Allg. Statistisches Archiv 81, 85 – 101, ISSN 0002-6018

Simple Diagnostic Procedures for Modeling Financial Time Series

Von FRANZ C. PALM, Maastricht, und PETER J.G. VLAAR, Amsterdam*

Summary: The finding that many financial data have a higher peak and fatter tails than implied by a normal distribution, has increased the use of non–normal distributions, possibly allowing for time–variation in the variance. However, most diagnostic tests are still based on the normal distribution. In this article, a simple transformation is proposed to obtain standard normally distributed residuals from residuals of a model estimated by pseudo maximum likelihood methods. An application to daily US dollar exchange rates for the period October 1983 to September 1992 and to weekly exchange rates for the European Monetary System, for the period April 4, 1979 to March 27, 1991, reveals that both the Student–t and the Bernoulli or Poisson mixtures of normals appropriately describe the leptokurtic behavior of daily data and that Bernoulli or Poisson mixtures account for the leptokurtosis and skewness of weekly data.

Zusammenfassung: Viele Kursänderungsraten bzw. Wertpapierrenditen weisen eine Dichte mit höherem Gipfel und dickeren Schwänzen als die normale Dichtefunktion auf. Die Modellierung dieses Phänomens hat zur Verwendung von nicht–Gausschen heteroskedastischen Verteilungen geführt. Für viele statistischen Verfahren gilt allerdings die Normalverteilung als Voraussetzung. In diesem Papier wird eine einfache Transformation der Residuen eines Modells für Finanzdaten vorgeschlagen mit dem Ziel, normalverteilte, homoskedastische Grössen zu erhalten, deren Eigenschaften mit Hilfe von üblichen Testverfahren überprüft werden können. In einer Analyse von Tageskursänderungsraten für den US Dollar im Zeitraum Oktober 1983 bis September 1992 werden die Student–t und Bernoulli– oder Poisson–Mischungen von Normalverteilungen für die Tageskursänderungsraten verifiziert. Wochenkursdaten für das Europäische Währungssystem im Zeitraum vom 4. April 1979 bis zum 27. März 1991 besitzen Bernoulli– oder Poisson–Mischungen.

1. Introduction

It is widely known that high frequency financial data series exhibit leptokurtic behavior (fatter tails than with the normal distribution). One way to explain this high kurtosis might be the presence of conditional heteroskedasticity, a feature also very common for these data series. This heteroskedasticity can be modeled by means of an ARCH (ENGLE, 1982) or GARCH (BOLLERSLEV, 1986) specification. These specifications have been very popular in econometrics since their introduction. A survey article by BOLLERSLEV/CHOU/KRONER (1992) cited more than 300 papers applying these or closely related techniques. A survey of more recent GARCH–models is given in PALM (1996). Most of these studies conclude that the GARCH specification is very successful in reducing conditional heteroskedasticity, but not so in explaining the high kurtosis.

* The research was sponsored by the Economics Research Foundation which is part of the Netherlands Organization for Scientific Research (NWO). The authors like to thank Jim Lindsey for helpful comments. Errors are all theirs.

That is why other distributions than the normal one, with or without GARCH specification, have been applied (for applications to daily exchange rate changes see e.g. McFarland/Pettit/Sung (1982, 1987), So (1987), Boothe/Glassman (1987), Tucker/Pond (1988), Akgiray/Booth (1988), Jorion (1988), Hsieh (1989) or Baillie/Bollerslev (1989)). Candidate distributions with fatter tails than the normal one are among others the Student-t, a normal-Poisson mixture, a discrete mixture of normals, a generalized error distribution, a normal-lognormal distribution and a stable Paretian distribution. Since the normal distribution is nested in these other distributions, one can easily test for normality. However, as the other distributions are not nested, one can not use the standard tests to discriminate between them.

The selection of an appropriate distribution is particularly difficult since most diagnostic tests are based on the normal distribution. Although the asymptotic properties under other distributions are often known for these diagnostics, their behavior in finite samples is mostly unknown. One can, of course, simulate the small sample properties, as has for instance been done for the skewness and kurtosis parameters for the Student-t distribution by De Ceuster/Trappers (1992), but especially for discrete mixture distributions it is simply impossible to tabulate confidence intervals since there are too many parameters determining the distribution. Alternatively, when standardizing the data by using (unconditional) interquantile ranges or standard-deviations, both higher and lower peaks than in the standard normal case can be obtained as has been shown by Krämer/Runde (1996).

Most studies use the value of the log-likelihood function to discriminate between candidate distributions. Tucker/Pond (1988) and Akgiray/Booth (1988) for instance, use the Schwarz (1978) criterion to discriminate between four different distributions. However, a high value of the log-likelihood function does not necessarily mean that the shape of the proposed distribution resembles the distribution of the data. Although the parameter estimates might be consistent under a wider class of distributions, see Gouriéroux/Monfort/Trognon (1984), a Pearson χ^2-goodness-of-fit test might be preferred since several important applications of these models, for instance option pricing (Ball/Torous, 1985), also rely on the proposed distribution. This test statistic, however, can not easily be computed for mixture distributions with a time-varying variance since the statistic requires independent and identically distributed observations.

In this paper we propose a simple transformation to overcome the problems related to non-normal distributions. This transformation is based on the values of the cumulative distribution function of the residuals. For a correctly specified likelihood function, these values should be independently, uniformly distributed between zero and one. The Pearson χ^2-goodness-of-fit test can serve as a direct test. From these values, we compute standard normal residuals by means of the inverse of the cumulative normal distribution function. The method is applied to daily US dollar rates of the British pound, the German mark, the Japanese yen and the Swiss franc from October 1983 to September 1992. Four different distributions are investigated, the normal, the Student-t, the Bernoulli–normal mixture and the Poisson–normal mixture. All the non-normal distributions seem to give appropriate results. For comparison, we also report some results for weekly exchange rates in the European Monetary System (EMS) for the period April 4, 1979 to March 27, 1991. The conclusions are similar to those for the daily rates.

The structure of the paper is as follows. Section 2 describes the test statistics. Section 3 contains the results for the US dollar rates and the EMS rates and section 4 concludes.

2. Test statistics

Since the properties of most procedures based on the likelihood method crucially rely on the correctness of the underlying distribution, the Pearson goodness-of-fit test statistic seems to be most useful. A Pearson goodness-of-fit test compares the empirical distribution of the (standardized) residuals with the theoretical distribution[1]. This is usually done by classifying the residuals in cells according to their magnitude. From the theoretical distribution function the borders of the groups are computed in such a way that the expected number in each group is the same (although this is not required). The choice of the number of groups is not evident, but should increase with the number of observations at a speed $T^{0.4}$ (KENDALL/STUART, 1967). For independent identically distributed observations it can be shown that

$$\sum_{i=1}^{g} \frac{(n_i - En_i)^2}{En_i} \sim \chi^2(g-1),$$

where g is the number of groups[2], and n_i is the number of observations in group i. A problem with this test statistic for models with a changing variance over time is that the residuals of these models are neither identically nor independently distributed. For the normal or Student-t distribution, this problem can

[1] The same null hypothesis can be tested by means of other test statistics as well. For an overview of the literature on the applicability of these tests in the case of the estimated parameter, see HECKMAN (1984).

[2] It is often argued that if some of the parameters are estimated, the asymptotic distribution is actually bounded between $\chi^2(g-1)$ and $\chi^2(g-k-1)$, where k is the number of estimated parameters in the likelihood function. The idea behind this adjustment is the assumption that the test statistic is minimized at the estimated parameter values, whereas these values need not be the true values. However, for ungrouped data the minimum need not be reached at the estimated parameter values, so the test statistic might even be biased upwards (see also CHERNOFF/LEHMANN (1954)). In the literature several adjustments of this test statistic have been proposed, see HECKMAN (1984), TAUCHEN (1985) and ANDREWS (1988) to compensate for the estimated parameters. In order to derive these adjustments, however, it is implicitly assumed that the estimated parameter values are the true ones.

be overcome by standardizing the residuals. For the discrete mixture distributions, however, such a standardization will not be of help since the standardized residuals of these models still have time dependent third and fourth moments which depend on the conditional variance.

This problem can be solved by redefining the classifying mechanism of the residuals. Instead of classifying according to their value, for each residual we calculate the probability of observing a smaller value. This procedure implies checking whether the empirical cumulative distribution function is uniformly distributed:

$$H_0 : G(res_t, \widehat{\Phi}) \sim U[0:1],$$

where $G(res_t, \widehat{\Phi})$ is the value of the cumulative distribution function for residual t (denoted as res_t), given the proposed distribution (G) and the estimated parameters ($\widehat{\Phi}$). The sorting mechanism for this test statistic is:

$$n_i = \sum_{t=1}^{T} I_{it}, \text{ where } I_{it} = \begin{cases} 1 \text{ if } (i-1)/g < G(res_t, \widehat{\Phi}) \leq i/g \\ 0 \text{ otherwise} \end{cases} \qquad \forall\, 1 \leq i \leq g.$$

For the normal or Student-t distribution, the results of this classifying mechanism are exactly the same as those obtained by grouping standardized residuals according to value. Instead of using the inverse of the cumulative distribution function to compute the borders of the groups given previously set probabilities, this mechanism computes probabilities given the values of the residuals using the cumulative distribution function.

The goodness-of-fit test requires independent observations. This means that a rejection of the null hypothesis of this test might be due to dependence in the data, which might be remedied by a modification in the ARMA-GARCH specification. If the data should be normally distributed, independence could be checked by means of a Ljung-Box or Box-Pierce statistic. Given the fact that there is a one to one relationship between the value of a standard normal residual and the value of its cumulative distribution function, we can compute „normalized" residuals in the following manner:

$$res_t^n \equiv F^{-1}(G(res_t, \widehat{\Phi})),$$

where $F^{-1}(.)$ is the inverse of the standard normal cumulative distribution function (see also PEARSON (1933) and SMITH (1985)). Under the assumption of a correctly specified model, res_t^n is standard normally distributed, and all the usual diagnostic tests which require normality can be applied to check this.

3. US dollar and EMS exchange rates

The test statistics are applied to daily US dollar rates, and weekly EMS rates, obtained from Datastream. The daily data are middle rate notations from Barclays bank international for the British pound (BP), the German mark (DM), the Japanese yen (JY) and the Swiss franc (SF) from October 12, 1983 to September 23, 1992 (2252 observations). The EMS rates consist of 626 weekly Wednesday closing rates from the London Eurocurrency market in terms of the German mark from April 4, 1979 to March 27, 1991. For the British pound, the sample runs until October 3, 1990. For comparison reason, weekly US dollar — German mark rates are also analyzed.

Table 1: Summary statistics for log US dollar exchange rates

		BP	DM	JY	SF
ADF	s	-2.388	-1.644	-1.185	-1.582
Mean ($\times\ 10^2$)	Δs	-.005	-.024	-.029	-.021
St.dev. ($\times\ 10^2$)	Δs	.736	.737	.632	.776
Min. ($\times\ 10^2$)	Δs	-5.557	-5.785	-4.909	-5.578
Max. ($\times\ 10^2$)	Δs	3.081	3.037	3.381	3.179
$Q_s(25)$	Δs	47.918***	43.605**	33.730	35.039*
$Q_s^a(25)$	Δs	35.820*	37.749**	27.107	30.501
$Q_{ss}(25)$	Δs	211.606***	108.698***	96.537***	100.222***
m_3	Δs	-.189***	-.255***	-.441***	-.278***
m_4	Δs	3.475***	2.935***	4.008***	2.128***

The sample consists of 2252 daily observations from October 12, 1993 to September 23, 1992. s is the logarithm of the spot rate expressed in domestic currency per US dollar. ADF is the augmented Dickey–Fuller test with a constant, a trend and 4 lags of the differenced series. $Q_s(25)$ and $Q_{ss}(25)$ are Box–Pierce statistics of the raw and squared data respectively. $Q_s^a(25)$ is a Box–Pierce statistic, adjusted for ARCH like heteroskedasticity (DIEBOLD, 1987). m_3 and m_4 are the skewness and excess kurtosis parameters respectively. Under the assumption of normality they are normally distributed with expectation 0 and variance $6/T$ and $24/T$ respectively.
* (**) [***] indicates significance at the 10% (5%) [1%] level (assuming normality).

In table 1 some summary statistics for the log dollar rates are given. The first line clearly shows that the unit root hypothesis can not be rejected for US dollar exchange rates. This confirms results of previous studies. The significant results for the Box–Pierce test for the squared data and for the excess kurtosis are also similar to the findings in previous studies. The findings for the Box–Pierce statistic of the raw data and for the skewness, however, differ from the usual results. There seems to be correlation in the mean since the Box–Pierce statistic, $Q_s(25)$, is significant for the British pound, the German mark and the Swiss franc at the 1, 5 and 10% level respectively. As these significant results might be due to the presence of conditional heteroskedasticity, we also computed the ARCH–adjusted Box–Pierce test, $Q_s^a(25)$ (DIEBOLD, 1987). Although less significant, correlations seem to persist, especially for the German mark. The (positive) serial correlation might be due to noise traders with positive feedback strategies (DE LONG/SHLEIFER/SUMMERS/WALDMANN, 1990) or to the use of stop-loss strategies (KRUGMAN/MILLER, 1993). Finally, the skewness parameter, m_3, indicates that the distributions of the series are not symmetric. The negative skewness is probably due to the 5% depreciation of the dollar relative to the other currencies after the Plaza agreement in September 1985. The negative skewness does not necessarily mean that the symmetric Student-t distribution cannot fit the data since even a skewness parameter of -.441 is not significant at the 5% level for a t distribution with 5.4 degrees of freedom (DE CEUSTER/TRAPPERS, 1992, table 9).

Since a GARCH specification might also explain the high kurtosis we first estimated models with a normal distribution. The serial correlation in the mean is modeled by means of a MA(1) specification, and the conditional heteroskedasticity by a GARCH(1,1) specification. Day of the week and holiday effects were investigated and found to be significant for the variance specification, but not for the mean equation.[3]

The log-likelihood of the resulting model for the normal distribution has the following form:

$$\ln(L_{norm}) = \sum_{t=1}^{T}(-0.5\,(\ln(2\,\pi\,h_t^2) + \varepsilon_t^2/h_t^2)), \tag{1}$$

where ε_t and h_t^2 are defined as

$$\varepsilon_t = \Delta s_t - \mu - \varphi\,\varepsilon_{t-1} \tag{2}$$

and

$$\begin{aligned} h_t^2 \;=\;& \gamma_{hol}\,N_{hol,t} + \gamma_M\,D_{M,t} + \gamma T\,D_{T,t} + \gamma_W\,D_{W,t} + \gamma_H\,D_{H,t} + \\ & \gamma_F\,D_{F,t} + \alpha_1\,\varepsilon_{t-1}^2 + \beta_1\,h_{t-1}^2. \end{aligned} \tag{3}$$

$N_{hol,t}$ is the number of days between two successive trading days, excluding weekends. $D_{M,t}$ to $D_{F,t}$ are dummy variables that refer to the day of the week. The results of maximum likelihood estimation are shown in table 2. The MA(1)-GARCH(1,1) specification seems to be appropriate to model the time dependence in the data since both the correlations in the residuals and in the squared residuals are insignificant.

Although the t-value for the MA parameter is not significant at the 5% level for the German mark and the Swiss franc, a likelihood ratio test indicated that the parameter should be included. Moreover, if the parameter is not included for the German mark, the residuals are significantly correlated. For the Japanese yen, the MA parameter is not significant.

[3] These findings are at odds with the results of MCFARLAND et al. (1982), who find significant daily differences in the mean, but not in the variance. In his comment, SO (1987) showed that the daily differences were primarily due to the neglected skewness in the data. Much more in line with our findings is the study by HSIEH (1988) and BAILLIE/BOLLERSLEV (1989), who also found significant daily effects in the variance but only minor differences in the mean.

Table 2: Maximum likelihood results for the normal distribution

	BP	DM	JY	SF
Coefficients				
μ ($\times 10^4$)	-1.223	-2.369	-2.515**	-1.090
	(0.83)	(1.61)	(2.07)	(0.69)
φ	0.065***	0.034		0.042*
	(2.87)	(1.45)		(1.90)
γ_{hol} ($\times 10^4$)	0.119***	0.133***	0.180***	0.155**
	(2.33)	(2.99)	(3.11)	(2.11)
γ_M ($\times 10^4$)	0.026	0.013*	0.044	0.069
	(0.54)	(1.32)	(0.70)	(1.14)
γ_T ($\times 10^4$)	0.000	0.000	0.000	0.000
	(0.04)	(0.00)	(0.00)	(0.00)
γ_W ($\times 10^4$)	0.000	0.000	0.000	0.000
	(0.04)	(0.01)	(0.10)	(0.00)
γ_H ($\times 10^4$)	0.000	0.000	0.012	0.000
	(0.01)	(0.00)	(0.27)	(0.03)
γ_F ($\times 10^4$)	0.090*	0.091**	0.123**	0.122**
	(1.63)	(1.97)	(2.17)	(1.66)
α_1	0.077***	0.081***	0.073***	0.068***
	(4.32)	(6.64)	(3.81)	(5.20)
β_1	.873***	.859***	.822***	.860***
	(20.74)	(34.85)	(14.88)	(17.86)
Diagnostics				
$\ln(L_{norm})$	7976.8	7951.6	8291.7	7809.1
$\chi^2(49)$	116.86***	100.78***	202.38***	79.81***
$Q_s(25)$	21.16	34.49*	33.16	29.91
$Q_{ss}(25)$	20.14	16.92	15.43	25.43
m_3	-.004	-.220***	-.617***	-.284***
m_4	1.852***	2.284***	4.760***	1.747***

The $\chi^2(49)$ statistic is a χ^2-goodness-of-fit test computed on 50 cells. Asymptotic absolute heteroskedasticity-consistent t-values are in parentheses.
* (**) [***] indicates significance at the 10% (5%) [1%] level.

The variance specification is quite successful. The GARCH parameters α_1 and β_1 are highly significant and the holiday effect is also very pronounced. The intercept differs very much for different days. Only the Friday parameter is significantly different from zero. The Monday coefficient is just not significant whereas the parameters for Tuesday, Wednesday and Thursday are almost zero. These results differ somewhat from the ones obtained by HSIEH (1988) who found a higher conditional variance on Mondays, but not on Fridays.

Looking at the diagnostics, it is clear that the normal distribution cum MA(1)-GARCH (1,1) does not account for important features of the data. The null hypothesis of the goodness-of-fit test is rejected for all currencies. Figure 1a shows the empirical distribution of the data over the cumulative distribution function, for the German mark.[4] If the model was correctly specified, the graph would show a uniform zero-one spread. This is clearly not the case for the normal distribution. There are more observations in the tails of the distribution and also more in the middle range. This leads to excess kurtosis, which is indeed significantly positive for all currencies. The skewness parameter is significant for three of the four currencies.

The second distribution considered is the Student-t. Since this distribution has fatter tails than the normal one, it seems to be a natural candidate. The log-likelihood function for this distribution has the following form:

$$\ln(L_T) = T\left[\ln\Gamma\left(\frac{v+1}{2}\right) - \ln\Gamma\left(\frac{v}{2}\right) - \frac{1}{2}\ln(\pi(v-2))\right] -$$

$$\frac{1}{2}\sum_{t=1}^{T}\left\{\ln\left(h_t^2\right) + (v+1)\ln\left(1+\frac{\varepsilon_t^2}{h_t^2(v-2)}\right)\right\}, \tag{4}$$

where $\Gamma(.)$ is the gamma function, v is the degrees of freedom parameter, and ε_t and h_t^2 are defined in (2) and (3) respectively. In table 3 the results for this distribution are shown. Since most of the coefficients are almost the same as for the normal distribution, they are not shown.

The number of degrees of freedom for the Student-t distribution ranges from 4.1 for the Japanese yen to 8.0 for the Swiss franc, indicating a finite fourth moment for all currencies. The t-value for v, which is the same as the t-value for $1/v$, is highly significant for all currencies. The importance of v can also be deduced from the values for the log-likelihood functions. The difference between these values and the corresponding values for the normal distribution is at least 36 points (Swiss franc).

m_3^{st} and m_4^{st} are the skewness and excess kurtosis for the standardized residuals. De Ceuster/Trappers (1992) constructed confidence intervals for these parameters for the Student-t distribution. Only the results for the skewness for the Japanese yen and the Swiss franc are outside the 90% confidence interval. If we compare these results for the same parameters, computed on the normalized residuals (m_3^n and m_4^n), we see that the results are quite similar.

The null hypothesis of the goodness-of-fit test is not rejected for the Student-t distribution for any of the currencies. The good fit for the German mark is visualized in figure 1b. All ranges of values of the cumulative distribution function appear to be observed equally often.

[4] Since the graphs for the other currencies were similar, only the results for the German mark are shown.

Figure 1: Frequency plots for the German mark under alternative distributions

Figure 1a: the normal distribution

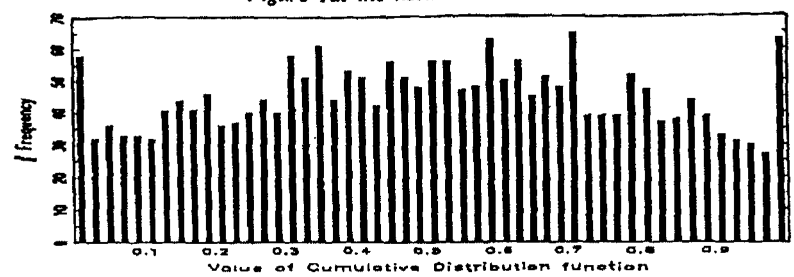

Figure 1b: the Student-t distribution

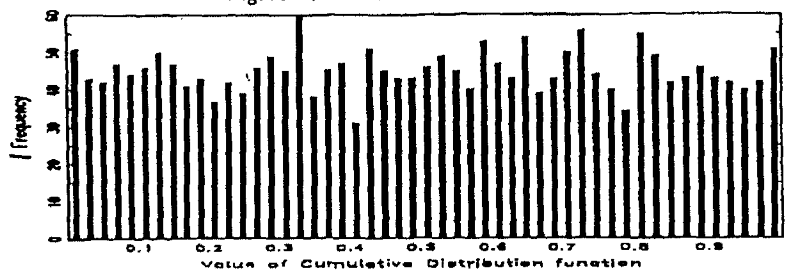

Figure 1c: the Bernoulli-normal mixture

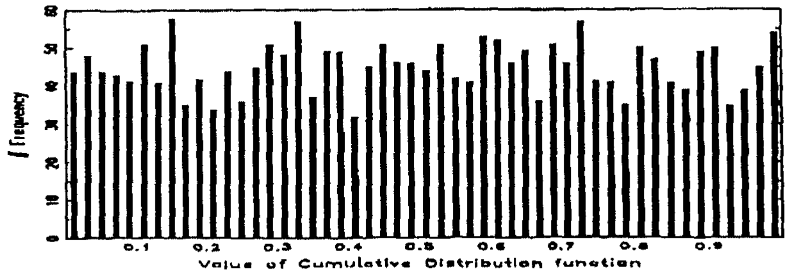

Figure 1d: the Poisson-normal mixture

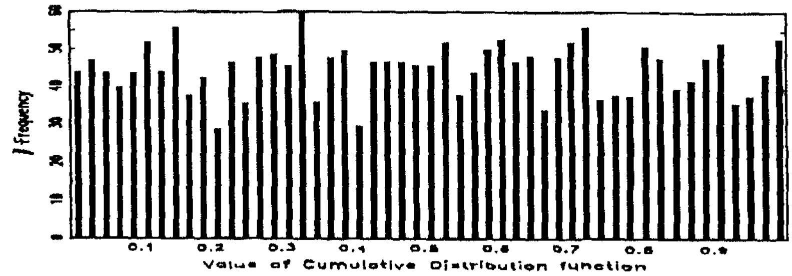

Table 3: Maximum likelihood results for the Student-t distribution

	BP	DM	JY	SF
v	5.745***	6.166***	4.102***	7.979***
	(8.31)	(8.02)	(10.88)	(5.88)
$\ln(L_T)$	8033.1	8006.2	8424.4	7845.1
m_3^{st}	-.005	-.230	-.737***	-.290**
m_4^{st}	1.905	2.356	5.954	1.785
$\chi^2(49)$	41.03	33.07	56.66	45.34
$Q_s^n(25)$	22.16	37.47*	34.29	32.56
$Q_{ss}^n(25)$	24.00	30.14	20.49	29.45
m_3^n	.050	-.045	-.118**	-.109**
m_4^n	-.049	-.030	-.059	-.010

For notes, see table 2.

The only problem with the Student-t distribution seems to be the skewness. The next distribution considered, the Bernoulli–normal mixture, accounts for skewness. If the mean of the second normal distribution differs from the first and the variance or the jump intensities differ, the resulting mixture is skewed (VLAAR/PALM, 1993). In order to derive the log-likelihood function, first the error term is defined in such a way that it has expectation zero:

$$\varepsilon_t = \Delta s_t - \mu - \lambda\,\theta - \varphi\,\varepsilon_{t-1}, \tag{2'}$$

where λ is the probability of a jump and θ is the expected jump size. The log-likelihood function has the following form:

$$\ln(L_B) = -\frac{T}{2}\ln(2\,\pi) + \sum_{t=1}^{T}\ln\left\{\frac{(1-\lambda)}{h_t}\exp\left(\frac{-(\varepsilon_t+\lambda\,\theta)^2}{2\,h_t^2}\right) + \right.$$
$$\left.\frac{\lambda}{\sqrt{h_t^2+\delta^2}}\exp\left(\frac{-(\varepsilon_t-(1-\lambda)\,\theta)^2}{2\,(h_t^2+\delta^2)}\right)\right\}, \tag{5}$$

where δ^2 is the variance of the jump size and ε_t and h_t^2 are defined in (2') and (3) respectively. In table 4 the estimation results are shown.

Although the t-values of the individual jump parameters are not always significant, their joined significance can be deduced from the likelihood ratio statistic of the Bernoulli–normal and normal distribution. The inclusion of the three parameters increases the log-likelihood value by at least 30.5 points (Swiss franc), which is highly significant at any conventional confidence level. The increase in log-likelihood value is lower for the Bernoulli–normal mixture than for the Student-t distribution. However, since these distributions are not nested it is difficult to draw conclusions from this observation.

As expected, the skewness is much less a problem for this distribution than for the Student-t distribution. The exception is the British pound, for which

Table 4: Maximum likelihood results for the Bernoulli–normalmixture

	BP	DM	JY	SF
λ	0.202**	0.215**	0.195***	0.094
	(2.14)	(1.93)	(3.44)	(0.64)
θ ($\times 10^2$)	0.112	-0.052	-0.179**	-0.304
	(1.49)	(0.63)	(2.35)	(0.96)
δ^2 ($\times 10^4$)	0.893***	0.885***	0.867***	1.169
	(3.04)	(3.04)	(4.09)	(1.22)
$\ln(L_B)$	8022.6	7996.2	8418.2	7839.6
$\chi^2(49)$	52.71	44.00	51.64	61.02
$Q^n_s(25)$	22.13	36.60*	33.83	31.43
$Q^n_{ss}(25)$	27.27	28.82	23.88	29.42
m^n_3	-.087*	-.064	-.037	-.018
m^n_4	.286***	.252**	.227**	.185*

For notes, see table 2.

the negative skewness was increased by the positive expected jump size. The main problem for the Bernoulli–normal mixture is the excess kurtosis, which is significantly positive for all currencies. As can be seen from figure 1c and the goodness–of–fit tests, this high kurtosis is not due to the overall fit. The number of observations in the upper and lower 2% tails of the distribution are not exceptionally high. It turns out that the significant skewness and kurtosis results are primarily caused by just one outlier.

Finally the Poisson–normal mixture is analyzed. This distribution models the possibility of more than one jump per unit of time. As such, the distribution has appealing martingale properties (HARRISON/PLISKA, 1981). The occurrence of several jumps might explain the very large outliers. The log–likelihood function has the following form:

$$\begin{aligned}\ln(L_P) &= -T\lambda - \frac{T}{2}\ln(2\pi) + \\ &+ \sum_{t=1}^{T} \ln\left\{\sum_{j=0}^{\infty} \frac{1}{\sqrt{h_t^2+\delta^2 j}} \exp\left(\frac{-(\varepsilon_t-(j-\lambda)\theta)^2}{2(h_t^2+\delta^2 j)}\right)\right\}. \quad (6)\end{aligned}$$

Although the possibility of a large number of jumps is theoretically appealing, it complicates the estimation considerably. In order to become estimable, the infinite sum has to be truncated. BALL/TOROUS (1985) give an upper limit for the truncation error. A truncation after eleven terms, which is also used in this study, seems to be appropriate for most applications.[5] However, even a

[5] The error increases with λ.

truncation after eleven terms leaves us with a likelihood function (combining (6), (2') and (3)) that is far more complex than the Bernoulli–normal mixture, which consists of only two terms.

In table 5, the maximum likelihood results are shown. The results resemble the Bernoulli–normal ones, but as expected the high kurtosis is reduced. Moreover, the values of the log–likelihood functions are somewhat higher than for the Bernoulli–normal mixture. Apart from the excess kurtosis for the British pound and the residual Box–Pierce test for the German mark, all diagnostic tests are passed for the Poisson–normal mixture.

Table 5: Maximum likelihood results for the Poisson–normal mixture

	BP	DM	JY	SF
λ	0.306**	0.341**	0.290***	0.256
	(2.11)	(2.21)	(3.07)	(0.71)
θ ($\times 10^2$)	0.072	-0.043	-0.135**	-0.178
	(1.24)	(0.79)	(2.35)	(1.09)
δ^2 ($\times 10^4$)	0.658***	0.605***	0.604***	0.640*
	(3.23)	(3.00)	(3.74)	(1.33)
$\ln(L_P)$	8024.3	7999.3	8421.8	7841.1
$\chi^2(49)$	47.07	48.80	41.91	54.93
$Q^n_s(25)$	22.30	37.02*	34.19	31.81
$Q^n_{ss}(25)$	32.80	31.42	24.81	28.67
m^n_3	-.068	-.040	-.019	-.007
m^n_4	.210**	.148	.164	.154

For notes, see table 2.

The results for the non–normal distributions are somewhat better than those obtained by other authors. For all the non–normal distributions investigated, the null hypothesis of the goodness–of–fit test is not rejected for any of the currencies. BOOTHE/GLASSMAN (1987) on the other hand, found that the Stable Paretian, Student–t and Bernoulli–normal mixture models for the currencies considered in this study did not pass the goodness–of–fit test, using daily data over the period 1973–1984. This rejection might be due to the lack of a GARCH specification. In the study of HSIEH (1989), only the normal–lognormal mixture distribution passes the goodness–of–fit test for all currencies, using daily data over the period 1974–1983. Apart from changes in the sample period, the differences might be due to the use of a two step procedure in Hsieh's study. In the first step the model is estimated, assuming a normal distribution, and in the second step the standardized residuals from the first step are confronted with other distributions. Although the maximum likelihood parameters for the normal distribution are also consistent for various other distributions under mild conditions (GOURIÉROUX/MONFORT/TROGNON, 1984), VLAAR/PALM (1993) showed that the parameter estimates may vary considerably with the assumed distribution. A direct estimation, as performed in this study, might be preferred.

Table 6: Maximum likelihood results for the Poisson-normal mixture for weekly data

Currency	μ ($\times 10^4$)	φ	α_0 ($\times 10^6$)	α_1	β_1	λ	θ ($\times 10^2$)	δ^2 ($\times 10^4$)
BF	-.595	-.358	.817	.198	.683	.036	.850	1.589
	(.67)	(7.95)	(1.92)	(2.94)	(6.58)	(2.01)	(1.47)	(1.12)
DG	-.501	-.333	.079	.259	.661	.082	.112	.199
	(1.04)	(6.47)	(1.12)	(3.34)	(6.12)	(2.04)	(1.39)	(1.87)
FF	1.129	-.139	4.196	.156		.037	.932	3.227
	(1.29)	(3.97)	(8.32)	(2.87)		(2.61)	(2.06)	(1.88)
DK	.542	-.155	7.053	.151		.043	1.031	1.198
	(.49)	(3.17)	(10.82)	(3.97)		(2.72)	(3.06)	(2.16)
IP	1.164	-.212	4.679	.296	.121	.039	.575	2.640
	(1.09)	(5.25)	(3.40)	(2.04)	(1.30)	(.83)	(1.12)	(.61)
IL	-1.493	-.146	2.467	.157	.428	.138	.547	.841
	(.90)	(3.63)	(2.17)	(2.49)	(3.07)	(2.19)	(3.11)	(1.66)
BP	-7.309	.065	1.226	.143	.585	.815	.159	.991
	(1.73)	(1.30)	(.43)	(1.84)	(1.49)	(1.88)	(1.40)	(2.80)
US$	-40.777		10.475	.118	.717	1.553	.280	.620
	(1.15)		(.58)	(2.58)	(9.45)	(.78)	(2.23)	(1.21)

Absolute asymptotic heteroskedasticity-consistent t-values are in parentheses.

Empirical results for weekly data DM-rates using a MA(1)-GARCH (1,1) model with a Poisson-normal mixture distribution are given in table 6. The countries under consideration are those participating in the exchange rate mechanism of the EMS since the beginning on April 4, 1979, i.e. Belgium, France, Denmark, Ireland, Italy and The Netherlands, and for reason of comparison, the United States and the United Kingdom. For the UK pound only the free-float period (until October 3, 1990) is considered, whereas for the ERM countries the sample ends on March 27, 1991 (626 weekly observations). The model consists of the equations (2) and (3) where the dummy variables have been deleted and replaced by a positive intercept α_0. The likelihood function is given in expression (6). Results for these data using the same model with a normal distribution and a Bernoulli-normal mixture can be found in VLAAR/PALM (1993) where also some summary statistics are reported for the data. The Bernoulli and Poisson specifications give similar results for all ERM currencies. The estimated jump intensities are slightly higher for the Poisson model, resulting in a lower jump size. The MA coefficient is negative and significant in most instances. For all currencies the expected jump size θ is positive, which is in accordance with positive skewness. The standard deviation of the jump size δ is much bigger than its mean θ. Apart from large depreciations, jumps might also represent large appreciations, for instance after an important policy announcement (German unification). The influence on volatility is most important. Model diagnostics are given in table 7. For comparison, diagnostic statistics for the model under a normal distribution are also included in table 7.

Table 7: Model diagnostics for weekly data

Distribution	ln(L)	$\chi^2(19)$	$Q_e(25)$	$Q_{e^2}(25)$	Skewness	Exc.kurt.
BFnormal	2560.2	66.5***	29.5	11.4	2.29***	21.31***
Pois/normal	2654.7	18.7	35.2*	22.4	-.02	.57***
DGnormal	3010.8	53.8***	32.0	14.9	.59***	5.09***
Pois/normal	3065.8	10.5	31.7	15.0	.06	.11
FFnormal	2522.6	210.0***	34.6*	1.9	4.36***	43.77***
Pois/normal	2802.6	38.7***	24.2	22.0	.11	.33*
DKnormal	2534.1	110.2***	28.1	21.8	1.92***	9.84***
Pois/normal	2656.4	21.8	25.0	40.3**	.07	.20
IPnormal	2534.8	139.8***	47.6***	4.4	2.81***	22.77***
Pois/normal	2675.6	34.8**	40.0**	40.4**	.13	.49**
ILnormal	2401.3	149.1***	34.2	36.6*	2.35***	20.31***
Pois/normal	2563.8	18.7	36.2*	22.5	.10	.51***
BPnormal	1856.4	47.8***	32.3	16.6	.39***	2.01***
Pois/normal	1887.8	14.7	35.3*	26.0	.05	.05
US$normal	1727.8	29.0*	25.7	22.1	.28***	.45**
Pois/normal	1731.4	23.6	26.6	23.7	.06	.03

- $\chi^2(19)$ is an adjusted Pearson goodness-of-fit test performed on a classification in 20 cells.
- The Box-Pierce statistics $Q_e(25)$ and $Q_{e^2}(25)$ and the skewness and excess kurtosis are computed on normalized residuals.
- * (**) [***] indicates significance at the 10% (5%) [1%] level.

The first row for each currency clearly shows the inappropriateness of the MA-GARCH normal model for ERM currencies. The smallest value of this $\chi^2(19)$ statistic, that for the Dutch guilder, is still 54. Also noticeable is the rejection according to this statistic of the model for the British pound. Although the excess kurtosis and skewness for this currency are much less extreme than for the European Exchange Rate (ERM) currencies, normality is still rejected.

For the Poisson-normal mixture (and the Bernoulli-normal mixture), the results improve tremendously. For 6 out of 8 countries, the mixture models are not rejected at the 5% level. Only for the French franc and the Irish pound, the goodness-of-fit test rejects the mixture model for reason of excess kurtosis.

The improvement of the fit for ERM currencies can also be seen from the first column of table 7. The log-likelihood function increases by values between 53 (Dutch guilder) and 280 (French franc) points when jumps are allowed for. Correlations of the residuals and the squared residuals are insignificant for all currencies, except the Irish pound. Correlations in the mean and time-varying volatility seem to be correctly modeled when using a Box-Pierce test based on standardized residuals in the case of a normal distribution and normalized residuals in the case of a Poisson-normal mixture distribution.

Using the skewness and excess kurtosis statistics, normality is clearly rejected for all MA-GARCH models, even for the US dollar. For the Poisson-normal distribution, the results are much better. Some excess residual kurtosis is found for the Belgian franc, the Irish pound and the Italian lira.

From these results, we conclude that the Poisson-normal mixture performs quite well for the weekly ERM exchange rates data, except for the French franc and the Irish pound. Results for Bernoulli-normal mixtures are similar. These findings for weekly exchange rate data are very much in accordance with the results for daily observations reported above.

4. Conclusions

In this paper a simple transformation is proposed to obtain standard normal residuals from models with non-normal distributions. This transformation makes it possible to use and interpret all kinds of diagnostic tests that assume the normal distribution. The method uses the cumulative distribution function to compute for each residual the probability of getting a smaller value than the one observed. If the model is correctly specified, these probabilities should be independently, uniformly zero-one distributed. A Pearson χ^2-goodness-of-fit test can serve as a direct test. Given these probabilities, the normalized residuals are computed by means of the inverse of the standard normal cumulative distribution function.

The method was first applied to daily US dollar rates for the British pound, the German mark, the Japanese yen and the Swiss franc over the period October 1983 to September 1992. An MA(1)-GARCH(1,1) specification was investigated using four different distributions: the normal, the Student-t, the Bernoulli-normal mixture and the Poisson-normal mixture. The normal distribution was rejected for its thin tails. For the other three, the null hypothesis of the goodness-of-fit test was not rejected. The results for the skewness and kurtosis were less successful. These parameters were influenced to a large extent by one big negative outlier. As a result of this, symmetry was rejected for the residuals of the Student-t model for the Japanese yen and the Swiss franc. Moreover, the excess kurtosis parameter for all Bernoulli-normal models and for the Poisson-normal model for the British pound was significantly positive.

Next, the method was applied to weekly DM-rates for ERM currencies, the British pound and the US dollar for the period April 4, 1979 to March 27, 1991 (October 3, 1990 for the British pound). Again normality was rejected in favor of a Poisson-normal or a Bernoulli-normal mixture distribution. Similar results were also found by VLAAR/PALM (1996) for excess returns in the European Monetary System.

5. References

ANDREWS, D.W.K. (1988): Chi-Square Diagnostic Tests for Econometric Models: Theory; Econometrica, 56, 1419–1453.

AKGIRAY, V. / BOOTH, G.G. (1988): Mixed Diffusion-Jump Process Modeling of Exchange Rate Movements; The Review of Economics and Statistics, 70, 631–637.

BAILLIE, R.T. / BOLLERSLEV, T. (1989): The Message in Daily Exchange Rates: A Conditional-Variance Tale; Journal of Business and Economic Statistics, 7, 297–305.

BALL, C. / TOROUS, W. (1985): On Jumps in Stock Prices and Their Impact on Call Option Pricing; Journal of Finance, 40, 155–173.

BOLLERSLEV, T. (1986): Generalized Autoregressive Conditional Heteroskedasticity; Journal of Econometrics, 31, 307–327.

BOLLERSLEV, T. / CHOU, R.Y. / KRONER, K. (1992): ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence; Journal of Econometrics, 52, 5–59.

BOOTHE, P. / GLASSMAN, D. (1987): The Statistical Distribution of Exchange Rates; Journal of International Economics, 22, 297–319.

CHERNOFF, H. / LEHMANN, E. (1954): The Maximum Likelihood Estimate in χ^2 Tests for Goodness of Fit; Annals of Mathematical Statistics, 25, 579–586.

DE CEUSTER, M. / TRAPPERS, D. (1992): Diagnostic Checking of Estimation with a Student-*t* Error Density; Paper presented at the Econometric Society European Meeting (ESEM), Brussels, August 1992.

DE LONG, J.B. / SHLEIFER, A. / SUMMERS, L.H. / WALDMANN, R.J. (1990): Positive Feedback Investment Strategies and Destabilizing Rational Speculation; Journal of Finance, 45, 379–395.

DIEBOLD, F.X. (1987): Testing for Serial Correlation in the Presence of ARCH; Proceedings from the American Statistical Association, Business and Economics Section, 323–328.

ENGLE, R. (1982): Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation; Econometrica, 50, 987–1007.

HARRISON, M. / PLISKA, S. (1981): Martingales and Stochastic Integrals in the Theory of Continuous Trading; Stochastic Processes and their Applications, 11, 215–260.

GOURIÉROUX, C. / MONFORT, A. / TROGNON, A. (1984): Pseudo Maximum Likelihood Methods: Theory; Econometrica, 53, 681–700.

HECKMAN, J.J. (1984): The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata; Econometrica, 52, 1543–1547.

HSIEH, D.A. (1988): The Statistical Properties of Daily Foreign Exchange Rates: 1974 – 1983; Journal of International Economics, 24, 129–145.

HSIEH, D.A. (1989): Modeling Heteroskedasticity in Daily Foreign-Exchange Rates; Journal of Business and Economic Statistics, 7, 307–317.

JORION, P. (1988): On Jump Processes in the Foreign Exchange and Stock Markets; The Review of Financial Studies, 1, 427–445.

KENDALL, M. / STUART, A. (1967): The Advanced Theory of Statistics, Vol. 2, Inference and Relationship; Haffner, New York.

KRÄMER, W. / RUNDE, R. (1996): Peaks or Tails – What Distinguishes Financial Data?; University of Dortmund, Research paper.

KRUGMAN, P. / MILLER, M.H. (1993): Why have a Target Zone?; Carnegie–Rochester Series on Public Policy, nr. 38, 279–314.

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, 37, 693–715.

MCFARLAND, J.W. / PETTIT, R.R. / SUNG, S.K. (1987): The Distribution of Foreign Exchange Price Changes: Trading Day Effects and Risk Measurement: A Reply; Journal of Finance, 42, 189–194.

PALM, F.C. (1996): GARCH Models of Volatility; forthcoming in Maddala, G.S. / Rao, C.R. / Vinod, H.D. (eds.), Handbook of Statistics, Vol. 14, Elsevier Science Publishers, 209–240.

PEARSON, K. (1933): On a Method of Determining whether a Sample of Size n Supposed to have been Drawn from a Parent Population Having a Known Probability Integral has Probably been Drawn at Random; Biometrika, 25, 379–410.

SCHWARZ, G. (1978): Estimating the Dimension of a Model; Annals of Statistics, 6, 461–464.

SMITH, J.Q. (1985): Diagnostic Checks of Non–Standard Time Series Models; Journal of Forecasting, 4, 283–291.

SO, J.C. (1987): The Distribution of Foreign Exchange Price Changes: Trading Day Effects and Risk Measurement — A Comment; Journal of Finance, 42, 181–188.

TAUCHEN, G. (1985): Diagnostic Testing and Evaluation of Maximum Likelihood Models; Journal of Econometrics, 30, 415–443.

TUCKER, A.L. / POND, L. (1988): The Probability Distribution of Foreign Exchange Price Changes: Tests of Candidate Processes; The Review of Economics and Statistics, 70, 638–647.

VLAAR, P.J.G. / PALM, F.C. (1993): The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity and Jumps; Journal of Business and Economic Statistics, 11, 351–360.

VLAAR, P.J.G. / PALM, F.C. (1996): Inflation Differentials and Excess Returns in the European Monetary System; paper forthcoming in Journal of International Financial Markets, Institutions and Money.

Prof. Dr. Franz Palm
Department of Quantitative Economics
Maastricht University
6200 MD Maastricht
P.O. Box 616

Dr. Peter J.G. Vlaar
Econometrics Department
De Nederlandsche Bank N.V.
1000 AV Amsterdam
P.O. Box 98