

# Estimating and forecasting asset volatility and its volatility: a Markov-switching range model

Jan Piplack <sup>\*†</sup>

*University of Maastricht*

WORKING PAPER. COMMENTS WELCOME

## **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:* ...

---

<sup>\*</sup>University of Maastricht, Department of Economics, PO Box 616, MD 6200, Maastricht, The Netherlands. Tel:+31 433883652, Fax: +31 433884864, Email: j.piplack@algec.unimaas.nl

<sup>†</sup>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. Volatility both 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). Volatility-modeling, therefore, 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 mostly accepted conviction that volatility is both time-varying and predictable (see also for example 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 seem to be affected by very high volatility 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)).

Essentially, 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), Bollerslev (1986), and Taylor (1982)). 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)). 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 where available at all. 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 can be important, for example, in option pricing, where an option trader might want 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 seem to be relatively clear periods of low or 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. We hereby combine 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 range data. First, we find that our model is well able to distinguish a low from a high volatility regime. A second finding is 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 in an absolute and mean square sense.

Our paper proceeds as follows. In Section 3 we describe the methodology used for estimations and also places it within the current literature. Section 4 presents the results of the application of our model to the data of the US S&P500 stock market index. We also present here results of the forecasting comparison exercise we perform. Finally, Section 5 summarizes our results, concludes and sketches directions for future research.

## 2 Theoretical foundations

In this section we will brief on 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. Furthermore we will argue that the range indeed might be modeled in terms of a Markov-switching model.

### 2.1 Theory of the range as a volatility estimator

The range as an estimate for the diffusion constant of stochastic processes like the continuous random walk have quite long history at least 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  $t = 1, \dots, n$  be a sequence of i.i.d. random variables with distribution  $F(u)$  with  $E(u_t) = 0$  and  $Var(u_t) = 1$ . Let then  $S_n = u_0 + \dots u_n$ ,  $M_n = \max(0, S_1, \dots S_n)$ , and  $m_n = \min(0, S_1, \dots S_n)$ . The range is then defined as  $R_n = M_n - m_n$ , which corresponds closely to our definition in section 3 (up to some monotonic transformations). Feller than derives formulas for the mean and the variance of the range  $R_t$ . 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. The argument for the use of the random walk based on the observations that is generally accepted that (at least to a good approximation) the log of stock prices follow a random walk. He further derives a function describing all the moments of the range distribution extending herewith the work of Feller (1951). According to Parkinson it can be written as:

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

where  $\Gamma$  is the gamma function,  $\zeta(x)$  is the Riemann zeta function,  $D$  is again the diffusion constant.  $p$  has to be real and  $\geq 1$ . The first two moments then are found to be:

$$E(R_n) = \sqrt{\frac{8Dn}{\pi}} \quad (1)$$

$$E(R_n^2) = [4\ln(2)]Dn. \quad (2)$$

## 2.2 Extension to non i.i.d. case

One might argue that the assumption of individually identically distributed increments ( $u_t$ ) is not a realistic one because there is clear evidence that the volatility of asset returns is changing over time in a (to some extent) predictable way. So, we want to relax the assumption that  $Var(u_t) = D$  being constant and change it to the assumption that  $Var(u_t) = \sigma_t^2 < \infty$ . Standard results as published in, for example, Davidson (1994) and Davidson (2000) can be used at this point. We will restate some results of the aforementioned author here.

**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  $\bar{\sigma}_n^2 = n^{-1} \sum_{t=1}^n \sigma_t^2$ . 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)  $\frac{\max_{1 \leq t \leq n} \|u_t\|_{2+\delta}}{\bar{\sigma}_n} \leq C < \infty \quad \delta > 0, \forall n \geq 1$

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

Let us also assume that

$$\frac{E(S_n^2)}{n} \rightarrow \sigma^2 < \infty \quad (\text{global wide-sense stationarity}). \quad (3)$$

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 in 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  $\sigma_t$  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. Heuristically speaking then, this theorem provides the basis to reason that the limit distribution of the range as in Equation (1) will stay the same with  $D = \bar{\sigma}_n^2$  even for the non i.i.d. case of the sequence  $u_t$ .

Such a result also covers the case of possible non-linear behavior of the diffusion constant  $D$  as long as Equation (3) is still assumed to hold, which should be reasonable considering the observation that the long-run variance of asset markets seems to not be an integrated process. So, possible Markov-switching behavior in the volatility of asset markets should not change the conclusion that the range is an efficient estimator of the diffusion constant for particular periods  $t$ . Therefore, under the assumption that changes in the diffusion constant  $D$  occur according to a Markov-switching process and that Equation (3) is still satisfied we can very well use the observed range as an estimator for  $D$ .

### 3 Methodology

This section outlines the general methodology proposed and used in this paper. We will introduce the exact estimation and forecasting technique that we apply for our Markov-Switching (MS) Range Model and all 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 data generating processes (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.<sup>1</sup> Hamilton and Susmel argue 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 and transitions between regimes 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:

$$\begin{aligned} u_t &= \sqrt{h(\theta_h, \Phi_{t-1})} v_t \\ &= \sqrt{h_t} v_t, \end{aligned} \tag{4}$$

---

<sup>1</sup>Add some references!

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

$$\begin{aligned} \text{Var}(u_t) &= h_t = f(u_{t-1}, u_{t-2}, \dots) \\ &= \omega + \sum_{i=1}^p \alpha_i u_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \end{aligned} \quad (5)$$

where  $\theta_h$  is a vector of parameters governing the variance equation and  $v_t$  is an i.i.d. sequence with zero mean and unit variance.  $\Phi_{t-1}$  is the information set generated by  $u_t$  and represents the available information set up to time  $t - 1$ . 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) argue that the extension of GARCH processes to the Markov-switching 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  $\Phi_{t-1}$ , depends directly on  $S_t$  but also indirectly on the whole history of regimes ( $S_{t-1}, S_{t-2}, \dots, S_0$ ) because of the path-dependence inherent in regime-switching GARCH models. Regime-dependence in MS-GARCH models comes through the lagged variance and lagged squared error terms. In a GARCH(1,1) model, 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  $S_t, S_{t-1}, \dots, S_0$  or  $\tilde{S}_t$ . In (quasi) maximum likelihood estimations 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 would be  $K^T$  possible regime-path realizations, which would quickly 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  $\tilde{S}_t$  where it emerges namely in Eq.(5) itself. To see this we have now to write Eq.(5) in a regime-dependent form:

$$\begin{aligned}
Var(u_t|\tilde{S}_t, \Phi_{t-1}) &= h_{k,t} \\
&= f(u_{t-1}, u_{t-2}, \dots; \tilde{S}_t) \\
&= \omega_k + \sum_{i=1}^p \alpha_{k,i} v_{t-i}^2 + \sum_{j=1}^q \beta_{k,j} h_{k,t-j},
\end{aligned} \tag{6}$$

where  $Var(u_t|\tilde{S}_t; \Phi_{t-1})$  denotes the variance of  $u_t$  conditional on observable information  $\Phi_{t-1}$  and the unobservable full regime path  $\tilde{S}_t$ . The parameters in the variance equation at time  $t$  are only determined by the current regime  $S_t$ . In Eq.(6) 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. In Hamilton and Susmel (1994) they circumvent the problem of path-dependence by excluding the lagged conditional variance terms  $h_{k,t-1}, \dots, h_{k,t-q}$  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  $\tilde{S}_t$ . In equation form this might be written like:

$$\begin{aligned}
Var(u_t|S_t, \Phi_{t-1}) &= \omega_k + \sum_{i=1}^p \alpha_{k,i} E_{t-2} v_{t-i}^2 \\
&\quad + \sum_{j=1}^q \beta_{k,j} E_{t-j-1} Var(u_{t-j}|S_{t-j}, \Phi_{t-j-1}),
\end{aligned} \tag{7}$$

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  $\Phi_{t-j-1}$ . 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  $v_{t-i}$  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  $t - j - 1$ .

In another paper Klaassen (2002) improves on Gray's (1996) method. Klaassen proposes 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, proposes to rather use the following representation:

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

where the expectation  $E_{t-1}$  again is across regimes  $\tilde{S}_{t-1}$  but now conditional on the information set  $\Phi_{t-1}$  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  $\tilde{S}_{t-1}$  and the path-dependence problem disappears.

In this paper we propose a new way to model the ‘‘observed’’ volatility of time series (in our case we focus on the US stock market) using the famous range estimator for volatility. Parkinson (1980) shows that the range is an effective extreme value theory estimator of the current volatility. In this paper we will use Klaassens (2002) approach and extend his method 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<sup>2</sup>, 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 to model it. This approach has two important advantages. First, we can essentially model the observed volatility and thereby an observed first moment. 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. This is an advantage inherent to the range estimator.<sup>3</sup> 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 version of the presented general model in order to find the best fit to the data. Let  $p_t$  denote the logarithm of the price of some speculative asset or asset index 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 an 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).

Our regime-switching-ARMA-GARCH range model consists of four elements. First there is the mean process that governs the dynamics of the

---

<sup>2</sup>The range obviously is an observed and not conditionally derived volatility measure.

<sup>3</sup>Actually the same would hold for the realized volatility estimator which is also a constructed estimator for volatility and thereby “observable”. Very interesting approaches to model the realized volatility include among others Andersen et al. (2003).

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. 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 regime process is assumed to follow an unobserved Markov chain. We will assume the errors to be i.i.d. standard Gaussian.

The mean of the range has also been modeled by Chou (2005) 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  $\lambda_t$  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 specify the log-range as a stochastic volatility model.

We propose basically a mixture of the above approaches with the additional assumption of an unobservable Markov chain. First 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. As we already shortly motivated above<sup>4</sup>, the data seem to be generated by different regimes. Here, we restrict ourselves to two regimes, namely a low and a high volatility regime. Extensions to more than two regimes are nevertheless possible. Let us start with specifying our mean equation:

$$r_t = \mu_k + \sum_{i=1}^a a_{k,i} r_{t-i} + \sum_{j=1}^b b_{k,j} E_{t-1}[(\epsilon_{t-j} | S_{t-j}, \Phi_{t-j}) | S_t] + \epsilon_t, \quad (9)$$

---

<sup>4</sup>The motivation in the introduction will say something about this point.

where

$$\begin{aligned}
\epsilon_t &= \sqrt{h_{k,t}} z_t \\
h_{k,t} &= \omega_k + \sum_{m=1}^p \alpha_{k,m} E_{t-1}[(\epsilon_{t-m} | S_{t-m}, \Phi_{t-1}) | S_t] \\
&\quad + \sum_{n=1}^q \beta_{k,n} E_{t-1}[(h_{t-n} | S_{t-n}, \Phi_{t-1}) | S_t].
\end{aligned} \tag{10}$$

In the mean equation  $\mu_k$  represents the constant term for all different regimes  $k = 1, 2, \dots, K$ ,  $a_{k,i}$  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.(10) we have  $\omega_k$  being the constant term of the variance equation,  $\alpha_{k,m}$  and  $\beta_{k,n}$  being 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<sup>5</sup>

$$p(S_t = j | S_{t-1} = i, S_{t-2} = k, \dots, \Phi_{t-1}) = p(S_t = j | S_{t-1} = i) = p_{ij}, \tag{11}$$

for  $i, j = 1, 2, \dots, K$ . 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:

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

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

---

<sup>5</sup>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).

### 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 only the information set available to the econometrician 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  $\Phi_T$  at the estimated coefficient vector  $\theta$ , 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  $t$ . The smoothed regime probabilities 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, as in our application in fact observe an estimator of the variance, are more focussing on the mean equation of course not neglecting the variance of the process. Above in Eq.(9) and (10) 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 essentially 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:

$$\begin{aligned} f(r_t|\Phi_{t-1}) &= \sum_{k=1}^K f(r_t, S_t = k|\Phi_{t-1}) \\ &= \sum_{k=1}^K f(r_t|S_t = k, \Phi_{t-1})p(S_t = k|\Phi_{t-1}), \end{aligned} \quad (12)$$

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:

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

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.(9) we can write:

$$f(r_t|S_t = k, \Phi_{t-1}) = \frac{1}{\sqrt{2\pi h_{k,t}}} \exp \left\{ \frac{-\epsilon_{k,t}^2}{2h_{k,t}} \right\}. \quad (13)$$

In the empirical implementation the residuals are relatively close to normality<sup>6</sup> but a Jarque-Bera test still formally rejects it. In general, the errors can for example also be assumed to follow a student-t distribution obviously changing (13) correspondingly.

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

---

<sup>6</sup>See Section 4.

at time  $t - 1$  one can write the ex-ante probability as:

$$\begin{cases} p(S_t = 1|\Phi_{t-1}) = \sum_{k=1}^K p(S_t = 1|S_{t-1} = k, \Phi_{t-1})p(S_{t-1} = k|\Phi_{t-1}), \\ p(S_t = 2|\Phi_{t-1}) = \sum_{k=1}^K p(S_t = 2|S_{t-1} = k, \Phi_{t-1})p(S_{t-1} = k|\Phi_{t-1}), \\ \vdots \\ p(S_t = K|\Phi_{t-1}) = \sum_{k=1}^K p(S_t = K|S_{t-1} = k, \Phi_{t-1})p(S_{t-1} = k|\Phi_{t-1}), \end{cases} \quad (14)$$

where, according to the Markov property,

$$p(S_t = j|S_{t-1} = i, \Phi_{t-1}) = p(S_t = j|S_{t-1} = i) = p_{ij}. \quad (15)$$

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.(11) and are summarized in the transition matrix  $\mathbf{P}$ .

Further note that the second part on the right hand side of Eq.(14),  $p(S_{t-1} = k|\Phi_{t-1})$  we can write, according to Bayes' Rule, as:

$$\begin{aligned} p(S_{t-1} = k|\Phi_{t-1}) &= p(S_{t-1}|r_{t-1}, \Phi_{t-2}) \\ &= \frac{p(r_{t-1}|S_{t-1}, \Phi_{t-2})p(S_{t-1}|\Phi_{t-2})}{p(r_{t-1}|\Phi_{t-2})} \\ &= \frac{p(r_{t-1}|S_{t-1}, \Phi_{t-2}) \sum_{S_{t-2}=1}^K p(S_{t-2}|\Phi_{t-2})p(S_{t-1}|S_{t-2})}{p(r_{t-1}|\Phi_{t-2})}. \end{aligned} \quad (16)$$

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:

$$\mathbf{L} = \sum_{t=\max(a,b,p,q)}^T f(r_t|S_t)p(S_t|\Phi_{t-1}) \quad (17)$$

where the first part on the right hand side is the conditional density at time  $t$  given in Eq.(13), when normality is assumed, and where the second part is the ex ante regime probability described in Eq.(14). 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.(9) and (10) 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 mechanism 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  $\Phi_{t-1}$ . It can be stated in the following way:

$$p(S_{t-1}|S_t, \Phi_{t-1}) = \frac{p(S_{t-1}|\Phi_{t-1})p(S_t|S_{t-1}, \Phi_{t-1})}{p(S_t|\Phi_{t-1})} \quad (18)$$

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

where  $p(S_{t-1}|\Phi_{t-1})$  is given by Eq.(16),  $p_{ij}$  are the fixed transition probabilities in Eq.(11), and  $p(S_t|\Phi_{t-1})$  is given by Eq.(14). If one wishes to estimate models with a lag structure  $max = max(b, p, q) > 1$  one obviously needs the corresponding probabilities ( $p(S_{t-2}|S_t, \Phi_{t-1}), p(S_{t-3}|S_t, \Phi_{t-1}), \dots, p(S_{t-max}|S_t, \Phi_{t-1})$ ), 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.(18). This completes the description of the estimation procedure.

### 3.4 Smoothed regime inference

In this subsection we brief the method of calculating the reported smoothed regime probabilities in Section 4 of this paper. As mentioned above, these

probabilities represent the econometricians best inference about the regime the process was in at time  $t$  using all available information up to time  $T$ . This section heavily draws on results by Hamilton (1989), Hamilton (1990), Gray (1996) and especially Klaassen (2002). In general, one can write  $p(S_t|\Phi_\tau)$  for all  $K$  regimes the ex post probability as:

$$\begin{aligned} p(S_t|\Phi_\tau) &= p(S_t|r_\tau, \Phi_{\tau-1}) \\ &= \frac{p(r_\tau|S_t, \Phi_{\tau-1})p(S_t|\Phi_{\tau-1})}{\sum_{S_{t-1}}^K p(r_{\tau-1}|S_t, \Phi_{\tau-1})p(S_t|\Phi_{\tau-1})}. \end{aligned} \quad (20)$$

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 ( $\tau = t+1, t+2, \dots, T$ ), 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.(20). 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  $p(r_{\tau-1}|S_t, \Phi_{\tau-1})$  for all  $K$  possible regime outcomes. In order to arrive at this density we have to go through some steps. First we can write is as:

$$p(S_\tau|r_t, \Phi_{\tau-1}) = \sum_{S_\tau=1}^K p(r_\tau|S_\tau, \Phi_{\tau-1})p(S_\tau|S_t, \Phi_{\tau-1}), \quad (21)$$

where one uses that the conditional distribution of  $r_\tau$  given  $S_\tau$  does not depend on the earlier regimes ( $S_t, S_{t-1}, \dots$ ) because we integrate out the path-dependence during the estimation procedure. In Eq.(21) we again have two parts on the right hand side. The first one is 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:

$$\begin{aligned}
p(S_\tau|S_t, \Phi_{\tau-1}) &= \sum_{S_{\tau-1}=1}^K p(S_\tau|S_{\tau-1}, \Phi_{\tau-1})p(S_{\tau-1}|S_t, \Phi_{\tau-1}) \\
&= \sum_{S_{\tau-1}=1}^K p_{ij}p(S_{\tau-1}|S_t), \tag{22}
\end{aligned}$$

where we again, as a first ingredient, have the one period ahead  $p_{ij}$  transition probabilities following from Eq.(11) or the transition matrix  $\mathbf{P}$ . The second part on the right hand side on Eq.(22) can be calculated recursively.

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

$$\begin{aligned}
p(S_{\tau-1}|S_t, \Phi_{\tau-1}) &= p(S_{\tau-1}|S_t, r_{\tau-1}, \Phi_{\tau-2}) \\
&= \frac{p(r_{\tau-1}|S_{\tau-1}, \Phi_{\tau-2})p(S_{\tau-1}|S_t, \Phi_{\tau-2})}{\sum_{S_{\tau-1}=1}^K p(r_{\tau-1}|S_{t_{au}-1}, \Phi_{\tau-2})p(S_{\tau-1}|S_t, \Phi_{\tau-2})}, \tag{23}
\end{aligned}$$

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  $\tau$  all ingredients in Eq.(23) 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 are going to present the results of fitting our model in Eq.(9) and (10) to stock market index range data. We first 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 best fits the data. 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 downloaded from the *yahoo.com* database. In order to arrive at the actual data we transformed the downloaded  $p_t^{Max}$  and  $p_t^{Min}$  being the highest and the lowest (log-)price index observation, respectively, like:

$$R_t = 100 * (p_t^{Max} - p_t^{Min}). \quad (24)$$

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<sup>7</sup>:

$$r_t = \ln(R_t), \quad (25)$$

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.

---

<sup>7</sup>In fact we use an outlier-adjusted version of the data series. We consider all realizations as outliers when the weekly range is either larger than 10% (8 cases) or smaller than 1% (30 cases). 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 but do not bias estimation results. A robustness check showed no significant changes (besides larger Jarque Bera test statistics) in estimation results, which are available upon request.

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 four series. So, there is no need for taking the first difference of the data. We can directly apply standard stationary time series analysis tools.

[Insert Table 1]

We show the range  $R_t$  and the log-range  $r_t$  time series in Figure 1. The data start on January 2<sup>nd</sup> 1962 and end on March 19<sup>th</sup> summing to 2359 observations in total (observations are on Mondays). The unconditional distributions of the range and log-range are shown in Figures 2. Obviously, the fact that  $r_t$  is close to normally distributed makes  $R_t$  appear to have a log-normal distribution.

[Insert Figure 1]

[Insert Figure 2]

We continue the data description with an informal time series analysis by having a closer look at the data. One can see in Figure 1 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, in 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 peak corresponding to “Black Monday” on October 19<sup>th</sup> 1987. On this day the main US stock markets dropped by ca.23% starting a period with extreme uncertainty in asset markets worldwide.<sup>8</sup> 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 burst of the “dot-com-bubble” and very much lasting until 2003 roughly corresponding to the end of the third Gulf War.

So, in sum 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 an 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 Markov-switching models only allowing for a change in the intercept of the

---

<sup>8</sup>See, for example, Shiller (1989) and Carlson (2007).

mean equation.<sup>9</sup> We indeed find significant results for the presence of at 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.(9) and (10).

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.(9) and (10), which are reproduced here for convenience:

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

---

<sup>9</sup>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 increased computing time without giving more insights.

where

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

So, 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) specifications 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 or foreseeability 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.(9) and (10) 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,  $\mu_k$ , in the mean equation

(Eq.(9)) 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. All considered estimation results we show 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. Constant terms in either regime 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.

[Insert Table 2]

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. 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.<sup>10</sup> 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.<sup>11</sup> We therefore consider the MS(2)-ARMA-C(1,1)-GARCH(2,1) model as fitting the data best. The fitted values of this model specification can be found in Figure 3. By inspecting the Jarque-Bera test statistic it is apparent that the normality assumption is very likely to be violated, though.

[Insert Figure 3]

By having a closer look at the coefficients of the MS-ARMA-C(1,1)-GARCH(2,1) model, one can see a quite clear difference in the constant terms of either regime. In the low volatility regime  $\mu_1$  is equal to 0.0545 whereas in the high volatility regime  $\mu_2$  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. The variance parameters  $\alpha_1$ ,  $\alpha_2$ , and  $\beta_1$  and the fact that they add up to 0.9988 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

---

<sup>10</sup>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.

<sup>11</sup>Because of space considerations we do not report all those test statistics in Table 2. They are available on request.

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.

[Insert Figure 4]

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.

#### **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.(9). 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. Again we follow the same approach as in the case where we only let the constant,  $\mu$ , change for identifying the appropriate model.

We present the estimation results in Table 3, the fitted values in Figure 5, and the ex ante and smoothed probabilities of the best fitting model in the corresponding Figure 6. 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 (9) to change, is the ARMA-X(1,1)-GARCH(2,1) specification.

[Insert Table 3]

[Insert Figure 5]

[Insert Figure 6]

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 tried to let all parameters of Eq.(9) and (10) vary with the state. Results were inconclusive though, which probably is due to large amounts of coefficients that need to be estimated.<sup>12</sup> 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. A further extension of the model for that possibility could be very interesting but is beyond the purpose of this paper and we leave it for further research along these lines.

### 4.3 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 Markov-switching model with a linear ARMA-GARCH specification. We estimate the best fitting linear<sup>13</sup> and the best fitting MS 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 so 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<sup>14</sup>. 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

---

<sup>12</sup>Results are available upon request.

<sup>13</sup>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.

<sup>14</sup>We performed estimations of the models only using sub-samples. It turned out that such an assumption appears to be justified.

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:

$$\hat{r}_{t+F} = \sum_{j_F=1}^K \hat{P}_{F,j_F} [\mu_{j_F} + \sum_{f=1}^a a_{m,j_F} \hat{r}_{F-f} + \sum_{f=F}^{F+a} a_{m,j_F} \hat{r}_{F-f}] \quad (26)$$

$$+ \sum_{f=\min(b,F)}^b b_{m,j_F} \hat{\epsilon}_{F-f}, \quad (27)$$

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

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

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.(26) to different forecasting horizons  $F = 1, 5, 10, 20, 25$  and show their relative performances with respect to the linear model in Table 4.

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 again 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

$$\bar{d} = \frac{1}{T_f} \sum_{f=1}^{T_f} d_t \tag{28}$$

$$Var_{\bar{d}} = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j, \quad \gamma_j = cov(d_t, d_{t-j}) \tag{29}$$

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 (28) 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  $S_1 \overset{a}{\sim} N(0, 1)$ .

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:

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

where

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

$S_2$  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:

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

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

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

where

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

Again in Table 4 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 6, 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 Concluding remarks

In this paper we propose a new non-linear volatility model based upon the observed volatility estimator range and/or log-range being defined as the spread in percent between the maximum and the minimum observed stock price index of the S&P500 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 quite 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 burst of the dot-com bubble.

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. The conditional volatility of the log-

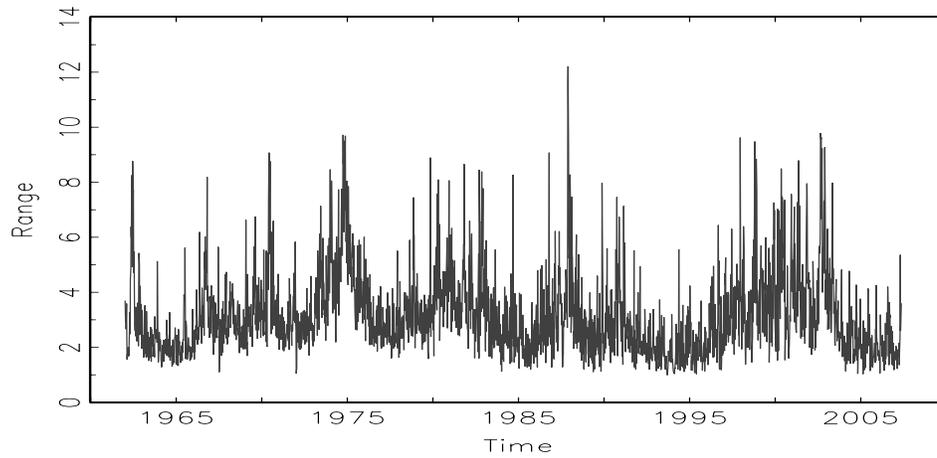
range (or the volatility of the volatility) is found to be described well by a GARCH structure with strong persistence, which is very robust over all different models considered. Such a fact means that shocks to the volatility of the volatility in the S&P500 stock index tend to be still present in the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

A 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.

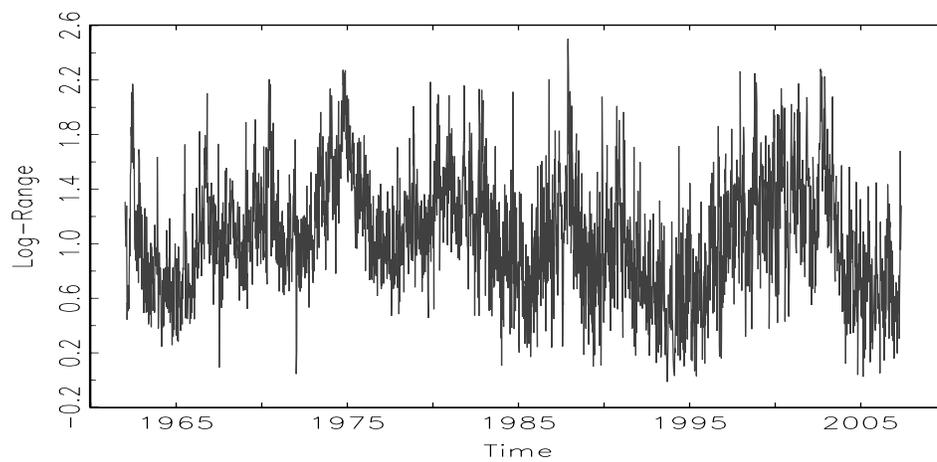
Much remains to be done in the area of volatility estimation and forecasting. Our model combining nonparametric volatility estimation with parametric Markov-switching time series methods is not the end of the story. Some very interesting extensions of our model might include the possibility of more than two volatility regimes. The transition probabilities between regimes need not to be constant either, but can be specified to be dependent on exogenous variables. 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 extent they might improve the estimation and forecasting of asset market volatility.

# Appendix

Figure 1: Weekly S&P500 range and log-range



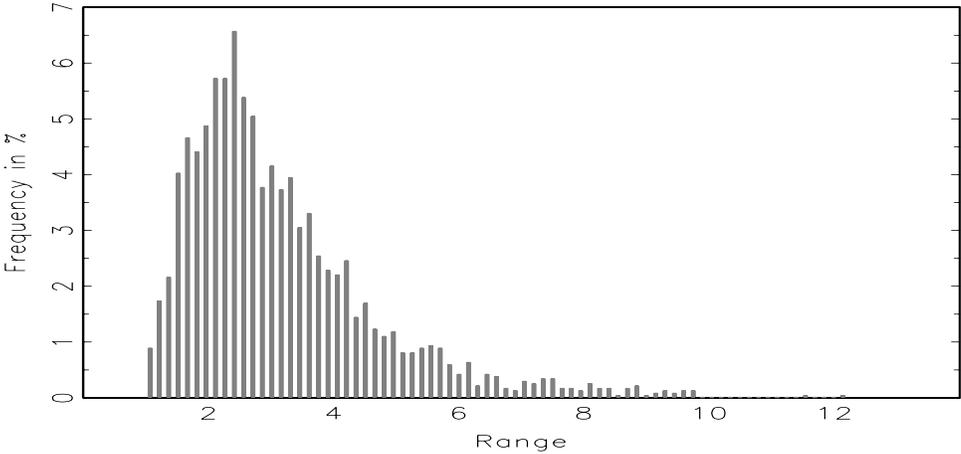
(a) Range



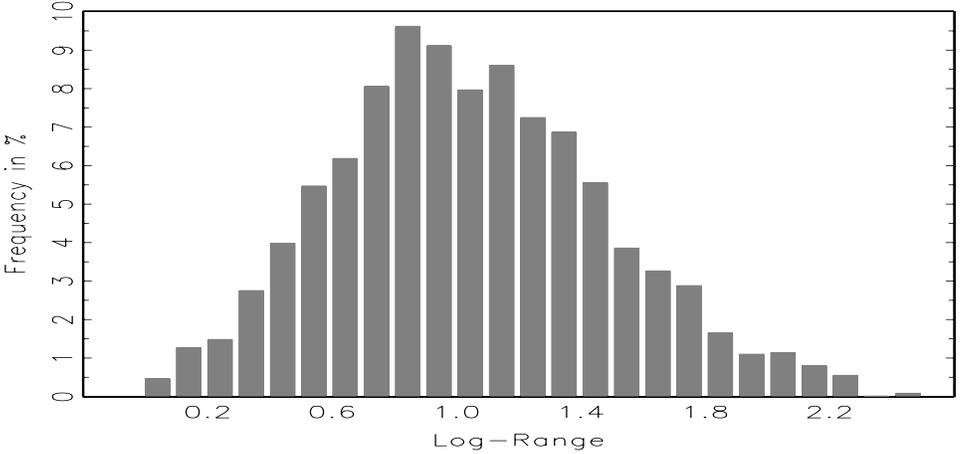
(b) Log-Range

**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 2: Weekly S&P500 range and log-range unconditional distributions



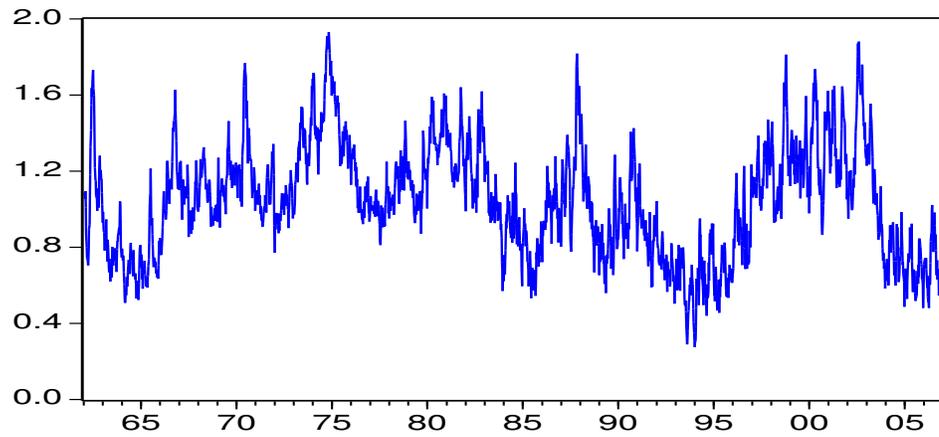
(a) Range



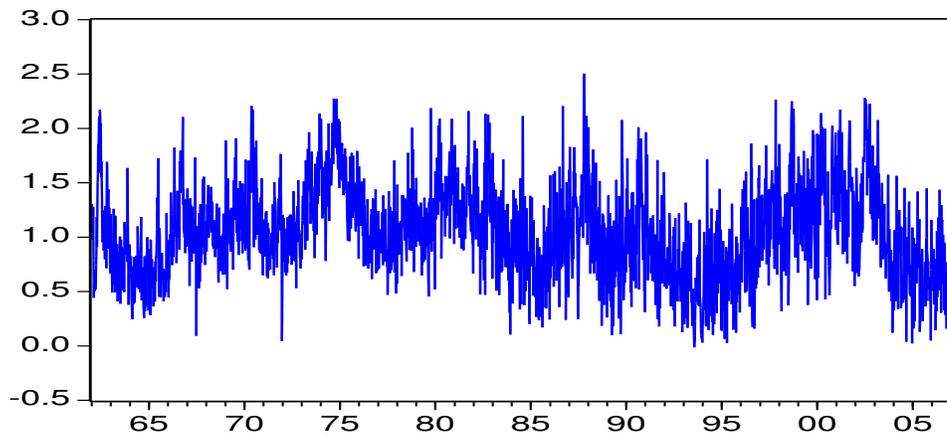
(b) Log-Range

**Note:** The range and log-range are calculated as in Figures ?? and 1.

Figure 3: Fitted and actual weekly S&P500 log-range values (only constant changes)



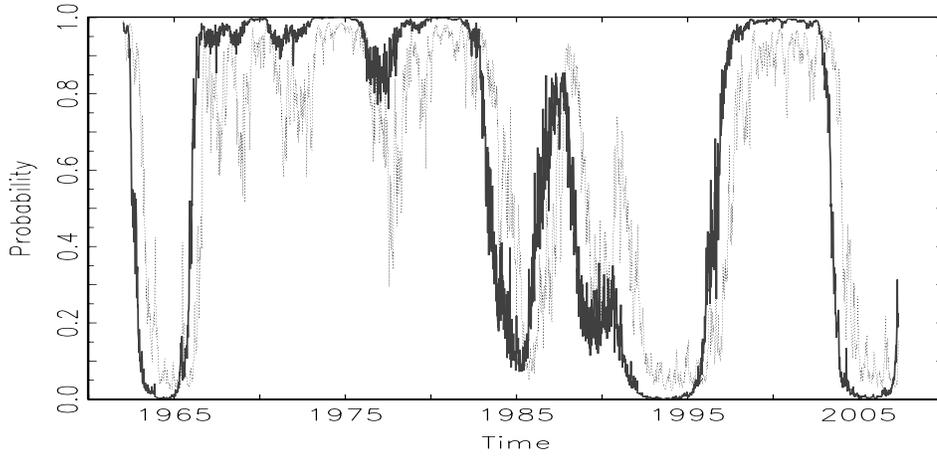
(a) Fitted Values



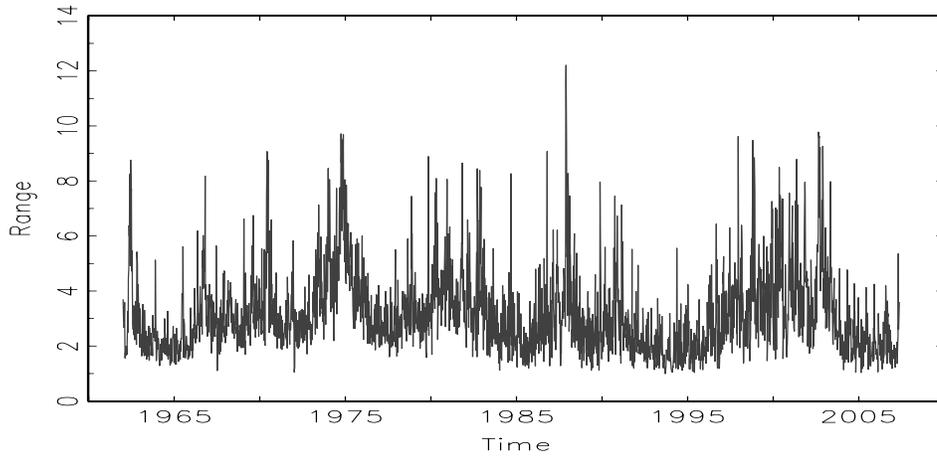
(b) Actual data

**Note:** The fitted values in Panel A are obtained from a MS-ARMA(1,1)-GARCH(2,1) specification. Panel B shows the observed data.

Figure 4: Weekly probabilities (only the constant changes)



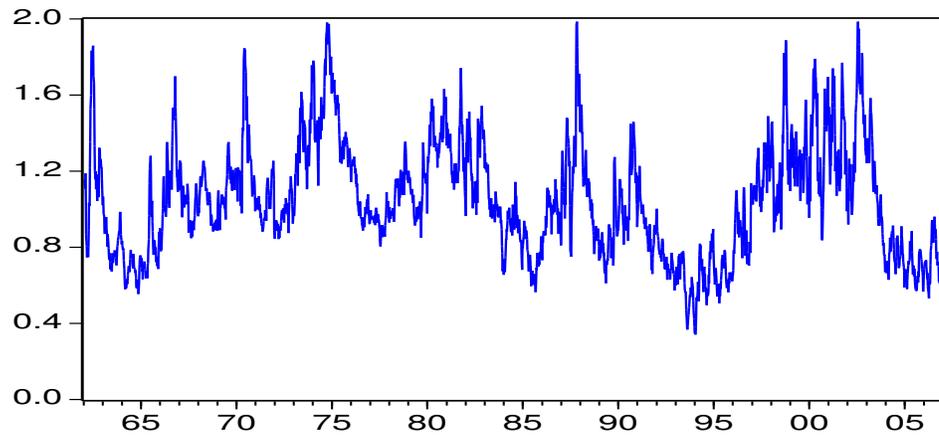
(a) Ex ante and smoothed probabilities



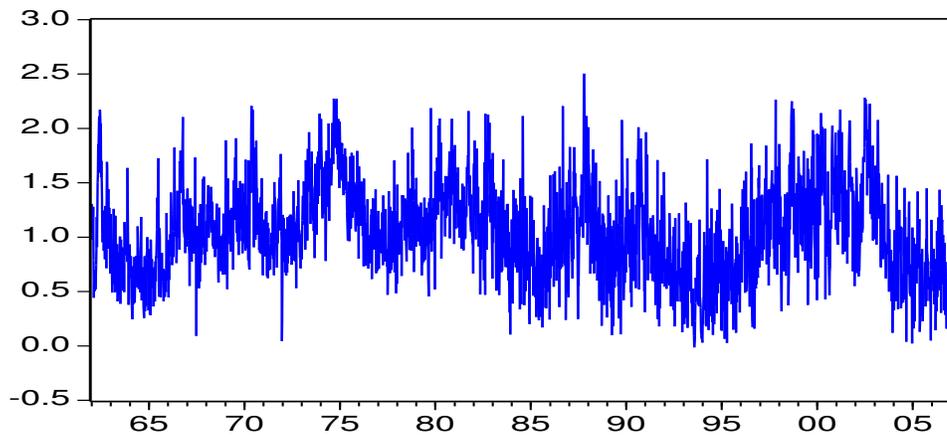
(b) Range

**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.

Figure 5: Fitted and actual weekly S&P500 log-range values (all mean equation parameters change)



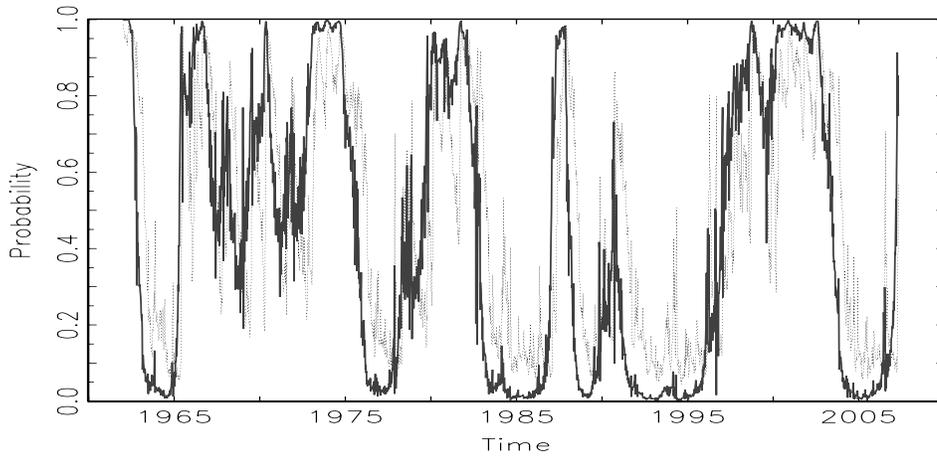
(a) Fitted Values



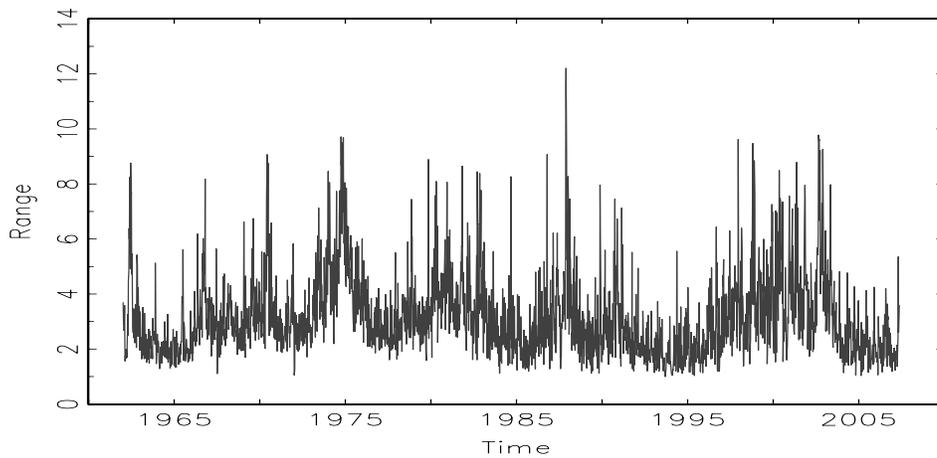
(b) Actual data

**Note:** The fitted values in Panel A are obtained from a MS-ARMA-X(1,1)-GARCH(2,1) specification. Panel B shows the observed data.

Figure 6: Weekly 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.

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.(24) and (25) respectively. Data are from January 2<sup>nd</sup> 1962 until March 19<sup>th</sup> summing to 2359 observations in total. The data are plotted in Fig. 1. Augmented Dickey-Fuller (ADF) test statistics and p-values are calculated based on an automatic lag-length selection using the Schwarz information criterion.

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

Parameters	AM(1,0)		AM(1,1)		AM(1,1)-G(1,0)		AM(1,1)-G(1,1)		AM(1,1)-G(2,1)	
	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.
$\mu_1$	0.5413	0.0004	0.0410	0.0073	0.0589	0.0050	0.0560	0.0053	0.0545	0.0006
$\mu_2$	0.8865	0.0009	0.1741	0.0554	0.0934	0.0114	0.0900	0.0148	0.0876	0.0009
$a_1$	0.3426	0.0003	0.9522	0.0030	0.9236	0.0096	0.9263	0.0123	0.9280	0.0008
$b_1$			-0.7533	0.0046	-0.6956	0.0008	-0.6808	0.0094	-0.6907	0.0033
$\omega$	0.1129	0.0000	0.1051	0.0007	0.1000	0.0318	0.0003	0.0000	0.0001	0.0001
$\alpha_1$					0.0939	0.0971	0.0128	0.0005	0.0892	0.0118
$\alpha_2$									-0.0797	0.0115
$\beta_1$							0.9842	0.0007	0.9893	0.0016
$P11$	0.9899	0.0000	0.9779	0.0196	0.9958	0.0003	0.9964	0.0011	0.9961	0.0004
$P22$	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.(9) and (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.  $LB_x$  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. 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.

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

Parameters	AM(1,0)		AM(1,1)		AM(1,1)-G(1,0)		AM(1,1)-G(1,1)		AM(1,1)-G(2,1)	
	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.	Estimate	Std.Er.
$\mu_1$	0.6000	0.0050	0.0288	0.0483	0.0281	0.0536	0.0227	0.0002	0.0244	0.0012
$\mu_2$	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
$\omega$	0.1130	0.0033	0.1055	0.0006	0.0957	0.0015	0.0002	0.0000	0.0001	0.0000
$\alpha_1$					0.0938	0.0302	0.0110	0.0000	0.0798	0.0103
$\alpha_2$									-0.0710	0.0104
$\beta_1$							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.(9) and (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.  $LB_x$  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. 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.

Table 4: Point forecast comparison Markov-switching vs. linear model

<u>Panel A</u>		Only constant changes			
Criterion	1	5	10	20	25
Absolute	0.584	1.997	3.238	4.076	3.679
Squared	1.073	3.386	5.695	6.544	6.545
$S1$	0.021	0.002	0.007	0.044	0.111
$S2$	0.031	0.001	0.000	0.000	0.000
$S2_a$	0.034	0.001	0.000	0.000	0.000
$S3_a$	0.026	0.000	0.000	0.000	0.000
<u>Panel B</u>		All mean equation parameters change			
Absolute	-26.557	-9.445	-3.215	1.528	2.974
Squared	-61.102	-21.285	-8.215	3.347	5.644
$S1$	1.000	1.000	0.922	0.287	0.185
$S2$	1.000	0.999	0.649	0.009	0.097
$S2_a$	1.000	0.999	0.665	0.010	0.105
$S3_a$	1.000	1.000	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 XY.

## References

- ALIZADEH, S., BRANDT, M., AND DIEBOLD, F. (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 in 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.

- BLACK, F. AND SCHOLLES, M. (1973), 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy*, 81, pp. 637–654.
- 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 & 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. AND ERLANDSSON, U. (2005), 'Exchange Rates and Markov Switching Dynamics', *Journal of Business & 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 processes', 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. AND FULLER, W. (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., 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, X.D. AND MARIANO, R.S. (1995), ‘Comparing Predictive Accuracy’, *Journal of Business & 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.
- KLAASSEN, F. (2002), ‘Improving GARCH Volatility Forecasts with Regime-Switching GARCH’, *Empirical Economics*, 27, pp. 363–394.
- LIEN, D. AND TSE, Y.K. (2002), ‘Some Recent Developments in Futures Hedging’, *Journal of Economic Surveys*, 16(3), pp. 357–396.
- 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.
- 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.
- 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.