Tijdschrift voor Economie en Management
Vol. XL, 2, 1995

Risk Assessments in a Markov Switching Framework

by H. DEWACHTER*

I. INTRODUCTION

How do foreign exchange rates evolve over time? Over the last decades an enormous literature has focused on this question. By now, there is a consensus that they follow, at least when high frequencies are concerned, a martingale model. According to this view, returns are unpredictable in mean, but higher moments may have some structured behaviour. Some of these structures have been filled in during the eighties. Returns are believed to display volatility clustering. More specifically, high and low returns tend to be clustered in time.

Despite the early reports on volatility clustering, see (Mandelbrot (1963)), models that incorporate this fact were only introduced in the mid eighties. With the introduction of ARCH and GARCH models by Engle (1982) and Bollerslev (1986), a new literature in empirical finance came to live. Currently, almost over fifty different variations exist on the first GARCH and ARCH specifications. While these models have been very successful, there remain some less desirable properties of the model to resolve. More specifically, according to the model, variances are sufficient statistics to model the variability of the return, i.e. news, distribution over time. Consider the simple GARCH model of Bollerslev for example. According to this model, returns are drawn from a normal distribution with time varying variances. An im-

* Rijksuniversiteit Limburg, Maastricht, Nederland.
The main part of this paper was written while I was at CEMFI (Madrid). The author would also like to thank Jelle van der Ploeg for valuable research assistance and the referee for interesting comments on an earlier draft of the paper. Of course all remaining errors are my own.

plication of this model is then that exceedance probabilities are determined by this variance and the normality assumption. In other words, one does not consider an independent and time varying model for moments higher than the variance. Within the framework of GARCH such a modelling exercise becomes extremely difficult. It entails a workable model for the time varying structure for the higher moments as well. No such model exists today.

Recently, a second generation of models has been used to analyse the volatility clusters. These models can (but do not yet), in contradistinction with the GARCH class, combine the volatility clustering and (variance) independent variation in the higher moments. This class of models is known as the class of Markov switching models, see for example Hamilton (1989). The underlying philosophy of this approach is that returns are drawn from different (normal) densities through time. Which density is activated in each period depends on an unobserved variable, called the state. This state follows a first order Markov process. Volatility clustering is obtained via a very inert Markov process and distinct variance levels for each of the distributions. These models can combine, unlike GARCH models, this volatility clustering with tractable expressions for the conditional return densities. The latter feature is currently used to obtain for example option prices in the context of Markov switching models, (see for example Kaelher and Marnet (1993)).

While this new approach has a lot of potential, there remain some unnecessary restrictive assumptions in its setup. The assumption of the normality of the underlying densities is, especially when high frequency returns are concerned, not tenable. This assumption implies two contradictions with the established facts for high frequency return series; One, there are only two variance levels (instead of a continuum as for example in the GARCH models) and two, all moments exist by assumption. The latter result is clearly in contradiction with the literature on the extremal behaviour of returns (see Koedijk et al (1990) or Jansen and de Vries (1991)). Moreover, it is evident that in this case too, the higher moments are totally determined by the respective variance levels through the assumption that each regime is defined by a normal density. In contrast with the GARCH literature, however, one could model the higher order time variation without major complications in the framework of Markov switching models.

The purpose of this paper is to adjust the standard Markov switching model to resolve the issues raised above. More specifically, the main objective of this paper is to come up with a model that combines volatility clustering with at least some variance independent variation in the higher moments. Given the switching structure of the model, this can be accomplished in a rather straightforward manner. Such a model is obtained by substitution of noncentral Student-t distributions for the normal distributions. This substitution allows a variance independent modelling (at least partially) of the higher moments via the determination of the respective degrees of freedom for the densities associated with each regime. Distributions can therefore differ in more than their scales across regimes. Here the models adds new insight that are not obtained in the standard literature. In section II of the paper we show how this slight extension also allows for a continuum of variance states, in contradistinction with the standard Markov switching model. Moreover, we discuss the implications of these substitutions for risk assessments purposes. Clearly, other applications such as option pricing are also feasible. However, they fall outside the scope of this paper. In section III estimates and tests for both models, the standard and the T-Markov switching model, are reported. To anticipate, we find that the extended model outperforms the standard one. Finally, conclusions summarise the most significant findings.

II. THE T-MARKOV SWITCHING MODEL

Consider the following model. There is an unobserved state variable s_t. This variable can take two values 1 or 2, referred to as 'regimes' or 'states', and follows a first order, discrete state space discrete time, Markov process. This variable, s_t, can be interpreted as the 'type' of news releases in the foreign exchange market[1]. News is thus either of 'type 1' or 'type 2'. The exact properties of each 'news type' are defined by the parameters of the density, associated with each regime, from which the return is drawn. By assumption both densities are constrained to the class of normals, denoted by $N(\mu(s_t), \sigma^2(s_t, t))$. These densities, and their parameters, define the way in which the markets transform news into (log) price changes, i.e. the return. For example, news can be important or not. If and when important news arrives the reaction of the financial markets is likely to change accordingly; large price changes are to be expected. Unimportant news will be neglected by the markets, changing prices only marginally. Within the Markov switching framework, the above characterisation could be filled in as follows: there are two news states, labelled one and two. The first state refers to releases of important news, the other to unimportant news. The market reaction is modelled through the respective densities associated with each of the states. The density for state one will have a large variance, indicating the considerable price movements in that state other density will be relatively small compared to the former. Other types of news classifications are possible too. For example good versus bad news classes which would be modelled through the respective means.

The density within each state is normal with means dependent on the state and a variance level that depends both on the state, s_t, as on time, t. . More specifically, we assume that the precision, the reciprocal of the variance, is an i.i.d. draw from a gamma distribution, $G(\sigma^{-2};\lambda(s_t),\kappa(s_t))$, with state dependent parameters. The variance can, within each state, take any value on R^+ since the support of the gamma distribution is R^+. The parameters, that depend on the state, define the moments of the precision. They define thus the average variance in the state as well as the variance of the variance. A more detailed account of this part of the model can be found below. Less attention is paid to the respective means of the normal densities. They are a direct function of the state. Given the quasi martingale behaviour, it is evident that this part of the model will be less important.

The conditional density for the return X conditional on the unobserved state variable s, denoted by $F(X_t | S_t)$ is a compounded mixture of a normal measurement equation and a gamma transition equation, i.e.

$$F(X_t|s_t) = \int_0^\infty N(X_t;\mu(s_t),\sigma^2)G(\sigma^{-2};\lambda(s_t),\kappa(s_t))d\sigma^{-2} \quad (1)$$

It is well known in the statistical literature that this compounded distribution reduces to the noncentral Studentát distribution with location $\mu(s_t)$, scale $\vartheta(s_t)=\frac{\kappa(s_t)}{\lambda(s_t)}$ and degrees of freedom $\nu(s_t)=2\kappa(s_t)$. The moments of the return X, conditional on s_t, are then given by $\mu(s_t)$ and $\vartheta^{-1}(s_t)[\frac{\nu(s_t)}{\nu(s_t)-2}]$ respectively for mean and variance. All the parts discussed above give the model presented in (2):

$$X_t|s_t = i \sim t'(X_t;\mu_i,\vartheta_i,\nu_i)$$

$$t'(X_t;\mu_i;\vartheta_i;\nu_i) = \frac{\nu^{(\frac{1}{2}\nu_i)}}{B(\frac{1}{2},\frac{1}{2}\nu_i)}[\nu_i + \vartheta_i(X_t - \mu_i)^2]^{-\frac{1}{2}(\nu_i+1)}\sqrt{\vartheta_i} \quad (2)$$

$$P = \begin{bmatrix} p & 1-q \\ 1-p & q \end{bmatrix} \text{with } P \text{ the Markov transition matrix for } s_t$$

$$P_{ij} = Pr\,[\, s_{t+1} = i|s_t = j\,]$$

The model thus copes with both objections raised in the introduction. First the variance is allowed to take any value on R^+ and secondly the degrees of freedom of the respective Student-t distributions can restrict the number of finite moments to any positive integer value. These are the most important differences with the standard Markov switching model: the specification of the underlying densities. For the extreme case where the degrees of freedom approach infinity both models yield identical results.

To further the understanding of the model it is instructive to analyse the system in (2) in greater detail. As mentioned above we identify two regimes by their differences in the type of the news released. Anticipating the empirical outcome, these types are distinguished by the statistics of the inverted gamma distribution, $\lambda(s_t)$ and $\kappa(s_t)$. In other words, news types are defined by the different characteristics of their sizes (important versus unimportant news), $\lambda(s_t)$ and $\kappa(s_t)$, not their means (good versus bad news, as is the case for the studies for lower frequencies see Engel (1994) and Engel and Hamilton (1990)). The parameters λ and κ are sufficient statistics to specify the characteristics of the size of the news component. More specifically, they define all the moments for the precision variable $\sigma^{-2\ 2}$. The mean and variance are $\frac{\kappa}{\lambda}$ and $\frac{\kappa}{\lambda^2}$, respectively. From these formulas it is easily seen that the precision, and thus the variance of the return, can have very different properties depending on the values of both statistics. For example, if the mean precision is A, such that $\kappa=A\lambda$, we can have a whole range of variances associated with this mean. Since the variance of the precision variable is then given by $\frac{A}{\lambda}$ all variances on R^+ can in principle be obtained. The setup considered here thus adds a dimension to the standard model in terms of the characteristics of the news types. These types are distinguished on their average precision and its variance. The former determines the average news content and the latter determines the degree of heterogeneity in the news content within one regime.

A concise measure for the amount of heterogeneity in each regime is given by the 'signal to noise ratio', i.e. the ratio of the mean over the standard deviation of the precision. This ratio is given by $\sqrt{\kappa}$. The higher κ, the higher this signal to noise ratio and the less important the idiosyncratic movements (heterogeneity) in the precision process. This characterisation of the process of the precision variable is transformed into the parameters of the noncentral Student-t distribution through the evaluation of the integral in (1). As was noted above, the scale of the Student-t distribution equals $\vartheta=\frac{\kappa}{\lambda}$ and measures the average mean precision. The degrees of freedom are $\nu=2\kappa$ and are as such one to one with the above discussed heterogeneity measure.

The unobserved component s determines the distribution from which the return is drawn. This variable follows a first order Markov process in discrete time and discrete state space. The dimension of the state space defines the number of 'types' of news distinguished. In line with the current literature we restrict this dimension to at most two. The actual characterisation of the different types is an empirical matter and is obtained by estimating the coefficients of the two Student-t distributions. The probabilistic law that governs this state variable is summarised in the transition matrix P. Each entry in this matrix defines a transition probability as follows $P_{i,j}=Pr[s_t+1=i\,|\,s_t=J]$. For example, in equation (2), p is the probability that one stays in regime 1 given that we are in regime 1. The diagonal elements thus measure the inertia of each of the regimes. The interpretation of this Markov process is in terms of the way that news evolves over time. It gives some insight in how news types are released through time. For example news of the same type can be clustered, independent or anticlustered in time if the diagonal elements of P are larger than, equal to or lower than $\frac{1}{2}$, respectively.

The practical relevance of this extension to the standard Markov switching model is an empirical matter and is linked to the level of the degrees of freedom. If the degrees of freedom turn out very high say thirty or higher, the model does not add anything significant to the standard mixtures of normals. However, when the degrees of freedom are low, there can be large differences in inferences. In order to illustrate this point we analyse the lower exceedance probabilities according to each of these models. These probabilities can be regarded as the probability assessed by the investor of an investment[3] return lower than a certain benchmark α. This probability can be written in function of s_t (for $\kappa=1,2$) as

$$Pr[X_{t+1} < \alpha | s_t = \kappa] = Pr[s_{t+1} = 1 | s_t = \kappa] \int_{-\infty}^{\alpha} F(X | s_{t+1} = 1) d X +$$

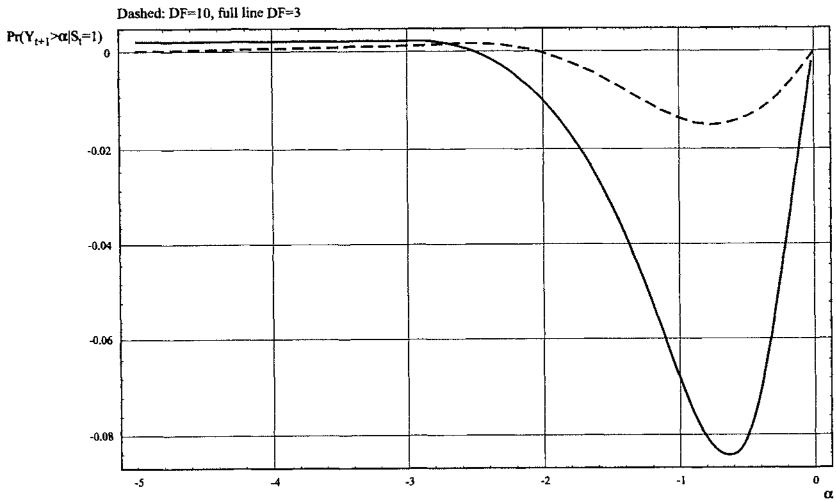
$$Pr[s_{t+1} = 2 | s_t = \kappa] \int_{-\infty}^{\alpha} F(X | s_{t+1} = 2) d X \qquad (3)$$

Figure 1 plots these probabilities for both the (standard) mixture of normals model and the extended one (with Student-t distributions). First, we concentrate on a simplified symmetric version of the model with $p=q=0.9$, $\nu_1=\nu_2$ and $\mu_1=\mu_2=0$ where the subscripts denote the value for the state variable s_t. The probabilities plotted in this figure represent the difference between the risk assessed using a mixture of normals model and the extended one. Both sets of parameters were chosen such that the variances within each regime are identical across models. This allows the interpretation of the probabilities in figure one as the error one makes when one would use the standard model if the DGP would be a mixture of Studentát distributions. The values are defined as the cumulative distribution function of the mixture of t's minus the cumulative distribution of the mixture of normals. It is very clear that the evaluation of risk differs drastically depending on the model used. Typically in this setting, the use of the mixture of normals leads to an underestimation of extreme events and an overestimation in the centre of the distribution.

These errors moreover tend to be larger the more the mixture of normals disagrees with the real underlying process, which is assumed to be the mixture of Student-t's.

FIGURE 1

Risk evaluation under different model specification
(relative to standard model)

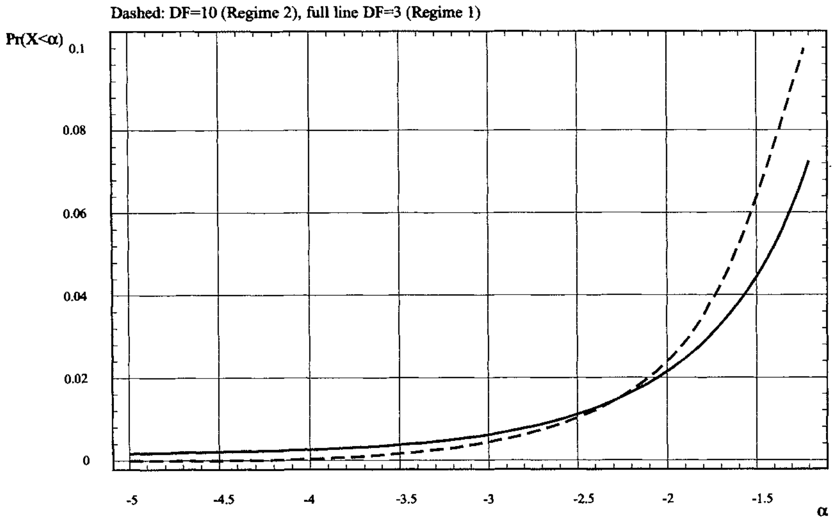


This can be seen by noting that the differences decline when one goes from the Student-t with three degrees of freedom to the one with ten. Eventually, the difference between the models converges to zero as the degrees of freedom approach infinity.

Other features of the mixtures of Student-t distributions are noteworthy. For example, in contradistinction with the mixture of normals or the GARCH models, this model allows for time variation in the higher moments which is not related to the variation in the scale. This is one of the most important assets of these models. This feature is illustrated in Figure 2.

FIGURE 2

Risk evaluation for T-Markov model with different degrees of freedom



In this figure, we plotted the cumulative distribution function for a T-Markov switching model where the only difference between the regimes is due to the degrees of freedom (regime one has three degrees of freedom, regime two has ten). More specifically, the variances do not differ across regimes. Since the variance does not change through time, an investor looking at the behaviour of the variance to assess the risk of his portfolio would find that risks do not change. This would also be the conclusion of applying standard Markov switching models or GARCH models. Yet, risks can differ significantly as is illustrated in Figure 2. If we are currently in regime one, risks of extreme events in the next period tend to be higher than if we were in regime two. This is so because if we are in regime one we have a.9 probability of drawing from the Student-t distribution with three degrees of freedom and only a.1 probability of drawing from the less leptokurtic distribution. In regime two the opposite occurs; a.9 probability of drawing from the less leptokurtic distribution and only.1 from drawing from the one with three degrees of freedom. These asymmetries are clearly reflected in the shape of the cumulative distribution function. Risk positions do differ even though variances are equalised across regimes. The T-Markov model can accommodate these higher order differences because it does not assume, unlike the mixture of normals model or GARCH, that the returns are drawn from distributions that only change in function of their scale. Instead, through the independent modelling of the degrees of freedom, return distributions can differ in other dimensions, defining the higher moments, as well. In the next section we proceed by estimating the this extended model.

III. EMPIRICAL RESULTS

In this section we estimate both the standard mixture of normals model and the extended model. We obtain the characterisations of the news types and of the way in which these news releases evolve over time. Next, both models are tested against each other which gives us a final description of the conditional densities for the return process.

Both models are estimated for three major free floating foreign exchange rates, i.e. the British Pound (BP), the French Franc (FF) and the Deutsche Mark (DM) all quoted against the US Dollar. The data series are weekly returns compiled by Datastream for the period 1973-1990. The return is defined as the first difference of the log prices. Different algorithms are available to estimate the model. We use the maximum likelihood method with numerical first and second order derivatives, see Hamilton (1989) and Kaminsky (1993)[4].

The estimation results are presented in Table 1. Both models are equivalent as far as the means are concerned. These means are, moreover, statistically insignificant from zero as one would expect on the basis of the quasi martingale behaviour of high frequency returns. News types are, therefore, not distinguished by their mean values but on the basis of the higher moments. In these moments the models differ. This is easy to see by concentrating on the estimated degrees of freedom in the extended model. For the extended T-Markov switching model, we find extremely low values for the degrees of freedom for both states. The standard model in contrast assumes infinite degrees of freedom.

TABLE 1

Estimates for standard and T-Markov switching model

		μ_1	μ_2	θ_1^{-1}	θ_2^{-1}	v_1	v_2	p	q	LR
DM	stand.	-.48	-.72	46.7	324.7	∞	∞	.896	.939	12.59
		(.55)	(.77)	(8.77)	(27.63)	(na)	(na)	(.028)	(.023)	
	T	-.97	-.8	36.2	237.0	5.15	9.42	.947	.976	
		(.57)	(.62)	(6.89)	(19.5)	(2.8)	(4.3)	(.028)	(.011)	
BP	stand.	.63	-.60	12.4	245.3	∞	∞	.846	.965	43.86
		(.37)	(.77)	(8.77)	(27.63)	(na)	(na)	(.037)	(.013)	
	T	.48	-.61	1.54	151.4	2.09	6.27	.900	.981	
		(.21)	(.48)	(.76)	(13.51)	(.63)	(1.34)	(.039)	(.004)	
FF	stand.	-.033	.389	69.3	425.1	∞	∞	.906	.880	57.07
		(.50)	(1.08)	(3.2)	(14.6)	(na)	(na)	(.02)	(.028)	
	T	-.55	.49	12.0	161.4	2.14	5.02	.966	.993	
		(.36)	(.48)	(3.25)	(16.2)	(.53)	(.91)	(.016)	(.004)	

Notes: Stand. refers to the estimates for the standard Markov switching model while T denotes the extended T Markov switching model. DM, FF and BP stand for Deutsche Mark, French Franc and British Pound, respectively. Returns are multiplied times 1000. Numbers within brackets are the standard errors. The entry LR is the likelihood ratio test of the null hypothesis of degrees of freedom equal to 240 in both states. The test is asymptotically chi-squared with two degrees of freedom.

Going from the mere description of differences in the estimates to rigorous statistical testing is not straightforward in this framework. The main reason for this is that the two models are not nested. True, the extended model has the mixtures of normals as a special case, i.e. for $v_i=\infty$. However, statistical tests such as the Likelihood ratio tests require a compact parameter space, excluding infinity as a valid alternative. Statistically speaking, the two models are not nested. If one pursues such tests one will have deformations of the asymptotic distribution of the test statistic. Fortunately these deformations are such that the test statistics become overconservative, i.e. they do not reject the null hypothesis often enough. For a lucid discussion on this issue see Bollerslev (1987). Therefore, if one proceeds by testing the null of normality using standard procedures and one reject the null hypothesis, then, a fortiori, one would reject the null hypothesis if the appropriate, but difficult, amendments were made on the asymptotic test distribution. We follow a slightly different approximation based on an equivalent argument. In this paper we test two extended Markov switching models against each other, the unrestricted one and one with a high but finite degrees of freedom (say 240 for each distribution).

In this framework the standard tests are valid. Given we reject the null hypothesis of 240 degrees of freedom we conclude that the normal model must be rejected as well. The argument being that the extended model with 240 degrees of freedom and the mixture of normals models are for all practical purposes equivalent[5]. The null hypothesis to contrast both models is thus given by $H_0 : \nu_i=240, i=1,2$. The likelihood ratio test is chi-squared with two degrees of freedom under the null hypothesis. Results for this test can be found under the entry LR in Table 1. As can be seen we reject the mixture of normals for each of the series at the weekly frequency. This establishes the superiority of the extended T-Markov switching model over the standard mixtures of normals model.

The retained T-Markov switching model agrees with three stylised facts of high frequency returns. First, the behaviour is martingale like, we could not reject means equal to zero. Secondly, the high inertia in the Markov process, i.e. high diagonal elements in the transition matrix, suggests a clustering of the news types through time. Since the types are characterised by (very,) very different variances we obtain volatility clustering through the news 'type'clustering. Finally, the number of finite conditional and unconditional moments is very low. This agrees with the literature on the extremal behaviour of exchange rate returns. More specifically, for most exchange rates we found only the first two moments to be finite. This, in conjunction with the high inert states, agrees with the GARCH versus IGARCH discussion in the more traditional literature.

The most prominent difference between the standard and the extended model is the way one models the behaviour of the variance within the same regime. According to the standard model this is a constant. In contrast, the extended T-Markov switching model allows for idiosyncratic noise in the variance. This amendment is essential given the relatively low value of the signal to noise ratio for the precision (i.e. the reciprocal of the variance) in each state. These values suggest considerable (and i.i.d.) deviations from the mean precision in each regime. These mean precisions differ drastically across regimes. Moreover, we found that the levels of heterogeneity as measured by the signal to noise ratio differ (LR statistics for the British Pound and the French Franc reject the null hypothesis of equality of degrees of freedom at 10%).

We can thus characterise the return (news) process as follows. News type releases evolve over time in a clustered way. Important news tends to be followed by important news until a shift occurs to less dramatic news period which remains then for quite a while until a juicy news period arrives again ... The variance of each regime is however not a full characterisation. Volatility levels, read news contents, differ drastically within each state. Typically for each exchange rate we found that the regime with the highest variance (news contents) also displays the highest degrees of freedom and thus the lowest heterogeneity. Therefore, it can happen in periods where important news is released that less important information arrives but on average the news content will be high. But the reverse, a one week important news release amidst a period of irrelevant news is more likely to occur as the heterogeneity measure in the low variance period is much higher. This contrast with the traditional approach were it was assumed that the news content, as measured by the variance, did not change within each regime.

Finally, the model has pointed to an important extra dimension for risk evaluation. For most currencies we found that regimes differ both in their scales and their degrees of freedom. The direct link between higher variance and higher risk is therefore no longer valid as was illustrated in Figures 1 and 2. An evaluation of the higher moments is needed as well.

IV. CONCLUSION

In this paper we extended the standard mixture model of Hamilton (1989) to incorporate both a continuum of variance states as well as a restricted number of finite moments. Both facts are not possible in the standard approach and this gives rise to some conflicts with established facts as far as high frequency returns are concerned. The way these facts are incorporated is by changing the standard normal densities to noncentral Student-t distributions. Via some well known results on compounded mixtures it is shown that this amendment is sufficient to accommodate both stylised facts.

Next we estimated both the extended and the standard model and tested both models against each other. We found, not surprisingly, that the extended model yielded extra information on the process driving weekly returns of major currencies. An interpretation of the different parameters of the mixture of t distribution in terms of news types was used. In this framework we conclude that the news types are characterised by their scale and degrees of freedom. These variables suggest that news is distinguished more on its impact (i.e. the relevant content of the news as measured by its variance) than on the direction of the news (i.e. good or bad news). Moreover, the low degrees of freedom suggest that, also within one regime heterogeneity in variances (news contents) are important. Moreover, we found statistically significant differences in degrees of freedom. These differences add an extra dimension to the evaluation of the risks of certain investment schedules. Variances no longer suffice, since densities differ across regime in other dimensions as well.

A part from a more detailed characterisation of the types of news, the paper contributes to the current debate on the appropriate specification of conditional distributions of returns. This specification-debate is currently reopened due to the increasing popularity of assets with high nonlinear payoff profiles. In order to assess some of the moments of investments in these assets one relies more and more on the Markov switching models. These models have the advantage that the distributions remain tractable. Of course the quality of the moment predictions rests on the quality of the description of the distribution. In this paper it is shown that the standard model, currently used in these types of exercises, is certainly not the optimal description. A better one is the mixture of Student-t distributions.

An illustration of the model was provided in terms of the evaluation of risks associated with a simple one period investment. Other examples, such as option pricing might be considered as well. for option pricing, results are likely to be even more pronounced as the inherent nonlinear payoff structure makes differences in higher moments even more important. However, this is left for future research. other potential applications of this model are numerous. To name only a few, one might evaluate, using this extended model, to evaluate the asymmetric behaviour of shocks over the business cycle or one might analyse stock market crashes in terms of the degrees of freedom of the underlying distributions.

NOTES

1. We use the terms news and returns interchangingly throughout the remainder of the paper. Such an equivalence will only be appropriate if the return process is a martingale. Although there may be some structure in the first moments, as is recently argued in the literature, this martingale assumption is not such a bad approximation. More specifically in the framework of the Markov switching models we can not reject this assumption.

2. Note that we have to do the interpretation in terms of the precision variable, the inverse of the variance. This is due to the Jensen inequality. If we were to do the interpretation in terms of the variance we would neglect the nonlinearity in the reciprocal transformation.
3. The investment schedule is very primitive. We analyse a one period investment in the underlying asset.
4. Other algorithms are also available: the EM algorithm developed by Hamilton (1991) or a method of Moment approach. The latter one is however not yet fully reliable as shown in the literature. The former seems to be the more robust one. However, the first and second order derivatives are not easy to obtain due to the nonlinear filtering problem and the use of noncentral Student-t distributions. We rely on the first algorithm presented in the literature. We tested the reliability of this method against the more robust EM approach and found that the former performs well. We were able to reproduce established results (which were based on the EM method) with the method used here.
5. If one does not feel comfortable with this testing procedure, one can interpret the test results as obtained from the other method discussed. It turns out that the loglikelihood of the extended model with 240 degrees of freedom each is almost exactly the likelihood of the mixtures of normals models. Therefore, test results will be identical for both approaches for these rates. In order to verify the overconservative character of the former approach a small monte carlo experiment was conducted. It was found that this feature was present for the relevant parameter values (i.e. values close to the estimated ones).

REFERENCES

Bollerslev T., 1986, Generalised Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31, 302-27.

Bollerslev T., 1987, A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Returns, *The Review of Economics and Statistics,* 69, 542-47.

Engel C., 1994, Can Markov Models Forecast Exchange Rates, *Journal of International Economics* 36, 151-66.

Engel C. and J.D. Hamilton, 1990, Long Swings in the Dollar: Are They in the Data and Do the Markets Know It?, *American Economic Review,* 80, 689, 713.

Engle R., 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, 50, 987-1007.

Hamilton J.D., 1989, A New Approach to the Economic Analysis of Nonstationary Time, Series and the Business Cycle, *Econometrica,* 57, 357-84.

Jansen, D.W. and C.G. de Vries, 1991, On the Frequency of Large Returns: Putting Booms and Busts into Perspective, *The Review of Economics and Statistics,* 1, 18-24.

Kaelher J. and V. Marnet, 1993, Markov Switching for Exchange Rate Dynamics and the Pricing of foreign Currency Options, (mimeo ZEW and University of Mannheim).

Kaminsky G., 1993, Is There a Peso Problem: Evidence from the Dollar/Pound Exchange Rate 1976, 1987, *American Economic Review,* 83, 450-72.

Koedijk K., M.M.A. Shafgans and C.G. de Vries, 1990, The Tail Index of Exchange Rate Returns, *Journal of International Economics,* 93-108.

Mandelbrot B., 1963, The Variation of Certian Speculative Prices, *Journal of Business,* 36, 394-419.