mussa (1984)) that the exchange rate is a forward-looking variable. However, the expectations of agents in foreign-exchange markets cannot be derived from published statistics. Hence the implementation of exchange-rate theories into econometric models is severely hampered if not impossible. In early 1988, the press release concerning a 9 billion dollar trade account deficit of the USA in February led to a sizeable appreciation of the dollar. At first sight, and in econometric models of exchange-rate determination, this appreciation seems to be a reaction of the wrong sign. It can, however, be explained by the fact that market participants expected the trade deficit for February 1988 to be much larger and were surprised by this positive news for the dollar.

In more formal terms, the principal difficulties of testing exchange-rate theories can be demonstrated with the aid of a simple reduced form of the asset-market approach to exchange rates (see Mussa (1984)):

(2) $$e_t = \lambda e_{t,t+1} + \beta x_t \quad \text{with} \quad 0 < \lambda < 1.$$

In (2) $e_{t,t+1}$ is the expectation of e_{t+1} formed at t, x_t is a column vector of exogenous variables, and β is a row vector of coefficients. Stepwise forward iteration and application of the properties of rational expectation yields:

(3) $$e_t = \lambda^\tau e_{t,t+\tau} + \sum_{i=0}^{\tau} \lambda^i \beta x_{t,t+i}$$

or

(4) $$e_t = \sum_{i=0}^{\infty} \lambda^i \beta x_{t,t+i}.$$

The change of the exchange rate is given by

(5) $$e_{t+1} - e_t = \beta(x_{t+1} - x_t) + \sum_{i=1}^{\infty} \lambda^{i-1} \beta(x_{t+1,t+i} - \lambda x_{t,t+i}) - \beta x_{t+1}.$$

This formula shows that it is not only the present change of x_t which moves the exchange rate but also revisions in the expectation of future values of x_t. These revisions of expectations have a geometrically declining weight, with variables far in the future receiving relatively small weights. Still, any news which brings about a revision of the whole future path of some driving forces can have strong effects on the present exchange rate, but these revisions are generally unobservable.

There is no obvious and easy way out of the measurement problem and hence there is no great hope for the econometric approach of testing exchange-rate theories. This dilemma directs empirical research on exchange rates to a more descriptive approach. If exchange-rate fluctuations cannot be explained, it is still worth exploring empirical properties of the data. This data-analysis approach may not seem to be directed to the research programme of "rerum causas cognoscere" but behind exploratory data analysis is the hope that the analysis will reveal unexpected properties of the phenomenon under investigation and will hence contribute to more knowledge about it. As the old saying goes: If you torture the data long enough, they will confess.

But this is not the only motivation for the study. 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. This result was confirmed by numerous studies, see e.g. Taylor (1986).

After the breakdown of the Bretton-Woods system of fixed exchange rates in the early 1970's it soon became apparent, to the surprise of many economists, that flexible exchange rates of the major currencies behaved just like other speculative prices. Therefore, the analytic tools developed for the study of price dynamics in speculative markets can readily be applied to the foreign-exchange market.

Two areas of exchange-rate economics, where the stochastic specification is of great importance, are the testing for market effiency and the pricing of foreign-exchange options. First, tests for the market efficiency in the context of uncovered interest parity require the specification of a risk premium if the unrealistic assumption of risk neutrality is abandoned; see the excellent review of this literature by Hodrick (1987). Second, option pricing along the lines of Black and Scholes (1973) requires the specification of the stochastic process for the price of underlying assets. Black and Scholes assume normality and constant volatility and it turns out that the normal assumption is quite important for their approach to construct a perfect hedge portfolio which eliminates all risk considerations from the pricing of options. It can be shown that a perfect hedge portfolio can only be constructed if the price of the underlying asset follows a Wiener process or a jump-diffusion process.

It is currently a very active research area to examine the implications for option pricing if the assumptions of normality, of constant volatility and of deterministic interest rates are relaxed. This topic will be taken up in Chapter 5.

This study is organized as follows. In Chapter 1, I will present a comprehensive statistical analysis of exchange-rate data. It is an exploratory data analysis which aims to identify the main statistical properties of the data in order to guide the stochastic modelling of exchange-rate dynamics. On several occasions, a single hypothesis will be tested by more than one method. This is a consequence of the general trade-off between efficiency and robustness in statistical testing. If the assumption of normality can be maintained, then there is usually an optimal parametric test. If however, there are serious doubts about the normality assumption, and this is the case for the exchange rate data, then it is often best to apply a robust non-parametric test.

I will apply two classes of stochastic modells to the exchange-rate data. In Chapter 2, static models will be applied which can capture the distributional properties of the data. In empirical finance, the static distribution models have often been analysed in isolation or without any unifying framework. I will provide such a framework by showing that the four models to be analysed can be interpreted as scale-compounded normal distributions. Some additional motivations for the distribution models will also be given. In Chapter 3, dynamic models will be applied which allow for dependence in the data in the form of heteroskedasticity. The static and dynamics models are compared in Chapter 4 with respect to fitting the data within the sample and with respect to forecast the data out of sample.

In Chapter 5, I will study the implications of the estimated exchange-rate models for the pricing of exchange-rate options. As mentioned above, the pricing of options along the lines of Black and Scholes is restricted to certain stochastic process and, therefore there is in general no direct solutions of the option pricing problems for the stochastic processes analysed in this study. It is possible, however, to determine for all stochastic processes the option prices in a simplified framework via Monte-Carlo simulation.

CHAPTER 1

STATISTICAL PROPERTIES OF EXCHANGE RATES

Due to the dominance of the random-walk model in comparison with econometric and time-series models of exchange-rate fluctuations, it is reasonable to take equation (1) of the Introduction as the starting point of the exploratory data analysis. The random-walk model incorporates only a few assumptions of statistical nature. It assumes that the u_t are independent and identically distributed, i.e. u_t is white noise. The following analysis takes a closer look at both assumptions. First, independence and time-series properties are examined in detail, both in the time domain and in the frequency domain. Second, the moments and the distributional properties of the data are studied.

1.1 DATA

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. A rise in the exchange rate signifies an appreciation of the dollar. 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, except for the sfr, whose exchange rate against the dollar from July 1974 to April 1980 was not published in the International Financial

Statistics and was therefore taken from the monthly reports of the Swiss National Bank. These exchange rates can be regarded as the four most important exchange rates on the international foreign exchange markets.

The data are mainly analysed in the form of first differences in the logarithm of exchange rates, i.e.

(6) $$u_t = \Delta e_t = e_t - e_{t-1} = \log \mathrm{E}_t - \log \mathrm{E}_{t-1}.$$

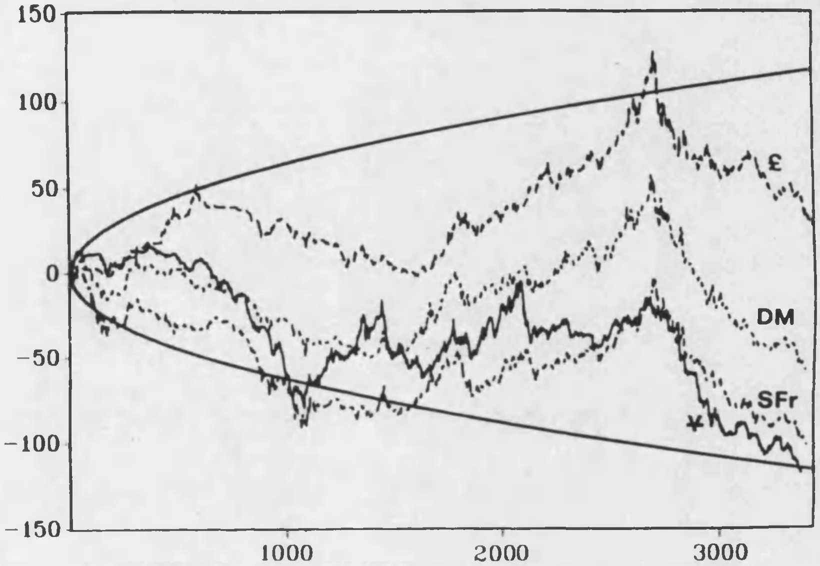
It is customary to analyse exchange-rate data in this form because it avoids Jensen's inequality which states that the expected value of the reciprocal of a random variable is greater than the reciprocal of the expected value. For small exchange-rate fluctuations, Δe_t is approximately equal to the percentage changes in E_t, since $x \approx \log(1+x)$ for small x. The variable Δe_t measures exchange-rate dynamics in the form of continuous growth rates of E_t. The data Δe_t were multiplied by the factor of 100 to express them in units of percentage change. In some models it turned out that multiplying the data by 100 increased numerical stability.

Figure 1 displays the evolution of the four daily exchange rates. Plotted are the values of $(\mathrm{E}_t - \mathrm{E}_0)/S_u$, where S_u is the standard deviation of $\mathrm{E}_t - \mathrm{E}_{t-1} = U_t$. This scaling produces unit variance in the innovations U_t of all four exchange rates. In addition, the lines $\pm 2\sqrt{t}$ are drawn into the figure to give 0.95 confidence limits for a random walk whose innovations have unit variance and an expected value of zero. Only the sfr-dollar rate falls out of these bounds somewhat excessively whereas the mark-dollar rate stays within the limits for the whole sample period.

As can be seen from figure 1, there is quite strong co-movement between the exchange rates, especially between the sfr and the mark. The visual display of the data cannot, of course, answer the question whether a random walk is an adequate model for exchange-rate movements. The task to answer this question and to give insight into various statistical properties of the data falls on the following tests.

Figure 1

Four daily exchange rates (standardized): July 1974 - Dec. 1987



1.2 INDEPENDENCE AND TIME-SERIES PROPERTIES

Under the assumption of identical distributions, $u_t = u_{t+j}$ implies $F(u_t) = F(u_{t+j})$ for all values of t and j, where $F(.)$ denotes the distribution function. The assumption of independence of all u_t can be formalized as $F(u_t \mid u_{t+j}) = F(u_t)$, where $F(u_t \mid .)$ denotes the conditional distribution function.

If one assumes that u_t has a normal distribution, the assumptions of independence and identical distributions simplify considerably since the normal distribtuion has two parameters, θ and σ^2 which coincide with the mean and the variance, respectively. Moreover, for normal variables, correlation of zero implies independence. Thus, under the assumption of Gaussian white noise it suffices to examine moments up to order 2. In the study of time-series properties, I will follow the custom of testing for indepence only up to moments of order 2 but in the study of distributional properties, I will examine moments up to order 4 (a justification will be given later).

If the random-walk model is written in matrix notation one gets

$$\Delta e = u, \tag{7}$$

where Δe is a $(T \times 1)$ vector with typical element $\Delta e_t = e_t - e_{t-1}$ and u is a $(T \times 1)$ vector with typical element u_t. The model of a random walk without drift assumes that

$$E(u) = 0, \tag{i}$$

and that

$$E(uu') = \sigma^2 \Omega = \sigma^2 I, \tag{ii}$$

where E is the expected-value operator, 0 is the $(T \times 1)$ null vector and I is the $(T \times T)$ identity matrix. Assumption (i) states that the expected values of all u_t's are zero, and assumption (ii) states that the variance-covariance matrix $\sigma^2\Omega$ of the u_t's is a scalar matrix, i.e. all off-diagonal elements are zero and the diagonal elements of Ω are constant.

A robust test for independence which imposes only very mild restrictions was introduced and recommended especially for studies of speculative prices by Dufour (1981). The test is directed towards an examination whether the off-diagonal elements of Ω are zero. Dufour's signed-rank test for serial independence is based on the assumption of a symmetrical continuous distribution with a median of zero. The test is distribution-free and in particular it does not assume identical distributions. The test statistic is

$$S_k = \sum_{t=1}^{T-k} g(v_{kt}) R_{kt}^{+}, \tag{8}$$

where $g(v_{kt})$ is an indicator function defined by

$$g(v_{kt}) = \begin{Bmatrix} 1 & \text{if} \quad v_{kt} \geq 0 \\ 0 & \text{otherwise} \end{Bmatrix} \tag{9}$$

and R_{kt}^{+} is the rank of a non-negative $v_{kt} = u_t u_{t+k}$ among the ranks of $|v_{kt}|$. Thus, S_k is the sum of the ranks of non-negative v_{kt}'s. This test statistic has the same form as the test statistic of the Wilcoxon signed-rank test of symmetry around zero and the distribution of S_k under the null hypothesis of independence is the same as the null distribution of the Wilcoxon statistic. In particular, the expected value of S_k is $T_k(T_k + 1)/4$ and the variance is given by $T_k(T_k + 1)(2T_k + 1)/24$. The distribution of S_k tends to the normal distribution as T_k, the number of products $u_t u_{t+k}$, goes to infinity (see e.g. Lehmann (1975), pp. 124-132). The result from applying Dufour's signed-rank test are summarized in table 1.

Table 1

Signed-rank test for serial independence

	mark	pound	sfr	yen	signif. coeff.
day	-/1/2	1/2/-	-/1/1	3/2/-	4/10/13
week	1/2/3	-/2/-	1/-/2	2/1/2	4/9/16
month	-/2/2	-/2/2	-/2/3	-/-/1	-/6/14
quarter	-/1/2	-/3/1	-/-/2	-/1/-	-/5/10

The test statistic was calculated for $k = 1, \ldots, 15$. In order to save space, for each of the 16 series only the numbers of significant statistics S_k are shown. The numbers reported refer to the number of statistics which are significant at the 1 percent, 5 percent and 10 percent level, respectively. The numbers in the last column summarize the results for all four exchange rates at the given period of time. Here the numbers refer to the number of statistics which are significant at least at the 1 percent level, at least at the 5 percent level and at least at the 10 percent level, respectively. These numbers can be compared with the expected numbers 0.6/3/6 under the null hypothesis. Overall, there are more significant S_k than expected, especially in the daily and weekly series. Noteworthy are two very large values of S_k for the weekly yen series at lags 1 and 2. Their standardized values are 4.96 and 4.38, respectively. All in all, there seems to be some moderate evidence against serial independence from the signed-rank test. It remains to be seen, however, whether this evidence can be attributed to dependence or to violations of the assumptions underlying the test, namely asymmetry or a non-zero median.

The nonparametric test of independence has the advantage of not requiring any specific distributional assumptions about u_t. As will be seen later, the assumption of normality is a quite critical one for exchange-rate data. It is still interesting, however, to supplement Dufour's test by parametric tests of independence. These tests can be put into a broad framework of time-series analysis.

A very general model for stationary time series is the autoregressive integrated moving-average (ARIMA) model which has been popularized by Box and Jenkins (1976). As a model for exchange-rate data, it can be written as

$$\Phi(L)\,(1-L)^d e_t = \Theta(L) u_t. \tag{10}$$

The autoregressive (AR) component $\Phi(L)$ is a polynominal of degree p in the lag operator L, i.e. $\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ and the moving-average (MA) component is a polynominal of degree q in L, i.e. $\Phi(L) = 1 - \theta_1 L - \ldots - \theta_q L^q$. The parameter d is the order of differencing which is necessary to make the series e_t stationary, d is also called the order of integration. A short-hand notation for this model is ARIMA (p, d, q). It is readily seen that the random-walk model fits into this framework and can be written as a ARIMA (0, 1, 0) model.

The test of $p = q = 0$ is a test of independence of the u_t's. Tests for the null hypothesis $(H_0) d = 1$ are called "tests for unit roots". Standard testing methods are not applicable for this H_0 since they require a stationary series. For testing $d = 1$, I use a Lagrange multiplier (LM) test, i.e. a test which is based on the likelihood of the model estimated under H_0. This test, which has been suggested by Solo (1984), is applicable to the ARIMA-framework. The test statistic is given by:

$$LM = \left(\sum_t \hat{u}_t \hat{\xi}_{t-1}\right)^2 \Big/ \left(\hat{\sigma}^2 \sum_t \hat{\xi}^2_{t-1}\right) \tag{11}$$

where the $\hat{u}_t$'s are the residuals from a fitted ARIMA (p, 1, q) model, $\hat{\xi}_{t-1} = [\hat{\Phi}(L)]^{-1} e_{t-1}$ and $\hat{\sigma}^2$ is the variance of the residuals. The asymptotic distribution of the test statistic is identical to that of τ^2 resp. τ_μ^2 given in Fuller (1976, p. 373). I chose a variant of this test which takes account of a non-zero mean and which leads to the τ_μ statistic.[2]

Table 2

ARIMA models and test for unit roots

	mark		pound		sfr		yen	
day	MA(3) 0.002	0.12 (6.60)	MA(2) 0.002	2.60 (6.60)	MA(9) 0.002	0.72 (6.60)	AR(4,9,10) 0.006	0.79 (6.60)
week	MA(19) 0.014	0.12 (6.60)	MA(7) 0.010	1.00 (6.60)	MA(4) 0.005	0.70 (6.60)	ARMA(1,19) 0.060	0.93 (6.60)
month	MA(2) 0.036	0.97 (6.62)	MA(18) 0.027	2.85 (6.62)	MA(16) 0.030	1.76 (6.62)	MA(3) 0.018	0.32 (6.62)
quarte r	ARMA(5,19) 0.569	4.80 (6.76)	MA(16) 0.487	1.38 (6.76)	MA(15) 0.037	1.59 (6.76)	MA(16) 0.427	3.37 (6.76)

Table 2 reports the main results from estimating and testing the ARIMA model. In each cell of the table, the order of ARIMA (p, 1, q) is given in the upper left hand side. The term MA(3), for example, denotes a moving-average polynomial of the form

[2] Hakkio (1986) compared four similar tests. Two of them are tests of a unit root as used by Meese and Singleton (1982) and Diebold (1988). The other two tests are tests for white noise: a standard F-Test and the Box-Pierce statistic. Hakkio tried to calculate the power of these tests against an empirical ARIMA (1,1,2) model by the Monte Carlo method. The results, however, are very confusing. No test attained approximately its nominal level under the null hypothesis. Since he used 1000 replications, the standard error of the estimated significance level α is $[\alpha(1-\alpha)/1000]^{1/2}$ which gives 0.0069 for $\alpha = 0.05$. The estimated levels, however, lay between 0.026 and 0.060. Very disturbing is the fact that for the Box-Pierce statistic Hakkio got an estimated level of 0.060 indicating an optimistic test. It is well-known that the Box-Pierce test is conservative, i.e. the estimated level is well below the nominal level in small samples (see e.g. Davies, Triggs and Newbold (1977)). In practical work, therefore, the Box-Pierce test has been replaced by the Ljung-Box test (see Ljung and Box (1978)) which is much closer to its nominal level than the former test. Hence, it is quite doubtful whether any safe conclusions can be drawn from Hakkio's study.

$\Theta(L) = 1 - \theta_3 L^3$, whereas the term AR(4, 9, 10) denotes an autoregressive polynominal of the form $\Phi(L) = 1 - \phi_4 L^4 - \phi_9 L^9 - \phi_{10} L^{10}$. The order of the ARIMA model was selected on the basis of the Schwarz information criterion (see e.g. Priestley (1981), pp. 372-376), which is given by $SIC = (p+q)\log T - 2\mathrm{L} - 2T$ and where $(p+q)$ is the number of AR- and MA-parameters estimated and L is the logarithm of the maximised likelihood. In most cases, the best model, i.e. the model with the lowest value of SIC is given by a simple MA-model. For daily, weekly and monthly data, the explanatory power of these univariate time-series models is very poor, as indicated by the R^2 coefficient of determination given below the order of the ARIMA model.[3] The quarterly exchange-rate dynamics of the mark, pound and yen against the dollar, however, are quite reasonably fitted by ARIMA models. In fact, the ARIMA(5,1,19) model for the quarterly mark-dollar exchange rate qualifies for canonisation, according to Mussa's criterion, and the quarterly models for the pound and yen are on the margin for qualifying. It would be interesting to examine the structural stability of these three quarterly ARIMA models. Since these data sets only have 53 observations each, it is not possible to do these tests in a meaningful way.

In table 2, the upper right hand number in each cell is the LM statistic of the unit-root test and below it the critical value of the statistic is given in brackets. Since H_0 states that $d = 1$, it is appropriate to choose a low significance level. The reported critical values

[3] Independence of exchange-rate dynamics has been examined by various authors employing different methods and deriving at contradictory conclusions. Rogalski and Vinso (1978) calculated the Box-Pierce statistic Q at lag 12 for 5 weekly exchange-rates in the post-Bretton-Woods era. They could not reject independence. It is quite typical for the empirical work on exchange rates in the 1970's that Rogalski and Vinso interpreted their findings in terms of market efficiency. It is now widely, but not universally, recognized that independence is neither a necessary nor a sufficient condition for market efficiency (see Levich (1985) pp. 1020-1025). Kim (1987) applied a F-test to daily, weekly and monthly dollar exchange rates against 7 currencies for the period from January 1973 to June 1985. He rejected the null hypothesis of independence for all daily (with exception of the yen) and weekly series but only for 2 monthly series (pound and yen). Baillie and McMahon (1987) performed a likelihood-ratio test with respect to AR (6) models for 6 monthly exchange rates and with respect to AR (2) models for 4 weekly exchange rates. The order of the AR processes was not derived from an optimal selection strategy. Baillie and McMahon rejected independence for the weekly series but could not reject it for any of the monthly series. Finally, Hsieh (1988) tested for serial independence in 5 daily exchange-rate series and could not reject the null hypothesis when the standard errors of the autocorrelation coefficient were adjusted for heteroskedasticity.

correspond to a significance level of 0.10. As can be seen, none of the LM statistics exceed their critical values. However, no strong conclusions can be drawn from not rejecting a null hypothesis.

Meese and Singleton (1982) and Diebold (1988) performed similar tests. They examined weekly and monthly exchange-rate movements and tested for unit roots in AR representations of the data.[4] Their results confirm my findings that the H_0 of a unit root cannot be rejected even at a low level of significance.[5] In addition, in both studies the hypothesis of two unit roots was tested, i.e. the model

$$\Phi(L)(1-L)^2 e_t = 0 \tag{12}$$

was considered. It can be argued that it is superfluous to test this model because even a visual inspection of the series Δe_t and their correlograms makes it abundantly clear that these series are non-integrated. Figure 2 shows one such series, the monthly movements (of the logarithm) of the yen-dollar rate. It is not very surprising therefore, that both Meese and Singleton and Diebold could formally reject $H_0{:}d = 2$ at very high significance levels.

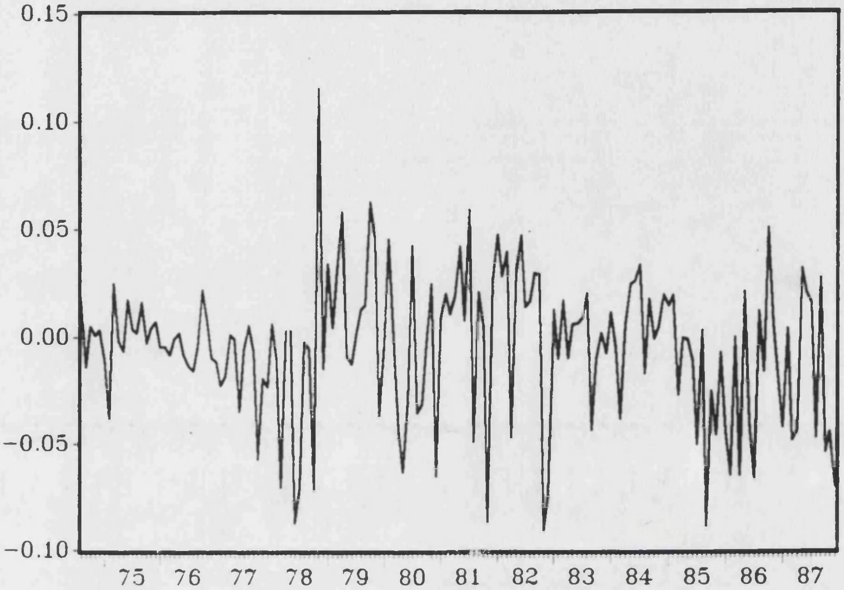
Sometimes, spectral analysis of time series helps to detect hidden periodicities in the data which violate the assumption of independence. It is therefore useful to complement the foregoing analysis in the time domain with some investigations in the frequency domain. The basic question is whether $\Delta e_t = u_t$ can be regarded as white noise. To test this hypothesis in the frequency domain, I chose a cumulative periodogram test which is based on the Kolmogorov-Smirnov statistic (see Priestley (1981), pp. 479-483).

[4] Diebold justified his testing of unit roots within AR repesentations with the fact that he found "no evidence of a moving average component in any of the seven series" he examined (Diebold (1988), p. 44). In addition to the four exchange rates included in my study, he also analysed the weekly dynamics of the US dollar against the Canadian dollar, the French franc and the Italian lira. His model specification procedure was also the Schwarz information criterion (SIC). So there is an obvious conflict with my results reported in table 3. The fact that Diebold used a different data source (i.e. International Monetary Markets Yearbook) cannot explain the discrepency in our results since data quality is no issue with exchange-rate data. One of the series used by Diebold has been listed in Engle and Bollerslev (1986). I analysed this dollar-sfr exchange rate and found that, according to SIC, the best model is a MA (14) repesentation. Using the same series and applying the extended sample autocorrelation function for identification, Tsay (1987) suggested a MA (4) model.

[5] See also Baillie and McMahon (1987) and Corbae and Ouliaris (1986).

Figure 2

Monthly exchange-rate movements: yen-dollar



The test statistic is

(13) $$C = \max \left| F_r - \frac{r}{M} \right|,$$

where F_r is the cumulated periodogram defined by

(14) $$F_r = \sum_{k=1}^{r} I(\omega_k) / \sum_{k=1}^{M} I(\omega_k),$$

with

(15) $$I(\omega_k) = c_0 + 2 \sum_{\tau=1}^{T-1} c_\tau \cos 2\pi\omega_k\tau.$$

The periodogram $I(\omega_k)$ is defined in (15) as the Fourier transform of the empirical covariance function c_τ. The periodogram is calculated at the Fourier frequencies $\omega_k = k/T$ for $k = 1, 2, \ldots, M$ where M is the integer part of $T/2$. The test statistic C is defined in (13) as the maximal absolute distance between the empirical cumulative periodogram and the theoretical cumulative periodogram. For a white-noise process, the periodogram has a uniform distribution and hence the cumulative periodogram is a straight line with slope $2/T$ if T is even (and approximately so if T is odd). The critical values for the cumulative periodogram test are given by $b_\alpha/(M-1)^{1/2}$ where b_α is the α-quantile of the Kolmogorov distribution tabulated in many statistics books. Table 3 summarizes the result from this test.

Table 3

Cumulative periodogram test

		mark	pound	sfr	yen
day	C $1/\omega^*$	0.020 (∞)	0.040 *** (2.2)	0.017 (23)	0.037 *** (76.5)
week	C $1/\omega^*$	0.067 * (9.8)	0.059 (3.7)	0.063 (3.7)	0.181 *** (16)
month	C $1/\omega^*$	0.114 (2.3)	0.097 (81.0)	0.091 (2.3)	0.126 (3.8)
quarter	C $1/\omega^*$	0.156 (27.0)	0.229 (27.0)	0.131 (27.0)	0.281 ** (10.8)

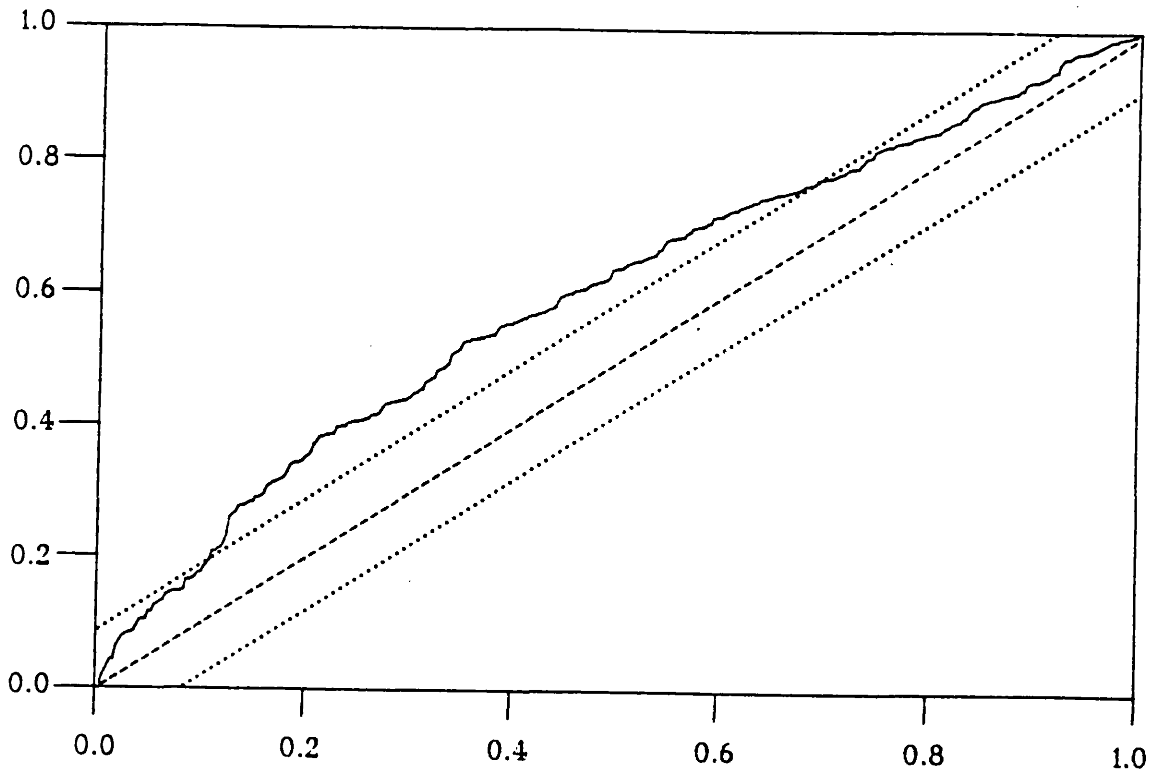
Significance levels: $\alpha = 0.01$ (***); $\alpha = 0.05$ (**); $\alpha = 0.10$ (*)

The cumulative periodogram test is basically a goodness-of-fit test to the uniform distribution under the null hypothesis of white noise. Since I am interested in rejections of H_0 it is appropriate to select a relatively low significance level of, say, 10 percent. According to this level, 5 out of 16 series show significant deviations from white noise. It is the yen-dollar rate which shows the most marked and systematic deviations. The numbers in parentheses are the periods $1/\omega^*$ where the estimated spectral density functions have their maximal value. These periods need not be the same as the periods where the Kolmogorov-Smirnov statistics attain their maximum. For the yen-dollar rate, there seems to be a strong and consistent 16 week cycle. There is also a surprising coincidence of maximal spectral densities in the quarterly exchange rates of the mark, pound and Swiss franc against the dollar at the frequency 0.037 which corresponds to a period of 6 years and 3 quarters.

The daily changes of the mark-dollar rate show a maximal density at frequency 0.0 which indicates a trend in this series. To summarize, it seems safe to conclude that there is no overwhelming evidence for hidden periodicities in the data with the exception of the yen-dollar rate.[6] There is, however, no obvious economic explanation for the 16 week cycle in this exchange rate. In figure 3, both F_r and r/M are plotted for the weekly changes of the yen-dollar rate. Also drawn into the figure are the dotted 0.99 confidence limits as parallels to the dashed r/M line.

Figure 3

Cumulative periodogram: weekly yen-dollar rate



[6] Logue and Sweeney (1977) calculated the spectrum of the daily French franc-dollar exchange rate for the period from January 1970 to March 1974 and concluded, without formal testing, that there were no marked deviations from white noise. Diebold (1988) applied Fisher's periodogram test and found no significant deviations from white noise for 6 weekly and monthly dollar exchange rates. However, Schlittgen, Hammann and Lepinat (1982) found strong evidence against white noise in 5 out of 7 daily mark exchange rates. They used the Watson test to compare the periodogram with the uniform distribution. In contrast to the Kolmogorov-Smirnov test, which is based on the maximum of deviations

It has been suggested by some economists that one should distinguish between periods of turbulence and periods of tranquility in the history of exchange-rate fluctuations (see e.g. Frenkel and Levich (1977)). It has to be noted that periods of tranquility and turbulence cannot be identified as the Bretton-Woods era and the post-Bretton-Woods era, respectively since both the Bretton-Woods era had turbulent periods and the system of generalized floating had tranquil periods. Figure 2, which plots the monthly fluctuations in the yen-dollar rate, shows that there were in fact periods of reduced volatility in the exchange rate in 1975, 1976, 1983, and 1984. On the other hand, in periods of turbulence, the sign of next periods exchange-rate changes seems not to be predictable from this periods change, i.e. a strong appreciation will be followed by a strong appreciation or a strong depreciation with roughly equal probability.[7]

The validity of this observation can be tested in a direct and simple manner within the framework of a discrete Markov-chain model.[8] Let the observations of the Δe_t's $(t = 1, \ldots, T)$ be classified into l quantiles with $\varepsilon_t = i (i = 1, \ldots, l)$. The empirical transition matrix N can be defined as the matrix whose typical element n_{ij} gives the number of cases in which pairs with $\varepsilon_t = i$ and $\varepsilon_{t+1} = j$ occur. The theoretical transition matrix is derived in the following way. Let $\tilde{n}_{ij}$ be the expected number of cases in row i and column j under a specified null hypothesis and let $\tilde{n}_{i.}$ and $\tilde{n}_{.j}$ be the corresponding row sums and column sums, respectively. Furthermore, let $\tilde{n}_{i.} = (T-1)/l$ for all i. If one assumes that

(16) $$P_{ij} = P(\varepsilon_{t+1} = j \mid \varepsilon_t = i) = 1/l \quad \text{for all } i \text{ and } j,$$

then the $\tilde{n}_{ij}$'s are obtained from

(17) $$\tilde{n}_{ij} = (T-1)/l^2.$$

between the theoretical and the empirical distribtuion, the Watson test is based on the sum of all squared deviations.

[7] Cornell and Dietrich (1978) found this property in their analysis of 6 daily spot exchange rates but they did not apply a formal statistical test. See also Taya (1980).

[8] A description of Markov-chain models can be found in many books on stochastic processes. See e.g. Grimmett and Stirzaker (1982).

Equation (16) states that P_{ij}, the conditional probability of ε_{t+1} being in quantile (or state) j if ε_t was in i, is the reciprocal of the number of equally likely states l, i.e. there is a uniform probability distribution along the rows of the transition matrix. In statistical terminology, (16) is the null hypothesis of independence of the ε_t's. The well-known χ^2 goodness-of-fit test and a likelihood-ratio (LR) test are applied to test this H_0 (see Chatfield (1973)). They are defined by

$$\chi^2 = \sum_{i=1}^{l} \sum_{j=1}^{l} (n_{ij} - \tilde{n}_{ij})^2 / \tilde{n}_{ij} \quad \text{and} \tag{18}$$

$$LR = 2(T-1)\log(T-1) - 2\sum_{i=1}^{l} n_{i.} \log n_{i.} - 2\sum_{j=1}^{l} n_{.j} \log n_{.j} + 2\sum_{i=1}^{l}\sum_{j=1}^{l} n_{ij} \log n_{ij}. \tag{19}$$

Both tests have $\nu = (l-1)^2$ degrees of freedom. The test results for all exchange-rate series are reported in table 4. For daily, weekly, and monthly series, l is equal to 5 and for quarterly series l is equal to 3 in order to conform with Cochrane's conservative rule of thumb that all $\tilde{n}_{ij}$ should be greater than 1 and at least 80 percent of the $\tilde{n}_{ij}$ should be greater than 5 (see Moore (1986, pp. 70-71)). The fact that both tests are asymptotically equivalent is born out by the great similarity of the test results. The upper number in each cell of table 4 gives the value of the $\chi^2(\nu)$ statistic and the lower number is the value of the LR statistic. There is strong rejection of independence for the daily and weekly data but only very weak evidence against independence for longer-term exchange-rate fluctuations. The choice of l does not seem to have an influence upon this result. Performing the same test with $l = 9$ for daily data and with $l = 7$ for weekly data resulted in the same rejections of H_0 at very high significance levels. In fact, increasing l brought an increase in all test statistics. The numerical values of these tests are shown in table 5. Again, the upper numbers in each cell are the χ^2 statistics and the lower numbers are the LR statistics. For all series, H_0 can be rejected at the 0.01 significance level.

Table 4

Test of independence in Markov chains (l = 5 or 3)

		mark	pound	sfr	yen
day	$\chi^2(16)$ LR	128.8 *** 127.1 ***	316.0 *** 303.2 ***	100.6 *** 100.6 ***	383.0 *** 365.8 ***
week	$\chi^2(16)$ LR	42.4 *** 42.6 ***	66.9 *** 61.9 ***	41.4 *** 41.6 ***	90.5 *** 87.3 ***
month	$\chi^2(16)$ LR	13.5 13.9	18.9 20.5	15.4 16.0	21.2 23.4
quarter	$\chi^2(4)$ LR	8.8 * 8.8 *	5.7 5.3	4.3 4.4	12.2 ** 12.1 **

Significance levels: see table 3

Table 5

Test of independence in Markov chains (l = 9 or 7)

		mark	pound	sfr	yen
day	$\chi^2(64)$ LR	188.0 *** 184.8 ***	454.1 *** 430.8 ***	180.6 *** 179.0 ***	510.6 *** 486.2 ***
week	$\chi^2(36)$ LR	75.6 *** 79.4 ***	83.8 *** 77.9 ***	65.8 *** 65.9 ***	110.8 *** 101.7 ***

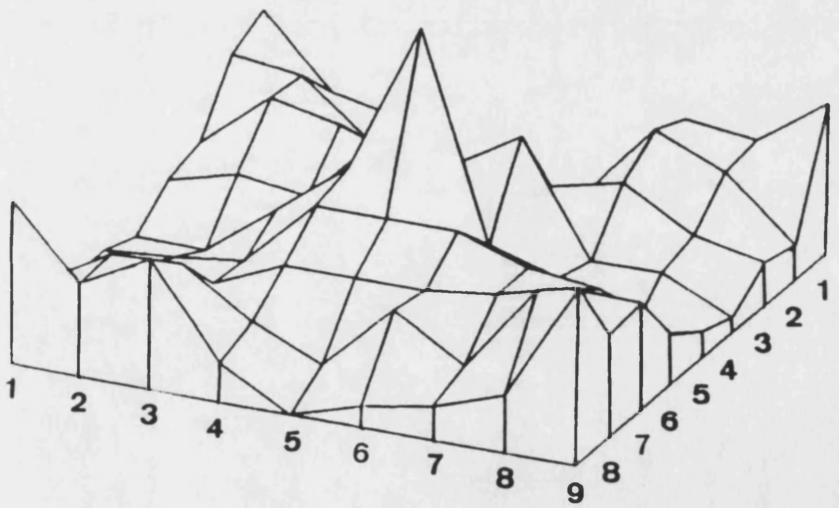
Significance levels: see table 3

Before one can start to speculate about the implication of these findings one would have to examine the deviations from independence in more detail. The results from tables 4 and 5 just reject independence without qualifying the kind of dependence. To gain some more insight into the dependencies, a typical empirical transition matrix is displayed in figure 4. It shows the data of daily changes in the pound-dollar rate classified into 9 quantiles. The height of the three-dimensional body is proportional to n_{ij}. There are 5 main peaks; a dominant one with $n_{55} = 109$ and 4 peaks in the corners with $n_{11} = 65$, $n_{19} = 66$, $n_{91} = 70$ and $n_{99} = 74$. There is also a side peak with $n_{46} = 69$. For all entries, the expected number is $\tilde{n}_{ij} = 42.2$. The main peaks can be interpreted in terms of periods of tranquility and periods of turbulence. The dominant peak n_{55} gives the number of cases where a small $|\Delta e_t|$ is followed by a small $|\Delta e_{t+1}|$. Likewise, n_{11} is the number of cases where a strong depreciation of the dollar was followed by another strong depreciation and n_{99} is the number of pairs of strong appreciation. On the other hand n_{19} cases could be counted where a strong appreciation followed a strong depreciation et vice versa for n_{91}.

This confirms the observation that in turbulent periods there can be a strong reaction in the foreign-exchange market of either sign, i.e. a strong exchange-rate movement in period t lowers the probability of small or moderate movements in t+1 and increases the probabilities both for a strong depreciation and a strong appreciation.[9] However, this phenomenon seems to vanish at longer time horizons, i.e. short term erratic exchange-rate movements are ironed out to some degree after a couple of weeks or months.

Figure 4

Markov transition matrix: daily pound-dollar rate



[9]Taya (1980) derived similar results for the daily mark-dollar rate.

In more formal terms, the findings from the Markov-chain analysis imply that the assumption of independence along the diagonal of Ω is violated. Classifying the exchange-rate data into quantiles, however, wastes a lot of information. The phenomenon that small fluctuations tend to be followed by small fluctuations and large fluctuations by large ones of either sign can be measured without this loss of information by estimating the autocorrelation function (ACF) of squared innovations $(\Delta e_t)^2 = u_t^2$. It is a very useful diagnostic tool to detect deviations from the random-walk model. Within the framework of the ARCH model, introduced by Engle (1982), and the bilinear model (see Granger, Anderson (1978)), the autocorrelation of squared data is used in the identification stage of model building. As the name indicates (ARCH stands for autoregressive conditional heteroskedasticity), the ARCH model assumes a time-varying variance. The ARCH(p) model is formally given by:

(20) $$\sigma_t^2 = \phi_0 + \sum_{i=1}^{p} \phi_i v_{t-i}^2,$$

where the distribution of v_t conditional on $(v_{t-1}, \dots, v_{t-p})$ is normal with zero mean and variance σ_t^2. Hence, σ_t^2 is the variance conditional on all information available at t. In the bilinear model, on the other hand, deviations from a white-noise process are caused by non-linearities which are introduced by product terms of a white-noise input series v_t and the output series u_t The general form of bilinear models is:

(21) $$u_t = \sum_{i=1}^{Q} \sum_{j=1}^{P} \beta_{ij} v_{t-i} u_{t-j} + v_t.$$

Both models can be embedded in the framework of ARIMA models. In the ARCH model, v_t can represent the residuals from an ARIMA model and the bilinear model can be extended to a bilinear ARIMA model (BARIMA) by adding the right-hand side of (21) to the right-hand side of (10). This extention to ARIMA models is quite instructive because it shows that ARCH effects would impinge on the conditional variance of the input series v_t of a ARIMA model, whereas bilinear terms would affect the mean of the output series u_t (see Weiss, 1986).

The AFC of a squared series, be it estimated input or output, is useful because for a white-noise series this function is by assumption zero at all lags $k \geq 1$ whereas both ARCH and bilinear models generally have theoretical ACF's which are positive at various lags. For the ARCH model, this is evident from (20), for the bilinear model, this can be demonstrated with a simple example. Assume the model

$$u_t = 0.5 v_{t-1} u_{t-2} + v_t. \tag{22}$$

It can be shown that the ACF of e_t^2 for $k = 1, \ldots, 4$ is given by the values 0.20; 0.25; 0.05 and 0.06 respectively.

The AFC for squared data $(\Delta e_t)^2 = u_t^2$ can be applied without modifications because McLeod and Li (1983) have established that under the H_0 of white noise, the standard errors of squared-data autocorrelations are the same as for the usual ACF. Hence, also the Ljung-Box statistic

$$Q(K) = T(T-2) \sum_{k=1}^{K} \hat{r}^2(k)/(T-k) \tag{23}$$

is applicable without modification. In (23), $\hat{r}(k)$ is the estimated autocorrelation coefficient at lag k, i.e.

$$\hat{r}(k) = \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) / \sum_{t=1}^{T} (x_t - \bar{x})^2. \tag{24}$$

The Ljung-Box test is a portmanteau test against white noise. It follows asymptotically a χ^2 distribution with degrees of freedom $(K-m)$, 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 as the upper number of each cell in table 6. 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. The number of significant autocorrelations and (after the slash) partial autocorrelations is given below the corresponding Q statistics.

Table 6

Results from the ACF of squared data

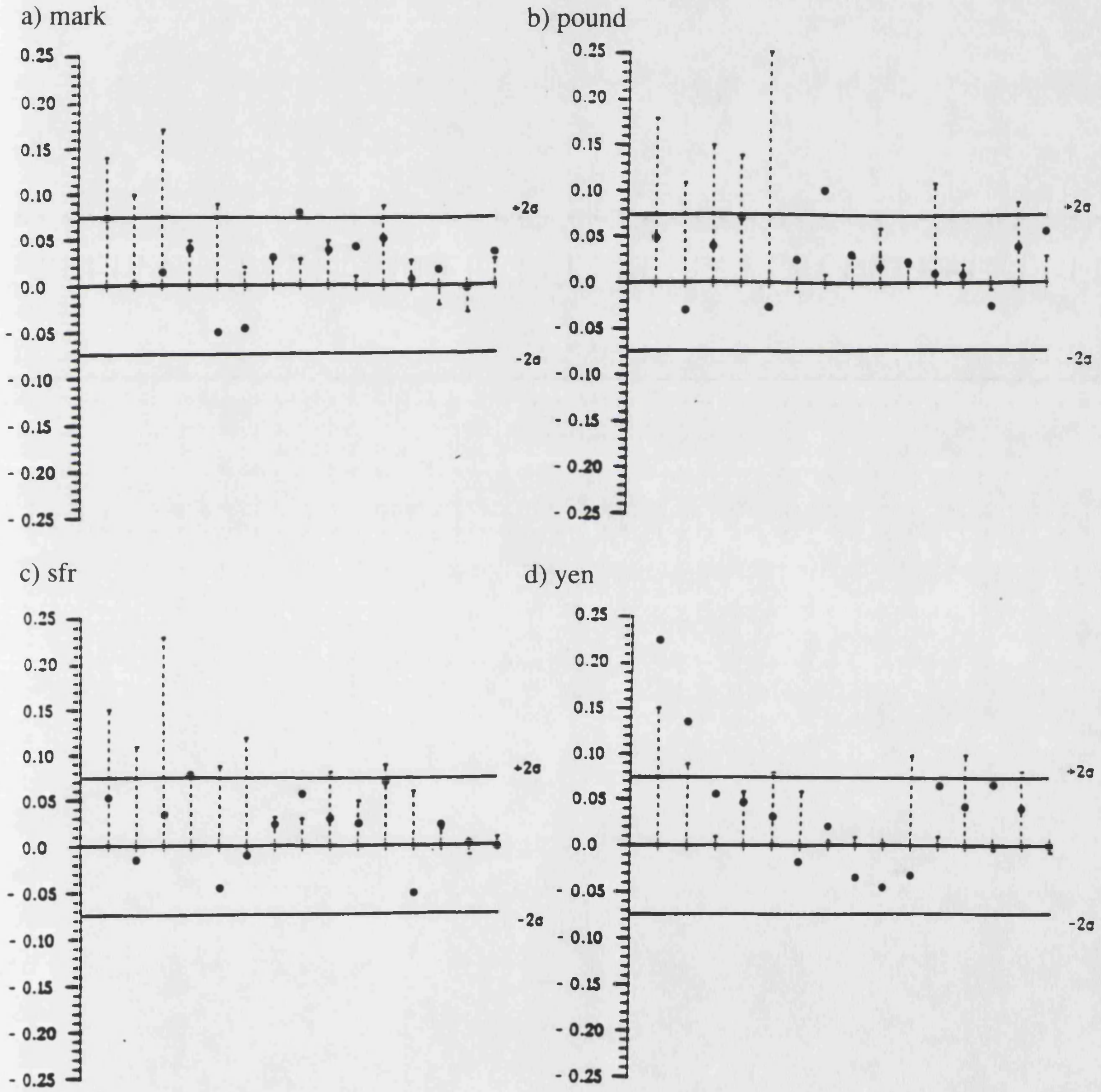
		mark	pound	sfr	yen
day	Q (15)	355.2 *** 15/6	507.3 *** 15/9	561.0 *** 15/10	432.2 *** 15/7
week	Q (15)	61.5 *** 5/4	123.9 *** 7/7	98.2 *** 8/3	52.8 *** 6/3
month	Q (15)	12.3 1/1	12.9 1/1	11.8 1/0	25.0 ** 2/1
quarter	Q (15)	12.0 0/0	7.4 0/0	12.1 0/0	5.4 0/0

Significance levels: see table 3

In order to give some impression about the numerical values of the autocorrelation coefficients and the pattern in the AFC, the AFC for Δe_t and $(\Delta e_t)^2$ is displayed in figure 5 for all four weekly exchange-rates. The AFC for squared data is given by the dashed lines and, for comparison, the AFC for Δe_t is given by the dots in the correlogram. As indicated by the fact that almost all dots lie within the 0.95 percent confidence limits, there is hardly any exploitable structure in the mean of exchange-rate fluctuations. The previously mentioned low R^2 values of ARIMA models for these series reflect this, too (see table 2). However, between 5 (for the mark-dollar rate) and 8 (for the sfr-dollar rate) autocorrelation coefficients among the first 15 are significantly different from zero for the squared data. This suggests that there is some structure in the time pattern of the variance which is worth exploring and modelling. This pattern could have been generated by ARCH processes, bilinear processes or some other processes. For longer time horizons, however, this pattern seems to disappear. With the exception of the monthly yen-dollar series, none of the monthly or quarterly series has significant Q statistics for squared data. The monthly yen-dollar series has significant autocorrelation coefficients $r(k)$ for $(\Delta e_t)^2$ at lags $k = 4$ and 5 and a Q statistic which is significant at the 0.05 level.

Figure 5

Correlogram for Δe_t and $(\Delta e_t)^2$: weekly exchange-rate dynamics



The final investigation of time-series properties of exchange-rate data will be concerned with a very simple test of the arc-sine law for stochastic processes. Let, as before, u_t be the first difference of the logarithm of exchange rates, i.e. $u_t = e_t - e_{t-1}$ is the innovation in a exchange-rate series at time t. Further, let h_τ be the sum of the u_t 's up to time τ , i.e.

(25) $$h_\tau = \sum_{t=1}^{\tau} u_t \tau = 1, \ldots, T$$

and let n_T^+ be the number of positive sums h_τ. In other words, n_T^+ is the number of times that e_t is above e_0 for $t = 1, \ldots, T$. Hence,

(26) $$0 \leq n_T^+ \leq T$$

and

(27) $$0 \leq n_T^+/T \leq 1.$$

For independent u_t having a symmetric continuous probability distribution with expected value of zero and finite variance, the arc-sine law gives the following probability limit for $T \to \infty$ (see Feller (1971), pp. 417-423):

(28) $$plim\left(\frac{n_T^+}{T} < \alpha\right) = \frac{1}{\pi}\int_0^\alpha [x(1-x)]^{-1/2} dx = \frac{2}{\pi} arcsin\, \alpha^{1/2}.$$

The essence of this law is that it is more probable to have n_T^+/T near to zero or near 1 than to have it near 0.5 which is a surprising result. This implies that it is quite likely for the stochastic process e_t to diverge from its starting value e_0 into one direction for quite some time with a low rate of recurrence to e_0.

Equation (28) shows that the arc-sine law can be applied directly to give an asymptotic non-parametric test of divergence from e_0 into the directions $e_t > e_0$ or $e_t < e_0$. Thus, it can be used as a two-sided test. The test results are reported in table 7. The bottom row should be understood to give the probability that n_T^+/T has a value smaller then the empirical one, given in the second row. It suffices to test with daily data.

Table 7

Arc-sine-law test of divergence from e_0 : daily series

	mark	pound	sfr	yen
n_T^+	956	3277	12	654
n_T^+/T	0.282	0.959	0.004	0.194
$Prob(n_T^+/T)$	0.357	0.870	0.037	0.291

At the conventional significance level of $\alpha = 0.05$, the H_0 is only rejected for the sfr-dollar rate. Refering back to figure 1, it can be seen that the sfr tended to appreciate against the dollar from the beginning of the sample period (July 1974) until the late seventies. The following depreciation until March 1985 did not fully compensate the early appreciation. The starting value of $E_0 = 3.0050$ has only been surpassed 12 times in the first couple of weeks but has never been reached again, since.

A rejection of the arc-sine law can be attributed to dependence, skewness, infinite variance, non-identical distributions of the u_t's or a non-zero mean, assuming that the distribution is continuous. Without further information, it is impossible to deduce whether the rejection was caused by dependence, by violations of the assumptions about moments or by non-identical distributions. It is the aim of the next section to provide this kind of information by examining in detail the moments of the data and the distributional properties.

1.3 MOMENTS AND DISTRIBUTIONAL PROPERTIES

The simplest extension of the random walk model for exchange rates can be obtained from the introduction of a non-zero mean; i.e. a random walk with drift:

(29) $$\Delta e_t = \mu + u_t,$$

where μ is the drift parameter. Economic theory suggests a phlethora of factors which has an influence on exchange rates. The two most basic ones can be derived from purchasing power parity (PPP) and interest rate parity (IRP). It is often argued that PPP exerts a long-run influence on exchange-rate movements according to inflation differentials. A short-run influence would only occur under hyperinflation. If, on the other hand, nominal interest rates were determined by a constant real interest rate plus rational inflationary expectation, then IRP would boil down to an ex-ante form of the PPP mechanism. Thus, if a sustained inflation differential between two countries exists, the most basic theory of exchange-rate determination would predict a corresponding depreciation of the high-inflation currency in the long run (see e.g. Mussa (1979), pp. 22-25).

I do not intend to add yet another test of PPP to the already existing numerous tests, but since PPP gives a most basic rationale for introducing a drift parameter into the random-walk model I have calculated the average depreciation rsp. appreciation (on a daily basis) implied by PPP for all four exchange rates. The "PPP means", derived on the basis of wholesale price indices, are given in the first line of table 8 for comparison with the empirical means reported below. This comparison shows that the appreciation of the mark, sfr and yen against the dollar was stronger, from July 1974 to December 1987, than predicted by PPP and the pound's depreciation was smaller than expected under PPP. Thus, with respect to the base period of July 1974, there was a real depreciation of the dollar against all four currencies.

Table 8

Tests for mean and median of zero

		mark	pound	sfr	yen
PPP mean (daily basis)		-0.008	0.018	-0.016	-0.015
mean		-0.014 (1.20)	0.007 (0.63)	-0.025* (1.79)	-0.025** (2.34)
median	day	-0.004 (0.38)	0.0 (0.50)	-0.009 (0.74)	0.0 (1.09)
	week	-0.045 (0.98)	-0.006 (0.26)	-0.053 (0.76)	0.0 (0.11)
	month	-0.237 (1.10)	0.046 (0.16)	-0.390 (1.26)	-0.099 (0.23)
	quarter	-1.317 (0.83)	-0.311 (0.55)	-0.763 (0.82)	-0.770 (0.27)

Significance levels: see table 3

Returning to equation (29), it is natural to test the H_0 that $\mu = 0$. It would be straightforward to apply the t-test for the mean [10], but experience shows that the mean performs quite poorly as an estimator of location in some non-normal distributions (see Rosenberger, Gasko (1983)). As will be shown later in this section, with exchange-rate data one has to take into account fat-tailed distributions. For these distributions, the median is superior to the mean, i.e. more efficient in the statistical sense. Since it is wise not to impose the assumption of normality or other distributional assumptions, a nonparametric test is called for. I apply the simple median sign test whose test statistic is given by (see Kendall, Stuart (1979), pp. 542-546):

$$z = \frac{|B - \frac{1}{2}T_{(0)}| - \frac{1}{2}}{\frac{1}{2}T_{(0)}^{1/2}}, \tag{30}$$

[10] Cornell (1977) performed this t-test for 7 monthly dollar exchange rates. He could reject H_0 only for the pound-dollar series.

where B is the number of observations which are smaller than zero and $T_{(0)}$ is T minus the number of observations which are exactly zero. The test statistic z follows asymptotically a standard normal distribution.

Table 8 reports both the t-test for a mean of zero and the sign test for the median. The t-values for means of the daily data are given in brackets below the means. The mean daily appreciation of 0.025 percent of the sfr and the yen are significant at the 0.10 level and the 0.05 level, respectively. It is superfluous to report the t-statistics for longer time horizons, since the mean of first differences is simply $(e_T - e_0)/T$ and it is only T which varies. Thus, changes of the t-statistics would stem from "unusual behaviour" of the variances under time aggregation. Since a detailed analysis of variances follows later, it suffices to mention that for quarterly data the significance levels drop in the sfr series under the 0.10 level and in the yen series under the 0.05 level. There are no other drastic changes in the results for longer time horizons compared to the t-values for daily data.

Even less evidence against a centre of distribution at zero can be derived from the sign test for the median. None of the z-values, given in brackets below the medians, is significant at the 0.10 level. The yen-dollar rate has a median of exactly zero for daily and weekly movements. There are, however, many zero elements in the yen series of Δe_t. On 146 days, there was no change to the exchange rate of the previous day. This might have been caused by an attempt of the Japanese central bank to peg the exchange rate.

From the point of view of order statistics, the mean and median are two extremes. While the mean attaches equal weight to all observations, the median uses only one or two observations depending on whether the number of observations is odd or even. In recent years, a rich literature on robust estimators of location has emerged. In order to test for a centre of distribution at zero with a robust estimator, I chose a member from the family of robust estimators (in fact, it is a so-called M-estimator) which is efficient in the statistical sense and easy to compute on a computer (see Iglewicz (1983)). This biweight mean b_m is defined by

(31) $$b_m = \tilde{u} + \sum_{|w_t| \le 1} (u_t - \tilde{u})(1 - w_t^2)^2 / \sum_{|w_t| \le 1} (1 - w_t^2)^2$$

where w_t is given by

(32) $$w_t = (u_t - \tilde{u})/9 \cdot MAD$$

and $\tilde{u}$ is the median and MAD is the median absolute deviation from the median. The corresponding robust biweight estimator of scale is

(33) $$s_b = \left[T \sum_{|w_t| \le 1} (u_t - \tilde{u})^2 (1 - w_t^2)^4 \right]^{1/2} / \left| \sum_{|w_t| \le 1} (1 - w_t)^2 (1 - 5w_t^2) \right| .$$

Using these statistics, a simple test of the hypothesis $H_0{:}\mu = 0$ can be performed. The distribution of the test statistic $\sqrt{T}\, b_m / s_b$ is well approximated by a t-distribution with $0.7(T-1)$ degrees of freedom. The results from this test are reported in table 9. The test statistics are given in brackets below the point estimates. Comparing the point estimates of the median with those of the biweight mean, there is a tendency for the biweight mean to give estimates that are further away from zero than those of the median. However, in only one instance is an estimated biweight mean different from zero at the 10 percent level.

Table 9

Test for biweight mean of zero

	mark	pound	sfr	yen
day	-0.004 (-0.39)	0.005 (0.56)	-0.012 (-0.97)	-0.010 (-0.11)
week	-0.033 (-0.67)	0.021 (0.44)	-0.078 (-1.33)	-0.034 (-0.81)
month	-0.288 (-1.13)	0.222 (0.90)	-0.531* (-1.80)	-0.377 (-1.48)
quarter	-1.061 (-1.23)	0.303 (0.38)	-1.268 (-1.26)	-1.370 (-1.60)

* significant at the 0.10 level

Therefore, both from the median test and from the biweight mean test, it seems safe to maintain the hypothesis of a centre of location at zero, i.e. no drift in the random walk of exchange rates. This assumption considerably simplifies the statistical analysis and modelling of exchange rates in various instances.

In the previous section it was shown by analysis in the time domain and in the frequency domain, that there is no strong systematic variation in the mean of u_t. However, there could still be unsystematic variation in the mean of u_t in the sense that within the sample there are periods where the exchange rate followed a random walk with different parameters of drift μ_j. This would still be compatible with an overall drift parameter μ of zero if some μ_j's were positive and some negative. In fact, the naked eye reads some phases of appreciation and depreciation into plots like those given in figure 1.[11]

The constancy of μ will be tested by two fairly standard methods: the K -sample version of the Brown-Mood median test and the Kruskal-Wallis test. The null hypothesis is: $\mu_1 = \mu_2 = \ldots = \mu_K$, where $k = 1, \ldots, K$ is the index for the k -th subgroup of the sample. Without an a-priori perception of the number and size of the subgroups, it is difficult to test H_0 in a rigorous manner. Hence, the following analysis should not be viewed as strict confirmatory testing but rather as exploratory data analysis. Since there is no natural division of observations into K subgroups, different division will be employed. The data will be subdivided into sequences of equal length with lengths T_k of 20, 60, 120, and 240, respectively. This corresponds roughly to time intervals of a month, a quarter, half a year and a year.

The Kruskal-Wallis test imposes the restriction that the u_t's are independent. In view of the results from the first section, this restriction is not too problematic. However, the

[11] McFarland, Pettit and Sung (1982) claimed to have found systematic day-of-the-week effects on the means of 7 dollar exchange rates. So (1987), however, questioned the validity of their results. Hsieh (1988) found mixed evidence on the day-of-the-week effect.

Kruskal-Wallis test also assumes that the u_t's have the same continuous distribution. Since the Kruskal-Wallis test is not based on the normality assumption, this test is nonparametric. The test statistic is based on order statistics of u_t and is calculated by

$$W = \frac{12}{T(T+1)} \sum_{k=1}^{K} T_k \left(\overline{R}_k - \frac{T+1}{2} \right)^2 \tag{34}$$

where $\overline{R}_k$ denotes the mean rank in subsample k. A correction factor for W has been included to take tied ranks into account (see e.g. Hollander and Wolfe (1973), pp. 115-119). The null distribution of W is asymptotically χ^2 with $K-1$ degrees of freedom.

Table 10 reports the results from the Kruskal-Wallis test. The first column gives the approximate period length to which the size of the subsample corresponds and in brackets below it gives the degrees of freedom. Note that since the series of daily observations is slightly longer for the pound-dollar rate and slightly shorter for the yen-dollar rate, the degrees of freedom for subsamples of length 20 (approximately monthly) for these exchange rates are 169 and 167, respectively. The table gives the values of W together with the corresponding probabilities of the upper tail of the χ^2 distribution in brackets. As can be read from the upper-tail probabilities, there is overall a very strong rejection of H_0 in all series. According to the Kruskal-Wallis test, the hypothesis of equal means is rejected in most cases at the 1 percent level of significance. The evidence against a constant mean seems to be strongest for the yen-dollar rate.

The Kruskal-Wallis test is a Chi-squared goodness-of-fit test based on the comparison of expected ranks with actual ranks within a subsample. As mentioned before, the test is based on the assumption that all u_t have the same distribution. Since, apart from variability in the means of u_t, there may be other distributional variability in u_t, it is desirable to employ a test which is less restrictive in its assumptions and more robust to other distributional variability. A fairly well-known test with these properties is the Brown-Mood median test.

Table 10

Kruskal-Wallis test for equality of means

	mark	pound	sfr	yen
month (168)	199.0 (0.051)	182.9 (0.221)	195.1 (0.075)	228.6 (0.001)
quarter (55)	91.2 (0.002)	75.2 (0.036)	83.2 (0.008)	84.2 (0.007)
half year (27)	48.4 (0.007)	49.6 (0.005)	46.2 (0.012)	61.6 (0.006)
year (13)	29.2 (0.006)	26.7 (0.013)	28.6 (0.007)	32.9 (0.002)

Like the Kruskal-Wallis test, it is a Chi-squared goodness-of-fit test. Within each subsample, it compares the actual number of observations below the grand median with its expected number. The test statistic can be written in the form of (see e.g. Conover (1971), pp. 167-172):

(35) $$BM = \frac{T^2}{a(1-T)} \sum_{k=1}^{K} \left(\frac{n_{ak}}{T_k} - \frac{T_k}{T} \right)^2,$$

where n_{ak} is the number of observations below the median in subsample k and a is the corresponding number for the whole sample. Under H_0, BM has an asymptotic χ^2-distribution with $K-1$ degrees of freedom.[12]

As can be seen from table 11 which reports the BM's and the upper-tail probabilities below the BM's in brackets, the results from applying the Brown-Mood median test differ quite substantially from those derived from the Kruskal-Wallis test. According to the

[12]Evans (1986) has developed and applied a similar test. However, Evans (1986) controlled for overlapping subsamples and for picking the subsample with maximum deviations from H_0. Evans rejected the H_0 of no speculative bubble for the dollar-pound exchange rate based on the fact that from 1981 - 1984 outright forward speculation would have been profitable (for the strategy to be short in pounds) in 39 months and resulted in a loss in 9 months only.

Brown-Mood test, there is only strong rejection of H_0 for all period lengths in the yen-dollar rate. There is also some evidence against a constant centre of location for longer time periods in the mark-dollar and in the pound-dollar rate. With the exception of the yen-dollar rate, however, the overall evidence against a non-constant mean is much weaker than from the Kruskal-Wallis test.

The discrepancies in the results from applying these two tests can have various causes. First, since the Kruskal-Wallis test is based on ranks and the Brown-Mood test is only based on signs, the later is certain to have less power than the former. On the other hand, the Brown-Mood test is more robust against variability of other distributional characteristics, such as heteroskedasticity, than the Kruskal-Wallis test.[13] Since the properties of variances in the exchange-rate data will be examined next, this can give a partial answer to the question whether heteroskedasticity might have biased the results from the Kruskal-Wallis test.

Table 11

Brown-Mood median test

	mark	pound	sfr	yen
month (168)	152.8 (0.794)	175.2 (0.357)	155.6 (0.745)	232.8 (0.001)
quarter (55)	69.2 (0.094)	68.5 (0.105)	54.9 (0.477)	92.4 (0.001)
half year (27)	33.6 (0.178)	46.4 (0.011)	26.6 (0.486)	69.3 (0.000)
year (13)	23.5 (0.036)	25.1 (0.022)	15.7 (0.264)	34.5 (0.001)

[13] Hsieh (1988) tested for equality of means within a regression framework employing 119 monthly dummy variables. For 4 out of 5 dollar exchange rates, he rejected the null hypothesis.

It should be recalled that it was shown in the previous section how the squared values of $\Delta e_t = u_t$ exhibit quite strong serial correlation. As already mentioned, this phenomenon can be explained in terms of an autoregressive pattern of heteroskedasticity or in terms of bilinearity. The question is whether other remarkable properties with regards to variances (or more generally: with regards to dispersion or scale) can be found in the exchange-rate data. The variance of a random variable, defined as the second central moment, is supposed to measure the dispersion of a variable.

Since some economists claimed that the distributions of exchange-rate dynamics do not have finite variances (see the next chapter), it is instructive to compare for u_t the variance with two other estimators of dispersion which are quite popular in exploratory data analysis and robust statistics, namely the so-called F-pseudovariance and the 4.2-percent pseudovariance. Both are special forms of the more general $p-$ percent pseudovariance defined by

(36)
$$\nu(p) = \left[\frac{F^{-1}(1-p) - F^{-1}(p)}{\Phi^{-1}(1-p) - \Phi^{-1}(p)} \right]^2,$$

where F^{-1} and Φ^{-1} are the inverse distribution functions of the empirical and a standard normal variable, respectively.[14] The F-pseudovariance employs a value of $p = 0.25$ and the 4.2-percent pseudovariance has, as the name indicates, a value of $p = 0.042$. It is easily seen that under normality $\nu(p) = \sigma^2$, i.e. the pseudo variance is equal to the variance for every p. In this respect, discrepancies between $\nu(p)$ and σ^2 would indicate non-normality. The F-pseudovariance is a standardized interquartile range, i.e. a interquartile range divided by 1.349. It is computed by dividing the difference between the (3T/4)-th and (T/4)-th order statistics of u_t by $\Phi^{-1}(0.75) - \Phi^{-1}(0.25) = 1.349$. In case T/4 and 3T/4 are not integers, the corresponding values of the inverse distribution function are determined by interpolation.

[14] Westerfield (1977) questioned the use of the variance to measure variability of exchange rates and suggested to measure it by Gini's mean difference and a statistic which is essentially a 0.28 pseudovariance.

The 4.2-percent pseudovariance is a very interesting measure of dispersion because it stays almost constant for members of the Pearson system of frequency curves (see Andrews et al. (1972), pp. 166 - 168). The Pearson system of frequency curves will be examined in more detail in the next chapter. Here it suffices to mention that many popular probability functions are subsumed under this system, e.g. the normal distribution, the beta distribution and the t-distribution. Because of its relation to the Pearson system, it has been suggested to call the 4.2.-percent pseudovariance a Pearsonian pseudovariance.

Table 12 compares all three measures of dispersion for the u_t -series. Several points are noteworthy. First, the Pearsonian pseudovariance (PPV) is, with a few exceptions like the case of quarterly mark-dollar fluctuations, very close to the estimated variance (V). This can be interpreted as being an indication for the fact that the probability distribution of exchange-rate fluctuations falls within the Pearsonian system. Second, for shorter time periods, i.e. daily and weekly data, the estimated variance is much greater than the F-pseudovariance (FPV). This result reveals a marked divergence of the distribution of the u_t's from a normal distribution. Since the variance exceeds the F-pseudovariance by a considerable amount for short-run exchange-rate dynamics, it can be conjectured that either the empirical distributions have heavy tails or that there are some outliers in the data. A closer examination of distributional properties will follow shortly. Finally, table 12 shows that for monthly and especially for quarterly data the F-pseudovariance is quite close to the variance. This suggests that discrepancies of the empirical distributions from the normal distribution disappear at longer time horizons of exchange-rate movements.

The apparent non-normality of short-run u_t's casts some doubt on a result derived in the previous section. In studying the autocorrelation function of squared exchange-rate changes, it was found that there is some time dependence in the u_t^2 -series. Since the Ljung-Box statistic Q and the standard errors of autocorrelation coefficients assume normality of the u_t's, the autoregressive pattern of heteroskedasticity may just be an artefact of non-normality. This calls for an investigation of heteroskedasticity with more robust methods. Levene's test for homogeneity of variances is such a robust method in the sense

Table 12

Variance, F pseudovariance and Pearsonian pseudovariance

		mark	pound	sfr	yen
day	V	0.46	0.44	0.67	0.38
	FPV	0.28	0.20	0.35	0.19
	PPV	0.45	0.43	0.62	0.38
week	V	2.16	2.06	2.95	1.64
	FPV	1.08	1.28	1.80	0.94
	PPV	2.26	2.02	3.38	1.62
month	V	11.5	10.5	15.0	11.4
	FPV	8.4	9.6	11.2	7.3
	PPV	12.6	10.0	13.4	11.6
quarter	V	38.9	31.5	54.1	38.2
	FPV	37.1	38.5	56.4	39.8
	PPV	56.2	26.3	61.7	37.1

that its actual size nearly equals its nominal significance level for a variety of underlying distributions. In comparing 56 tests for homogeneity of variances, Conover et al. (1981) found that a version of Levene's test defined by

$$\Lambda = \frac{(T-K)\sum_{k=1}^{K} T_k(\overline{w}_k - \overline{w})^2}{(K-1)\sum_{k=1}^{K}\sum_{t=1}^{T}(w_{kt} - \overline{w}_k)^2} \tag{37}$$

was among the best 3 tests in terms of robustness and power. It is apparent from (37) that Levene's test is based on a one-way analysis of variance for $w_{kt} = | u_{kt} - \tilde{u}_k |$, where $\tilde{u}_k$ is the median of the u_t's in the k -th subsample. The null hypothesis of equal variances in the K subsamples, $H_0{:}\sigma_1^2 = \ldots = \sigma_K^2$, will be rejected if Λ exceeds the $(1-\alpha)$ -quantile of the F-distribution with $(K-1)$ and $(T-K)$ degrees of freedom. The selection of subsamples is the same, quite arbitrary, one as in the case of testing for constancy of means. The results are reported in table 13. The numbers in brackets below the estimated values of Λ are the numerator and denumerator degrees of freedom, respectively. Only for subsamples of length $T_k = 20$ are there small variations in these degrees of freedom due to slightly different overall sample sizes.

Table 13

Levene's test for homogeneity of variance

	mark	pound	sfr	yen
month ($T_k = 20$)	5.2 (168, 3211)	5.7 (169, 3230)	6.1 (168, 3211)	6.5 (167, 3192)
quarter ($T_k = 60$)	11.0 (55, 3304)	13.3	10.4	13.3
half year ($T_k = 120$)	15.6 (27, 3332)	19.0	14.1	22.9
year ($T_k = 240$)	25.8 (13, 3356)	32.7	21.8	42.8

For all entries of table 13 the Λ -estimates fall far into the upper tail of the corresponding F-distribution.[15] In fact, for all 16 Λ -values of this table, the empirical significance level is at least of order 10^{-11}. However, apart from indicating that there is extremely strong evidence for heteroskedasticity in the data, the Levene test does not identify any structure of heteroskedasticity, nor does the test identify the subsamples with abnormal variance.[16]

[15] This confirms and extends Hsieh's (1988) result based on monthly subperiods for 5 daily dollar exchange rates.

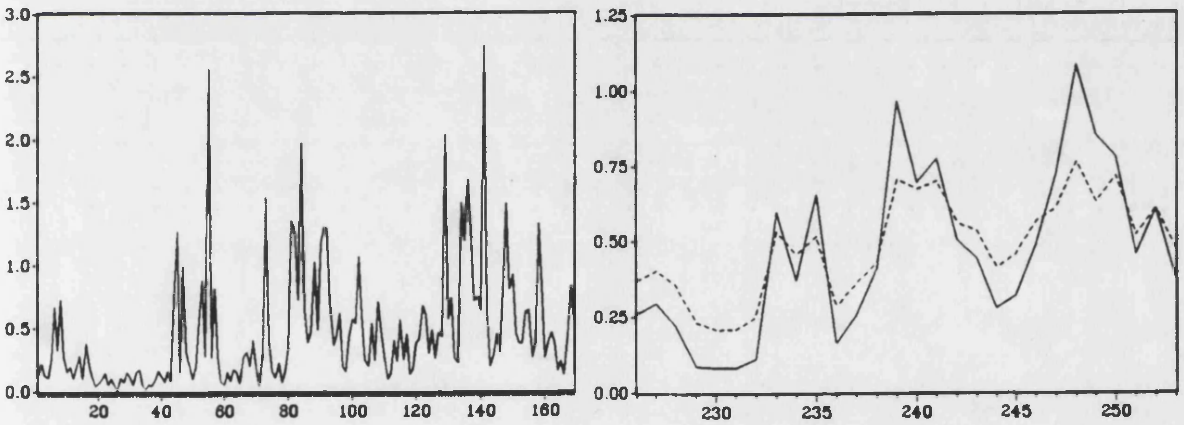
[16] In an early study, Logue, Sweeney and Willet (1978) calculated variances of daily exchange-rate changes for 3 equal subperiods of their total sample which ranged from April 1973 through January 1976. For all 7 exchange rates analysed, they observed a decline in the variance from the first to the last subperiod. They attributed this to the fact that the foreign-exchange markets became deeper and more liquid. However, with a longer sample period (June 1973 - September 1979), Friedman and Vandersteel (1982) showed that there is substantial but unsystematic heteroskedasticity.

Figure 6 provides some insight into the extent and pattern of heteroskedasticity for the mark-dollar exchange rate. The variance for subperiods of length 20, displayed in figure 6a, indeed show marked variability. However, there does not appear to be a clear pattern for these approximately monthly periods. Periods of turbulence and periods of tranquillity seem to follow one another in a rather random way. Looking at the display of variances (solid line) and the subperiod averages of w_{kt} (dashed line) in figure 6b, one tends to detect a positive trend and cyclical variation in volatility. With only 28 observations, however, this probably reads too much into the data.

Figure 6

Variances in subperiods: mark-dollar

a) monthly subperiods b) half-year subperiods



Another, but related, aspect which is interesting to investigate is the influence on the variance when the u_t's are summed. Note that in the previous analysis, daily observations within different time intervals were compared whereas now the observations are summed to get exchange rate movements over longer time spans. Define the variable $h(\tau)$ by

(38) $$h_n(\tau) = \sum_{i=1}^{\tau} u_{n\tau+i} \qquad \text{for} \quad n = 0, 1, 2, \ldots,$$

then it is straightforward to show that

(39) $$\sigma^2_{h(\tau)} = \tau\sigma^2_u$$

or that one gets for the variance ratio

(40) $$VR(\tau) = \frac{\sigma^2_{h(\tau)}}{\sqrt{\tau}\sigma^2_u} = 1$$

if the u_t 's are stationary, independent and homoscedastic. This means, for instance, that, under the stated assumptions, the variance of weekly exchange-rate movements would be expected to have approximately five times the variance of daily data.

There are different ways to view and test (40). It can be shown (see e.g. Cochrane (1988)) that

(41) $$VR(\tau) = 1 + 2\sum_{j=1}^{\tau-j} \frac{\tau-j}{\tau}\rho_j$$

where ρ_j is the autocorrelation coefficient of u_t. Furthermore, $VR(\tau)\sigma^2_u$ is equal to the Bartlett estimator of the spectral density at frequence zero.

There has recently been a renewed interest in the statistic $VR(\tau)$ in connection with the issues of stationarity and mean reversion. Cochrane (1988), for instance, argued that the variance ratio with a long lag τ provides a better measure of persistance of shocks than the application of parsimonious ARIMA models which are based on low order autocorrelations. It can be seen from (41) that the variance ratio is related to autocorrelations with linear declining weights. Under independence $\rho_j = 0$ for all j and, therefore, $VR(\tau) = 1$ for all τ. On the other hand, positive low-order autocorrelations may be compensated by negative high-order autocorrelations and this would imply eventual mean reversion, at least in part.

As a final interpretaion of the variance-ratio statistic, it can be shown that $VR(\tau)$ is equal to the ratio of the variance of a permanent innovation (i.e. the variance of the random walk component) to the instantaneous variance of the series (i.e. the variance σ_u^2). This decomposition of variances is independent of the specification of the time-series model (see Cochrane (1988)).

These interpretations of the VR statistic reveal that it is closely related to time-series properties. From the analysis of the previous section one would conclude that, first, the series e_t has one and only one unit root, second, there is only minor evidence for serial dependence in the series $u_t = e_t - e_{t-1}$ and, third, there is strong evidence for heteroskedasticity. Therefore, I concentrate on the assumptions of independence and homoskedasticity in testing the null hypothesis $VR(\tau) = 1$.

It is quite straightforward to derive an asymptotic distribution for $VR(\tau)$ but Lo and MacKinlay (1988) introduced some interesting modifications of the VR statistic. First, instead of defining $VR(\tau)$ with respect to non-overlapping intervals, as in (38), they suggest to define $VR(\tau)$ with respect to overlapping τ-th differences of e_t in order to extract all information from the series and to obtain a more efficient estimator. The variance of the τ-th difference is, therefore, defined by

(42)
$$\sigma^2_{h(\tau)} = \frac{1}{n} \sum_{t=\tau}^{T} (e_t - e_{t-\tau} - \tau\tilde{\mu})^2$$

where $\tilde{\mu}$ is the estimated mean of the series u_t and $n = \tau(T - \tau + 1)(1 - \tau/T)$. Asymptotically, $VR(\tau)$ has a normal distribution and the test statistic

(43)
$$z_1(\tau) = (VR(\tau) - 1)\sqrt{\frac{3\tau T}{2(2\tau - 1)(\tau - 1)}}$$

has asymptotically a standard normal distribution.

The second modification of the VR statistic allows for heteroskedasticity under the null hypothesis. The corresponding test statistic is [17]

(44)
$$z_2(\tau) = \frac{VR(\tau) - 1}{\sqrt{\gamma(\tau)}}$$

where

[17] Note that Lo and MacKinlay (1988, 1989) report an erroneous formula for $z_2(\tau)$.

(45)
$$\gamma(\tau) = \sum_{j=1}^{\tau-1}\left[\frac{2(\tau-j)}{\tau}\right]^2 \delta(j)$$

and

(46)
$$\delta(j) = \frac{\sum_{t=j+1}^{T} (e_t - e_{t-1} - \hat{\mu})^2 (e_{t-j} - e_{t-j-1} - \hat{\mu})^2}{\left[\sum_{t=1}^{T} (e_t - e_{t-1} - \hat{\mu})^2\right]^2} .$$

Equation (45) is derived from (41) and $\delta(j)$ is the estimated variance of the autocorrelation coefficients. Asymptotically, $z_2(\tau)$ has also a standard normal distribution.

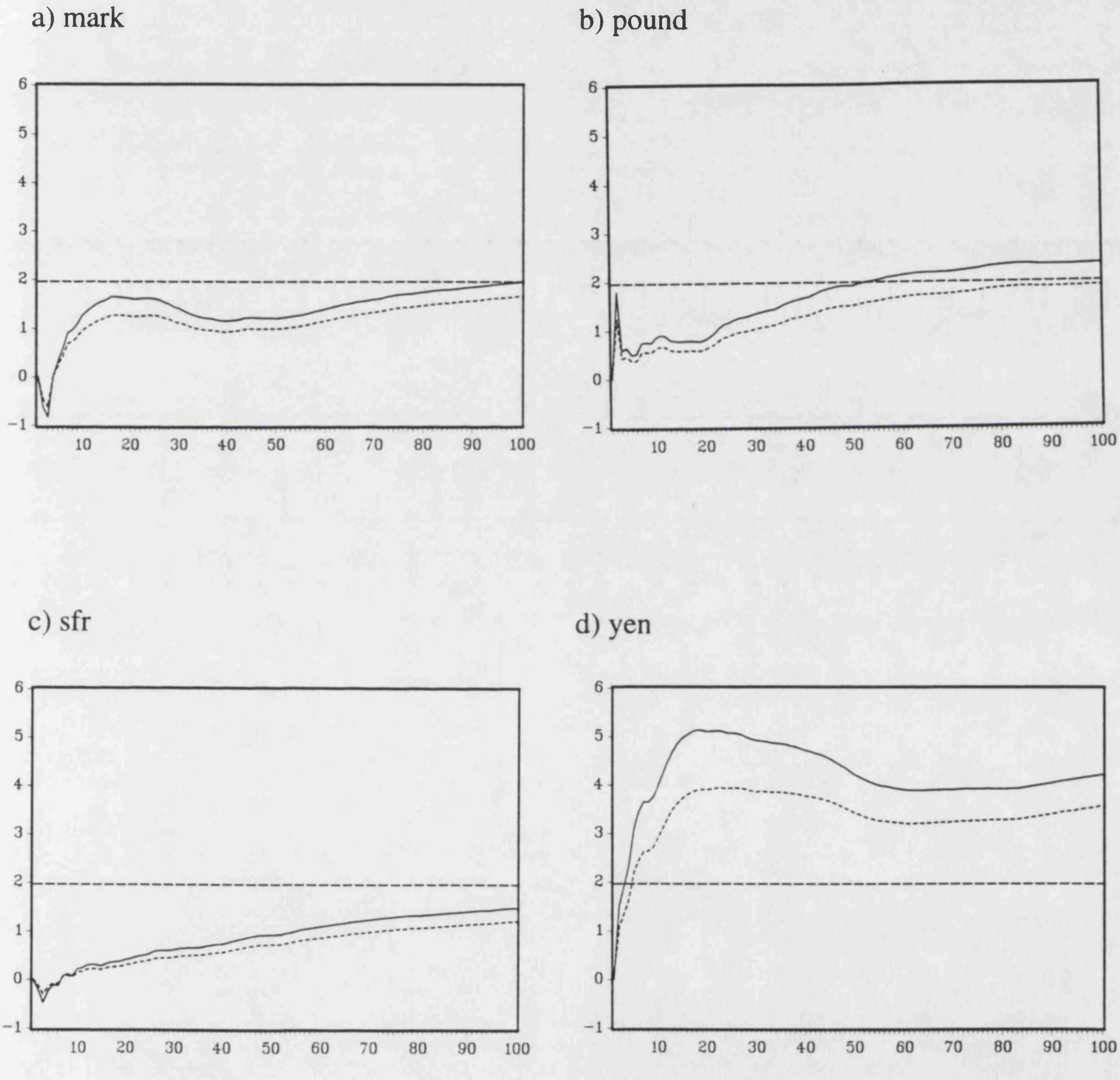
A crucial point in the application of the variance-ratio test is the choice of τ. In extensive simulation studies, Lo and Mackinlay found that both $z_1(\tau)$ and $z_2(\tau)$ are closest to their nominal size when τ is small. More specifically, their simulations showed that for $1 \leq \tau \leq 100$ one gets reliable VR statistics $z_1(\tau)$ and $z_2(\tau)$ for sample sizes of a magnitude similar to the ones of this study.

The function $VR(\tau)$ is sometimes called the variogram. Figure 7 displays, for ease of interpretation, the variogram statistics $z_1(\tau)$ as solid lines and $z_2(\tau)$ as dashed lines. Also shown is the upper critical value of 1.96 corresponding to a significance level of 5 percent. The plots of the variogram statistics show that, with a few exceptions for the mark and sfr series, nearly all variogram statistics are positive, i.e. the variance grows faster than expected under time aggregation. However, none of the $z_1(\tau)$ or $z_2(\tau)$ statistics are significant for the mark and sfr series, although $z_1(100) = 1.95$ for the mark is at the margin to be significant. For the pound series, $z_1(\tau)$ is significant, at the 5 percent level, for $\tau \geq 54$ but $z_2(\tau)$ is not significant for any $\tau \leq 100$. Since $z_2(\tau)$ is robust to heteroskedasticity, this indicates that for the pound, the rejections of H_0 can be attributed to heteroskedasticity. Furthermore, $z_2(\tau)$ is always smaller in absolute value than $z_1(\tau)$ for all four series. In the plots for the mark, pound and Swiss franc, there is a clear tendency for both $z_1(\tau)$ and $z_2(\tau)$ to increase with τ . The upper limit of $\tau = 100$ was not set in order to produce a desired result but in order to conform with the simulation results of Lo and MacKinlay

(1989). They showed that the size of the VR test increases substantially with τ and surpasses the nominal level of $\alpha = 0.05$ between $\tau = 64$ and $\tau = 128$ when $T = 1024$. Furthermore, nearly all rejections come from the upper tail when τ is large relative to T.

Figure 7

Variogram statistics $z_1(\tau)$ and $z_2(\tau)$



For the yen series, $z_1(\tau)$ is significant for all $\tau \geq 4$ and $z_2(\tau)$ is significant for all $\tau \geq 5$. The deviation from the expected variance ratio in the case of the yen is probably not very surprising. Recall from the previous analysis that there was some evidence against the null hypotheses of white noise and of a constant mean for the yen.

One may conclude from the variance-ratio tests that they provide no additional evidence for serial dependence for the mark, pound and sfr, but confirm earlier findings of serial dependence in the yen series. In addition, the discrepancies between the $z_1(\tau)$ and $z_2(\tau)$ statistics for all four series indicate substantial heteroskedasticity in these series.[18]

The final investigation into the behaviour of variances will be concerned with sequential variances. I calculated the variances

(47) $$s_{T_n}^2 = \frac{1}{T_n - 1} \sum_{t=1}^{T_n} (u_t - \overline{u}_n)^2$$

stepwise by setting $T_1 = 10$ and adding 10 more observations at each following step. This gives a sequence of variances for the first 10, 20, 30, ... observations. Granger and Orr (1972) proposed to plot this sequence of variances against T_n. If all u_t come from the same distribution with variance σ_u^2, then $s_{T_n}^2$ should converge to σ_u^2. A failure of convergence could be a sign that u_t does not have finite variance. Since the publication of Mandelbrot's seminal paper (Mandelbrot (1963)), the hypothesis of infinite variance has been very popular in the empirical financial-markets literature. This hypothesis has been put forward to explain fat tails in the frequency distributions of stock-price changes. As will be shown in more detail later, this fat-tail property is also typical for exchange-rate data. Granger and Orr noted, however, that their graphical 'converging variance test' does not give a sufficient condition for the presence of infinite variance since non-convergence

[18] Lin and He (1991) did a similar analysis with weekly exchange rates. They got mixed results but their interpretation is unclear. Furthermore, the relationship between $z_1(\tau)$ and $z_2(\tau)$ seems to be implausible in their study.

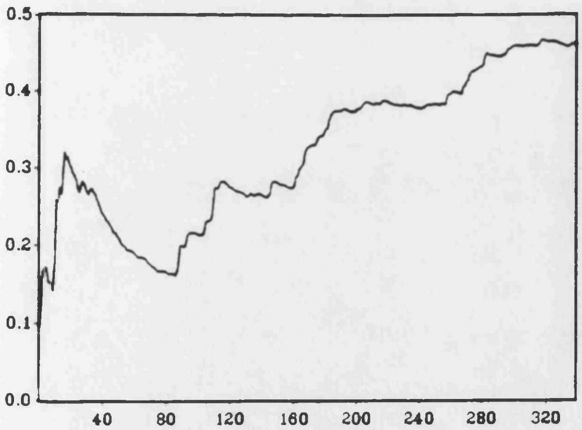
can also be caused by non-stationarity or non-independence. Although no strong conclusion can be drawn from the sequential variances, it is still instructive to look at the plots within an exploratory data analysis. The plots are displayed in figure 8.

The sfr-dollar series and the yen-dollar series, somewhat later than the sfr-dollar rate, appear to have reached almost stationary values of their variances. The mark-dollar series and the pound-dollar series, on the other hand, show an upward tendency in the sequential variances. Note, too, that the sfr-dollar series of u_t has the highest variance among all four series (cf table 12). The upward trend of variances in the mark-dollar and pound-dollar series could thus be interpreted as a convergence to a stationary variance which is similar to that of the sfr-dollar series. On the other hand, if an economic explanation is sought for the fact that the yen-dollar series exhibits a lower level of variance than the other three series, it might be the case that the Japanese authorities restricted the erratic movements of the exchange rate by intervention or other means. As this stage, however, all this is speculative.

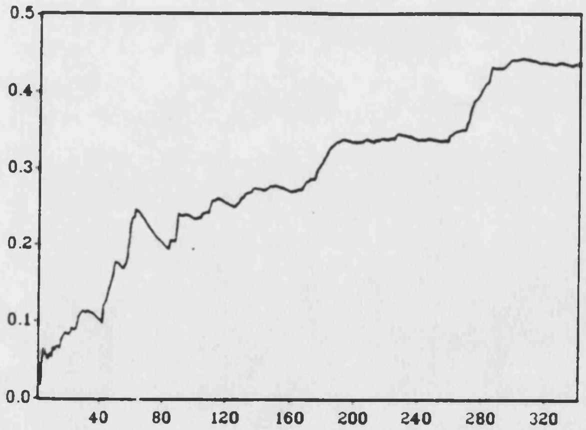
Figure 8

Sequential variances: daily exchange rates

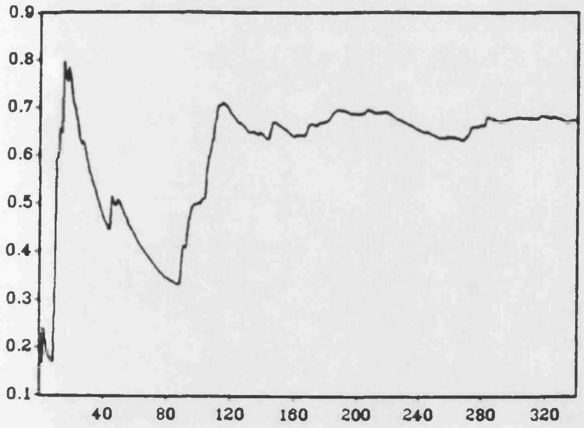
a) mark



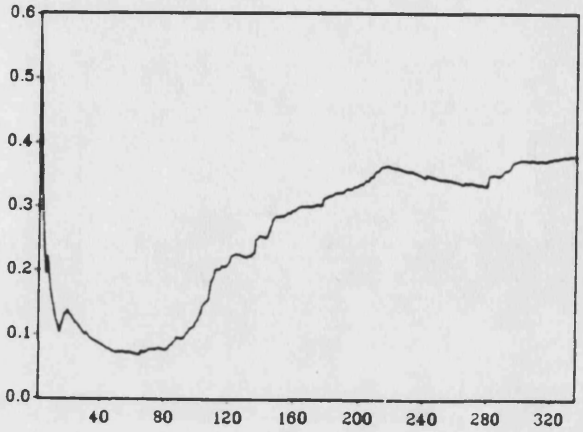
b) pound



c) sfr



d) yen



Turning next to third moments, the skewness of a variable measures the asymmerty of the underlying distribution. The most popular measure of asymmetry is Charlier's skewness measure defined by

(48) $$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} \quad ,$$

where μ_i is the i-th central moment of the distribution. In an empirical measure of skewness, the theoretical moments of (48) are replaced by the empirical ones, hence the empirical counterpart to (48) is

(49) $$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \quad .$$

The normal distribution, like any other symmetrical distribution, has $\sqrt{\beta_1} = 0$. So it is natural to regard a test of $\sqrt{\beta_1} = 0$ as a test for symmetry. Even for a large number of observations, the distribution of the standardized value of $\sqrt{\beta_1}$ does not follow a standard normal distribution under the null hypothesis. However, there is a quite complex transformation available which leads approximately to the standard normal distribution (see Bowman, Shenton (1986)).[19]

The estimated values of $\sqrt{b_1}$ for different time spans of Δe_t are reported as the upper numbers in each cell of table 14. It is apparent from the attached stars, which indicate significance levels, that a test for symmetry based on Charlier's skewness leads to strong rejections of H_0 for daily and weekly data. With the exception of the monthly pound-dollar dynamics, the empirical distributions of Δe_t for longer time spans seem to be free of asymmetry.[20] Furthermore, all estimates of skewness, which are significantly different from zero, are negative, i.e. the frequency distributions are skewed with respect to a depreciation of the dollar.

At this point, it is interesting to reconsider the form in which exchange rate dynamics are analysed. The most widely used form is that of first differences of logarithms, which is also adopted throughout this study, i.e.

(50) $$\Delta e_t = \ln \mathrm{E}_t - \ln \mathrm{E}_{t-1} \quad .$$

Some economists use percentage changes of the exchange rate, i.e.

(51) $$r_t = \frac{\mathrm{E}_t}{\mathrm{E}_{t-1}} - 1 \quad .$$

Since Δe_t is simply the logarithm of $r_t + 1$, the former variable is derived from the latter by a concave transformation. This transformation has the effect of decreasing the coefficient of skewness. Thus, it is interesting to study to what extent the significantly negative values of $\sqrt{b_1}$ are a consequence of using (50) rather than (51) as a measure of

---

[19] I apply here an approximation by Johnson S_μ curves. In large samples, the distribution of the transformed statistic $\sqrt{\beta}'_1$ is approximately normal with mean zero and variance 6/T.

[20] For similar results, but with a somewhat vague interpretation, see Boothe, Glassman (1987).

exchange-rate movements. In each cell of table 14, the lower number is the skewness for r_t. As the comparision with the skewness of Δe_t shows, there are several series (daily pound and sfr, weekly sfr) in which the logarithmic transformation is responsible for significant negative skewness and there is only one case (monthly sfr) where the transformation removes a positive skewness in r_t which was significant at the 0.05 level. However, even for the r_t series, there remains strongly significant negative skewness for the mark-dollar and the yen-dollar dynamics at shorter time spans (day, week).

Table 14

Test of skewness of zero: Δe_t and r_t

		mark	pound	sfr	yen
day	$\sqrt{b_1}(\Delta e_t)$ $\sqrt{b_1}(r_t)$	-0.37*** -0.30***	-0.14*** -0.07*	-0.11** -0.01	-0.61*** -0.55***
week	$\sqrt{b_1}(\Delta e_t)$ $\sqrt{b_1}(r_t)$	-0.30*** -0.19**	0.01 0.15	-0.25*** -0.15	-0.99*** -0.89***
month	$\sqrt{b_1}(\Delta e_t)$ $\sqrt{b_1}(r_t)$	-0.02 0.12	-0.55*** -0.42**	0.19 0.38**	-0.27 -0.14
quarter	$\sqrt{b_1}(\Delta e_t)$ $\sqrt{b_1}(r_t)$	0.15 0.30	-0.004 0.14	-0.43 -0.25	-0.46 -0.33

Significance levels: see table 3

It is instructive to examine transformations to symmetry in more detail for the exchange-rate data r_t . The family of power transformations can be defined by:

(52)
$$y_\zeta(x) = \begin{cases} (x^\zeta - 1)/\zeta & \text{for} \quad \zeta \neq 0 \\ \ln x & \text{for} \quad \zeta = 0. \end{cases}$$

Sometimes, (52) is called the Box-Cox transformation. The aim is to find a value of ζ which is optimal in some sense. For the objective of obtaining symmetry, Emerson and Stoto (1982) proposed a simple exploratory method (see als Emerson (1983)) which is based on so-called letter values (see Hoaglin (1983)). Letter values, which are much used in exploratory data analysis, are approximately equal to the quantiles $\varepsilon_k = 2^{-k}$ with $k = 1, 2, 3, \ldots$. The letter values are order statistics defined by their so-called depth d which is recursively determined by

$$d_k = \frac{1 + \mathrm{int}(d_{k-1})}{2} \tag{53}$$

where $\mathrm{int}(d_{k-1})$ denotes the integer part of the previous depth. The initial depth d_0, corresponding to the median (ε_1), is $(1+T)/2$. To each depth, there corresponds a lower letter value $x_l(k)$ which is the d_k -th order statistic and an upper letter value $x_u(k)$ which is the $(T - d_k + 1)$ -th order statistic. Emerson and Soto propose to plot

$$\frac{x_l(k) - x_u(k)}{2} - \tilde{x} \tag{54}$$

on the ordinate against

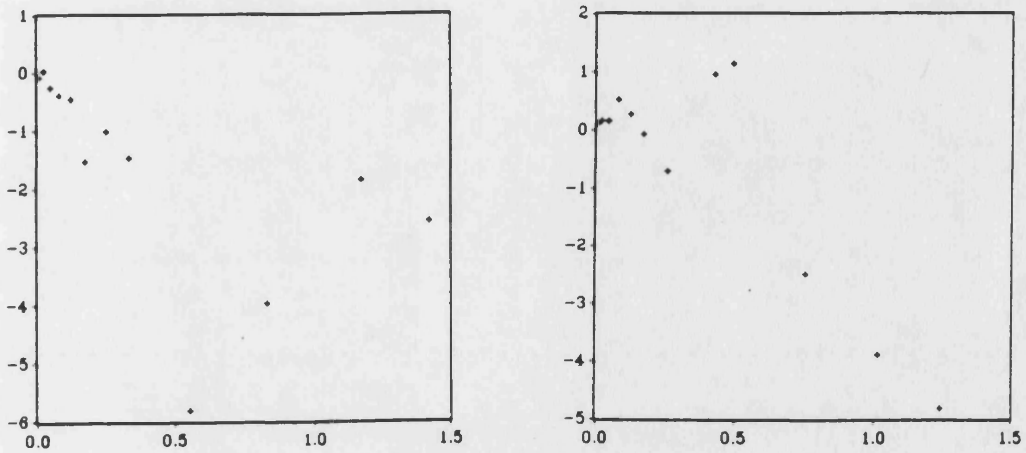
$$\frac{(x_l(k) - \tilde{x})^2 + (x_u(k) - \tilde{x})^2}{4\tilde{x}} \tag{55}$$

on the abscissa. On the ordinate, the difference between the midsummaries and the median $\tilde{x}$ is plotted. For a symmetric distribution, this difference is zero for all values of k .

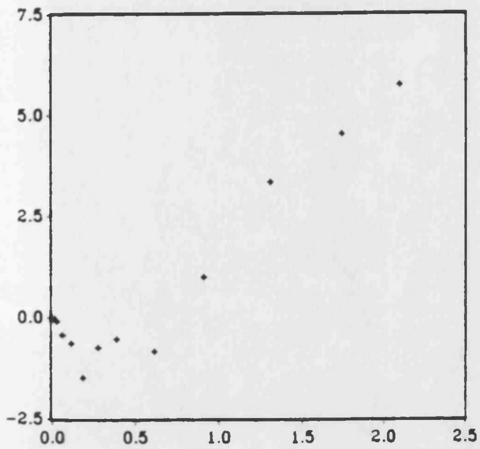
Figure 9

Transformation plots for symmetry: daily series of r_t

a) mark b) pound



c) sfr



d) yen

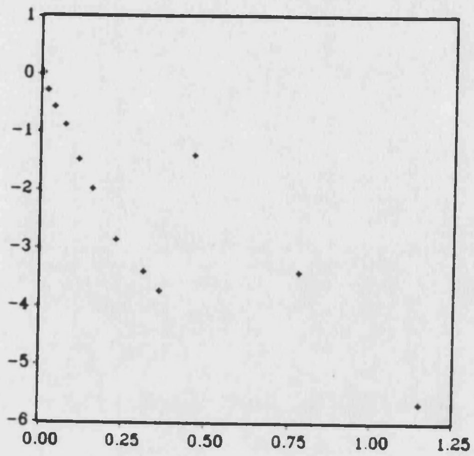


Figure 9 shows the transformation plots for symmetry of daily exchange-rate changes in the form of r_t. If logarithmic transformations were necessary to achieve symmetry, the points in the plot would be close to a straight line with slope 1. If, on the other hand, no transformation to symmetry was necessary, all points would lie around the abscissa. However, the plots suggest that lines with negative slopes would fit the points in the plot best, the only exception being the plot for the sfr whose highest four letter values are somewhat out of line. An estimate of ζ can be obtained from the relation

(56) $$\frac{x_l(k) - x_u(k)}{2} - \tilde{x} \approx (1-\zeta)\frac{(x_l(k)-\tilde{x})^2 + (x_u(k)-\tilde{x})^2}{4\tilde{x}} \quad .$$

If all points in a transformation plot are close to a straight line, a good estimate of ζ is given by 1 minus the slope of the line, as (56) suggests. With some outliers in the plot, like those in figure 9, it is best to choose a robust estimate of ζ. Since (56) can serve to give an estimate of ζ for each value of $k = 1, \ldots, K$, the median of all K estimates ζ_k is a quite robust estimator of ζ. These median estimates are reported in table 15 for all 16 series of r_t. The ζ estimates are highest for the yen-dollar series which also shows stronger skewness to the left in r_t than the other series. In general , there is a quite marked correspondence between the ζ estimates in table 15 and the skewness coefficients in table 14. Large values of ζ correspond to small values of $\sqrt{b_1}$, as it should be. Only one of the estimated power coefficients is near 1. The fact that all other coefficients are quite distinct from 1 should not be used for a recommendation to actually apply the respective transformations. It would certainly seem odd to work with a series of $(E_t/E_{t-1})^{11.7}$ for the weekly yen-dollar rate. Rather, the transformation plots and the estimation of power coefficients are applied here to highlight symmetry or the lack thereof.

Table 15

Estimates of power coefficients ζ

	mark	pound	sfr	yen
day	5.6	-1.2	2.4	12.6
week	3.1	-1.6	4.2	11.7
month	0.3	0.8	0.5	6.4
quarter	-1.0	-0.9	3.1	4.9

In order to round off the analysis of symmetry, I apply a test which is more robust than the skewness test and more rigorous than the transformation plots. It is more robust than the skewness test because it does not assume normality under H_0. This test, suggested by Randles et al. (1980), is based on the U statistics

$$(57) \qquad \hat{\eta} = \frac{1}{3}\binom{T}{3}^{-1} \sum_{i<j<k}\sum\sum \mathrm{sign}(x_{(i)} + x_{(k)} - 2x_{(j)}).$$

In (57), $x_{()}$ is an order statistic of the variable Δe_t. The summation in (57) is over the so-called kernel functions. This sum is more easily understood to be the number of right triples minus the number of left triples, where a right (left) triple is defined by the condition that the mean of the 3 order statistics $x_{(i)}, x_{(j)}, x_{(k)}$ is greater (smaller) than their median. A standardized form of $\hat{\eta}$ has the standard normal distribution as a limiting distribution. Since the test is only based on signs, it is robust to outliers. In particular, it does not assume normality as the skewness test does. Furthermore, it compared favourably to the skewness test in terms of empirical power and empirical size in simulations (see Randles et al. (1980)). The estimates $\hat{\eta}$ are shown in table 16. The results of this U-test are drastically different from those of the skewness test. The evidence of asymmetry disappears virtually. Only in the daily yen-dollar series is $\hat{\eta}$ different from zero at the 0.05 level and at the 0.10 level, the daily pound-dollar series has an $\hat{\eta}$ which is significantly different from zero . All other coefficients are not significantly different from zero. However, for most series is the sign of $\hat{\eta}$ equal to the sign of $\sqrt{b_1}$.

Table 16

U-test for symmetry

	mark	pound	sfr	yen
day	-0.010	-0.031*	-0.015	-0.028**
week	-0.007	0.005	-0.010	-0.028
month	-0.010	-0.017	-0.015	-0.039
quarter	0.015	0.011	-0.039	-0.051

For some statistical models of exchange-rate dynamics, which will be introduced in the next chapter, it is difficult or impossible to incorporate an asymmetric distribution. It is thus fortunate that the U-test provides an alibi for ignoring asymmetry.[21]

There is no intention to examine all moments of order up to k as k goes to infinity, but an analysis of fourth moments, as the highest to be examined, gives some useful insight into distributional aspects of the data. Chapter 2 will provide a justification for analysing only moments up to order 4. The most popular statistic based on fourth moments is the kurtosis β_2 defined by

[21] Calderon-Rossell and Ben-Horim (1982) and Westerfield (1977) tested for symmetry with a sign test on the empirical mean. Westerfield found no evidence against symmetry in 5 weekly dollar exchange rates but Calderon-Rossell and Ben-Horim reported some evidence against symmetry for some of the daily exchange rates they analysed.

(58) $$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad ,$$

i.e. it is the fourth central moment divided by the square of the variance. 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 β_2^* is defined as $\beta_2^* = \beta_2 - 3$.[22] 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, 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 β_2 is standardized, is to remain constant.

In the empirical measure of kurtosis (b_2), the theoretical moments are replaced by empirical ones. A test of the null hypothesis H_0:$\beta_2 = 3$ is, of course, a test of normality, but more specific it is a test for mesokurtosis with the two-sided alternatives of platykurtic $(\beta_2 < 3)$ and leptokurtic $(\beta_2 > 3)$ distributions, i.e. the alternatives are either strong shoulders or fat tails and / or peakedness. A quite complex transformation of b_2 leads to a test statistic which has an approximate standard normal distribution under H_0 (see D'Agostino (1986)) [23].

The values of b_2 are reported as the upper number in each cell of table 17 for the series of Δe_t. As the table shows, there is extremely strong leptokurtosis in the daily and

[22] There is a considerable amount of confusion in the literature about the meaning and even about the definition of kurtosis. Often, kurtosis and excess are confused. Some standard statistical software packages (like SAS and RATS) report skewness and kurtosis statistics but actually do not compute them as they are defined in (49) and (58). Instead they compute Fisher's k statistics which are only asymptotically equal to skewness and kurtosis, respectively (see D'Agostino et al. (1990)).

[23] Even for large samples, the distribution of β_2 is not normal under the H_0 of normality. I apply here an approximation due to Anscome and Glynn. Sometimes the normal approximation with a mean of 3 and a variance of 24/T is used. However, even for samples with $T = 1000$ this is a poor approximation because, due to the lower bound of β_2 at 1, the distribution of β_2 is skewed.

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.[24]

Table 17

Test for mesokurtosis and octile measure of shape

		mark	pound	sfr	yen
day	b_2 M	8.32*** 1.49	8.36*** 1.75	8.89*** 1.54	8.00*** 1.64
week	b_2 M	5.84*** 1.69	7.36*** 1.47	4.96*** 1.45	7.03*** 1.56
month	b_2 M	3.87** 1.34	4.15*** 1.32	4.19*** 1.47	3.62 1.60
quarter	b_2 M	2.67 1.24	2.72 1.20	2.77 1.34	2.62 1.01

Significance levels: see table 3.

Due to the fact that β_2 is based on the standardized fourth moment, this measure of distributional shape is clearly outlier-prone. It would, therefore, be desirable to have a robust measure of shape, i.e. a measure of concentration of mass in the tails and near the centre of distribution which is more resistant to extreme observations. Recently Moors (1988) suggested such a measure based on the octiles of a distribution. The i-th octile O_i of a random variable x is defined by $P(X \leq 0_i) = i/8$, for $i = 1, \ldots, 7$. Since peakedness and heavy tails imply a reduction of mass in the shoulders of the distribution, Moors proposed to measure these distributional shapes by the concentration of mass around the 2nd and 6th octile. Moors' shape measure is defined by

[24] See also Boothe and Glassman (1987).

(59) $$M = \frac{O_7 - O_5 + O_3 - O_1}{O_6 - O_2}$$

The approximation of the shoulders of the distribution by the distance between O_1 and O_3 and between O_5 and O_7 can probably be optimized, but it seems to capture the essential shape characteristics well. Moors (1988) showed that there is in general a positive but non-linear association between β_2 and M for several families of distributions.

The normal distribution has $M = 1.23$ To facilitate the interpretation of the M measure for the exchange-rate series reported in table 17, some further benchmarks are needed. The double exponential (or Laplace) distribution has $M = 1.59$ and $\beta_2 = 6$ (see Moors (1988), p. 29). For the t-distribution with ν degrees of freedom, the kurtosis is given by $\beta_2(\nu) = 3\ (\nu - 2)/(\nu - 4)$ $\nu \geq 5$. This yields $\beta_2\ (5) = 9$, $\beta_2\ (7) = 5$, $\beta_2\ (10) = 4$, $\beta_2\ (30) = 3.38$, for example. On the other hand, the corresponding Moors measures of shape for the t-distribution with different degrees of freedom are: $M\ (5) = 1.33$, $M\ (7) = 1.30$, $M\ (10) = 1.28$, and $M\ (30) = 1.25$. Moors M coefficients for the 16 exchange-rate series are reported in the bottom row of each cell in table 17. For each exchange-rate, there is a tendency for M to decrease as the time interval increases. Without formal testing of M, which is not yet available, it is difficult to make precise statements about convergence to normality under time aggregation as measured by M. However, quarterly series would clearly be classified as showing no sign of peakedness or heavy tails, whereas the evidence for peakedness or heavy tails is quite strong in all daily and weekly series. For monthly series, M is not quite so high but, as compared with the benchmark values from the double exponential and the t-distribution, there are still some marked deviations from the shape of a normal distribution. On the whole, the results from the outlier-resistent measure M

confirm the results derived from b_2: short run exchange-rate dynamics are characterized by an excessive amount (compared with the normal distribution) of very small and of very large fluctuations.[25] Under time aggregation this property disappears.

Finally, I will examine the distributional properties of exchange-rate data in more detail and in a more general form. If one wishes to test the empirical distribution of exchange rates against a specific probability distribution, one would certainly pick the normal distribution for the null hypothesis first. This choice can be justified in two respects. First, the central limit theorem gives this distribution as the limiting distribution for a sum of independent and identically distributed random variables with finit mean and variance. The foregoing analysis has shown that the assumptions of mean-independence and finite mean and variance are probably unproblematic. The critical assumption, however, seems to be the one of identical variances. It also fits well into the asset-market theory of exchange-rate determination to regard exchange-rate dynamics as being caused by a large number of random variables, see equation (5). Second, the normal distribution plays a central role in statistical testing and many of the previous tests are based on the normality assumption. Hence, testing for normality can shed some light onto the question whether these tests are reliable or whether the non-parametric tests, which also have been applied, are more appropriate.

Based upon previous skewness and kurtosis tests, one can produce a powerful omnibus test of normality. The test statistic is

(60) $$G^2 = z_1^2(\sqrt{b_1}) + z_2^2(b_2)$$

[25] Friedman and Vandersteel (1982) found excessive mass in the tails of the distribution by counting the number of daily observations beyond ± 3 standard deviations. They also detected abnormally long tails by applying the studentized range statistic.

where z_1 and z_2 are derived from the approximations to the normal distributions discussed above (see also D'Agostino (1986)). Under the H_0 of normality, G^2 has approximately a χ^2 distribution with 2 degrees of freedom[26]. The results are reported as the upper number in each cell of table 18.

Overall, the same pattern emerges as in the kurtosis test. The omnibus test rejects normality for all daily and weekly data at very high significance levels. On the other hand, there is only moderate evidence against normality in the monthly data and no evidence at all in the quarterly data. Interestingly, the omnibus test does not reject normality for the monthly mark series, although this series has significant leptokurtosis (at the 5 percent level). For the monthly yen series, on the other hand, the omnibus test rejects normality at the 10 percent level, although none of the two component tests is significant at this level.

There are two important classes of goodness-of-fit tests for normality: the moments based tests, to which the omnibus test belongs, and tests based on empirical distribution functions (EDF). Stephens (1986) recommends the Anderson-Darling test from the class of EDF tests. Like the popular Kolmogorov-Smirnov test, the Anderson-Darling test is based upon the vertical difference between the empirical distribution function and the theoretical distribution function F, but it has more power than the former.[27] The test statistic is given by:

$$AD = -T - \frac{1}{T}\sum_{i=1}^{T}[(2i-1)\ln F_i + (2T+1-2i)\ln(1-F_i)]. \tag{61}$$

[26] Some standard econometric software packages like Microfit and PC-GIVE employ the Jarque-Bera test of normality which is also an omnibus test of normality based on the skewness and kurtosis statistics (see Jarque and Bera (1980)). However, the Jarque-Bera test employs the simple normal approximation to the distribution of $\sqrt{b_1}$ and b_2. As mentioned before, this approximation is poor even for large samples with T = 1000. The same applies to the very similar test of Kiefer and Salmon (1983).

[27] As D'Agostino (1986, p. 406) put it: "For testing for normality, the Kolmogorov-Smirnov test is only a historical curiosity. It should never be used." Giddy and Dufey (1975) applied the Kolmogorov-Smirnov test for 3 daily exchange-rate series of the early 1970's. They rejected the H_0 of normality for all series. Boothe and Glassman (1987) used a Chi-squared test and the Jarque-Bera test. Their results agree with mine reported below.

In order to perform the test, the standardized observations have to be put in ascending order $(i = 1, \ldots, T)$ and the corresponding value of the cumulative normal distribution function has to be calculated for all observations. Note that the Anderson-Darling test does not assume the mean and variance to be known. For a small number of observations, there is a modification of the test statistic given by

$$AD^* = AD\left(1 + \frac{0.75}{T} + \frac{2.25}{T^2}\right). \tag{62}$$

Table 18 shows the results for all 16 series. As the asterisks, representing significance levels, indicate, the results from the AD statistic are quite similar to those from the omnibus test. Normality is overwhelmingly rejected for daily and weekly data. It is quite peculiar that the only monthly series for which the normality of u_t is rejected is the yen-dollar series whereas in the omnibus test, normality could not be rejected at the 5 percent level for this series. This demonstrates that these tests are sensitive to different distributional aspects. With the exception of the monthly yen-dollar series, the normality assumption seems to be a good approximation for monthly and quarterly exchange-rate dynamics.

Table 18

Tests for normality

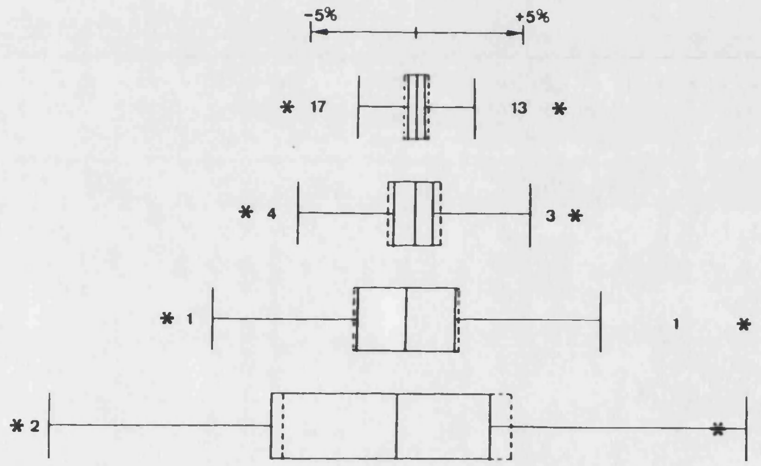
		mark	pound	sfr	yen
day	G^2	470.4 ***	414.5 ***	434.9 ***	556.2 ***
	AD^*	24.1 ***	45.5 ***	32.9 ***	43.2 ***
week	G^2	62.3 ***	76.6 ***	42.3 ***	115.1 ***
	AD^*	6.5 ***	5.6 ***	5.3 ***	8.9 ***
month	G^2	4.0	13.7 ***	7.1 **	4.7 *
	AD^*	0.6	0.5	0.6 *	2.2 ***
quarter	G^2	0.3	0.1	1.9	2.5
	AD^*	0.2	0.4	0.5	0.7 *

Significance levels: see table 3

In order to gain more insight into the distributional properties, box-plots are displayed in figure 10 for the four sfr-dollar series. Boxplots have become a popular graphical tool in exploratory data analysis. It is a display based on seven order statistics. A box is drawn such that its left and right ends are at the values of the first and third quartile, respectively, while the crossbar represents the median. Next, a line is drawn from each end of the box to the most remote data point which is not regarded as an outlier. This point is marked by a fence (sometimes also called whisker). The fences represent those data points which have probability $1/T$ of being surpassed if the data came from a normal distribution.

Figure 10

Boxplots for sfr-dollar exchange-rates



The parameters of the normal distribution are the empirical median and Pearsonian pseudovariance. The asterisks show the two extreme observations, i.e. the minimum and maximum. The number between an asterisk and a fence indicates the total number of observations beyond that fence. Finally, dashed boxes are drawn into the plots to indicate the first and third quartiles which would be obtained under a normal distribution with the variance estimated by the Pearsonian pseudovariance. The scale above the boxplots indicates the magnitude of a 5 percent appreciation and a 5 percent depreciation, respectively.

Several features of the data can be read from the boxplots of the sfr-dollar series which are representative of boxplots of all series. First, there is no apparent asymmetry in the distributions since the median lies roughly half way between the quartiles. Second, since the actual first and third quartiles are, in general, somewhat nearer to the median than the normal quartiles, there must be some concentration of mass around the centre. However, peakedness cannot come out clearly from these boxplots because the quartiles extent right into the shoulders of a distribution. Third and most important, for short-run exchange-rate dynamics, especially for daily data, there is a great number of observations which can be classified as outliers if reference is made to the normal distribution. For instance, the lower and upper quantiles marked by the smallest and largest data in the daily sfr-dollar series have probabilities of the order 10^{-7}, i.e. observations of this magnitude would be expected to occur once, on average, among 10 million observations. On the other hand, monthly and quarterly data do not contain many outliers. There is one extreme monthly depreciation of the sfr [28] but there is no excessive number of observations beyond the fences in any of the monthly and quarterly series.

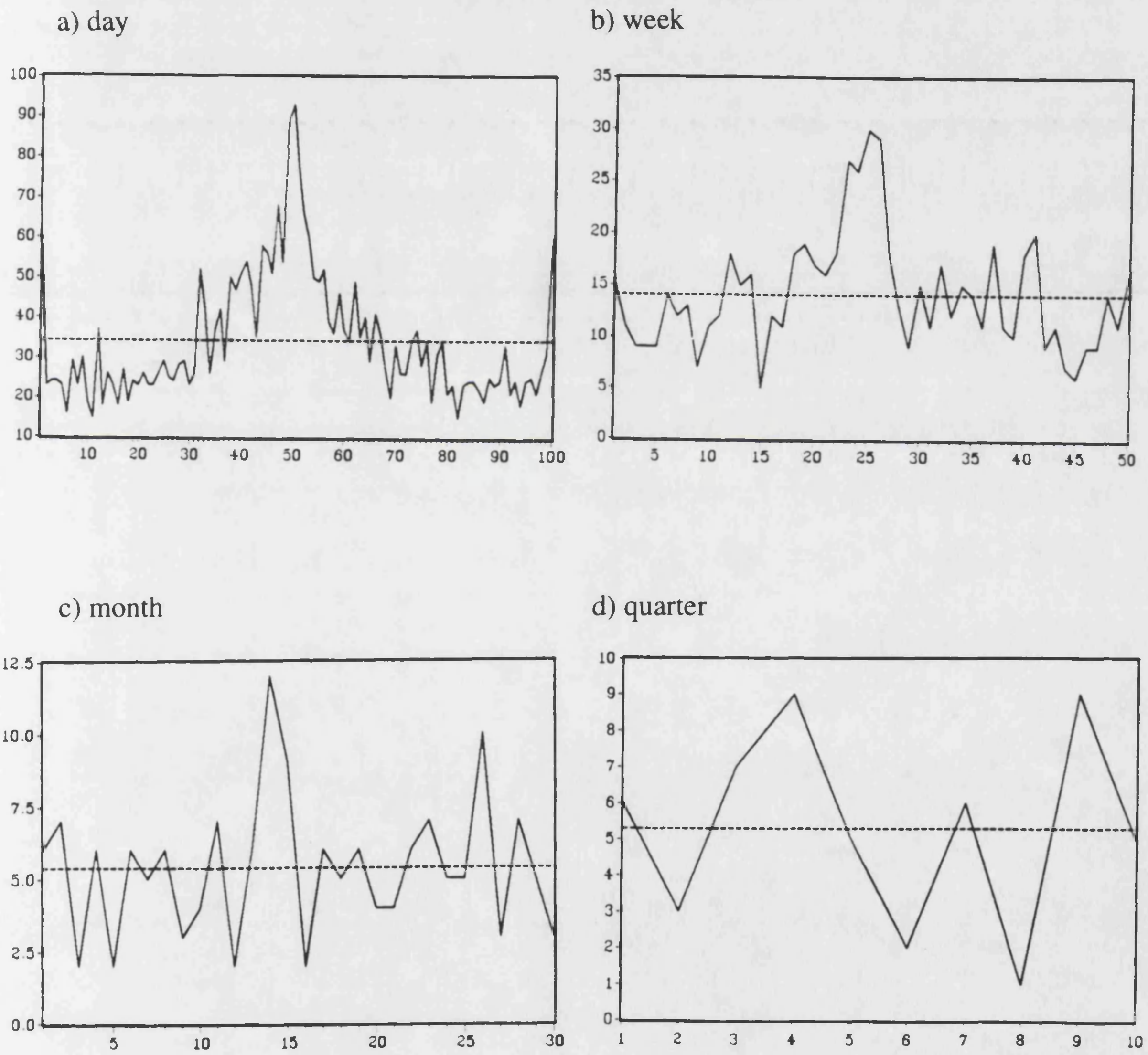
More detailed information about the shape of the empirical distributions in comparison with a normal distribution is provided by figure 11. As in a Chi-squared test of fit, the normal distribution is divided into K equi-probable quantiles. A rule of thumb for the optimal number of quantiles is given by $K = 4(2T^2/c(\alpha)^2)^{1/5}$, where $c(\alpha)$ is the upper α-quantile of the standard normal distribution and α is a chosen significance level (see Moore (1986), pp. 69-70). For $\alpha = 0.05$ one gets approximately 100, 50 and 30 quantiles

[28] This depreciation by 16.7 percent occurred in November 1978. It was caused by strong support measures to stop the previous decline of the dollar. They were announced on November, 1st by the Carter administration and within a day the dollar appreciated by 5.6 percent.

for the daily, weekly and monthly series, respectively. In order to conform with Cochrane's rule, the number of quantiles is 10 for quarterly series. For each quantile, the actual number of observations (solid line) is compared with the expected number under a normal distribution (dashed line). To save space, only the comparisons for the four pound-dollar series are displayed in figure 11.

Figure 11

Empirical and normal distribution: pound-dollar



The previous analysis showed that there is significant departure from normality in short-run exchange-rate dynamics, i.e. in daily and weekly series. The non-normality was revealed both by the kurtosis test and by the Anderson-Darling test. Leptokurtosis, however, can be caused by peakedness or by heavy tails. Figure 11 shows that in daily pound-dollar series, excessive mass can be found in the tails and at the centre of the distribution. In the weekly series, however, non-normality seems to be entirely due to peakedness. Only a mild form, if any, of peakedness can be seen in the monthly pound-dollar series. In the few quantiles of quarterly series, no pattern of discrepancies between actual and expected numbers of observations can be detected. It should be emphasized that the other three exchange rates show exactly the same pattern in their quantiles with the exception that there is no sign of peakedness in the other three monthly series.

1.4 PROPERTIES OF EQUIDISTANT DATA

The stylized facts produced in this chapter shall serve as guidelines for the stochastic modelling of the exchange-rate data. However, before this modelling is undertaken, I want to check whether there is a trivial explanation for heteroskedasticity and leptokurtosis in the data. It should be borne in mind that the observations in all series are not strictly equidistant in time. If one assumes that a stationary stochastic process, which is continuous in real time, generates the exchange-rate dynamics then gaps in the time series caused by weekends and holidays provide the most obvious explanation for leptokurtosis. To take the simplest possible case, assume that daily, or more precisely: 24 hour, exchange-rate dynamics are generated by an independent and stationary Gaussian process with variance σ_u^2. Changes in the exchange-rate over the weekend would then have variance $3\sigma_u^2$. It is straightforward to show that a mixture of both distributions exhibits leptokurtosis (see the next chapter). Hence, it should be checked whether non-normality in short-run exchange-rate dynamics is due to the lack of equidistance. Table 19 provides an answer.

Table 19

Statistics for equidistant daily data

	mark (2628)	pound (2699)	sfr (2663)	yen (2594)
variance	0.43	0.40	0.64	0.33
implied variance	0.33	0.31	0.48	0.24
F-Pseudo-variance	0.25	0.18	0.33	0.17
kurtosis	7.86 ***	7.92 ***	9.04 ***	6.99 ***
Anderson-Darling test	20.13 ***	34.92 ***	27.78 ***	30.14 ***

Significance levels: see table 3.

It reports some statistics for the 24-hour changes in the daily series. The number of observations, after eliminating all weekends and holidays from the daily series, is reported in brackets below the name of the series. The variances for the equidistant observations are slightly smaller than the variances for all daily observations (cf table 12), but they are much larger than the F-pseudovariances. As a first approximation, implied variances are calculated under the assumption that the data were generated by a mixture of normal distributions with variances σ_u^2 and $3\sigma_u^2$ and probabilities $P_1 = 4/5$ and $P_2 = 1/5$, respectively, i.e. the implied variance is 5/7 times the variance of all daily observations. The variances of all equidistant series are much larger than the corresponding implied variances, indicating that weekend and hoiliday effects do not provide a satisfactory explanation of the observed phenomena. This is also confirmed by the kurtosis test and the Anderson-Darling test: both reject on extremely high significance levels the null hypotheses of mesokurtosis and normality.

1.5 SUMMARY

The aim of this chapter was to provide a comprehensive analysis of the statistical properties of exchange-rate dynamics. The analysis was basically exploratory in nature. Since interest in empirical exchange-rate modelling seems to switch from econometric models to statistical (probability) models, it is essential to have a clear idea about the empirical regularities of exchange rates which such a model has to embody.

In his well-known survey of empirical regularities in the behaviour of exchange rates, Mussa (1979) stated 19 general regularities. Only one of those was concerned with a statistical property, namely: the natural logarithm of the spot exchange rate follows approximately a random walk. The analysis of this chapter provides a more detailed and accurate account of the statistical regularities. The results are summarised in table 20. The null hypotheses which were tested, sometimes by more than one method, are reported in the first column. A plus sign indicates that the null hypothesis cannot be rejected and a minus sign indicates rejection. The table is only meant to give a broad overview of the results and to filter out the strong properties. Note, too, that a certain number of rejections of null hypotheses has to be expected due to type I errors.

The evidence concerning serial mean-independence is mixed (for this reason the pluses have been put into brackets). Within ARIMA models, departure from serial mean-independence can only be found in quarterly series. On the other hand, Dufour's signed-rank test reveals moderate serial mean-dependence in short-run exchange-rate dynamics only. Since Dufour's test imposes very weak assumptions and is thus robust, more reliance should be placed on this test than on the resullts derived in the ARIMA framework. The hypothesis of one unit root cannot be rejected for any of the 16 series. This justifies the use of first differences (in the logarithm) in the time-series analysis of exchange-rates.

No hidden periodicities can be found for three out of four exchange-rates. Spectral analysis detects significant deviations from white noise only for the yen. The observant reader will have noticed that for several statistical properties there is a sharp difference between short-run (i.e. daily and weekly) and medium-run (i.e. monthly and quarterly) exchange-rate dynamics. Two such properties are first order independence in distribution (Markov property) and serial homoskedasticity. Both properties are very strongly rejected for short-run dynamics. The rejections are obviously due to the fact that there is a tendency for large exchange-rate movements to be followed by large movements of either sign. To put this into economic terms: there is evidence for short-run periods of turbulence and tranquility.

Table 20

Summary of results on statistical properties

Null hypothesis	day	week	month	quarter	remark
Mean-independence	(+)	(+)	+	(+)	
Unit root	+	+	+	+	
White noise	+	+	+	+	rejection for yen
Markov property	-	-	+	+	
Serial homoskedasticity	-	-	+	+	
Arc-sine law	+				rejection for sfr
Zero mean	+	+	+	+	
Constant mean	+	+	+	+	rejection for yen
Homogeneity of variance	-				
Symmetry	+	+	+	+	
Mesokurtosis	-	-	-	+	
Normal distribution	-	-	+	+	

For short-run dynamics, there is also highly significant non-normality and excess kurtosis. For daily changes, this can be attributed to fat tails and peakedness but for weekly data, and to some extent for monthly data, this seems to be due only to peakedness.

Another very remarkable property is the heteroskedasticity which can be found in all exchange-rates at all time horizons. On the other hand, there is conflicting evidence concerning the constancy of the mean. The Kruskal-Wallis test generally rejects this null hypothesis but the alternative Brown-Mood test rejects it only for the yen. Since the Brown-Mood test is less restrictive than the Kruskal-Wallis test, in particular is does not assume homoskedasticity, it appears to be safer to put more weight on the results from the Brown-Mood test than on those from the Kruskal-Wallis test.

Conflicting results are also obtained on the symmetry of distribution. The coefficient $\sqrt{b_1}$ shows significant negative skewness at short periods. This result is to some extent due to the specific form in which exchange-rate dynamics are analysed, namely the form of first differences in logarithms. Applying the asymptotically distribution free U-test, however, gives no evidence of asymmetry. Again, more reliance should be placed on the U-test than on the $\sqrt{b_1}$ -test because the U-test is more robust than the latter.

Finally, the convenient hypothesis of a zero mean cannot be rejected by either of the two robust tests applied (median test and biweight mean test). With the exception of the sfr-dollar rate, there is also reversion to the mean in the series which is compatible with the assumption that the exchange-rate dynamics follow a symmetric stochastic process with mean zero. The main results can be succinctly summarized in the following three statements:

i) All series of exchange-rate dynamics show approximate serial mean-independence, no periodicities (with the exception of the yen), a constant mean at zero (with the exception of the yen) and symmetry in distribution.

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

iii) Medium-run exchange-rate dynamics (i.e. monthly and quarterly changes) show no serial heteroskedasticity and have a frequency distribution which is approximately normal.

CHAPTER 2

COMPOUND-DISTRIBUTION MODELS OF EXCHANGE RATES

In this chapter, I shall introduce and estimate several stochastic models of exchange-rate dynamics. These models are supposed to capture the main empirical regularities of short-run exchange-rate data. The analysis in the previous chapter has demonstrated that daily and weekly data show a significant amount of heteroskedasticity and leptokurtosis. On the other hand, monthly and quarterly exchange-rate data cannot, in general, be distinguished from Gaussian white noise, i.e. there is convergence towards independence, stationarity and normality under time-aggregation.

The models to be analysed in the next two chapters are compatible with some or all of these empirical properties. The models can be classified into two groups. The first group (which will be analysed in this chapter) consists of four models which are all static in the sense that the probability distribution at time t is constant for all t. Thus, these models cannot capture the property of serial dependence in variances. The four models in this group are the finite 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 I will show that they can all be viewed as compound normal distributions where an independent probability distribution is attached to the variance of a normal variable. They are, therefore, called scale-compounded distribution models. Although these models are based on the notion of stochastic variance, it would be wrong to say that they imply heteroskedasticity since this term is usually only used for cases where the variance is a function of time or some other variables. The choice of the normal distribution as the compounded distribution is somewhat arbitrary, of course, but one can always refer to the central limit theorem as a justification of this choice, whereas it is unclear how alternative distributions could be selected for application.

The second group of models (which will be analysed in chapter 3) is designed to capture heteroskedasticity. Markov-switching models and ARCH-type models belong to this group. Markov-switching models are rather straightforward extensions of finite mixture distributions and ARCH models were introduced in the previous chapter. These models are not only able to capture heteroskedasticity, but it can also be shown that they imply leptokurtosis.

I shall test these models separately against the null hypothesis of Gaussian white noise but it is also essential to compare the two groups since they model different aspects of the data. In chapter 4, the comparisions will be made in terms of goodness-of-fit, in terms of likelihoods, and in terms of predictive power.

The parameters of the stochastic models will generally be estimated by maximum-likelihood (ML) methods. The application of ML methods is often justified on the grounds that, under certain conditions, ML estimators are asymptotically efficient. In this investigation of exchange-rate dynamics it is obvious, however, that there is a substantial model uncertainty. Hence, I would simply refer to ML methods as a natural approach to estimate probability models.

In general, the estimation and test results will only be reported for daily and weekly data since only these data showed empirical properties that are worth exploring in a formal model. I estimated the models for monthly and quarterly data, too, but in many cases the estimates were rather poor or inconsistent with the model and, as expected, the null hypothesis of normality could often not be rejected.

2.1 FINITE MIXTURES OF NORMAL DISTRIBUTIONS

The model of finite mixtures of normal distributions has a long tradition. It has been introduced by Karl Pearson in 1894 (see Everitt and Hand (1981) or Titterington et al.

(1985)) but economists who have applied this model to financial data (e.g. Boothe and Glassman (1987)) seem to have been unaware of this literature. A finite mixture of J normal distributions is defined by

$$f(x \mid \psi) = \sum_{j=1}^{J} p_j f_j(x \mid \theta_j, \sigma_j) \tag{1}$$

where

$$f_j(x) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(x-\theta_j)^2}{2\sigma_j^2}\right\} \tag{2}$$

and ψ is the vector of parameters: $\psi = (p_1, \ldots, p_J, \theta_1, \ldots, \theta_J, \sigma_1, \ldots, \sigma_J)$. The p_j 's are assumed to be positive and to sum to 1. Hence they can be interpreted as probabilities. The model in (1) gives the density $f(x \mid \psi)$ as a weighted sum of normal densities f_j with different means θ_j and different variances σ_j^2. As formulated in (1), the model is, of course, too general to be directly applicable. In order to estimate the model, one has to specify J. I shall apply the principle of parsimony and set $J = 2$ initially. Since the analysis in the previous chapter has shown that the null hypothesis of a constant mean at zero cannot be rejected for the exchange-rate series except the yen-dollar rate, I shall also set $\theta_j = 0$ initially. Thus, the initial model can be written as

$$f(x \mid p, \sigma_1, \sigma_2) = \frac{p}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{x^2}{2\sigma_1^2}\right\} + \frac{1-p}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{x^2}{2\sigma_2^2}\right\}. \tag{3}$$

It is a scale mixture of two normal distributions and it incorporates stochastic variance in the simplest manner. Without loss of generality, one may assume that $\sigma_1 < \sigma_2$.

It should also be examined whether this model is compatible with the major stylized facts of exchange-rate dynamics. Stochastic variance is directly incorporated in the model but heteroskedasticity does not follow from this model since all parameters are assumed to be invariant over time. Therefore, it does not capture the clustering of small and of large exchange-rate fluctuations as is evident in the Markov transition matrix (see Chapter 1, figure 4 and table 4-5), in the ACF of the squared data, and in the Levene test. On the other hand, it is straightforward to prove leptokurtosis in the more general scale mixture of J normal distributions.[1] It is easy to show that

(11) $$\mu_2^2(X) = \left(\sum_{j=1}^{J} p_j \sigma_j^2 \right)^2$$

and that

(12) $$\mu_4(X) = 3 \sum_{j=1}^{J} p_j \sigma_j^4$$

where μ_i is the i-th central moment. Leptokurtosis is defined by the condition

(13) $$\beta_2 = \frac{\mu_4}{\mu_2^2} > 3$$

or, equivalently, by

(14) $$\kappa_4 = \mu_4 - 3\mu_2^2 > 0 \quad ,$$

where κ_4 is the fourth cumulant. Inserting (11) and (12) into (14) yields

(15) $$\kappa_4(X) = 3\left[\sum_{j-1}^{J} p_j \sigma_j^4 - \left(\sum_{j=1}^{J} p_j \sigma_j^2 \right)^2 \right]$$

(16) $$3\mathrm{Var}(\sigma^2) > 0$$

since σ^2 (which is now a stochastic parameter) is non-degenerate by assumption. Finally, the convergence to normality follows from the central limit theorem. As regards the more general model given in (1), in which the means θ_j can vary between the components of the mixture, no general result concerning kurtosis can be derived. Take, for simplicity, a mixture of two normal distributions with $\sigma_1 = \sigma_2 = 1$ and $\mu_1 = -\mu_2 = 1$. This mixture has a kurtosis of 1.25, i.e. it is platykurtic.

[1] Gridgeman (1970) proved only the peakedness of general scale-mixtures of normal distributions.

When Karl Pearson introduced the model of finite mixtures of normal distributions, he proposed to estimate the parameters of the model by the method of moments.[2] Since all odd central moments of the normal distribution are zero and the even central moments are given by

$$\mu_{2r} = \frac{(2r)!}{2^r r!}\sigma^{2r} \tag{17}$$

the 3 parameters of the model in (3) can be estimated by the 3 equations

$$p\sigma_1^2 + (1-p)\sigma_2^2 = m_2 \tag{18}$$

$$3p\sigma_1^4 + 3(1-p)\sigma_2^4 = m_4 \tag{19}$$

$$15p\sigma_1^6 + 15(1-p)\sigma_2^6 = m_6 \quad , \tag{20}$$

where m_i is the i-th empirical central moment. Defining

$$\xi_1 = \frac{k_6}{10k_4} + \left[\frac{k_6^2}{100k_4^2} + \frac{k_4}{3}\right]^{1/2} \tag{21}$$

and

$$\xi_2 = \frac{k_6}{100k_4} - \left[\frac{k_6^2}{10k_4^2} + \frac{k_4}{3}\right]^{1/2} \quad , \tag{22}$$

where k_i is the i-th empirical cumulant [3], explicit solutions for the moment estimators can be derived (see Cohen (1967))

$$\sigma_1^2 = \xi_1 + m_2 \tag{23}$$

$$\sigma_2^2 = \xi_2 + m_2 \tag{24}$$

$$p = -\xi_2/(\xi_1 - \xi_2) \quad . \tag{25}$$

[2] Ball and Torous (1983), obviously being unaware of the literature on mixture distributions, introduced this model under the name "Bernoulli mixture of Gaussian densities" and suggested to estimate the parameters by, what they called, the "method of cumulants" but which is identical with the moment estimators.

[3] The fourth and sixth empirical cumulants are given by $k_4 = m_4 - 3m_2^2$ and by $k_6 = m_6 - 15m_4m_2 - 10m_3^2 + 30m_2^3$, respectively.

Because of its desirable asymptotic properties, ML estimators of this model would be prefered to the moment estimators (see Everitt and Hand (1981) and Titterington et al. (1985)). However, the ML estimators cannot be derived in an explicit form and hence it is convenient to employ the moment estimators as starting values in an algorithm for the ML estimators. Everitt and Hand (1981) and Titterington et al. (1985) recommend the Newton-Ralphson algorithm [4] which requires the gradient vector

(26) $$\frac{\partial L}{\partial p_1} = \sum_{t=1}^{T} \frac{f_{1t} - f_{2t}}{f_t}$$

(27) $$\frac{\partial L}{\partial \sigma_j^2} = p_j \sum_{t=1}^{T} \frac{f_{jt} z_{jt}}{f_t} \qquad j = 1,2$$

(28) with $$z_{jt} = \frac{1}{f_{jt}} \frac{\partial f_{jt}}{\partial \sigma_j^2} = \frac{x_t^2}{2\sigma_j^4} - \frac{1}{2\sigma_j^2}$$

(where L denotes the log-likelihood function; f_{jt} is a short-hand notation for $f_j(x_t \mid \sigma_j)$) and the Hessian matrix whose elements are

(29) $$\frac{\partial^2 L}{\partial p_1^2} = -\sum_{t=1}^{T} \frac{(f_{1t} - f_{2t})^2}{f_t^2}$$

(30) $$\frac{\partial^2 L}{\partial (\sigma_j^2)^2} = p_j \sum_{t=1}^{T} \frac{f_{jt}}{f_t^2} [z_j^2 (f_t - p_j f_{jt}) + f_t y_{jt}] \qquad j = 1,2$$

(31) with $$y_{jt} = \frac{\partial z_{jt}}{\partial \sigma_j^2} = \frac{1}{2\sigma_j^4} - \frac{x_t^2}{\sigma_j^6}$$

(32) $$\frac{\partial^2 L}{\partial p_1 \partial \sigma_j^2} = -\sum_{t=1}^{T} \frac{f_{jt} z_{jt}}{f_t^2} [(-1)^j f_t - p_j (f_{1t} - f_{2t})] \qquad j = 1,2$$

(33) $$\frac{\partial^2 L}{\partial \sigma_1^2 \partial \sigma_2^2} = -p_1 p_2 \sum_{t=1}^{T} \frac{f_{1t} z_{1t} f_{2t} z_{2t}}{f_t^2}$$

[4] Tucker and Pond (1988) applied this algorithm to the estimation of normal mixtures for daily exchange rates. Akgiray and Booth (1988) applied the EM algorithm for the same kind of data, but experience shows that the EM algorithm is inferior to the Newton-Ralphson algorithm because the former converges very slowly (see Everitt (1984) and Kaehler (1988)).

A third estimation method was suggested by Quandt and Ramsey (1978) and is based on the minimization of the squared distance between the empirical and the theoretical moment generating function. The moment generating function of a normal distribution with a mean of zero is given by

$$E\exp\{ux\} = \exp\left\{\frac{1}{2}\sigma^2 u^2\right\} \tag{34}$$

where E is the expected-value operator and u is an auxiliary variable. The auxiliary variable introduces an element of arbitrariness into the analysis. Experience also showed that the Quandt-Ramsey method is inferior to methods based on the characteristic function $E\exp\{iux\}$, where $i = \sqrt{-1}$, and to ML estimators (see Everitt and Hand (1981) and Titterington et al. (1985)). Boothe and Glassman (1987), nevertheless, applied the Quandt-Ramsey method to the modelling of the distribution of exchange rates.

The parameter estimates of the scale mixture of two normal distributions are reported in table 1. Asymptotic standard errors, obtained from the Hessian matrix at the maximum, are given in brackets. It is clear from the entries of the table that there are quite marked differences in the variances of the two components. In general, the variances and the probability p can be estimated with quite high precision. However, the point estimates of p are somewhat in conflict with the economic interpretation of the model. It was argued above that the second component distribution with the higher variance should be associated with rare shocks and hence it would be expected that p is close to the value of 1. However, for two series (daily pound, weekly mark) the estimated p is even below 0.5 and the average of the 8 estimates is 0.65. The average of the 8 moment estimates of p is 0.88 which is more compatible with the given interpretation of the model.

Also reported in table 1 is the likelihood-ratio statistic (LR) for a test against the null hypothesis $H_0{:}f(x \mid \psi) = f(x \mid 0, \sigma)$ which serves as a benchmark model. The LR statistic has a χ^2 distribution with 1 degree of freedom (either from the restriction $\sigma_1 = \sigma_2$ or from $p = 1$ or $p = 0$). The LR test shows that for all series, the scale mixture of two normal distributions fits the data decisively better than a simple normal distribution. In addition,

Table 1

Estimates of scale mixtures of two normal distributions

		mark	pound	sfr	yen
day	σ_1^2	0.24 (0.02)	0.06 (0.01)	0.33 (0.02)	0.12 (0.01)
	σ_2^2	1.37 (0.19)	0.76 (0.04)	2.24 (0.26)	0.87 (0.07)
	p	0.80 (0.05)	0.46 (0.03)	0.82 (0.03)	0.65 (0.04)
	LR	370.2***	589.1***	503.4***	526.1***
	$\chi^2(96)$	167.8***	233.9***	153.0***	645.2***
week	σ_1^2	0.28 (0.16)	1.09 (0.22)	0.90 (0.20)	0.63 (0.16)
	σ_2^2	3.26 (0.43)	6.23 (1.96)	5.52 (0.80)	4.12 (0.04)
	p	0.37 (0.10)	0.81 (0.10)	0.55 (0.09)	0.71 (0.07)
	LR	78.8 ***	77.2***	67.1***	97.1***
	$\chi^2(46)$	49.0	79.7***	64.6**	64.6**

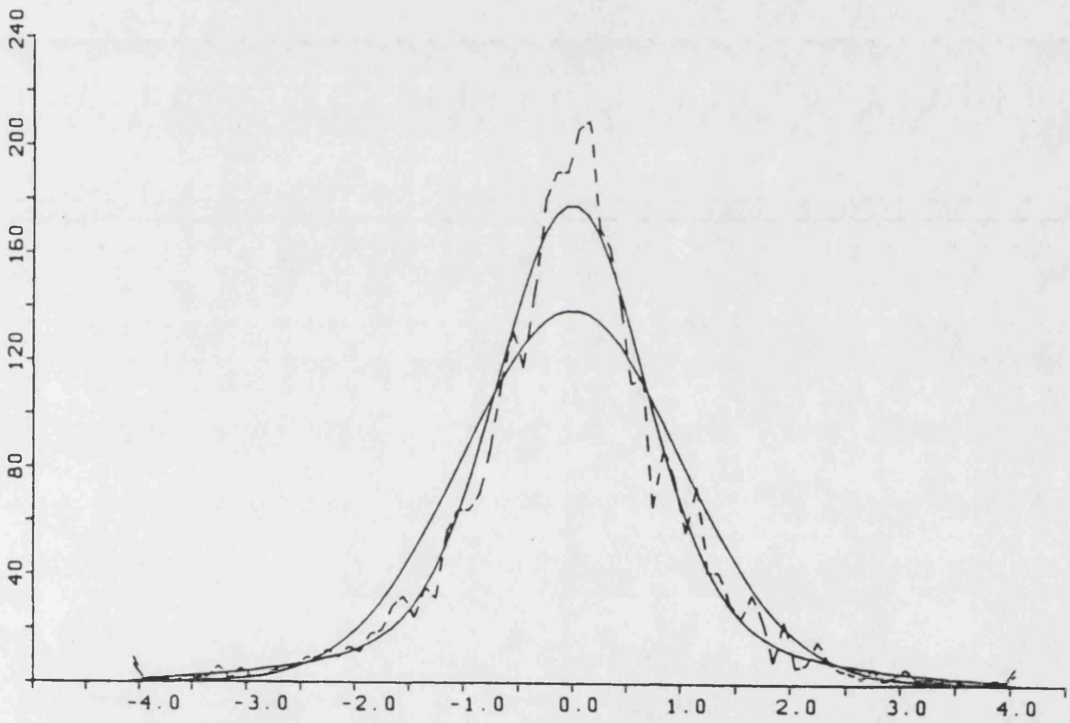
Significance levels: $\alpha = 0.01$ (***); $\alpha = 0.05$ (**); $\alpha = 0.10$ (*)

I applied a standard χ^2 goodness-of-fit test to see whether this model achieves a satisfactory fit to the data. As explained in the previous chapter, a rule of thumb for the optimal number of equi-probable quantiles for this test gives 100 quantiles for daily data and 50 quantiles for weekly data.Therefore, for daily data the χ^2 goodness-of-fit test has 96 degrees of freedom and for weekly data it has 46 degrees of freedom. The χ^2 test rejects this model quite strongly. For 5 series, the hypothesis that the data follow a 2-component scale mixture can be rejected at the 1 percent significance level and for another 2 series, H_0 can be rejected at the 5 percent level. Only for the weekly mark-dollar series is it not possible to reject H_0 at conventional significance levels.

In order to gain more insight into the results of estimating this model, figure 1 displays the empirical and theoretical frequencies for the daily mark-dollar rate. The empirical frequencies are given by the dashed line, the peaked solid line gives the frequencies of the scale mixture and the third line gives the frequencies of a normal distribution with parameters $\theta = 0$ and $\sigma = s$, where s is the empirical standard deviation. It is clear from this figure that the scale mixture is both more peaked and has fatter tails as compared with the simple normal distribution, i. e. the scale mixture is leptokurtic. Since the empirical distribution is also highly leptokurtic, the scale mixture gives a much better fit to the data than the normal distribution. However, the scale mixture is obviously not able to capture the full amount of peakedness in the data.

Figure 1

Empirical frequencies and scale mixture: daily mark-dollar rate



The fact that, according to the χ^2 goodness-of-fit test, a convincing fit to the data cannot be achieved for 7 of the 8 series, suggests that the more general model of J mixtures of normal distributions with different variances and means, as given in (1), should be employed. I identified the dimension of the model with respect to the parameter vector ψ by SIC. It turned out that in 3 cases (daily mark, weekly pound and sfr), a 2-component scale mixture,as estimated before, was optimal, in 3 cases (daily pound and sfr, weekly mark) a 3-component scale mixture was optimal. For the yen series, mixtures of normal distributions with different means and variances were identified as optimal. The daily series had 3 components and the weekly series had 2 components. The details are given in table 2.

Table 2

Estimates of optimal mixtures of normal distributions

	mark	pound	sfr	yen
day	$\sigma_1^2 = 0.24(0.02)$ $\sigma_2^2 = 1.37(0.19)$ $p = 0.80(0.05)$ LR = 370.2*** χ^2= 167.8***	$\sigma_1^2 = 0.03(0.00)$ $\sigma_2^2 = 0.43(0.02)$ $\sigma_3^2 = 2.50(0.28)$ $p_1 = 0.27(0.00)$ $p_2 = 0.67(0.00)$ LR = 710.8*** χ^2 = 174.3***	$\sigma_1^2 = 0.07(0.01)$ $\sigma_2^2 = 0.56(0.02)$ $\sigma_3^2 = 3.69(0.41)$ $p_1 = 0.23(0.00)$ $p_2 = 0.70(0.01)$ LR = 563.8*** χ^2 = 95.0	$\theta_1 = 0.014(0.009)$ $\theta_2 = 0.027(0.013)$ $\theta_3 = -0.239(0.064)$ $\sigma_1^2 = 0.01(0.00)$ $\sigma_2^2 = 0.22(0.02)$ $\sigma_3^2 = 1.16(0.13)$ $p_1 = 0.14(0.02)$ $p_2 = 0.68(0.03)$ LR = 664.7*** χ^2 = 717.6***
week	$\sigma_1^2 = 0.13(0.03)$ $\sigma_2^2 = 1.77(0.16)$ $\sigma_3^2 = 7.37(1.25)$ $p_1 = 0.21(0.00)$ $p_2 = 0.66(0.01)$ LR = 97.4*** χ^2 = 38.0	$\sigma_1^2 = 1.09(0.22)$ $\sigma_2^2 = 6.23(1.96)$ $p = 0.81(0.10)$ LR = 77.2*** χ^2 = 79.7***	$\sigma_1^2 = 0.90(0.20)$ $\sigma_2^2 = 5.52(0.80)$ $p = 0.55(0.09)$ LR = 67.1*** χ^2 = 64.6**	$\theta_1 = 0.103(0.053)$ $\theta_2 = -0.585(0.189)$ $\sigma_1^2 = 0.58(0.12)$ $\sigma_2^2 = 3.54(0.58)$ $p = 0.68(0.08)$ LR = 107.0*** χ^2 = 41.6

Significance levels: See table 1

In the 3-component scale mixtures, the components are well separated, i.e. the variances are clearly distinct. Also, the probability of drawing from the high-variance (σ_3^2) component is, as expected, quite low. This probability is 6 percent for the daily pound series, 7 percent for the daily sfr series and 13 percent for the weekly mark series. For 3 exchange rates (pound, sfr, yen), the dimension of the parameter vector ψ is smaller for the weekly data than for the daily data. However, for the mark-dollar rate, the dimension of the parameter vector increases quite surprisingly under time aggregation. As in table 1, there is no apparent pattern for the probability p of drawing from the low-variance component in the 2-component models. For the 3-component models, however, there is some similarity in the results. They all have p_1 around 20 percent and p_2 at roughly 70 percent.

Table 2 also reports results from an application of LR tests with the H_0 of white noise. It should be noted that there are some technical problems with applications of LR test, and also with the application of ML methods in general, to these mixture models. First, the likelihood function is unbounded when $\sigma_1 \to 0$. If both means and variances were allowed to vary between components, this problem of singularities would be even more severe (see Titterington et al. (1985), pp. 93-93). Hamilton (1991a) suggested to apply Bayesian methods as a remedy against this problem.

Second, the regularity conditions are violated for the application of the LR test since p_j is on the boundary of the parameter space under H_0. However, according to Everitt (1981), this only leads to a downward biased size of the test statistic in multivariate cases and when the sample size is less than 100. Third, it is sometimes unclear how the degrees of freedom are to be determined. If, for instance, the alternative is a two-component mean-variance mixture, then one can either obtain Gaussian white noise with the restriction $p = 1$ (giving one degree of freedom) or with the restrictions $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$ (giving two degrees of freedom). A pragmatic solution in this case would be to choose a conservative test with two degrees of freedom. Finally, there is the closely related problem of nuisance parameters which are not identified under H_0 (an equivalent way to pose the problem is to say that the score is identically zero under H_0). For instance, under the H_0 of $\sigma_1 = \sigma_2$, the value of p is not identified in a two-component scale mixture. Hansen (1992) discussed this problem more generally and proposed to derive bounds of the asymptotic distribution of the LR statistic through the covariance function of the empirical likelihood surface. He also applied this approach to Markov-switching models which are a generalization of mixture models and which will be introduced in Chapter 3.

For the applications of the LR test as reported in tables 1 and 2, these problems are certainly inconsequential. The lowest LR statistic in these tables is 67.1 and it would make no difference whether one would choose 1 or 2 degrees of freedom for the χ^2 distribution. There is also no reason to suppose that the distribution of the LR statistic is so drastically different from a χ^2 distribution as to invalidate the statistical inferences.

As regards goodness-of-fit, there is some improvement in the generalized normal mixtures of table 2 as compared with the scale mixtures of 2 normal distributions of table 1. At the 10 percent significance level, the H_0 that the data come from the specified normal mixtures cannot be rejected for 3 series (daily sfr, weekly mark and yen). However, the H_0 must be rejected at the 1 percent significance level for 4 series and for another series H_0 must be rejected at the 5 percent level. Therefore, the mixture of normal distribution does not seem to be a very satisfactory model of exchange-rate dynamics. The model does capture leptokurtosis to a certain degree (see also Chapter 4) but overall the fit to the data is not convincing. One reason for the lack of fit may be due to the fact that the normal mixture is not capable of capturing more complex models of stochastic variance in a parsimonious way. Viewed slightly differently, the scale mixture of normal distributions is a model of a random variable with a conditional normal distribution whose parameter σ^2 has itself a probability distribution. Therefore, this model belongs to the family of "compound" distributions [5] and it can be written as

(35) $$f(x) = \int_{\Re^+} f(x \mid \sigma^2) dF(\sigma^2) \quad ,$$

where $F(\sigma^2)$ has the step function

[5] See Douglas (1980, pp. 21-22, 75-76) on the various, and sometimes confusing, uses of terminology for these kinds of models.

(36) $$F(\sigma^2) = \begin{cases} 0 & \text{for} \quad 0 < \sigma^2 < \sigma_1^2 \\ p_1 & \text{for} \quad \sigma_1^2 \le \sigma^2 < \sigma_2^2 \\ 1 & \text{for} \quad \sigma_2^2 \le \sigma^2 \end{cases}$$

This form of the model suggests to look for more complex but parsimonious functions $F(\sigma^2)$. Note that every additional value in the domain of $F(\sigma^2)$, i.e. every increase in the number of components of the mixture, increase the numbers of parameters to be estimated by 2. Hence, it is an obvious strategy to find suitable, but parsimonious, probability distributions for σ^2 in order to model stochastic variance. In the next three sections, I will introduce both a discrete distribution and two continuous distributions for σ^2 which also allow to cast the models into a more general framework.

2.2 COMPOUND POISSON PROCESS AND RANDOM SUMS

There are two ways in which the compound Poisson process can be introduced as a model of exchange-rate dynamics. First, the compound Poisson process can be derived in the form of a compound distribution where a parameter of the "compounded" distribution is random and has itself a Poisson distribution. Second, it can be derived in a very general way as a model of a sum of random variables where the number of summands is random and has itself a Poisson distribution. Both ways of deriving the compound Poisson process will be described, but first it is instructive to recall from the basic theory of stochastic processes that the Poisson process is a very fundamental and general form of stochastic processes. I will briefly sketch the derivation of the Poisson process (see e.g. Cox and Miller (1965), pp. 146-153) because as a by-product one obtains the formula for the generating function of a Poisson distribution which is useful later in this section.

Let $N(s,t)$ be the number of occurrences of a random event in the time interval (s,t). $N(s,t)$ is a Poisson process if

i) $N(s,s)=0$,

ii) $N(s,t)$ has stationary and independent increments,

iii) $P[N(t,t+\Delta t)\geq 2]=o(\Delta t)$,

iv) $P[N(t,t+\Delta t)=1]=\lambda\Delta t+o(\Delta t)$.

It follows from the assumption of stationary increments that one may set $s=0$ without loss of generality. Conditions (iii) and (iv) state that the probability of having more than one occurrence in a small interval Δt is negligible (o denotes the order of magnitude) and that the probability of the occurrence of exactly one event is approximately $\lambda\Delta t$. Therefore,

$$P[N(t,t+\Delta t)=0]=1-\lambda\Delta t+o(\Delta t). \tag{37}$$

In order to simplify notation, define

$$P_n(t+\Delta t)=P[N(0,t+\Delta t)=n]. \tag{38}$$

Then one gets from the assumption of independence

$$P_n(t+\Delta t)=P_n(t)(1-\lambda\Delta t)+P_{n-1}(t)\lambda\Delta t+o(\Delta t). \tag{39}$$

It follows from (39) that

$$\frac{d}{dt}P_n(t)=-\lambda P_n(t)+\lambda P_{n-1}(t). \tag{40}$$

A solution to the above ordinary differential equation can be obtain from applying the probability generating function

$$G(z,t)=\sum_{n=0}^{\infty}P_n(t)z^n, \tag{41}$$

where z is an auxiliary variable, to (40) which gives

$$\frac{\partial}{\partial t}G(z,t)=-\lambda G(z,t)+\lambda zG(z,t). \tag{42}$$

A solution to the differential equation (42) is given by

$$G(z,t)=A(z)\exp\{-\lambda t+\lambda tz\}. \tag{43}$$

Condition (i) is a natural assumption, given that time is a continuous variable in this model. This implies that $P_0(0) = 1$ and that $P_n(0) = 0$ for $n = 1, 2, \ldots$. It follows from (41) that $G(z, 0) = 1$ and from (43) that $A(z) = 1$. Thus,

(44) $$G(z,t) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} z^n$$

and

(45) $$P_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!},$$

which is the familiar probability function of the Poisson distribution. The compound Poisson process is now given by the model

(46) $$X_t = Y_{1,t} + Y_{2,t} + \ldots + Y_{N(t),t}$$

where the $Y_{j,t} (j = 1, \ldots, N)$ are independent and identically distributed. Furthermore, $N(t)$ is random and has a Poisson distribution given in (45). Note that the domain of the random variable N are the integers $N \geq 1$ since $N = 0$ would leave X_t in (46) undefined. If, however, N has a Poisson distribution, as assumed, there is a conflict in domains since the domain of the Poisson distribution includes the value of zero. This problem is usually not addressed in the literature and many authors seem not to be aware of it. Thus, I will digress here to discuss alternative ways to avoid the problem of conflicting domains. A more elaborate treatment of this issue can be found in Kaehler (1990b).

The easiest way to augment the model of the compound Poisson process is to redefine (46) as (I shall drop the time index t in this digression in order to simplify notation):

(46') $$X = \begin{cases} Y_1 + Y_2 + \ldots + Y_N & \text{if } n \geq 1 \\ 0 & \text{if } n = 0 \end{cases}$$

and to let N have the Poisson distribution given in (45). The following economic interpretation can be ascribed to this model. Let the foreign-exchange market be subject to random shocks which are independent and identically distributed. In a small time period, the occurrence of a shock is a rare event in the sense that the probability of more than one shock is negligible. Further, let the total time period for which exchange-rate movements are recorded (say, a day) be fixed. Under these assumptions, the number of shocks hitting the market, and causing exchange-rate movements Y_j, is random with a Poisson distribution. It is then also natural to assume, as in (46'), that in the absence of shocks the exchange rate remains constant.

Next, I shall examine whether this model is compatible with leptokurtosis. It follows from (46') that

(47) $$\mathrm{Var}(X) = n\sigma_Y^2$$

where σ_Y^2 denotes the common variance of the Y_j in (46'). Since n is random, X has a compound distribution with a stochastic variance. The distribution function of the variance, $F(\sigma_X^2)$, is now

(48) $$F(n\sigma_Y^2) = e^{-\lambda} \sum_{i=0}^{n} \frac{\lambda^i}{i!} \quad \text{with} \quad n = 0, 1, 2, \ldots.$$

Leptokurtosis of the model can be established by applying the characteristic function which is defined by

(49) $$\Phi_X(u) = E(\exp\{iux\})$$

where $i = \sqrt{-1}$. The characteristic function is easily derived to be

(50) $$\begin{aligned}\Phi_X(u) &= E_N(\Phi_{X|N}(u)) \\ &= E_N(\Phi_Y^n(u)) \\ &= \sum_{n=0}^{\infty} \Phi_Y^n(u) e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \exp\{-\lambda[1 - \Phi_Y(u)]\}\end{aligned}$$

where Φ_Y is the characteristic function of Y_j. Note that $\Phi_Y^o(u) = E(e^0) = 1$. It is now more convenient to proceed with the log-characteristic function (also called cumulant-generating function)

(51) $$\log \Phi_X(u) = -\lambda[1 - \phi_Y(u)].$$

The m-th cumulant is defined by

(52) $$\kappa_m(X) = \frac{1}{i^m}\frac{d^m}{du^m}\log\Phi_X(u)\bigg|_{u=0}.$$

From (51) and (52) one obtains

(53) $$\kappa_m(X) = \lambda\mu_m^{'}(Y)$$

where $\mu_m^{'}(Y)$ is the m -th moment about zero of Y_j. Thus,

(54) $$\kappa_4(X) = \lambda\mu_4^{'}(Y) > 0,$$

and according to (14), leptokurtosis obtains. This is a strong result because relatively mild restrictions have been put on the Y_j. In addition to the assumption that the Y_j are independent and identically distributed, one only has to assume that $\mu_4^{'}(Y)$ exists.[6]

Press (1967, 1968) introduced a simple variant of the basic model (45) and (46). He simply added a normal variable to the right-hand side of (46) to get

(46'') $$X = Y_1 + Y_2 + \ldots + Y_n + V \quad n = 0, 1, \ldots$$

where V follows a normal distribution with constant mean θ and variance σ_V^2. Press also used the restriction $E(V) = 0$ but I will not need this restriction here. In addition, V is assumed to be independent of the Y_j. The probability distribution of the random parameter N is given by (45) as in the basic model. In economic terms, V represents background noise in the foreign-exchange market. Without any news Y_j hitting the market, exchange-rate dynamics are determined by normal (in a double sense) fluctuations V of demand and supply. One would conjecture, therefore, that $\sigma_V^2 \ll \sigma_Y^2$.

The stochastic variance of this variant of the basis model is obvious. The distribution function $F(\sigma_X^2)$ is now

(48') $$F(n\sigma_Y^2 + \sigma_V^2) = e^{-\lambda}\sum_{i=0}^{n}\frac{\lambda^i}{i!} \quad \text{with} \quad n = 0, 1, \ldots.$$

[6] Zimmermann (1985) obtained a similar result but he had to assume that the Y_j have moments of all orders. An alternative proof, applying conditional moments, is provided by Kaehler (1992). The formulae for conditional central moments, on which this proof is based, are derived in Kaehler (1990a).

Leptokurtosis can be proved via the characteristic function:

(55)
$$\Phi_X(u)=\left[\sum_{n=0}^{\infty}\phi_{X\mid N}(u)e^{-\lambda}\frac{\lambda^n}{n!}\right]\exp\left\{i\theta u-\frac{1}{2}\sigma_V^2u^2\right\}$$
$$=\exp\left\{-\lambda[1-\Phi_Y(u)]+i\theta u-\frac{1}{2}\sigma_V^2u^2\right\}$$

Differentiating the log-characteristic function four times, one obtains at $u=0$ the same result as in (54):

(54')
$$\kappa_4=\lambda\mu_4'(Y)>0.$$

Thus, one gets the same strong result concerning leptokurtosis as in the basic model. Note that Press (1968) proved leptokurtosis only for the special case where the Y_j follow a normal distribution.

The third variant of the basic model adopts the version of the random sum as given in (46) but assumes a "shifted" Poisson distribution for N, i.e.

(45')
$$P_n=\frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda}\quad n=1,2,\ldots\quad .$$

The model of (46) and (45') assumes that there is at least one realization of the Y_j variables. It is quite reasonable to assume that foreign-exchange markets are subject to certain recurring news which occur in every time span, say a business day. Think of it, for instance, as the closing price of the Tokio market when the London market opens. In contrast to the Press model, this model assumes that there is only one kind of news relevant to the market.

The stochastic variance of the variant is, again, obvious. It's distribution function for the variance of X is

(48'')
$$F(n\sigma_y^2)=e^{-\lambda}\sum_{i=1}^{n}\frac{\lambda^{i-1}}{(i-1)!}\qquad \text{with}\quad n=1,2,\ldots\quad .$$

As regards kurtosis, the characteristic function of this model is

(56) $$\Phi_X(n) = \sum_{n=1}^{n} \Phi_{X|N}(n) e^{-\lambda} \frac{\lambda^{n-1}}{(n-1)!}$$
$$= \Phi_Y(u) \exp\{-\lambda[1 - \Phi_Y(u)\} \quad .$$

From the log-characteristic function one obtains

(57) $$\kappa_4(X) = \kappa_4(Y) + \lambda\mu_4'(Y) \quad .$$

Thus the condition for leptokurtosis in this version is $\kappa_4(Y) > -\lambda\mu_4'(Y)$. If it is assumed that the Y_j have a normal distribution, as in the model by Press, then this version implies leptokurtosis since $\kappa_4 = 0$ for a normal distribution. Alternatively, if it is assumed that $E(Y_j) = 0$, then it can be shown that $\lambda > 2$ is a sufficient condition for leptokurtosis.

From an economic point of view, the Press model is probably the most attractive of the three variants of the compound Poisson process because it allows to distinguish between random shocks and gradual movements in tranquil periods. The first variant ((45) and (46')) and the third variant (the shifted Poisson distribution (45') and (46)) are very similar and they differ only in the probabilities they attach to $n\sigma_Y$. These three variants do not, of course, exhaust all possible variations of the compound Poisson process. One could, for instance, apply the summation of random variables in (46) (for n = 1, 2...) and add a zero-trancated Poisson distribution for N. Similar to the model with the shifted Poisson distribution, this variant exhibits leptokurtosis only under certain parameter values of λ and $\beta_2(Y)$ (see Kaehler (1990b) for more details).

I applied variants 1 and 2 to the exchange-rate data. It turned out that variant 2 (the Press model) is clearly superior to variant 1 (the basic model) in terms of goodness-of-fit. Hence, only the results from applying the Press model are reported. The model was estimated by maximizing the log-likelihood function based on the normal distribution for Y_j and V:

(58) $$L = \sum_{t=1}^{T} \log \sum_{n=0}^{\infty} \frac{1}{[2\pi(\sigma_V^2 + n\sigma_Y^2)]^{1/2}} \exp\left\{-\frac{x_t^2}{2(\sigma_V^2 + n\sigma_Y^2)}\right\} \frac{\lambda^n}{n!} e^{-\lambda} \quad .$$

The infinite sum in (58) is, of course, awkward and must be truncated at some point. Ball and Torous (1985) showed that an upper bound for the truncation error is

(59) $$B(n) \leq (2\pi\sigma_V^2)^{-1/2} \frac{\lambda^{n+1}}{(n+1)!} \quad .$$

Imposing the restriction that $B(n) \leq (prec)^{1/2}$, where $prec$ is the machine precision, one can endogenously determine the truncation point n^* as a function of λ from

(60) $$\lambda \leq \exp\left\{\frac{1}{n^*+1}[\log(n^*+1)! + \log c]\right\}$$

where $c = (2\pi\sigma_Y^2 \cdot prec)^{1/2}$ and the machine precision on the Siemens mainframe 7570CX was $0.22204 \cdot 10^{-15}$. Application of this rule let to values of n^* between 8 and 13 for daily series and values of n^* between 9 and 12 for weekly series at the final values of the estimated parameter $\hat{\lambda}$.

A further complication arises from the fact that one cannot explictly derive the ML estimators of the three parameters from (58). This requires to apply an iterative procedure and to supply starting values. It is not difficult to derive first and second derivatives from (58) for the application of gradient methods, but here I used numerical derivatives. Starting values can be obtained from the method of moment estimators, proposed by Press (1967, 1968) for estimation. Equating empirical and theoretical cumulants yields

(61) $$k_2(X) = \lambda_Y^2 + \sigma_V^2$$

(62) $$k_4(X) = 3\lambda\sigma_Y^4$$

(63) $$k_6(X) = 15\lambda\sigma_Y^6 \quad .$$

This system of 3 equations in 3 unknowns can be easily solved for the moment estimators

(64) $$\tilde{\lambda} = 25k_4^3/3k_6^2$$

(65) $$\tilde{\sigma}_Y^2 = k_6/5k_4$$

(66) $$\tilde{\sigma}_V^2 = k_2 - 5k_4^2/3k_6 \quad .$$

Press (1967) reported that in his applications of the moment estimator to stock-price data, some of the variance estimates turned out to be negative. Using (64)-(66) for the daily and weekly exchange-rate series, I only obtained a negative estimate for σ_V^2 in the weekly sfr series and used instead $\lambda = 1$ and $\sigma_Y^2 = \sigma_V^2 = \sigma_X^2/2$ as starting values . For all series, the moment estimates $\tilde{\lambda}$ were very much smaller than the ML estimates $\hat{\lambda}$. The values of $\tilde{\lambda}$ ranged between 0.042 and 0.061 for the four daily series and between 0.109 and 0.413 for the weekly series. As a consequence, the moment estimates of the variances, $\tilde{\sigma}_Y^2$ and $\tilde{\sigma}_V^2$, are much larger than their ML counterparts. This means that from the moment estimates one would conclude that shocks to the foreign-exchange market are quite rare but strong and that the background noise has rather great variability.

The ML estimates are reported in table 3. According to the estimates of λ , the mean number of shocks is around 1 for both daily series and weekly series. The sfr-dollar rate seems to be subject to fewer shocks than the other three series but on the other hand, it seems to have higher background variance than the other series. The fact that for all 8 data sets, $\hat{\sigma}_Y^2$ is much greater than $\hat{\sigma}_V^2$ fits very well into the economic interpretation given for this model. For some series, the variance of the shock variables Y_j is greater than the variance of the background noise by more than a factor of 10 and it is at least greater by a factor of 3. The asymptotic standard errors of the point estimates are given in brackets. Only the precision of $\hat{\sigma}_V^2$ in the weekly data is somewhat unsatisfactory.

Also reported in table 3 are the values of a likelihood-ratio test against the H_0 of a simple normal distribution with mean zero which serves as a benchmark for all stochastic models introduced in this chapter. H_0 is overwhelmingly rejected for all series. Note, too, that the LR statistic for the compound Poisson process is higher than that for the 2-component normal mixture for all series except the weekly sfr series. Both models have 3 parameters to be estimated.

Finally, table 3 shows the values of a χ^2 goodness-of-fit. For this model, it turned out to be much easier and computationally much cheaper to determine the quantiles for

Table 3

Estimates of the compound Poisson process

		mark	pound	sfr	yen
day	λ	1.025 (0.239)	1.262 (0.103)	0.442 (0.140)	1.592 (0.111)
	σ_Y^2	0.345 (0.059)	0.307 (0.025)	0.941 (0.221)	0.216 (0.016)
	σ_V^2	0.093 (0.027)	0.027 (0.005)	0.236 (0.043)	0.014 (0.003)
	LR	403.0***	681.6***	516.8***	601.4***
	χ^2	129.7**	130.6***	146.5***	146.0***
week	λ	1.470 (0.217)	0.796 (0.444)	0.818 (0.322)	1.212 (0.404)
	σ_Y^2	1.333 (0.204)	1.816 (0.735)	2.728 (0.850)	1.149 (0.323)
	σ_V^2	0.143 (0.047)	0.545 (0.270)	0.731 (0.267)	0.187 (0.116)
	LR	92.2***	81.9***	64.5***	98.3***
	χ^2	34.1	76.2***	61.4*	53.2

Significance levels: see table 1

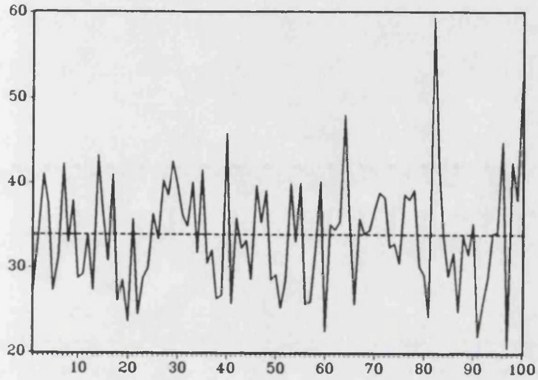
the empirical distribution first and to compare in every quantile the constant empirical frequencies with the corresponding theoretical frequencies. An additional complication was caused by the fact that the daily pound and yen series contain many values of zero (43 zeros in the pound series and 146 zeros in the yen series). Setting the number of quantiles for daily series at 100 (as in Chapter 1 and in Section 2.1), would have implied theoretical frequencies of zero for some quantiles. Therefore, the number of quantiles was reduced to 80 for the daily pound series and to 23 for the daily yen series. The degrees of freedom in the χ^2 test for these daily series are, therefore, 76 and 19. For all weekly series, 50 quantiles were used.

The χ^2 goodness-of-fit test rejects the H_0 that the data were generated by a compound Poisson process quite strongly for daily data. For the mark-dollar series, H_0 is rejected at the 5 percent level of significance and for the other three daily series, H_0 is rejected at the 1 percent level. However, for weekly series the compound Poisson process achieves a much better fit than for daily series. The χ^2 test only rejects the H_0 for the pound-dollar series at the 1 percent level. The χ^2 -value of 61.4 for the weekly sfr series is significant at the 10 percent level.

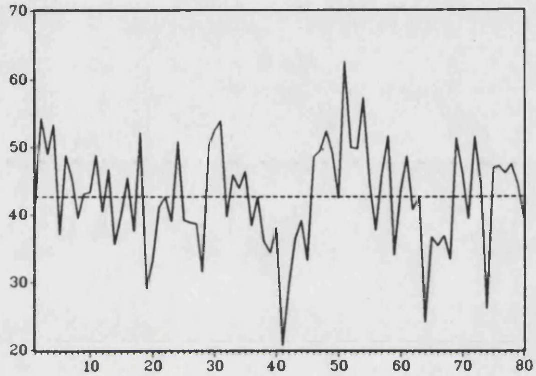
Figure 2

Compound Poisson process: empirical and theoretical frequencies for daily data

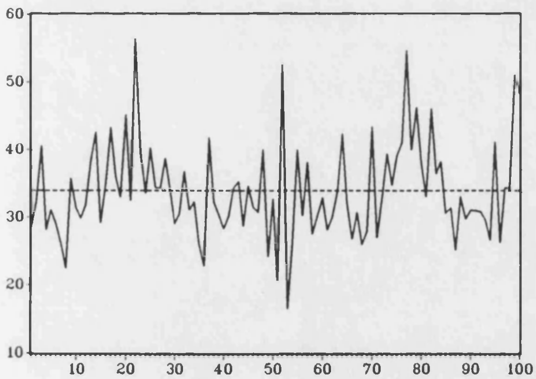
a) mark



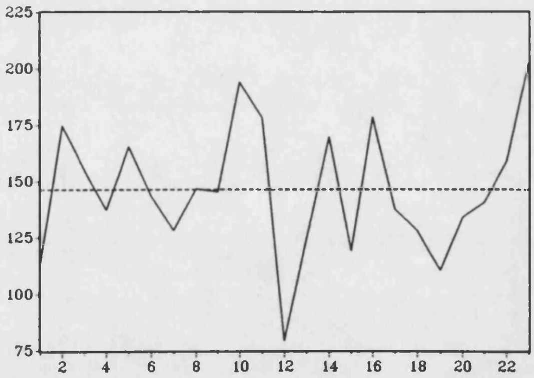
b) pound



c) sfr



d) yen



In order to better understand why the χ^2 test rejects the model for daily exchange-rate data, figure 2 displays the empirical frequencies (dashed line) and theoretical frequencies (solid line) for the four daily series.

For the pound series and the yen series it is quite obvious from figure 2 that the compound Poisson process cannot capture the full amount of peakedness in the data. The empirical distribution of daily sfr-dollar exchange-rate fluctuations, on the other hand, seems to have less mass in the shoulders than predicted by the model. No clear pattern of discrepancies between empirical and theoretical frequencies appear in the plot for the mark series.

Compared with the scale mixture of two normal distributions in terms of likelihood ratios and in terms of goodness-of-fit, the compound Poisson process is superior to the former. The quite strong rejection by the goodness-of-fit test for the daily series, however, casts some doubts on the appropriateness of the compound Poisson process. From a theoretical point of view, the model is very attractive because of its generality as a random sum of random variables. The alternative way to view this model is to regard it as a compound normal distribution where the variance follows a Poisson distribution. The fact that this model can be viewed in these two different ways is a direct consequence of Gurland's theorem. Gurland (1957) proved in a very general setting the equivalence in distribution between random convolutions (like the Poisson sum) and compound distributions.In the next section, I will introduce a model which can also be derived as a compounded normal distribution but with a continuous distribution for the variance.

2.3 STUDENT's DISTRIBUTION AND THE PEARSON FAMILY

The Student distribution, also called t -distribution, has been introduced by Praetz (1972) into the modelling of financial data. He derived a (one parameter) t -distribution as a compound normal distribution where the variance follows an inverted gamma distribution.

Thus, the t -distribution can be directly related to the mixture of normal distributions and to the compound Poisson process which can also be viewed as scale-compounded normal distributions. The inverted gamma distribution, however, is a continuous distribution.

There are, of course, numerous candidate distributions which could serve as a continuous compounding distribution for the variance of normal distribution. Clark (1973), for instance, suggested a lognormal distribution for the variance to model speculative prices. In order to restrict arbitrariness in the choice of the compounding distribution, one should be able to motivate the choice in some way. For the inverted gamma distribution this motivation is possible along two lines. First, in Bayesian statistics the natural conjugate for the distribution of the variance of a normal distribution is the inverted gamma distribution (and for the precision it is the gamma distribution, see e.g. Raiffa and Schlaifer (1961), p. 291). Second, the fact that the adoption of the inverted gamma distribution leads to the t -distribution allows one to put the compound distribution into a broad framework of an important family of distributions, namely the Pearson family of distributions.

In order to relate the t -distribution to the Pearson family, I shall slightly generalize the model adopted by Praetz and derive a 2-parameter (or generalized) t -distribution. Assume, first, that the variable X (i.e. the exchange-rate dynamics) follows a normal distribution conditional on h:

(67) $$f_N(x \mid 0, h\sigma^2) = \frac{1}{\sqrt{2\pi}(h\sigma^2)^{1/2}} \exp\left\{-\frac{x^2}{2h\sigma^2}\right\}$$

where $h > 0$ is a random factor of the variance. Assume further that h follows an inverted gamma distribution, i.e.

(68) $$f_{I\gamma}(h \mid \nu) = \frac{e^{-1/h} h^{-\nu-1}}{\Gamma(\nu)} \quad \text{with} \quad \nu > 0$$

where $\Gamma(\nu)$ is the gamma function defined by

(69) $$\Gamma(\nu) = \int_0^\infty e^{-u} u^{\nu-1} du.$$

The marginal distribution of X is now given by

(70)
$$\int_0^{\infty} f_N(x \mid 0, h\sigma^2) f_{I\gamma}(h \mid \nu) dh$$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi}(h\sigma^2)^{1/2}} \exp\left\{-\frac{x^2}{2h\sigma^2}\right\} \frac{e^{-1/h} h^{-\nu-1}}{\Gamma(\nu)} dh$$

$$= \frac{1}{\Gamma\left(\frac{1}{2}\right)\Gamma(\nu)} (2\sigma^2)^{-1/2} \int_0^{\infty} h^{-(\nu+3/2)} \exp\left\{-\frac{1}{h}\left(1+\frac{x^2}{2\sigma^2}\right)\right\} dh$$

where $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. By the change of variables $h = (1 + x^2/2\sigma^2)y$ one gets

(71)
$$f(x \mid \sigma^2, \nu) = \frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\Gamma(\nu)\Gamma\left(\frac{1}{2}\right)} (2\sigma^2)^{-1/2} \left(1+\frac{x^2}{2\sigma^2}\right)^{-\nu-\frac{1}{2}} \int_0^{\infty} \frac{\left(\frac{1}{y}\right)^{\nu+\frac{3}{2}} e^{-\frac{1}{y}}}{\Gamma\left(\nu+\frac{1}{2}\right)} dy.$$

It is obvious from (68) that the integral in (71) is 1. Hence one obtains

(72)
$$f(x \mid \sigma^2, \nu) = \frac{1}{B\left(\frac{1}{2}, \nu\right)} (2\sigma^2)^{-1/2} \left(1+\frac{x^2}{2\sigma^2}\right)^{-\nu-\frac{1}{2}}$$

where $B\left(\frac{1}{2}, \nu\right)$ is the beta function defined by $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$.

In a slightly different parametrization of (72) one gets

(73)
$$f(x \mid \eta, \gamma) = \frac{1}{\gamma B\left(\frac{1}{2}, \eta-\frac{1}{2}\right)} \left(1+\frac{x^2}{\gamma^2}\right)^{-\eta}$$

where $\eta = \nu + \frac{1}{2}$ and $\gamma = \sqrt{2}\,\sigma$. Equation (72) is the form in which the density of the Pearson type VII distribution is given. I will come back to this distribution later in this section. The one-parameter t-distribution obtains if $\eta = (1+\gamma^2)/2$ where the degrees-of-freedom parameter is given by $2\eta - 1$.

From its derivation as a compound normal distribution it is evident that this model implies stochastic variance under the assumption that the underlying distribution is normal.

Since the generalized t -distribution is also leptokurtic (see Elderton and Johnson (1969), p. 45), this model captures an important stylized facts of exchange-rate dynamics. Blattberg and Gonedes (1974) showed by numerical examples that the generalized t -distribution is more peaked than the normal distribution if both are standardized such that their variances are unity.

Recall from chapter 1 that the empirical values of the Pearsonian pseudovariance indicated that the distribution of exchange-rate dynamics might fall within the Pearson family of frequency curves. As stated above, the generalized t -distribution is a member of the Pearson family which is characterized by the differential equation for the density function

(74) $$\frac{d}{dx}f(x) = -\frac{(x-a)f(x)}{c_0 + c_1 x + c_2 x^2},$$

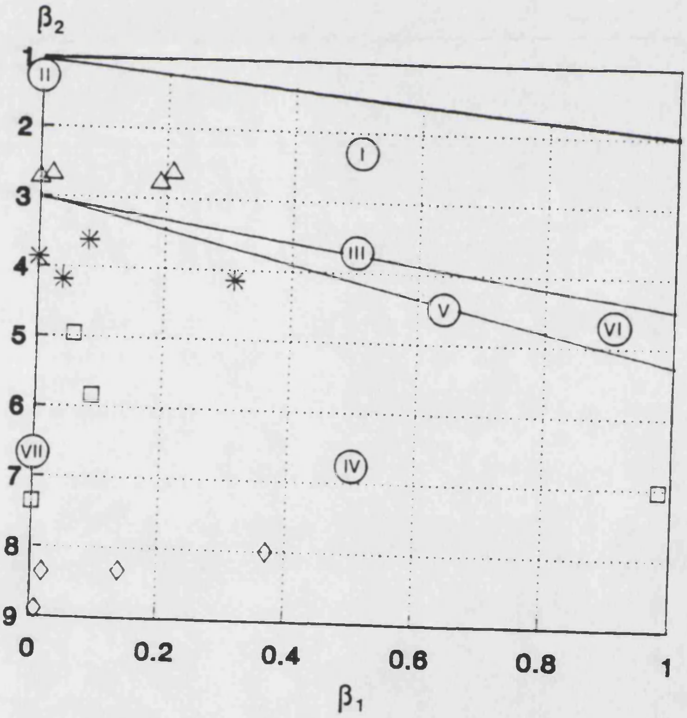
where a, c_0, c_1 and c_2 are constants. The equation might be criticized because of its ad hoc nature but it is related to a continuous-time, continuous-state birth-and-death process and Pearson derived it with a limiting argument from a difference equation satisfied by the hypergeometric distribution (see Ord (1985)). This family encompasses many important distributions, like the normal, the gamma, the beta, the uniform, the exponential and the F-distribution. A great variety of distributional shapes can be found in this family and the distributions are divided into 7 major types according to the parameter values in (74). The generalized t -distribution is the so called type VII distribution and obtains for $a = c_1 = 0, c_0 > 0, c_2 > 0$.

There is an alternative way to classify the types of distribution within this family according to the squared skewness (β_1) and kurtosis (β_2), as illustrated in figure 3. The roman numbers within the circles indicate the types of distribution. There are 3 main types (types I, IV and VI) covering areas in the $\beta_1 - \beta_2$ -diagram and 4 "transitional" types (types II, III, V and VII) covering line segments. The normal distribution obtains for $\beta_1 = 0$ and $\beta_2 = 3$ (or $c_1 = c_2 = 0$ and $c_0 > 0$) and is a limit distribution for all 7 types.

The identification of an appropriate type is now possible through the empirical values of β_1 and β_2. In figure 3, daily series are denoted by rhombi, weekly series by squares, monthly series by stars and quarterly series by triangles. The tendency towards normality under time aggregation for the exchange-rate series is obvious from the diagram. The daily and weekly series fall into the region of type IV and are near to the line segment of type VII with $\beta_1 = 0$ and $\beta_2 > 3$. As detailed in the previous chapter, the null hypothesis of symmetry cannot be rejected for the exchange-rate data. Hence, it is reasonable to set $\beta_1 = 0$ and to pick the type VII distribution for short-run exchange-rate dynamics since this is the only symmetric and leptokurtic distribution within this family.

Figure 3

Identification of Pearson type



With this additional motivation from the Pearson family of distributions, I next turn to the issue of estimating the generalized Student distribution [7]. As shown by Johnson and Kotz (1970, p. 115), the ML-estimators of the 2-parameter t -distribution satisfy the equations

$$(75) \qquad \sum_{t=0}^{T} \log\left(1+\frac{x_t^2}{\hat{\gamma}^2}\right) = T\left[\Gamma'(\hat{\eta}) - \Gamma'\left(\hat{\eta}-\frac{1}{2}\right)\right]$$

$$(76) \qquad \sum_{t=1}^{T} \left(1+\frac{x_t^2}{\hat{\gamma}^2}\right)^{-1} = T - \frac{T}{2\hat{\eta}}$$

where $\Gamma'(.)$ denotes the digamma function defined by $\Gamma'(\eta) = d\log\Gamma(\eta)/d\eta$. For $\eta > 2$, a very good approximation is given by $\Gamma'(\eta) \cong \log\left(\eta - \frac{1}{2}\right)$ (see Johnson and Kotz (1969), p. 7). The approximate variances of the estimators are given by

$$(77) \qquad Var(\hat{\eta}) = \frac{1}{T}\left[\Gamma''\left(\hat{\eta}-\frac{1}{2}\right) - \Gamma''(\hat{\eta}) - \frac{\hat{\eta}+1}{\hat{\eta}^2(2\hat{\eta}-1)}\right]^{-1}$$

$$(78) \qquad Var(\hat{\gamma}) = \frac{1}{T}\frac{(\hat{\eta}+1)\hat{\gamma}^2}{2\hat{\eta}-1}\frac{\hat{\eta}^2(2\hat{\eta}-1)}{\hat{\eta}^2(2\hat{\eta}-1)-(\hat{\eta}+1)\left[\Gamma''\left(\hat{\eta}-\frac{1}{2}\right)-\Gamma''(\hat{\eta})\right]^{-1}}$$

where $\Gamma''(.)$ is the trigamma function which is derived from the digamma function by differentiation. I used the approximation $\Gamma''(\eta) \cong \left(\eta - \frac{1}{2}\right)^{-1}$ which is quite accurate unless η is small.

I determined the parameter estimates by numerical calculation of the roots for the system of equations (75) and (76). The results are shown in table 4. In addition to the estimates of η and γ (and their standard errors in brackets), the degrees-of-freedom parameter df = $(2\eta - 1)$ is reported because this parameter might be easier to interpret than $\hat{\eta}$. For all series, η, and with it the degrees of freedom, can be estimated with rather great precision. The degrees of freedom range between 3.84 and 7.12 which implies that

[7] It is interesting to note that the issue of estimating the generalized Student distribution led to a violent clash between R.A. Fisher and Karl Pearson in 1922, following a paper of Fisher's where he introduced the maximum-likelihood method for estimation and showed that this method is superior to Pearson's method of moments for this distribution in terms of efficiency.

the distributions are strongly fat-tailed and peaked. Note, too, that the degrees of freedom for the daily pound series implies that the kurtosis for this series does not exist. The parameter γ, which is a scale factor, can only be estimated with satisfactory precision for the weekly data. There is a clear increase of $\hat{\gamma}$ under time aggregation but no such increase for $\hat{\eta}$ (with the exception of the sfr).

Table 4

Estimates of the generalized Student distribution

		mark	pound	sfr	yen
day	η	3.03 (0.06)	2.42 (0.04)	2.72 (0.05)	2.50 (0.04)
	df	5.07	3.84	4.45	4.01
	γ	1.16 (0.74)	0.89 (0.83)	1.25 (0.64)	0.86 (0.87)
	LR	416.5***	586.0***	541.7***	542.2***
	χ^2	153.6***	308.5***	129.8**	710.2***
week	η	2.79 (0.11)	3.01 (0.12)	4.06 (0.18)	2.66 (0.10)
	df	4.58	5.02	7.12	4.31
	γ	2.34 (0.35)	2.43 (0.35)	3.90 (0.26)	1.92 (0.41)
	LR	77.9***	85.4***	50.9***	101.3***
	χ^2	64.6**	67.1**	79.0***	57.6

Significance levels: see table 1

The likelihood-ratio statistics are, again, computed with regard to the benchmark model of the normal distribution which obtains for $\eta \rightarrow \infty$. It is clear that the LR test strongly rejects this restriction and thus also the normal distribution. The high values of the LR statistic obtained are comparable in size to those for the normal mixture and the compound Poisson process.

Finally, a χ^2 test was applied to check the goodness of fit. As for the previous two models, the χ^2 test strongly rejects the generalized t -distribution for daily data. The rejection is very strong for the daily pound and yen. In both cases, the rejection is primarily due to a lack of peakedness in the model as compared with the empirical distribution. The fit to the weekly data is marginally better but, still, the H_0 of a generalized t -distribution is rejected at the 5 percent level for 2 series and at the 1 percent level for the weekly sfr.

The overall rejection of the generalized t-distribution by the goodness-of-fit test suggests one looks further for alternatives to the scale-compounded distributions in order to model stochastic variance and, possibly, leptokurtosis. As mentioned above, the choice of the normal distribution as the compounded distribution guarantees leptokurtosis but the choice of the compounding distribution must also be motivated in order to avoid arbitrariness. In the next section, I will introduce another stochastic model which can be viewed as a scale-compounded normal distribution, but which also describes a very general and important stochastic process.

2.4 STABLE DISTRIBUTIONS AND REGULARLY VARYING TAILS

The family of stable distributions was introduced into economics and finance by the eminent mathematician Benoit Mandelbrot in the 1960's. In a series of papers he applied stable distributions to the modelling of income distribution and speculative prices. In finance, the model soon became popular because of its generality and because it was consistent with leptokurtosis found in many distributions of speculative price dynamics. The model of stable distributions has also been the most popular stochastic model in exchange-rate economics and was applied to exchange-rate data by Westerfield (1977), Rogalski and Vinso (1978), Friedman and Vandersteel (1982), Calderon-Rossell and Ben-Horim (1982), McFarland, Pettit and Sung (1982), So (1987), Boothe and Glassman (1987), Akgiray and Booth (1988) and Tucker and Pond (1988). I will come back to these studies later in this section and will explain the discrepancies between their results and mine.

The great popularity of the stable distributions is probably due to the fact that they are related to a generalization of the central limit theorem. Somewhat loosely formulated, the central limit theorem says that the sum of appropriately standardized, independent and identically distributed random variables with finite variance has a normal distribution in the limit as the number of summands goes to infinity. This strong result, of course, explains the dominant role of the normal distribution in probability theory and statistics. If one drops the assumption of finite variance one arrives at the family of stable distributions as the only limit distributions for sums of independent and identically distributed random variables (with appropriate standardization) and the normal distribution is just a special member of this family.

As explained above, a stochastic model which is based on the summation of random variables fits well into the broad framework of asset-market theories of the exchange rate. As the analysis in chapter 1 showed, the conditions of the central limit theorem are obviously not met by short-run exchange-rate dynamics since there are significant deviations from normality in these series. With respect to the central limit theorem, there are four different explanations for non-normality in the short-run data. First, non-normality might be caused by dependence in the series. In fact, some dependence of second order, i.e. in variances, has been found in the autocorrelation function of squared data. Also, the analysis of Markov chains revealed clustering of small and of large exchange-rate movements. Second, the assumption of identical distributions might be violated. As the result from the Levene test of homogeneity in variance indicate, there is in fact very strong evidence against homogeneity. Third, non-normality might be simply due to the fact that in short-run data the number of random components in the sum is too small to give a good approximation to the normal distribution. This explanation, however, does not seem to be very convincing given the steady flow of new information to the foreign-exchange market and continuous trading. Finally, non-normality might be caused by infinite variances of the random variables which determine exchange-rate fluctuation.

The last explanation leads directly to the family of stable distributions since they are the only limiting distributions under infinite variance. There is another property of stable distributions which makes them attractive to model speculative prices. In continuous time, a stochastic process which is driven by Gaussian increments, the so-called Wiener process, is itself continuous and cannot, therefore, explain jumps in the observed series. The sample path of a stochastic process driven by increments from a stable distribution, on the other hand, is everywhere discontinuous. Most of the variation is due to non-infinitesimal jumps. An almost trivial reason for jumps in price series is due to the fact that prices are always quoted in integer multiplies of certain currency units or base points, i.e. a price variable is always a discrete variable. More substantially, however, it is a common property of short-run speculative price series that they include great jumps which are incompatible with a normal distribution in the sense that they would be extremely unlikely under this distribution. For instance, the U.S. dollar depreciated against the Deutsch mark on Monday, 23rd September 1985 by 5.75 percent following the Plaza-Agreement at the weekend before to bring the dollar down. Under the normal distribution, with the population variance replaced by the sample variance, a depreciation of this magnitude would be expected to occur once in about 70,000 years. Similarly, the appreciation of the dollar against the mark by 4.95 percent on Thursday, 2nd November 1978 (following the announcement of strong support measures for the dollar by the Carter administration) would occur once in about 2200 years. Thus, there is an obvious need to choose a model which attaches more probability to extreme observations, i.e. to adopt a fat-tailed distribution. As I will show shortly, stable distributions are in fact fat tailed.

On the other hand, there are two reasons which appear to make stable distributions not very attractive for applications. First, not everybody is ready to accept the implications of infinite variance, since variance is a widespread concept in statistics and economics. One would have to rework many areas of statistic and economics to incorparate stable distributions. In fact, there were some attempts to formulate portfolio analysis for underlying stable distributions of returns by Fama and Samuelson following the apparent success of these distributions to fit stock-price distributions. However, stable distributions are rather awkward to work with analytically as will be shown shortly. Also, finite variance is sometimes seen to be implausible. It is true that every empirical variance must be finite, because every empirical support is finite, but this cannot really be advanced as an argument against stable distribution because they share with many other distributions the property of infinite support, including the normal distribution. The difference to finite-variance distributions is simply the increase of probability in the tails of the distributions.

As a heuristic way to examine whether infinite variance is present, Granger and Orr (1972) suggested to plot sequential variances. Under non-normal stable distributions, the sequential variances do not converge to a stationary value since the population variance for these distributions is infinite. Figure 8 in chapter 1 plots sequential variances for the four daily exchange-rates. There appears to be convergence in sequential variances for the sfr and yen series but not in the mark and pound series. This finding, however, may only be interpreted as a rough indication that infinite variances might drive the stochastic process in some series. As Granger and Orr already noted, shifts in sequential variances may also be caused by non-stationarity.

Second, stable distributions are not only awkward to work with analytically but also empirically because they cannot, in general, be described in closed forms of the density or the distribution function. Instead, they are usually described by their log-characteristic function

(79) $$\log \Phi_X(u) = i\delta u - |\gamma u|^{\alpha}[1 + i\beta(u/|u|)\omega(u,\alpha)]$$

where

(80) $$\omega(u,\alpha) = \begin{cases} \tan(\pi\alpha/2) & \text{if} \quad \alpha \neq 1 \\ 2\log(|u|)/\pi & \text{if} \quad \alpha = 1 \end{cases} .$$

and u is an auxiliary variable.

The characteristic function is determined by four parameters which can be related to the first four moments. First, δ is a location parameter $(-\infty < \delta < \infty)$ which is equal to the expected value of X if $1 < \alpha \leq 2$. It is equal to the median or mode if $\beta = 0$. Second,

γ is a scale parameter $(\gamma > 0)$ which measures the spread of the distribution. If $\alpha = 2$ (the case of the normal distribution), $\gamma = \sigma/\sqrt{2}$, where σ is the standard deviation. For $\alpha < 2, \gamma$ is some other measure of spread, for instance if $\alpha = 1$ and $\beta = 0$ (the case of the Cauchy distribution) γ is the semi-interquartile range. Third, β is a skewness parameter $(-1 \leq \beta \leq 1)$. If $\beta = 0$, then the distribution of X is symmetric, for $\beta > 0$ it is skewed to the right and for $\beta < 0$ it is skewed to the left. Together with the characteristic exponent $\alpha (0 < \alpha \leq 2)$, β determines the type of distribution. The characteristic exponent determines the highest order of finite moments within this family. If $\alpha < 2$, then the variance is infinite, i.e. the normal distribution with characteristic function

(81) $$\Phi_X(u) = \exp\left\{ i\theta u - \frac{1}{2}\sigma^2 u^2 \right\}$$

is the only member in this family with finite variance and finite moments of any (positive integer) order . The expected value is not finite (and a fortiori all higher moments are not finite) if $\alpha \leq 1$.

The characteristic exponent is related to kurtosis in the following way. Recall that kurtosis measures both peakedness and tail weight. For a symmetric stable random variable which is standardized by $x' = (x - \delta)/\gamma$, the density at the origin is given by (see Holt and Crow (1973))

(82) $$f(0 \mid \alpha) = \frac{1}{\pi\alpha} \Gamma\left(\frac{1}{2}\right)$$

and, obviously, this is a decreasing function in α. Hence, all symmetric non-Gaussian stable distributions are peaked as compared with the normal distribution.

The tail behaviour of the symmetric non-Gaussian stable distributions can be described by

(83) $$F(x) = Cx^{-\alpha} \quad \text{for} \quad x \to -\infty$$

(84) $$1 - F(x) = Cx^{-\alpha} \quad \text{for} \quad x \to +\infty \quad .$$

where F denotes the distribution function and $C > 0$, whereas $C = 0$ for the normal distribution (see Mandelbrot (1963)). This means that non-Gaussian stable distributions have fatter tails than normal distributions and it follows that the tail tickness is a decreasing function of the characteristic exponent α. The properties of (83) and (84) are quite important in economics and probability theory. In economics, the Pareto distribution, which is often applied to model income distributions, has the distribution function $F(x) = 1 - (k/x)^a$ with $a > 0, k > 0$ and $x \geq k$ and thus this distribution satisfies (84). This prompted Mandelbrot to introduce the name stable-Paretian distributions for the non-Gaussian stable distributions.

In probability theory, distribution functions which satisfy (83) and (84) are called distributions with regularly varying tails. They play an important role in the concept of the domain of attraction. The common distribution F of independent random variables X_j is defined to belong to the domain of attraction of a distribution G if the sum of the appropriately standardized X_j tends in distribution to G. The classical central limit theorem is based on the fact that F belongs to the domain of attraction of the normal distribution if F has finite variance [8]. On the other hand, a distribution belongs to the domain of attraction of stable Paretian distributions if it satisfies (83) and (84) with $0 < \alpha < 2$. In more informal terms, this implies that stable Paretian distributions can only attract distributions which are "similar" to themselves whereas the normal distribution can attract distributions with widely varying shapes (see Galambos (1988), chapter 6). I will come back to the concept of regularly varying tails later in this section.

Stable Paretian distributions are "self-attracting" in the sense that the sum of independent and identically distributed (i.e. with the same α and β) stable variables has also a stable distribution. This is easily seen from the log-characteristic function (79). Since

[8] Actually, the domain of attraction of the normal distribution is a bit wider than this, i.e. finite variance is only a sufficient and not a necessary condition (see Galambos (1988), section 6.4).

the characteristic function of a sum S of τ independent random variables X_j is equal to the product of the characteristic functions of the X_j, one gets

(85) $$\log \Phi_S(u) = i\tau\delta u - \tau \mid \gamma u \mid^{\alpha} [1 + i\beta(u/\mid u \mid)\omega(u,\alpha)].$$

Thus, the summation affects δ and γ but not α and β. This stability of the shape parameters α and β under addition gave rise to the name of this family of distributions. It follows from (85) that the standardizing factor for the scale parameter is $\tau^{-1/\alpha}$ whereas it is well-known that it is $\tau^{-1/2}$ for the normal distribution. Recall from Chapter 1 that under time aggregation the standard deviations of all four exchange-rate series increased by more than expected under the $\sqrt{\tau}$-law. This result can now be interpreted to have been caused by applying the wrong standardizing factor and hence as a possible indication of the presence of stable distributions.

Before I turn to estimators and estimates, I want to present an alternative interpretation of the stable Paretian distributions which permits to relate this model to the scale-compounded models of the previous three sections. I will show that a stable Paretian distribution can be obtained as a scale-compounded normal distribution (with zero mean) where the compounding distribution is positive stable. A distribution is said to be positive if it is concentrated on the non-negative real line. For a stable distribution to be positive, one needs to impose the additional parameter restrictions $\alpha < 1, \beta = 1$ and $\delta \geq 0$. Thus the model can be formulated as

(86) $$X \mid h \sim N(0; h\sigma^2)$$

where h is a random factor in the variance of the conditional distribution and H has a positive stable distribution. In terms of characteristic functions the model is given by

(87) $$\Phi_{X\mid h} = E(\exp\{iuX \mid h\}) = \exp\left\{-\frac{1}{2}h\sigma^2 u^2\right\}$$

(88) $$\Phi_H(v) = \exp\{-\mid \gamma v \mid^{\alpha} [1 + i(v/\mid v \mid)\tan(\pi\alpha/2)]$$

where v is an auxiliary variable. By the change of variable

(89) $$iv = -\frac{1}{2}u^2\sigma^2$$

one gets

(90) $$\begin{aligned}\Phi_X(u) &= \Phi_H\left(\frac{1}{2}iu^2\sigma^2\right)\\ &= \exp\left\{-\left|\frac{1}{2}i\gamma u^2\sigma^2\right|^\alpha [1 + i^2\tan(\pi\alpha/2)]\right\}\\ &= \exp\{-|\gamma' u|^{2\alpha}\}\end{aligned}$$

with $\gamma' = \frac{1}{2}\gamma\sigma^2[1 - \tan(\pi\alpha/2)]$. Therefore, one obtains a symmetric stable distribution with $\delta = 0$.

What makes stable Paretian distributions awkward to work with empirically, is the fact that, in general, closed forms for the corresponding densities are not available. Apart from the Cauchy distribution, which was mentioned above, the only other non-normal stable distributions with known closed-form densities are the Holtsmark-Levy-Smirnov distribution with $\alpha = 1/2$ and $\beta = \pm 1$ [9] and the Mitra distributions with $\alpha = 2^{-k} (k = 1, 2, \ldots)$ and $\beta = 0$ (see Csörgö (1984)). Of course, the probability law of a random variable can be described by the distribution function, by the density function, or by the characteristic function, and all three ways are perfectly equivalent. Furthermore, the three functions are related through the operations of differentiation, integration and Fourier transform. However, in order to apply ML methods, one needs to compute the densities.

The lack of closed forms of the densities has led to the suggestion of numerous estimators for the parameters, especially for α which is the decisive parameter [10]. For exchange-rate data and other speculative prices, the most popular estimator of α has been the one suggested by Fama and Roll (1968, 1971). Their estimator is based on the matching of empirical and theoretical fractiles and exploits the fact that tail weight is a function of

---

[9] With $\alpha = 1/2$ and $\beta = -1$, one gets the reciprocal of a chi-square variable with one degree of freedom. This distribution belongs to type V of the Pearson family.

[10] A good overview is provided by Csörgö (1984).

α. McCulloch (1986) generalized the Fama-Roll estimator to cover non-symmetric distributions and also removed the asymptotic bias in the Fama-Roll method.[11] McCulloch's method has been applied to exchange-rate data by So (1987) and Tucker and Pond (1988).

A method based on the empirical characteristic function has been introduced by Koutrouvelis (1980) and was applied to exchange-rate data by Akgiray and Booth (1988). This method depends crucially on the values of the auxiliary variable u chosen. Koutrouvelis suggested to select different values of u_i and to estimate α from a regression of $\log(-\log|\Phi_T(u_i)|^2)$ on $\log|u_i|$, where $\Phi_T(u_i)$ is the empirical characteristic function based on T observations.[12]

Finally, Feuerverger and McDunnough (1981) proposed to overcome the problem of lacking densities by estimating the densities via the fast Fourier transform (FFT). This method implies some computational burden, but it is the most elegant and convincing method of all the ones which have been proposed and it permits to apply ML methods on the estimated densities. This estimator has been applied by Boothe and Glassman (1987) to exchange-rate data.

Following the analysis of chapter 1, I restrict the model of stable distributions to the symmetric case, i.e. $\beta = 0$, with $\delta = 0$ and $1 < \alpha \leq 2$. An estimated value of α in the interval $(0,1]$ would obviously put me in an unpleasant position of having to reconcile such a result with the finding in Chapter 1 that the H_0 of a constant mean at zero cannot be rejected. Anyway, in the actual estimations of α there was never a convergence to the value of 1. Also, in previous applications of stable distributions to exchange rates, all estimates of α were above 1. Note that the restriction on α implies that the model of stable distributions and the model of the generalized Student distribution have the same "boundary distributions" since the Student distribution with 1 degree of freedom is the

[11] There is a downward bias in the Fama-Roll estimator of α.

[12] Akgiray and Lamoureux (1989) compared the McCulloch estimator and the Koutrouvelis estimator in a Monte-Carlo study and found that the Koutrouvelis estimator performed better than the McCulloch estimator in terms of bias and precision for any sample size and values of α and β. Both methods were quite accurate in estimating α.

Cauchy distribution. The "upper boundary distribution" is in both cases the normal distribution. The restriction $\beta = 0$ is imposed because the H_0 of symmetry could not be rejected in chapter 1.

In order to understand the Feuerverger-McDunnough approach, note that a density function $f(x)$ can be obtained from a characteristic function $\Phi(u)$ via the Fourier transform (see e.g. Parzen (1960), chapter 9)

(91) $$f(x) = \frac{1}{2\pi} \int^{\infty} \exp\{-iux\}\Phi(u)du.$$

$$f(x) = \frac{1}{2\pi} \int \exp\{-iux\}\Phi(u)du.$$

If the FFT algorithm is applied to evaluate the intergral in (91) then x takes on values on the equi-spaced grid $0, \pm\Delta x, \pm 2\Delta x, \ldots, \pm N\Delta x/2$. For the auxiliary variable one gets $\Delta u = 2\pi/(N\Delta x)$ and the algorithm is applied to the sequence $1/2, \Phi(\Delta u), \ldots, \Phi((N-1)\Delta u)$. If the output sequence is multiplied by $2/(N\Delta x)$, one obtains $f(0), f(\Delta x), \ldots, f(N\Delta x/2)$ and the corresponding densities at the negative values of the Δx grid points (which do not contribute additional information under symmetric distributions). In order to apply this method, one has to choose values of N and Δx. Following the suggestions of Feuerverger and McDunnough, I set $N = 1024$ and $\Delta x = 0.05$.

The consequence of applying the discrete FFT approximation

(92) $$f(k\Delta x) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \Phi(j\Delta u)\exp\{-i2\pi jk/N\} \quad k = 0, 1, \ldots, N-1$$

of the Fourier integral is the so-called aliasing effect (see e.g. Fuller (1976) pp. 119-120). Feuerverger and McDunnough found that the aliasing error stays essentially constant and can thus be determined by the difference between the estimated density and the exact density at $x = 0$ which is given in (82). This requires to standardize the data by x_t/γ with the current estimate of γ. In order to get the densities at the actual values of x_t, I used cubic Hermite interpolation (see Hamming (1973), pp. 277-287). I applied these methods and checked the calculated densities with the densities tabulated by Holt and Crow (1973) and found complete agreement. Having obtained densities, I then used a numerical gradient method with local search at suspected maxima to get the ML estimates $\hat{\alpha}$ and $\hat{\gamma}$.

A central issue in the fitting of stable distributions to speculative prices is the stability of α under time aggregation. According to this model, only γ and δ (if non-zero) should change under time aggregation. In earlier applications to exchange-rate data, however, there was an apparent increase in the estimated α's under time aggregation. Table 5 gives an overview of the studies.[13] McFarland et al. (1982) and Boothe and Glassman (1987) observed an "instability" of α and interpreted it as evidence against this model. At this stage, however, this conclusion is not very convincing. First, an increase of α could also occur in a model which is a mixture of stable distributions. That is, one does not have to abandon the family of stable distributions in order to reconcile rising α's with the model. Second and more importantly, those earlier studies listed in table 5 do not present any test statistics on which proper statistical inference could be based to test the stability of α. The only study which reports standard errors of α is the one by So but he analyses only daily data. Therefore, the stability of α, and thus the applicability of this model, is still an open question.

My results from applying the Feuerverger-McDunnough approach to the estimation of stable Paretian distributions are reported in table 6. Starting values for $\hat{\alpha}$ and $\hat{\gamma}$ in the iterations were obtained from the Koutrouvelis estimators. The estimates of α from the Koutrouvelis method ranged between 1.70 and 1.78 for daily data and 1.73 and 1.80 for weekly data.

[13] From the nine studies mentioned at the beginning of this section, three studies are not included in the table. In the article of Rogalski and Vinso (1978) it is unclear which data they used to estimate the stable distributions. Calderon-Rossell and Ben-Horim (1982) did not estimate the characteristic exponent. Instead, they applied a Kolmogorov-Smirnov goodness-of-fit test to the estimated Cauchy and normal distribution and an arbitrary stable distribution with $\alpha = 1.5$. Akgiray and Booth (1988) reported only likelihoods for the estimated stable distributions. However, these likelihoods are very unplausible. For other candidate models, the log-likelihoods are between 7944.2 and 8192.5 but for the stable distribution they are between -24302.5 and -17307.1. It is quite likely that they made a serious error in estimating the stable distributions. I will show in chapter 4 that the likelihoods of all stochastic models applied in this study are very similar in magnitude. The candidate models of Akgiray and Booth (1988) are a subset of these stochastic models.

Table 5

Studies applying stable distributions to exchange-rate data

Authors	*Estimates of* α	*Method*	*Currencies*	*Period*
Westerfield	1.33-1.51	Fama-Roll	DM, £, SFr	day
Friedman and Vandersteel	1.11-1.45 1.30-1.50 1.33-1.63	Fama-Roll	DM, £, SFr, ¥	day week † month †
McFarland, Pettit and Sung	1.12-1.40 1.50-1.92	Fama-Roll	DM, £, SFr, ¥	day week †
So	1.10-1.16	McCulloch	£, ¥	day
Boothe and Glassman	1.27-1.62 1.54-1.72 1.37-2.00	Feuerverger-McDunnough	DM, £, ¥	day week month
Tucker and Pond	1.12-1.39 1.26-1.55	McCulloch	DM, £, SFr, ¥	day week †

† Quasi-weekly or quasi-monthly data obtained from sums of 5 or 20 daily data, respectively. Note that only those currencies are listed in column 4 which are also analysed in this study.

Compared with these earlier studies, my Koutrouvelis estimates of α for daily data are surprisingly high. I also applied the Fama-Roll method to estimate α and got values between 1.41 and 1.55 for the four daily series. These estimates from the Fama-Roll method are more in line with earlier studies. Since the comparative simulation study by Akgiray and Lamoureux (1989) showed that the Koutrouvelis estimators of α is superior to the one by McCulloch (and hence also to the one by Fama and Roll) in terms of bias and precision, one may conclude that earlier studies probably underestimated α.

As table 6 shows, the ML estimates of the Feuerverger-McDunnough approach do not differ very much from the Koutrouvelis estimates but the estimates of α tend to be somewhat smaller. Standard errors of the parameters are reported in brackets. I also estimated stable Paretian distributions for monthly data because stability of α under time aggregation is central to this model.

Table 6

Estimates of stable distributions by the Feuerverger-McDunnough method

		mark	pound	sfr	yen
day	α	1.74 (0.03)	1.56 (0.03)	1.68 (0.03)	1.60 (0.03)
	γ	0.40 (0.01)	0.35 (0.01)	0.46 (0.01)	0.33 (0.01)
	LR	371.8 ***	511.1 ***	497.1 ***	473.7 ***
	χ^2 (97)	226.7 ***	418.4 ***	220.0 ***	741.7 ***
week	α	1.68 (0.07)	1.74 (0.07)	1.68 (0.07)	1.65 (0.07)
	γ	0.84 (0.04)	0.84 (0.03)	1.00 (0.05)	0.71 (0.03)
	LR	60.6 ***	74.4 ***	38.9 ***	81.6 ***
	χ^2 (47)	57.8	77.5 ***	67.0 **	61.9 *
month	α	1.81 (0.14)	1.92 (0.07)	1.91 (0.08)	1.87 (0.17)
	γ	2.18 (0.19)	2.16 (0.13)	2.60 (0.16)	2.26 (0.21)
	LR	3.56 *	0.99	4.31 **	7.49 ***
	χ^2 (27)	21.4	27.8	33.7	47.9 ***

Significance levels: see table 1

There are several remarkable findings. First, for short-run exchange-rate dynamics, the estimates of α are significantly below 2 and there is no obvious increase of $\hat{\alpha}$ in weekly data as compared with daily data. According to the likelihood-ratio statistic, stable distributions achieve a much better fit in comparison with the normal distribution. However, the χ^2 test of goodness-of-fit rejects all daily models at the 1 percent significance level and one of the weekly models at the same level. This rejection by the χ^2 test is very similar to the rejection of the previous three models (mixture of normal distributions, compounded Poisson distribution and generalized t -distribution) by this test. The results for the monthly data, however, are drastically different. None of the $\hat{\alpha}$'s is significantly different from 2 as judged from their standard errors. Accordingly, the LR test rejects the H_0 of normality only for the yen. Thus, there is strong evidence for convergence towards normality. I also estimated the model with quarterly data and got point estimates of $\hat{\alpha}=2$

for all four series. For the quarterly mark and pound, the starting values from the Koutrouvelis method were also equal to 2. The results from table 6 are broadly in line with earlier studies.

The fact that α is significantly smaller than 2 for short-run data but not for monthly or quarterly data indicates that the exchange rates do not follow stable Paretian distributions but other fat-tailed distributions. In order to examine this possibility, I come back to the concept of regularly varying tails. As mentioned above, only (symmetric) distributions whose tail probabilities follow the function $Cx^{-\alpha}$, with $0 < \alpha < 2$, belong to the domain of attraction of (symmetric) stable Paretian distributions. If $\alpha > 2$, then the distribution belongs to the domain of attraction of the normal distribution. I, therefore, extend the model of stable distributions to the class of distributions with regularly varying tails of which stable distributions are a sub-class which obtains when the tail probabilities follow $Cx^{-\alpha}$ with $\alpha < 2$ (recall that stable distributions are self-attracting). All other fat-tailed distributions with tail probability $Cx^{-\alpha}$ and $\alpha > 2$ do not belong to the family of stable Paretian distributions. In contrast to stable Paretian distributions they converge to the normal distribution under addition.

Within the class of distributions with regularly varying tails one can, therefore, discriminate between fat-tailed stable and non-stable distributions by estimating the coefficient of regular variation α. One can reject the model of stable Paretian distributions if α turns out to be larger than 2. Note that α can have two meanings in this context. First, it denotes the coefficient of regular variation which determines the tail behaviour of the distributions function and, second, it denotes in addition the characteristic exponent of stable distributions if $\alpha < 2$.

Analysing the family of distributions with regularly varying tails can enable us to reject the sub-class of stable Paretian distributions, but it does not help to identify a specific distribution function if α is estimated to be greater than 2. Some distributions like Student's distribution are known to have regularly varying tails with $\alpha > 2$ but one cannot associate a specific $\alpha > 2$ with a specific distribution function.

Hill (1975) proposed a conditional ML method to estimate α. The estimator is given by:

$$\hat{\alpha}(q) = \left[\frac{1}{q} \sum_{j=1}^{q} \log | x_{(T-j+1)} | - \log | x_{(T-q)} | \right]^{-1} \tag{93}$$

where $x_{(i)}$ denotes the i-th order statistic in descending order. It is called a conditional ML estimator because it is a function of the chosen integer q. Hall (1982) established the asymptotic normality of $\hat{\alpha}(q)$ and showed that the asymptotic standard errors of $\hat{\alpha}(q)$ under the H_0 of $\alpha = \alpha_0$ are equal to $\hat{\alpha}(q)/\sqrt{q}$. In the estimation of α, the choice of q is crucial. Following the work of Hall (1982), Phillips and Loretan (1990) suggested to apply a range of values of q centred around $q^* = T^{2/3}/\log(\log T)$.

The estimator in (93) can be applied to both the lower and the upper tail of a distribution but also to both tails simultaneously. In the latter case, one has to take absolute values first before the observations are ordered according to magnitude. For symmetric distributions it is preferable to apply the 2-tailed version. In the sequel, I will only report results from this 2-tailed version, because the results from analysing the lower and upper tails separately do not differ substantially from the results of the 2-tail version.[14]

The rule that q should be centred around q^* leads to approximate values of q^* of 100, 40, 20, and 10 for the daily, weekly, monthly and quarterly data respectively. Figure 4 plots values of $\hat{\alpha}(q)$ centred around the approximate values of q^* with the upper bound of q equal to $2q^*$. Some low values of q have been truncated because $\hat{\alpha}(q)$ is very erratic for these values of q. To save space, only the plots for the four yen series are shown.

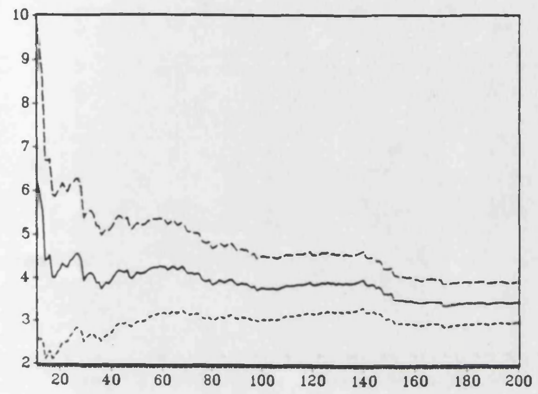
The solid lines in figure 4 show $\hat{\alpha}$ as a function of q and the dashed lines mark 95-percent confidence intervals. In all four plots, every single value of $\hat{\alpha}$ is above the critical line of $\alpha = 2$. The estimated $\hat{\alpha}$'s converge apparently to values significantly above 2 in short-run data. In monthly and quarterly data, however, the confidence intervals include

[14] There is only a tendency in short-run data for $\hat{\alpha}(p)$ to be somewhat lower when it is estimated from the lower tail than when it is estimated from the upper tail or both tails.

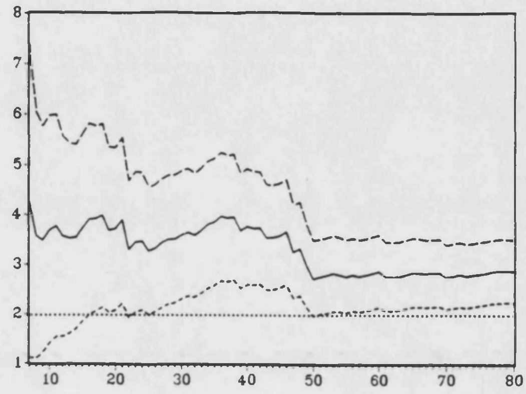
Figure 4

Coefficient of regular variation as a function of q : yen-dollar series

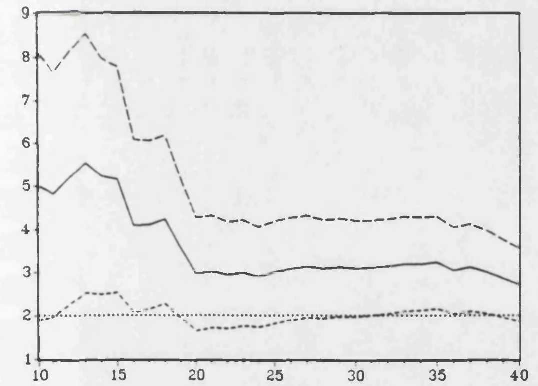
a) day



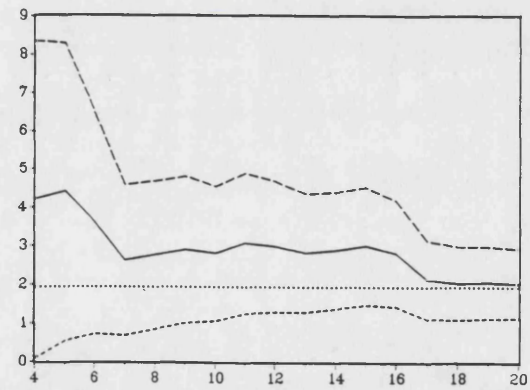
b) week



c) month



d) quarter



the value of 2. This could be attributed to a decrease in power of the test under decreasing sample size or to convergence to normality under time-aggregation. The decisive result is, however, that the H_0 of $\alpha < 2$ can be firmly rejected and this is very strong evidence against stable Paretian distributions.

The estimates of $\hat{\alpha}(q^*)$, based on the 2-tailed version, for all series, together with their standard errors, are reported in table 7. For all series, $\hat{\alpha}(q^*)$ is above 2 and for all short-run data, it is significantly above 2. In most cases, the exponent of regular variation lies in the interval from 3 to 5.

There is another interesting aspect in the estimation of α. For a distribution which satisfies (83) and (84), one obtains (see Brockwell and Davies (1987), p. 479)

(94) $$E\,|X|^b = \infty \qquad \text{if} \quad b \geq \alpha$$

(95) $$E\,|X|^b < \infty \qquad \text{if} \quad b < \alpha,$$

i.e. α also determines the maximal moment exponent. If $\alpha \leq 3$, then the third moment is not finite and when $\alpha \leq 4$ then the fourth moment is not finite. For most series, α is not significantly smaller than 4 but significantly smaller than 5. This implies that the kurtosis is finite and gives a late justification for analysing only moments up to order 4 in Chapter 1.

Table 7

Estimates of the exponent of regular variation

		mark	pound	sfr	yen
day	$\hat{\alpha}$ (100)	3.81 (0.38)	3.86 (0.39)	3.50 (0.35)	3.77 (0.38)
	χ^2 (18)	26.0*	19.6	21.6	1.24
week	$\hat{\alpha}$ (40)	3.51 (0.56)	3.35 (0.53)	4.80 (0.76)	3.76 (0.59)
	χ^2 (6)	3.6	4.0	5.2	6.0
month	$\hat{\alpha}$ (20)	3.02 (0.68)	4.34 (0.97)	5.37 (1.20)	2.96 (0.66)
	χ^2 (2)	2.8	1.2	0.0	1.6
quarter	$\hat{\alpha}$ (10)	3.46 (1.09)	4.97 (1.57)	2.16 (0.68)	2.80 (0.88)
	χ^2 (2)	0.8	2.0	0.4	1.2

Also reported in table 7 is a test for goodness of fit. Hill (1975) showed that the variable $W_q = \alpha q \log(X_{(q)}/X_{(q+1)})$ has an exponential distribution with expectation 1 under the assumption of regularly varying tails. This suggests to examine the approriateness of the model with a χ^2 goodness-of-fit test for $W_1, \ldots, W_{q*}$. The results from the χ^2 test in table 7 show that there is no evidence against this model. The only rejection of an exponential distribution for W_q is for the daily mark series at the 10 percent significance level.

In summary, both the estimates of the exponent of regular variation and the convergence of $\hat{\alpha}$ to the value of 2 in unrestricted ML estimation cast serious doubts on the applicability of stable Paretian distributions to exchange-rate dynamics. On the other hand, this result is good news for those who have feared that traditional statistical methods and concepts in financial economics are not applicable to speculative prices because they were thought to follow distributions with infinite variance.

2.5 SUMMARY

In this chapter, I have applied four stochastic models (mixture of normal distributions, compound Poisson distributions, Student's distributions and stable Paretian distributions) to capture the statistical properties of exchange-rate fluctuations. These models are compatible with leptokurtosis. But they do not imply heteroskedasticity and in this sense they are static models. Furthermore, stable Paretian distributions do not converge to normality under time aggregation.

Distributional models of exchange-rate dynamics have also been compared by Boothe and Glassman (1987), Akgiray and Booth (1988), and Tucker and Pond (1988). This study differs from those studies in several important aspects. First, I provide a unifying framework for the application of the distributional models by showing that all four models can be derived as scale-compounded normal distributions, i.e. as models of a normal distribution with a stochastic variance. With this unifying framework lacking, the selection of distributional models would seem to be arbitrary. Second, the application of distributional models is motivated by and related to the stylized facts of the data, i.e. to the highly significant leptokurtosis. It is well-known that the kurtosis of Student's distribution is larger than 3 and that it is infinite for stable Paretian distributions. Here I show that very general versions of scale-mixtures of normal distributions and of compound Poisson distributions also imply leptokurtosis. Third, the models have not been estimated by a single statistical method in those earlier models. In general, models have been estimated by maximum likelihood but Akgiray and Booth (1988) and Tucker and Pond (1988) did not estimate stable Paretian models by ML whereas Boothe and Glassman (1987) did not estimate normal mixtures by ML. This is important for likelihood based comparisons of models by likelihood ratios, AIC and SIC (as applied by these authors) since models which have not been estimated by ML have an obvious disadvantage in these comparisons. I will present comparisons, based on SIC, of these models together with the models of the next chapter in chapter 4.

The main results of this chapter may be summarized as follows. If one compares the four stochastic-variance models with the null hypothesis of Gaussian white noise by a likelihood-ratio statistic, the null hypothesis is very strongly rejected in all four cases. However, a χ^2 goodness-of-fit test rejects most models, especially for daily data. Furthermore, the model of stable Paretian distribution is rejected by direct ML estimation of the characteristic exponent (for monthly and quarterly data) and of the coefficient of regular variation. To the best of my knowledge, this is the first direct inference-based rejection of this model that has been reported.

A major drawback of all four stochastic-variance models is the fact that they do not imply heteroskedasticity which is another highly significant empirical regularity of exchange-rate data. The next chapter introduces two classes of models which are able to capture this property.

CHAPTER 3

MODELS OF EXCHANGE-RATE HETEROSKEDASTICITY

In order to incorporate heteroskedasticity, the variance of a random variable has to be a function of time or of some other variables. As explained in the introduction, I will restrict this study to univariate analysis. Therefore, the variance of exchange-rate dynamics Δe_t will only be modelled as a function of the history of e_t.

The difference between the heteroskedasticity models of this chapter and the stochastic-variance models of the previous chapter lies in the fact that for the stochastic-variance models the conditional variance $\sigma^2(\Delta e_t \mid \mathrm{I}_t)$, given information I available at time t, is equal to the unconditional variance $\sigma^2(\Delta e_t)$, whereas this equality will not, in general, hold for the heteroskedasticity models. In other words, the variances $\sigma^2(\Delta e_t)$ are serially independent and independent of Δe_t in the models of scale-compounded distributions, but not so in the heteroskedasticity models. The probability models of the stochastic-variance distributions assume that there are two independent random variables. The first random variable determines the variance of the normal distribution and this random variable has a multinomial distribution for the mixture of normal distributions, it has a Poisson distribution for the compound Poisson process, it has an inverted Gamma distribution for the (generalized) Student distribution, and it has a positive stable Paretian distribution with $\alpha < 1$ for the family of stable Paretian distributions. The second random variable, given the drawing from the variance distribution, is determined by drawing from a normal distribution with the given variance. The crucial assumptions are that the two random variables are independent and that the first random variable is serially uncorrelated.

In order to be consistent with the clustering of large and small exchange-rate fluctuations (of either sign) and with the significant autocorrolation of squared data (see Chapter 1), the models of this chapter either dispense with the assumption that the stochastic variance be serially uncorrelated (as in the Markov-switching model of the next section) or with the assumption that the variance variable be independent of the exchange-rate variable. Both assumptions are relaxed in the family of ARCH-type models (see section 3.2).

3.1 MARKOV-SWITCHING MODELS

As described above, the finite-mixture model may be decomposed analytically into two independent random variables where the first variable is a stochastic variance with a multinomial distribution (or a Bernoulli distribution if we have a two-component scale mixture) and the second variable has a conditional normal distribution. One may, therefore, regard the first variable as a state variable s_t which can take on values $j = 1, \dots, J$, where J is the number of components in the mixture. The probability of drawing from component j is p_j where component j is a normal distribution with mean Θ_j and variance σ_j^2. As in Section 2.1, this model may be written as

(1) $$f(x) = \sum_{j=1}^{J} p_j f_j(x \mid \theta_j, \sigma_j)$$

where

(2) $$f_j(x \mid \theta_j, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(x-\theta_j)^2}{2\sigma_j^2}\right\}.$$

The Markov-switching model is an extension of this mixture model where it is assumed that the state variable s_t follows a time-homogeneous first-order Markov process characterized by the transition probabilities

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

For a Markov-switching model with two states, one gets a 2x2 transition matrix of states $P = \{p_{ij}\}$, with two independent probabilities p_{11} and p_{22}. Of course, it follows that $p_{12} = 1 - p_{11}$ and that $p_{21} = 1 - p_{22}$. The Markov chain in (3) together with the mixture model in (1) and the specification of the normal distribution in (2) gives for J=2 a 7 parameter model with parameter vector $\Psi = \{p_{11}, p_{22}, \theta_1, \theta_2, \sigma_1, \sigma_2, \rho\}$ where $\rho = p(s_1 = 1)$.

One needs ρ, the probability of being in state 1 in time period $t = 1$, to start off the Markov chain and a natural choice is to set ρ equal to the stationary probability of being in state 1 which is given by

$$\rho = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}. \tag{4}$$

The basic idea of the model is due to Baum et al. (1970) who suggested to estimate the model with the expectation-maximization (EM) algorithm. Furthermore, they derived the essential properties of the EM algorithm within a general model with Markov-chain dependence. They showed that, under certain regularity conditions, the EM algorithm increases the likelihood function monotonically and that it converges to the ML estimates. Lindgren (1978) detailed the steps needed to implement the EM algorithm for the Markov-switching model, extended the model to the case of switching regressions, and examined the statistical properties of the model and its ML estimators. In a series of papers, Hamilton recently extended the model and adapted it to the modelling of interest rates, exchange rates and the business cycles (see Hamilton (1988, 1989, 1990, 1991 a, b) and Engel and Hamilton (1990)).

If $J = 2$ and $p_{11} = p_{21}$ or, what is equivalent, if $p_{12} = p_{22}$ then the Markov-switching model reduces to the mixture model since these equalities imply that the s_t are independent, i.e. $p(s_t \mid s_{t-1}) = p(s_t)$. Since the statistical analysis in chapter 1 showed that there is significant clustering of large and of small exchange-rate movements, one expects that $p_{11} > p_{21}$, that $p_{22} > p_{12}$, and that the variance effects are much stronger than the mean effects. The condition that the diagonal elements of the transition matrix be larger than the off-diagonal elements together with the assumption of scale mixtures (i.e. $\theta_1 = \ldots = \theta_J$), reflects the notion of persistence in variability (or volatility, as it is called nowadays in financial markets) or of periods of turbulence and tranquillity.

The economic interpretation of the model is straightforward. Assume that there are two kinds of "news variables" hitting the foreign exchange market and causing exchange-rate movements. One kind of news causes low variance movements in the exchange rate and is associated with tranquil periods. The second kind of news represents shocks to the foreign-exchange market which may be unexpected regime shifts, massive interventions by central banks, the outbreak of economic or political crises and the like. These shocks would be associated with turbulent periods and a higher variance σ_2^2 of exchange-rate movements. From its very nature of shocks, it would also follow that the parameter p would be expected to be near the value of 1.

As a simple economic illustration of the model, take the general asset-market model of exchange-rate determination in the form of

(5) $$e_t = \phi e_{t,t+1} + \alpha x_t$$

(see chapter 1 equations (2) -(5)) with $\phi + \alpha = 1$ and $|\phi| < 1$. A solution to (5) is given by

(6) $$e_t = \alpha \sum_{i=0}^{\infty} \phi^i x_{t,t+i} \quad .$$

Shifting (6) forward by one unit of time and substracting (6) from it, gives the exchange-rate dynamics

(7) $$e_{t+1} - e_t = \alpha[\sum_{i=0}^{\infty} \phi^i x_{t+1,t+1+i} - \sum_{i=0}^{\infty} \phi^i x_{t,t+i}] \quad .$$

Assume further that the exogenous variable x_t follows the stochastic process

(8) $$x_{t+1} = x_t + \omega_{t+1}$$

and ω_{t+1} follows the first-order autoregressive process

(9) $$\omega_{t+1} = \rho\omega_t + v_{t+1} \quad ,$$

where $0 \leq \rho < 1$ and v_{t+1} is white noise. It is now easily seen that

(10) $$e_{t+1} - e_t = \omega_t$$

with $E(\omega_t) = 0$ and

(11) $$\mathrm{Var}(\omega_t) = \frac{1}{1-\rho}\sigma_v^2 \quad .$$

According to (8), x_t is an infinite-memory process like a random walk but the persistence of the shocks ω_{t+1} depends upon the value of ρ. For $\rho = 0$, the shocks are only transitory and the closer ρ is to the value of 1, the stronger is the persistence of the shocks. As a consequence, it can be seen from (11) that an increase in the persistence of the shocks increases the variance of exchange-rate movements.

In this framework, a Markov switching model can be viewed as a regime shift caused by a change in the parameter ρ. It should be stressed, however, that this simple example is meant to serve as an illustration only of a model which is compatible with a Markov switching model. There is no intention to identify or estimate the structural model.

It has to be examined whether the Markov-switching model captures not only heteroskedasticity but also the other two major statistical properties of exchange-rate dynamics, namely leptokurtosis and convergence to normality. The issue of leptokurtosis is easily dealt with because the stationary distribution of a Markov-switching model is a mixture distribution if the transition matrix is not degenerate. For a 2 x 2 transition matrix, it is sufficient to assume that $0 < p_{11}, p_{22} < 1$. It then follows from the stationary distribution that the proof of leptokurtosis for arbitrary scale mixtures of normal distributions (see section 2.1) may be applied without modifications.

The issue of convergence to normality under time aggregation is less straightforward. The central limit theorem is not directly applicable since the condition of independence is violated. However, Lindgren (1978) established asymptotic independence for general versions of the Markov-switching model.[1] It then follows from generalized versions of the central limit theorem that convergence to normality occurs (see White (1984)).

Estimation of the Markov-switching model is quite involved since the state variable s_t is not observable. I applied the EM algorithm as suggested by Baum et al. (1970),

[1] Lindgren proved asymptotic independence by showing that "mixing" conditions are satisfied. The use of the term "mixing" might cause some confusion in this context since the "mixing" conditions are not related in any way to the mixing of densities as in (1).

Lindgren (1978), and Hamilton (1990). The principle of the EM algorithm, which has a wide range of applications in statistics to models with incomplete and unobservable data, can be described as follows. Decompose the log-likelihood function into

$$(12) \qquad \log p(\Psi \mid X) = \log p(\Psi \mid S, X) - \log p(S \mid \Psi, X) + C$$

where X denotes the data $\{\Delta e_1, \ldots, \Delta e_T\}$, S denotes the unobservable states $\{s_1, \ldots, s_T\}$, and $C = \log p(S \mid X)$ is independent of Ψ and may, therefore, be disregarded in the maximization. Equation (12) can be rewritten as

$$(13) \log p(\Psi \mid X) = \int_S \log p(\Psi \mid S, X) p(S \mid \Psi, X) dS - \int_S \log p(S \mid \Psi, X) p(S \mid \Psi, X) dS .$$

Now define

$$(14) \qquad Q(\Psi, \Psi^*) = \int_S \log p(\psi \mid S, X) p(S \mid \Psi^*, X) dS$$

and

$$(15) \qquad H(\Psi, \Psi^*) = \int_S \log p(S \mid \psi, X) p(S \mid \Psi^*, X) dS .$$

One can then describe a change in the log-likelihood in an iteration from k to $k+1$ as

$$(16) \log p(\Psi^{k+1} \mid X) - \log p(\Psi^k \mid X) = Q(\Psi^{k+1}, \Psi^k) - Q(\Psi^k, \Psi^k) - [H(\Psi^{k+1}, \Psi^k) - H(\Psi^k, \Psi^k)] .$$

The algorithm consists of an expectations step (E), in which $Q(\Psi, \Psi^*)$ is calculated, and of a maximization step (M), in which $Q(\Psi, \Psi^*)$ is maximized with respect to Ψ. Baum et al. (1970) showed that choosing Ψ^{k+1} such that $Q(\Psi^{k+1}, \Psi^k) > Q(\Psi^k, \Psi^k)$ implies that $H(\Psi^{k+1}, \Psi^k) - H(\Psi^k, \Psi^k) \leq 0$ and therefore, according to (13), the log-likelihood increases. As mentioned before, they also showed that under weak regularity conditions, the EM algorithm converges to the ML estimates.

The essence of the EM algorithm is, therefore, to replace the maximization of the log-likelihood function by the maximization of $Q(\Psi, \Psi^*)$ which is a weighted conditional log-likelihood function where the likelihood is conditional on the unobserved state variable S and the weights $p(S \mid \Psi^*, X)$ are the conditional probabilities for S, conditional on the previous estimate of Ψ and on the data X. The weights $p(S \mid \Psi^*, X)$ are a useful by-product of the EM algorithm since they give inferences about the state variable s_t, given the full sample X. They are called smoothed probabilites, whereas the probabilities $p(s_t \mid x_1, \ldots, x_t, \Psi)$ are called filter probabilities.

The most time-consuming computations in this application of the EM algorithm are the E steps in which the smoothed probabilities are computed. Lindgren (1978) and, in a different but equivalent form, Hamilton (1990) described how the smoothed probabilities can be computed in a recursive way. For daily data, the computational burden is quite substantial even if only a two-component model without mean effects, i.e. with $\Psi = \{p_{11}, p_{22}, \sigma_1, \sigma_2\}$, is estimated. A typical iteration from k to $k+1$ takes about 45 minutes and often many iterations are required to satisfy the convergence criterion. Starting values for the parameters of the normal densities can be obtained from estimates of mixture distributions.

In Table 1, I report estimates of a two-state Markov-switching model without mean effects but I also estimated models with more than two states and with mean effects. As Kaehler and Marnet (1993) show, the mean effects are in general not significant for daily, weekly and monthly data, with the exception for the daily yen series. Similarly, a third component contributed little to the goodness of fit of these models. Engel and Hamilton (1990) applied a two-state Markov-switching model to quarterly exchange rates and found significant mean effects in two out of three series.

Table 1 shows that the estimated diagonal transition probabilities p_{11} and p_{22} are in all cases very high, i.e. they are always larger than 0.9. The expected duration of state j, δ_j, can be calculated from $\delta_j = (1 - p_{jj})^{-1}$ and one finds, for instance, that the expected duration of state 1 is 12 weeks for the weekly mark series. For the daily data, the expected duration of states varies between 14.9 days (state 1 for pound) and 57.5 days (state 2 for mark), for weekly data δ_j varies between 12.0 weeks and 44.2 weeks (state 2 for pound). For the mark series, the δ_j's are rougly consistent across time horizons since $\delta_1 = 56.6$ for daily data with $\delta_1 = 12.0$ for weekly data and also $\delta_2 = 57.5$ for daily data with $\delta_2 = 15.6$ for weekly data. On the other hand, $\delta_2 = 20.8$ for the daily pound but $\delta_2 = 44.2$

for the weekly pound. With this exception, however, all δ_j for weekly data are smaller than their corresponding values of δ_j for daily data (where the former is measured in weeks and the latter in days). In general, the estimated probabilities p_{11} and p_{22} indicate high persistence of states. The asymptotic standard errors, reported in brackets, also show that the transition probabilities have been estimated with quite high precision.

Table 1

Estimates of the Markov-switching model

		mark	pound	sfr	yen
day	σ_1^2	0.137 (0.007)	0.073 (0.008)	0.203 (0.010)	0.068 (0.006)
	σ_2^1	0.786 (0.034)	0.697 (0.031)	1.228 (0.055)	0.595 (0.023)
	p_{11}	0.982 (0.004)	0.933 (0.013)	0.981 (0.004)	0.971 (0.006)
	p_{22}	0.983 (0.004)	0.952 (0.009)	0.978 (0.005)	0.981 (0.004)
	LR_n	874.0 ***	985.6 ***	954.4 ***	1046.5 ***
	$\chi^2(95)$	130.5 **	220.5 ***	157.4 ***	715.2 ***
	LR_m	503.8 ***	396.5 ***	451.0 ***	520.4 ***
	W	19359.4 ***	1934.8 ***	14249.3 ***	11836.1 ***
	β_2	4.60 ***	5.30 ***	4.47 ***	5.17 ***
	$Q(15)$	55.0 ***	215.4 ***	132.6 ***	155.3 ***
week	σ_1^2	0.435 (0.068)	0.232 (0.066)	0.818 (0.105)	0.265 (0.053)
	σ_2^1	3.497 (0.305)	2.588 (0.183)	4.872 (0.464)	2.440 (0.189)
	p_{11}	0.917 (0.025)	0.924 (0.035)	0.948 (0.019)	0.949 (0.026)
	p_{22}	0.936 (0.021)	0.977 (0.011)	0.954 (0.018)	0.972 (0.014)
	LR_n	151.2 ***	121.3 ***	132.4 ***	181.9 ***
	$\chi^2(45)$	73.2 ***	73.8 ***	168.6 ***	70.4 **
	LR_m	72.4 ***	44.1 ***	65.3 ***	84.8 ***
	W	470.2 ***	451.7 ***	756.1 ***	612.8 ***
	β_2	3.81 ***	6.23 ***	3.21	4.75 ***
	$Q(15)$	14.9	94.9 ***	18.9	20.4

Significance levels: $\alpha = 0.01$(***); $\alpha = 0.05$(**); $\alpha = 0.10$(*)

It is also informative to compute the stationary state probabilities ρ as in (4). It is surprising that the low-variance states (this is always state 1), have in almost all cases a smaller stationary probability than the high-variance states. Only for the daily sfr series is ρ greater than 0.5. For daily data ρ varies between 0.417 and 0.539 and for weekly data it varies between 0.228 and 0.470. This result is in contrast to estimates of p for 2-component scales mixtures as reported in section 2.1, table 1. Those estimates were in general much larger than the estimates of p . One would conclude, therefore, from the estimates of p_{11} and p_{22} that both states have roughly the same probability of occurring.

Similar to the results for the scale mixtures, the two states of the estimated Markov-switching models are well separated by their variances: σ_1^2 is larger than σ_2^2 by at least a factor of 5 and, as indicated by the standard errors reported in brackets, the variances can be estimated with quite high precisions. The variances are roughly in the same order of magnitude as those of the scale mixtures. The most noteable difference is for the weekly pound series where $\sigma_1^2=1.09$ and $\sigma_2^2 = 6.23$ in the scale-mixture model but $\sigma_1^2 = 0.232$ and $\sigma_2^2 = 2.588$ in the Markov-switching model. The fact that the variances in both states of the Markov-switching model are much smaller than the corresponding variances in the components of the scale mixture is compensated by the fact that $\rho = 0.23\#2$, whereas the probability of component 1 is 0.81 in the mixture model.

The Markov-switching model is also tested, like all other candidate models, against a H_o of Gaussian white noise with a LR test. The same theoretical caveats concerning the applicability of the LR apply here as in the case of mixture models (see section 2.1), but from a practical point of view those caveats seem to be immaterial here. As table 1 shows, the LR statistics (which are here denoted as LR_n) are so high that it appears to be futile to worry whether 1 or 2 degrees of freedom should be applied to the χ^2 distribution. The LR tests rejects clearly a simple normal distribution in favour of the Markov-switching model. This rejection is not a great surprise since for all models of Chapter 2, this rejection occurred, too. Instead it is more interesting to test the Markov-switching model against the scale-mixture model of Section 2.1 since it is nested within the Markov-switching model. As noted above, the Markov-switching model reduces to the mixture model if $p_{11} = 1 - p_{22}$. The LR test with the mixture model as the H_o, denoted as LR_m in table 1, is therefore a test of independence of the draws from the two states. Since heteroskedasticity was found to be a strong property of the exchange-rate data, one expects to find strong rejections of the restriction $p_{11} = 1 - p_{22}$. This is indeed the case, as table 1 shows. Assuming that the LR statistic has asymptotically a $\chi^2(1)$ distribution, independence is strongly rejected for all daily and weekly series.

Alternatively, one can test the restriction $p_{11} = 1 - p_{22}$ with a Wald test. The test statistic is

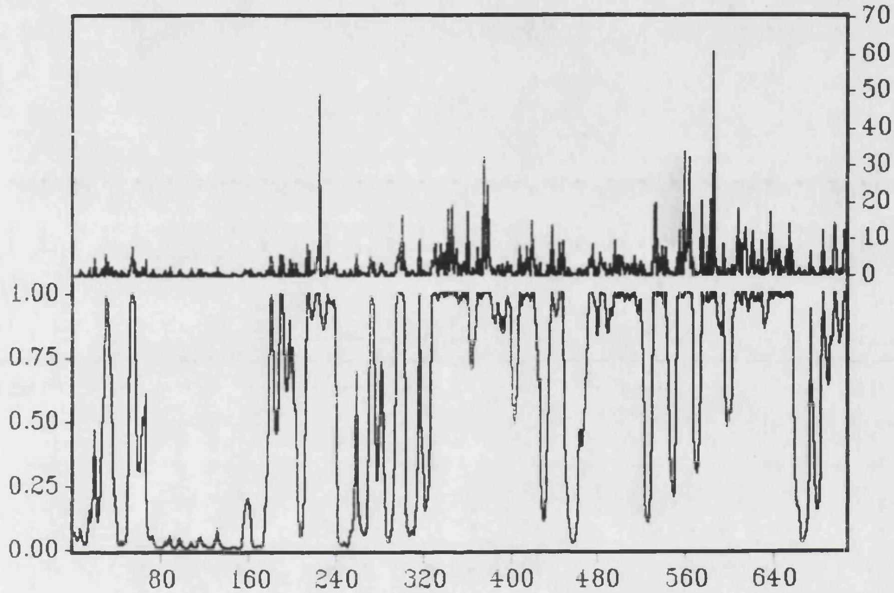
$$W = \frac{[p_{11} - (1 - p_{22})]^2}{var(p_{11}) + var(p_{22}) + 2cov(p_{11}, p_{22})} \tag{17}$$

which has asymptotically a $\chi^2(1)$ distribution. The test statistics which are reported in table 1 appear to be pathologically high but note that the standard errors of p_{11} and p_{22} are very small in table 1. Since both p_{11} and p_{22} are quite close to 1 for all series, it is clear that independence of draws will be rejected very strongly.

As mentioned above, a useful by-product of applying the EM algorithm is the calculation of the smoothed probabilities $p(s_t \mid x_1, \ldots, x_T, \Psi)$ which are conditioned on the information from the full sample. Figure 1 plots the smoothed probabilities of state 2 for the weekly mark series together with the squared exchange-rate dynamics. It is apparent from this figure that state 2 is associated with turbulent periods (i.e. high-variance episodes). The probability smoother identifies the period from November 1975 until November 1977 as the only prolonged period being associated with state 1 if the criterion for identifying state 1 is that $p(s_t = 1 \mid x_1, \ldots, x_T, \Psi) > 0.5$. The corresponding plot of squared data shows that this was also a period of relatively small weekly exchange-rate fluctuations. The only other longer periods of tranquillity are the ones from July until December 1974 and from February until September 1979. This seems to indicate that the early period of the post-Bretton-Woods era was more tranquil than the more recent one. It is also interesting to note that the smoother attaches a probability of 1 to $p(s_t \mid .)$ for both the strongest appreciation of the mark in the sample (the 7.8 percent appreciation in the wake of the Plaza agreement in September 1985) and its strongest depreciation (the 7.0 percent depreciation after the introduction of support measures for the dollar in November 1978).

Figure 1

Squared data and smoothed probabilities of state 2: weekly mark



In order to examine the adequacy of the model, I performed a χ^2 goodness-of-fit test for the implied stationary distribution of the Markov-switching model. As with previous applications of this goodness-of-fit test, the results are quite disappointing. The test rejects the model for all daily and weekly data at least at the 5 percent level. There is even no clear improvement as compared to the fit of the mixture models to the empirical frequency curves. Two cases are quite extreme: the daily yen and the weekly franc. As noted before, the distribution of the daily yen data is strongly peaked since there are 146 days within this sample period on which the yen-dollar rate did not change. The interval which includes the value of zero contributes alone 381.8 to the excessively large value of the χ^2 statistic.

In the case of the weekly franc, the strong rejection of the goodness-of-fit test is largely due to an underestimation of the lower tail probability. The Markov-switching model predicts 14.1 depreciations (of the dollar) of 2.78 percent or more in this weekly data set but 45 depreciations of this magnitude were actually observed.

It is also possible to analyse the "residuals" of the model. The expected variance of each x_t can be derived form the estimated smoothed probabilities according to

$$\tilde{\sigma}_t^2 = p(s_t = 1 \mid x_1, \ldots, x_T, \Psi)\sigma_1^2 + p(s_t = 2 \mid x_1, \ldots, x_T, \Psi)\sigma_2^2. \tag{18}$$

and one can standardize the data x_t to obtain an independent and normally distributed series $\tilde{x}_t = x_t/\tilde{\sigma}_t$ with unit variance. Table 1 shows that there is still significant leptokurtosis in the residuals since the kurtosis β_2 of $\tilde{x}_t$ is significantly greater than 3 for all series. Compared with the kurtosis of the unstandardized data x_t, however, the leptokurtosis is strongly reduced (cf. table 17 in Section 1.3) and the excess kurtosis is more than halved in all cases but one.

Finally, I computed the Ljung-Box statistic $Q_{xx}(15)$ for the squared residuals in order to examine whether the Markov-switching model captures the full amount of heteroskedasticity in the data. Table 1 indicates that this is not the case. There is still significant autocorrelation in the squared residuals of the daily series although it is much smaller than in the unstandardized data where it varies between 355.2 and 561.0 (cf table 6, Section 2.2). In the weekly data, however, it is only the pound series which shows residual heteroskedasticity.

To summarise, the Markov-switching model provides a reasonably good fit to the daily and weekly data although the fit is not completely satisfactory on all accounts, but the model is an improvement to the compound-distribution models of Chapter 2 since, contrary to those models, it captures heteroskedasticity. The next section introduces ARCH models which incorporate heteroskedasticity explicitly and became very popular in empirical finance in recent years.

3.2 *ARCH*-TYPE MODELS

Engle (1982) introduced the model of autoregressive conditional heteroskedasticity (ARCH) which provides a direct approach to model serial dependence in variances. The model can be formulated in the following way. Let a random variable x_t be factored into

(19) $$x_t = \varepsilon_t h_t^{1/2}$$

where ε_t is white noise with unit variance and h_t is a linear variance function given by

(20) $$h_t = \alpha_0 + \alpha_1 x_{t-1}^2 + \ldots + \alpha_s x_{t-s}^2,$$

then this ARCH model is said to be of order s, ARCH (s) for short. In its simplest version, it is assumed that x_t does not have any structure in its mean, i.e. that x_t is the residual of a regression or of a time-series model like the ARIMA model. One does not have to worry too much about the estimation of the mean equation because Engle (1982) established the independence of the mean equation from the variance equation given in (20) by showing that the Hessian matrix is block-diagonal in the parameters of the two equations. Furthermore, the exploratory analysis of chapter 1 has shown that there are no strong and persistent patterns in the mean of exchange-rate dynamics and that the H_0 of a constant mean at zero cannot, in general, be rejected. Hence, for the purpose of this study I take x_t to be the first difference in the logarithm of exchange rates.

Although the model assumes that there is a time pattern in variances, it does not introduce explanatory variables for the variance. The causes of heteroskedasticity are assumed to be unobserved variables which are slowly varying or, to put it differently, which exert persistent effects in variability. The framework of the ARCH model is very useful for empirical applications in financial economics when risk is measured by the variance of a random variable and the model has been applied for this purpose repeatedly (see the comprehensive survey by Bollerslev et al. (1992)). It is applicable to univariate models of speculative prices since the model implies that small and large fluctuations tend to cluster together. Empirically, this phenomenon of speculative price dynamics was already noticed by Mandelbrot (1963). With the Markov chain model of chapter 1, I could show that this property is highly significant in short-run exchange-rate data.

A good example of turbulent periods is provided by the yen-dollar rate in August 1981. On August 4th, the dollar appreciated by 2.21 percent (the 7th largest appreciation in the daily yen series) despite massive intervention by central banks. According to newspaper reports, the dollar appreciated because of a sharp rise in U.S. interest rates and because of political unrest in Poland. The next day, the dollar depreciated by 2.92 percent (the 6th largest depreciation in the series) following continued central bank intervention and a decline in U.S. interest rates but speculators were also reported to take profits.

The clustering of large fluctuations of either sign is often due to great uncertainty among foreign-exchange dealers but sometimes a strong appreciation follows a strong depreciation, or vice versa, because government authorities step in to stop an exchange-rate crisis. A case in point is the dollar crisis in July, 1973. On July, 4th the dollar declined by 3.08 percent against the mark (the 8th largest depreciation in this series) in the wake of the Watergate affair and the imposition of wage and price controls. Following an agreement between central banks to raise swap facilities and the announcement of interventions, the dollar appreciated against the mark by 3.04 percent (the 3rd largest appreciation in this series) on July, 10th.

A priori, nothing can be said about the lengths of periods of turbulence and periods of tranquillity. With the Markov-chain model I only examined first order dependence in variability but the results from the empirical autocorrelation functions for squared data indicate that higher order dependence is present in short-run dynamics. Therfore, it is best to choose the lag-length in the variance function heuristically.

Before I turn to the issue of estimation, it should be mentioned that the ARCH model implies that x_t in (19) has an unconditional leptokurtic distribution. Leptokurtosis for the ARCH(1) model was shown by Engle (1982) and the generalization for the ARCH(s) model was established by Milhoj (1985). Thus, the ARCH model seems to be ideal to be fitted to short-run exchange-rate dynamics since it incorporates both the properties of heteroskedasticity and leptokurtosis. Furthermore, Diebold (1988) showed that ARCH models converge to normality under temporal aggregation.

As regards the estimation of the model, Engle (1982) proposed to estimate the model with ML methods and derived the gradient vector and the Hessian matrix under the assumption that ε_t has a standard normal distribution. The model requires, of course, that h_t be positive. A sufficient condition to ensure positive conditional variance is that all α_j's in (20) are positve. However, I found it advantageous not to impose this restriction because violations of it may indicate abnormalities or a lack of fit of the ARCH model to the data. It turned out that non-positive variance was never a problem in the application to exchange rates.

As regards the identification of the order s, I applied both the Akaike information criterion (AIC), defined by $\text{AIC} = 2s - 2L^*$, and the Schwarz information criterion $\text{SIC} = s \log T - 2L^*$, where L^* is the logarithm of the maximised likelihood. It is well-known that SIC tends to identify models of smaller order than AIC. This is also true in this case as shown in table 2.

Table 2

ARCH models: identification of order s

		mark	pound	sfr	yen
day	AIC	11	20	14	24
	SIC	11	20	12	11
week	AIC	22	17	20	23
	SIC	3	4	6	12

I estimated ARCH models up to order s =25 and identified models of quite high order for daily series according to both information criteria. Only for the daily yen series is the difference in the identified order quite substantial. For all weekly series, however, SIC identifies a much smaller order than AIC. According to AIC, on the other hand, the order increases for two series in weekly data as compared with daily data. Whatever information criterion is selected to identify s, the order s is quite high for both daily and weekly series.[2] This suggests to find a more parsimonious parametrization of the model. There are basically two possibilities. First, one can impose a restriction on the α_j's in the form of linearly or geometrically declining weights (or some other functional form). This approach was suggested by Engle (1982) and was applied to univariate exchange-rate models by Diebold (1988), Hsieh (1988, 1989a) and Lastrapes (1989). Hsieh (1989a), however, found that the restrictions of linearly and of geometrically declining weights were rejected by LR tests for daily exchange-rate data.

The alternative is to find a more parsimonious parametrization similar in spirit to the approximation of a high order MA process by a low order ARMA process. This approach was worked out by Bollerslev (1986). Bollerslev's generalized ARCH model (GARCH) is obtained when (20) is replaced by the GARCH(s,r) model

$$h_t = \alpha_0 + \alpha_1 x_{t-1}^2 + \ldots + \alpha_s x_{t-s}^2 + \beta_1 h_{t-1} + \ldots + \beta_r h_{t-r} \tag{21}$$

but (19) is retained. The idea here is to choose low orders of s and r to approximate a high order ARCH process. Bollerslev's presentation of the GARCH model closely follows the one by Engle for the ARCH model and Bollerslev established for the GARCH(1,1) model the leptokurtosis of x_t.

The choice between the ARCH(s), where s is determined by SIC and the GARCH(1,1) model can again be made by applying information criteria. Table 3 reports the comparison between ARCH and GARCH for the exchange-rate data by SIC Only in

[2] Similar results were obtained by Diebold (1988) and Hsieh (1989a). Diebold estimated ARCH models with fixed $s = 12$ for seven weekly exchange-rate series and found statistically significant α_j s up to this order. Hsieh identified for five daily series the order s with likelihood ratios, AIC and SIC. According to the first two criteria, the optimal s was between 10 and 24 and according to SIC it was between 7 and 23.

one out of 8 cases is SIC lower, and hence better, for the ARCH model than for the GARCH model and this result would clearly favour the GARCH model. However, if AIC is applied then the choice between ARCH and GARCH is less clear-cut. For weekly data, AIC favours GARCH but for daily data it favours ARCH in all series. On the whole, however, the GARCH model seems to be slightly preferable to an ARCH(s) model.[3]

Table 3

Comparison between ARCH(s) and GARCH(1,1) models by SIC

		mark	pound	sfr	yen
day	ARCH	6097.5	6030.6	7323.9	5427.1
	GARCH	6064.3	5981.2	7299.7	5430.1
week	ARCH	2463.5	2454.5	2701.6	2300.3
	GARCH	2442.4	2422.7	2663.0	2237.4

ML estimation of the GARCH(1,1) model is quite straightforward. I applied a gradient method based on analytical first derivatives which are given in Bollerslev (1986). The recursion requires values of x_0^2 and h_0 and I followed Bollerslev's suggestion to set both values equal to the mean of $x_t^2(t = 1, \ldots, T)$. The estimates are shown in table 4.

The parameters α_1 and β_1 can be estimated with quite high precision, especially in daily data. In most cases, α_1 is close to 0.1 and β_1 is close to 0.9. Since the mean lag of conditional variance effects is given by $(1-\beta_1)^{-1}$, the high value of β_1 implies that there is strong persistence in variances. The fact that the sum of α_1 and β_1 is close to 1 leads to the issue of stationarity. Bollerslev (1986) proved that a GARCH(s,r) process is second-order stationary if and only if

[3] Hsieh (1989a) obtained a similar result. For daily series, GARCH(1,1) was better than ARCH(s) in terms of SIC for the yen, the sfr and the Canadian dollar, whereas ARCH(s) was better than GARCH(1,1) for the pound. For the mark, SIC led to indifference.

Table 4

Estimates of the GARCH(1,1) model

		mark	pound	sfr	yen
day	α_0	0.656 (0.128)	0.675 (0.109)	0.856 (0.179)	0.766 (0.120)
	α_1	0.169 (0.015)	0.135 (0.013)	0.143 (0.013)	0.190 (0.021)
	β_1	0.833 (0.013)	0.864 (0.011)	0.857 (0.012)	0.820 (0.017)
	LR	970.3 ***	908.5 ***	1015.1 ***	901.6 ***
	$\chi^2(96)$	181.9 ***	405.9 ***	188.0 ***	699.9 ***
	β_2	4.80 ***	10.58 ***	6.29 ***	17.44 ***
	$Q_{zz}(15)$	22.86 *	8.43	54.08 ***	3.19
week	α_0	1.71 (1.34)	4.16 (1.68)	1.83 (1.22)	0.52 (0.38)
	α_1	0.095 (0.027)	0.114 (0.029)	0.095 (0.017)	0.076 (0.013)
	β_1	0.906 (0.027)	0.878 (0.025)	0.907 (0.016)	0.932 (0.011)
	LR	111.4 ***	103.0 ***	119.3 ***	133.8 ***
	$\chi^2(46)$	98.4 ***	73.83 ***	98.4 ***	88.2 ***
	β_2	4.29 ***	7.03 ***	4.26 ***	9.81 ***
	$Q_{zz}(15)$	29.56 ***	8.03	11.66	4.43

Significance levels: see table 1
Note: the values of α_0 and their standard errors are multiplied by 100.

(22) $$\alpha_1 + \ldots + \alpha_s + \beta_1 + \ldots + \beta_r < 1.$$

According to table 4, there are several series for which $\alpha_1 + \beta_1 > 1$. This violation of the stationarity condition, caused by high values of β_1, has been observed repeatedly in applications of the GARCH model to financial data. This led Engle and Bollerslev (1986) to extend the GARCH model to the case where variances are non-stationary. This integrated GARCH model, IGARCH for short, obtains if the polynomial equation

(23) $$1 - \alpha_1 z - \ldots - \alpha_s z^s - \beta_1 z - \ldots - \beta_r z^r = 0$$

(where z is an auxiliary variable) has at least one unit root, whereas the GARCH model requires that all roots lie outside the unit circle of the complex plane.

I do not want to pursue the idea of integration in variance here further because the statistical properties of the IGARCH model are not yet fully developed (see also the discussion of Engle and Bollerslev (1986) in volume 1 of Econometric Reviews). Furthermore, the analysis of sequential variances in chapter 1 seemed to indicate that, at least for the sfr and the yen series, the variance converges to a finite value. A GARCH model with infinite variance was proposed by de Vries (1991) but he did not derive estimators of the full model.

I also performed a LR test for the GARCH(1,1) model against the H_0 of a (conditionally) stationary normal distribution, i.e. the H_0 implies that $s = 0$ and $r = 0$. The LR test rejects Gaussian white noise very strongly against the GARCH(1,1) model for all series (see table 4). The likelihood ratios for this model are also much higher than those for the scale mixture of two normal distributions and for the compound Poisson process although all three models have three parameters to be estimated. When compared with the four-parameter Markov-switching model, the LR statistic of the GARCH model is smaller for two daily series and all weekly series; see also the next chapter.

On the other hand, the χ^2 goodness-of-fit test strongly rejects the GARCH(1,1) model. Here it is tested whether $\hat{\varepsilon}_t = x_t/h_t^{1/2}$ has a standard normal distribution. The test statistic is significant at the 1 percent level for all data sets. For the previous models there was also strong rejection by the goodness-of-fit test for daily data, but for weekly data, the fit seemed to be much better. A closer look at the discrepancies between expected and actual frequencies reveals that the GARCH(1,1) model underestimates both the mass in the tails and at the centre of the distribution. The extreme test statistic of nearly 700 for the daily yen series is again due to the great number of zeros in this series. For the interval [0.0, 0.0251), the expected frequency is 33.65 but the actual frequency is 148.

The standardized data $\hat{\varepsilon}_t$ are further analysed for residual leptokurtosis and heteroskedasticity. Table 4 shows that the H_0 of mesokurtosis can be rejected at very high significance levels for all series. Even more surprisingly, leptokurtosis increases substantially in $\hat{\varepsilon}_t$ as compared with x_t for both yen series and the daily pound series (cf

Section 2.3, table 17). For the daily yen series, kurtosis more than doubles. This surprising result casts doubts on the appropriateness of the assumption that the conditional distribution of x_t is normal.

Table 4 also reports the Ljung-Box statistic Q for squared residuals at lag 15. As compared with the same statistics for the raw data (cf Section 2.3, table 6), there is a dramatic drop in Q for the residuals of the GARCH(1,1) model. In all cases, the Q for the residuals is less than 1/10 of the Q for the raw data. Only for the daily sfr series is the Q of the residuals significant at the 1 percent level. One may conclude, therefore, that the orders $s = 1$ and $r = 1$ are sufficient to capture the serial dependence of variances.

The fact that both the goodness-of-fit test and the test for mesokurtosis reject the distributional assumptions of the GARCH(1,1) model suggests to replace the conditional normal distribution by some other distribution. Since the comparison between actual and expected frequencies revealed that the model underpredicts the mass in the tails and at the centre of the distribution, it seems appropriate to replace the normal distribution by a leptokurtic distribution.

Bollerslev (1987) introduced a GARCH model with a conditional Student distribution to model speculative prices. In the context of this study of exchange-rate dynamics, it seems very fitting to adopt the Student distribution because this connects the static leptokurtic models with the dynamic process which captures serial dependence in variances. Bollerslev (1987) suggested to estimate the model with an algorithm based on numerical derivatives. Since analytical derivatives should be used, if available, in iterative optimazation, I derive them here for the GARCH(1,1) model

(24) $$h_t = \alpha_0 + \alpha_1 x_{t-1}^2 + \alpha_2 h_{t-1},$$

where, for notational simplicity, $\alpha_2 = \beta_1$. As stated above, it is natural to start the recursion with

(25) $$h_1 = \alpha_0 + \alpha_1 \overline{x^2} + \alpha_2 \overline{x^2}$$

where $\overline{x^2}$ denotes the sample mean of x_t^2.

The derivatives of (24) with respect to the 3 parameters lead to the simple recursions

$$\frac{\partial h_t}{\partial \alpha_0} = 1 + \alpha_2 \frac{\partial h_{t-1}}{\partial \alpha_0} \tag{26}$$

$$\frac{\partial h_t}{\partial \alpha_1} = x_{t-1}^2 + \alpha_2 \frac{\partial h_{t-1}}{\partial \alpha_1} \tag{27}$$

$$\frac{\partial h_t}{\partial \alpha_2} = h_{t-1} + \alpha_2 \frac{\partial h_{t-1}}{\partial \alpha_2} \tag{28}$$

for $t \geq 2$. It follows from (25) that the recursions (26) - (28) are started by

$$\frac{\partial h_1}{\partial \alpha_0} = 1 \tag{29}$$

$$\frac{\partial h_1}{\partial \alpha_1} = \frac{\partial h_1}{\partial \alpha_2} = \overline{x}^2. \tag{30}$$

Applying the approximation of the digamma function introduced in Chapter 2, one obtains for the first derivatives of the log-likelihood at t, l_t, after some arithmetic

$$\begin{aligned} \frac{\partial l_t}{\partial \alpha_i} &= \frac{\partial l_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial \alpha_i} \\ &= -\frac{1}{2}\left[\frac{1}{h_t} - \frac{(\nu+1)x_t^2}{x_t^2 h_t + (\nu-2)h_t^2}\right] \cdot \frac{\partial h_t}{\partial \alpha_i} \end{aligned} \tag{31}$$

for $i = 0, 1, 2$ and

$$\frac{\partial l_t}{\partial \nu} = \log\left(\frac{\nu}{\nu-1}\right) - \frac{1}{2(\nu-2)} - \frac{1}{2}\left[\log\left(1 + \frac{x_t^2}{(\nu-2)h_t}\right) - \frac{\nu+1}{\nu-1} \frac{x_t^2}{x_t^2 + (\nu-2)h_t}\right] \tag{32}$$

where ν denotes the degrees-of-freedom parameter of the Student distribution. Obviously, one has to restrict the parameter space of ν by the condition $\nu > 2$.

The results from estimating this model are shown in table 5 and it is interesting to compare these results with the ones for the GARCH model with a conditional normal distribution as given in table 4. First, the estimated value of α_0, i.e. the constant in the conditional variance equation, is much smaller in the GARCH-t model than in the GARCH model; but this does not indicate that the implied stationary variance of the GARCH-t models is smaller than the stationary variance of the GARCH models. In the case of

Table 5

Estimates of the GARCH (1,1)-*t* model

		mark	pound	sfr	yen
day	α_0	0.004 (0.001)	$3.6 \cdot 10^{-5}$ $(4.3 \cdot 10^{-5})$	0.005 (0.002)	$2.2 \cdot 10^{-4}$ $(2.0 \cdot 10^{-4})$
	α_1	0.164 (0.017)	0.174 (0.023)	0.131 (0.017)	0.217 (0.031)
	α_2	0.847 (0.014)	0.873 (0.012)	0.874 (0.014)	0.842 (0.017)
	ν	7.240 (0.828)	3.737 (0.261)	6.468 (0.664)	3.939 (0.272)
	LR	1105.6 ***	1594.5 ***	1219.9 ***	1614.1 ***
	$\chi^2(95)$	735.0 ***	136.3 ***	922.2 ***	1192.4 ***
	β_2	4.74 ***	146.7 ***	6.53 ***	14.71 ***
	$Q_{xx}(15)$	28.1 **	0.4	59.0 ***	3.3
week	α_0	0.079 (0.058)	$4.0 \cdot 10^{-4}$ $(1.4 \cdot 10^{-3})$	0.034 (0.029)	0.004 (0.005)
	α_1	0.223 (0.080)	0.430 (0.112)	0.175 (0.050)	0.105 (0.029)
	α_2	0.792 (0.072)	0.762 (0.030)	0.847 (0.038)	0.913 (0.021)
	ν	4.088 (0.757)	3.188 (0.451)	4.729 (0.945)	4.183 (0.710)
	LR	163.3 ***	196.0 ***	158.5 ***	236.8 ***
	$\chi^2(45)$	40.9	362.6 ***	197.0 ***	232.6 ***
	β_2	4.38 ***	7.80 ***	5.06 ***	11.18 ***
	$Q_{xx}(15)$	24.0 ***	8.1	13.5	4.8

Significance levels: see table 1

GARCH models, the stationary condition $\alpha_1 + \beta_1 < 1$ is violated for the mark, sfr, and yen series, however, the stationarity condition $\alpha_1 + \alpha_2 < 1$ is violated for all daily and weekly series in the case of the GARCH-t models. Also, the sum of α_1 and α_2 for the GARCH-t models is in all eight series larger than the sum of α_1 and β_1 for the GARCH models. The largest value obtains for the weekly pound where $\alpha_1 + \alpha_2 = 1.192$.

Second, in testing the GARCH-t Model against the H_0 of Gaussian white noise, I obtained remarkably large values of the LR statistics. The values are much larger than the corresponding values of the previous static and dynamic models.

Third, although the LR statistic seems to indicate a very good fit, the χ^2 goodness-of-fit test strongly rejects the GARCH-t model for seven of the eight series. The test statistics are extremely large for the daily mark, the weekly pound and both sfr and yen series.

Fourth, the "residuals" $\hat{\varepsilon}_t = x_t/\sqrt{h_t}$ are analysed for residual heteroskedasticity and leptokurtosis. As regards heteroskedasticity, the results are quite similar to those obtained for the GARCH model. The Ljung-Box statistic $Q_{xx}(15)$ is strongly reduced when compared with the corresponding values of the raw data (cf. Section 2.3, table 6) and it is only the daily sfr series where there is highly significant heteroskedasticity in the residual. It should also be noted, however, that the Ljung-Box statistic is extremely small for the daily pound series, but, of course, the Ljung-Box test is right-sided.

As regards residual leptokurtosis, the results for the GARCH-t model are even worse than the results for the GARCH model. This is somewhat surprising since the conditional Student distribution, as a fat-tailed distribution, was chosen to reduce this residual kurtosis. But, in fact, the leptokurtosis in the residuals of the GARCH-t is for five of the eight series model larger than the leptokurtosis of the data (cf Section 2.3, table 17). An extreme case is the daily pound series where the leptokurtosis of the residuals is 146.7 which is more than 16 times larger than the leptokurtosis of the sample data. Figure 2 explains how this occured.

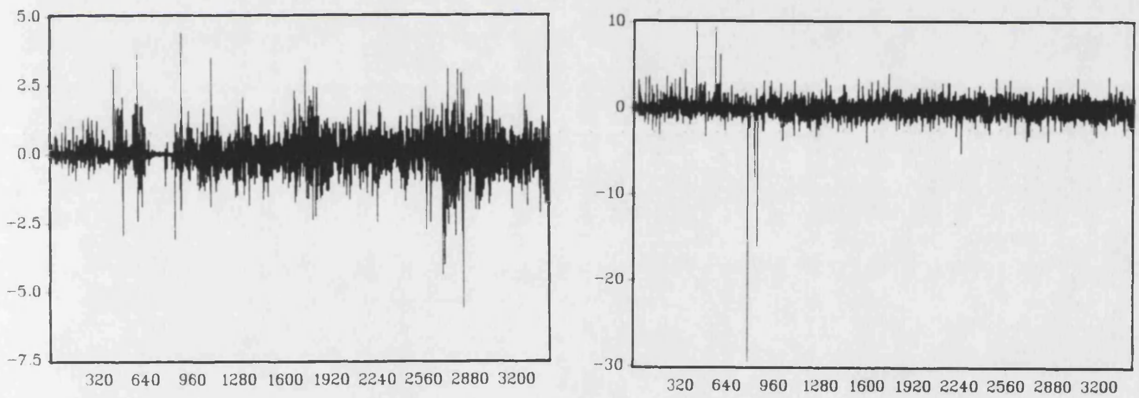
The standardized exchange-rate fluctuations of the daily pound are plotted in the upper panel (with a standardization by the sample standard deviation) and the residual series is shown in the lower panel of figure 2. Both panels have the same vertical scale. A comparison of the two plots shows that the standardization by the conditional variance increases the fluctuations in the residual series compared to the raw data in periods of tranquillity, as in the period from t = 700 to t = 779 (1 April - 27 August, 1977) where the maximum daily exchange-rate movement was 0.087 percent. In this same period the average conditional variance $\bar{h}$ was 0.0026, which is substantially smaller than the average

$\bar{h}$ of 0.5681 for the whole sample and also much smaller than the sample variance of 0.4367. Therefore, in this period the series $\hat{\varepsilon}_t = x_t/\sqrt{h_t}$ shows much larger fluctuation than the series $\varepsilon_t = x_t/\sigma$. At t = 780 (28 August, 1977), however, the dollar depreciated by 1.16 percent against the pound. From the conditional variance equation in (24) it is clear that h_t can only react with a time lag to this sudden volatility shock. In t = 780, h_t was still 0.00155 and this produced a value of $\hat{\varepsilon}_{780} = -29.48$ which appears as a large outlier in the lower panel of figure 2. There are a few other outliers in the residual series which were caused in a similar way and which, together, produce the large leptokurtosis.

Figure 2

Exchange-rate dynamics and GARCH-t residuals: daily pound

a) Exchange-rate dynamics b) GARCH-t residuals



Finally, it is interesting to compare the estimates of ν with the estimated degrees-of-freedom paprameter (df) of the static Student distribution applied in Chapter 2 (cf Section 2.3, table 4). Although there are some differences between these estimates, they are similar in range. The estimates ν'*s* vary between 3.2 and 7.2 whereas the df's vary between 3.8 and 7.1. This implies that the conditions distributions are strongly fat-tailed and peaked. Note, too, that for three series ν is smaller than 4 which means that the conditional kurtosis is not finite. As mentioned above, the stability condition is violated for all eight estimated GARCH-t models and, therefore, the unconditional kurtosis is infinite for all series.

To sum up, the results leave doubts about whether the GARCH-t model is adequate for the exchange-rate data. Although the large LR statistics are quite impressive and there is few residual heteroskedasticity, the goodness-of-fit test produces very poor results and some values of residual leptokurtosis are alarmingly large.

These results are similar to those obtained by Hsieh (1989b) who found that for the daily pound and yen series, the residuals of GARCH(1,1)-*t* models were "extremely ill behaved". In the applications of Bollerslev (1987) and Baillie and Bollerslev (1989), on the other hand, no such extreme results occurred. However, in both studies no goodness-of-fit test was applied.

Another interesting variant of the GARCH model has been introduced by Nelson (1991). He suggested some modifications in the 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. Nelsons exponential GARCH model (EGARCH(r,s)) can be written as:

(33) $$h_t = \exp\left\{a_0 + \sum_{i=1}^{s} a_{ia}\varepsilon_{t-1} + a_{ib}(|\varepsilon_{t-i}| - E|\varepsilon_{t-i}|) + \sum_{j=1}^{r} b_j \log h_{t-j}\right\}$$

where ε_t is white noise with unit variance as in (19). I shall also assume that ε_t has a normal distribution. It then follows that $E|\varepsilon_t|$, the expected value of a half-normal distribution, is equal to $\sqrt{2/\pi}$.

There are several arguments for prefering the functional form of the EGARCH model to the functional form of a simple GARCH model as given in (21). First, the fact that an exponential form is used for the conditional variance h_t guarantees that h_t is always positive. In GARCH models, on the other hand, one has to restrict all coefficients α_i and β_j to be non-negative (and α_o to be positive) in order for h_t to be positive. As a consequence, EGARCH models permit oscillating variance effects, i.e. an increase in h_t may actually reduce some future h_{t+n}. Second, instead of making h_t a function of squared data from the underlying series x_t, as in the GARCH model, the EGARCH has a standard normal variable ε_t as an argument in the conditional variance function. This is a subtlety to simplify the determination of the moments of h_t and of stationarity conditions for h_t. Third, ε_t enters into the conditional variance function in its level and as a deviation of the absolute value of ε_t from its expected value. Both components have, of course, an expected value of zero. The second component captures the size effect of shocks. Since the size effect enters in the form of an absolute value and not as a square, the volatility effects are dampened in comparison to GARCH models. Finally, the fact that the level of ε_t is also included in (33) allows to introduce asymmetric volatility effects. If a_{ia} is negative, then a decrease in ε_{t-1} will have a stronger effect on volatility than an increase. It has been argued that such asymmetric effects can be expected on stock markets (see Nelson (1991)).

However, not all of these advantages of the EGARCH model in comparison with the GARCH model are important for this application to the modelling of exchange-rate volatility. First, negative h_t's were never a problem in the ARCH, GARCH and GARCH-t estimates. Second, the restriction that the coefficients α_i and β_j be positive is only a serious restriction in ARCH models with long lags. In GARCH (1,1) models, which are usually sufficient to capture the dynamics of volatility, the problem of negative coefficients appears virtually never. Third, one would not expect strong asymmetric volatility effects in exchange-rate data since theoretical arguments for these effects are missing and since one cannot find strong skewness in the distribution of the data (see Chapter 1). However, it is probably important that in EGARCH models size effects appear in the form of absolute values instead of squares since the violation of stationarity conditions was the major drawback of GARCH and GARCH-t estimates. For the applications to the exchange-rate data it proved to be sufficient to apply EGARCH (1,1) models which have the stationarity condition $b_1 < 1$. Table 6 reports the estimates.

Table 6

Estimates of the EGARCH (1,1) model

		mark	pound	sfr	yen
day	a_0	-0.027 (0.005)	-0.038 (0.003)	-0.009 (0.004)	-0.055 (0.005)
	a_{1a}	0.005 (0.008)	0.026 (0.006)	-0.022 (0.006)	-0.031 (0.008)
	a_{1b}	0.319 (0.018)	0.300 (0.008)	0.293 (0.014)	0.322 (0.013)
	b_1	0.963 (0.004)	0.940 (0.002)	0.968 (0.003)	0.934 (0.004)
	LR	968.6 ***	930.3 ***	1021.3 ***	978.4 ***
	$\chi^2(95)$	158.1 ***	379.3 ***	178.8 ***	691.8 ***
	β_2	4.99 ***	11.10 ***	6.53 ***	13.65 ***
	$Q_{u}(15)$	24.10 *	8.62	67.17 ***	4.08
week	a_0	0.048 (0.012)	0.050 (0.010)	0.034 (0.009)	0.014 (0.003)
	a_{1a}	0.016 (0.016)	0.030 (0.015)	-0.015 (0.012)	-0.029 (0.011)
	a_{1b}	0.279 (0.035)	0.268 (0.035)	0.198 (0.020)	0.151 (0.017)
	b_1	0.953 (0.013)	0.945 (0.009)	0.975 (0.008)	0.988 (0.004)
	LR	124.1 ***	120.9 ***	118.6 ***	149.7 ***
	$\chi^2(95)$	79.8 ***	75.4 ***	72.0 ***	82.9 ***
	β_2	4.48 ***	6.63 ***	4.47 ***	8.73 ***
	$Q_{u}(15)$	20.13	9.73	9.74	6.25

Significance levels: see Table 1

In contrast to the GARCH models, the EGARCH estimates show no violations of the stationarity conditions. All point estimates of b_1 are below 1 and since the corresponding standard errors, reported in brackets, are small, one may also conclude that the interval estimates are below 1. This is a very reassuring result. The estimates of a_{1b} indicate that there are strong and highly significant scale effects in all daily and weekly series. On the other hand, the asymmetry effects, as given by the estimates of a_{1a}, are much smaller and unsystematic. For the sfr and the yen, the estimates are negative but for the mark and the pound they are positive. When compared with their asymptotic standard error, one can conclude that some of these coefficient are in fact significant. Overall, however, the asymmetry effects are minor.

The LR statistics indicate that Gaussian white noise would very strongly be rejected in favour of the EGARCH model. As with previous models, a χ^2 goodness-of-fit test shows that there are still large differences between the frequency distribution of the data and the frequency distribution implied by the EGARCH model. The standardized variable $\tilde{\varepsilon}_t = x_t / h_t^{1/2}$ should have a standard normal distribution but the χ^2 test rejects this assumption for all daily and weekly series. This rejection is mainly due to an underestimation of peakedness by the EGARCH models, an underestimation of tail probabilities is of secondary importance.

Table 6 presents also some analysis of the residuals $\tilde{\varepsilon}_t$. The most interesting question is, of course, whether there is some residual leptokurtosis and heteroskedasticity in $\tilde{\varepsilon}_t$. Since the χ^2 test rejected the distribution implied by the EGARCH model, mainly due to a lack of peakedness but to some extent also due to lacking fatness of tails, it is not surprising to find significant leptokurtosis in the "residuals" $\tilde{\varepsilon}_t$. Even more disconcerting is the fact that for three series (daily pound and yen, weekly yen) the kurtosis is higher in $\tilde{\varepsilon}_t$ than in x_t. On the other hand, the EGARCH models capture heteroskedasticity quite well. As the Ljung-Box statistic $Q_{xx}(15)$ shows, the squared residuals show no significant autocorrelation up to lag 15, with the daily sfr series being the only exception to this rule.

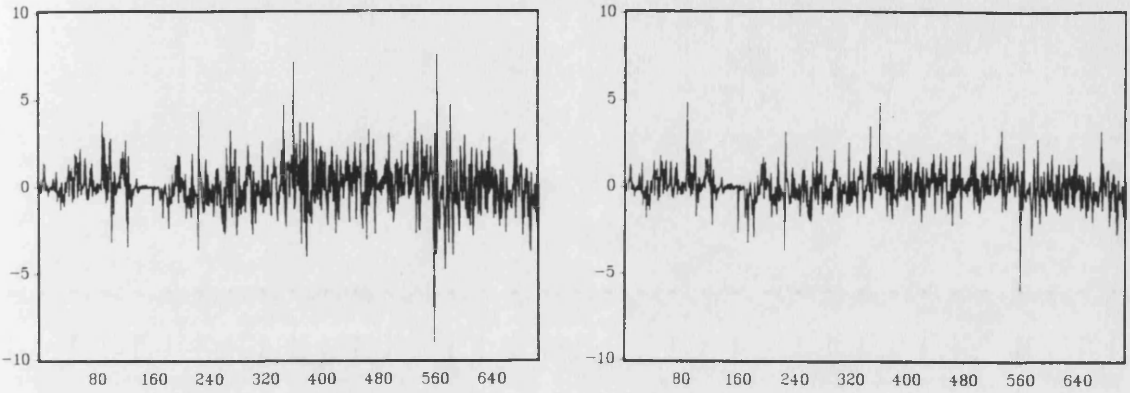
In order to gain more insight into the properties of the EGARCH model, figure 2 plots the exchange-rate fluctuations x_t, for the weekly pound series along with the corresponding "residuals" $\tilde{\varepsilon}_t$. According to the model, $\tilde{\varepsilon}_t$ should have a normal distribution. Figure 2 shows that the magnitude of large exchange-rate fluctuations is indeed very much reduced in the "residuals" as compared with exchange-rate series.

A case in point is the largest weekly depreciation of the dollar against the pound in this sample of -9.01 at $t = 559$. This is reduced to -5.92 in the residual series. Even more dramatic is the reduction of the largest appreciation of 7.60 at $t = 564$ to 2.53 in the residuals. However, there are also a few large residuals which are even larger than the

Figure 3

Exchange-rate dynamics and EGARCH residuals: weekly pound

a) Exchange-rate dynamics　　　　b) EGARCH residuals



corresponding values in the exchange-rate series. An example is at $t = 88$ where $x_t = 3.75$ but the residual is 4.84. The overall impression from figure 2 is, however, that the fluctuations in the residuals are more homogeneous than those in the exchange-rate series.

All in all, the EGARCH model can be regarded as a quite satisfactory model of exchange-rate dynamics which captures both leptokurtosis and heteroskedasticity of the data, although the model does not capture the full amount of peakedness. There seems to be two properties of the functional form of the conditional variance which give the EGARCH models an advantage over GARCH models by dampening the effect of very large exchange-rate movements on the conditional variance. First, the conditional variance process is driven by the standardized variable ε_t and not by the raw data as in GARCH models. The fact that the EGARCH models do not violate the stationarity for the variance can probably be attributed to these two properties. Second, the size effect enters the conditional-variance function in the form of absolute values, whereas it enters the function in the form of squared values in GARCH models.

3.3 SUMMARY

In this chapter, four models have been applied to the modelling of exchange-rate dynamics. In particular, the models are designed to capture the heteroskedasticity of the data but they also imply a leptokurtic conditional distribution. The analysis showed that both the Markov-switching model and all three ARCH-type models (GARCH, GARCH-t, and EGARCH) do indeed capture heteroskedasticity quite well. The heteroskedasticity in the residuals of these models is very much reduced as compared with the heteroskedasticity of the exchange-rate series and most of the Ljung-Box statistics for the squared residuals are insignificant.

Whereas the GARCH and EGARCH models provide a satisfactory fit to the data, the GARCH-t models imply strong violations of stability conditions and give some pathological values for the goodness-of-fit test and residual leptokurtosis.

Although the models imply leptokurtosis, there remains significant residual leptokurtosis for all models. An obvious strategy is to replace the conditional normal distribution by a leptokurtic conditional distribution to improve the fit but this strategy failed with an application of Student's distribution.

One has to conclude from the analysis in this chapter and the previous chapter that none of the models is completely satisfactory on all accounts. It is especially the residual leptokurtosis which is disturbing. However, all models are clearly and very significantly superior to a model of Gaussian white noise. Even if the ideal model did not emerge from this analysis, it is interesting to examine which of the models is best. The next chapter will compare the candidate models according to different criteria.

CHAPTER 4

COMPARISON OF MODELS AND OUTLIER ANALYSIS

In the previous two chapters eight stochastic models of exchange-rate dynamics have been applied. In Chapter 2 four static models, which can be subsumed under scale-compounded normal distributions, have been analysed and in Chapter 3 four dynamic models (the Markov-switching model and three variants of ARCH-type models). It is probably fair to say that none of the models turned out to be fully satisfactory on all accounts. The static models of Chapter 2 lack the ability to capture heteroskedasticity, but although the dynamic models of Chapter 3 are compatible both with leptokurtosis and heteroskedasticity, they did not pass the goodness-of-fit tests very well.

Even if the analysis showed that there is not an ideal model among the eight candidate models, the question remains which of the models is the best. The answer to this question would obviously depend upon the criteria applied. In this chapter 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 I would suggest to dismiss the stable Paretian distributions, because the analysis of distributions with regularly varying tails revealed that the coefficient of regular variation is not smaller than 2 for the exchange-rate series. 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.

The point of departure for the statistical analysis in Chapter 1 was the random-walk model with Gaussian white noise. The random-walk is often applied in international economics and finance to model exchange rates. Therefore, the Gaussian random walk (Gauss, for short) shall serve as a benchmark model to judge the performance of the seven candidate models.

4.1 GOODNESS OF FIT

The candidate models of Chapter 2 and 3 were built around the idea to capture the main stylized facts of the exchange-rate data, namely leptokurtosis and heteroskedasticity. It is, therefore, essential to examine whether the models capture these regularities adequately. In addition, I will summarize the results of the χ^2 goodness-of-fit tests and present a comparison of the models according to the Schwarz information criterion (SIC).

As shown in Chapters 2 and 3, 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. The implied leptokurtosis of the candidate models can be computed from the following formulae:

(1) $$\tilde{\beta}_2(\text{mixture}) = 3\frac{p\sigma_1^4 + (1-p)\sigma_2^4}{[p\sigma_1^2 + (1-p)\sigma_2^2]^2}$$

(2) $$\tilde{\beta}_2(\text{Poisson}) = 3 + \frac{3\lambda\sigma_Y^2}{(\lambda\sigma_Y^2 + \sigma_V^2)^2}$$

(3) $$\tilde{\beta}_2(\text{Student}) = 3\frac{2\eta - 3}{2\eta - 5}$$

(4) $$\tilde{\beta}_2(\text{Markov}), \text{ as in (1) with } \quad p = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}$$

$$(5) \qquad \beta_2(\text{GARCH}) = 3\frac{1-(\alpha_1+\beta_1)^2}{1-\beta_1^2-2\alpha_1\beta_1-3\alpha_1^2}.$$

The formula for the kurtosis of the EGARCH model is more involved. Nelson (1991) derived a formula for the p-th moment in terms of parabolic cylinder functions but, due to computational constraints, I computed kurtosis of the EGARCH models by Monte-Carlo methods.

Table 1 shows the kurtosis of the exchange-rate samples (as in Chapter 1, Table 17) 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 generally underestimate 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 the kurtosis. The estimates of the GARCH model and all estimates of the GARCH-t model with the exception of the daily and weekly pound implied non-stationarity of variances. Therefore, kurtosis cannot be finite for those models. Table 1 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, from (3), that $\eta > 2.5$. This condition is only violated for the daily pound series were $\hat{\eta} = 2.42$. 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 of 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. However, there are some reasons to conjecture that the data-generating process has finite

Table 1

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

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 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.

As explained in Chapter 1, leptokurtosis obtains when there is excessive probability mass, compared to the normal distribution, either in the tails or at the centre of the distribution. Since the non-normality of the distribution is a strong property of short-run exchange-rate dynamics, it is also desirable to have a broader distributional test of the candidate model's adequacy. This is provided in the form of a conventional χ^2 goodness-of-fit test. The results of these tests, which were reported in Chapters 2 and 3, are summarized in Table 2 and are compared with a χ^2 test where the underlying model is a normal distribution with a mean of zero and with the sample variance (denoted as Gauss). The latter test corresponds to the graphical display of figure 11 in Chapter 1.

The quantiles are equiprobable under the corresponding models and there are 100 quantiles for daily data and 50 quantiles for weekly data. The only exception to this rule is the compound Poisson process where it is much easier to make the quantiles equiprobable under the empirical distribution. For this model, the number of quantiles had to be reduced for the daily pound and yen series, and therefore those two numbers are not directly comparable to the other numbers in the table.

The degrees of freedom of the χ^2 test are determined by $n-r-1$, where n is the number of quantiles and r is the number of estimated parameters. The degrees of freedom are reported in brackets in table 2. Since the value of r varies between models, the χ^2 statistics cannot be compared directly but the asterisks indicate significance level.

Table 2 shows again the disappointing performance of all candidate models in the χ^2 goodness-of-fit test. The fit is especially poor for the daily data where all models are rejected at least at the 5 percent level. However with the exception of the GARCH-t model, the candidate models perform much better than the benchmark of Gaussian white noise. The fit of the GARCH-t model is surprisingly poor. For six of the eight series, the χ^2 is even higher than that of the normal distribution.

For weekly data, there is a satisfactory fit for some series with static models whereas all dynamic models are rejected. It is certainly surprising to find the static models outperforming the dynamic models. Among all models, the compound Poisson process seems to achieve the best fit. The poor performance of the dynamic models is mainly caused by an underestimation of peakedness. This is not in conflict with the result that the GARCH models overestimate leptokurtosis since leptokurtosis is either due to fat tails or to peakedness.

Table 2

χ^2 goodness-of-fit tests of candidate models

		mark	pound	sfr	yen
day	Gauss (98)	354.8 ***	705.9 ***	425.7 ***	1042.1 ***
	mixture (96)	167.8 ***	233.9 ***	153.0 ***	645.2 ***
	Poisson (96)	129.7 **	130.6 ***[a)]	146.5 ***	146.0 ***[b)]
	Student (97)	153.6 ***	308.5 ***	129.8 **	710.2 ***
	Markov (95)	130.5 **	220.5 ***	157.4 ***	715.2 ***
	GARCH (96)	181.9 ***	405.9 ***	188.0 ***	699.9 ***
	GARCH-t (95)	735.0 ***	136.3 ***	922.2 ***	1192.4 ***
	EGARCH (95)	158.1 ***	379.3 ***	178.8 ***	691.8 ***
week	Gauss (48)	93.7 ***	103.0 ***	115.0 ***	123.7 ***
	mixture (46)	49.0	79.7 ***	64.6 **	64.6 **
	Poisson (46)	34.1	76.2 ***	61.4 *	53.2
	Student (47)	64.6 **	67.1 **	79.0 ***	57.6
	Markov (45)	73.2 ***	73.8 ***	98.6 ***	70.4 **
	GARCH (46)	68.4 ***	73.8 ***	98.4 ***	88.2 ***
	GARCH-t (45)	40.9	362.6 ***	197.0 ***	232.6 ***
	EGARCH (45)	79.8 ***	75.4 ***	72.0 ***	82.9 ***

Significance levels: $\alpha = 0.01$(***);$\alpha = 0.05$(**);$\alpha = 0.10$(*)
a) based on 80 quantiles and 76 degress of freedom
b) based on 23 quantiles and 19 degrees of freedom

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 3 summarizes the results on the residual heteroskedasticity and compares it to the heteroskedasticity of the data. Heteroskedasticity is here measured as the Ljung-Box statistic at lag 15 of the standardized data

$$\tilde{\varepsilon}_t = \frac{x_t}{\sqrt{\tilde{h}_t}} \tag{6}$$

where x_t is the first difference in the logarithm of the exchange rate and $\tilde{h}_t$ is the estimated conditional variance. As table 3 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.

Table 3

ACF of squared data and residual heteroskedasticity of dynamic model

		mark	pound	sfr	yen
day	data	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	data	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

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 also based on likelihoods and it also corrects for the number of estimated parameters. Table 4 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 4. 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.

Table 4

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)

To summarize the results from the four goodness-of-fit criteria, there is no clear overall ranking of the models. The generalized Student distribution, the GARCH model and the GARCH-t model overestimate the magnitude of leptokurtosis. All models perform quite poorly in the χ^2 goodness-of-fit test but, quite surprisingly, the static models are better than the dynamic models on this account. On the other hand, the dynamic models dominate the static models with respect to the modelling of heteroskedasticity, where the static models fail completely, and with respect to SIC. An overall ranking within the group of dynamic models is not possible since the ARCH-type models capture heteroskedasticity better than the Markov-switching model, but the Markov-switching models achieve on average a better SIC than the GARCH and EGARCH models.

4.2 FORECASTING PERFORMANCE

In this section 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, the 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, usually calculated from the Black-Scholes model (see the next chapter), is smaller than the expected future volatility and they sell options when the implicit volatility is larger than the expected volatility.[1]

Since this study has concentrated on the modelling of variance effects and has neglected mean effects, the forecasting performance will only be evaluated with respect to volatility.

[1] The actual strategy would be that of a straddle where one either simultaneously buys a call and a put option or one sells a call and a put.

Recall, too, from the Introduction that the study of Meese and Rogoff (1983) and many subsequent studies have shown that structural exchange-rate models are not able to outperform the random-walk model in the forecasting of exchange-rate levels.

Although the following forecasting experiments are concerned with the volatility of exchange rates and not with their levels, the methodology of the forecasting exercises is the same as in Meese and Rogoff (1983). 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 τ for the next k periods is given by

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

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 produce constant volatility forecasts. It would be possible to estimate each static model up to time τ and to compute the implied variances from the parameter estimates; I will show, however, in the next chapter that the implied variances of the static models are very close to the historical variances as defined in (8).

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

(9) $$\begin{aligned}\hat{\sigma}_{t+k} &= \sigma_1^2 p(s_{\tau+k} = 1 \mid x_\tau) + \sigma_2^2 p(s_{\tau+k} = 2 \mid x_\tau) \\ &= (\sigma_1^2 - \sigma_2^2) p(s_{\tau+k} = 1 \mid x_\tau) + \sigma_2^2\end{aligned}$$

where σ_1^2 and σ_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_τ. This probability may be decomposed into

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

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

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

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

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

one gets the first-period forecast

$$(13) \qquad \hat{h}_{\tau+1} = a_0 + a_1 x_\tau + b_1 h_\tau$$

which involves only observable variables. For the periods k > 2, the forecasts are

$$(14) \qquad \hat{h}_{\tau+k} = a_0 + a_1 E(x_{\tau+k-1}^2) + b_1 E(h_{\tau+k-1}) = a_0 + (a_1 + b_1)\hat{h}_{\tau+k-1}.$$

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 $\tau = 1$ 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. The forecast horizon includes each of the next 20 weeks for every forecast experiment. This gives 24 forecast experiments for each of the weekly series.

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 5 and 6. Note that the mean errors and RMSE are averaged over all 20 forecast horizons.

Table 5 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 5

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 (2)	-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 (1)	-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 four out of eight cases. If the rankings are aggregated over all eight series, the EGARCH model obtains the best overall ranking of 15 [2], followed by the Markov-switching model with 22, the GARCH model with 24, the GARCH-t model with 29 and the naive model with 30.

Table 6 reports the results for the RMSE criterion. The ranking among models is here 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.

Table 6

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)

[2] It is four times the best model, once the second best, and finishes three times in the third place.

In a way these results extend the results of Meese and Rogoff (1983) for the forecasting of exchange-rate means to the case of exchange-rate volatility. At least with respect to the ARCH-type models, tables 5 and 6 show that non-forecastability extends to volatility. On average, the Markov-switching model performs better than the benchmark model of static variance but numerically this improvement is rather small. It is also interesting to note that the superiority of the random-walk model over asset-market models in forecasting the mean is more obvious with respect to the RMSE than with respect to mean errors in the study of Meese and Rogoff (1983). Since tables 5 and 6 show that the same applies to the forecasting of exchange-rate volatility, there is another correspondence between their results and the results reported here.

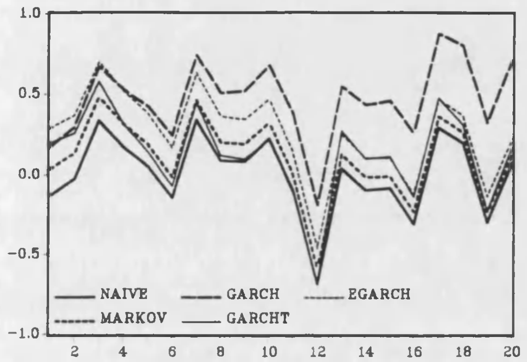
In order to gain more insight into the forecasting performance, figure 1 plots the mean errors and RMSE at forecast horizons 1 to 20 for the daily Swiss franc. It is quite striking how similar the patterns of mean errors and RMSE are across forecast horizons. The plot of mean errors shows how the GARCH models tend to overestimate volatility. Recall from the previous chapter that the GARCH model of the daily Swiss franc 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 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

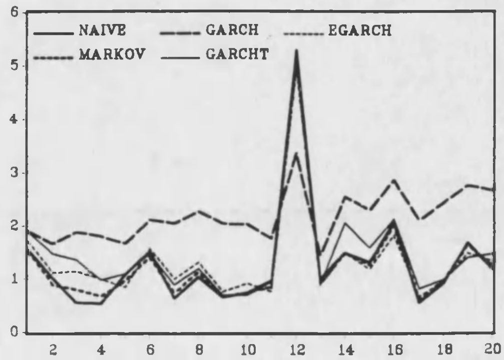
Figure 1

Forecast errors of volatility at different time horizons: daily sfr

a) mean errors



b) RMSE



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. The next section provides a more detailed and systematic examination of outliers.

4.3 OUTLIER ANALYSIS

In the previous two sections, I analysed seven classes of stochastic models which were supposed to capture the main empirical regularities of exchange-rate dynamics. All models are compatible with leptokurtosis but only the dynamic models imply heteroskedasticity. The analysis showed that for short-run dynamics, all models achieve a much better fit to the data than the Gaussian white-noise model. However, the χ^2 goodness-of fit test rejects the models, especially for daily data. It is also quite surprising that the ARCH-type models are not able to outperform naive forecasts of volatility. This might be attributed to a lack of stability in the parameters of these models.

The fact that none of the models is fully satisfactory at closer examination raises the more fundamental question of the adequacy of the approach taken. The basic idea of stochastic models relies on the assumption that there is a unique data-generating process which can be approximated by a simple model. As an alternative to this view, one could regard large exchange-rate fluctuations as being caused by unique events which cannot be subsumed under an all-embracing data-generating process. As an example, the dollar depreciated against the mark on Thursday, April 16, 1986 from 2.3317 to 2.2662 following rumors that Col. Qadhafi died after the bombing of Libyan targets by the U.S. It was argued that "Qadhafi's death would reduce the anxiety in Europe about terrorism and provide less of an incentive to move into the dollar as a refuge" (Wall Street Journal Europe, April 17, 1986, p.13). The fall in the exchange rate by 2.8 percent represents the 7-th largest depreciation in the daily mark series. Under the normal distribution, this depreciation would be regarded as an outlier. Given the empirical standard deviation of 0.68, the lower tail probability of the normal distribution at the standardized value of -4.18 is $0.14*10^{-4}$. Should a stochastic model be able to emcompass events of this kind or is it more appropriate to regard them as outliers caused by unique circumstances?

At this point it is interesting to take a closer look at the extreme observations in the data series in order to understand better the mechanisms which caused them. Table 7 lists the 10 largest appreciations and the 10 largest depreciations of the dollar against the pound in the daily series. The explanations of the strong movements are taken from the foreign-exchange-market reports of the Financial Times. The table reveals that the most extreme currency fluctuations were caused by activities of monetary authorities. The two strong daily appreciations of the dollar in 1978 by 4.39 and 3.49 percent can be attributed to attempts to bring a dollar crisis to an end. The two strongest depreciations of the dollar against the pound occurred in 1985. In February 1985, massive interventions brought the dollar down although the dollar kept on rising for a short while in March. The Plaza agreement in September only served to assure that the dollar continued to depreciate.

The strong appreciations can be attributed to 'regime shifts'. As regards to an increase in swap arrangements, they only strengthen the ability to intervene but within the framework of the 'news model', sketched in Chapter 3, it is readily understood how the creation of a 'policy potential' can have immediate effects. The drastic depreciations of the dollar in 1985 can be explained as a consequence of a bursting bubble. Analysing the strong appreciation of the dollar in the first half of the 1980's, Evans (1986) found strong evidence for the view that the dollar was on a bubble path during this period. The official interventions and agreements triggered the bursting of the bubble and accelerated the depreciation.

Another factor which often appears as a cause for large movements in exchange rates are actual or expected changes in interest rates. A rise in American interest rates is associated with an appreciation of the dollar. Connected to this factor is the influence of economic growth. Agents in foreign-exchange markets seem to relate an increase in growth rates with an increase in interest rates et vice versa.

There is another factor, profit taking by speculators, which is not present in this table but which is given repeatedly as an explanation for large fluctuations in other exchange rates. For instance, 3 of the 10 largest depreciations of the dollar against the yen were associated with the unwinding of speculative positions. These 3 large depreciations occured in December 1978, in December 1979 and in August 1981. Again, this factor can be attributed to the bursting of bubbles but now it is due to market forces.

This examination of outliers shows that the forces behind the most extreme exchange-rate movements can in fact be related to general economic mechanisms. Thus, there is no need to

Table 7

The largest positive and negative exchange-rate changes: daily pound

$100^*\Delta e_t$	*Date*	*Explanation in the press*
4.39	Thr 05.01.78	US Treasury and Fed announce intervention to support the dollar; swap arrangement with Bundesbank
3.65	Mon 25.10.76	Speculation in the press that the terms of Britain's loan from the IMF will include a pound devaluation
3.49	Thr 02.11.78	Sweeping moves to halt dollar decline; Fed increases discount rate, announces supplementary reserve requirements, increases swap arrangements with Germany, Japan and Switzerland
3.20	Thr 04.06.81	High US interest rates; expectations of cut in price of Britain's North Sea Oil
3.11	Mon 08.03.76	Speculation against pound in view of higher UK inflation rate
3.07	Wed 24.04.85	Expectation of higher inflation and interest rates in US; speculators cover short dollar positions
3.03	Fri 02.08.85	Fall in UK interest rates; expectations of lower oil price
2.93	Fri 06.09.85	Financial crisis in South Africa; recovery of economic growth and diminishing prospect of fall in interest rates in US
2.44	Tue 25.08.81	Expectations of higher US interest rates
2.42	Tue 18.09.84	Speculation in favour of dollar ('market impulse'), confidence in strength of US economy
-2.56	Mon 29.04.85	Slow down in US growth; expectation of very large US trade deficit
-2.76	Mon 24.09.84	Intervention of the Bundesbank against the dollar
-2.81	Fri 13.09.85	Unexpected slow down in US industrial production and retail sales
-2.95	Tue 08.06.76	Standby credit to the Bank of England from G10, Switzerland and BIS; UK miners approve pay accord
-2.95	Wed 10.07.85	Expectation of lower US interest rates due to slower economic growth and of cut in US discount rate; expectation of high UK interest rates as long as M3 is outside its target
-3.02	Wed 27.03.85	Expectation of lower US interest rates due to concern on stability of US banking system; slower US economic growth
-3.07	Mon 31.10.77	UK Government stops intervention for holding pound down; sharp rise in official reserves in the UK
-4.03	Tue 19.03.85	Expectation of lower US interest rates due to concern on stability of US savings institutions; tight monetary and fiscal policy announced in the UK budget
-4.41	Wed 27.02.85	Massive interventions by European central banks bring an end to speculation in dollars
-5.60	Mon 23.09.85	Plaza agreement of G5 on concerted intervention

regard the extreme movements as outliers which are unique in their occurrence and cannot be modelled in an economic or stochastic model. The analysis revealed that a successful model should be able to distinguish between different regimes of exchange-rate dynamics because there is strong evidence for periods of tranquility and periods of turbulence.

4.4 SUMMARY

In this chapter, seven candidate models have been compared with respect to their ability to capture the stylized facts of the data and with respect to forecast performance. Only the stable Paretian distributions were dismissed from these comparisons since the analysis in Chapter 2 showed that this model can be rejected to represent the data-generating mechanism of exchange rates.

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. On the other hand, the dynamic models have a natural advantage over the static models because not only do they capture leptokurtosis but also heteroskedasticity. However, it is quite surprising to observe that in forecasting experiments, the dynamic models can only clearly outperform a naive model of constant variances with respect to mean error but not with respect to RMSE.

CHAPTER 5

OPTION-PRICE EFFECTS OF VOLATILITY MODELS

As noted in the Introduction, it is common practice in finance to assume that rates of return and price dynamics in speculative markets follow a normal distribution. However, Mandelbrot (1963) and Fama (1965) produced early evidence against this assumption 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 ground (see e.g. Fama (1976), Ch. 1).

More recently, a renewed interest in distributional properties of financial data emerged. This renewed interest stemmed 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.

Since option pricing is probably that area in finance where the distributional assumptions and the assumption of constant variance are most critical, in this chapter I will be studying the implications of the models introduced in Chapters 2 and 3 for the pricing of foreign-currency options. After a brief description of approaches to price foreign-currency options and of price biases found in analysing the currency-options market, the spot-rate effects and maturity effects of applying three static models of alternative distributions and three dynamic models of heteroskedasticity to the pricing of options will be studied in more detail.

5.1 PRICING OF FOREIGN-CURRENCY OPTIONS

Foreign-Currency options are among those derivative financial instruments which gained enormous popularity and market size during the 1980's. Although over-the-counter options on foreign currencies have been written for many years, it was only in 1978 that foreign-currency options were first traded on an organized exchange (on the European Options Exchange in Amsterdam). In 1982 the first traded contracts were introduced in the USA, on the Philadelphia Stock Exchange. Today, the Philadelphia Stock Exchange is the most important exchange worldwide for the trading of foreign-currency options, both in terms of range and in terms of traded volume. There are not only options of the major currencies (German mark, British pound, Swiss franc, Japanese yen, Australian dollar, ECU, and Canadian dollar against the US dollar) but also recently some cross currency options (British pound vs. German mark; German mark vs. Japanese yen) were introduced. Although most traded options are American style options (options can be exercised at any time up to the expiration date), some European style options (options can only be exercised at the expiration date) are also traded (for the British pound, the German mark, the French franc, and the Japanese yen).

The modern theory of option pricing is based on the approach of risk-neutral valuation introduced in the seminal paper of Black and Scholes (1973). This approach soon became very popular both in the academic world and on financial markets. The academic world was probably most attracted by the fact that the Black-Scholes model derives closed-form option prices without any assumptions about investors' preferences or demand and supply in the market. Practioners from financial markets, on the other hand, found this model probably so attractive because it produced simple option-pricing formulae which could easily be programmed, even on a pocket calculator.

The Black-Scholes model was originally formulated for the case of non-dividend-paying equities but the case of foreign-currency options is a simple extension of it. It is, therefore, instructive to start with the initial model for equities paying no dividends. The model is set up in continous time and the share price S is assumed to follow an Ito process in the form of geometric Brownian motion:

(1) $$dS = \mu S dt + \sigma S dz$$

where μ and σ are parameters and dz denotes a Wiener process with

(2) $$dz = \varepsilon\sqrt{dt}$$

and ε is a random variable with a standardized normal distribution. The model in (1) and (2) says that the instantaneous rate of change in the share price (dS/S) has a normal distribution with a mean of μdt and a variance of $\sigma^2 dt$.

It follows from Ito's lemma that $\ln S\, dt$ has a normal distribution with mean $(\mu - \sigma^2/2)dt$ and variance $\sigma^2 dt$. It can also be shown that S follows a lognormal distribution with mean $S\exp\{\mu dt\}$ and variance $S^2\exp\{2\mu dt\}(\exp\{\sigma^2 dt\} - 1)$. The geometric Brownian motion is the continuous-time equivalent of the discrete-time model of Gaussian white noise which served as the starting point and reference model in the previous chapters.

A call option gives the buyer of the option the right to buy a certain amount of the underlying asset at a predetermined exercise price X. If the price of the call option, c, is only a function of S and time t, Ito's lemma can be applied to (1) to get

(3) $$dc = \left(\frac{\partial c}{\partial t} + \frac{1}{2}\frac{\partial^2 c}{\partial S^2}\sigma^2 S^2\right)dt + \frac{\partial c}{\partial S}dS$$

The crucial idea is now to form a portfolio P of the call option and the equity as in

(4) $$P = \alpha_1 S + \alpha_2 c\ ,$$

where α_1 is the quantity of the equities and α_2 is the quantity of the call options in the portfolio. A change in the value of the portfolio is given by

(5) $$dP = \alpha_1 dS + \alpha_2 dc \,.$$

Inserting (3) into (5) yields

(6) $$dP = \alpha_1 dS + \alpha_2 \left[\left(\frac{\partial c}{\partial t} + \frac{1}{2} \frac{\partial^2 c}{\partial S^2} \sigma^2 S^2 \right) dt + \frac{\partial c}{\partial S} dS \right]$$

It is now possible to choose α_1 and α_2 such that the stochastic terms, which involve dS, are eliminated from (6). This requires that

(7) $$\alpha_1 = -\alpha_2 \frac{\partial c}{\partial S}$$

holds. Condition (7) implies that the portfolio consists either of α_1 units in a long position of the equity and $\alpha_2 \partial c / \partial S$ units in a short position of the call or, alternatively, of α_1 units in a short position of the equity and $\alpha_2 \partial c / \partial S$ units in a long position of the call. Therefore, (6) simplifies to the deterministic equation

(8) $$dP = \alpha_2 \left(\frac{\partial c}{\partial t} + \frac{1}{2} \frac{\partial^2 c}{\partial S^2} \sigma^2 S^2 \right) dt \,,$$

i.e. the portfolio is instantaneously riskless. In a frictionless market, arbitrage will ensure that the rate of return of the portfolio P is equal to the riskless interest rate r, i.e.

(9) $$dP/P = r dt \,.$$

Inserting (4) and (8) into (9) and normalizing α_2 by $\alpha_2 = -1$ yields

(10) $$\frac{\partial c}{\partial t} = r \left(c - \frac{\partial c}{\partial S} S \right) - \frac{1}{2} \frac{\partial^2 c}{\partial S^2} \sigma^2 S^2 \,.$$

This is the Black-Scholes partial differential equation which can be solved with the additional boundary condition

(11) $$c = \max(S - X, 0) \quad \text{when} \quad t = T \,,$$

i.e. at maturity T of the option, the option value will be $S-X$ if the option is in the money and it will be zero otherwise. Solving (10) and (11) is non-trivial but the result is beautifully simple:

(12) $$c = SN(d_1) - e^{-r(T-t)}XN(d_2)$$

with

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

and

(14) $$d_2 = d_1 - \sigma\sqrt{T-t}.$$

In equation (12), N(.) denotes the standard normal distribution function.

It is also instructive to consider an alternative derivation of (12) which is useful later in this chapter. The Black-Scholes differential equation (10) does not include any term which reflects investors preference towards the trade-off between risk and return. In particular (10) is independent of μ, the expected return of the equity. Since a solution to the option-value problem can be found, which is independent of risk preferences, one may also employ a particular preference assumption which simplifies the solution, knowing that the resulting solution generalizes to any arbitrary preference assumption. This is the basic idea of the risk-neutral-valuation principle.

In a risk neutral world, the returns of all assets are equal. The call option price may, therefore, be computed as the present value of its expected value at maturity:

(15) $$c = e^{-r(T-t)}\int_X^{\infty}(S-X)f_L(S)dS$$

where f_L denotes the density of the lognormal distribution. An evaluation of the integral in (15) leads again to (12)-(14).

The valuation of an European call option is easily extended to the valuation of an European put option, which gives the buyer of the option the right to sell a certain amount of the underlying asset at a predetermined exercise price, by means of the put-call parity (see e.g. Hull (1993)):

$$p = c + Xe^{-r(T-t)} - S \,. \tag{16}$$

The extension to American style options is less straightforward. However, if an equity does not pay dividends then it is never optimal to exercise a call option on this equity prior to maturity. Therefore, the right of early exercise is worthless and the American style and the European style call options will have the same price. On the other hand, early exercise can be optimal for all American put options and for American call options on dividend-paying equities. For those cases, the simple Black-Scholes formula does not apply.

If, however, the underlying equity pays a constant continuous dividend q, one obtains a straightforward modification of the Black-Scholes formula for European calls and puts. This follows from the fact that an European option on an equity with price S paying a continuous dividend must have the same price as an European option on a non-dividend paying equity with price $Se^{-q(T-t)}$. Therefore, a simple substitution of S by $Se^{-q(T-t)}$ in (12)-(13) and (16) will give the European call and put option prices for the case of constant continuous dividends.

This reasoning can also be applied to options on foreign currency. An investment in foreign currency earns the continuous interest rate r^* which is assumed to be fixed. Therefore, r^* corresponds to the constant continuous dividend payments q in the case of equities. It follows that the price of an European option on foreign currency is

$$c = e^{-r^*(T-t)} S N(d_1) - e^{-r(T-t)} X N(d_2) \tag{17}$$

with

$$d_1 = \frac{\ln(S/X) + (r - r^* + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \tag{18}$$

and

(19) $$d_2 = d_1 - \sigma\sqrt{T-t}\,.$$

The price of an European put on foreign currency can also be determined by a simple modification of the put-call parity (16):

(20) $$p = c + Xe^{-r(T-t)} - Se^{-r^*(T-t)}\,.$$

The formulae (17)-(20) for the pricing of European options on foreign currency were developed independently by Biger and Hull (1983), Garman and Kohlhagen (1983), and Grabbe (1983); but they are, as shown above, a simple corollary of Merton's model of an equity with constant continuous dividend payments. It is, therefore, somewhat odd that (17)-(20) are often called the Garman-Kohlhagen formulae in the literature and in the markets.

Due to the constant interest receipts of an investment in foreign currency, the pricing of American options is more complicated. It can be shown that early exercise is optimal for American call and put options if the options are sufficiently far in the money (see Grabbe (1983)). Therefore, call and put prices of these American options on foreign currency should be higher than the prices of the corresponding European options but closed form formulae for American options are not available. Furthermore, it can be shown that the American premium above the European option price is a positive function of $r^* - r$ for calls and a positive function of $r - r^*$ for puts. This is easily explained by the fact that the exercise of a call leads to an investment in the foreign currency which yields the interest rate r^* whereas the exercise of a put leads to an investment in the domestic currency which yields the interest rate r.

Table 1 shows option prices from the trading at the Philadelphia Stock Exchange (PHLX) on Friday, 13th August 1993 and is taken from the Wall Street Journal. It reports the spot rate (in cents) of the foreign currency in the first column, the exercise prices in the second column, call-option prices for the expiry months August, September, and December in columns 3-5 and the corresponding put-option prices in columns 6-8. The letter r denotes the case that an option is not traded, whereas s means that the option is not offered. If not otherwise stated, the options are American style.

There are 22 pairs of American and European options with the same maturity and exercise price in the table but only in 10 cases is the American option price larger than the European counterpart. For instance, the European call option to buy 62.500 German marks at 57 cents/mark in August was traded at 1.42 cents/mark whereas the corresponding American option cost 1.47 cents/mark. However, 10 cases violate the condition that the European option price should be the lower bound of the corresponding American option price. A case in point is the call option on the British pound with an exercise price of 1.50 dollars and maturity in September. The European version of this option was priced at 0.65 cents and the American version at 0.52.

There are three explanations for the "mispricings". The first one is trivially that some option prices might be misprints. It is hardly conceivable that this could explain all 12 "mispricings" but there are some obvious candidates for misprints in the table. Take, for instance the American call options on the British pound for delivery in August. The option with an exercise price of $X = 1.45$ traded at $c = 0.80$, for $X = 1.475$ it was $c = 0.05$, and for $X = 1.50$ it was $c = 1.55$. This is grossly inconsistent, and inconsistent is also the option price of $c = 1.55$ for $X = 1.50$ with the option price of $c = 0.52$ for a call option with the same exercise price but with delivery in September.

The second explanation would be that the option prices do not necessarily come from contemporaneous trades. If there are large exchange-rate changes during the day and one option price refers to a trade in the morning and another one to a trade in the afternoon, then this can explain the apparent "inconsistency" of option prices in the table.

Thirdly, the market may not be liquid enough to accommodate large orders and, therefore, option prices would fluctuate strongly. In addition, one would have to assume that transaction costs or a lack of arbitrageurs would let option prices remain inconsistent.

A comparison of option prices in Table 1 certainly cannot provide a rigorous empirical test of price biases in the option market since the prices are not associated with contemporaneous trades. Therefore, I turn to studies which address this issue more rigorously.

Table 1

Prices of Foreign-Currency Options at the Philadelphia Stock Exchange

Friday, August 13, 1993

OPTIONS

PHILADELPHIA EXCHANGE

Option & Underlying	Strike Price	Calls—Last Aug	Calls—Last Sep	Calls—Last Dec	Puts—Last Aug	Puts—Last Sep	Puts—Last Dec
50,000 Australian Dollars-cents per unit.							
ADollr.....	66	r	r	r	r	0.25	r
67.68	67	r	1.02	r	r	r	r
67.68	68	0.05	r	r	0.35	r	r
67.68	70	r	r	0.62	r	r	r
67.68	71	r	0.05	0.37	r	r	r
31,250 British Pound-German Mark cross.							
BPd-GMk	250	r	2.22	r	r	r	r
251.14 ..	252	r	r	r	2.36	2.74	r
251.14 ..	254	r	0.86	r	4.42	4.00	r
251.14 ..	256	r	0.44	r	r	r	r
31,250 British Pounds-European Style.							
BPound .	150	r	0.65	r	4.20	r	r
145.46 ..	137½	r	9.02	r	r	r	r
145.46 ..	145	1.00	r	r	r	r	r
31,250 British Pounds-cents per unit.							
BPound .	145	0.80	2.18	r	0.03	1.73	4.10
145.46 ..	147½	0.05	1.14	r	1.60	3.13	r
145.46 ..	150	1.55	0.52	r	3.90	4.50	r
145.46 ..	152½	r	0.22	r	6.20	6.75	r
145.46 ..	160	r	r	r	r	14.45	r
50,000 Canadian Dollars-European Style.							
CDollar....	77½	r	r	r	1.56	r	r
50,000 Canadian Dollars-cents per unit.							
CDollr	74	r	r	r	r	r	0-
76.19	75	r	r	r	r	r	0-
76.19	75½	r	r	r	r	0-	r
76.19	76	r	r	r	r	r	0-
76.19	77	r	0.17	r	1.00	1.23	r
76.19	77½	r	r	r	1.27	r	r
76.19	78	r	0.05	r	r	2.10	2.60
250,000 French Francs-10ths of a cent per unit.							
FFranc....	16	r	r	r	r	0.76	r
165.18 ...	16¼	r	r	r	r	1.50	4.16
165.18 ...	16½	0.16	1.68	r	0.08	2.60	r
165.18 ...	16¾	r	1.02	r	2.44	4.00	r
165.18 ...	17	r	r	r	4.92	5.96	8.80
165.18 ...	17½	r	0.12	r	r	r	r
165.18 ...	18	r	1.02	r	r	r	r
165.18 ...	18¾	r	r	r	r	1.32	r
250,000 French Francs-European Style.							
FFranc....	16½	0.12	r	r	r	r	r
165.18 ...	18¼	r	r	r	17.32	r	r
62,500 German Mark-Japanese Yen cross.							
GMk-JYn .	61½	r	0.27	s	r	r	s
62.04	62	r	r	0.75	r	r	r
62.04	63	r	r	0.47	r	r	r
62,500 German Marks-European Style.							
DMark	55½	r	r	s	r	0.06	s
58.43	56	r	r	2.56	r	r	r
58.43	57	1.42	1.53	r	r	0.25	r
58.43	58	0.36	r	r	r	r	1.70
58.43	58½	0.06	r	s	0.15	r	s
58.43	59	r	0.40	r	0.60	r	r
58.43	60	r	r	r	1.54	1.88	r
58.43	61	r	0.07	r	r	r	r
58.43	61½	r	0.05	s	r	r	s
62,500 German Marks-cents per unit.							
DMark	54	r	r	4.45	r	r	r
58.43	55	r	r	r	r	r	0.54
58.43	56	r	r	r	r	0.10	r
58.43	57	1.47	r	r	r	0.26	1.22
58.43	57½	r	r	s	r	0.40	s
58.43	58	0.49	0.89	r	0.02	0.57	1.70
58.43	58½	0.03	0.56	s	0.03	0.83	s
58.43	59	0.01	r	1.04	0.57	r	r
58.43	60	0.01	0.16	r	1.57	1.88	r
58.43	60½	r	r	s	2.16	2.40	s
58.43	61	r	0.08	r	2.55	2.68	r
6,250,000 Japanese Yen-100ths of a cent per unit.							
JYen......	81	s	16.70	r	s	r	r
97.98	82	r	15.58	r	r	r	r
97.98	85	r	r	r	r	0.85	r
97.98	88	r	r	r	r	r	0.24
97.98	91	6.99	r	r	r	r	0.51
97.98	91½	r	r	s	r	0.10	s
97.98	92	5.98	r	r	r	r	r
97.98	93½	r	r	s	r	0.23	s
97.98	94	4.00	4.26	r	r	0.32	1.22
97.98	94½	r	3.65	s	r	0.41	s
97.98	95	2.70	3.44	r	r	0.56	1.44
97.98	95½	1.90	r	s	0.03	0.64	s
97.98	96	1.92	2.41	3.72	r	r	1.89
97.98	96½	1.21	r	s	r	0.88	s
97.98	97	0.73	1.79	3.04	0.03	1.03	r
97.98	97½	0.51	r	s	r	r	s
97.98	98	0.08	1.34	2.70	0.03	r	2.68
97.98	99	r	0.92	r	r	r	r
97.98 ...	100	0.01	0.73	r	r	r	r
97.98 ...	102	r	0.31	r	r	r	r
97.98 ...	103	s	r	1.04	s	r	r
97.98 ...	104	s	0.13	r	s	r	r
6,250,000 Japanese Yen-European Style.							
JYen......	94	3.50	r	r	r	r	1.05
97.98	95	2.50	r	r	r	r	r
97.98	96	1.62	r	r	r	r	r
97.98	97	0.95	r	3.27	r	r	r
97.98 ...	100	r	r	1.77	r	r	r
62,500 Swiss Francs-European Style.							
SFranc....	63	2.70	r	r	r	r	r
65.63	64	1.70	r	r	r	r	r
65.63	65½	r	r	s	0.20	r	s
62,500 Swiss Francs-cents per unit.							
SFranc....	64½	r	r	s	r	0.46	s
65.63	65½	0.12	0.92	s	0.02	r	s
65.63	66	0.03	r	r	0.42	r	r
65.63	66½	r	0.55	s	0.80	1.55	s
65.63	67	r	r	r	r	1.72	r
65.63	70	r	0.03	r	r	r	r

THE WALL STREET JOURNAL EUROPE

5.2 PRICE BIASES OF FOREIGN-CURRENCY OPTIONS

Empirical studies on the pricing of foreign-currency options can be subsumed under three headings. First, there are studies which examine the potential mispricing of options according to the modified Black-Scholes formula (17) or the corresponding formula for American-Style options. Differences between model prices and market prices are analysed for systematic price biases and profit opportunities from investment strategies exploiting the price differences. These studies can shed light on the questions of whether the modified Black-Scholes model is appropriate for the pricing of foreign-currency options and whether the foreign-currency options market is efficient.

Second, several studies deal with the differences between American and European foreign-currency option prices. Table 1 shows that the majority of foreign-currency options traded at the PHLX are American style. As mentioned above, there are no closed-form price formulae for American-style options and, therefore, the calculation of option prices is more involved for these options than for European-style options. Since prices of American-style options are often approximated by the European price formulae, it is interesting to ask how big the approximation error is.

Third, several authors investigate the consequences of replacing the assumption of geometric Brownian motion in (1) (resp. of Gaussian white noise for the discrete-time exchange-rate fluctuations) by alternative processes or distributional forms. These studies are motivated by the fact that there is strong empirical evidence against the assumption of Gaussian white noise.

Turning first to the questions of market efficiency and price biases, a straightforward approach is to search for unexploited arbitrage conditions. It is only with arbitrage conditions that market efficiency can be studied without the auxiliary hypothesis of a specific market model. Bodurtha and Courtadon (1986) examine two boundary conditions which are satisfied in an arbitrage-free market: (a) the time value of an option is non-negative and (b) the put-call parity holds. The time value is defined as the difference between the option price and the "moneyness" of the option, where moneyness measures how much the option is in-the-money.[1] For American foreign-currency options the put-call parity takes the form of inequalities (see Grabbe (1987)). Bodurtha and Courtadon (1986) report that only 31 option trades out of 52509 violate the time-value condition and only 1 put-call pair out of 3998 violates the put-call boundary if the data on the options, the spot exchange rates and the interest rates are simultaneous and if transaction costs are taken into account. However, many violations of the boundary conditions were found if closing prices were used and if transaction costs were neglected. Their results demonstrate that rigorous efficiency test cannot be conducted with data based on the table of the Wall Street Journal because spot-rate and option-price data would in general be non-synchronous.

The question of systematic option-price biases has mainly been analysed with data from the PHLX. Goodman, Ross and Schmidt (1985) use daily closing prices from the Wall Street Journal to compare market prices. They find that the market overprices options for all currencies they analyse (mark, pound, Swiss franc, Yen, and Canadian dollar) relative to the theoretical model. However, they use the wrong model (European options) for the data (American options) and this may explain their findings to a certain extent.

Shastri and Wethyavivorn (1987) also use the European pricing formula for American-style call options at the PHLX but they secure against approximation biases by excluding foreign currencies whose interest rate is higher than the domestic interest rate. The authors examine price biases in terms of implied volatilities, calculated from an inverted Black-Scholes formula, and find a U-shaped pattern of implied volatility with respect to the moneyness ratio S/X for short-maturity call options, i.e. the implied volatility is relatively

[1] It is defined as $S-X$ for in-the-money call options and as $X-S$ for in-the-money put options. Otherwise it is zero.

low for at-the-money options compared to out-of-the-money options and in-the-money options. This pattern of implied volatilities is sometimes called the "smile effect". Shastri and Wethyavivorn claim that a diffusion-jump process (i.e. the compound Poisson process of Section 2.2) would be compatible with this implied-volatility pattern.

Tucker (1985) and Shastri and Tandon (1986a, 1987) implement trading strategies to exploit differences between theoretical prices and actual market prices at the PHLX. However, Tucker (1985) and Shastri and Tandon (1986a) apply the European option-pricing formulae to these American options and only Tucker (1985) and Shastri and Tandon (1987) use synchronized data and consider transaction costs. Whereas Shastri and Tandon (1986a) claim to have found abnormal profit opportunities for traders in the market if they trade at prices reported in the Wall Street Journal, Tucker (1985) and Shastri and Tandon (1987) show that excess profits disappear when transaction costs are taken into account.

The appropriate American option pricing model is used in the study of Bodurtha and Courtadon (1987) together with synchronized data to examine the pricing of foreign-currency options at the PHLX. The authors find that the model prices are on average higher than market prices but that out-of-the-money calls are undervalued by the American option-pricing model relative to the market. Furthermore, the relative pricing error decreases with maturity. Bodurtha and Courtadon conjecture that the pricing biases are related to leptokurtic distributions of the exchange-rate fluctuations or by the presence of a diffusion-jump process.

Similarly, Chesney and Loubergé (1987) find for an over-the-counter options market in Geneva that the modified Black-Scholes formula tends to overprice European call options on the dollar/Swiss franc spot rate. Chesney and Loubergé offer three explanations for their findings: (a) traders in the market systematically underestimated future volatility in the market; (b) the assumptions of constant interest rates and volatilities are invalid; (c) the exchange rate follows a diffusion-jump process.

Turning next to studies which deal with the price differences between American and European options on foreign exchange, one finds quite uncontroversial results. Shastri and Tandon (1986b) and Adams and Wyatt (1987, 1989) compare American and European option prices and find, not unexpectedly, that the American premium is larger for in-the-money options than for out-of-the-money options. Furthermore, the premium is a positive function of r^* for calls and a positive function of r for puts. Shastri and Tandon (1986b) also claim that the premium is larger for puts than for calls but this result is probably due to the fact that they use higher domestic interest rates than foreign interest rates. Similarly, Fabozzi, Hauser and Yaari (1990) show that the percentage premium of the American foreign-currency option over the European option is a positive function of maturity and a negative function of volatility.

Hilliard and Tucker (1991) analyse transactions data from the PHLX to see whether the price differences between American and European calls and puts are non-negative. The authors find that in only 25 out of 5886 cases this condition is violated if transaction costs are neglected but that no violation occurs if transaction costs are included. One can deduce from this result that the "violations" of the American-European inequalities for option prices in table 1 are obviously caused by the non-synchronity of the data. Hilliard and Tucker also report that the average market premia of American options over European options are 2.17 percent for calls and 1.38 percent for puts and that the correlation between market premia and model premia is only 0.46.

Mixed results are obtained by Fabozzi, Hauser and Yaari in their comparison of the American and European price formulae to fit to market prices at the PHLX. Whereas the American formula is superior to the European one for the pricing of in-the-money calls when $r^* > r$, it is inferior to the European formula for the pricing of in-the-money puts when $r > r^*$, and the later result is somewhat surprising.

Finally, there are several studies which compare option prices calculated from the modified Black-Scholes formulae with option prices under alternative stochastic assumptions for the exchange-rate fluctuations. Tucker, Peterson and Scott (1988) and Melino and Turnbull (1991) applied the constant-elasticity-of-variance (CEV) model to the pricing of foreign-currency options. The CEV model generalizes the Black-Scholes model by replacing the geometric Brownian motion of (1) by the process

$$dS = \mu S dt + \sigma S^{\beta/2} dz \,. \tag{21}$$

Obviously, the geometric Brownian motion obtains when $\beta = 2$. If $\beta > 2$, there is a positive association between exchange-rate movements dS and their volatility, whereas this association is negative for $\beta < 2$. A further motivation for the CEV model is derived from the fact that (21) implies a leptokurtic distribution of dS/S.

In their estimation of β, Tucker, Peterson and Scott find that β is significantly different from 2 for 26 out of 30 cases (6 exchange rates in 5 years each) and that β is larger than 2 for 22 cases. Melino and Turnbull, on the other hand, report values of β which are mainly below 2 but they do not estimate β directly. In a comparison between market prices and model prices, Tucker, Peterson and Scott show that the CEV model predicts option prices better than the Black-Scholes model for time horizons of 1-3 days but not so for longer horizons. They attribute this finding for longer horizons to the fact that β appears to be intertemporally unstable.

In some of the studies on price biases (Shastri and Wethyavivorn (1987), Bodurtha and Courtadon (1987), and Chesney and Loubergé (1987)), it was conjectured that an option pricing model, which replaces the assumption of geometric Brownian motion (equation (1)) by the assumption of a diffusion-jump process, would fit market prices better than a modified Black-Scholes model as given in (17)-(19) or a corresponding American price formula. A diffusion-jump process is given by:

$$dS/S = (\mu - \lambda v)dt + \sigma dz + dy \tag{22}$$

where λ is the frequency of jumps, ν is the average jump size and dy is a Poisson jump process independent of the Wiener process dz. As mention above, the discrete-time equivalent of this continuous-time process is the compound Poisson process which was analysed in Section 2.2. If it is assumed that the jump risk associated with dy is diversifiable, i.e. that it represents nonsystematic risk, European call option prices can be obtained from

$$c_p = \sum_{j=0}^{j} \frac{e^{-\lambda'\tau}(\lambda'\tau)^j}{j!} c_j \tag{23}$$

where $\lambda' = \lambda(1+\nu)$ and $\tau = T - t$. In (23) c_j is the Black-Scholes option price for a volatility of $(\sigma^2 + j\delta^2/\tau)$ where δ^2 is the variance of the normally distributed jump process. Borenzstein and Dooley (1987), Jorion (1988), and Tucker (1991) apply this model to the pricing of foreign-currency options and compare it to the modified Black-Scholes model.

Actually, Borenzstein and Dooley (1987) use a pure jump model in which $\mu = 0$ and $\sigma dz = 0$. Analysing option prices from the PHLX, they observe that the modified Black-Scholes model underprices out-of-the-money call options and that the price biases are substantially reduced if option prices are calculated assuming a pure jump process. Jorion (1988) and Tucker (1991) show that a model of Gaussian white noise is rejected in favour of the compound Poisson process by a likelihood-ratio test and confirm the finding of Borenzstein and Dooley that a jump model gives option prices which are closer to market prices than those derived from the Black-Scholes formula. However, Jorion reports that only for in-the-money options with short maturities is the option price from the compound Poisson process significantly different from, i.e. larger than, the Black-Scholes price.

A third class of alternative models for the underlying exchange-rate process has been applied by Chesny and Scott(1989) and Melino and Turnbull (1990) for the pricing of foreign-currency options. They assume that volatility is stochastic and that it can be described by the Ornstein-Uhlenbeck process.

(24) $$d \ln \sigma = (\phi + \beta \ln \sigma) dt + \gamma dw$$

where ϕ, β and γ are parameters and dw is a Wiener process. Equation (24) supplements (1) which describes the dynamics of the exchange-rate level.

With volatility σ being stochastic a complication arises in the pricing of options because volatility is not a traded asset. Therefore, no portfolio can be constructed which could eliminate the volatility risk and the principle of risk-independent valuation, which is the major achievement of the Black-Scholes approach, would have to be abandoned. It can be shown, however, that European option prices under stochastic volatility can be obtained from an integration of the Black-Scholes price over the distribution of the average volatility during the life of the option if dw and dz are uncorrelated (see Hull (1993)). Chesney and Scott impose this assumption but Melino and Turnbull examine the more general case where the risk premium on volatility is non-zero.

Melino and Turnbull motivate the application of this stochastic-variance model by the fact that the model of (1) and (24) implies both leptokurtosis and heteroskedasticity. However, they admit that the specification of the volatility as an Ornstein-Uhlenbeck process is ad hoc. In their estimation of the model by the generalized method of moments they find strong evidence for the stochastic specification of volatility and a negative correlation between dz and dw. Melino and Turnbull report that the stochastic volatility model gives a better fit to actual option prices of the Canadian dollar at the PHLX than a constant-volatility model if the risk premium on volatility is assumed to be negative. However, Chesney and Scott find for European-style options of the US dollar against the Swiss franc traded in Geneva that a stochastic-volatility outperforms the Black-Scholes model only if historical estimates of volatility are used in the Black-Scholes model, whereas the Black-Scholes model has smaller price errors than the stochastic-volatility model if implied volatility estimates are used in the Black-Scholes formula and are revised daily

To summarize, the empirical studies of foreign-currency option prices, mainly at the PHLX, have revealed that substantial and systematic differences exist between model prices and market prices and that these differences cannot only be attributed to the pricing of American options with models of European options. It seems rather that there is a systematic smile effect even when the options are evaluated with an American option model. According to this smile effect, implied volatilities of at-the-money-options, as calculated from the Black-Scholes model, are lower than implied volatilities of out-of-the-money and in-the-money options. In terms of price biases, the smile effect implies that market prices are systematically above Black-Scholes prices for out-of-the-money and in-the-money options whereas the opposite holds for at-the-money options.

5.3 SIMULATION OF OPTION PRICES

In this section I will examine the implications for the pricing of foreign-currency options which follow from the stochastic models of Chapter 2 and 3. As in Chapter 4, I will drop the stable Paretian distributions from the list of candidate models because they were clearly rejected by the analysis in Section 2.4. In addition it was necessary to drop the GARCH-t model because this model produced highly erratic option prices. This was caused by the fact that the estimates of this model implied non-stationary variances for all exchange-rate series. Since the conditional Student distribution has fatter tails than a normal distribution, any draws from the tails will eventually lead to explosive behaviour of the simulated series. A case in point are the parameter estimates of $a_1 = 0.174$ and $a_2 = 0.873$ for the daily pound series violating the stationary condition $a_1 + a_2 < 1$. The simulation (with 20 000 repetitions) of option prices from these parameters for at-the-money options with a maturity of 84 days gave an unrealistic option price of \$ 43.96 whereas the simulated Black-Scholes price was \$ 0.04 and the simulated GARCH option price was \$ 0.14. For longer maturities, the simulations produced some even more unrealistic prices.

On the other hand, the model of stochastic volatility, introduced in the previous section, will not be included in the analysis of this section for two reasons. First, the specification of the continuous-time model in (24) is ad hoc, as admitted by Melino and Turnbull (1990), and motivated by its simplicity rather than by a study of observed prices, as noted by Taylor (1992). Second, the stochastic-volatility model is very close to GARCH models in its statistical properties (see Taylor (1992)) and, in fact, in its discrete-time version it is compatible with an GARCH (0,1) model. A discrete-time version of (1) and (24) can be formulated as:

(25) $$\Delta \ln S_t = \eta + \sigma_{t-1} u_t$$

(26) $$\ln \sigma_t = \zeta + \theta(lu\sigma_{t-1} - \zeta) + \psi v_t$$

where η, ζ, θ and ψ are parameters and u_t and v_t are white-noise error terms. Note that

(27) $$E(\ln \sigma_t \mid I_{t-1}) = \zeta + \theta(\ln \sigma_{t-1} - \zeta)$$

where I_{t-1} is the information set at $t-1$. Equations (25) and (27) come close to a GARCH (0,1) model. The major difference between a stochastic-volatility model and a GARCH model lies in the way in which the price dynamics $\Delta \ln S_t$ impinge on the volatility process. In the stochastic-volatility model, $\Delta \ln S_t$ is only related to σ_t if u_t and v_t are correlated, whereas in GARCH models lags of the squared price dynamics enter directly into the volatility equation. A definite advantage of ARCH-type models is the fact that they can easily accomodate a rich dynamic structure.

As mentioned in the previous section, a major drawback of replacing the assumption of Brownian motion in option-pricing models by alternative distribution models or hetero-skedastistic proceses is the fact that the Black-Scholes approach of constructing a perfect-hedge strategy and deriving pricing formulae, which are independent of investors' preferences, is not feasible under these alternative assumptions. The common feature of the distribution models of Chapter 2 and of the heteroskedasticity models of Chapter 3 is the stochastic nature of variance in these models. As explained in Chapter 2, the scale-mixture of distribution, the compound Poisson process, the generalized Student distribution and the symmetric stable Paretian distributions can be viewed as scale-compounded normal distributions where the variance has an independent distribution. In this sense both the static models of Chapter 2 and the dynamic models of Chapter 3 introduce an additional source of uncertainty into the modelling of the price process. However, since the variance is an unobservable variable and cannot be traded on financial markets[2], there is no way to eliminate this risk by a simple hedge strategy.

The fact that preferences of investors come back into option-pricing models has also been realized in the stochastic-volatility literature (see Wiggins (1987)). These preference parameters can only be eliminated with additional assumptions. Hull and White (1987) assume that volatility is uncorrelated with aggregate consumption, i.e. that volatility risk is not priced, and that volatility is uncorrelated with the price of the underlying asset. Wiggins (1987) deals with the more general case where volatility and price movements are correlated, but imposes the restrictions that in investors have logarithmic utility functions and that the partial correlation between the market return and the volatility of the asset is zero. It is interesting to note that Hull and White find a "smile effect" with their stochastic -variance model, i.e. relative to Black-Scholes prices, their model produces lower prices for at-the-money call options and higher prices for in-the-money and out-ot-the-money call options. On the other hand, Wiggins finds negative correlations between American stock returns and their volatilities. Wiggins' stochastic-volatility model with this negative correlation implies that the Black-Scholes model

[2] Actually, in the jargon of market participants some investment strategies are described as "trading volatility", but these strategies mean that assets whose implied volatilities appear to be low, are bought, and those whose implied volatilities are regarded to be high are sold.

would overprice out-of-the-money call options and underprice in-the-money call options. Recall that the EGARCH model of Section 3.2 allows for correlation between price movements and volatility, but that no significant effect was found in the exchange-rate data.

The easiest way to avoid the analytical problems of having an additional source of risk which cannot be hedged, is simply to assume that investors are risk neutral and this approach will be adopted here. It is possible to impose alternative assumptions which would lead to the risk-neutral valuation principle. In particular, Duan (1991) has shown for the GARCH model that the risk-neutral valuation principle would also hold under the conditions that either utility functions imply constant relative risk aversion and changes in logarithmic aggregate consumption follow a GARCH process or that utility functions imply constant absolute risk aversion and changes in aggregate consumption follow a GARCH process.

In this section I will derive prices of European call options along the lines of (15) where the option price is computed as the present value of the option's expected value at maturity. For simplicity, I also set the domestic and foreign interest rates equal to zero since they would only enter as scale factors into option prices. Of course, one cannot hope to derive analytically the density function of the spot exchange rate at maturity for the distributional models and heteroskedasticity models. Therefore, option prices are computed by simulation based on the expected value of the boundary condition at maturity, i.e. European call option prices were computed as

(25) $$c = \frac{1}{K}\sum_{k=1}^{K} \max\{S_k - X;0\}$$

where S_k is the terminal spot rate in the k-th simulation, X is the exercise price and K=20000 is the number of repetitions in every experiment.

The main ojective is to examine whether the static models of Chapter 2 and the dynamic models of Chapter 3 imply any systematic differences from Black-Scholes prices. The Black-Scholes prices can be computed analytically but in order to reduce the impact of sample variation they, too, were simulated in Monte-Carlo experiments. The simulations require draws from a normal distribution for the mixture distribution, the compound Poisson process, the Markov-switching model, the GARCH model, and the EGARCH model. In addition, draws from a uniform distribution are required for the mixture distribution and the Markov-switching model. An algorithm for sampling from a Student distribution was taken from Kinderman et al. (1977).

The simulations are based on the parameter estimates of the daily pound series[3] because this is the only daily series for which the condition of finite stationary variances of the GARCH model was not violated[4]. The parameter estimates along with the (stationary) variances of the estimated models are shown in table 2. The implied variances were computed from the following formulae:

(29) $$\tilde{\sigma}^2(\text{mixture}) = p\sigma_1^2 + (1-p)\sigma_2^2$$

(30) $$\tilde{\sigma}^2(\text{Poisson}) = \lambda\sigma_Y^2 + \sigma_V^2$$

(31) $$\tilde{\sigma}^2(\text{Student}) = \frac{\gamma^2}{2\eta - 3}$$

(32) $$\tilde{\sigma}^2(\text{Markov}):\quad \text{as}\quad \text{in (29)}\quad \text{with}\quad p = 1 - \frac{p_{22}}{2 - p_{11} - p_{22}}$$

(33) $$\tilde{\sigma}^2(\text{GARCH}) = \frac{\lambda}{1 - \alpha_1 - \beta_1}$$

(34) $$\tilde{\sigma}^2(\text{EGARCH}) = \exp\left\{\frac{a_0}{1 - b_1}\right\}.$$

---

[3] Since the data were analysed in the form of $100\Delta \ln S_t$, one has to rescale some parameter values to remove the impact of the factor 100. For most models, this is done in an obvious way. Note, however, that the rescaling is $a'_0 = a_0 - (1 - b_1)\ln(10000)$ in the case of the EGARCH model.

[4] Simulations were also performed with parameter estimates from the weekly pound series. The results do not differ in any important way from the results of the daily series.

Table 2 also shows the kurtosis of the models as reported in table 1 of Chapter 4. For comparison, table 2 gives also the sample variance and the kurtosis of 3 for the normal distribution under the heading of Gauss.

Table 2

Parameter estimates, variance and kurtosis: daily pound

Model	Parameter estimates	Variance	Kurtosis
Gauss	$\sigma^2 = 0.437$	(0.437)	3.00
Mixture	$p = 0.460$ $\sigma_1^2 = 0.061$ $\sigma_2^2 = 0.756$	0.437	4.86
Poisson	$\lambda = 1.262$ $\sigma_\gamma^2 = 0.307$ $\sigma_v^2 = 0.027$	0.414	5.07
Student	$\mu = 2.423$ $\gamma = 0.887$	0.426	∞
Markov	$p_{11} = 0.933$ $p_{22} = 0.952$ $\sigma_1^2 = 0.073$ $\sigma_2^2 = 0.697$	0.436	4.99
GARCH	$\alpha_0 = 0.007$ $\alpha_1 = 0.135$ $\alpha_2 = 0.864$	4.299	∞
EGARCH	$\alpha_0 = -0.038$ $\alpha_{1a} = 0.026$ $\alpha_{1b} = 0.300$ $b_1 = 0.940$	0.528	4.35

Of course, all distributional models and heteroskedasticity models are leptokurtic. It is noteworthy that both the generalized Student distribution and the GARCH model imply infinite kurtosis and that the kurtosis of the EGARCH model is the smallest among these candidate models.

On the other hand, the variances of the mixture distribution and the Markov-switching model are very close to the sample variance of 0.437 and the variances of the compound Poisson process and the Student distribution are somewhat smaller than the sample variance. Both ARCH-type models have larger variances and the variance of the GARCH model exceeds the sample variance by nearly a factor of 10.

For the understanding of price differences between Black-Scholes prices and simulated option prices of the three static and three dynamic models, 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

$$
\begin{aligned}
(35) \qquad C_A = C_L &+ \frac{1}{2!e^{r\tau}}[\sigma^2(A) - \sigma^2(L)]f_L(X) \\
&- \frac{1}{3!e^{r\tau}}[\mu_3(A) - \mu_3(L)]\frac{df_L(X)}{dS} \\
&+ \frac{1}{4!e^{r\tau}}[\kappa_4(A) - \kappa_4(L) + 3(\sigma^2(A) - \sigma^2(L))]^2\frac{d^2f_L(X)}{dS^2} \\
&+ \varepsilon(X)
\end{aligned}
$$

where C_L is the Black-Scholes price (based on the log-normal distribution L), $e^{-r\tau}$ is the discount factor (which will be neglected here), $\sigma^2(A)$ and $\sigma^2(L)$ are the variances of the alternative (true) distribution and the log-normal distribution, respectively, μ_3 is the third central moment (which is related to the skewness β_1 by $\mu_3 = \beta_1\sigma^3$), κ_4 is the 4-th cumulant.

f_L is the density of the log-normal distribution, and $\varepsilon(X)$ is the approximation error as a function of the exercise price X. Note that $\kappa_4 = \mu_4 - 3\sigma^4$ where μ_4 is the 4-th central moment, i.e. $\kappa_4 > 0$ if the distribution is leptokurtic.

With reference to (35), option price biases can be decomposed into three components. First, the price bias will be a weighted function of the difference between $\sigma^2(A)$ and $\sigma^2(L)$, where the weights are given by the density of the log-normal distribution. Since the density is always positive, option prices under the true distribution will, ceteris paribus, be higher than Black-Scholes prices if $\sigma^2(A) > \sigma^2(L)$.

Second, there is a weighted skewness effect where the weights are given by the first derivative of the density f_L with respect to the spot rates. Figure 1, which plots the density of the log-normal distribution along with its first and second derivative, shows that the first derivative changes sign. Note that in (35) the density f_L is a function of the exercise price X. An option is said to be at the money if the spot rate S_t is equal to the discounted exercise price $e^{-r\tau}X$. Since $e^{-r\tau}X$ is the mean of the distribution, points to the right of the mean (which is to the right of the mode) classify as out-of-the-money, and points to the left as in-the-money.

The skewness term in (35) has a negative sign and the first-derivative weight has a negative (positive) sign for out-of-the-money (in-the-money) options. Therefore, out-of-the-money option prices under an alternative distribution would, ceteris paribus, be higher than Black-Scholes prices if this distribution is more skewed to the right than f_L, i.e. $\mu_3(A) > \mu_3(L)$. This, of course, is a very intuitive result.

The third effect is related to kurtosis and has weights given by the second derivative of f_L. The second derivative changes sign twice and is plotted in the lower panel of figure 1. The second derivative is positive for in-the-money and out-of-the-money options. If the kurtosis of A is larger than the kurtosis of L, then (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.

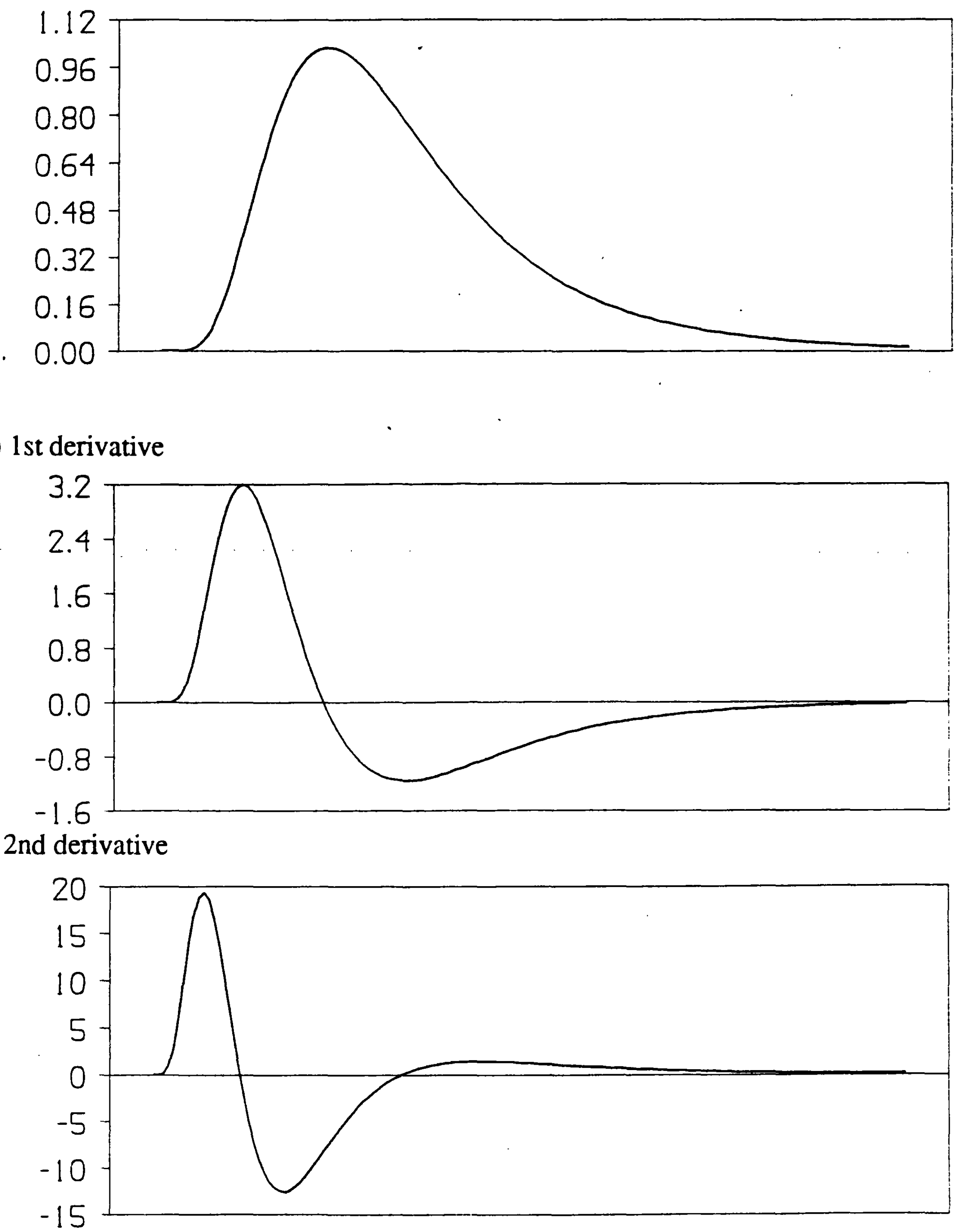
Figure 1

Density of the log-normal distribution and its 1st and 2nd derivative

a) density

b) 1st derivative

c) 2nd derivative



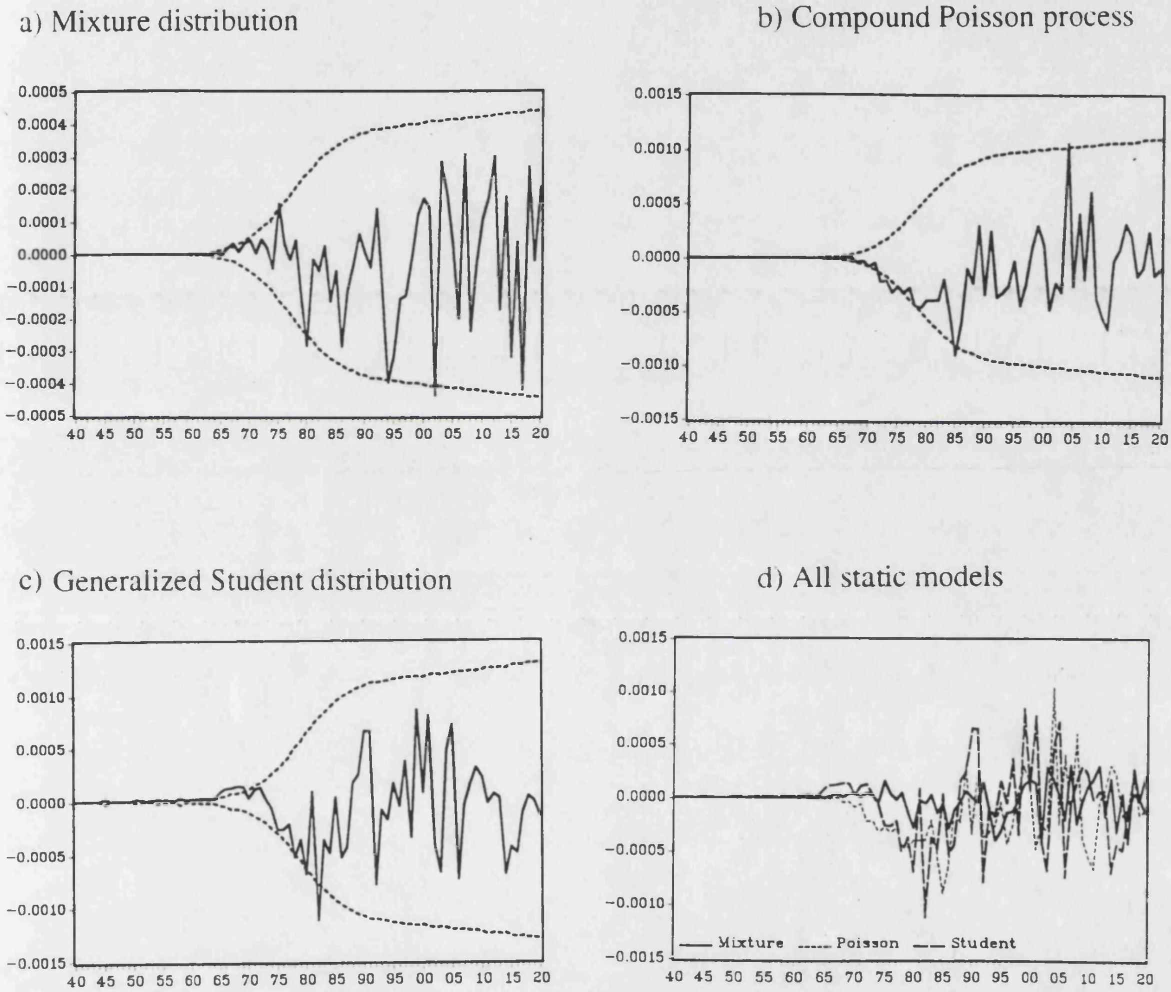
Simulations of European call option prices, according to the boundary condition in (28), were performed to study both the spot-rate effects and the maturity effects of the six candidate models. In the study of spot-rate effects, the exercise price X is set to 1.80 and the current spot rate S_t is varied between 1.40 and 2.20, i.e. the moneyness ratio varies between $0.\overline{7}$ and $1.\overline{2}$. The time to maturity is set to 20 days (roughly a month).

Figure 2 plots the spot-rate effects of biases in European call-option prices for the three static (distributional) models. In addition, the dashed lines show the confidence intervals of ± 2 standard errors around zero. A price bias is defined as the difference between the simulated option price C_k of the alternative model and the Black-Scholes price C_{BS}, i.e. a positive bias indicates that the "true" option price would be higher than the Black-Scholes price. Options with $S_t < 1.75$ are denoted as out-of-the-money options and options with $S_t > 1.85$ as in-the-money options.

Panel a of figure 2 displays the simulated price biases of the scale-mixture of normal distributions as a function of the current spot rate. There is no clear pattern in the price biases and in only 5 out of 81 cases is the bias outside the confidence interval by a small margin. Two of these biases are positive and three are negative. There is only a weak tendency for out-of-the-money options, i.e. those with a current spot rate of less than 1.75 (denoted as 75 on the horizontal axis), to have positive biases and for at-the-money options (between 75 and 85 on the horizontal axis) to have negative biases. In terms of the decomposition of price biases, according to (35), one would not expect to find variance and skewness effects because the variance of the mixture model is virtually identical with the sample variance and because a scale mixture is symmetric. Furthermore, the kurtosis effect seems to be too small to produce significant biases.

Figure 2

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



Turning to the spot-price effects of the compound Poisson process, one finds some significant negative biases for spot rates around the exercise price of 1.80 (denoted as 80)[5].

---

[5] The significant positive bias at $S_t = 2.04$ can safely be attributed to sample variation.

This can be attributed to both variance and kurtosis effects. Note that, according to table 1, the variance of 0.414, implied by the parameter estimates of the compound Poisson process, is smaller than the sample variance of 0.437. One may deduce, therefore, from (35) that the variance effect will be negative. In addition, leptokurtosis effects receive a negative weight for at-the-money options since the second derivative is negative for these options. Thus, both the variance and the kurtosis effect have a negative sign.

The negative effect on at-the-money options for leptokurtic distributions can be associated with peakedness. Recall that leptokurtosis can be caused by either peakedness or fat tails. If a distribution is more peaked than a normal distribution, i.e. there are more small-size price movements than expected under a normal distribution, then the probability of being deep-in-the-money at maturity will decrease for at-the-money options and, therefore, the option price should be lower. This is the economic intuition behind the negative peakedness effect.

The third static model is the generalized Student distribution whose price biases are plotted in panel c. Although the price biases are not large in comparison with the standard errors, there appears to be a systematic pattern for out-of-the-money options and at-the-money options. With one exception, all price biases between $S_t = 1.59$ and 1.69 are significantly positive and, with two exceptions, all biases between 1.75 and 1.85 are negative (although only three of them are significant). Like with Student's distribution, negative biases would be caused by the variance effect and the peakedness effect. Table 2 reports that a variance of 0.426 follows from the parameter estimates of the Student distribution and this is smaller than the sample variance of 0.437. As explained above, the peakedness effect is negative for at-the-money options.

The positive price biases for out-of-the-money options are readily explained in statistical and economic terms. Statistically, these biases can be described as fat-tail effects. Recall from figure 1 that the 2nd derivative of f_L, which gives the weights of the kurtosis effect, is positive in the domain of out-of-the-money options. Thus, the kurtosis effect will be positive for leptokurtic distributions. The economic intuition is simple: the Student distribution has fatter tails than a normal distribution. The fatter right tail implies that out-of-the-money options have a greater probability of finishing in the money than under a normal distribution and, therefore, the option price should be higher than the Black-Scholes price.

Since the vertical scales of panels a-c are not the same, the magnitude of price biases for all three static (distributional) models are compared in panel d. In general, the price effects are not very strong, especially for in-the-money options where the sample variation is rather high[6,7].

The results for the compound Poisson process are in agreement with the study of Jorion (1988) who found only statistically significant (negative) price effects for at-the-money options. Figure 2 shows that the price effects of the Student distribution are somewhat stronger than those of the compound Poisson process and that the price effects of the mixture distribution are the weakest amongst the three statis models. It seems that these models would not be able to explain the differences between actual prices and Black-Scholes prices observed on foreign-currency option markets. In particular, the observed smile effects cannot be derived from these models. Only the Student distribution produces vaguely a one-sided smile effect.

[6] Note that the draws from the normal distribution are the same for the simulation of the mixture model and the compound Poisson process (and the following dynamic models) at a given spot rate S_t but that the draws vary with S_t.

[7] In the literature, option-price biases are often displayed in percentage terms and in those terms the price effect for out-of-the money options is invariably the strongest. For two reasons it was decided to plot the price effect in money terms and not in percentage terms. First, high-percentage price effects can be very misleading when the option price is very low, as is typical for out-of-the-money options. For instance, in percentage terms the price bias for $S_t = 1.58$ is 2800 for the Student distribution and this number would dominate the plot of percentage biases although it is not statistically significant. Secondly, the larger percentage biases of out-of-the-money options are also often economically misleading. In the above example, the price effect is only 0.000028 in money terms and this would probably also be economically insignificant, even without transaction costs.

Figure 3 plots the spot-rate effects of the dynamic models. Panel a shows the simulated price biases for the Markov-switching model. For out-of-the-money options, the price biases are systematically positive and significantly so for $S_t > 1.60$, i.e. for these options are the option prices under a Markov-switching model higher than Black-Scholes prices. Note that the variance of the Markov-switching model is very close to the sample variance. Therefore, this price bias is only due to the fat-tail effect. For at-the-money options there is a strong and significant negative price bias which can be attributed to the peakedness effect. On the other hand, the price biases for in-the-money options is only significant at $S_t = 1.86$ and $S_t = 1.88$. However, all price biases between 1.85 and 2.00 (with one exception at 1.92) are positive. Thus, the Markov-switching model is broadly consistently with the observed smile effects.

How can the positive price effect for in-the-money options be explained? Statistically it follows from the kurtosis effect which has positive weights from the second derivative. Economically, the left-tail effect can be explained by the fact that it enhances the insurance value of deep-in-the-money options by increasing the probability of finishing out of the money. Or, alternatively, the price effect may be explained with the put-call parity of (16). Note that this parity is independent of distributional assumptions. If a call option is in the money, the corresponding put option with the same exercise price will be out of the money. The preceding description of the fat-tail effect for out-of-the-money call options can be extended to out-of-the-money put options if both the right and left tail are fat (which is the case here because the stationary distribution is symmetric). A fatter left tail implies for the out-of-the-money option that the probability of moving into the money increases and, therefore, the option price should be higher. It now follows from the put-call parity that also the price of the corresponding in-the-money call option with the same spot rate should be higher.

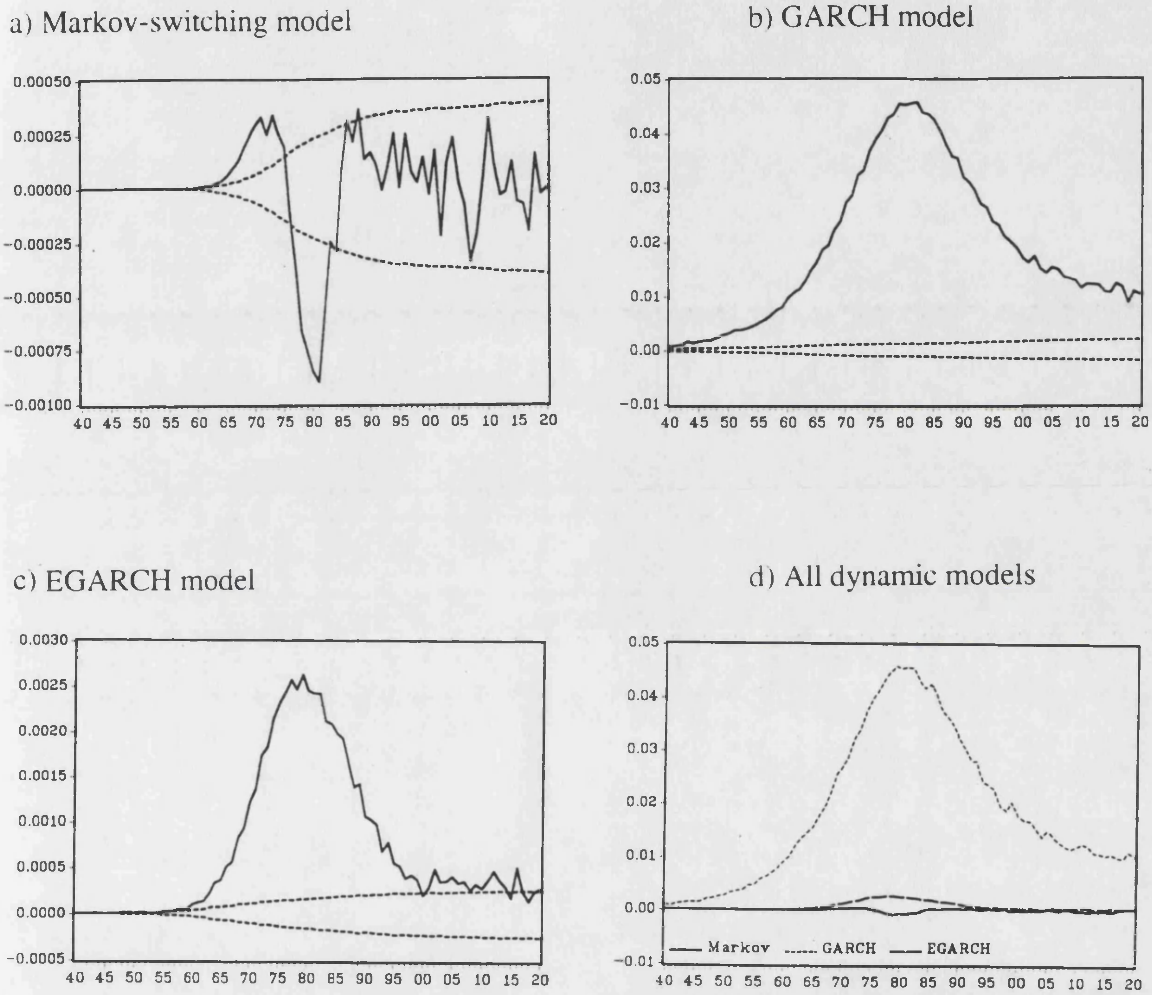
The price effects of the GARCH and EGARCH models, displayed in panels b and c, respectively, are drastically different from the pattern of the Markov-switching model. For both models, the simulated option prices exceed the Black-Scholes prices for the full range of spot rates. The price effect is strongest for at-the-money options. At $S_t = 1.80$, the Black-Scholes price is 0.0215, whereas it is 0.0234 for the EGARCH model and 0.0668 for the GARCH model. These large positive price biases are obviously related to the fact that both models imply stationary variances which are much larger than the sample variance of 0.437 (see table 2). Since, according to the decomposition of price biases in (35), the variance effect is weighted by the density f_L, the price effect is strongest at the money.

Panel d of figure 3 plots the spot-rate effects of all three dynamic models to demonstrate that the GARCH model produces biases which are multiples of those from the other two models. Although the price effects of at-the-money options are opposite in sign for the Markov-switching and the EGARCH model, both models display biases which are roughly similar in magnitude. With these enormous differences in option prices between the GARCH model and the other two models, the question arises, which of the models gives the most reliable option prices. One way to answer this question would be to compare model prices with market prices to find the best fit. This, however, would be beyond the scope of this study. Instead, one could argue that the GARCH model is likely to overestimate option prices seriously. The application of the GARCH model in Section 3.2 has made it clear that for the exchange-rate data, the condition of finite stationary variance is often violated and the parameter estimates are only marginally below the stationarity condition $\alpha_1 + \beta_1 < 1$ for the daily pound series.

Turning next to a study of maturity effects, the aspects of temporal aggregation is important. The models, as estimated from daily data, are strictly valid only for this time period of one day. Only stable distributions have the property of additivity. If, e.g., we model the distribution of daily exchange-rate fluctuation by the Student distribution and aggregate over time to get from daily fluctuations to monthly fluctuations, then the monthly fluctuations will not have a Student distribution since the sum of Student-distributed random variables does not have a Student distribution.. The same argument applies to the other non-stable dis-

Figure 3

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



tributions and heteroskedasticity models, too.

For the three static models it follows from the central limit theorem that they converge to normality under aggregation and the results of Lindgren (1978) and Diebold (1988) indicate that this convergence to normality holds also for Markov-switching and ARCH models. As a corollary, one expects to find stronger divergence from Black-Scholes prices for short-maturity than for long-maturity options and for long enough maturities the option prices should converge to Black-Scholes prices.

Maturity-effects will be studied for at-the-money options with $S_t = X = 1.80$ and the maturity is varied between 1 day and 100 days (roughly 5 months).

Figure 4 plots the maturity effects of the static models. Panel a shows that there is indeed a negative bias for short maturities of up to 6 days for the mixture distribution. Since the mixture distribution does not have any variance or skewness effect, the negative bias can be attributed to the peakedness effect of at-the-money options as described above. The convergence to normality is obviously so fast for the mixture distribution that one cannot find significant price effects for longer maturities. Figure 4a is, of course, consistent with Figure 2a where no strong biases were found for a maturity of 20 days.

The negative peakedness effect is also borne out in figures 4b and 4c. Note, however, that the peakedness effect of the compound Poisson process and the Student distribution is enhanced by a negative variance effect, since the variance of these distributions are smaller than the sample variance. It is also interesting to observe that quite a few price biases for long maturities are also significantly negative for both distributions. The convergence to normality seems to be relatively slow for the compound Poisson process.

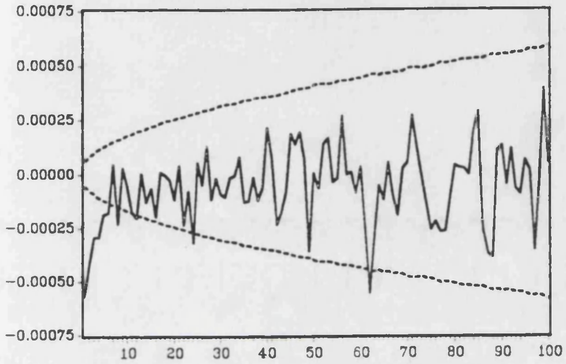
Figure 4d compares the maturity effects of the three static models in percentage terms. In general, the biases of the compound Poisson process and of the Student distribution are much stronger than those of the mixture distribution and they also have only very few positive biases whereas the biases of the mixture distribution quickly fluctuate around zero.

Finally, the maturity effects of the heteroskedasticity models are displayed in figure 5. As noted above, at-the-money options have only a peakedness effect for the Markov-switching model. Figure 5a illustrates that the negative peakedness effect increases in size when maturity increases from 1 day to 10 days and the negative bias levels off rather slowly. The slow

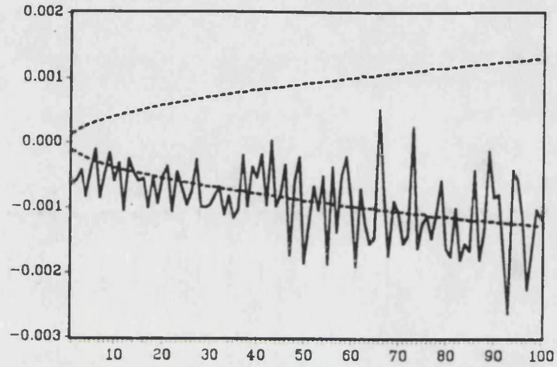
Figure 4

Maturity effect of biases in option prices for static models

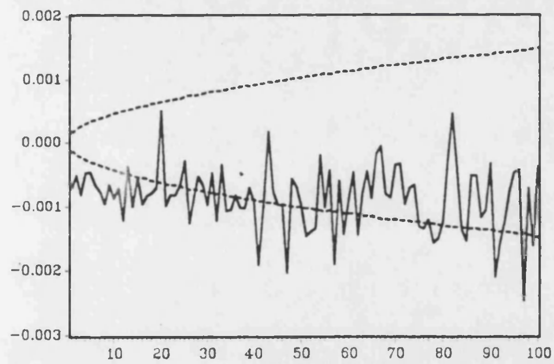
a) Mixture distribution



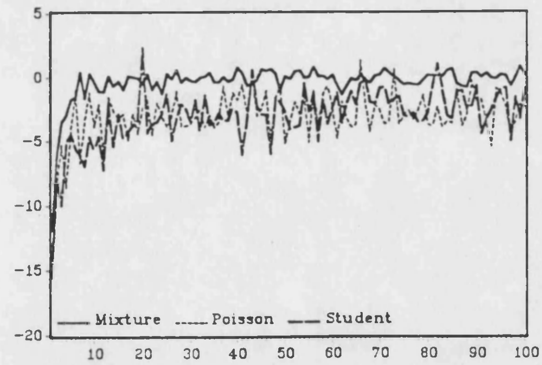
b) Compound Poisson process



c) Generalized Student distribution



d) All static models

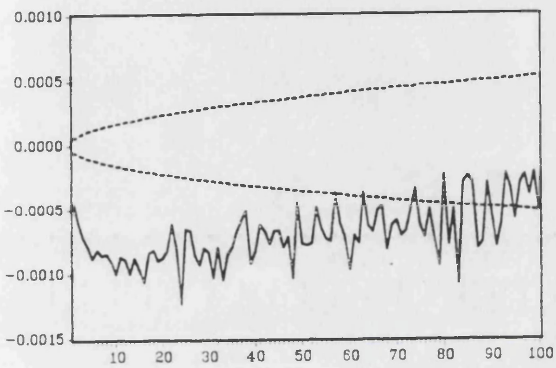


convergence to Black-Scholes prices is caused by the large persistence of states which follows from the large estimates of the transition probabilities $p_{11} = 0.933$ and $p_{2} = 0.952$. Therefore, the transition matrix will converge rather slowly to its stationary values.

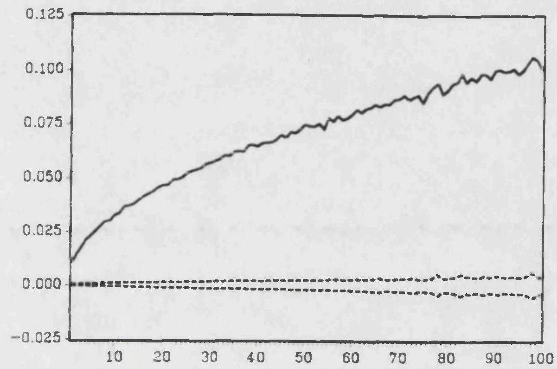
Figure 5

Maturity effect of biases in option prices for dynamic models

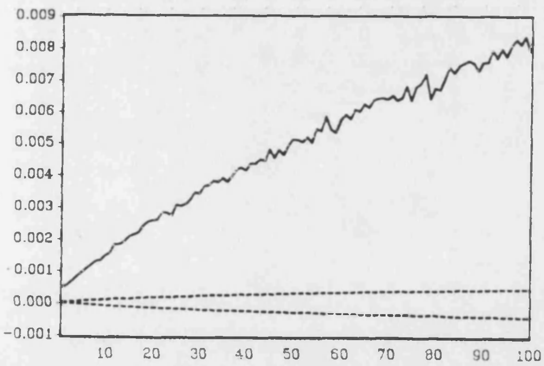
a) Markov-switching model



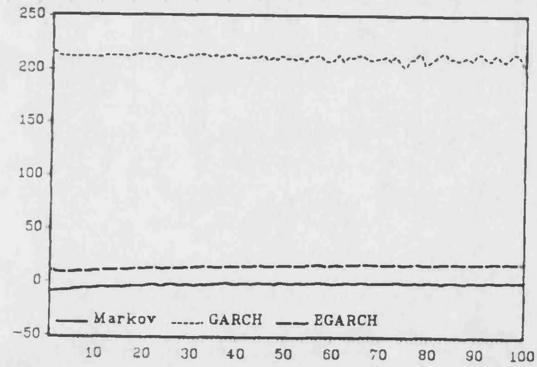
b) GARCH model



c) EGARCH model



d) All dynamic models



Whereas the Markov-switching model implies that at-the-money options are overpriced by the Black-Scholes model, the opposite conclusion would be drawn from the GARCH and EGARCH models. It was explained above that the price biases of the latter models are dominated by the positive variance effect. Figures 5b and 5c show that the biases in money terms increase both for the GARCH and EGARCH model with maturity. Since both models have stationary variances which exceed the sample variance, one would not obtain a convergence to Black-Scholes prices based on the sample variance, but to those based on the stationary variances when maturity increases. If, in fact, the Black-Scholes prices are calculated with the stationary variance $\bar{\sigma}^2 = 4.299$ of the GARCH model, one would find at a maturity of 10 days that the GARCH option price of 0.0475 is 12.6 percent smaller than the Black-Scholes price (this is the peakedness effect), whereas at a maturity of 100 days the GARCH option price of 0.1503 exceeds the Black-Scholes price by only 1.1 percent. Figure 5d, however, shows that the percentage bias of the GARCH model in terms of Black-Scholes prices based on the sample variance is remarkably constant at around 210 percent when maturity is varied.

5.4 SUMMARY

Empirical studies of prices on foreign-currency option markets (mainly at the PHLX) have revealed that there are some systematic differences between market prices and (modified) Black-Scholes prices, especially for options with short maturities. In particular, smile effects were observed, i.e. implied volatilities of at-the-money options are systematically smaller than those of in-the-money options and out-of-the-money options. It was conjectured that the smile effects were caused by leptokurtic distributions. Some researchers applied the constant-elasticity-of-variance model, the compound Poisson process, and stochastic-volatility models to exchange-rate data in order to capture these effects. It was found, however, that these models produced only rather small option price effects which could not explain the large observed biases.

In this chapter, I have applied three static (distributional) models and three dynamic (heteroskedastic) models to simulate option prices by the Monte-Carlo method. It might appear to be confusing that the models produce so different price effects. The static models have only significant price biases for short maturities, where the models predict lower at-the-money option prices than the Black-Scholes model. On the other hand, both ARCH-type models produce very large option prices, compared to Black-Scholes prices, for all moneyness-ratios (when the maturity is 20 days) and all maturities (when the option is at the money). Only the Markov-switching model is consistent with the observed smile effects.

In order to understand the option-price effects produced by the models, it is useful to decompose the price biases into variance effects and kurtosis effects. It is shown that these effects are weighted, in a generalized Edgeworth-series expansion, by the density of the log-normal distribution and its second derivative, respectively. The variance effect is always positive when the model's variance is larger than the sample variance, i.e. the simulated option prices are larger than Black-Scholes prices. The kurtosis effect can be further decomposed into a peakedness effect and fat-tail effects. For leptokurtic distributions, the peakedness effect of at-the-money options is negative (the simulated prices are smaller than Black-Scholes prices) whereas the fat-tail effect is positive for both out-of-the-money options and in-the-money options.

The Markov-switching model does not have a variance effect because its variance is virtually identical with the sample variance. Therefore, the smile effect is entirely caused by the leptokurtosis of the model. Although the static models have larger leptokurtosis than the Markov-switching model (see table 2), they only have strong price effects for short maturities since they converge quite rapidly to a normal distribution under time aggregation. On the other hand, the parameter estimates of the Markov-switching model imply strong persistence of states and, therefore, a slow convergence to the stationary distribution. As a consequence, significant price effects can also be obtained for relatively long maturities of up to 90 days or so.

CONCLUSIONS

The aim of this study was to find a suitable approach to model econometrically the process of exchange-rate dynamics. Eight models have been introduced to capture the statistical properties of leptokurtosis and heteroskedasticity. Two of these models can be rejected. First, the stable Paretian distributions are not compatible with convergence to normality under temporal aggregation and were rejected both by direct estimation via the Fast-Fourier-Transform and by putting them into the broader framework of distributions with regularly varying tails. Second, the GARCH-t model violates strongly the stationarity condition for the variance and does not produce well-behaved residuals. As a matter of fact, the leptokurtosis of residuals is in most cases larger than the sample leptokurtosis.

The obvious question is: which of the remaining six models gives the best representation of the exchange-rate data. It might appear to be obvious that the dynamic heteroskedasticity models (Markov-switching, GARCH, and EGARCH) perform much better than the static distributional models (mixture distribution, compound Poisson process, and Student distribution) since the latter cannot capture heteroskedasticity. However, Chapter 4 showed that this is not necessarily so. First, the static models gave, in general, better results in χ^2 goodness-of-fit tests than the static models. Second, dynamic models had only superior forecasting performance of volatility with respect to the mean forecast error, but not so with respect to the precision of the forecasts. On the other hand, the dynamic models are clearly superior with respect to the adjusted likelihood criterion SIC.

Another important aspect of model comparison is the option-price effect of the models. The Markov-switching model is the one which can best mimic the smile effect of implied volatilities. GARCH and EGARCH models probably overestimate option prices because they have strong variance effects caused by the near violation of the stationarity condition. On the other hand, static models have only significant option-price effects when the maturity is very short (i.e. less than 10 days). Therefore, on this account, the Markov-switching model seems to be the most adequate model.

It is more difficult to give an overall ranking of all models because none of them was superior in all criteria of comparison. If forced to pick a favourite model, I would probably pick the Markov-switching model because it is compatible with the major stylized facts of the exchange-rate data and achieves satisfactory results in forecasting experiments and option-price simulations. It might be argued that the Markov-switching model is difficult to adapt to a richer dynamic structure (second or higher order Markov chain) and that generalizations to more than 2 or 3 states can lead to numerical problems. On the other hand, this study has shown that for exchange-rate data a parsimonious four-parameter model is adequate.

It is not difficult to adapt the approach of this study to other financial markets like stock markets and commodity markets since the price dynamics on these markets also exhibit the properties of heteroskedasticity and leptokurtosis. The comprehensive analysis of statistical properties in Chapter 1 has shown that it is sufficient to concentrate the modelling of exchange-rate data on the even moments, but for share prices it would presumably be necessary to extend the modelling to the odd moments. For most models of this study, this is not a difficult extension.

Chapter 5 indicated that the stochastic modelling of the dynamic and distributional properties of financial time series is important for option markets. For the pricing of options it is essential to have reliable estimates and forecasts of volatility. It is encouraging to see that the modelling of volatility is not only of academic interest.

REFERENCES

Adams, P.D., Wyatt, S.B. (1987), On the pricing of European and American foreign currency call options. Journal of International Money and Finance. Vol. 6, 315-338.

Adams, P.D., Wyatt, S.B. (1989), On the pricing of European and American foreign currency options: A clarification. Journal of International Money and Finance, Vol. 8, 305-311.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J.W. (1972), Robust Estimation of Location. Princeton: Princeton University Press.

Akgiray, V., Booth, G.G.(1988), Mixed diffusion-jump process modeling of exchange rate movements. Review of Economics and Statistics, Vol. 70, 631-637.

Akgiray, V., Lamoureux, C.G. (1989), Estimation of stable-law parameters: A comparative study. Journal of Business and Economic Statistics, Vol. 7, 85-93.

Backus, D. (1984), Empirical models of the exchange rate: Separating the wheat from the chaff. Canadian Journal of Economics, Vol. 17, 824-846.

Baillie, R.T., Bollerslev, T. (1989), The message in daily exchange rates: A conditional variance tale. Journal of Business and Economic Statistics, Vol. 7, 297-305.

Baillie, R.T., McMahon, P.C. (1987), Empirical regularities in exchange rate behaviour. In: K.A. Chrystal, R. Sedgwick (eds.), Exchange Rates and the Open Economy, Brighton: Wheatsheaf, 7-29.

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

Ball, C.A., Torous, W.N. (1983), A simplified jump process for common stock returns. Journal of Financial and Quantitative Analysis, Vol. 18, 53-61.

Ball, C.A., Torous, W.N. (1985), On jumps in common stock prices and their impact on call option pricing. Journal of Finance, Vol, 155-173.

Beale, E.M., Mallows, C.L. (1959), Scale mixing of symmetric distributions with zero means. Annals of Mathematical Statistics, Vol. 30, 1145-1151.

Biger, N., Hull, J. (1983), The valuation of currency options. Financial Management, Vol. 12, 24-28.

Black, F., Scholes, M.J. (1973), The pricing of options and corporate liabilities. Journal of Political Economy, Vol. 81, 637-654.

Blattberg, R.C., Gonedes, N.J. (1974), A comparision of the stable and Student distributions as statistical models for stock prices. Journal of Business, Vol. 47, 244-280.

Bodurtha, J.N.Jr., Courtadon, G.R. (1986), Efficiency tests of the foreign currency options market. Journal of Finance, Vol. 41, 151-162.

Bodurtha, J.N.Jr., Courtadon, G.R. (1987), Tests of an American option pricing model on the foreign currency options market. Journal of Financial and Quantitative Analysis, Vol. 22, 153-167.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, Vol. 31, 307-327.

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.

Boothe, P., Glassman, D. (1987), The statistical distribution of exchange rates. Journal of International Economics, Vol. 22, 297-319.

Borensztein, E., Dooley, M.P. (1987), Options on foreign exchange and exchange rate expectations. IMF Staff Papers, Vol. 34, 643-680.

Bowman, K.O., Shenton, L.R. (1986), Moment $(\sqrt{b_1}, b_2)$ techniques. In: R.B. D'Agostino, M.A. Stephens (eds.), 279-329.

Box, G.E.P., Jenkins, G.M. (1976), Time Series Analysis: Forecasting and Control. Rev. ed., Oakland: Holden-Day.

Brockwell, P.J., Davis, R.A. (1987), Time Series: Theory and Methods. New York: Springer.

Burt, J., Kaen, F.R., Booth, G.G. (1977), Foreign exchange market efficiency under flexible exchange rates. Journal of Finance, Vol. 32, 1325-1330.

Calderon-Rossell, J.R., Ben-Horim, M. (1982), The behavior of foreign exchange rates. Journal of International Business Studies, Vol. 13, 99-111.

Chatfield, C. (1973), Statistical inference regarding Markov-chain models. Applied Statistics, Vol. 22, 7-20.

Chesny, M., Lonbergé, H. (1987), The pricing of European currency options: Empirical tests based on Swiss data. Aussenwirtschaft, Vol. 42, 213-228

Clark, P.K. (1973), A subordinated stochastic process model with finite variance for speculative prices. Econometrica, Vol. 41, 135-155.

Cochrane, J.H. (1988), How big is the random walk in GNP? Journal of Political Economy, Vol. 96, 893-920.

Cohen, A.C. (1967), Estimation in mixtures of two normal distributions. Technometrics, Vol. 9, 15-28.

Conover, W.J. (1971), Practical Nonparametric Statistics. New York: Wiley.

Conover, W.J., Johnson, M.E., Johnson, M.M. (1981), A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. Technometrics, Vol. 28, 351-361.

Corbae, D., Ouliaris, S. (1986), Robust tests for unit roots in the foreign exchange market. Economics Letters, Vol. 22, 375-380.

Cornell, W.B. (1977), Spot rates, forward rates and exchange market efficiency. Journal of Financial Economics, Vol. 5, 55-65.

Cornell, W.B., Dietrich, J.K. (1978), The efficiency of the market for foreign exchange under floating exchange rates. Review of Economics and Statistics, Vol. 60, 111-120.

Cox, D.R., Miller, H.D. (1965), The Theory of Stochastic Processes. London: Chapman and Hall.

Csörgö, S. (1984), Adaptive estimation of the parameters stable laws. In: R. Revesz (ed.), Limit Theorems in Probability and Statistics, Vol. 1, Amsterdam: North-Holland, 305-367.

D'Agostino, R.B. (1986), Tests for the normal distribution. In: R.B. D'Agostino, M.A. Stephens (eds.), 367-419.

D'Agostino, R.B., Belanger, B., D'Agostino, R.B.Jr. (1990), A suggestion for using powerful and informative tests of normality. American Statistician, Vol. 44, 316-321.

D'Agostino, R.B., Stephens, M.A. (eds.) (1986), Goodness-of-Fit Techniques. New York: M.Dekker.

Davies, N., Triggs, C.M., Newbold, P. (1977), Significance levels of the Box-Pierce portmanteau statistic in finite samples. Biometrica, Vol. 64, 517-522.

Diebold, F.X. (1988), Empirical modeling of Exchange Rate Dynamics. Berlin: Springer.

Douglas, J.B. (1980), Analysis with Standard Contagious Distributions. Fairland: International Co-operative Publishing House.

Duan, J.-C. (1991), The GARCH option pricing model. Faculty of Management, McGill University, mimeo.

Dufour, J.-M. (1981), Rank tests for serial dependence. Journal of Time Series Analysis, Vol. 2, 117-128.

Elderton, W.P., Johnson, N.L. (1969), Systems of Frequency Curves. Cambridge: Cambridge University Press.

Emerson, J.D. (1983), Mathematical aspects of transformation. In: D.C. Hoaglin, F. Mosteller, J.W. Tukey (eds.), 247-282.

Emerson, J.D., Stoto, M.A. (1982), Exploratory methods for choosing power transformations. Journal of the American Statistical Association, Vol. 77, 103-108.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, Vol. 50, 987-1007.

Engle, R.F., Bollerslev, T. (1986), Modelling the persistence of conditional variances. Econometric Reviews, Vol. 5, 1-50.

Evans, G.W. (1986), A test for speculative bubbles in the sterling-dollar exchange-rate: 1981-1984. American Economic Review, Vol. 76, 621-636.

Everitt, B.S. (1984), Maximum likelihood estimation of the parameters in a mixture of two univariate normal distributions: A comparison of different algorithms. The Statistician, Vol. 33, 205-215.

Everitt, B.S., Hand, D.J. (1981), Finite Mixture Distributions. London: Chapman and Hall.

Fabozzi, F.J., Hauser, S., Yaari, M. (1990), Early exercise of foreign currency options. Advances in Futures and Options Research, Vol. 4, 219-236.

Fama, E.F. (1965), The behavior of stock market prices. Journal of Business, Vol. 38, 34-105.

Fama, E.F. (1970), Efficient capital markets: A review of theoretical and empirical work. Journal of Finance, Vol. 25, 383-417.

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

Fama, E.F., Roll, R. (1968), Some properties of symmetric stable distributions. Journal of the American Statistical Association, Vol. 63, 817-836.

Fama, E.F., Roll, R. (1971), Parameter estimates for symmetric stable distributions. Journal of the American Statistical Association, Vol. 66, 331-338.

Feller, W. (1971), An Introduction to Probability Theory and Its Applications. Vol. 2, 2nd ed., New York: Wiley

Feuerverger, A., McDunnough, P. (1981), On efficient inference in symmetric stable laws and processes. In: M. Csörgö, D.A. Dawson, J.N.K. Rao, A.K.M. Saleh (eds.), Statistics and Related Topics, Amsterdam: North-Holland, 109-122.

Frenkel, J.A., Levich, R.M. (1977), Transaction costs and arbitrage: Tranquil versus turbulent periods. Journal of Political Economy, Vol. 13, 171-186.

Friedman, D., Vandersteel, S. (1982), Short-run fluctuations in foreign exchange rates. Journal of International Economics, Vol. 13, 171-186.

Fuller, W.A. (1976), Introduction to Statistical Time Series. New York: Wiley.

Galambos, J. (1988), Advanced Probability Theory. New York: M. Dekker.

Garman, M.B., Kohlhagen, S.W. (1983), Foreign currency option values. Journal of International Money and Finance, Vol. 2, 231-237.

Giddy, I.H., Dufey, G. (1975), The random behavior of flexible exchange rates: Implications for forecasting. Journal of International Business Studies, Vol. 6, 1-32.

Grabbe, J.O. (1983), The pricing of call and put options on foreign exchange. Journal of International Money and Finance, Vol. 2, 239-253.

Grabbe, J.O. (1987), Foreign currency options: A survey. In: S.J. Khoury and A. Ghosh (eds.), Recent Developments in International Banking and Finance, Vol. 1, Lexington.

Granger, C.W.J., Andersen, A.P. (1978), An Introduction to Bilinear Time Series Models. Göttingen: Vandenhoeck and Ruprecht.

Granger, C.W.J., Orr, D. (1972), "Infinite variance" and research strategy in time series analysis. Journal of the American Statistical Association, Vol. 67, 275-285.

Grimmett, G.R., Stirzaker, D.R. (1982), Probability and Random Processes. Oxford: Clarendon Press.

Gurland, J. (1957), Some interrelations among compound and generalized distributions. Biometrika, Vol. 44, 265-268.

Hakkio, C.S. (1986), Does the exchange rate follow a random walk? A Monte Carlo study of four tests for a random walk. Journal of International Money and Finance, Vol. 5, 221-229.

Hall, P. (1982), On some simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society Series B, Vol. 44, 37-42.

Hamming, R.W. (1973), Numerical Methods for Scientists and Engineers. 2nd ed., New York: McGraw-Hill.

Hill, B.M. (1975), A simple general approach to inference about the tail of a distribution. Annals of Statistics, Vol. 3, 1163-1174.

Hillard, J.E., Tucker, A.L. (1991), Market-determined permia for American currency spot options. Advances in Futures and Options Research, Vol. 5, 227-240.

Hoaglin, D.C. (1983), Letter values: A set of selected order statistics. In: D.C. Hoaglin, F. Mosteller, J.W. Tukey (eds.), 33-57.

Hoaglin, D.C., Mosteller, F., Tukey , J.W. (eds.) (1983), Understanding Robust and Exploratory Data Analysis. New York: Wiley.

Hodrick, R.J. (1987), The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets. Chur: Harwood.

Hollander, M., Wolfe, D.A. (1973), Nonparametric Statistical Methods. New York: Wiley

Holt, D.R., Crow, E.L. (1973), Tables and graphs of the stable probability density functions. Journal of Research of the National Bureau of Standards, Vol. 77B, 143-198.

Hsieh, D.A. (1988), The statistical properties of daily foreign exchange rates: 1974-1983. Journal of International Economics, Vol. 24, 129-145.

Hsieh, D.A. (1989a), Modelling heteroskedasticity in daily foreign-exchange rates. Journal of Business and Economic Statistics, Vol. 7, 307-317.

Hsieh, D.A. (1989b), Testing for nonlinear dependence in daily foreign exchange-rates. Journal of Business, Vol. 62, 339-368.

Hull, J. (1987), The pricing of options on assets with stochastic volatilities. Journal of Finance, Vol. 42, 281-300.

Hull, J. (1993), Futures, and Other Derivative Securities. Englewood Cliffs: Prentice-Hall, 2nd ed.

Iglewicz, B. (1983), Robust scale estimators and confidence intervals for location. In: D.C. Hoaglin, F. Mosteller, J.W. Tukey (eds.), 404-431.

Jarque, C.M, Bera, A.K. (1980), Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters, Vol. 8, 255-259.

Jarrow, R., Rudd, A. (1982), Approximate option valuation for arbitrary stochastic processes. Journal of Financial Economics, Vol. 10, 347-369.

Johnson, N.L., Kotz, S. (1969), Distributions in Statistics: Discrete Distributions. New York: Houghton Mifflin.

Johnson, N.L., Kotz, S. (1970), Distributions in Statistics: Continuous Univariate Distributions. Vol. 1. New York: Houghton Mifflin.

Jorion, P. (1988), On Jump Processes in the foreign exchange and stock markets. Review of Financial Studies, Vol. 1, 427-445.

Kaehler, J. (1988), Statistische Modelle für den DM-Dollar Wechselkurs. In: H. Kräger (ed.), Empirische Wirtschaftsforschung, Frankfurt: Campus.

Kaehler, J. (1990a), Laws of iterated expectations for higher order central moments. Statistical Papers, Vol. 31, 295-300.

Kaehler, J. (1990b), A new proof of the leptokurtosis theorem for Poisson sums of random variables. Jahrbücher für Nationaökonomie und Statistik, Vol. 209, 337-342.

Kaehler, J. (1990c), Some variations on the compound Poisson process as a model of price dynamics in financial markets. Institut für Volkswirtschaftslehre und Statistik, University of Mannheim, Discussion Paper No. 423-90.

Kendall, M., Stuart, A. (1979), The Advanced Theory of Statistics. Vol. 2, 4th ed., London: Griffin.

Kiefer, N.M., Salmon, M. (1983), Testing normality in econometric models. Economics Letters, Vol. 11, 123-127.

Kim, B.J.C. (1987), Do the foreign exchange rates really follow a random walk? Economics Letters, Vol. 23, 289-293.

Kinderman, A.J., Monahan, J.F., Ramage, J.G. (1977), Computer methods for sampling from Student's t distribution. Mathematics of Computation, Vol. 31, 1009-1018.

Koutrouvelis, I.A. (1980), Regression-type estimation of the parameters of stable laws. Journal of the American Statistical Association, Vol. 75, 918-928.

Lastrapes, W.D. (1989), Exchange rate volatility and U.S. monetary policy: An ARCH application. Journal of Money, Credit, and Banking, Vol. 21, 66-77.

Lehmann, E.L. (1975), Nonparametrics, New York: McGraw-Hill.

Leventakis, J.A. (1987), Exchange rate models: Do they work? Weltwirtschaftliches Archiv, Vol. 123, 363-376.

Levich, R.M. (1985), Empirical studies of exchange rates: Price behavior, rate determination and market efficiency. In: R.W. Jones, P.B. Kenen (eds.), Handbook of International Economics, Vol. 2, Amsterdam: North-Holland, 979-1040.

Liu, C.Y., He, J. (1991), A variance-ratio test of random walks in foreign exchange rates. Journal of Finance, Vol. 46, 773-785.

Ljung, G.M., Box, G.E.P. (1978), On a measure of lack of fit in time series models. Biometrika, Vol. 65, 297-303.

Lo, A.W., MacKinlay, A.C. (1988), Stock market prices do not folow random walks: Evidence from a simple specification test. Review of Financial Studies, Vol. 1, 41-66.

Lo, A.W., MacKinlay, A.C. (1989), The size and power of the variance ratio test in finite samples. Journal of Econometrics, Vol. 40, 203-238.

Logue, D.E., Sweeney, R.J. (1977), "White-noise" in imperfect markets: The case of the franc/dollar exchange rate. Journal of Finance, Vol. 32, 761-768.

Logue, D.E., Sweeney, R.J., Willett, T.D. (1978), Speculative behavior of foreign exchange rates during the current float. Journal of Business Research, Vol. 6, 159-174.

Mandelbrot, B. (1963), The variation of certain speculative prices. Journal of Business, Vol. 36, 394-419.

McFarland, J.W., Pettit, R.R., Sung, S.K. (1982), The distribution of foreign exchange price changes: Trading day effects and risk measurement. Journal of Finance, Vol. 37, 693-715.

McCulloch, J.H. (1986), Simple consistent estimators of stable distribution parameters. Communications in Statistics - Simulation and Computation, Vol. 15, 1109-1136.

McLeod, A.I., Li, W.K. (1983), Diagnostic checking ARIMA time series models using squared-residual autocorrelations. Journal of Time Series Analysis, Vol. 4, 269-273.

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.

Meese, R.A., Singleton, K.J. (1982), On unit roots and the empirical modelling of exchange rates. Journal of Finance, Vol. 37, 1029-1035.

Melino, A., Turnbull, S.M. (1991), The pricing of foreign currency options. Canadian Journal of Economics, Vol. 24, 251-281.

Milhoj, A. (1985), The moment structure of ARCH processes. Scandinavian Journal of Statistics, Vol. 12, 281-292.

Milhoj, A. (1987), A conditional variance model for daily deviations of an exchange rate. Journal of Business and Economic Statistics, Vol. 5, 99-103.

Moore, D.S. (1986), Tests of chi-squared type. In: R.B. D'Agostino, M.A. Stephens (eds.), 63-95.

Moors, J.J.A. (1988), A quantile alternative for kurtosis. The Statistician, Vol. 37, 25-32.

Mussa, M. (1979), Empirical regularities in the behavior of exchange rates and theories of the foreign exchange market. Carnegie-Rochester Conference Series on Public Policy, Vol. 11, 9-57.

Mussa, M. (1984), The theory of exchange rate determination. In: J.F.O. Bilson, R.R. Marston (eds.), Exchange rate theory and policy. Chicago: University of Chicago Press, 13-58.

Ord, J.K. (1985), Pearson system of distributions. In: S. Kotz, N.L. Johnson (eds.), Encyclopedia of Statistical Sciences, Vol. 6, New York: Wiley, 655-659.

Parzen, E. (1960), Modern Probability Theory and its Applications. New York: Wiley.

Pentecost, E.J. (1991), Econometric approaches to empirical models of exchange rate determination. Journal of Economic Surveys, Vol. 5, 71-96.

Phillips, P.C.B., Loretan, M. (1990), Testing covariance stationarity under moment condition failure with an application to common stock returns. Cowles Foundation for Research in Economics, Yale University, mimeo.

Praetz, P.D. (1972), The distribution of share price changes. Journal of Business, Vol. 45, 49-55.

Press, S.J. (1967), A compound events model for security prices. Journal of Business, Vol. 40, 317-335.

Press, S.J. (1968), A modified compound Poisson process with normal compounding. Journal of the American Statistical Association, Vol. 63, 607-613.

Priestley, M.B. (1981), Spectral Analysis and Time Series. Vol. 1, London: Academic Press.

Quandt, R.E., Ramsey, J.B. (1978), Estimating mixtures of normal distributions and switching regressions. Journal of the American Statistical Association, Vol. 73, 730-738.

Raiffa, H., Schlaifer, R. (1961), Applied Statistical Decision Theory. Boston: Harvard University.

Randles, R.H., Fligner, M.A., Policello, G.E., Wolfe, D.A. (1980), An asymptotically distribution-free test for symmetry versus asymmetry. Journal of the American Statistical Association, Vol. 75, 168-172.

Rogalski, R.J., Vinso, J.D. (1978), Empirical properties of foreign exchange rates. Journal of International Business Studies, Vol. 9, 69-79.

Rosenberger, J.L., Gasko, M. (1983), Comparing location estimators: Trimmed means , medians, and trimean. In: D.C. Hoaglin, F. Mosteller, J.W. Tukey (eds.), 297-338.

Schlittgen, R., Hammann, D.,Lepinat, K. (1982), Zur Verteilung von Wechselkursveränderungen. Allgemeines Statistisches Archiv, Vol. 66, 162-173.

Shastri, K., Tandon, K. (1986a), Valuation of foreign currency options: Some empirical tests. Journal of Financial and Quantitative Analysis, Vol. 21, 145-160.

Shastri, K., Tandon, K. (1986b), On the use of European models to price American options on foreign exchange. Journal of Fututes Markets, Vol. 6, 93-108.

Shastri, K., Tandon, K. (1987), Valuation of American options on foreign exchange. Journal of Banking and Finance, Vol. 11, 245-269.

Shastri, K., Wethyavivorn, K. (1987), The valuation of currency options for alternate stochastic processes. Journal of Financial Research, Vol. 10, 283-392.

So, J.C. (1987), The distribution of foreign exchange price changes: A comment. Journal of Finance, Vol. 42, 181-188.

Solo, V. (1984), The order of differencing in ARIMA models. Journal of the American Statistical Association, Vol. 79, 916-921.

Stephens, M.A. (1986), Tests based on EDF statistics. In: R.B. D'Agostino, M.A. Stephens (eds.), 97-193.

Taya, T. (1980), Variability of exchange rates and the efficiency of the foreign exchange market. In: D.Bigman, T. Taya (eds.), The Functioning of Floating Exchange Rates, Cambridge, Mass.: Ballinger, 141-167.

Taylor, S. (1986), Modelling Financial Time Series. Chichester: Wiley.

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

Titterington, D.M., Smith, A.F.M., Makov, K.E. (1985), Statistical Analysis of Finite Mixture Distributions. Chichester: Wiley.

Tucker, A.L. (1985), Empirical tests of the efficiency of the currency otpion market. Journal of Financial Research, Vol. 8, 275-285.

Tucker, A.L. (1991), Exchange rate jumps and currency options pricing. In: S.J. Koury (ed.), Recent Developments in International Banking, Elsevier, 441-459.

Tucker, A.L., Peterson, D.R., Scott, E. (1988), Tests of the Black-Scholes and constant elasticity of variance currency call option valuation models. Journal of Financial Research, Vol. 11, 201-213.

Tucker, A.L., Pond, L. (1988), The probability distribution of foreign exchange price changes: Tests of candidate processes. Review of Economics and Statistics, Vol. 70, 638-647.

Tsay, R.S. (1987), Conditional Heteroscedastic Time Series Models. Journal of the American Statistical Association, Vol. 82, 590-604.

Weiss, A.A. (1986), ARCH and bilinear time series models: Comparison and combination. Journal of Business and Economic Statistics, Vol. 4, 59-70.

Westerfield, J.M. (1977), An examination of foreign exchange risk under fixed and floating rate regimes. Journal of International Economics, Vol. 7, 181-200.

Wiggins, J.B. (1987), Option values under stochastic volatility: Theory and empirical estimates. Journal of Financial Economics, Vol.19, 351-372.

Zimmermann, H. (1985), Eine Illustration zur Verteilung spekulativer Preise. Jahrbücher für Nationalökonomie und Statistik, Vol. 200, 420-427.