Tjalling C. Koopmans Research Institute

Discussion Paper Series nr: 09-08

Estimating and Forecasting Asset Volatility and Its Volatility: A Markov-Switching Range Model

Jan Piplack

Universiteit Utrecht

Utrecht School of Economics

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Janskerkhof 12
3512 BL Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics.
It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to J.M.vanDort@uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Jan Piplack
Utrecht University
Utrecht School of Economics
Janskerkhof 12
3512 BL Utrecht
The Netherlands.
E-mail: J.Piplack@uu.nl
RABO bank
Croeselaan 18
3521 CB Utrecht
The Netherlands.

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 09-08

Estimating and Forecasting Asset Volatility and Its Volatility: A Markov-Switching Range Model

Jan Piplack

Utrecht School of Economics
Utrecht University

RABO bank
Utrecht, the Netherlands

May 2009

Abstract

This paper proposes a new model for modeling and forecasting the volatility of asset markets. We suggest to use the log range defined as the natural logarithm of the difference of the maximum and the minimum price observed for an asset within a certain period of time, i.e. one trading week. There is clear evidence for a regime-switching behavior of the volatility of the S&P500 stock market index in the period from 1962 until 2007. A Markov-switching model is found to fit the data significantly better than a linear model, clearly distinguishing periods of high and low volatility. A forecasting exercise leads to promising results by showing that some specifications of the model are able to clearly decrease forecasting errors with respect to the linear model in an absolute and mean square sense.

Keywords: Volatility, range, Markov-switching, GARCH, forecasting.

JEL classification: C53, G15, G17

Acknowledgements

The author thanks Bertrand Candelon, Yin-Wong Cheung, Ai-Ru Cheng, Alain Hecq, Franz Palm, Jean-Pierre Urbain and the participants of the workshop of the Methods in International Finance Network 2007 in Maastricht for their very useful comments and suggestions. Further, I would like to thank the University of California in Santa Cruz, where this project commenced, for their hospitality. I also want to thank the Maastricht Research School of Economics of Technology and Organizations (METEOR) for their financial support. All errors and omissions are my own.

1 Introduction

Asset-volatility is of outstanding importance in finance. It directly and indirectly influences asset pricing (for example options prices directly depend on the underlying asset's volatility), the optimal hedge ratio, portfolio decompositions, and risk management among others (Alizadeh, Brandt and Diebold, 2002). Therefore, volatility-modeling has been a focus of much academic research in the last decennia. Early contributions assumed constant asset-volatility (e.g., Merton (1969) or Black and Scholes (1973)). Especially the work of Engle (1982) and Bollerslev (1986), however, contributed to the nowadays widely accepted conviction that volatility is both time-varying and predictable (see also Andersen and Bollerslev (1997)). Engle introduced the autoregressive conditional heteroscedasticity (ARCH) models whereas Bollerslev extended those to the class of generalized ARCH (GARCH) models. The observation that some time periods are affected by very high while others by relatively low volatility fostered the more recent development of regime-switching models. Building on Hamiltons (1989) work, Hamilton and Susmel (1994), Gray (1996) and Klaassen (2002) further introduced regime-switching ARCH and GARCH models improving further the modeling of volatility. ARCH and GARCH models are today the workhorse of asset-volatility modeling both in academics and industry (see for example Ghysels et al. (2006)).

Volatility in economics is defined as the variability of a random variable of a time series. Hence, this volatility is "...inherently unobservable, or latent, and develops stochastically through time" (Brandt and Diebold, 2006, p. 1). Volatility is inherently latent because the true data generating process of asset prices is not known, making it impossible to quantify unambiguously the "random component" of a time series and even more difficult to pin down its instantaneous variability or volatility. There appear to be two solutions to this problem. First, one can try to model the latent variable volatility as the conditional second moment/variance of an observed return series parametrically (e.g., Engle (1982), Taylor (1982), and Bollerslev (1986)). Or second, one uses nonparametric estimators for the volatility. The range, defined as the difference between the maximum and the minimum log asset prices over a fixed interval, appears here as a natural estimator and has indeed been the subject of much academic research (e.g., Garman and Klass (1980), Parkinson (1980), Beckers (1983), Ball and Torous (1984), Rogers and Satchell (1991), Andersen and Bollerslev (1998), Yang and Zhang (2000), Alizadeh et al. (2002), Chou (2005), Brandt and Diebold (2006)). Another interesting and related approach is introduced in Cheung (2007) where the daily highs and lows of the price process are modeled by means of cointegration analysis.

In contrast to the conditional variance modeling, the range is directly observable from the data and does not need to be estimated. Apart from being a very intuitive and directly observable volatility estimator, the range also is very efficient. Indeed, as noted in Brandt and Diebold (2006, pp.61):

> As emphasized most recently by Alizadeh et al. (2002), the range is a highly efficient volatility proxy, distilling volatility information from the entire intraday price path, in contrast to volatility proxies based on the daily return, such as the daily squared return, which use only the opening and closing prices.

Moreover, as has also been mentioned by many authors (e.g., Alizadeh et al. (2002)), range data on the one hand are available for many different assets such as individual stocks, stock indices, currencies, and Treasury securities, and on the other hand these data series often have a history of many decades. This constitutes a strong advantage over another nonparametric estimator of the variance, namely the realized volatility, which uses high-frequency data at say 5-minute intervals. Those data often only start in the middle of the 1990s if available at all. Furthermore, Alizadeh et al. (2002) clearly show that market microstructure problems like bid-ask spreads can severely bias the realized volatility estimator. Primarily theoretical references regarding realized volatility include, among others, Barndorff-Nielsen and Shepard (2001; 2003; 2004), and Andersen, Bollerslev, Diebold and Labys (2003).

A further advantage of using an observed volatility estimator is that it can be modeled in the mean equation. This enables the econometrician to model the volatility of the volatility as the conditional second moment of the range in contrast to having to model it as the conditional fourth moment of a return series. Modeling the volatility of the volatility is important, for example, in option pricing, where an option trader wants to know the probability that the volatility, a direct price determinant, changes in order to optimize his pricing decision. Additionally, changes in volatility also have an influence on optimal hedge ratios (e.g., Ederington (1979), Lien and Tse (2002)). Therefore, predictable volatility of volatility can help in making better hedging decisions.

In the literature asset markets have been found to show regime-switching behavior. There are clear periods of low/normal volatility but also longer-lasting periods where asset market volatility is significantly higher than in the low-volatility periods. Such regime-switching volatility behavior has usually been modeled with first-order Markov processes. See, for example, Hamilton and Susmel (1994), Gray (1996), and Klaassen (2002).

Motivated by these points, we propose a simple yet efficient way of modeling asset market volatility and its volatility. We suggest to fit the log range of assets to a Markov-switching-(MS-)ARMA-GARCH time series model. This combines the advantages of the range as a nonparametric yet highly efficient volatility estimator with well established time-series modeling techniques in order to estimate and forecast asset volatilities. We fit our proposed model to weekly S&P500, ten year T-notes, three months T-bills, FTSE100, and Nickei225 data. Our findings are: First, our model is well able to distinguish low from high volatility regimes. Second, that volatility dynamics change with the regime, which has important effects for forecasting purposes, confirming results found in Gray (1996). Third, a forecasting exercise leads to promising results by showing that some specifications of the model are able to clearly decrease forecasting errors with respect to a linear model.

This paper proceeds as follows. After presenting the theoretical background about the range in Section 2 we describe the methodology for model estimations in Section 3. Section 4 presents the results of the application of our model to the data of different assets and markets. We also present results of a forecasting comparison. Finally, Section 5 summarizes our results, concludes and sketches directions for future research.

2 Theoretical foundations

In this section we will present the theoretical distributional foundation of the range as an estimator for the diffusion constant of a continuous random walk that has already been derived in the literature. Next we heuristically argue that those distributional results should basically remain unchallenged once we allow for non i.i.d. errors. We also conduct a short Monte Carlo experiment to show the superiority of a simple range model compared to a GARCH specification.

2.1 Theory of the range as a volatility estimator

Taking the range as an estimate for the diffusion constant of stochastic processes like the continuous random walk has a long history going back to Feller's (1951) seminal paper where he derived the asymptotic distribution of the range for the sum of independent variables using the theory of Brownian motion. The assumptions Feller used are as follows. Let $[u_t]$ for $t = 1, ..., n$ be a sequence of individually identically distributed (i.i.d.) random variables with distribution $F(u)$ with $E(u_t) = 0$ and $Var(u_t) = 1$. Let then $S_n = u_0 + \ldots u_n$, $M_n = max(0, S_1, \ldots S_n)$, and $m_n = min(0, S_1, \ldots S_n)$. The range is then defined as $R_n = M_n - m_n$, which corresponds closely to our definition in Section 3. Feller than derives formulas for the mean and the variance of the range R_n. For details we refer to Feller (1951).

Parkinson (1980) extended the work of Feller (1951) to the case where $Var(u_t) = D$ and D being the random walk diffusion constant. Additionally he applies the framework to the stock market and shows that the range is a far superior estimate for the diffusion constant than the traditional estimates using closing prices only. Parkinson further derives a function describing all the moments of the range distribution, which yields:

$$E(R^p) = \frac{4}{\sqrt{\pi}}\Gamma(\frac{p+1}{2})(1-\frac{4}{2^p})\zeta(p-1)(2Dn)^{p/2,}$$

where Γ is the gamma function, and $\zeta(x)$ is the Riemann zeta function. p has to be real and ≥ 1. The first two moments are found to be:

$$E(R_n) = \sqrt{\frac{8Dn}{\pi}}$$
$$E(R_n^2) = [4ln(2)]Dn. \tag{1}$$

Another contribution showing the superiority of the range estimator over the standard squared returns is the work of Martens and van Dijk (2007). They show with simulations that the so-called realized range has a lower mean-squared error than the realized variance, where the realized range is defined as the sum of all observed equidistant intra-day ranges adjusted by some factor to account for microstructure issues. For details we refer to the paper.

2.2 Extension to non i.i.d. case

One can argue that the assumption of i.i.d. increments (u_t) is not a realistic one because there is clear evidence that the volatility of asset returns is changing over time. So, it would be nice to extend the results of Feller (1951) and Parkinson (1980) to the non i.i.d. case. Such an exercise is not trivial because the proofs used in mentioned sources rely on the i.i.d. assumption, since they proceed by reordering the observations.[1] In the non i.i.d. case such an arbitrary reordering is for obvious reasons not possible.[2] We could not find a way to solve this problem, which makes it impossible for us to come up with moments like (1) for the non i.i.d. case. We can only provide the reader with some intuition and heuristic reasoning to argue that the results found by Feller (1951) and Parkinson (1980) should basically remain valid (probably with some scaling parameter accounting for the dependency). Thereby, the following is by no means a formal proof but can help to get some intuition

[1]For details we refer the reader to Feller (1951).

[2]Therefore, the proofs cannot be used to extend the results to the more general case of dependency.

about how such a proof might proceed. An exact proof we leave for further research.

At this point we state some standard results as published in, for example, Davidson (1994) and Davidson (2000). Davidson derived the following theorem:

Let S_n be defined as above and u_t be a martingale difference sequence with $E(u_t) = 0$ and $E(u_t^2) = \sigma_t^2 < \infty$ with $\overline{\sigma}_n^2 = n^{-1} \sum_{t=1}^n \sigma_t^2$. If u_t meets the additional conditions that

a) $n^{-1} \sum_{t=1}^n (u_t^2 - \sigma_t^2) \xrightarrow{pr} 0$, and

b) either

  i) the sequence is strictly stationary or

  ii) $\frac{max_{1 \leq t \leq n} ||u_t||_{2+\delta}}{\overline{\sigma}_n} \leq C < \infty \quad \delta > 0, \forall n \geq 1$

then $\nu_n = \frac{\sqrt{n}\overline{u}_n}{\overline{\sigma}_n} \xrightarrow{d} \nu \sim N(0, 1)$.

Let us also assume that

$$\frac{E(S_n^2)}{n} \rightarrow \sigma^2 < \infty \quad \text{(global wide-sense stationarity)}. \tag{2}$$

If we then define $X_n(r) = \frac{S[nr]}{\sqrt{n}\sigma}$, then $X_n \xrightarrow{d} B$. $\triangleright$

Here $\xrightarrow{pr}$ and $\xrightarrow{d}$ stand for convergence in probability and distribution, respectively. B stands for Brownian motion and r in between 0 and 1. A proof of this theorem can be found in Davidson (1994), Theorems 27.14 and 29.6.

This theorem then states that under the condition of not too strong dependence in the sequence of σ_t the correctly weighted partial sum S_n still converges to Brownian motion. Such a convergence is the basis of the proofs in Feller (1951) and Parkinson (1980) for the distribution of the range estimate for the diffusion constant. Heuristicly speaking then, this theorem provides the basis to reason that the limit distribution of a correctly scaled range as in Equation (1) will stay the same with $D = \overline{\sigma}_n^2$ even for the non i.i.d. case of the sequence u_t. Such a scaling is likely to depend on the structure of the dependency.

Another intuitive way to justify that one can use the results of Feller (1951) and Parkinson (1980) is provided in Alizadeh et al. (2002). They propose to simplify the underlying continuous time process by mapping it into a discrete time process. In practice the econometrician does not observe a continuous but a discrete time price process. One can divide the sample period [0 T] into n intervals with equal length $L = T/n$ corresponding to say one day or one week. Then we replace the continuous volatility dynamics with a piecewise-constant process, where we assume that the volatility within every interval i is constant, i.e. $\sigma_t = \sigma_{iL}$. From one interval to the next we allow the volatility to change. This implies that within each interval i the price process follows a geometric Brownian motion such that Feller's (1951) and Parkinson's (1980) results apply again for that interval.

2.3 Monte Carlo experiment

As already mentioned above, Parkinson (1980) and Alizadeh et al. (2002) find that the range and log range are more efficient estimators for the diffusion constant or underlying instantaneous volatility D. These results are well-known for the i.i.d. case. We now perform a small-scale Monte-Carlo simulation in order to show that this also holds for the non i.i.d. case. We assume a piecewise-constant volatility process but allow the volatility to change from one interval to another. The exact process we assume here is a standard GARCH(1,1) with a constant mean equation. Coefficients are based upon US stock market data and the model looks as follows:

$$\begin{aligned} r_t &= 0.0018 + \epsilon \quad \text{where} \\ \epsilon &= \sigma_t u_t \quad \text{with } u_t \sim N(0,1). \end{aligned} \tag{3}$$

We assume the variance to follow a standard GARCH(1,1) model, which has become the workhorse in the volatility literature[3]:

$$\sigma_t^2 = 0.0000159 + 0.119\epsilon^2 + 0.851\sigma_{t-1}^2 \tag{4}$$

[3]See, for example, Ghysels et al. (2006).

With this process we simulate $T = 2500$ observations of r_t and σ_t^2, roughly corresponding to the sample length of our US stock market data. In order to be able to also simulate T range observations, we assume that during every interval the volatility is constant. We then simulate $T_{t,int}$ intermediate returns $r_{t,int}$ with an underlying volatility $\sqrt(\sigma_t^2/T_{int})$ giving us in total $T \times T_{int}$ returns and T variance observations. With these intra-period observations we can construct a quasi-continuous price process and obtain our T range estimates. In order to justify the usage (1), which is based on the assumption of Brownian motion, we take $T_{int} = 1000$. Larger values for T_{int} would only make the results more accurate but would not add much to the point we want to make.

In order to show the efficiency gains of the range estimator and our model we propose to use later in this paper, we compare the fitted values of two different models. After one simulation of the outlined process we have T observations of intra-period returns and T observations of ranges. Now we would like to compare the standard GARCH model to a very basic model based upon the range.[4] To this end we estimate a standard GARCH(1,1) model like in (3) and (4), which also serves as the data generating process (DGP). This model is the benchmark model. In a second step we use the simulated range observations and take the log of it to arrive at the log-range. Such a monotonic transformation is done to make estimation of the model easier and will be reversed after estimation in order to arrive again at the range. We then fit a standard ARMA(1,1) model on the log-range, which has the same amount of parameters as the GARCH(1,1) model. Taking the fitted values of this model we can calculate the expected values for the underlying diffusion constant of the DGP by using (1).

We now compare the fitted values of the GARCH(1,1) based on the noisy measure of squared returns to the fitted values of the range-based ARMA(1,1) model, by calculating the mean square errors (MSE). We just calculate for every t the difference between the model-induced estimates and the true simulated values, square them and take the average. In our exercise we do

[4] Later in this paper we further develop such a model extending it to be Markov-switching. At this point we continue to assume that it is not Markov-switching but linear.

this 1000 times in order to get a distribution of MSEs for the GARCH and the for the ARMA-range model. Those two can now be compared. Results show that the MSE of the range-based model is on average 27.63% lower than that of the GARCH model. The minimum and maximum of MSE improvements were 10.79% and 45.84%, respectively.

Already this small-scale Monte-Carlo experiment shows the superiority of range-based estimates of volatility compared to squared return estimates using a standard GARCH model even though the data have been simulated by such a GARCH model. This improvement is mainly due to the fact that the range is the more efficient volatility proxy compared to the noisy proxy of squared returns. These results confirm, for example, those obtained by Parkinson (1980) and Alizadeh et al. (2002) but extend them towards the non i.i.d. case. We, therefore, propose to use the range as a basis for volatility models instead of squared inter-interval returns.

3 Methodology

Taking the results from Section 2 about the range as a volatility estimator into account, we now outline the general methodology proposed in this paper. We will introduce the estimation and forecasting technique that we apply for our Markov-switching (MS) Range Model and its different specifications.

Markov-switching time series models in econometrics today draw heavily on Hamilton (1989) and Hamilton (1990) where he develops the idea that output and business cycles in an economy may be subject to discrete changes in regimes underlying their DGPs. Hamilton argues that during economic expansions the average GDP growth rate should be different compared to times of recessions and that such a behavior might best be described by a Markov chain that governs switches from regime 1, expansion say, to regime 2, recession, and vice versa. In his paper he proposes to model the GDP growth rate as a Markov-switching autoregressive process of order q (MS-AR(q)).

3.1 The path to the model

Hamilton and Susmel (1994) and Cai (1994) argue that in financial time series often observed volatility clusters or volatility persistence can be modeled in a similar fashion as in Hamilton (1989). In their paper, they develop a MS-ARCH model. ARCH models go back to the pioneering work of Engle (1982) which Bollerslev (1986) extended to generalized-ARCH (GARCH) models. Those models are designed to model the conditional second moment or variance of time series and usually fit, for example, stock market returns very well.[5] Hamilton and Susmel find that ARCH models often impute much persistence to stock volatility but fail to give good forecasts. They pose that this might be due to large shocks that arise from different "regimes" rather than normal shocks. One finding is that the parameters of an ARCH process seem to come from different regimes where transitions between regimes are governed by an unobserved Markov chain.

An important advantage of GARCH models over ARCH models is that they usually capture much better the time dependence in the volatility. In order to be more precise we introduce a very general $GARCH(p,q)$ model. We refrain for the moment from specifying a mean equation but will do so in a later section. The GARCH model can be written as:

$$\begin{aligned} u_t &= \sqrt{h(\theta_h, \Phi_{t-1})} v_t \\ &= \sqrt{h_t} v_t, \end{aligned} \tag{5}$$

with the conditional variance of u_t specified as a function like:

$$\begin{aligned} Var(u_t) = h_t &= f(u_{t-1}, u_{t-2,\dots}) \\ &= \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \end{aligned} \tag{6}$$

where θ_h is a vector of parameters governing the variance equation and v_t is an i.i.d. sequence with zero mean and unit variance. Φ_{t-1} is the information set generated by u_t and represents the available information up to time

[5] See, for example, Sabbatini and Linton (1998).

$t-1$. Other assumptions for the error distributions are generally possible. A $GARCH(p,0)$ model is equal to an $ARCH(p)$. So, the GARCH representation allows for a richer parametrization of the conditional variance and facilitates modeling the observed volatility persistence.

Both Cai (1994) and Hamilton and Susmel (1994) claim that the extension of GARCH processes to the Markov-switching (MS) framework is intractable because of its path-dependence. Path dependence here means that the distribution at time t, if made conditional on regime (S_t) and on the available information set Φ_{t-1}, depends directly on S_t but also indirectly on the whole history of regimes $(S_{t-1}, S_{t-2}..., S_0)$. Regime-dependence in MS-GARCH models arises through the lagged variance and lagged squared error terms. In such models, conditional variance at time t depends on the squared error and the conditional variance at time $t-1$, which obviously depends on regime S_{t-1} and the squared errors and conditional variance at time $t-2$ and so forth. This introduces an infinite path dependence on the unobserved regimes $S_t, S_{t-1}, ..., S_0$ or $\widetilde{S}_t$. In (quasi) maximum likelihood estimations (QMLE) the likelihood function could only be constructed by integrating out all possible regime paths. If we denote K as the number of regimes and T the full sample time dimension, then there are K^T possible regime-path realizations, which would make estimation impossible as the time dimension increases.

In order to avoid this path-dependence problem present in GARCH models, Gray (1996) and Klaassen (2002) introduce ways to integrate out the path-dependence inherent in GARCH models avoiding the integration over K^T possible likelihoods. Gray's idea is to integrate out the unobserved regime-path $\widetilde{S}_t$ where it emerges namely in Eq.(6) itself. To see this we now have to write Eq.(6) in a regime-dependent form:

$$\begin{aligned} Var(u_t|\widetilde{S}_t, \Phi_{t-1}) &= h_{k,t} \\ &= f(u_{t-1}, u_{t-2,...}; \widetilde{S}_t) \\ &= \omega_k + \sum_{i=1}^{p} \alpha_{k,i} u_{t-i}^2 + \sum_{j=1}^{q} \beta_{k,j} h_{k,t-j}, \end{aligned} \tag{7}$$

where $Var(u_t|\widetilde{S}_t;\Phi_{t-1})$ denotes the variance of u_t conditional on observable information Φ_{t-1} and the unobservable full regime path $\widetilde{S}_t$. The parameters in the variance equation at time t are only determined by the current regime S_t. In Eq.(7) there is still the full regime-path-dependence present and it is not possible to estimate its parameters.

Different ways of integrating out the path-dependence have been suggested in the literature. Hamilton and Susmel (1994) circumvent the problem of path-dependence by excluding the lagged conditional variance terms $h_{k,t-1},...,h_{k,t-q}$ in the variance equation. Hereby they only need to integrate K^p different pathes out of the likelihood function in order to estimate the parameters. Gray (1996) uses a different idea. As already mentioned above he integrates out the path dependence in the GARCH by taking expectations of the conditional variances over all possible regimes. Hereby, he makes the conditional variance at time t only dependent on the current regime S_t but not the full path $\widetilde{S}_t$. In equation form this might be written like:

$$\begin{aligned} Var(u_t|S_t,\Phi_{t-1}) = \quad & \omega_k + \sum_{i=1}^{p} \alpha_{k,i} E_{t-2} u_{t-i}^2 \\ & + \sum_{j=1}^{q} \beta_{k,j} E_{t-j-1} Var(u_{t-j}|S_{t-j},\Phi_{t-j-1}), \end{aligned} \tag{8}$$

where E_{t-j-1} means that expectations are taken at time $t-j-1$ over all possible regimes and conditional on the information set Φ_{t-j-1}. This basically means that every period ex-ante probabilities are calculated (we will show the whole estimation algorithm later in this section) which are then used to weigh all possible values of u_{t-i}^2 and $Var(u_{t-j}|S_{t-j})$. In the next period those weighted values are used as inputs for the variance equation. So, essentially the regimes S_{t-j} are integrated out at time $t-j-1$.

Klaassen (2002) improves on Gray's (1996) method by proposing to wait with integrating out the past regimes until they are really needed and that is at time $t-1$. Hereby more observations can be used in order to draw inferences about the probabilities of regimes at different points in time. If for example it is very likely that the observation at time t comes from regime k

and regimes are very persistent, then this adds information to the calculation of the state probabilities in periods before. In other words Klaassen proposes to use the fact that the regime at time t essentially is in the conditioning information of $Var(u_t|S_t, \Phi_{t-1})$ particularly if regimes are highly persistent. He, therefore, suggests to rather use the following representation:

$$\begin{aligned} Var(u_t|S_t, \Phi_{t-1}) = \quad & \omega_k + \sum_{i=1}^{p} \alpha_{k,i} E_{t-1}[(u_{t-i}|S_{t-i}, \Phi_{t-i})|S_t]^2 \\ & + \sum_{j=1}^{q} \beta_{k,j} E_{t-1}[Var(u_{t-j}|S_{t-j}, \Phi_{t-j})|S_t], \end{aligned} \tag{9}$$

where the expectation E_{t-1} again is across regimes $\widetilde{S}_{t-1}$ but now conditional on the information set Φ_{t-1} and the regime S_t. Constructed like this, $Var(u_t|S_t, \Phi_{t-1})$ again only depends on the current regime S_t and not on the full regime path $\widetilde{S}_{t-1}$ and the path-dependence problem disappears.

Chou (2005) proposes another way to model the volatility of asset markets by using the range of the price process as an observable estimator for volatility. We already showed in Section 2 that the range is a very efficient volatility estimator. Chou (2005) then suggests to model the mean of the range in the following way:

$$\begin{aligned} R_t &= \lambda_t \epsilon_t \\ \lambda_t &= \omega + \alpha R_{t-1} + \beta \lambda_{t-1}, \end{aligned}$$

where $\epsilon_t \sim F(1, .)$. Here λ_t can be interpreted as the expectation of the range at time t and is modeled in an autoregressive fashion very much like a GARCH model. As can be easily seen, this model is from the multiplicative class of models and asks for an error distribution with a non-negative support in order to guarantee positivity of the range. Chou shows that this model fits the S&P500 range data quite well. Another approach is due to Alizadeh et al. (2002) who use the log-range in a stochastic volatility model.

In this paper we present a new way to combine the range volatility estimator introduced by Feller (1951) and Parkinson (1980) with the approaches of Klaassen (2002) and Chou (2005). We extend Klaassens (2002) approach to a more general MS-ARMA(a,b)-GARCH(p,q) case in order to model the log-range. As our main focus is modeling asset volatility with the help of the range estimator, we have to focus also on the mean equation and not only on the variance equation. As already mentioned above, the advantage of "observing" the volatility makes it possible to use standard time series methods, as in Chou (2005), to model it. This approach has two important benefits. First, we can essentially model the observed volatility. In the ARCH and GARCH literature, the volatility is not observed but rather derived as the conditional second moment from a series of asset returns. Second, this approach allows us to model the volatility of the volatility as a conditional second moment of the range. We do not need to estimate a conditional fourth moment as would be the case if we used return data. So, we can also model the dynamics and persistence of the volatility of the volatility of assets relatively easily.

3.2 The model

In this subsection we present the model we would like to fit to the data in its most general form. In Section 4 we fit different versions of such a model to the data. Let p_t denote the logarithm of the price of some speculative asset at time t. Then the range of that asset over a certain period, say a week, can be defined as $R_t = 100 * (p_t^{Max} - p_t^{Min})$. Here p_t^{Max} and p_t^{Min} denote the highest and the lowest observed price of that asset over the considered time period, respectively. In other words, the range measures the maximum spread in percent of an asset's price over a specified period. Let r_t denote the logarithm of R_t. We, thereby, use the same definition of the range and its logarithm as in Alizadeh et al. (2002). Contrary to Chou (2005) we model the log-range instead of the range in order to allow also for negative observations. This basically changes a multiplicative into an additive model and facilitates estimation.

In the following we will use the terms range and log-range interchangeably. We only make a clear distinction at those points where it is essential. Our MS-

ARMA-GARCH range model consists of four elements. First there is a mean process that governs the dynamics of the conditional mean of the range. The second element is the process for the variance specifying the dynamics of the conditional variance of the error terms. Third we have to identify the process governing the regimes. Here, we restrict ourselves to two regimes, namely a low and a high volatility regime. By allowing for two regimes only we follow contributions like Hamilton (1989), Gray (1996), Bollen et al. (2000) and Klaassen (2002). Extensions to more than two regimes are nevertheless possible. A last ingredient is the assumed error distribution. As already indicated before, the mean equation is assumed to follow an ARMA(a,b), the variance a GARCH(p,q) process and the regimes we assume to follow an unobserved first order Markov chain. We will assume the errors to be i.i.d. standard Gaussian.

Let us start with specifying the mean equation:

$$r_t = \mu_k + \sum_{i=1}^{a} a_{k,i} r_{t-i} + \sum_{j=1}^{b} b_{k,j} E_{t-1}[(\epsilon_{t-j}|S_{t-j}, \Phi_{t-j})|S_t] + \epsilon_t, \tag{10}$$

where

$$\begin{aligned} \epsilon_t = & \sqrt{h_{k,t}} z_t \\ h_{k,t} = & \omega_k + \sum_{m=1}^{p} \alpha_{k,m} E_{t-1}[(\epsilon_{t-m}|S_{t-m}, \Phi_{t-1})|S_t] \\ & + \sum_{n=1}^{q} \beta_{k,n} E_{t-1}[(h_{t-n}|S_{t-n}, \Phi_{t-1})|S_t]. \end{aligned} \tag{11}$$

In the mean equation μ_k represents the constant term for all different regimes $k = 1, 2, ..., K$, $a_{k,i}$ are all autoregressive coefficients, $b_{k,j}$ are all moving average coefficients and z_t is assumed to be i.i.d. with a $N(0,1)$ distribution. In Eq.(11) ω_k is the constant term of the variance equation, $\alpha_{k,m}$ and $\beta_{k,n}$ are the lagged squared error and lagged variance coefficients, respectively. By this it is clear that S_t fully determines the parameters of the conditional distribution of r_t.

As, for example, in Hamilton (1989) we assume that the regimes S_t follow a first-order Markov process with constant transition probabilities[6]

$$p(S_t = j|S_{t-1} = i, S_{t-2} = j, ...; \Phi_{t-1}) = p(S_t = j|S_{t-1} = i) = p_{ij}, \qquad (12)$$

for $i, j = 1, 2, ..., K$. So, as required by the Markov property the probability of state $S_t = j$ only depends on S_{t-1}, namely the state the process was in at time $t-1$. All these probabilities can be summarized in a $(K \times K)$ transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{K1} \\ p_{12} & p_{22} & \cdots & p_{K2} \\ \vdots & \vdots & \dots & \vdots \\ p_{1K} & p_{2K} & \cdots & p_{KK} \end{bmatrix},$$

where each column of $\mathbf{P}$ sums to unity.

3.3 Estimation

In the regime-switching literature, models are usually estimated by quasi maximum likelihood (QMLE). Gray (1995) proves for some regime-switching models the consistency and asymptotic normality of the QML estimator under relatively mild regularity conditions. We, therefore, follow this path with our MS-ARMA-GARCH range model. As in Gray (1996) and Klaassen (2002), our likelihood has a first-order recursive structure and can be estimated similar to a normal single regime GARCH model. At the same time one can calculate probabilities that the process is in a particular regime at a specific time t, which is very useful if we want to classify our series into periods with low and high volatility. Also following Gray and Klaassen we use two different types of regime probabilities. The first is the ex ante probability of a certain regime. It will be denoted as $p(S_t|\Phi_{t-1})$ and is the conditional probability that the process is in a certain regime at time t given

[6] In general it is also possible to model the transition probabilities as time-varying. Examples are the contributions of Diebold, Lee and Weinbach (1994) and Gray (1996).

only the information set available at time $t-1$. Second, we also calculate the smoothed regime probabilities $p(S_t|\Phi_T, \theta)$ or in short $p(S_t|\Phi_T)$ which use the complete data and information set Φ_T at the estimated coefficient vector θ, thereby smoothing the ex ante probabilities. These smoothed regime probabilities give the econometrician's best inference about the probability of the regime the process was in at time t and will be calculated from the ex ante probabilities we obtain during estimation of the model.

We now introduce the estimation procedure by extending the work of Klaassen (2002) and Gray (1996) to the general case of a MS(K)-ARMA(a,b)-GARCH(p,q) model. Klaassen and Gray are mostly concerned with the Markov-switching aspects in the conditional variance equation. We observe an estimator of the variance and are rather focussing on the mean equation of course not neglecting the variance of the process. Above in Eq.(10) and (11) we already presented a more general model essentially using the same ideas as in Klaassen's paper. Now we turn to the estimation procedure for those models.

In order to obtain the full sample likelihood function we basically have to model the density of every range observation at time t for all possible regimes conditional on only observable information. So, we write that density as:

$$\begin{aligned} f(r_t|\Phi_{t-1}) &= \sum_{k=1}^{K} f(r_t, S_t = k|\Phi_{t-1}) \\ &= \sum_{k=1}^{K} f(r_t|S_t = k, \Phi_{t-1}) p(S_t = k|\Phi_{t-1}), \end{aligned} \tag{13}$$

where we take the sum $\sum_{k=1}^{K}$ of the regime conditional densities over all possible regimes weighted by their respective ex ante probabilities of occurrence $p(S_t = k|\Phi_{t-1})$. Therefore, we can write the distribution of r_t conditional on available information like:

$$r_t|\Phi_{t-1} \sim \begin{cases} f(r_t, S_t = 1|\Phi_{t-1}) \text{ with probability } p(S_t = 1|\Phi_{t-1}), \\ f(r_t, S_t = 2|\Phi_{t-1}) \text{ with probability } p(S_t = 2|\Phi_{t-1}), \\ \quad \vdots \\ f(r_t, S_t = K|\Phi_{t-1}) \text{ with probability } p(S_t = K|\Phi_{t-1}). \end{cases}$$

In the empirical section of this paper we restrict ourselves to the case of $K = 2$. If we assume conditional normality for the error distribution in Eq.(10) we can write:

$$f(r_t|S_t = k, \Phi_{t-1}) = \frac{1}{\sqrt{2\pi h_{k,t}}} exp \left\{ \frac{-\epsilon_{k,t}^2}{2h_{k,t}} \right\}. \tag{14}$$

In general, the errors can for example also be assumed to follow a student-t distribution obviously changing (14) accordingly.

As in Gray (1996) and Klaassen (2002) and according to the assumed first-order Markov structure, the probability $p(S_t = k|\Phi_{t-1})$ depends only on the regime the whole process is in at time $t-1$. If we condition on the regime at time $t - 1$ one can write the ex-ante probability as:

$$\begin{cases} p(S_t = 1|\Phi_{t-1}) = \sum_{k=1}^{K} p(S_t = 1|S_{t-1} = k, \Phi_{t-1}) p(S_{t-1} = k|\Phi_{t-1}), \\ p(S_t = 2|\Phi_{t-1}) = \sum_{k=1}^{K} p(S_t = 2|S_{t-1} = k, \Phi_{t-1}) p(S_{t-1} = k|\Phi_{t-1}), \\ \quad \vdots \\ p(S_t = K|\Phi_{t-1}) = \sum_{k=1}^{K} p(S_t = K|S_{t-1} = k, \Phi_{t-1}) p(S_{t-1} = k|\Phi_{t-1}), \end{cases} \tag{15}$$

where, according to the Markov property,

$$p(S_t = j|S_{t-1} = i, \Phi_{t-1}) = p(S_t = j|S_{t-1} = i) = p_{ij}. \tag{16}$$

So, the probabilities $p(S_t = j|S_{t-1} = i, \Phi_{t-1})$ only depend on S_{t-1} and are equal to the fixed transition probabilities in Eq.(12) which are summarized in the transition matrix $\mathbf{P}$.

Further note that the second part on the right hand side of Eq.(15),

$p(S_{t-1} = k|\Phi_{t-1})$ we can write, according to Bayes' Rule, as:

$$\begin{aligned} p(S_{t-1} = k|\Phi_{t-1}) &= p(S_{t-1}|r_{t-1}, \Phi_{t-2}) \\ &= \frac{p(r_{t-1}|S_{t-1}, \Phi_{t-2})p(S_{t-1}|\Phi_{t-2})}{p(r_{t-1}|\Phi_{t-2})} \\ &= \frac{p(r_{t-1}|S_{t-1}, \Phi_{t-2}) \sum_{S_{t-2=1}}^{K} p(S_{t-2}|\Phi_{t-2})p(S_{t-1}|S_{t-2})}{p(r_{t-1}|\Phi_{t-2})}. \end{aligned} \tag{17}$$

Here, the variables needed to compute $p(S_{t-1} = k|\Phi_{t-1})$ are its previous values $p(S_{t-2} = k|\Phi_{t-2})$, the constant transition probabilities p_{ij} and the densities $p(r_{t-1}|S_{t-1}, \Phi_{t-2})$ and $p(r_{t-1}|\Phi_{t-2})$ from the same calculation one step before. So, the computation of $p(S_{t-1} = k|\Phi_{t-1})$ is a first-order recursive process.

Before we move to the last ingredient for the calculation of the likelihood function we present the full-sample log-likelihood which can be obtained as:

$$Ł = \sum_{t=max(a,b,p,q)}^{T} f(r_t|S_t)p(S_t|\Phi_{t-1}) \tag{18}$$

where the first part on the right hand side is the conditional density at time t given in Eq.(14) and where the second part is the ex ante regime probability described in Eq.(15). Unfortunately, the density $f(r_t|S_t)$ cannot be calculated in a straightforward fashion because of the path dependency in the moving average and variance part of a plain ARMA-GARCH model. So, we have to use Eq.(10) and (11) which necessitates the calculation of the expectations of lagged error and variance terms across regimes. Klaassen (2002) proposed to use all available information up to time $t-1$ to calculate the expected lagged variance in the variance equation of a Markov-switching GARCH(1,1) model. We propose to use the same probability measure to also weigh the lagged error terms in the MA-part and the lagged squared errors in the ARCH-part of our proposed model. In his paper Klaassen proposes a weighing mchanism which gives the probability that the previous regime was S_{t-1} given that the current regime is S_t and given the information set Φ_{t-1}.

It can be stated in the following way:

$$p(S_{t-1}|S_t, \Phi_{t-1}) = \frac{p(S_{t-1}|\Phi_{t-1})p(S_t|S_{t-1}, \Phi_{t-1})}{p(S_t|\Phi_{t-1})} \tag{19}$$

$$= \frac{p(S_{t-1}|\Phi_{t-1})p_{ij}}{p(S_t|\Phi_{t-1})}, \tag{20}$$

where $p(S_{t-1}|\Phi_{t-1})$ is given by Eq.(17), p_{ij} are the fixed transition probabilities in Eq.(12), and $p(S_t|\Phi_{t-1})$ is given by Eq.(15). If one wishes to estimate models with a lag structure $max = max(b, p, q) > 1$ one obviously needs the corresponding probabilities $(p(S_{t-2}|S_t, \Phi_{t-1}), p(S_{t-3}|S_t, \Phi_{t-1}), ..., p(S_{t-max}|S_t, \Phi_{t-1}))$, in order to get the expected values of those lagged error and variance terms as well. Those probabilities can be calculated in a similar way as in Eq.(19). This completes the description of the estimation procedure.

3.4 Smoothed regime inference

As mentioned above, the smoothed regime probabilities represent the econometricians best inference about the regime the process was in at time t using all available information up to time T.[7] In general, one can write $p(S_t|\Phi_\tau)$ for all K regimes the ex post probability as:

$$\begin{aligned} p(S_t|\Phi_\tau) &= p(S_t|r_\tau, \Phi_{\tau-1}) \\ &= \frac{p(r_\tau|S_t, \Phi_{\tau-1})p(S_t|\Phi_{\tau-1})}{\sum_{S_t=1}^{K} p(r_{\tau-1}|S_t, \Phi_{\tau-1})p(S_t|\Phi_{\tau-1})}. \end{aligned} \tag{21}$$

When $\tau = t$, then $p(S_t|\Phi_\tau)$ follows directly because we already know $p(S_t|\Phi_{\tau-1})$ and $p(r_{\tau-1}|S_t, \Phi_{\tau-1})$ from the foregoing maximum likelihood estimation process. For all the following times $(\tau = t+1, t+2, ..., T)$, the calculation of the smoothed probabilities is a first-order recursive process.

If $\tau > t$ we basically need two inputs in order to compute Eq.(21). The first ingredient is the previous K ex post probabilities $p(S_t|\Phi_{\tau-1})$, which are known from the previous iteration. Second, we need to compute the density

[7] This section heavily draws on results by Hamilton (1989), Hamilton (1990), Gray (1996) and especially Klaassen (2002).

$p(r_{\tau-1}|S_t, \Phi_{\tau-1})$ for all K possible regime outcomes. In order to arrive at this density we have to go through some steps. First we can write it as:

$$p(S_\tau|r_t, \Phi_{\tau-1}) = \sum_{S_\tau=1}^{K} p(r_\tau|S_\tau, \Phi_{\tau-1})p(S_\tau|S_t, \Phi_{\tau-1}), \tag{22}$$

where one uses the insight that the conditional distribution of r_τ given S_τ does not depend on the earlier regimes $(S_t, S_{t-1}, ...)$ because we integrate out the path-dependence during the estimation procedure. In Eq.(22) we again have two parts on the right hand side. The first one contains the densities $p(r_\tau|S_\tau, \Phi_{\tau-1})$ for all K regimes, which are known from the estimation procedure. The second part, $p(S_\tau|S_t, \Phi_{\tau-1})$, consists of the $\tau - t$ period transition probabilities of the Markov chain for all possible regime outcomes. By using the Markov property we can rewrite it as:

$$\begin{aligned} p(S_\tau|S_t, \Phi_{\tau-1}) &= \sum_{S_{\tau-1}=1}^{K} p(S_\tau|S_{\tau-1}, \Phi_{\tau-1})p(S_{\tau-1}|S_t, \Phi_{\tau-1}) \\ &= \sum_{S_{\tau-1}=1}^{K} p_{ij}p(S_{\tau-1}|S_t), \end{aligned} \tag{23}$$

where we again, as a first ingredient, have the one period ahead p_{ij} transition probabilities following from Eq.(12). The second part on the right hand side on Eq.(23) can be calculated recursively.

Let us write $p(S_{\tau-1}|S_t, \Phi_{\tau-1})$ for all K^2 regime combinations as:

$$\begin{aligned} p(S_{\tau-1}|S_t, \Phi_{\tau-1}) &= p(S_{\tau-1}|S_t, r_{\tau-1}, \Phi_{\tau-2}) \\ &= \frac{p(r_{\tau-1}|S_{\tau-1}, \Phi_{\tau-2})p(S_{\tau-1}|S_t, \Phi_{\tau-2})}{\sum_{S_{\tau-1}=1}^{K} p(r_{\tau-1}|S_{tau-1}, \Phi_{\tau-2})p(S_{\tau-1}|S_t, \Phi_{\tau-2})}, \end{aligned} \tag{24}$$

where we use the fact that the conditional density $p(r_{\tau-1}|S_{\tau-1}, \Phi_{\tau-2})$ is independent of all earlier regimes once $S_{\tau-1}$ is given. For iteration τ all ingredients in Eq.(24) are known either from the foregoing estimation procedure (the conditional density $p(r_{\tau-1}|S_{\tau-1}, \Phi_{\tau-2})$) or the previous iteration in the calculation of the smoother (the $(\tau - t - 1)$-period ahead transition probability $p(S_{\tau-1}|S_t, \Phi_{\tau-2})$). The ex post probability for $\tau = T$ then gives the smoothed regime probability $p(S_t|\Phi_T)$, which completes the calculation of the smoothed probabilities.

4 Application and results

In this section of the paper, we present the results of fitting our model in Eq.(10) and (11) to the data. First we concentrate ourselves on the S&P500. Later we extend the analysis to the ten year T-notes, three months T-bills, the FTSE100, and the Nikkei225. As a first step we present the data themselves, some descriptive statistics and evidence indicating that there very well might be a hidden Markov process underlying the data causing the data generating process to switch between a low and a high volatility state. As already mentioned above, we assume a two regime Markov process. We then further present the results of fitting different versions of our MS-ARMA-GARCH range model and find that model which fits the data best. We will end this section by briefing possible interpretations of the results.

4.1 The data

The data we use are weekly ranges for the US stock market index S&P500 from January 2^{nd} 1962 until February 11^{th} 2008 summing to 2406 observations in total (observations are on Mondays). We downloaded them from the *yahoo.com* database. In order to arrive at the actual data we transformed the downloaded p_t^{Max} and p_t^{Min}, the highest and the lowest (log-)price index observation, respectively, like:

$$R_t = 100 * (p_t^{Max} - p_t^{Min}). \tag{25}$$

The range R_t is by definition a positive variable and would ask for either a multiplicative model and/or an error distribution that has a lower bound at zero. Furthermore, its unconditional distribution is highly skewed further complicating its modeling. We, therefore, use the log-range[8]:

$$r_t = ln(R_t), \tag{26}$$

which unconditional distribution is surprisingly close to a normal distribution. This result confirms results of, for example, Alizadeh et al. (2002) who also find that the log-range can very well be described as normally distributed - a fact that is uncommon in financial time series, which are usually skewed and show excess kurtosis. Furthermore, Andersen et al. (2003) find that forecasting the log transformation of volatility yields better in- and out-of sample forecasts of the variance because it puts less weight on extreme realizations of the volatility.

Table 1 shows descriptive statistics of our range and log-range data. The Jarque-Bera test statistics reject normality for both series. In the case of log-ranges the statistic still rejects normality but is already very much closer to non-rejection than in the case of the range. We also perform an augmented Dickey and Fuller (1979) test (ADF) with lag-length selection using the Schwarz (1978) information criterion. The null hypothesis of a unit root is clearly rejected for all series. So, there is no need for taking the first difference of the data.

In Figure 1 we show the autocorrelation function and the partial autocorrelation function for the lags $k = 1, ..., 52$, respectively. We observe a quite slow decay in the autocorrelation of the S&P500 range data, which might be interpreted as long memory or, in other words, as a fractionally integrated data generating process. Nevertheless, there is the possibility of confusing long memory with structural breaks or regime-switching behavior. For contributions linking structural breaks and fractional integration see for exam-

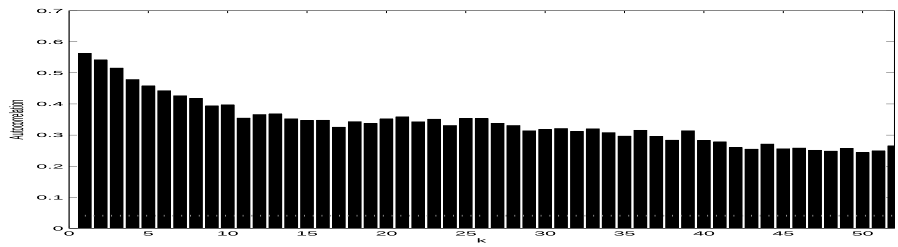
[8]In fact we use an outlier-adjusted version of the data series. We consider all realizations as outliers that are in the upper and the lower 1% percentile of the unconditional distribution. Less than five trading days per week are often responsible for lower tail outliers. Identified outliers are eliminated by taking the average of five consecutive observations, namely the two observations before and after the outlier and the outlier observations itself. By this method we make sure that extreme observations remain extreme. A robustness check showed no significant changes (besides larger Jarque Bera test statistics) in estimation results.

Table 1: Descriptive statistics

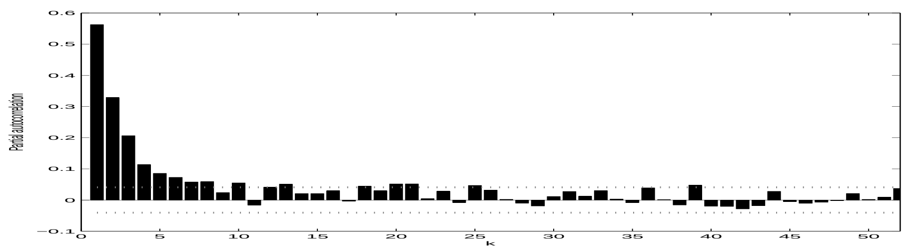
	Weekly observations	
	Range	Log-Range
Mean	3.145	1.045
Median	2.745	1.010
Maximum	12.215	2.503
Minimum	0.988	-0.012
Std.Dev.	1.530	0.441
Skewness	1.576	0.295
Kurtosis	6.250	2.825
Jarque-Bera	2015.069	37.173
P-value	0.000	0.000
ADF test	-9.203	-8.574
P-value	0.000	0.000

Note: Descriptive statistics relating weekly range and log-range observations as derived from Eq.(25) and (26) respectively. Data are from January 2^{nd} 1962 until February 11^{th} 2008 summing to 2406 observations in total. The data are plotted in Fig. 2. Augmented Dickey-Fuller (ADF) test statistics and p-values are calculated based on an automatic lag-length selection using the Schwarz information criterion.

Figure 1: S&P500 (partial) autocorrelation



(a) Autocorrelation



(b) Partial autocorrelation

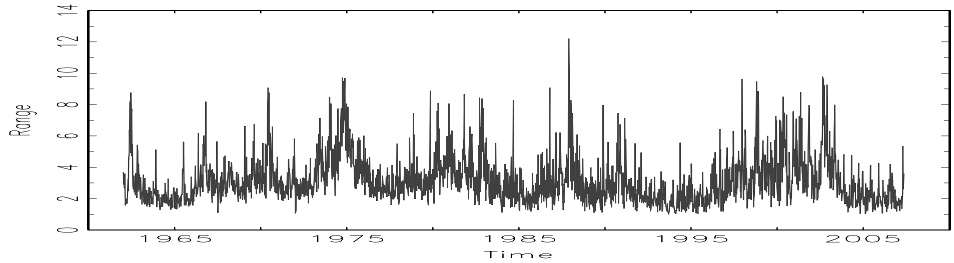
Note: Dotted lines represent the two standard deviation critical values.

ple Bhattacharya et al. (1983), Künsch (1986), and Teverovsky and Taqqu (1997). Even more relevant here is the work of Diebold and Inoue (2001). They find theoretically and by means of simulation that "...structural change in general, and stochastic regime switching in particular, are intimately related to long memory and easily confused with it, so long as only a small amount of regime switching occurs in an observed sample path" (Diebold and Inoue (2001, p.153)). Anticipating on our results, we find such a "small amount" of regime switches in the data. Furthermore, we are interested in the co-occurrence of high and low volatility periods across international financial markets. We also believe that investors are likely to be driven by psychological factors more in line with regime-switching behavior than by long-memory processes. Therefore, we aim for a means of volatility regime classification and thus opt for a Markov-switching model representation as outlined before.

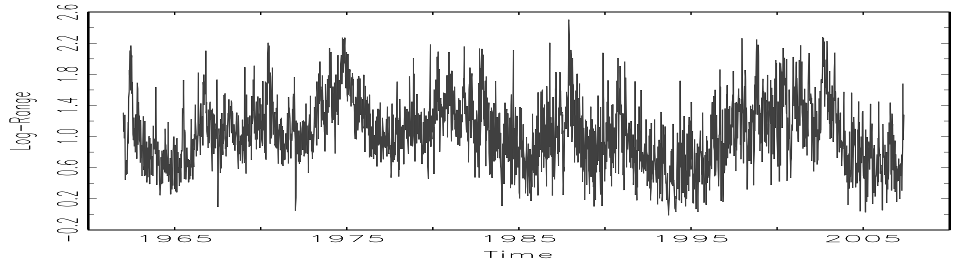
We show the range R_t and the log-range r_t time series in Figure 2. The unconditional distributions of the range and log-range are shown in Figures 3. Obviously, the observation from Table 1 that r_t is close to normally distributed makes R_t appear to have a log-normal distribution.

We continue the data description with an informal time series analysis by having a closer look at the data. One can see in Figure 2 quite clearly that there are periods of relatively low volatility and periods of high volatility. Especially the periods in the middle of the 1970s, the beginning of the 1980s, the late 1980s and from 1998 until 2003 are marked by clearly higher average volatilities measured by the range and/or log-range. High volatility in the early and middle of the 1970s coincides with the break down of the Bretton Woods gold system and the first oil crisis starting in 1973, which was followed by strong reactions of world financial markets. The high volatility period in the beginning of the 1980s corresponds to the second oil crisis, where between 1980 and 1981 the price of crude oil more than doubled within a period of 12 months. In the late 1980s there is another very pronounced but relatively short period of high volatility with a pronounced peek corresponding to "Black Monday" on October 19^{th} 1987. On this day the main US stock markets dropped by ca. 23%, starting a period with extreme uncertainty in

Figure 2: S&P500 range and log-range



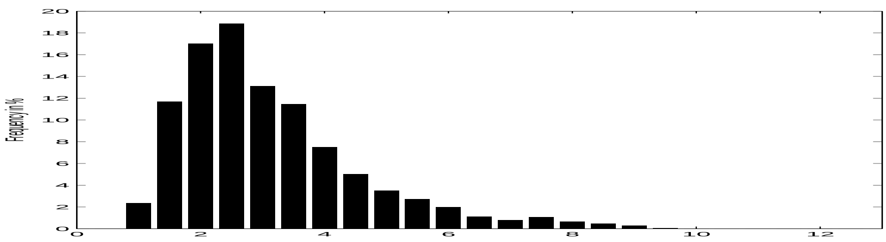
(a) Range



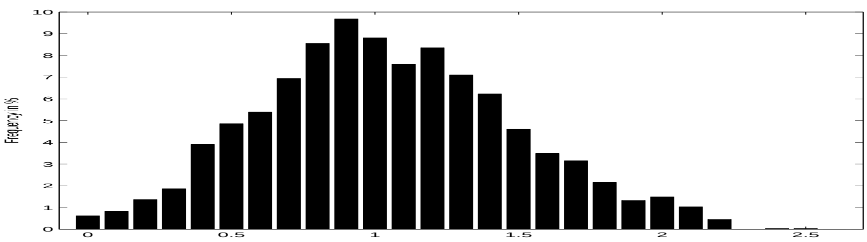
(b) Log-Range

Note: Range is equal to $R_t = 100 * (p_t^{Max} - p_t^{Min})$. The Log-Range is calculated as $r_t = ln(R_t)$.

Figure 3: S&P500 range and log-range unconditional distributions



(a) Range



(b) Log-Range

Note: The range and log-range are calculated as in Figure 2.

asset markets worldwide.[9] This period of increased asset market volatility did not last very long though and markets returned to pre-crash volatility levels before showing some increased volatility again in the beginning of the 1990s during and after the second Gulf War. A further period of higher than normal volatility starts in 1998/1999 probably corresponding to the Russian crisis and the build-up and later burst of the "dot-com-bubble". This period lasts until 2003 which roughly corresponds to the end of the third Gulf War.

In summary, there appear to be quite distinct periods of high and low market uncertainty corresponding to high and low volatility, as measured by the range and log-range, respectively. We think that this is strong evidence for an underlying regime-switching process that might very well be described as a Markov chain. In order to formally test for the presence of a low and high volatility regime we use the testing procedure introduced by Cheung and Erlandsson (2005). They propose a Monte Carlo based testing procedure to simulate an empirical finite sample test statistic for the null hypothesis of one regime (no Markov-switching) against the alternative of two regimes. Such a testing procedure comes in handy because standard statistical procedures fail here. Under the null hypothesis of a linear model with only one regime the nuisance parameters P11 and P22, which are present under the alternative, are not defined making the distribution of the asymptotic log-likelihood ratio test statistic non-standard. Contributions like Hansen (1992; 1996) and Garcia (1998) derived such asymptotic distributions. But still not much is known about their finite sample behavior. We therefore opt for the procedure proposed by Cheung and Erlandsson (2005) which they show to have good power also in finite samples.

We apply the Cheung and Erlandsson (2005) testing procedure to the data by comparing the best fitting linear model with the best fit of the Markow-switching models only allowing for a change in the intercept of the mean equation.[10] We indeed find significant results for the presence of at

[9]See, for example, Shiller (1989) and Carlson (2007).

[10]Such a test can easily be applied to different alternative model specifications. Nevertheless, it appears sufficient to us at this point to take the simplest Markov-switching as an alternative model because it already showed up to be sufficient to generate significant results. Furthermore, any more complicated alternative model specification would have

least two regimes. The p-value of the likelihood ratio test was found to be at 2%, clearly rejecting the null hypothesis of a linear specification in favor of the alternative Markov-switching hypothesis justifying the further procedure of modeling the log-range according to Eq.(10) and (11).

Another criterion for a well fitting regime-switching model should be that it is capable of at least also identifying some of the periods of high and low volatility visually found in the graphs before. Therefore, we present the results of fitting the considered MS-ARMA-GARCH models to the data in Section 4.2.

4.2 Estimation results

In Section 4.1 we showed that the weekly S&P500 range and log-range are very likely to be drawn from at least two different densities and thereby from more than one volatility regime. In this section we aim at finding the best fitting, parsimonious model from our proposed class of MS(2)-ARMA(a,b)-GARCH(p,q) models generally described in Eq.(10) and (11), which are reproduced here for convenience:

$$r_t = \mu_k + \sum_{i=1}^{a} a_{k,i} r_{t-i} + \sum_{j=1}^{b} b_{k,j} E_{t-1}[(\epsilon_{t-j}|S_{t-j}, \Phi_{t-j})|S_t] + \epsilon_t,$$

where

$$\begin{aligned} \epsilon_t = & \sqrt{h_{k,t}} z_t \\ h_{k,t} = & \omega_k + \sum_{m=1}^{p} \alpha_{k,m} E_{t-1}[(\epsilon_{t-m}|S_{t-m}, \Phi_{t-1})|S_t] \\ & + \sum_{n=1}^{q} \beta_{k,n} E_{t-1}[(h_{t-n}|S_{t-n}, \Phi_{t-1})|S_t]. \end{aligned}$$

We will have to fit different specifications of the MS(2)-ARMA(a,b)-GARCH(p,q) models in order to be able to decide upon which one fits the data best. We will proceed in a bottom-up way. We start with MS(2)-ARMA(a,b) specifica-

increased computing time without giving more insights.

tions without looking at possible GARCH or volatility of volatility clustering effects. In order to make sure that the QMLE estimation arrived at the global maximum likelihood, we estimate the models with 100 different randomly drawn starting values. To check for a good fit we will employ different means. A very important criterium will obviously be to check, if there is any autocorrelation in the standardized residuals and/or the squared standardized residuals left. Any remaining autocorrelation in the residuals asks for an increase in the amount of ARMA-terms. Any remaining autocorrelation in the square of the standardized residuals hints at GARCH-effects not sufficiently accounted for by the model, and we might need to add more ARCH or GARCH terms. The best fitting model will not have any remaining autocorrelation in the standardized residuals or squared standardized residuals. So, the following subsections analyze the data in more detail.

4.2.1 Only the intercept changes with the regime

In the empirical implementation we allow different parts of Eq.(10) and (11) to change with regimes. An ARMA-C or an ARMA-X specification mean that only the constant or all parameters in that part of the model are allowed to change, respectively. In this subsection we concentrate on the different MS(2)-ARMA-C(a,b)-GARCH(p,q) model specifications. In all the coming models we let only the constant or intercept, μ_k, in the mean equation (Eq.(10)) change with the regime. Later, we also experiment with regime dependent ARMA and GARCH parameters in order to find out if the volatility of volatility is changing with time as well. We present all relevant estimation results in Table 2. The columns represent all different specifications with parameter estimates and standard errors reported. We also show the value of the maximized log-likelihood function and Ljung-Box (LB) and Jarque-Bera statistics in order to check for residual and squared standardized residual autocorrelation and normality of the residual distributions.

We start with the most parsimonious specification being the MS(2)-AR(1) model. Here, we can already see that there are clearly two different volatility regimes in the S&P500 data over the considered sample period. The constant terms in either regime (regime 1 and 2) differ significantly from each other. Checking for correct model specification by inspecting the Ljung-Box statistics both for the standardized residuals and squared residuals it becomes apparent that the simple MS(2)-AR(1) specification cannot completely eliminate autocorrelation in the residuals and their squares. Two points arise from this. First, we need to increase the order of ARMA-terms in the mean equation. Second, there is evidence for conditional heteroscedasticity in the residuals asking for the inclusion of some ARCH and/or GARCH terms in order to allow for a time-varying variance. Though, before specifying the conditional variance of the range, we first proceed in finding an ARMA-specification that is able to account for the autocorrelation in the residuals. Afterwards we continue with modeling the conditional heteroscedasticity.

Table 2: Estimation results for weekly S&P500 data only allowing the constant (μ) to change

	AM(1,0)		AM(1,1)		AM(1,1)-G(1,0)		AM(1,1)-G(1,1)		AM(1,1)-G(2,1)	
Parameters	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.
μ_1	0.5413	0.0004	0.0410	0.0073	0.0589	0.0050	0.0560	0.0053	0.0545	0.0020
μ_2	0.8865	0.0009	0.1741	0.0554	0.0934	0.0114	0.0900	0.0148	0.0876	0.0041
a_1	0.3426	0.0003	0.9522	0.0030	0.9236	0.0096	0.9263	0.0123	0.9280	0.0032
b_1			-0.7533	0.0046	-0.6956	0.0008	-0.6808	0.0094	-0.6907	0.0039
ω	0.1129	0.0000	0.1051	0.0007	0.1000	0.0318	0.0003	0.0000	0.0001	0.0001
α_1					0.0939	0.0971	0.0128	0.0005	0.0892	0.0111
α_2									-0.0797	0.0109
β_1							0.9842	0.0007	0.9893	0.0002
$P11$	0.9899	0.0000	0.9779	0.0196	0.9958	0.0003	0.9964	0.0011	0.9961	0.0003
$P22$	0.9867	0.0004	0.6789	0.2956	0.9969	0.0000	0.9947	0.0002	0.9962	0.0002
Log-Likelihood	-838.060		-728.313		-717.875		-695.227		-688.276	
					P-Values					
LB_1	0.000		0.513		0.975		0.625		0.781	
LB_5	0.000		0.286		0.726		0.775		0.808	
LB_{10}	0.000		0.524		0.871		0.909		0.917	
LB_1^2	0.000		0.000		0.612		0.000		0.969	
LB_5^2	0.000		0.000		0.003		0.002		0.375	
LB_{10}^2	0.000		0.000		0.000		0.007		0.318	
Jarque-Bera	13.314		6.774		36.263		33.699		32.855	
P-value	0.001		0.034		0.000		0.000		0.000	

Note: AM(a,b)-G(p,q) is short for an ARMA(a,b)-GARCH(p,q) specification. Parameters are estimated using the GAUSS6.0 conditional optimization package (co) under the constraints of all ARMA and GARCH roots lying outside the unit circle. Additionally, we impose a positivity constraint for the variance and conditional variance. We apply the standard convergence criteria. The parameters are as in Eq.(10) and (11) for the respective model specifications. $P11$ and $P22$ are the Markov chain transition probabilities for every period for staying in the low and in the high volatility regime, respectively. LB_x stands for the Ljung-Box test at x lags of the standardized residuals. LB_x^2 is the same but for squared standardized residuals. For the Ljung-Box test we only report p-values for the null hypothesis of no autocorrelation. Ljung-Box test statistics are not adjusted for ARCH effects as suggested in Diebold (1986). The Jarque-Bera test tests for standard normality in the standardized residuals. For the Jarque-Bera test we report the test statistics and corresponding p-values.

Already an ARMA(1,1) specification is able to deliver insignificant autocorrelation levels in the residuals, which can be checked by looking at the Ljung-Box statistics.[11] But we still have clear evidence for remaining conditional heteroscedasticity. The addition of an equation specifying the conditional variance solves this problem.

In order to take care of the conditional heteroscedasticity in the data we specify different GARCH(p,q) models. We only report the results for the GARCH(1,0), GARCH(1,1) and the GARCH(2,1) cases in Table 2. The GARCH(1,0) specification for the variance equation does not seem to be sufficient to justify the i.i.d. assumption for the residuals because the Ljung Box statistics for the standardized squared residuals are still significant.[12] We therefore try two different approaches, namely augmenting the conditional variance with a lagged conditional variance term (GARCH(1,1)) and augmenting it with higher order ARCH-terms (GARCH(2,1)). Also the assumption of a GARCH(1,1) specification does not fully solve the problem of not having i.i.d. residuals because the Q-statistic at one lag is still significant at a 5% level. The GARCH(2,1) model though delivers insignificant autocorrelations for the squared residuals at a 10% significance level.[13] We therefore consider the MS(2)-ARMA-C(1,1)-GARCH(2,1) model as fitting the data best. By inspecting the Jarque-Bera test statistic it is apparent that the normality assumption is likely to be violated, though.

By having a closer look at the coefficients of the MS-ARMA-C(1,1)-GARCH(2,1) model in Table 2, one can see a quite clear difference in the constant terms of either regime. In the low volatility regime μ_1 is equal to

[11] The Ljung-Box statistics in Table 2 and Table 3 have been calculated without accounting for ARCH effects. Diebold (1986) suggested to adjust the standard Ljung-Box statistics in the presence of ARCH effects. Such an adjustment would nevertheless only decrease the test statistics and increase the p-values to some extend. Such an adjustment is therefore not likely to affect the choice of the appropriate model significantly and we therefore only report the unadjusted Ljung-Box test statistics. Adjusted test statistics as suggested in Diebold (1986) are available upon request.

[12] By inspecting the Ljung-Box Q-statistics more closely, we find that especially the first four lags cause the rejection of the no autocorrelation null hypothesis. Detailed results are not reported here, but are available upon request.

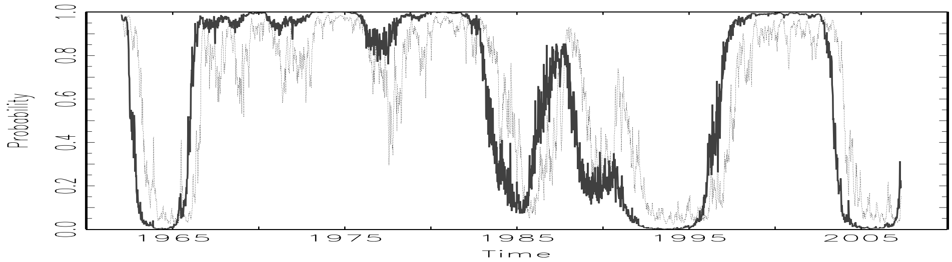
[13] Because of space considerations we do not report all those test statistics in Table 2. They are available on request.

0.0545 whereas in the high volatility regime μ_2 is equal to 0.0876. These intercepts and the AR-coefficient of 0.928 give us the unconditional log-range values of 0.757 and 1.217 for the low and the high volatility regime, respectively. Such log-range values translate into ranges of 2.132 and 3.376, respectively, which corresponds to a, on average, 61% larger volatility during periods with high volatility as compared to those periods with low volatility. Furthermore, the parameters in the variance equation α_1 and β_1 add up to 0.9988, which suggest a quite persistent conditional volatility of the log-range, where the biggest contribution of this persistence comes from the GARCH and not the instantaneous ARCH parameters. This suggests that shocks to the volatility of the volatility die out quite slowly.

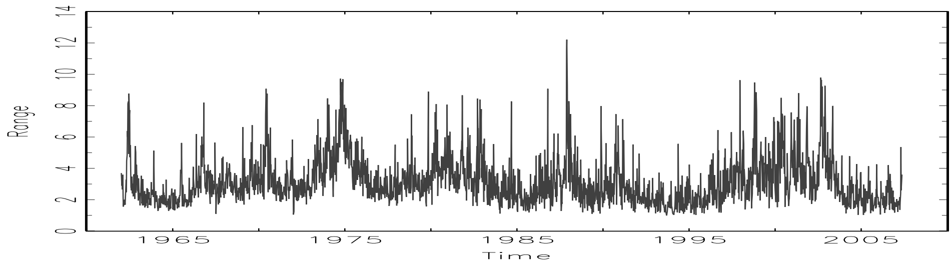
We also present the ex ante and smoothed regime probabilities derived from the best fitting model for the weekly data. Figure 4 shows the ex ante and the smoothed regime probabilities in Panel (a), and the corresponding range observations in Panel (b). There is a clear peak in the smoothed and ex ante probabilities around the 1987 stock market crash. Also the high volatility period from 1997 until 2003 is clearly identified. Interestingly, the weekly data ranging back to the beginning of the 1960s also identify a longer period of high volatility from the end of the 1960 until the beginning of the 1980s. As already mentioned above, this period was characterized by many world economic changes and crises, as for example the first and the second oil shock and the collapse of the Bretton Woods system.

In sum, our proposed model for the weekly log-range S&P500 data do a good job in terms of identifying important periods of financial uncertainty and increased volatility in a very important US stock market index. They are capable of distinguishing quite clearly low- and high-volatility periods from each other. Also standardized residuals do not show important signs of autocorrelation or remaining unexplained conditional heteroscedasticity, which justifies the i.i.d. assumption important for quasi maximum likelihood estimation.

Figure 4: Regime probabilities (only the constant changes)



(a) Ex ante and smoothed probabilities



(b) Range

Note: In Panel (a) we show the ex ante (dotted line), $p(S_t = 2|\Phi_{t-1})$, and smoothed probabilities (solid line), $p(S_t = 2|\Phi_T)$, which are calculated as in Section 3.3 and 3.4 respectively. Both show the probability that the data at time t are drawn from the high volatility regime distribution. Panel (b) shows the corresponding observed range data for which the probabilities are calculated. All probabilities are obtained from the daily MS(2)-ARMA-C(1,1)-GARCH(4,0) model, where the C stands for only constant, meaning that only the constant is allowed to change with the regime.

4.2.2 Allowing all mean equation parameters to change

Up to now we only allowed for changes in the constant term of the mean equation in Eq.(10). We also would like to check the evidence for changes in the dynamics. It might be that the dynamics of the range as a time series change with the regime. One might argue that in a high volatility regime the dependence of the volatility today on the volatility in the past is different compared to the low volatility regime because investors could change their behavior according to their perception of what volatility regime markets are in. In order to check for differences in the dynamics across regimes we let all parameters of the mean equation free to change with the regime. A note of caution here is that there might be other Markov-chains governing the switches in the intercept, the AR and MA parameters. If the true data generating process is governed by more than one Markov-chain, the estimation results and classification of regimes would be distorted. That is also why we present Sections 4.2.1 and 4.2.2 separately. In the estimation we follow the same approach as in the case where we only let the constant, μ, change for identifying the appropriate model.

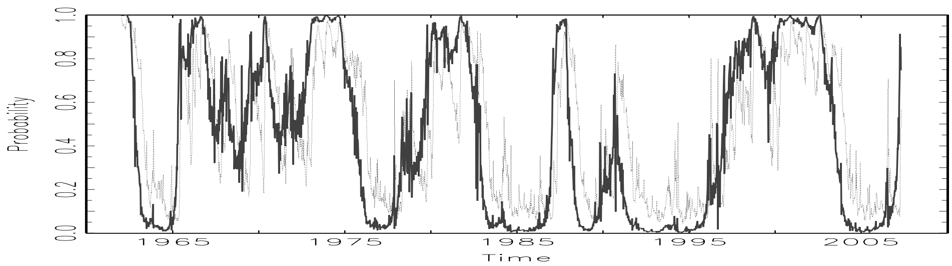
We present the estimation results in Table 3 and the ex ante and smoothed probabilities of the best fitting model in the corresponding Figure 5. Again we take the same approach for model selection as before. The best fitting model here, where we allow all parameters of the mean equation (10) to change, is the ARMA-X(1,1)-GARCH(2,1) specification.

Table 3: Results for weekly S&P500 data allowing all mean equation parameters to change

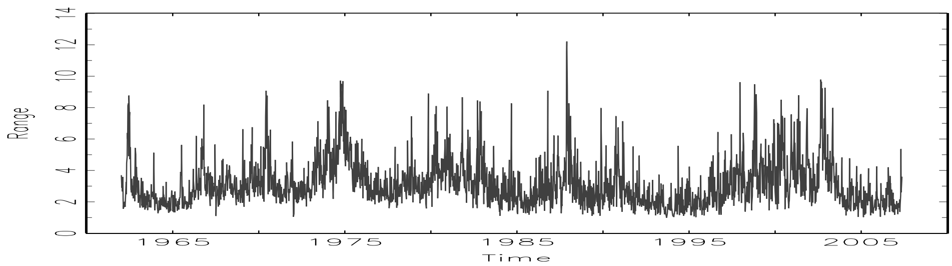
	AM(1,0)		AM(1,1)		AM(1,1)-G(1,0)		AM(1,1)-G(1,1)		AM(1,1)-G(2,1)	
Parameters	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.
μ_1	0.6000	0.0050	0.0288	0.0483	0.0281	0.0536	0.0227	0.0002	0.0244	0.0012
μ_2	0.7005	0.0723	0.1711	0.0066	0.1555	0.3773	0.1357	0.0013	0.1103	0.0015
$a_{1,1}$	0.2326	0.0174	0.9655	0.0617	0.9651	0.0493	0.9745	0.0003	0.9670	0.0017
$a_{2,1}$	0.4550	0.0361	0.8714	0.0004	0.8809	0.2548	0.8969	0.0008	0.9132	0.0011
$b_{1,1}$			-0.8223	0.0546	-0.8347	0.1628	-0.8255	0.0009	-0.8388	0.0014
$b_{2,1}$			-0.5257	0.1408	-0.5537	0.3383	-0.4987	0.0005	-0.5838	0.0077
ω	0.1130	0.0033	0.1055	0.0006	0.0957	0.0015	0.0002	0.0000	0.0001	0.0000
α_1					0.0938	0.0302	0.0110	0.0000	0.0798	0.0103
α_2									-0.0710	0.0104
β_1							0.9866	0.0002	0.9901	0.0048
$P11$	0.9914	0.0033	0.9932	0.0020	0.9931	0.0149	0.9927	0.0000	0.9908	0.0000
$P22$	0.9903	0.0031	0.9869	0.0125	0.9887	0.0117	0.9810	0.0001	0.9922	0.0005
Log-Likelihood	-828.018		-716.746		-708.093		-685.380		-680.768	
					P-Values					
LB_1	0.000		0.716		0.988		0.877		0.917	
LB_5	0.000		0.772		0.774		0.907		0.887	
LB_{10}	0.000		0.847		0.825		0.965		0.936	
LB_1^2	0.000		0.002		0.000		0.046		0.164	
LB_5^2	0.000		0.000		0.000		0.107		0.243	
LB_{10}^2	0.000		0.000		0.000		0.094		0.177	
Jarque-Bera	30.241		27.562		38.670		34.113		34.713	
P-value	0.001		0.000		0.000		0.000		0.000	

Note: AM(a,b)-G(p,q) is short for an ARMA(a,b)-GARCH(p,q) specification. Parameters are estimated using the GAUSS6.0 conditional optimization package (co) under the constraints of all ARMA and GARCH roots lying outside the unit circle. Additionally, we impose a positivity constraint for the variance and conditional variance. We apply the standard convergence criteria. The parameters are as in Eq.(10) and (11) for the respective model specifications. $P11$ and $P22$ are the Markov chain transition probabilities for every period for staying in the low and in the high volatility regime, respectively. LB_x stands for the Ljung-Box test at x lags of the standardized residuals. LB_x^2 is the same but for squared standardized residuals. For the Ljung-Box test we only report p-values for the null hypothesis of no autocorrelation. Ljung-Box test statistics are not adjusted for ARCH effects as suggested in Diebold (1986). The Jarque-Bera test tests for standard normality in the standardized residuals. For the Jarque-Bera test we report the test statistics and corresponding p-values.

Figure 5: Regime probabilities (all mean equation parameters change)



(a) Ex ante and smoothed probabilities



(b) Range

Note: In Panel (a) we show the ex ante (dotted line), $p(S_t = 2|\Phi_{t-1})$, and smoothed probabilities (solid line), $p(S_t = 2|\Phi_T)$, which are calculated as in Section 3.3 and 3.4 respectively. Both show the probability that the data at time t are drawn from the high volatility regime distribution. Panel (b) shows the corresponding observed range data for which the probabilities are calculated. All probabilities are obtained from the daily MS(2)-ARMA-X(1,1)-GARCH(1,1) model, where the X stands for all, meaning that all parameters in the ARMA equation are allowed to change with the regime.

In Table 3 three interesting results, compared to the earlier results where we only allowed the constant in the mean equation to change, appear. First, there seems to be a quite clear difference in the autoregressive coefficients across the regimes. In the case of the ARMA-X(1,1)-GARCH(2,1) specification we estimate the AR-coefficients for the low- and the high volatility regime to be equal to 0.9670 and 0.9132, respectively. This means that the half-life of a shock to the volatility is 21 weeks in the case of the low and 7 weeks for the high-volatility regime. So, in the low-volatility even 21 weeks after a shock around 50% of it is still present in the actual volatility. In the high-volatility regime markets seem to “forget” much more quickly. Here the half-life of a shock is around seven weeks. This confirms the results of Gray (1996) and Klaassen (2002) who also find that volatility persistence in the high volatility regime is lower. A second result is that also the moving-average parameter in the high-volatility regime are lower in absolute value than those in the low-volatility regime. The third interesting results is that the GARCH structure, being a GARCH(2,1), does not change compared to the results before and thereby appears to be very robust through different regimes and specifications. So, also when we allow all parameters of the mean equation to change with the regime, there still is strong evidence for quite persistent volatility of the volatility.

We also experimented with specifications allowing all parameters of Eq.(10) and (11) to vary with the state. Results were inconclusive though, which probably is due to large amounts of coefficients that need to be estimated.[14] Another possibility is that the variance equation might be governed by a second Markov chain that not necessarily coincides with the Markov chain governing the parameters of the mean equation. We therefore, do not report those results here.

4.3 Comovements of volatility across asset markets

Above we showed how our Markov-switching model can be applied to the stock market data of the US. For investors, for example, it is important to know if different asset markets tend to be in the same volatility regime at the same time because this has repercussions on hedging and portfolio diversification effects. We therefore apply the basic model to different asset markets in order to find possible comovements in the volatility regimes across countries and assets. Although this analysis will be limited in scope (for example, we limit ourselves to the most simple model specifications) we find evidence for interesting changes of volatility comovement in the sample.

The assets we are going to analyze are three months US T-bills, ten years US T-notes both starting in October 1965, the FTSE100 UK stock market index starting in April 1984, and the Japanese stock market index NIKKEI225 starting in August 1986. All sampled series run until February 2008. In the case of ten year T-notes we split the sample into two subsamples and fit each one separately to the model. The split is chosen to be in August

[14] Results are available upon request.

1979 the month when Paul Volcker became the Fed's chairman. In Figure 8 one can see a clear and lasting change in the behavior of the range of T-notes. It is well known that Volcker ended the Fed's policy of targeting the interest rate but rather focused on the money supply by limiting the money growth. We take this as sufficient evidence for an exogenous change in the policy regime.

In the estimation procedure we follow exactly the same setup as before in the S&P500 case. We choose that model as the best fit, which delivers standardized and squared standardized residuals that do not show any sign of remaining autocorrelation. In Table 4 we present the results of fitting our Markov-switching model to the different assets. We focus here on the model where we only allow the intercept to change with the regime because we only want to draw attention on the level of the volatility and not on changes in dynamics. As already explained in Section 3, also allowing the other parameters to change with the regime may distort the separation of regimes when the level of the volatility and it's dynamics are governed by different Markov chains. In order to prevent such effects we limit ourselves on the level of volatility only. The structure of the table is similar to Tables 2 and 3 before. Here we only neglect the test statistics on the autocorrelations. The Jarque-Bera test statistics for normality for T-bills, T-notes, FTSE and Nickei cannot reject the null hypothesis at a 1% level. Results for the S&P500 stay as before and are repeated for convenience. Parameter estimates concerning the mean equations do not show strong qualitative differences across markets. S&P500, T-bills and the T-notes sample starting in 1979 clearly show significant GARCH effects. In the case of the other assets an ARCH(1) model is sufficient to account for the conditional variance of the range. Further, the Markov transition probabilities indicate that both the low and the high volatility regimes are persistent in all the cases.

In Figures 6 to 10 we report the models' estimates for the regime probabilities and the underlying range data. The figures have again the same interpretation as Figures 4 and 5. Figure 6 just repeats the results of Figure 4 for convenience. At first sight there are obvious comovements of volatility regimes across asset markets. For example, all markets are found to be in

Table 4: Estimation results for different weekly asset markets

Parameters	S&P500	TB	TN(79)	TN(08)	FTSE100	NIKKEI225
μ_1	0.0545	0.1024	-0.0683	0.0210	0.1487	0.1461
μ_2	0.0876	0.2452	0.0429	0.0641	0.2534	0.2350
a_1	0.9280	0.8049	0.7895	0.9479	0.8141	0.8355
b_1	-0.6907	-0.5514	-0.6530	-0.7863	-0.5554	-0.5743
ω	0.0001	0.0087	0.3683	0.0001	0.1337	0.1430
α_1	0.0892	0.0261	0.0992	0.0136	0.0270	0.0453
α_2	-0.0797					
β_1	0.9893	0.9502		0.9860		
$P11$	0.9961	0.9936	0.9879	0.9984	0.9953	0.9846
$P22$	0.9962	0.9899	0.9853	0.9987	0.9904	0.9926
Log-Likelihood	-688.28	-2034.15	-708.52	-932.92	-547.03	-537.36
Jarque-Bera	32.86	2.98	7.97	2.95	5.90	5.89
P-value	0.00	22.55	1.87	22.89	5.24	5.24

Note: TB stands for T-bills. TN(79) and TN(08) refer to T-notes samples from 1962 until 1979 and from 1979 until 2008, respectively. Parameters are estimated using the GAUSS6.0 conditional optimization package (co) under the constraints of all ARMA and GARCH roots lying outside the unit circle. Additionally, we impose a positivity constraint for the variance and conditional variance. We apply the standard convergence criteria. The parameters are as in Eq.(10) for the respective model specifications. $P11$ and $P22$ are the Markov-chain transition probabilities for every period for staying in the low and in the high volatility regime, respectively. The Jarque-Bera test tests for standard normality in the standardized residuals. For the Jarque-Bera test we report the test statistics and corresponding p-values.

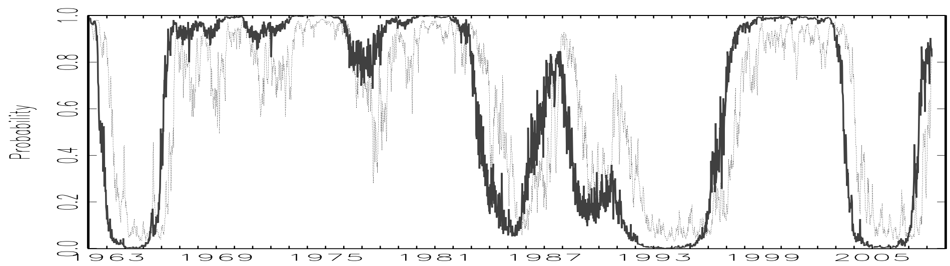
a high volatility regime in the end of 1990 and in the beginning of the new millennium. Further we find that the US stock market, T-bills and T-notes markets are in the high volatility regime around the the first and the second oil crises and the collapse of the Bretton Woods system.

In order to quantify the degree of comovements in the volatility regimes across asset markets we calculate concordance indices for different asset combinations. The concordance index can be calculated as follows:

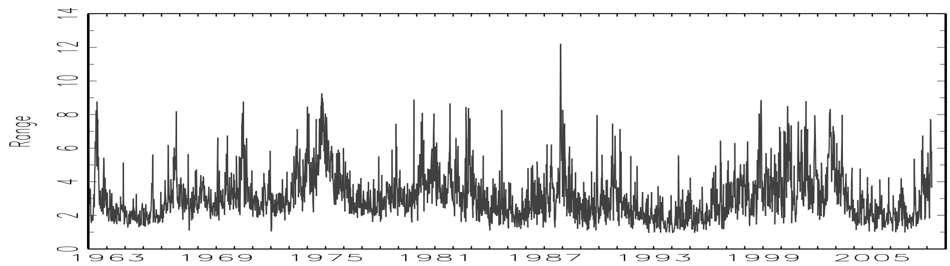
$$CI_{i,j} = \frac{1}{T}\sum_{t=1}^{T}[H_t^i H_t^j + L_t^i L_t^j], \tag{27}$$

where H_t^i is an indicator function that equals one if market i is in the high volatility regime at time t and L_t^i is equal to one if it is in the low volatility regime. Thereby, the index can only assume values between 0 and 1. It equals 0 if there is absolutely no concordance and it equals 1 if there is perfect concordance. The extension to the case with more than two markets is straightforward. In order to distinguish low and high volatility regimes we use the estimated smoothed probabilities and follow the rule that $H_t^i = 1$ when $p(S_t|\Phi_T) > 0.5$, where $p(S_t|\Phi_T)$ is the probability that the observation of the range at time t came from the high volatility regime. When that probability is smaller than 0.5 we classify it as coming from the low volatility regime. Other classification algorithms are obviously possible, but the cutoff at 50% appears to be the most natural one.

Figure 6: S&P500 range and smoothed probabilities



(a) Smoothed probabilities



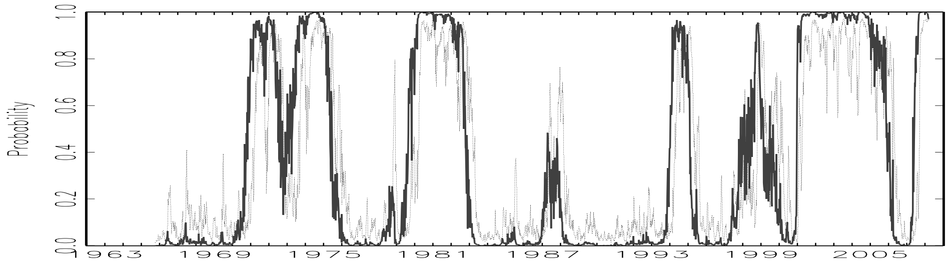
(b) Range

Note: Range is equal to $R_t = 100 * (p_t^{Max} - p_t^{Min})$.

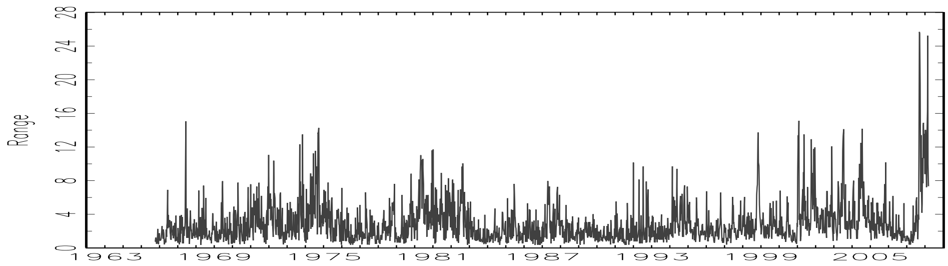
In Figure 11 we show the results of recursive concordance index calculations. This means that at time t the value of the concordance index reflects the concordance of the sample up to and including point t. Therefore, the last observation then gives the concordance of the considered assets for the full sample. Like this we can see the development of the degree of comovement of assets' volatilities over time but also keep the overall comparability over time. By construction the recursive concordance index shows larger variability at the beginning than at the end of the sample.

In general, over time the figures show a tendency of increasing comovements of the volatility regimes across assets and countries. Especially the

Figure 7: T-bills range and smoothed probabilities



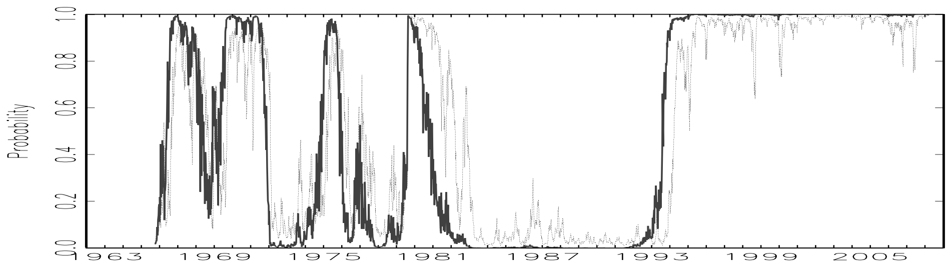
(a) Smoothed probabilities



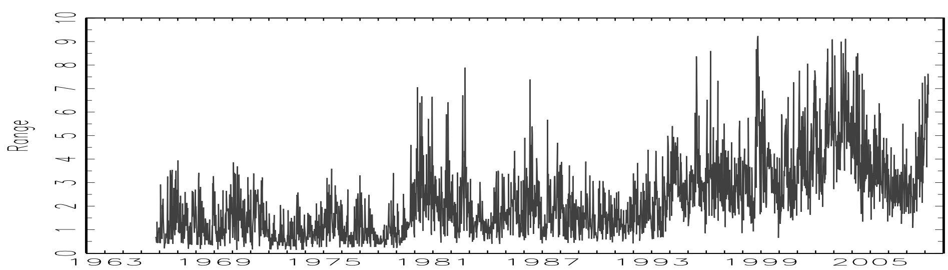
(b) Range

Note: Range is equal to $R_t = 100 * (p_t^{Max} - p_t^{Min})$.

Figure 8: T-notes range and smoothed probabilities



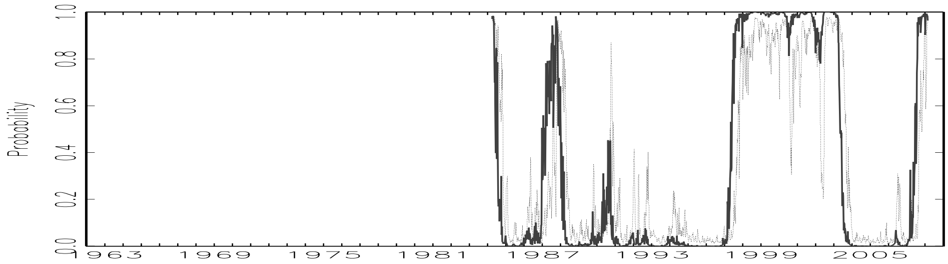
(a) Smoothed probabilities



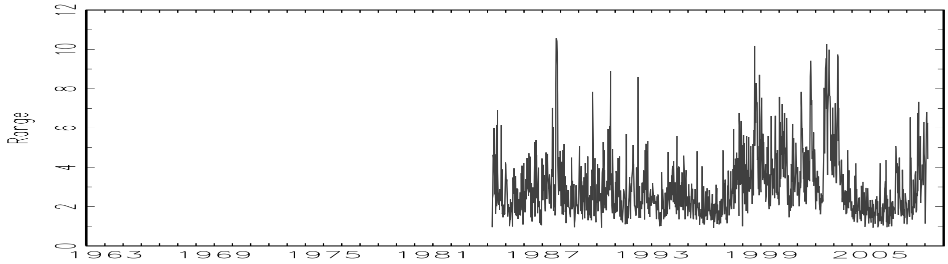
(b) Range

Note: Range is equal to $R_t = 100 * (p_t^{Max} - p_t^{Min})$.

Figure 9: FTSE range and smoothed probabilities



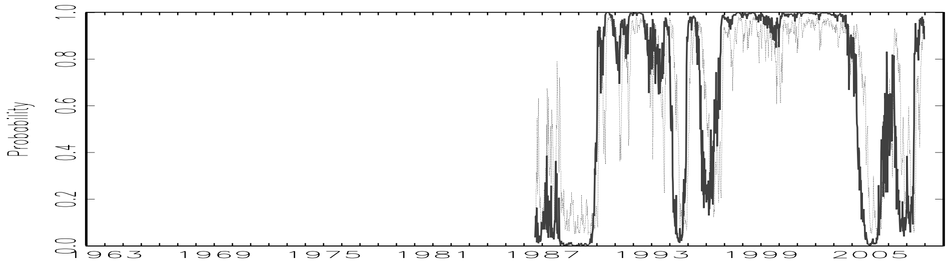
(a) Smoothed probabilities



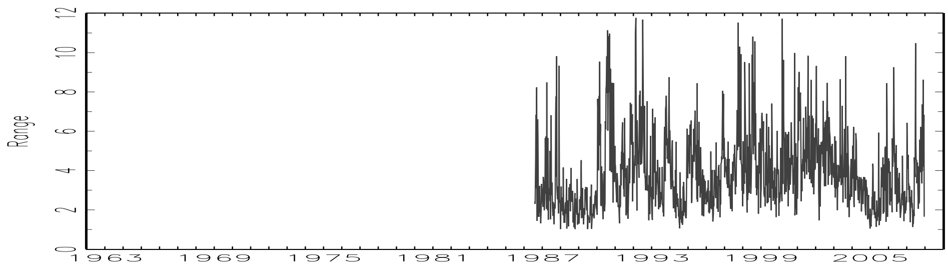
(b) Range

Note: Range is equal to $R_t = 100 * (p_t^{Max} - p_t^{Min})$.

Figure 10: NIKKEI range and smoothed probabilities



(a) Smoothed probabilities



(b) Range

Note: Range is equal to $R_t = 100 * (p_t^{Max} - p_t^{Min})$.

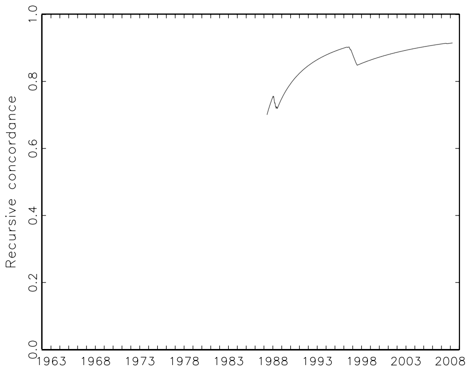
S&P500 with the FTSE100 and T-notes with T-bills clearly show such a tendency. Here also S&P500 and FTSE100 show very high absolute values of concordance even reaching higher than 0.9 at the end of the sample period. This clearly shows a strong comovement of volatility regimes of the US stock market with the UK stock market, whereas the comovement of the US with the Japanese stock market is much lower in absolute value. Another clear point is the strong increase in comovement of volatility regimes among the considered stock markets beginning in the mid and late 1990s. Such an increase in concordance is strongly driven by the high volatility regime starting at the end of the nineties and reaching far into the new millennium. This long period of high volatility we find in virtually all stock markets and seems to be an international or even global phenomenon. A further interesting observation is also the clear increase in comovement of volatility regimes among the US stock, T-notes, and T-bills markets. It appears that also those markets show a tendency towards stronger co-volatility and thereby increasing integration through time. Such an tendency of increasing comovements of volatility across asset classes and asset markets can either be due to some global factors affecting all markets similarly or it reflects a stronger international and inter-asset integration of the markets. An identification of the precise causes of the increased comovements is not possible with these methods and beyond the scope of this paper.

4.4 Forecasting performance

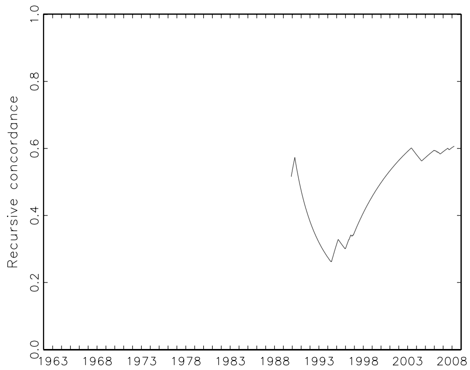
A better fit to the data is already an own end for modeling the data generating process of volatility as a Markov-switching model instead of a linear one in order to identify high and low volatility periods within the sample. But another interesting point is a comparison of forecasting performances. In this subsection we present the results of an in-sample forecasting comparison of our proposed Makrov-switching model with a linear ARMA-GARCH specification. We estimate the best fitting linear[15] and the best fitting MS

[15] The linear model we consider is an ARMA(1,1)-GARCH(2,1) specification without Markov-switching. For brevity we do not show the details of this model here, but they are available upon request.

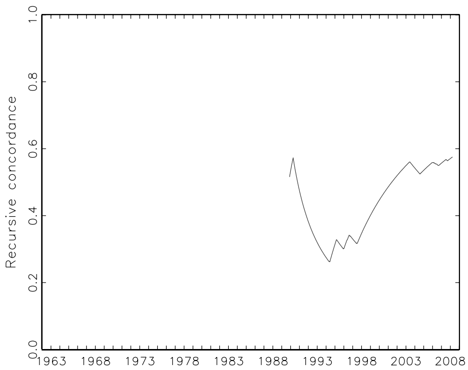
Figure 11: Concordances



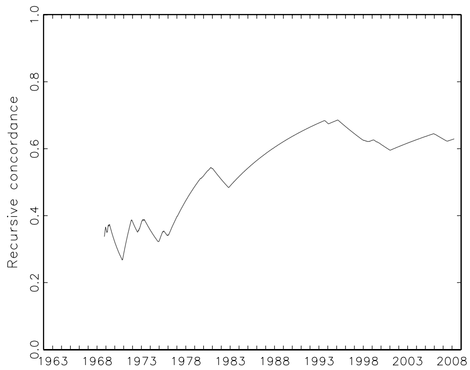
(a) S&P500,FTSE



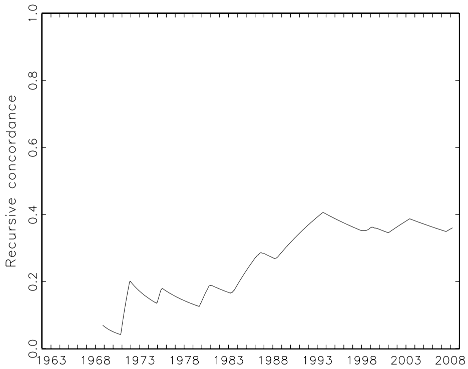
(b) S&P500,NIKKEI



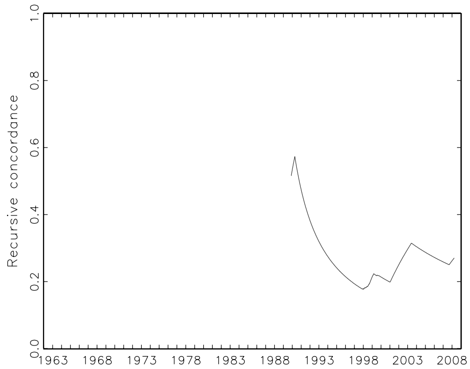
(c) S&P500,FTSE,NIKKEI



(d) T-bills,T-notes



(e) S&P500,T-bills,T-notes



(f) S&P500,T-bills,T-notes,FTSE,NIKKEI

Note: Concordances are calculated using (27) recursively.

model using the full sample. Then we pick a starting point t in the sample and forecast F periods into the future. After obtaining such a forecast we go to observation $t+1$ and do the same again rolling through the sample until we arrive at period $T-F$ which is the period of the last forecast. An underlying assumption of such a procedure is that the parameter estimates do not change much by either estimating the models with the full sample or by always re-estimating it.[16] For calculating our forecasts we follow the methods developed in Davidson (2004) where he proposes a method for multi-period forecasting with a Markov-switching dynamic regression model accounting for conditional heteroscedasticity.

Imagine that we want to forecast r_{t+F} for $F \geq 1$ given observations on the process up to date t. With other words the object of interest is $E(r_{t+F}|\Phi_t)$. Davidson (2004) develops a recursion for computing $E(r_{t+F}|\Phi_t)$, which we denote by $\hat{r}_{t+F}$ for brevity. Such a recursion involves only K terms at each iteration. The terms are the probability-weighted averages of the one-step contingent forecasts. We can rewrite Davidson's recursion slightly by adapting it to our case as:

$$\hat{r}_{t+F} = \sum_{j_F=1}^{K} \hat{P}_{F,j_F}[\mu_{j_F} + \sum_{f=1}^{a} a_{m,j_F}\hat{r}_{F-f} + \sum_{f=F}^{F+a} a_{m,j_F}\hat{r}_{F-f}] \tag{28}$$

$$+ \sum_{f=min(b,F)}^{b} b_{m,j_F}\hat{\epsilon}_{F-f}, \tag{29}$$

where $\hat{P}_{f,j} = Pr(S_f = j|\Phi_t)$ and is generated from

$$\hat{P}_{f,j} = \sum_{i=1}^{K} P_{ji}\hat{P}_{f-1,i}, \quad \text{for} \quad j = 1, ..., K \quad \text{and} \quad f = 1, 2, ..., F.$$

For a proof see Davidson (2004, p.3-4).

As a penalty function we use the mean absolute and the mean squared errors. We apply Eq.(28) to different forecasting horizons $F = 1, 5, 10, 20, 25$

[16] We performed estimations of the models only using sub-samples. It turned out that such an assumption appears to be justified.

and show their relative performances with respect to the linear model in Table 5.

Table 5: Point forecast comparison

Panel A	Only constant changes				
Criterion	1	5	10	20	25
Absolute	0.611	2.140	3.379	4.367	3.932
Squared	1.119	3.555	5.932	6.996	6.893
S_1	0.021	0.002	0.006	0.038	0.099
S_1^{adj}	0.021	0.002	0.007	0.044	0.110
S_2	0.045	0.001	0.000	0.000	0.001
S_{2a}	0.049	0.001	0.000	0.000	0.001
S_{3a}	0.025	0.000	0.000	0.000	0.000
Panel B	All mean equation parameters change				
Absolute	-26.524	-9.420	-3.215	1.528	2.974
Squared	-60.996	-21.229	-8.390	3.347	5.644
S_1	1.000	1.000	0.922	0.287	0.185
S_1^{adj}	1.000	1.000	0.918	0.294	0.197
S_2	1.000	0.999	0.649	0.009	0.097
S_{2a}	1.000	0.999	0.665	0.010	0.105
S_{3a}	1.000	0.999	0.929	0.087	0.013

Note: Values are the improvements in the forecasting performance of the Markov-switching model compared to the linear model. We compare the ARMA(1,1)-GARCH(2,1) linear model with the MS-ARMA(1,1)-C-GARCH(2,1) model starting the forecasts at $t = 1800$ which corresponds to ca. the 75^{th} percentile of the sample. All test statistics but $S1_{adj}$ are as in Diebold and Mariano (1995). S_1^{adj} is the adjusted $S1$ test statistic from Diebold and Mariano (1995) as suggested in Harvey et al. (1997). Test statistics are based on the absolute forecasting error.

In order to compare the forecasting accuracy of the linear and the Markov-switching model more formally we perform the statistical tests proposed in Diebold and Mariano (1995). They develop different test statistic allowing to compare forecasts from two competing models against each other. Such a comparison is based on a loss function $g(\cdot)$ that can take a variety of forms. We opt for an absolute forecasting error loss function. The Diebold and Mariano procedure tests the null hypothesis of equality of the two competing forecasts against the alternative that one forecasting model outperforms the other in its forecasting accuracy. In equation form the null hypothesis may be written as:

$$E[g(e_{it})] = E[g(e_{jt})], \quad \text{or} \quad E(d_t) = 0,$$

where e_{it} is model i's forecasting error and $d_t \equiv [g(e_{it}) - g(e_{jt})]$ is the loss differential. They propose different test statistics one of them being an asymptotic test, which they call S_1:

$$S_1 = \frac{\bar{d}}{\sqrt{\widehat{Var}_{\bar{d}}/T_f}},$$

where

$$\bar{d} = \frac{1}{T_f} \sum_{f=1}^{T_f} d_t \qquad (30)$$

$$Var_{\bar{d}} = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j, \quad \gamma_j = cov(d_t, d_{t-j}) \qquad (31)$$

and where $\widehat{Var}_{\bar{d}}$ is a consistent estimate of the asymptotic variance of $\sqrt{T}\bar{d}$ as proposed in Diebold and Mariano. The infinite sum of covariances in Equation (30) is difficult to estimate. (Diebold and Mariano, 1995, p.254) state that "optimal k-step ahead forecast errors are at most (k-1)-dependent...(k-1)-dependence implies that only (k-1) sample autocovariances need to be used..." They further show that $S_1 \overset{a}{\sim} N(0,1)$.[17]

One might argue if an asymptotic test is applicable to our data. So, we also calculate the "finite-sample tests" proposed by (Diebold and Mariano, 1995), which we will show below:

$$S_2 = \sum_{t=1}^{T} I_+(d_t),$$

[17] We also report the test statistic $S_1^{adj} = \left[\frac{T_f + 1 - 2F + T_f^{-1}F(F-1)}{T_f}\right]^{1/2} S_1$ as suggested in Harvey et al. (1997) to account for finite sample bias and heavy-tailed error distributions. Results do not differ much from S_1.

where

$$I_+(d_t) = 1 \quad \text{if} \quad d_t > 1$$
$$= 0 \quad \text{otherwise.}$$

S_2 may be assessed using the cumulative binomial distribution with a success probability of $p = 0.5$ under the null. In large samples another version of the S_2 sign test is:

$$S_{2a} = \frac{S_2 - 0.5T}{\sqrt{0.25T}} \overset{a}{\sim} N(0,1).$$

The last test statistic we will use is based on a rank-test and is also standard normally distributed under the null:

$$S_{3a} = \frac{S_3 - \frac{T(T-1)}{4}}{\sqrt{\frac{T(T+1)(2T+1)}{24}}} \overset{a}{\sim} N(0,1),$$

where

$$S_3 = \sum_{t=1}^{T} I_+(d_t) rank(|d_t|).$$

Again in Table 5 we show the results of forecasting comparison between the linear and the non-linear Markov-switching model. It is apparent that in the weekly dataset the Markov-switching model, where only the constant term in the mean equation changes, outperforms the linear alternative significantly at any forecasting horizon considered. It is interesting but not surprising to see that the forecasting accuracy of the Markov-switching compared to the linear model improves the longer the forecasting horizon. Such a behavior was to be expected because the change in the absolute difference in the intercepts between the low and the high volatility state is not large and the processes need some time after state-switches to "burn-in" towards the new unconditional volatility level.

When the Markov-switching model, where all mean parameters are free to change, is the competing one, we can see that the linear model forecasts better at short horizons and marginally worse at longer horizons. The better forecasting performance of the MS model at longer horizons is at most small and not very significant. There are a couple of possible explanations for such an outcome. One explanation might be that the differences between the linear and the non-linear MS model are not very large not leading to any significant improvements. Another reason can be that the process does not remain long enough in one regime or another in order to take full advantage of the difference in constants across regimes. This would not allow the forecast to burn in towards the respective unconditional mean in order to obtain a better forecast performance. Such a reason might be justified by again having a look at Figure 5, where it is apparent that the average time the process is estimated to stay in one of the two regimes is much shorter than for the MS model which only allows for changes in the intercept of the mean equation.

5 Conclusions

In this paper we propose a new non-linear volatility model based upon the log-range being defined as the log of the spread (the range) between the observed maximum and the minimum price of assets within a trading week. The results of such an analysis are of potential interest for option pricing, hedging decisions, VaR calculations, but also for policy making. We find strong evidence for an underlying and unobservable Markov chain governing the parameters of the ARMA-GARCH specification that fits the log-range data best. We clearly identify two, a high and a low, volatility regimes. Smoothed regime probabilities that are obtained during the estimation of the models also very well coincide with periods of either low or high volatility observed in the data. Periods most likely to show stronger than average volatility correspond to the collapse of the Bretton Woods system, the first and the second oil crisis, to a (surprisingly) lower extend the period around "Black Monday" in October 1987, and the time from 1998 until 2003 with the Russian crisis, the burst of the dot-com bubble and the second Gulf War.

We further find evidence for different volatility dynamics across different volatility regimes. Volatility appears to be more persistent when the average level of it is relatively low, but seems to be less persistent when it is high. Such results confirm those of Gray (1996) and Klaassen (2002) and hint at the fact that asset market participants act differently during normal versus very volatile periods. In high volatility periods they seem to "forget" quicker than during low volatility periods. Such results have to be interpreted with caution, though, as the dynamics can be governed by another Markov chain as the change in the intercept.

The conditional volatility of the log-range (or the volatility of the volatility) is found to be described well by a GARCH structure with strong persistence, which is very robust over all different models considered. Such a fact means that shocks to the volatility of the volatility in the S&P500 stock index tend to be still present in the market many periods after they happened.

A comparison across different asset markets revealed pronounced comovements in the volatility regimes, especially among stock markets like the S&P500 and the FTSE100. These concordances in volatility also show an increasing tendency over the sample period suggesting the conclusion that there are either more pronounced global factors affecting all markets at the same time or that asset markets have become more integrated. Obviously this does not preclude a combination of both.

A forecasting comparison between a linear model and the proposed Markov-switching models shows promising results. Whereas the Markov-switching model allowing all mean equation parameters to change performs only marginally better at longer horizons than the linear model, we find that the Markov-switching specification only allowing the constant term to change with the regime performs significantly better than the linear competitor at all horizons considered.

Much remains to be done in the area of volatility estimation and forecasting. Our model combining nonparametric volatility estimation with parametric Markov-switching time series methods is not the end of the story. Some very interesting extensions of our model might include the possibility of more than two volatility regimes. The transition probabilities between regimes do not need to be constant either, but can be specified to be dependent on exogenous variables. We might also want to allow for more than one Markov chain governing different aspects of the model. Another very interesting extension of our model would be to check if the forecasting performance of the

Markov-switching model may be improved by assuming that the unconditional mean of the volatility changes with regimes and not only the constant term. Such a behavior would cause the forecasts to move much quicker to the new mean of the volatility corresponding to the respective regime the process is forecast to be in. We are working on some of these extensions and it will be interesting to see to what extend they might improve the estimation and forecasting of asset market volatility.

References

Alizadeh, S., Brandt, M.W., and Diebold, F.X. (2002), 'Range-based estimation of stochastic volatility models', *The Journal of Finance*, 57(3), pp. 1047–1091.

Andersen, T.G. and Bollerslev, T. (1997), 'Heterogeneous information arrivals and return volatility dynamics: Uncovering the long run in high frequency returns', *Journal of Finance*, 52, pp. 975–1005.

Andersen, T.G. and Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review*, 39, pp. 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica*, 71, pp. 579–625.

Ball, C.A. and Torous, W.N. (1984), 'The maximum likelihood estimation of security price volatility: Theory, evidence, and application to option pricing', *Journal of Business*, 57, pp. 97–112.

Barndorff-Nielsen, O.E. and Shepard, N. (2001), 'Econometric analysis of realized volatility and its use un estimating stochastic volatility models', *Journal of the Royal Statistical Society, Series B*, 64, pp. 253–280.

Barndorff-Nielsen, O.E. and Shepard, N. (2003), 'Realized power variation and stochastic volatility', *Bernoulli*, 9, pp. 243–265, (correction, pp.1109–1111).

Barndorff-Nielsen, O.E. and Shepard, N. (2004), 'Power and bipower variation with stochastic volatility and jumps (with discussion)', *Journal of Financial Econometrics*, 2, pp. 1–48.

Beckers, S. (1983), 'Variance of security price returns based on high, low, and closing prices', *Journal of Business*, 56, pp. 97–112.

Bhattacharya, R.N., Gupta, V.K., and Waymire, E. (1983), 'The Hurst effect under trends', *Journal of Applied Probability*, 20, pp. 649–662.

Black, F. and Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy*, 81, pp. 637–654.

Bollen, N.P.B., Gray, S.F., and Whaley, R.E. (2000), 'Regime switching in foreign exchange rates: Evidence from currency option prices', *Journal of Econometrics*, 94, pp. 239–276.

Bollerslev, T. (1986), 'Generalised autoregressive conditional heteroscedasticity', *Journal of Econometrics*, 31, pp. 307–327.

Brandt, M.W. and Diebold, F.X. (2006), 'A no-arbitrage approach to range-based estimation of return covariances and correlations', *Journal of Business*, 79(1), pp. 61–73.

Cai, J. (1994), 'A Markov model of unconditional variance in ARCH', *Journal of Business and Economic Statistics*, 12, pp. 309–316.

Carlson, M. (2007), 'A brief history of the 1987 stock market crash with a discussion of the Federal Reserve response', Staff working paper 2007-13, Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C.

Cheung, Y.W. (2007), 'An empirical model of daily highs and lows', *International Journal of Finance and Economics*, 12, pp. 1–20.

Cheung, Y.W. and Erlandsson, U. (2005), 'Exchange rates and Markov switching dynamics', *Journal of Business and Economic Statistics*, 23(3), pp. 314–320.

Chou, R.Y. (2005), 'Forecasting financial volatilities with extreme values: The conditional autoRegressive range (CARR) model', *Journal of Money Credit and Banking*, 37, pp. 561–582.

Davidson, J. (2004), 'Forecasting Markov-switching dynamic, conditional heteroscedastic precesses', Discussion paper, Cardiff University.

Davidson, J.E.H. (1994), *Stochastic limit theory*, Oxford: Oxford University Press.

Davidson, J.E.H. (2000), *Econometric theory*, Oxford: Blackwell Publishing.

Dickey, D.A. and Fuller, W.A. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, 74, pp. 427–431.

Diebold, F.X. (1986), 'Testing for serial correlation in the presence of ARCH', *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 323–328, washington, DC: American Statistical Association, 1987.

Diebold, F.X. and Inoue, A. (2001), 'Long memory and regime switching', *Journal of Econometrics*, 105, pp. 131–159.

Diebold, F.X., Lee, J.H., and Weinbach, G.C. (1994), 'Regime switching with time-varying transition probabilities', in: C. Hargreaves (Editor), 'Nonstationary Time Series Analysis and Cointegration: Advanced Texts in Econometrics', Oxford and New York: Oxford University Press.

Diebold, F.X. and Mariano, R.S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics*, 13(3), pp. 253–263.

Ederington, L.H. (1979), 'The hedging performance of the new futures market', *The Journal of Finance*, 34(1), pp. 157–170.

Engle, R.F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation', *Econometrica*, 50, pp. 987–1007.

Feller, W. (1951), 'The asymptotic distribution of the range of sums of independent random variables', *The Annales of Mathematical Statistics*, 22, pp. 427–432.

Garcia, R. (1998), 'Asymptotic null distribution of the likelihood ratio test in Markov-switching models', *International Economic Review*, 39, pp. 763–788.

Garman, M.B. and Klass, M.J. (1980), 'On the estimation of price volatility from historical data', *Journal of Business*, 53, pp. 67–78.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006), 'Predicting volatility: Getting the most out of return data sampled at different frequencies', Working Paper.

Gray, S.F. (1995), 'An analysis of conditional regime-switching models', Working paper (Fuqua School of Business, Duke University, Durham, NC).

Gray, S.F. (1996), 'Modeling the conditional distribution of interest rates as a regime-switching process', *Journal of Financial Economics*, 42, pp. 27–62.

Hamilton, J.D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, 57(2), pp. 357–384.

Hamilton, J.D. (1990), 'Analysis of time series subject to changes in regime', *Journal of Econometrics*, 45, pp. 39–70.

Hamilton, J.D. and Susmel, R. (1994), 'Autoregressive conditional heteroskedasticity and changes in regime', *Journal of Econometrics*, 64, pp. 307–333.

Hansen, B.E. (1992), 'The likelihood ratio test under non-standard conditions: Testing the Markov trend model of GNP', *Journal of Applied Econometrics*, 7, pp. S61–S82, supplement.

Hansen, B.E. (1996), 'Erratum: The likelihood ratio test under non-standard conditions: Testing the Markov trend model of GNP', *Journal of Applied Econometrics*, 11, pp. 195–198.

Harvey, D., Leybourne, L., and Newbold, P. (1997), 'Testing the equality of prediction mean squared errors', *International Journal of Forecasting*, 13, pp. 281–291.

Klaassen, F. (2002), 'Improving GARCH volatility forecasts with regime-switching GARCH', *Empirical Economics*, 27, pp. 363–394.

Künsch, H.R. (1986), 'Discrimination between monotonic trends and long-range dependence', *Journal of Applied Probability*, 23, pp. 1025–1030.

Lien, D. and Tse, Y.K. (2002), 'Some recent developments in futures hedging', *Journal of Economic Surveys*, 16(3), pp. 357–396.

Martens, M. and van Dijk, D. (2007), 'Measuring volatility with the realized range', *Journal of Econometrics*, 138, pp. 181–207.

Merton, R.C. (1969), 'Lifetime portfolio selection under uncertainty: The continuous-time case', *Review of Economics and Statistics*, 51, pp. 247–257.

Parkinson, M. (1980), 'The extreme value method for estimating the variance of the rate of return', *Journal of Business*, 53, pp. 61–65.

Rogers, L.C.G. and Satchell, S.E. (1991), 'Estimating variance from high, low, and closing prices', *Annals of Applied Probability*, 1, pp. 504–512.

Sabbatini, M. and Linton, O. (1998), 'A GARCH model of the implied volatility of the Swiss market index from option prices', *International Journal of Forecasting*, 14, pp. 199–213.

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics*, 6(2), pp. 461–464.

Shiller, R. (1989), 'Investor behavior in the October 1987 stock market crash: Survey evidence', in: 'Market Volatility', Boston: Massachusetts Institute of Technology.

Taylor, S.J. (1982), 'Financial returns modeled by the product of two stochastic processes - a study of daily sugar prices 1961-79', in: O.D. Anderson (Editor), 'Time Series Analysis: Theory and Practice', Amsterdam: North Holland, pp. 203–226.

Teverovsky, V. and Taqqu, M.S. (1997), 'Testing for long-range dependence in the presence of shifting means or a slowly-declining trend, using a varaince-type estimator', *Journal of Time Series Analysis*, 18, pp. 279–304.

Yang, D. and Zhang, Q. (2000), 'Drift-independent volatility estimation based on high, low, open, and close prices', *Journal of Business*, 73, pp. 477–491.